ENRE 447/602 Reliability Analysis

Module 1: Course Overview, Reliability Engineering Perspective & Fundamentals

Professor:

Dr. Katrina M. Groth

Mechanical Engineering Department

Associate Director, Center for Risk and Reliability

University of Maryland

kgroth@umd.edu

Co-Instructor:

Dr. Mohammad Modarres

Center for Risk and Reliability

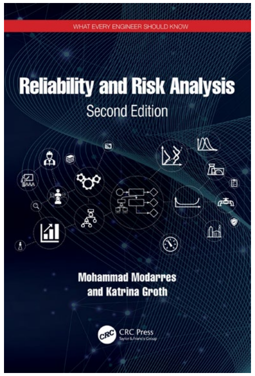
University of Maryland

modarres@umd.edu

Objectives for part 1 of this module

- To introduce ourselves
- To explain how this course is designed
- To set expectations for students and instructors
- To discuss logistics & answer questions about the course

Course Topics (Modules) for ENRE447 & ENRE602



Prof. Katrina M. Groth

Mechanical Engineering,
Center for Risk and Reliability

University of Maryland

By the end of this course, you will understand the fundamental qualitative and quantitative methods for conducting reliability and risk analysis for engineering systems.

Modules

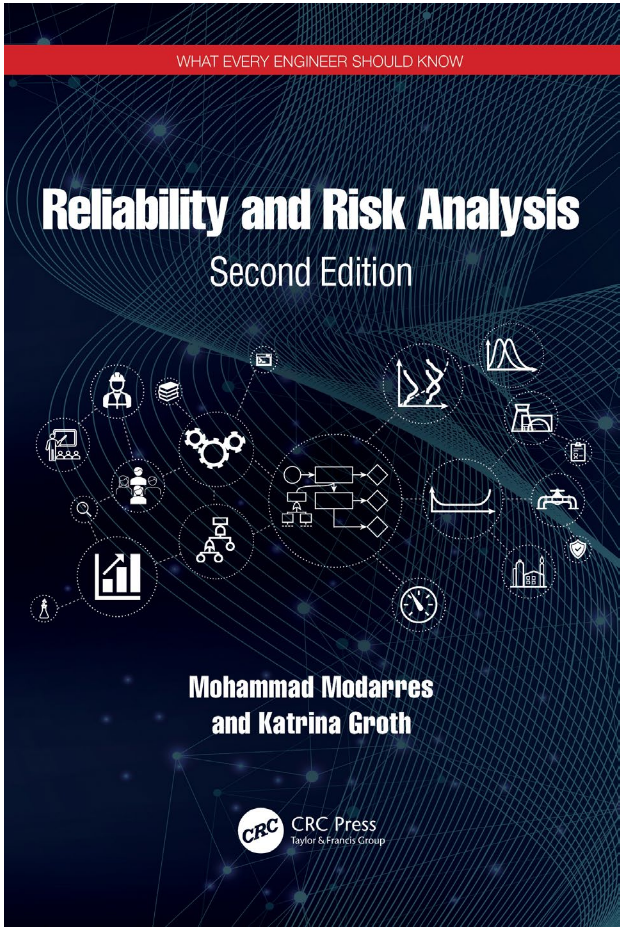
ENRE447: Modules 1-6, 9

1. Reliability Engineering in Perspective
2. Reliability Math: Probability
3. Elements of Component Reliability
4. Reliability Math: Statistics
5. Reliability Data Analysis & Model Selection
6. System Reliability Analysis

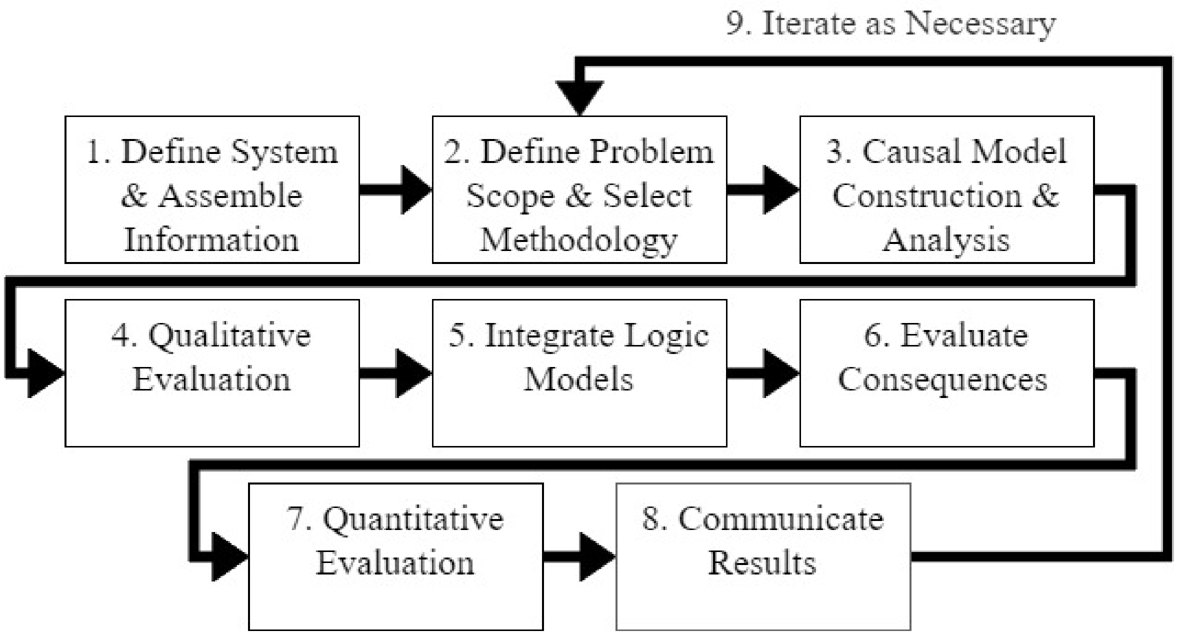
ENRE602: All Modules

7. Reliability and Availability of Repairable Components and Systems
8. Advanced Topics: Uncertainty Analysis, Importance Measures, Human Reliability Analysis, Bayesian Networks, Reliability Research
9. Risk Analysis

Textbook- PRINTED version needed.



Paperback pictured; Hardcover is light blue plain cover



Modarres & Groth, *Risk and Reliability Analysis,* Taylor and Francis, **April 2023**. https://www.taylorfrancis.com/books/mono/10.1201/9781003307495/reliability-risk-analysis-mohammad-modarres-katrina-groth

General Info

- Open the syllabus now. We'll go over it in a few slides.
  - The syllabus is the authoritative source of information about this course. In case of any differences between this presentation and syllabus, the syllabus takes precedence.
- Course format & attendance:
  - Lectured-based with frequent interaction and exercises; On campus & recorded for distance education students (typically asynchronous).
  - On campus students (0101) - attend in person unless you have an excused absence; excused: watch recorded video from this week's lecture. Available on ELMS\Video Lectures.
  - If instructors are unable to attend: class will be asynchronous using previously recorded lecture.
  - Lectures are recorded for future educational purposes. Microphones in this room are always on.
  - Do not distribute course materials or lecture videos to anyone.

Center for Risk and Reliability (CRR)

Home to UMD's Reliability Engineering Program. Location: 0151 Martin Hall

Our Mission

We advance reliability and risk analysis for complex engineering systems through innovative research, education, and collaboration with industry partners.

Our Approach

We *research* why systems fail, how they fail, when they fail, how to prevent failure, and how to mitigate consequences.

We *educate* through coursework, research, and stakeholder engagement. We *engineer* solutions.

Our Impact

We prevent losses and protect life, property, and the environment. Our work improves systems and processes in energy, transportation, defense, space, information systems, and civil infrastructures.

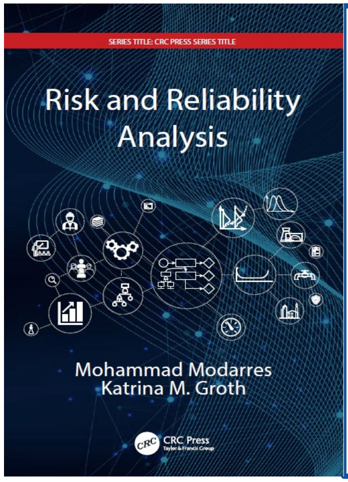
Fast Facts about UMD’s Reliability Engineering program & Center for Risk and Reliability

20+
Core, Affiliate, and Adjunct Faculty

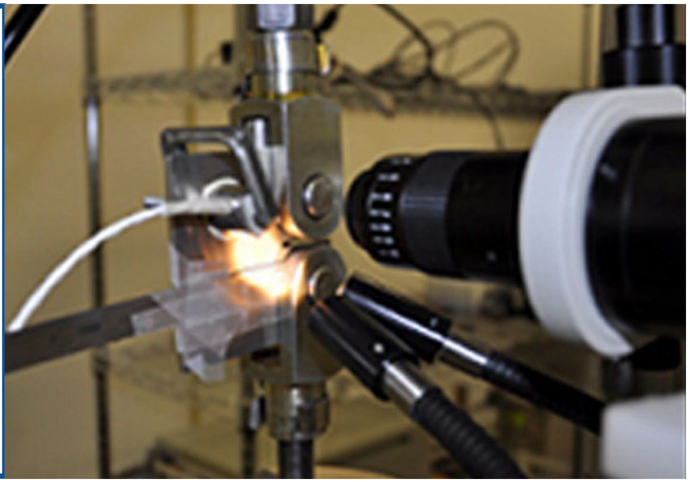
6
Cutting-Edge Research Laboratories

4
Degrees Offered (Ph.D. M.S., M.Eng, Certificate)

500+
Graduates since 1991



- Systems Risk and Reliability Analysis Lab (SyRRA)
- Probabilistic Physics of Failure and Fracture
- Cybersecurity Quantification Lab
- Risk And Decision Analysis Lab (RADA)
- Design Decision Support Lab
- Laboratory for Reliable Nanoelectronics
- Risk-Informed Solutions in Engineering (RISE)



The **#1** Reliability Engineering program in the U.S. (Source: Scopus)

Our alumni are making impact

Government

U.S. FOOD & DRUG ADMINISTRATION
U.S.NRC
Australian Government Defence
Office of Naval Research Science & Technology
NASA
NIH National Institutes of Health

Industry & Tech

BELL
AcuTech
Google
Ford
metro
Capital One
ExxonMobil
amazon
Chevron
FM Global
intel
NORTHROP GRUMMAN
Marathon
SIEMENS
RIVIAN

Academia & Research

JOHNS HOPKINS APPLIED PHYSICS LABORATORY
COLORADO STATE UNIVERSITY
RUTGERS
UNIVERSITY OF MARYLAND
UCLA
CHINA UNIVERSITY OF PETROLEUM
UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Teaching philosophy: why am I a professor?

- To expand the field of reliability engineering via research & students
- Realization that risk & reliability are essential for many disciplines, but that few people have more than one method in their toolbox…e.g. entirely quantitative, entirely qualitative..
- To bring you a spectrum of methods – qualitative & quantitative – to apply to your problems.



Introductions

Instructor & TA

Various administrative contacts:

- M.S. and Ph.D. students: Your advising is through the Mechanical Engineering (and Reliability Engineering) department graduate office (megrad@umd.edu) and proctoring is through DETS.
- M.Eng. and Certificate Students: Your advising (and proctoring) is via the MAGE office mage-advising@umd.edu

- Who are you and why are you here? (Speak loudly! And give me a moment to check my class roster)
  - Ask TA first.

My approach to this class

- The class is an environment, a variety of materials and assignments in multiple formats, which together create a learning experience.
- Math + theory + discussion + engagement + real applications
  - Opportunity to really dig into both the math and the theory and align these
  - Use discussion to provide opportunity for you all to get what you want and need out of this class
- Active learning – Learning is more than listening. You complete problems, some of which will be designed to be hard. You must try (and in some cases fail) now before you are faced with these problems in the field.
- Needs 10-20 hours per week from you.

My approach to this class

- ***You are my product, not my customer.***
  - I have high standards for my product! I want you to become capable engineers and maintain (or increase) the quality and integrity of a UMD engineering degree.
  - I want you all to succeed and have designed the class carefully.
  - You must study & work hard to be a product I can be proud of.
  - As an engineer, I must uphold the fundamentals of the profession; I cannot “release” a faulty product. I don’t give passing grades just for showing up.
- If you are struggling, please reach out and we will figure something out together.

Key elements of syllabus

- Pull up your syllabus and review:
  - Grading
  - Exam dates
  - Schedule
  - Policies and expectations
  - Communication
  - Academic integrity
  - Copyright policy

Grading Rubric – available on ELMS – read it!

- Show me your understanding, not just your answer.
- Most problems will be broken down into multiples of 5 points
- Shows example of good vs. bad solution.

5	**The student clearly understands how to solve the problem and the solution is correct.**
4.75	**The student clearly understands how to solve the problem.** Minor mistakes and careless errors (akin to typos) may have led to an incorrect answer, but they do not indicate a conceptual misunderstanding.
4	**The student understands the main concepts and problem-solving techniques, but has some minor yet non-trivial gaps in their** understanding, reasoning, or explanation.
3	**The student has partially understood the problem.** The student is not completely lost, but requires tutoring in some of the basic concepts. The student may have missed a critical step or aspect of the problem. The student may have started out correctly, but gone on a tangent or may have not finished the problem.
2	**The student has a poor understanding of the problem.** The student may have gone in a not-entirely-wrong but unproductive direction, or attempted to solve the problem using pattern matching or by rote. The student may have shown so little work, or made major errors in notation that the understanding cannot be determined even if the answer is correct. The answer may appear somewhere but cannot be clearly identified.
1	**The student did not understand the problem.** They may have written some appropriate formulas or diagrams, but nothing further. Or, they may have done something entirely wrong.
0	The student wrote nothing or almost nothing.

Other elements of grading: following directions, submitting neat, professional work on time.

Participation / coursework – did your participation enhance the class & the learning experience for the class?

Academic integrity

- Under the Code of Academic Integrity, there are five types of academic dishonesty: **cheating, fabrication, facilitation, plagiarism, and self-plagiarism.**
- **Don't do it.**
  - It hurts all UMD grads and the engineering profession. It devalues your degree and mine.
  - Your academic & professional integrity is more important than any grade or any assignment.
- I will report it. Please report it to me if you find it.
  - Apathy in the presence of academic dishonesty is NOT a neutral act.
  - All of us - students, faculty, and staff - share the responsibility to challenge and make known acts of apparent academic dishonesty.
- *Do you want to get on an airplane designed by someone who cheated to get through engineering school?*

Clarifications on academic dishonesty in this class.

- If you haven't done it: Read the whole Code of Academic Integrity.
- These are NOT a violation:
  - Collaborations on homework is ok! See details on syllabus.
  - Exams – TBD, but they will be either open book/notes OR allow at least 1 page of notes.
  - Asking and answering questions about homework on our course website is ok (within moderation. I.e., not every problem, every week).
  - Using solutions I give you as part of your notes for subsequent open note assignments.
- Examples of violations:
  - Distributing course materials to any person or entity who is not registered in this semester of this course. (Don't upload them to a website, test bank, group chat, etc.)
  - Using any course materials or solutions from a previous semester or any source other than the course website. (Only use solutions that I give you.)
  - Using an outside source without citing it. (Cite any outside sources you use in Chicago or IEEE style.)
  - Changing your answer on a quiz while we are grading it together in class. (Don't!)
  - Copying homework #2 from your friend because you forgot it was due today (Don't! Both of you are cheating now! I drop a homework for a reason!)
  - Using specialized engineering software from work.
  - Using ChatGPT or similar. Using Chegg or similar.

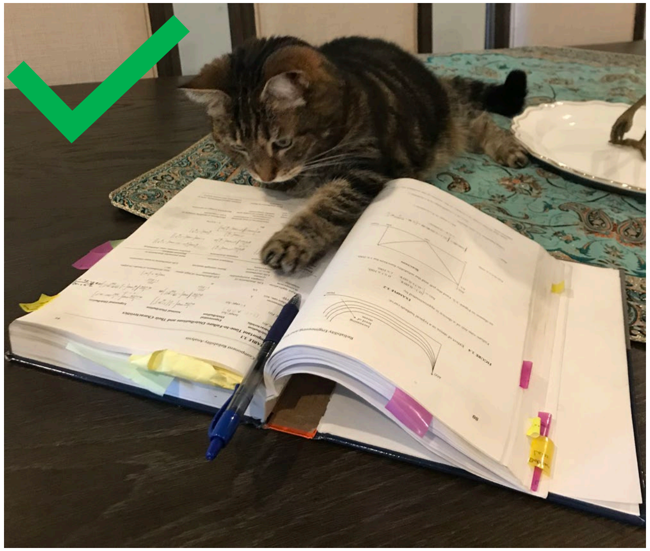
NSPE Code of Ethics: Fundamental Canons

- Engineers, in the fulfillment of their professional duties, shall:

1. Hold paramount the safety, health, and welfare of the public.
2. Perform services only in areas of their competence.
3. Issue public statements only in an objective and truthful manner.
4. Act for each employer or client as faithful agents or trustees.
5. Avoid deceptive acts.
6. Conduct themselves honorably, responsibly, ethically, and lawfully so as to enhance the honor, reputation, and usefulness of the profession.

Tips for success

- **Put in the time outside of class.**
- *Work through* the problems in the textbook & slides. (don't passively read)
- Do the assignments. Work through the solutions after they are released.
- Create topic summaries & reference sheets
- Read the textbook. Read the cited references if you have additional questions about specific topics.
- Connect with each other
- Go to the library & read reliable sources:
  - Reliable: Textbooks, journal articles, government reports, software documentation & some forums
  - Unreliable: blogs, homework "help" sites (aka cheating websites), twitter, or un-verified information from search engines



Let's have a great semester!

- Any questions?

Breakpoint
Begin Reliability Engineering Perspective & Fundamentals

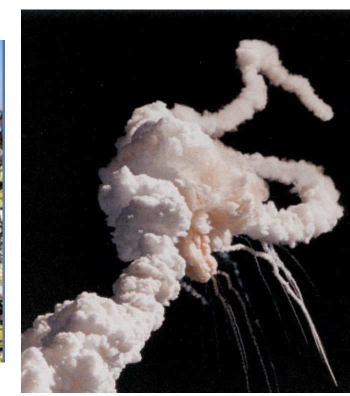
Objectives for this module

- At the end of this module you will be able to:
  - Identify common motivations for studying reliability engineering
  - Identify multiple approaches used in reliability analysis
  - Define key terms: Reliability, risk, availability, maintainability, failure, performance, system component
  - Define and use failure modes, failure mechanisms, and causes to discuss failure scenarios.

Why study risk & reliability?

- The core of U.S. economy, security and quality of life depends on *complex engineering systems (CES)* that range from power plants, energy systems, and pipelines to aircraft, defense, and transportation.
- Engineers create transformative technologies …but the engineering doesn’t end when the product is delivered or the lights come on.
- Systems can be engineered for safety & reliability
- Engineers need insight into how to prevent, mitigate, and recover from system failures, accidents



Things Fail

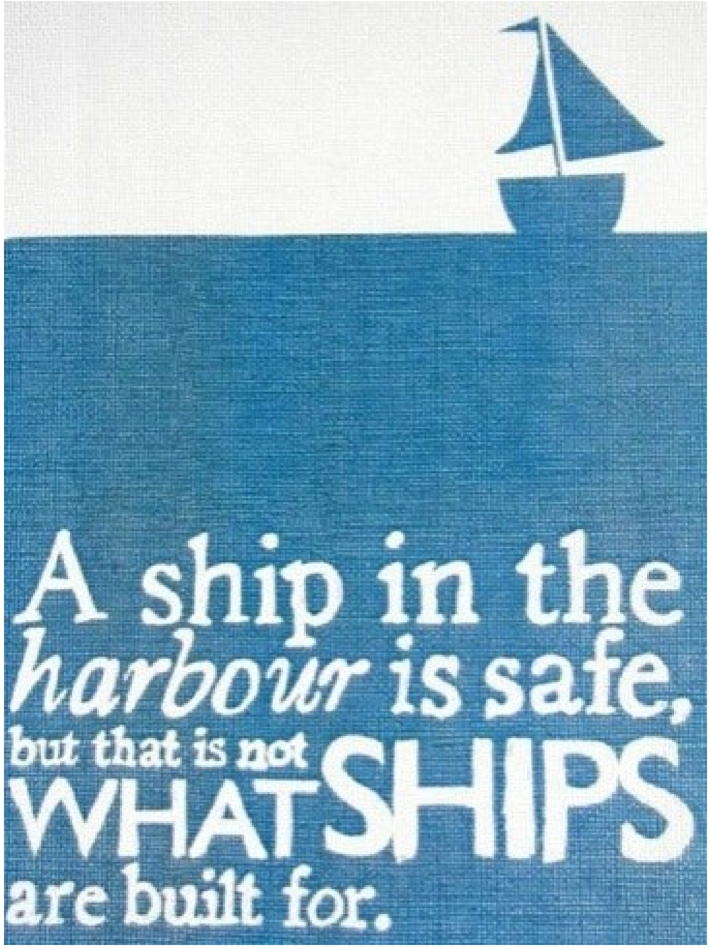
- 2011– Fukushima Daiichi– Earthquake and tsunami induce loss of reactor cooling, resulting in meltdown; radiation release and environmental contamination; evacuation
- 2003 – Space Shuttle Columbia disintegrates on reentry to Earth;
  - Loss of all 7 crew members
  - US space program grounded
  - Small piece of insulation broke off during takeoff –due to aerodynamic forces at takeoff.
- 2009 Air France 447 crashed into Atlantic Ocean en route from Rio de Janeiro to Paris
  - Inaccurate airspeed measurements due to frozen pitot tubes
  - 228 fatalities

…and they keep failing

- Nuclear
  - Three Mile Island
  - Fukushima accident
  - Davis Besse
- Aviation & Aerospace
  - Challenger
  - Columbia
  - Tenerife
- Chemical, oil, and gas
  - Bhopal
  - Piper alpha
  - Deep water horizon
  - Bruno canyon
  - Aliso canyon
  - Texas city
- Buildings & Infrastructure
  - Grenfell fire
  - Triangle shirtwaist factory
  - Rhode Island Station nightclub
- Many less public examples in defense, intelligence

Reliability engineering supports decision-making

- A process **to create knowledge, explore priorities, encourage discourse** and **build a common basis for safety and reliability decisions**
- By creating an understanding of:
  - What the system is supposed to do (performance)
  - Why and how failures occur (e.g., the sources, causes, and likelihood of failures (physics based, human, computational, etc.)
  - How often failures occur (e.g., likelihood)
  - Strategies to reduce failure (e.g., design, operation, maintenance)

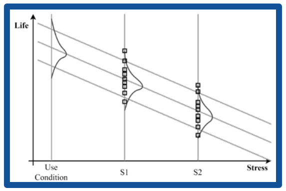


www.curiositiesbydickens.com/wp-content/uploads/ship-in-the-harbor.jpg

History and Evolution of the Field

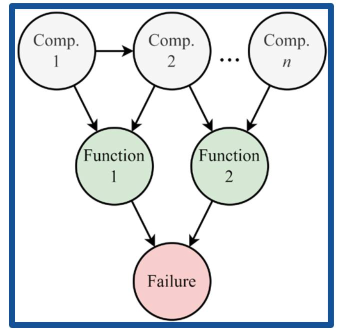
1950s-60s

TTF Distributions
RBDs & FMEA
Physics of Failure



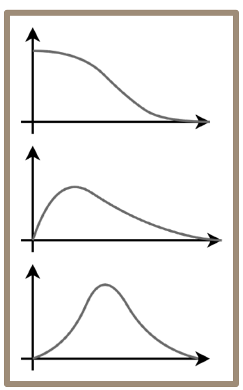
1990s

Probabilistic Physics of Failure
HALT



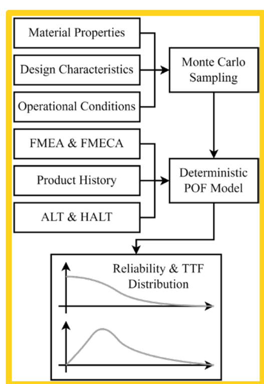
2010s

Bayesian, AI, ML methods
Big Data
PHM & BNs



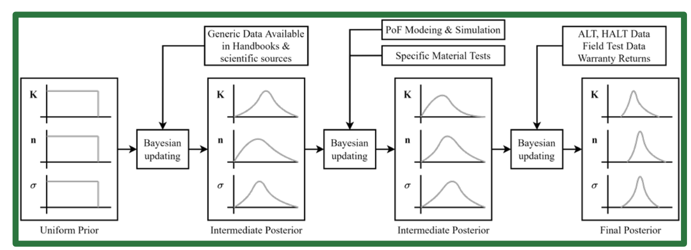
Acc. Life Tests (ALT)
Fault Trees
MLE Methods
Bayesian Estimation

1970s-80s



Hybrid Modeling Approaches (BNs)
MCMC Simulation
CBM & PHM

2000s



Defining Reliability

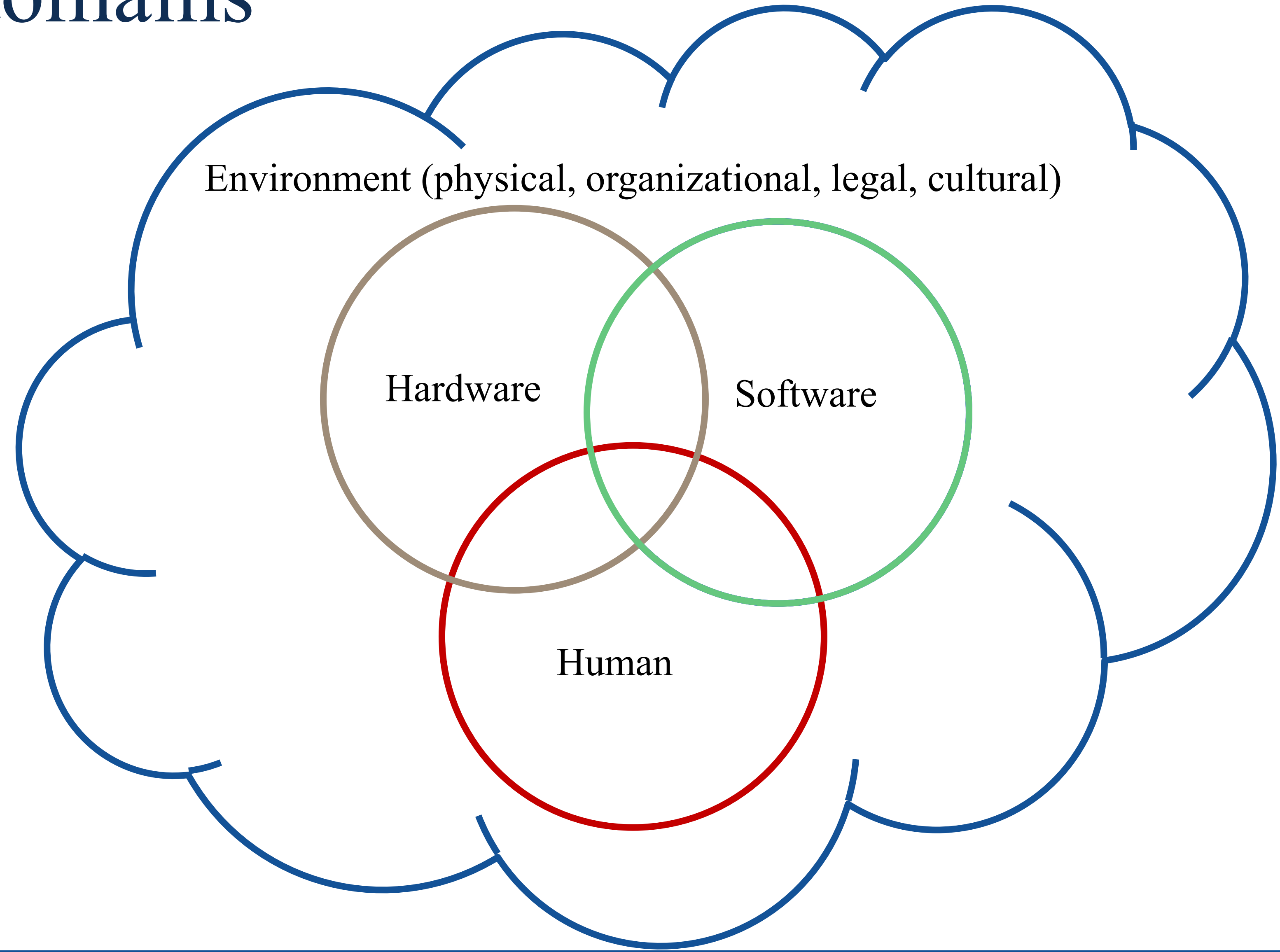
- ***Reliability*** *is the ability of an item to operate without failure under specified operating conditions to attain a mission having the desired length of time or number of cycles.*
- To assess reliability, we must
    1. Define the item (e.g., part, component, subsystem, system, or structure)
    2. Define the expected mission of the item and what constitutes "success" (or failure)
    3. Define the operating conditions and environments of use
    4. Specify mission variable (e.g., time, # cycles, stress)
    5. Assess ability (e.g., through testing, modeling, data collection, analysis)

Reliability Engineering Approaches

Failure / Damage Data from Field or Test → Failure Model (Distribution) or Physics of Failure Model → Estimate Model Parameters & Uncertainties → Predict Performance / Reliability

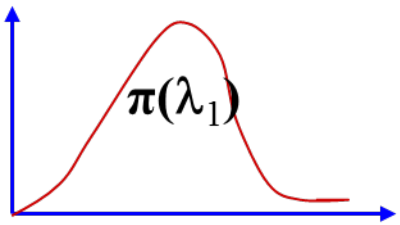
- Deterministic - Understand and model failure mechanisms, physics of failure
  - ENRE600 introduces these in depth.
  - Connects fundamental science and empirical approaches
  - However, it quickly becomes apparent that *predicting failure* inherently involves uncertainty.
- Thus:
- Probabilistic - **Reliability is a *probability***
  - ENRE602 uses the above and adds likelihood (probability) and uncertainty of events.
  - Thus, the probabilistic approach to reliability includes:
    - Understanding failure mechanisms and physics of failure
    - Connecting data and the fundamental sciences (sometimes called a “hybrid” approach.)
    - Using past data to predict future reliability, performance, etc.
    - Using probabilistic and statistical analysis of data
    - Quantifying of uncertainties
      - Random variability (Aleatory)
      - Lack of data, information, knowledge (Epistemic)

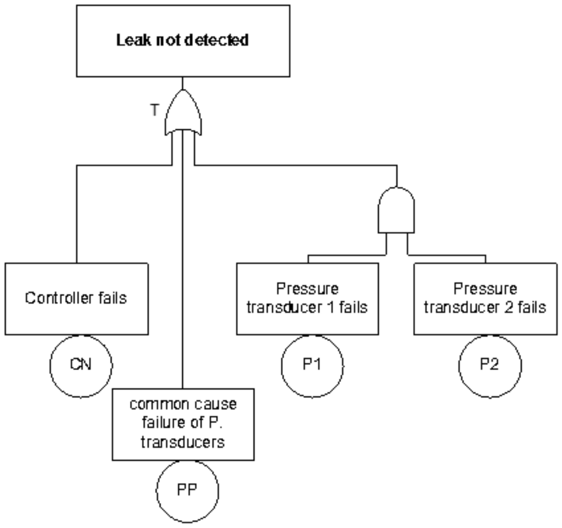
Reliability Engineering covers multiple domains



Another facet of data-driven approaches to reliability modeling

- Approaches to building probabilistic models
  - Statistical models: "How often?"
    - Predictions for static, uncertain conditions
    - Require data
      - Classical statistics: large (infinite) number of exchangeable observations
      - Bayesian statistics: sparse data
  - Causal models: "Why?"
    - Predictions for changing (uncertain) conditions
    - May or may not use "traditional" statistical data





Physics of Failure (PoF) Approaches to Reliability

PoF views failure as a *challenge* exceeding an item's *capacity*.

- Challenges and capacities affected by internal and external conditions (**influencing factors**)
- Challenges are **agents of failure** activated by the influencing factors

- **Performance-Requirement Model**
  - Reliability = performance (e.g., efficiency, output) w/i acceptable limits
- **Stress-Strength Model**
  - Reliability = challenge (stress) within capacity (strength)
  - Stress = aggregate of challenges and external conditions
  - Strength = random variable (r.v.) or uncertainty
- **Damage-Endurance Model**
  - Stress causes accumulating damage over time

Agents of Failure

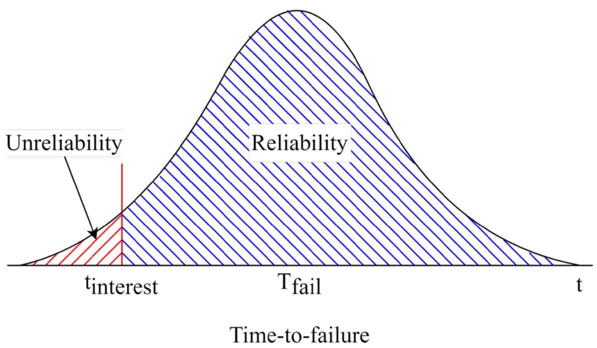
- **Failure Agent**: conditions/occurrences that *cause* items to fail
- PoF approach: challenges are caused by failure agents
- Two most important failure agents are **time** and **stress**
- **Time:** component aging, small cumulative damages over time
- **Stress:** mechanical, thermal, electrical, chemical, radiation
  - Mechanical: loading, cyclic loading
  - Thermal: high heat, thermal cycling
  - Electrical: high voltage
  - Chemical: salt, corrosion
  - Radiation: neutron, ionizing radiation

Measures of performance

- **Reliability:** Item’s ability to operate without failure under specified operating conditions to attain a mission having the desired length of time or number of cycles.
- **Availability:** Probability that an item, when used under stated conditions and support environment will be operational at a given time. (i.e., ability to become & remain operational following a failure)
- **Maintainability:** Item's ability to be quickly restored following a failure.
- **Capability:** Item's ability to attain its functional requirements.
- **Efficiency:** Item's ability to attain its functions economically and quickly with minimum waste.

Probabilistic definition of reliability

- *Reliability (R) is the ability of an item to operate without failure under specified operating conditions to attain a mission having the desired length of time or number of cycles.*
- $R(t_{interest}) = \Pr(T_{fail} \geq t_{interest} | c_1, c_2, \ldots)$



- Where:
  - $t_{interest}$ = time of interest (e.g., mission time) or aggregate agent of failure
  - T_{fail} = a random variable time-to-failure (cycle-to-failure, stress-to-failure, etc)
  - $c_i, c_2, \ldots$ = Designated operating conditions, environments, etc.

Definitions: Availability & Maintainability

- **Availability & Maintability consider repair:**.
  - Conditions & environment include perfect spare parts, personnel, diagnosis equipment, procedures, etc.
  - Availability is a *probability*, will be discussed further in Module 7
- Notionally, Availability (A) is: $A = \frac{U}{U+D}$, where:

  A = Availability

  U = uptime during time *T*

  D = downtime during time *T*

Discussion questions:

- In what types of industries, applications is reliability engineering especially prevalent? Why?

- What are some consequences of unreliability?




Definitions: Risk & Risk Analysis

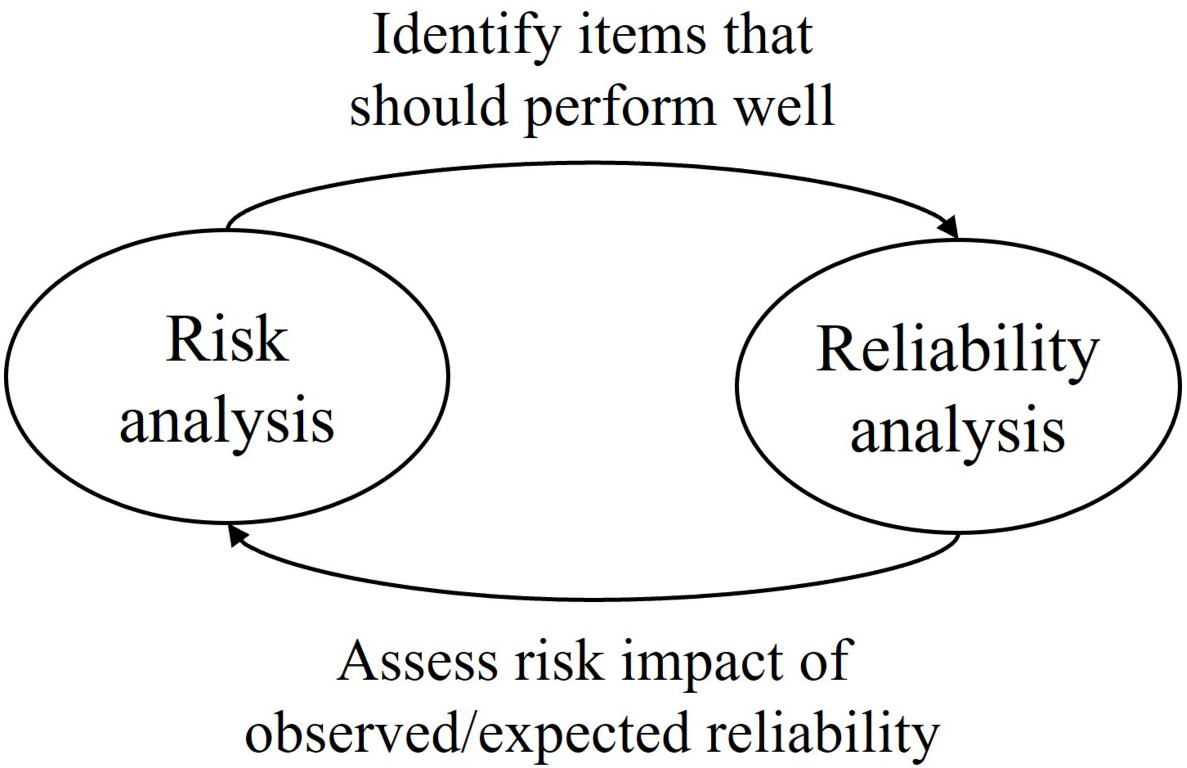
Risk: *"the potential to cause a loss" (more specifically, "the potential of loss resulting from exposure to a hazard"* which is understood to embody "uncertainty about the potential for and severity of loss(es))

Hazard: *a source of damage, harm, or loss*

Risk Assessment		Risk Management
• A process used to identify and characterize risk in a system • What could go wrong? • How likely is it? • What are the consequences?	+	• Provide inputs to decision makers on: • Sources of risk • Strategies to reduce risk • Priorities

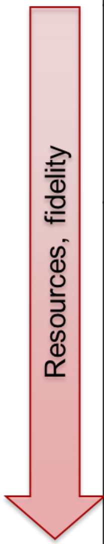
$$\boldsymbol{Risk \propto scenarios, probabilities, consequences}$$

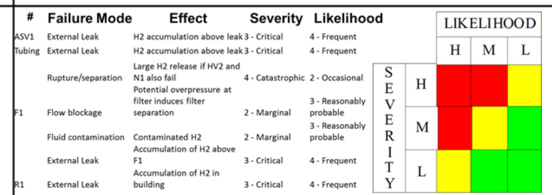
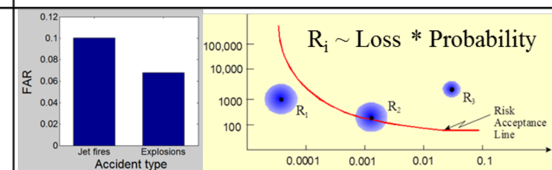
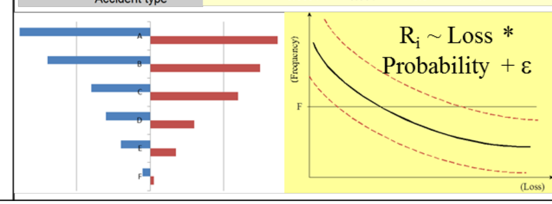
Reliability and risk analysis are closely related



There are many ways to do Reliability & Risk Analysis

- **Caution:** The term "Risk analysis" or "Risk assessment" is associated with over 10,000 methods/models/equations!



Type	Example methods	Example outputs
Qualitative to semi-quantitative	• FMEA • HAZOP • PHA	
Quantitative	**QRA [Quantitative Risk Assessment]** • Fault Trees • Event Trees • Bayesian Networks • Simulations	 

- Rigorous QRA (Quantitative Risk Assessment) methods involve a wide range of inputs, models, & data
- Requirements & practices vary across industry, location, application, & technology – terminology, methods, goals, criteria, & rigor varies widely.

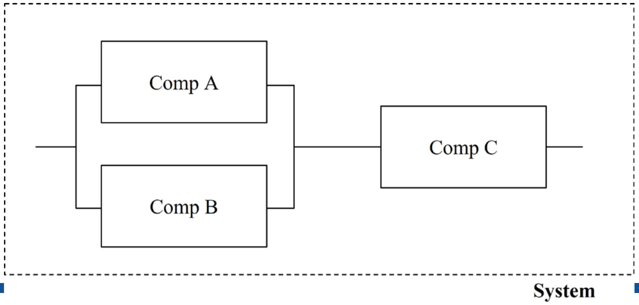
Exercise (5 min)

- Break into groups of 2-4 and pick an accident or failure you're familiar with
  - List a hardware failure, a human failure, and an environmental factor which contributed to the accident.
  - Name one consequence of the accident (other than loss of life).

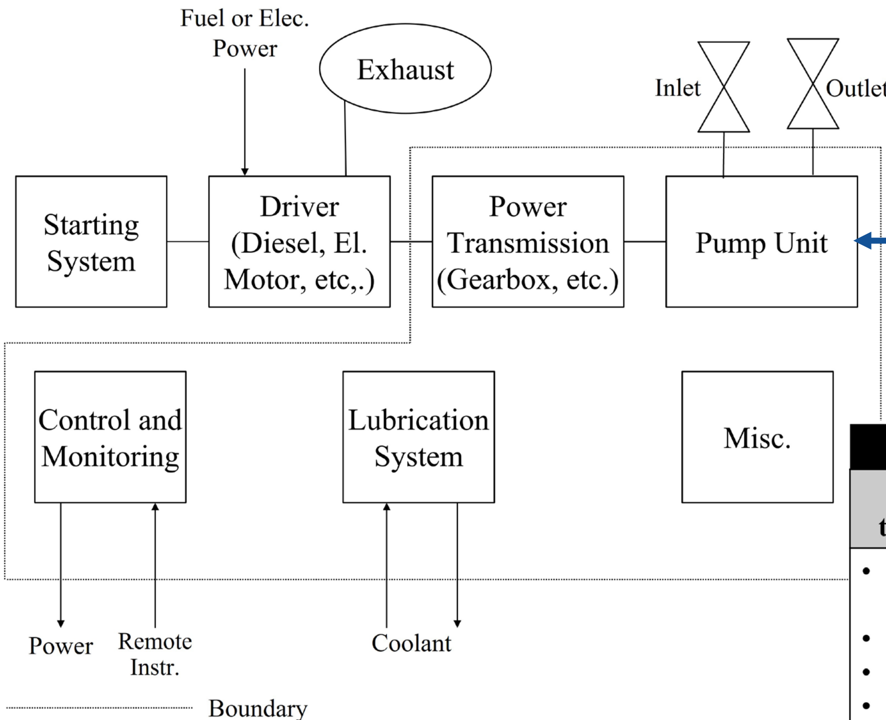
  - (Wait to report out, we will come back to this).

Component vs System vs Function

- **Component:** Basic physical entity of the system analyzed from a reliability perspective (i.e., not further divided into more abstract entities).
- **System:** Collection of items whose coordinated operation achieves a specific function (or functions).
  - Items may be subsystems, structures, components, software, algorithms, human operators, users, maintenance programs, etc.
- The line between component and system is arbitrary and varies depending on objectives, scope, and resources of analysis
- **Function:** The purpose served by an item to meet some system requirement



Example: System & component definition for a pump

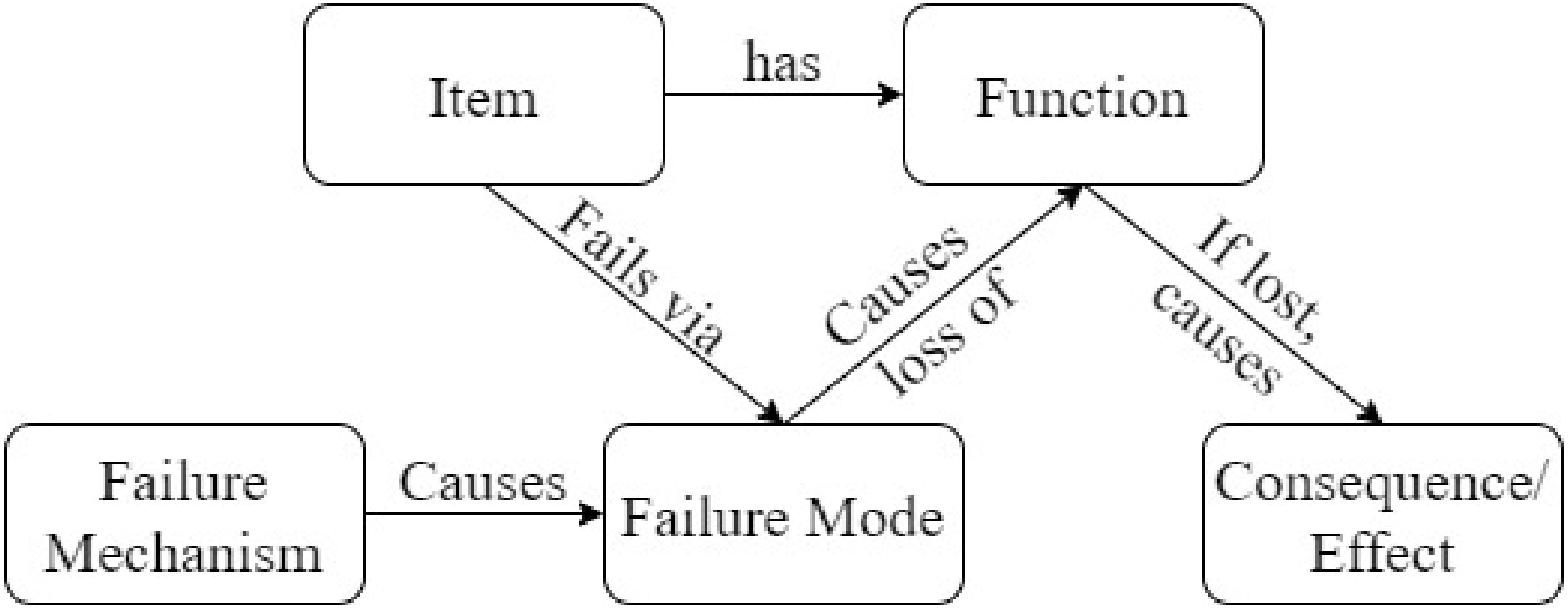


Pump unit can be treated as a *component* in a broader system

Pump System				
Power transmission	**Pump**	**Control and monitoring**	**Lubrication system**	**Miscellaneous**
• Gearbox/var. drive • Bearing • Seals • Lubrication • Coupling to driver • Coupling to driven unit • Instruments	• Support • Casing • Impeller • Shaft • Radial bearing • Thrust bearing • Seals • Valves & piping • Cylinder liner • Piston • Diaphragm • Instruments	• Instruments • Cabling & junction boxes • Control unit • Actuating device • Monitoring • Internal power supply • Valves	• Instruments • Reservoir w/ heating system • Pump w/ motor • Filter • Cooler • Valves & piping • Oil • Seals	• Purge air • Cooling/heating system • Filter, cyclone • Pulsation damper

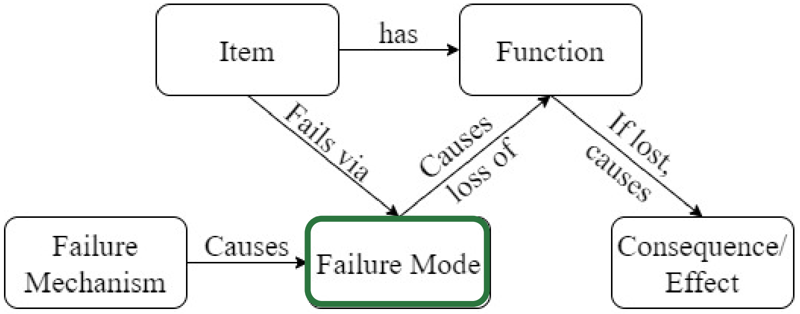
Pump can be treated as a *subsystem* and broken down into smaller *components*

Failure modes, failure mechanisms, and functions are building blocks of risk & reliability scenarios



Failure Modes

- The **Failure Mode** is the manner or way in which an item fails
  - Functional Manner of item (component, system) failure
  - *How* does the item fail?
  - *What* does the item fail to do?
- Some typical failure modes for active hardware components:
  1. Premature operation,
  2. Failure to start operation when needed,
  3. Failure to stop operation at the prescribed time,
  4. Failure to continue operation after start,
  5. Degraded operation.



Example: List of pump failure modes

AIR	Abnormal instruments reading
BRD	Breakdown
ERO	Erratic output
ELP	External leakage - Process medium
ELU	External leakage - Utility medium
FTS	Fail to start on demand
STP	Fail to stop on demand
HIC	High output
INL	Internal leakage
LOO	Low output
SER	Minor in service problems
NOI	Noise
OTH	Other
OHE	Overheating
PDE	Parameter deviation
UST	Spurious stop
STD	Structural deficiency
UNK	Unknown
VIB	Vibration

Source: OREDA – Offshore Reliability Data Handbook, 4th Edition, 2002

Example: Pump failure mode data

Taxonomy no	Item		
1.3	Machinery Pumps		

Population	Installations	Aggregated time in service (10^6 hours)		No of demands
		Calendar time *	Operational time †	
449	61	19.0224	8.6743	11200

Failure mode	No of failures	Failure rate (per 10^6 hours). Lower	Mean	Upper	SD	n/τ	Active rep.hrs	Repair (manhours) Min	Mean	Max
Critical	**524***	**0.00**	**20.52**	**108.44**	**49.34**	**27.55**	**37.3**	**1.0**	**53.1**	**1025.0**
	524†	**1.14**	**65.40**	**204.64**	**72.93**	**60.41**				
Breakdown	45*	0.00	1.27	6.56	5.17	2.37	16.1	3.0	52.5	766.0
	45†	0.01	3.85	15.72	5.95	5.19				
Erratic output	2*	0.00	0.14	0.72	0.58	0.11	19.8	11.0	39.5	68.0
	2†	0.00	0.38	2.00	0.91	0.23				
External leakage - Process medium	86*	0.00	2.38	12.29	9.53	4.52	28.4	2.0	38.3	444.0
	86†	0.00	7.07	33.87	13.94	9.91				
External leakage - Utility medium	46*	0.00	1.20	5.04	5.60	2.42	16.0	2.0	29.8	90.0
	46†	0.00	3.59	16.82	6.84	5.30				
Fail to start on demand	50*	0.01	2.52	9.77	3.62	2.63	52.0	1.0	56.6	551.0
	50†	0.08	13.75	48.28	17.83	5.76				
Fail to stop on demand	2*	0.00	0.10	0.21	0.54	0.11	3.5	3.0	3.5	4.0
	2†	0.00	0.26	1.30	0.56	0.23				
High output	3*	0.00	0.67	3.51	2.44	0.16	-	1.0	3.3	6.0
	3†	0.00	2.31	12.00	5.32	0.35				
Internal leakage	8*	0.00	0.34	1.39	0.52	0.42	95.5	3.0	48.3	188.0
	8†	0.16	0.98	2.37	0.72	0.92				
Low output	46*	0.00	2.50	3.96	15.25	2.42	35.4	3.0	41.2	508.0
	46†	0.00	4.57	13.58	22.90	5.30				
Noise	6*	0.15	0.33	0.56	0.13	0.32	23.3	16.0	60.5	122.0
	6†	0.01	1.03	3.73	1.38	0.69				
Other	8*	0.00	0.57	2.99	2.43	0.42	275.5	2.0	424.5	734.0
	8†	0.00	1.53	7.57	3.21	0.92				
Overheating	5*	0.00	0.27	0.95	0.35	0.26	183.2	3.0	265.0	1025.0
	5†	0.00	6.41	32.56	14.04	0.58				
Parameter deviation	18*	0.00	0.66	3.49	2.31	0.95	11.0	1.0	20.8	88.0
	18†	0.14	1.96	5.66	1.87	2.08				
Spurious stop	133*	0.00	5.69	27.65	11.50	6.99	37.5	1.0	42.1	714.0
	133†	1.57	19.07	53.52	17.47	15.33				
Structural deficiency	33*	0.00	0.41	0.51	4.91	1.73	20.6	5.0	40.5	211.0
	33†	0.00	1.24	3.74	6.18	3.80				
Unknown	1*	0.00	0.05	0.15	0.05	0.05	-	-	-	-
	1†	0.00	0.11	0.33	0.12	0.12				
Vibration	32*	0.00	1.67	7.70	3.10	1.68	81.2	5.0	118.3	896.0
	32†	0.47	5.11	14.03	4.53	3.69				
Degraded	**754***	**0.00**	**44.20**	**210.34**	**86.32**	**39.64**	**20.2**	**0.3**	**26.4**	**798.0**
	754†	**11.39**	**238.41**	**714.72**	**239.40**	**86.92**				
Abnormal instrument reading	9*	0.00	0.80	4.56	2.45	0.47	9.0	2.0	16.0	65.0
	9†	0.00	2.53	11.22	4.42	1.04				
Erratic output	23*	0.00	2.27	12.50	6.03	1.21	14.8	2.0	16.8	65.0
	23†	0.00	7.88	35.25	13.95	2.65				

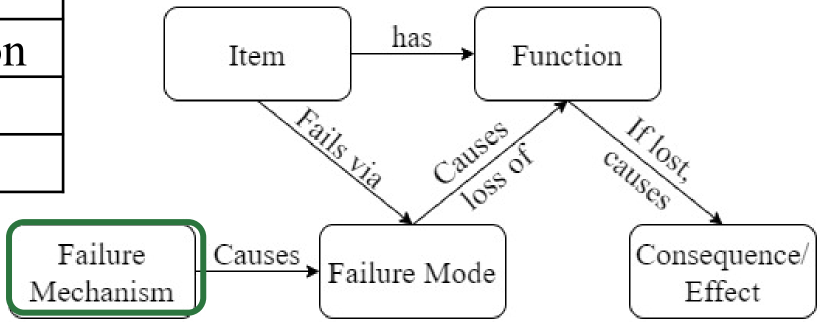
Source: OREDA – Offshore Reliability Data Handbook, 4th Edition, 2002

Failure Mechanisms

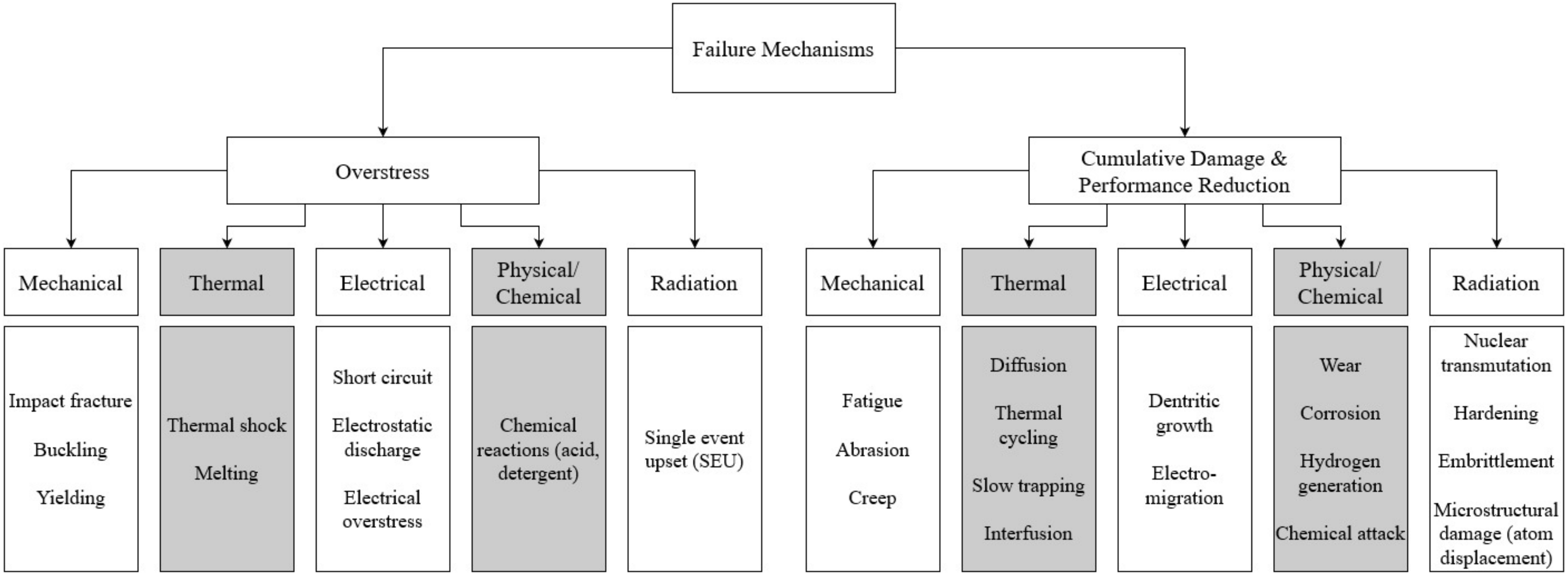
- **Failure Mechanisms** are physical processes through which damage occurs.
- Damage can occur rapidly (abruptly) or slowly (cumulatively).
  - Rapid: brittle fracture, melting, and yielding.
  - Cumulative: fatigue, wear, and corrosion.

Failure occurs when resulting damage exceeds item's capacity

Damage-inducing	**Capacity-reducing**
Wear	Fatigue
Corrosion	Embrittlement
Cracking	Thermal shock
Diffusion	Diffusion
Creep	Grain boundary migration
Fretting	Grain growth
Fatigue	Precipitation hardening



Examples of different types of failure mechanisms



Example: Failure mechanisms & modes

- Failure mechanisms are physical processes, failure modes are the **outcome** of the physical process
- **Example:** neutron irradiation causes embrittlement (cumulative failure mechanism), which results in fracture (failure mode)
- **Example:** short circuit (overstress failure mechanism) in pump results in failure to start on demand (failure mode)
- **Example:** corrosion (cumulative failure mechanism) in pump results in external leakage of process medium (failure mode)

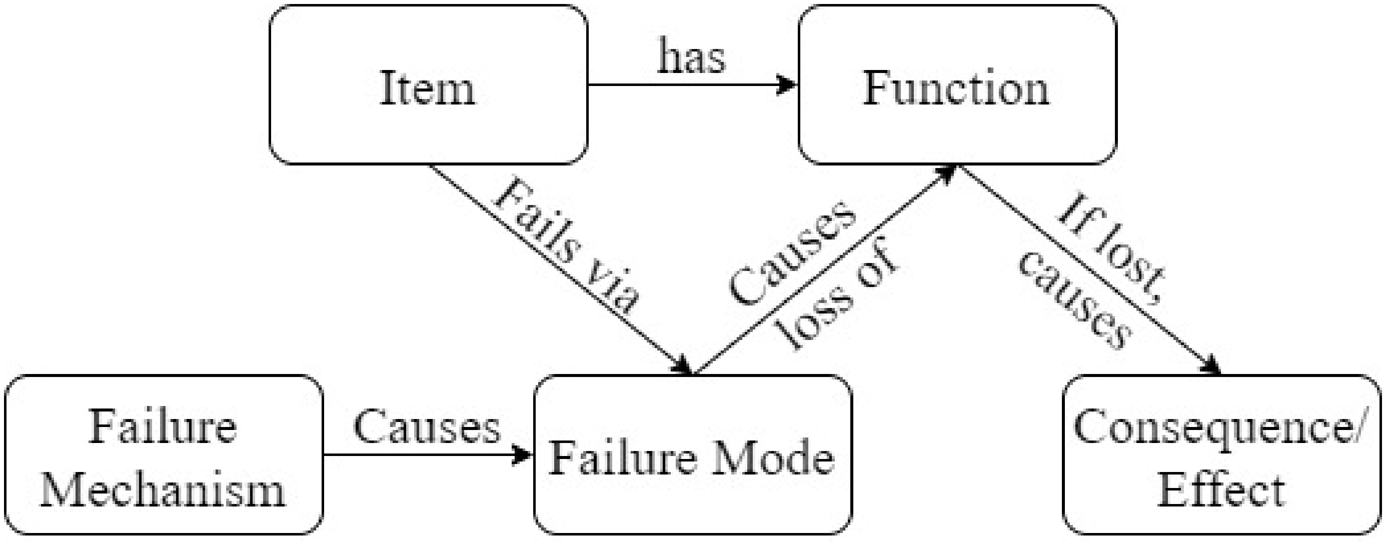
Human functions, failure, and reliability

- Systems are not only hardware and software – **humans** play a critical role in a functioning engineered system.
- Humans are not hardware:
  - **Multiple roles**: design, operation, maintenance, decommissioning
  - Achieve various functions at several levels of abstraction
- **Human reliability analysis** (HRA) provides methodologies to understand and model human failures in a system.
  - **Major human tasks**: Information Gathering, Diagnosis & Decision Making; Action Execution
  - Over 50 methods exist for engineering applications
  - Methods differ in the use of **performance influencing factors** (PIFs)
- Similar importance and modeling for **software reliability**
- We'll dive deeper in Module 8.

Example hardware failure scenario

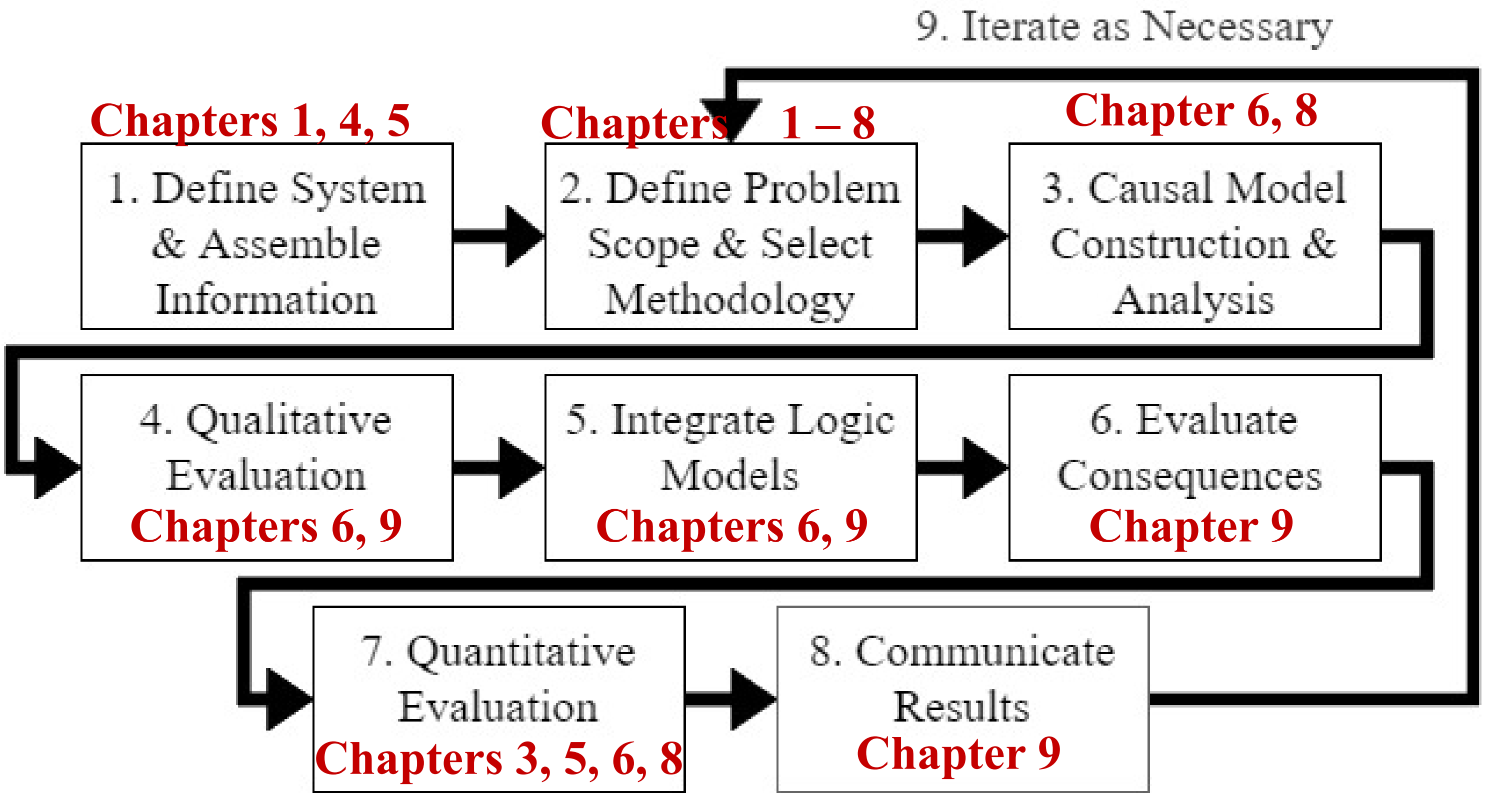
Example for an aircraft's airframe:

Fatigue (failure mechanism) of airframe (component or system), caused by the cyclic stress, cyclic temperature (influencing factors), activated by random flight loads (influencing factor) external to airframe, leading to crack initiation and growth (degradation process), and ultimately causing structural failure during operation (system failure mode). Structural failure causes loss of airworthiness (function), which causes crash (effect).



Risk Assessment Modeling

Putting the pieces together enables us to do risk assessment – and forms the outline of the modules in this course



This just the beginning. ENRE00, ENRE640, ENRE641, ENRE645, ENRE655, ENRE670 and more will delve even deeper into specific facets of the discipline.

Quiz 1: Write a failure scenario for the event you discussed earlier.

- Your scenario **must include**:
  - Event/accident name, date, description (1-2 sentences), and photo
  - System failure mode, operating environment, consequence(s)
  - At least one of: hardware failure mode, failure mechanism, environment, and human failure
- Example: Air France 447 Airbus A330 (July 1, 2009)

Component & failure mode | Component failure mechanism

Subsystem failure mode

Flow blockage of **pitot tubes,** caused by **freezing,** led to **erroneous airspeed indication** resulting in **loss of control,** **loss of aircraft, fatality of all 228 passengers and crew,** **with pilot failure to diagnose** the erroneous output due to inadequate information, and time load. Occurred at **high altitude, low temperature, in stormy weather.**

System failure mode

Consequences

Human failure mode & PIFs

Environment

Image: Brazilian Navy/ Getty Images

Reliability Analysis

Module 2: Basic Reliability Math: Probability

Prof. Katrina M. Groth
Mechanical Engineering Department
Center for Risk and Reliability
University of Maryland
kgroth@umd.edu

Objectives for this module

- At the end of this module you will [be able to]:
  - Manipulate events (and data) using sets & Boolean algebra
  - Know laws of probability
  - Know and use Bayes' Theorem
  - Solve problems using common parametric probability distributions

Random variables (events) of interest

- Notation: A capital letter (e.g., X) denotes a random variable, a lower-case letter (x) denotes a value that the r.v. can take.
- X could be an r.v representing:
  - Time-to-failure of an item: $x \in [0, \infty)$
  - Number of failures occurring in some time: $x \in [0, 1, 2, \dots, n]$
  - Number of cycles until first failure occurs: $x \in [0, 1, 2, \dots, n]$
  - Number of failed components drawn from a population: $x \in [0, 1, 2, \dots, n]$
  - And more

Sets and Boolean algebra

- A **set** is a collection of items or elements, each with some specific characteristics.
- A **universal set** Ω: is a set that includes all items of interest
- A **null set or empty set ,** $\varnothing$ is a set with no items.
- A **subset,** ($\subset$, or *subset equal to* $\subseteq$) refers to a collection of items that belong to a set
- $B \subset A \subset \Omega$.

In engineering notation, Ω is often replaced by 1 and $\varnothing$ by 0.

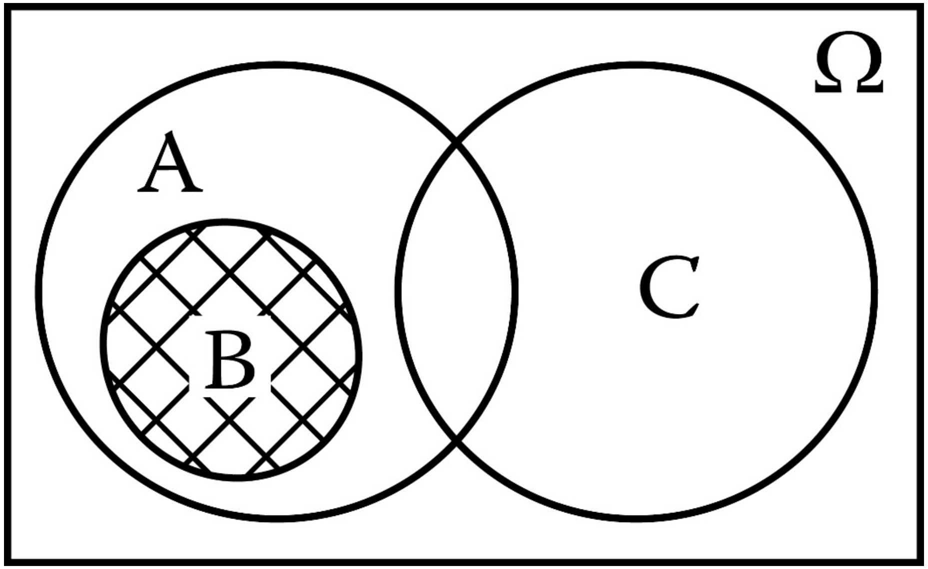
Example sets

- Example continuous set:
  - For T, failure times for light bulbs.
    - $A = \{t \mid t > 0\}$ could show the universal set
    - $B = \{t \mid 0 < t < 100\}$ shows a subset of light bulbs that operate for $t > 0$ but fail prior to 100 hours.

- Example discrete sets:
  - States of a pump A: {Failed, working}
    - We may be interested in analyzing only those that are failed…
  - A={1,2,3,4,5,6} is the universal set of outcomes from rolling a die once.
    - B={4,5,6} is the subset of dice outcomes that are *greater than 3*.

Sets and Boolean algebra (cont.)

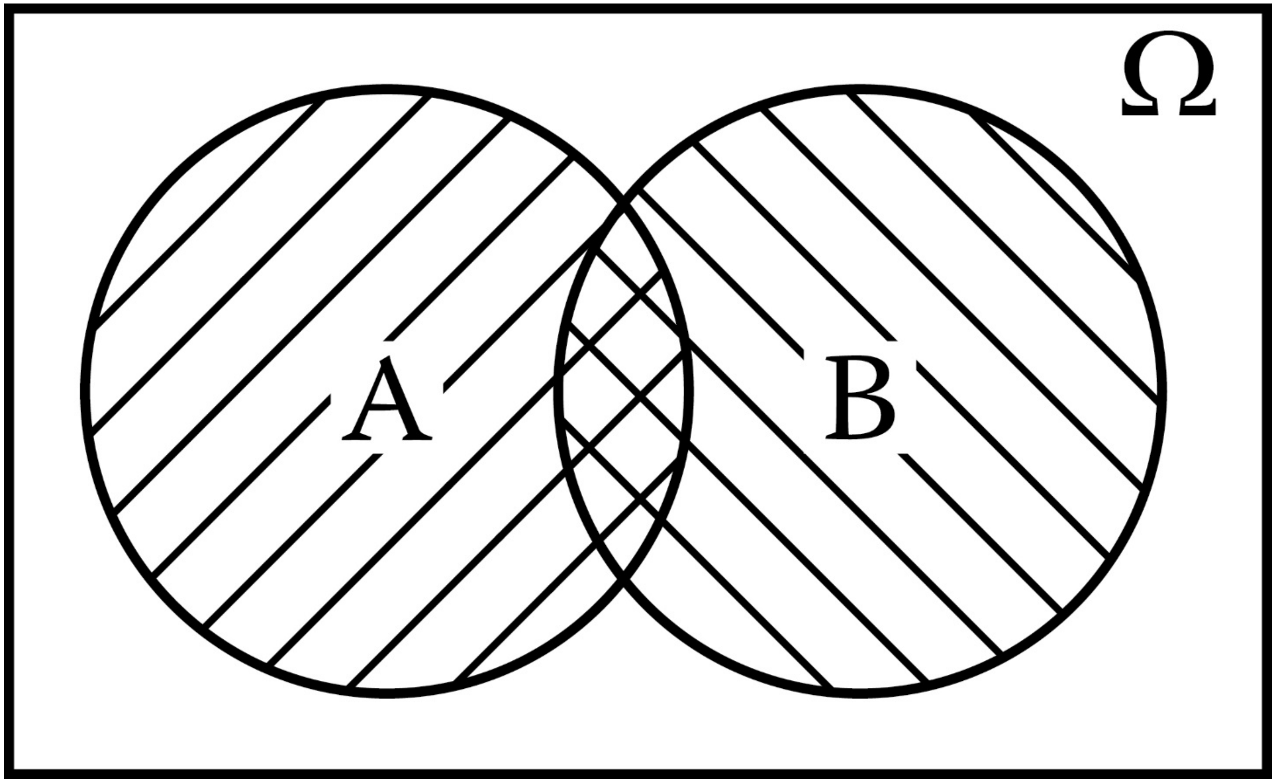
- The **Venn diagram** shows the relationship between sets and subsets.
- Universal set Ω by a rectangle, and its subsets A and B by circles. It can also be seen that B is a subset of A. The relationship between subsets A and B and the universal set can be symbolized by $B \subset A \subset \Omega$ (or *subset equal to*, $\subseteq$).
- In this figure:
  - C has no elements in common with B.
  - A and C have at least one element in common.



Set operation

- The **union** of two sets, A or B, is a set of element that belong to A or B or both.

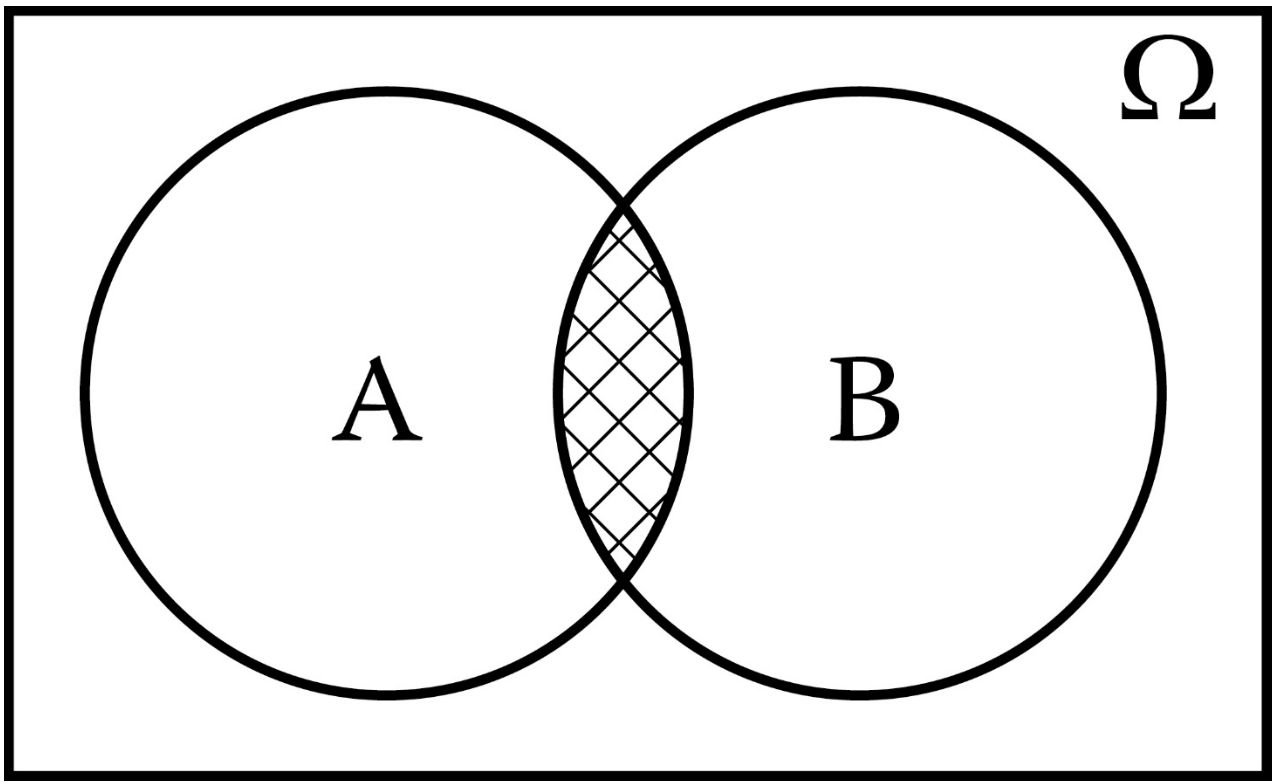
$$(A \cup B) \text{ or } (A + B)$$



Set operation

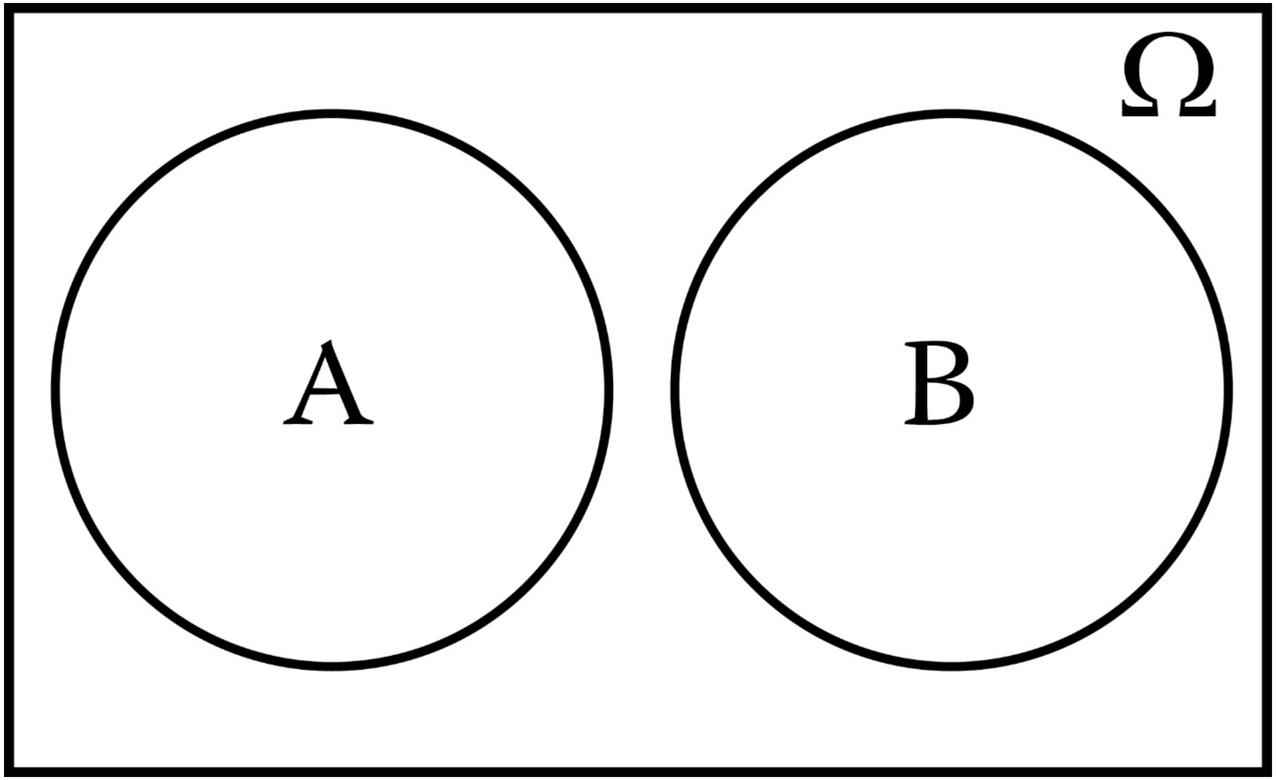
- **Intersection** of two sets called (A and B) is the set of elements belonging to *both* A and B

$$(A \cap B) \text{ or } (A \cdot B)$$



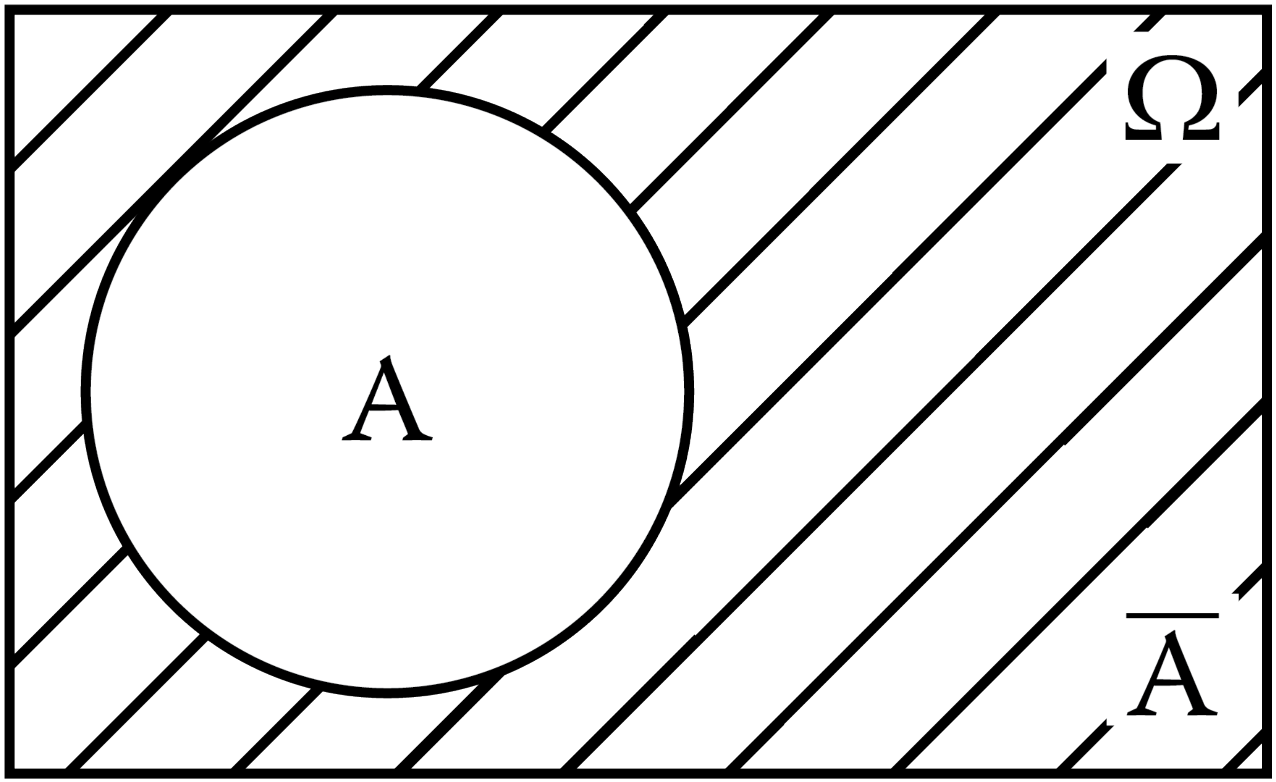
Set operation

- If $A \cap B = \varnothing$, then the sets A and B are **mutually exclusive (or disjoint)** i.e., they have no elements in common.



Set operation

- The **complement** of a subset is all the elements in the universal set that do not belong to subset. The complement of A is $\bar{A}$.



Example: Set operation

- **Example 1**: Consider the discrete set S and subset A. What is the complement of A?

$$S = \{1,2,5,10\} \qquad A = \{2,5\}$$

$Solution$: $\bar{A} = \{1,10\}$

- **Example 2**: Consider the continuous set S, which represents lightbulb failure times in our dataset, and subset A, which we will call "early failures." What is the complement of A?

$$S = \{t|t > 10\} \qquad A = \{t|100 \geq t > 10\}$$

$Solution$: $\bar{A} = \{t|t > 100\}$

Why we use sets

- We can manipulate interesting subsets of sample spaces or talk about interesting combinations of events.
  - For example, $B = \{t \mid 0 < t < 100\}$ shows a subset of light bulbs that operate for $t > 0$ but fail prior to 100 hours.
  - A = {Failed}
- For example, take the two events: A = pump A has failed (off), and B = pump B has failed (off).
- We might want to know when:
  - $A \cap B$ = Both pumps failed.
  - $A \cap \overline{B}$ = A has failed, B did not fail.
  - $\overline{A} \cap B$= A did not fail, B did fail.
- Boolean algebra is extensively used in reasoning about **events & sets.**

Boolean Algebra Laws

Designation	Mathematical Notation	Engineering Notation
Identity laws	$A \cup \emptyset = A$ $A \cup \Omega = \Omega$ $A \cap \emptyset = \emptyset$ $A \cap \Omega = A$	$A + 0 = A$ $A + 1 = 1$ $A \cdot 0 = 0$ $A \cdot 1 = A$
Idempotent laws	$A \cap A = A$ $A \cup A = A$	$A \cdot A = A$ $A + A = A$
Complement laws	$A \cap \overline{A} = \emptyset$ $A \cup \overline{A} = \Omega$	$A \cdot \overline{A} = 0$ $A + \overline{A} = 1$
Law of Absorption	$A \cup (A \cap B) = A$ $A \cap (A \cup B) = A$	$A + (A \cdot B) = A$ $A \cdot (A + B) = A$
de Morgan's Theorem	$\overline{(\mathrm{A} \cap B)} = \overline{A} \cup \overline{B}$ $\overline{(\mathrm{A} \cup B)} = \overline{A} \cap \overline{B}$	$\overline{(A \cdot B)} = \overline{A} + \overline{B}$ $\overline{(A + B)} = \overline{A} \cdot \overline{B}$
Commutative laws	$A \cap B = B \cap A$ $A \cup B = B \cup A$	$A \cdot B = B \cdot A$ $A + B = B + A$
Associative laws	$A \cup (B \cup C) = (A \cup B) \cup C = A \cup \mathrm{B} \cup C$ $A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$	$A + (B + C) = (A + B) + C = A + B + C$ $A \cdot (B \cdot C) = (A \cdot B) \cdot C = A \cdot B \cdot C$
Distributive laws	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$A \cdot (B + C) = (A \cdot B) + (A \cdot C)$ $A + (B \cdot C) = (A + B) \cdot (A + C)$

Boolean algebra: Example 1

Example: Show that $\overline{(A \cdot B) + \left(A \ \cdot \ \overline{B}\right)} = \ \overline{A}$

(KG reminder: work this out on board immediately; runs too long otherwise).

Boolean algebra: Example 1 solution

Example: Show that $[\overline{(A \cdot B) + \left(A \cdot \overline{B}\right)} = \overline{A}$

$$\overline{(A \cdot B) + \left(A \cdot \overline{B}\right)} = \left(\overline{A \cdot B}\right) \cdot \left(\overline{A \cdot \overline{B}}\right) \qquad \text{(de Morgans)}$$

$$= \left(\overline{A} + \overline{B}\right) \cdot \left(\overline{A} + B\right) \qquad \text{(de Morgans)}$$

$$= \overline{A} + \underbrace{\left(\overline{B} \cdot B\right)} \qquad \text{(Distributive)}$$

$$= \overline{A} + 0 \qquad \text{(Complement)}$$

$$= \overline{A} \qquad \text{(Identity)}$$

Extra example: Solve this on your own and show that

$$(X \cdot Y) + \left(\overline{X} \cdot Z\right) + (Y \cdot Z) = (X \cdot Y) + \left(\overline{X} \cdot Z\right)$$

Terminology

- **Probability**: A numerical measure of the chance that an event occurs (or that a hypothesis is true); used to quantitatively express uncertainty about the occurrence of an event
- **Frequency:** The rate of occurrence of events – that is, the number of times an event occurs in a given period of time [or space, or number of trials]
- **Likelihood:** The probability of observing evidence or data when given an event or proposition.
- **Statistics**: Collecting, analyzing, and interpreting data

Probability interpretations used in reliability

- Take an event, X. We want to discuss $\Pr(X = x)$
  - E.g., $\Pr(x) = 0.80$
- Two interpretations are used in aspects of reliability engineering.
  - **Frequentist interpretation: estimate the value of $\Pr(\boldsymbol{x})$ with observed data**
    - Appropriate when the number of observations is large, when the observed data are relevant for (and interchangeable with) the situation at hand.
  - **Subjectivist (Bayesian) interpretation: probability as a degree of belief (state of knowledge)**
    - Allow us to take additional information into account; to deal with rare events, unique systems, one-time events, etc.
- The same laws of probability and Boolean methods apply regardless of interpretation

Philosophical moment….

Probability is not really about numbers;
it is about the structure of reasoning.

Glenn Shafer
Rutgers University

Probability defined

Classical Definition of Probability (sample space partition)

- Associated with any event (outcome) E in a sample space is a probability denoted by:

$$\Pr(E) = \frac{n_E}{n} \text{ (Note that } \Pr(S) = \frac{n}{n} = 1, \Pr(\emptyset) = 0\text{)}$$

- Where:

  n_E = number of times event E occurs in sample space
  n = total number of events in the sample space

Example: All dice outcomes n = 6, All odd dice outcomes $n_E = 3$

Probability defined (cont.)

Frequentist (or Frequency) Definition of Probability

- This interpretation of Pr(E) assumes that the total number of *identical* events in the sample space (n) is unknown. Therefore, the probability of event E may be defined as a limit of n_E/n as n becomes large, i.e. the *proportion (or frequency) of outcomes*:

$$\Pr(E = e) = lim_{n \to \infty} \left(\frac{n_E}{n} \right)$$

- Where:

  n_E = number of times event E occur in sample space
  n = total number of events in the sample space

Example: Number of times a pump is started n = 2000, Number of failed pump starts $n_E = 20$.

Then Pr(E=e) = 20/2000=0.01

Probability defined (cont.)

Subjectivist (or Bayesian) Definition of Probability

- Bayesian probability, also called evidential probability or subjectivist probability), can be assigned to any statement whatsoever, as a way to represent its plausibility, or the degree to which the statement is supported by the available evidence (*degree of belief*).

  - Probability is the **degree of belief** in the truth of a proposition.
  - It is a representation of an individual's state of knowledge
  - The main requirement is **coherence**: One's subjective probability of an event must be consistent with their knowledge (evidence) & the laws of probability
  - ***Assumption***: Any two individuals with the same knowledge, information, and biases will assign the same probability value.

Examples: probability interpretation

Probability Statement	Which probability interpretation could this be? Frequentist or Subjectivist
The probability of flipping a [fair] coin and getting heads is 0.50	
There is an 0.80 probability of rain in College Park tomorrow.	
There is a 0.019 probability that a bolt is defective; the bolt is from a supply of 1000 bolts manufactured today at process plant X.	
The probability that you will receive an A in this class	
The probability of a loss of containment event this year at nuclear power plant X is $1.7*10^{-5}$.	

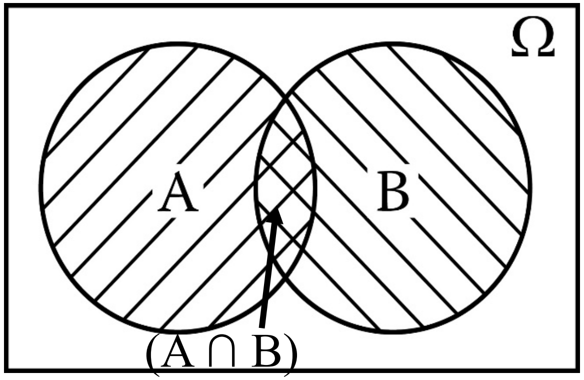
Probability terminology & notation

- **Marginal (unconditional) probability: Pr(A)**
  - The probability of event A occurring
- **Joint probability: $\Pr(A \cap B) = \Pr(A, B) = \Pr(AB)$**
  - The probability of events A and B both occurring
- **Conditional probability: $\Pr(A|B)$**
  - The probability of event A occurring given that B has occurred.

$$\mathbf{Pr}(\boldsymbol{A}|\boldsymbol{B}) = \frac{\mathbf{Pr}(\boldsymbol{A} \cap \boldsymbol{B})}{\mathbf{Pr}(\boldsymbol{B})}$$

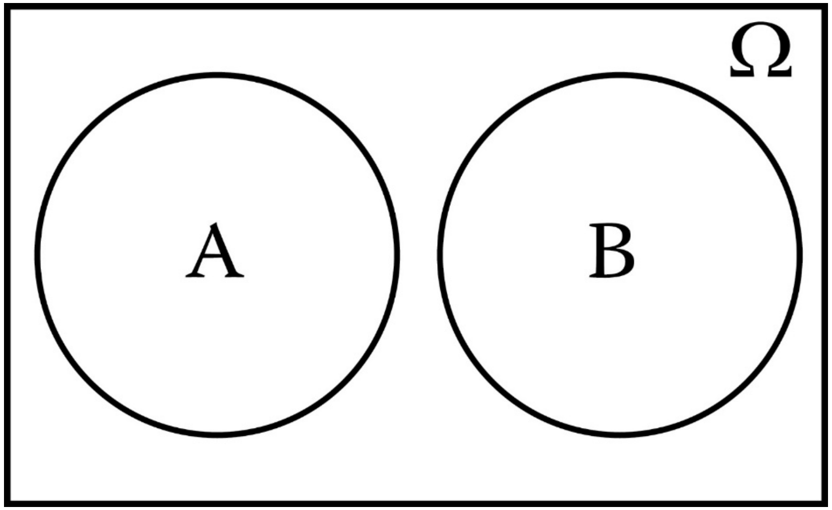
Definition of independent events

- **Definition of independent events:** Two events are independent if the occurrence or nonoccurrence of one does not depend on or change the probability of the occurrence of the other.
  - Mathematically this means: $\Pr(A|B) = \Pr(A)$
    - Which also means that $\Pr(A \cap B) = \Pr(A)\Pr(B)$
  - E.g., A = rolling a 1 on a die (first roll), B = rolling a 1 on the second roll
  - E.g., A = valve A fails to open, B = pump B fails to start (assuming no common cause)



Definition of mutually exclusive (disjoint)

- **Definition of mutually exclusive events:** $A \cap B = 0$
  - Two events are *mutually exclusive* if they can't both happen at the same time
    - Which means that $\Pr(A \cap B) = 0$
  - E.g.: A = rolling an even number on a die, B= rolling an odd number on a die
  - E.g., E1 = Valve E is operational, E2 = Valve E is failed



Axioms of probability

(Kolmogorov 1933):

1. $Pr(E_i) \geq 0$, for every event E_i

2. $Pr(\Omega) = 1$

3. $Pr(E_1 \cup E_2 \cup \ldots) = Pr(E_1) + Pr(E_2) + \ldots,$
   when $E_1, E_2, \ldots$ are mutually exclusive
   - Reminder: Mutually exclusive means: i.e., $E_1 \cap E_2 = \varnothing$, no common points exist between $E_1, E_2, \cdots$

Additional implications of the axioms

- $\Pr(\emptyset) = 0$
- $0 \leq \Pr(E_i) \leq 1$
- If $A \subseteq B$, then $\Pr(A) \leq \Pr(B)$
- $\Pr(\overline{E}) = 1 - \Pr(E)$

Chain rule of probability

- The **chain rule of probability (aka multiplication rule):** defines the relationship between joint probability and conditional probability

$$\Pr(E_n \cap E_{n-1} \cap \ \dots \cap E_2 \cap E_1) = Pr(E_n | E_{n-1}, \dots E_2, E_1) \cdot Pr(E_{n-1} | \dots E_2, E_1) \cdot \dots \cdot Pr(E_2 | E_1) \cdot Pr(E_1)$$

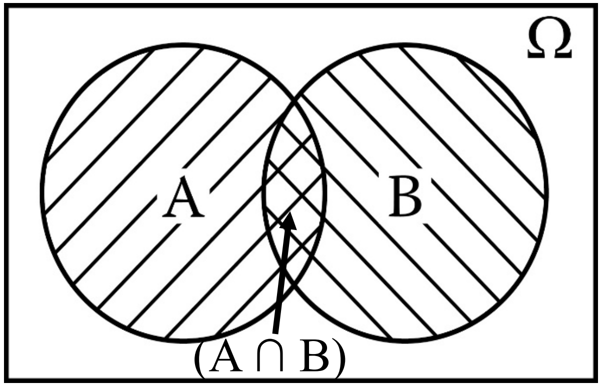
- If all events are **independent** (that is, $E_n \perp E_{n-1} \perp \ \dots \ \perp E_2 \perp E_1$), this simplifies to:

$$Pr(E_n \cap E_{n-1} \cap \ \dots \cap E_2 \cap E_1) = \prod_{i=1}^{n} Pr(E_i) = Pr(E_n) \cdot Pr(E_{n-1}) \cdot \dots \cdot Pr(E_2) \cdot Pr(E_1)$$

Addition law of probability

The **addition law of probability** (**inclusion-exclusion principle**)

If there are common element between A and B (i.e., they are not mutually exclusive, A ∩ B ≠ ∅) then:
$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.



If all E_i events are independent this is written in compact form as:

$$\Pr(E_1 \cup E_2 \cup \cdots \cup E_n) = 1 - \prod_{i-1}^{n} (1 - \Pr(E_i))$$

Law of total probability

- The **law of total probability** defines the relationship between joint and marginal distributions:

$$\Pr(A) = \sum_{i=1} \Pr(A \cap B_i) = \sum_{i=1} \Pr(A|B_i) \Pr(B_i)$$

- For example, to marginalize out a binary variable B:

$$\Pr(A) = \Pr(A \cap B) + \Pr(A \cap \overline{B})$$
$$\Pr(A) = \Pr(A|B) \Pr(B) + \Pr\left(A\middle|\overline{B}\right) \Pr(\overline{B})$$

Example

- **Example:** You receive bolts from two *independent* suppliers. The probability of the event D_i, "selecting a defective bolt from supplier i" is characterized by:

$$\Pr(D_1) = 0.05, \;\; \Pr(D_2) = 0.07$$

- If one bolt is selected from each supply, find:
  a) The probability that both selected bolts are defective?

  b) The probability that at least one is defective?

Hint: start by writing the LHS side of the questions in set & probability notation.

Example Solution

Example: Given $\Pr(D_1) = 0.05$, $\Pr(D_2) = 0.07$, and $D_1 \perp D_2$. One bolt is selected from each supplier.

a) The probability that both selected bolts are defective, since the bolts are independent, is:

$$\boldsymbol{Pr(D_1 \cap D_2)} = \Pr(D_1)\Pr(D_2) = (0.07)(0.05) = \mathbf{0.0035}$$

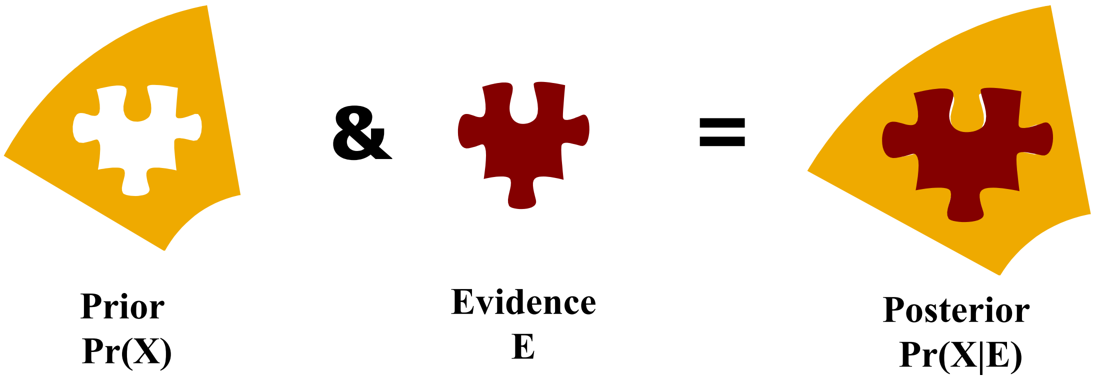
b) The probability that at least one bolt is defective is:

$$\boldsymbol{Pr(D_1 \cup D_2)} = \Pr(D_1) + \Pr(D_2) - \Pr(D_1 \cap D_2)$$

$$= 0.05 + 0.07 - 0.0035 = \mathbf{0.1165}$$

Bayes' Theorem: Conceptual

- Probability is used as the foundation for *reasoning*
- Provides a framework for updating beliefs (probabilities) in light of new evidence.



Bayes' Theorem (discrete)

- If A_j is an event of interest, and if E is any other event (or set of evidence) such that Pr(E) > 0, then:

$$\mathbf{Pr}(\boldsymbol{A_j}|\boldsymbol{E}) = \frac{\mathbf{Pr}(\mathbf{A}_{\boldsymbol{j}})\mathbf{Pr}(\boldsymbol{E}|\boldsymbol{A_j})}{\mathbf{Pr}(\mathbf{E})}$$

  - $\Pr(\mathrm{A_j}|\mathrm{E})$ is called **posterior (or a posteriori) probability of $\mathbf{A_j}$** (given evidence E)
  - $\Pr(\mathrm{A_j})$ is called **prior (or a priori) probability of $\mathbf{A_j}$**
  - $\Pr(\mathrm{E}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \Pr(\mathrm{A_i})\Pr(\mathrm{E}|A_i)$ from the law of total probability.
  - $\Pr(E|A_j)/\Pr(E)$ **is the relative likelihood** (the model by which the probability is revised)

Bayes' Theorem (continuous)

- For **continuous probability functions**, Bayes' Theorem is given:

$$f(\theta|E) = \frac{\Pr(E|\theta)\Pr(\theta)}{\Pr(E)} = \frac{f(\theta)f(E|\theta)}{f(E)}$$

$$f(E) = \int f(E|\theta)f(\theta)d\theta \longleftarrow \text{marginal pdf of } E$$

Example: Events, probability & Bayes' Theorem

- Tractors made by a company are from three assembly lines: Red, White, and Blue. (R, W, B)
- Suppose 48% of the company's tractors are made on the Red line and 31% are made on the Blue line.
- Let *D* be the event that the new tractor is defective (i.e., it wont start)
- The probability that a tractor will not start when it rolls off of a line are 6%, 11%, and 8% given that it comes from Red, White, and Blue, respectively.

a) Translate the problem statement & questions into probability notation.

b) What is the probability that a randomly selected tractor is defective?

c) What is the probability that a tractor came from the red line given that it is defective?

Solution part a: Define events

Events:

- Let D be the event that the tractor won't start
- Let R be the event that the tractor was made by the red assembly
- Let W be the event that the tractor was made by the white assembly
- Let B be the event that the tractor was made by the blue assembly

$$\Pr(\mathrm{R}) = 0.48$$

$$\Pr(\mathrm{W}) = 0.21$$

$$\Pr(\mathrm{B}) = 0.31$$

$$\Pr(\mathrm{D}|\mathrm{R}) = 0.06$$

$$\Pr(\mathrm{D}|\mathrm{W}) = 0.11$$

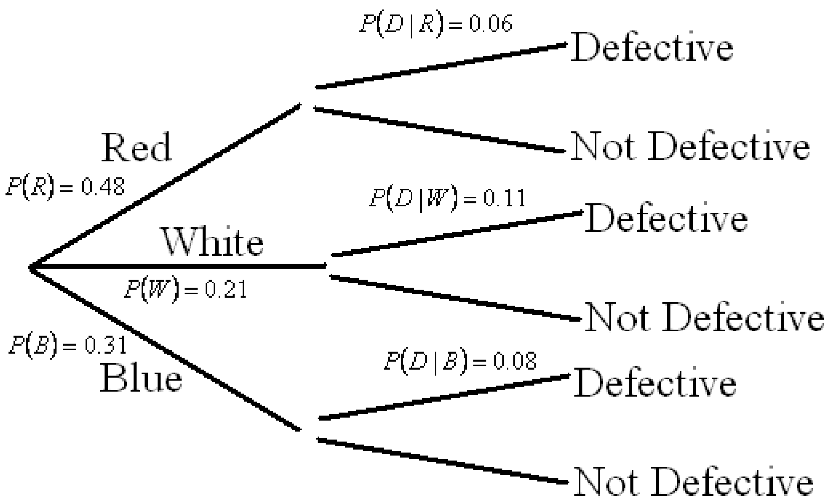
$$\Pr(\mathrm{D}|\mathrm{B}) = 0.08$$

- Our problem:

  b) What is the probability that a randomly selected tractor is defective? In other words: **Find $\mathbf{Pr(D)}$**.

  c) What is the probability that a tractor came from the red line given that it was defective? **Find $\boldsymbol{Pr(R|D)}$**.

Solution part b & c: Tree diagram



$\Pr(R \cap D)$?

$\Pr(W \cap D)$?

$\Pr(B \cap D)$?

$$\Pr(R \cap D) = \Pr(D|R)\Pr(R)$$

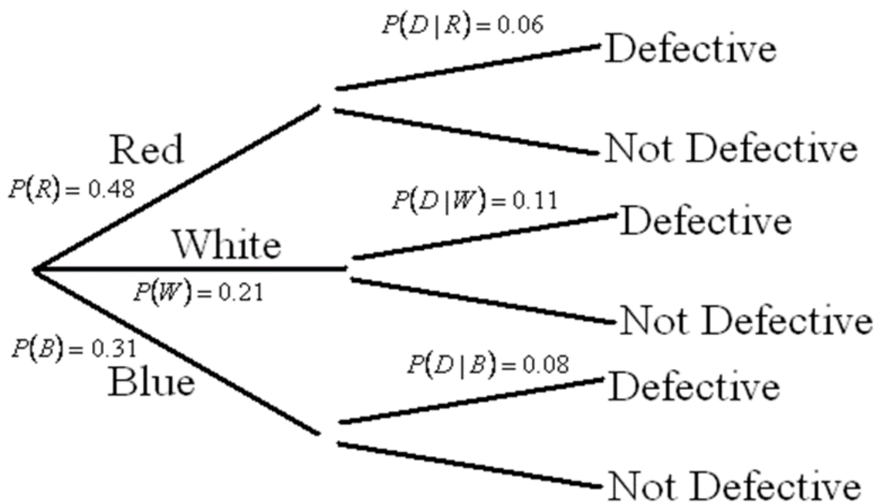
$$\Pr(W \cap D) = \Pr(D|W)\Pr(W)$$

$$\Pr(B \cap D) = \Pr(D|B)\Pr(B)$$

- Because each of these events represents an instance where a tractor is defective, to find the total probability that a tractor is defective, regardless of who made it:

$$\Pr(D) = \Pr(D|R)\Pr(R) + \Pr(D|W)\Pr(W) + \Pr(D|B)\Pr(B)$$

Solution parts b & c



$Pr(R \cap D)$?
$Pr(W \cap D)$?
$Pr(B \cap D)$?

$$Pr(R \cap D) = Pr(D|R)\,Pr(R)$$
$$Pr(W \cap D) = Pr(D|W)\,Pr(W)$$
$$Pr(B \cap D) = Pr(D|B)\,Pr(B)$$

- **Solution:**

$$\mathrm{Pr(D) = Pr(D|R)Pr(R) + Pr(D|W)Pr(W) + Pr(D|B)\,Pr(B)}$$
$$Pr(D) = (0.06)(0.48) + (0.11)(0.21) + (0.08)(0.31)$$
$$\boxed{\mathbf{Pr}(\boldsymbol{D}) = \mathbf{0.077}}$$

$$\mathbf{Pr}(\boldsymbol{R}|\boldsymbol{D}) = \frac{Pr(R \cap D)}{Pr(D)} = \mathbf{0.374}$$

Probabilities—where do they come from? (1)

- **The probability of an event is not directly observable**
- **Probability models help translate data into probabilities.**
  - We can observe if events (e.g., failures) occur (about x failures out of n trials), and the times of failures $(t_1, t_2, t_3, \ldots)$
  - We can count combinations of items or options
  - We can collect measurements, run large sets of simulation models, etc.
  - We can ask experts to assign probabilities ("implicit models")

Probabilities – where do they come from? (2)

- Random Variables may be **Discrete** or **Continuous**.
- Similarly, probability distributions representing the Random Variables are also discrete or continuous
  - Binary outcomes (e.g., success/fail, on/off)
  - Discrete data (e.g., countable items)
  - Continuous quantities (e.g., measured variables)

Probability distributions

- **<u>Definition</u>**: A valid **probability distribution** model for a random variable is a function which assigns a total probability of 1.0 to the set of all points in the sample space S.

$$S = \{x_1, x_2, \ldots, x_k\} \text{ where } k \text{ is either finite or infinite}$$

The probability model (**probability density function**) is then one for which:

$$\mathrm{f}(x_i) \geq 0 \text{ for } i = 1,2,\ldots,k \quad \text{or } -\infty < x < \infty$$

and

$$\sum_{i=1}^{k} f(x_i) = 1 \quad \text{or } \int_{-\infty}^{\infty} f(x)dx = 1 \text{ (i.e., the area under the curve is 1)}$$

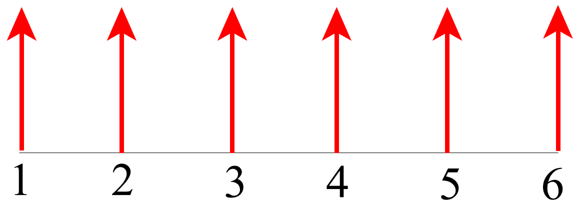
Discrete uniform distribution

- For a sample space $S = \{x_1, x_2, \dots, x_k\}$ with k equally likely outcome; that is, the probability of each outcome is:

$$\mathrm{f}(\mathrm{x_i}) = \Pr(X = x_i) = \frac{1}{k}, \qquad i = 1, 2, \dots, k$$

- Example: Take a die. The probability that any one of the die options $\Omega = \{1,2, \dots, 6\}$ will be rolled is:

$$\Pr(x_i) = \frac{1}{6}$$



Probability density function (PDF)

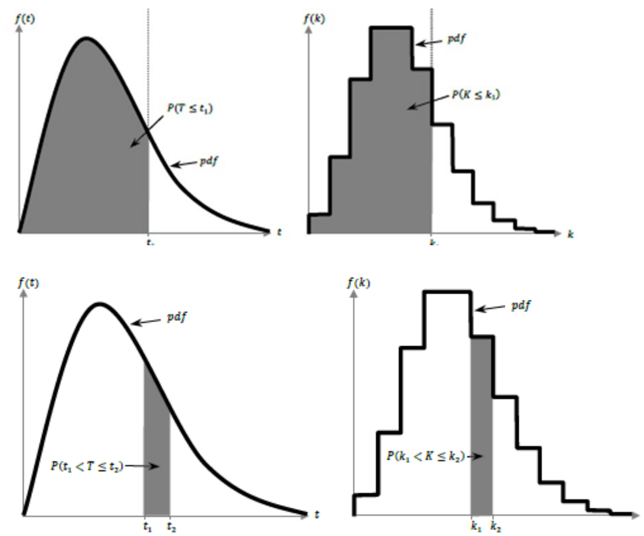
- The function **f(x)** is the **probability density function** (pdf).
  - For discrete distributions, **f(x)** is the **probability mass function** (pmf).
- The pdf/pmf provides point probabilities (discrete r.v.) or point densities (continuous r.v.).
- For a random discrete variable X:

$$\boldsymbol{f(x) = \mathrm{Pr}(X = x)}$$

- For a continuous r.v. X, we use the pdf to solve for the **cdf**:

$$\boldsymbol{Pr(X \le x_i) = \int_{-\infty}^{x_i} f(x)dx = F(x_i)}$$

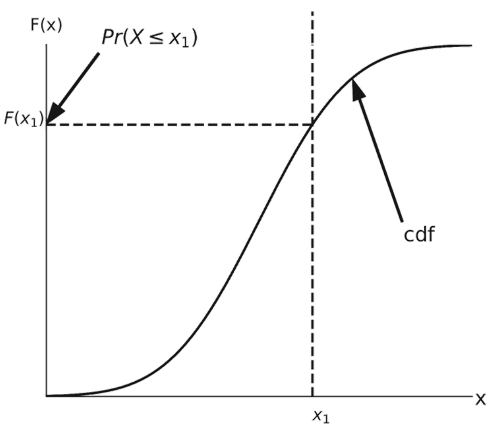
$$\boldsymbol{\mathrm{Pr}(a \le X \le b) = \int_a^b f(x)dx = F(b) - F(a)}$$

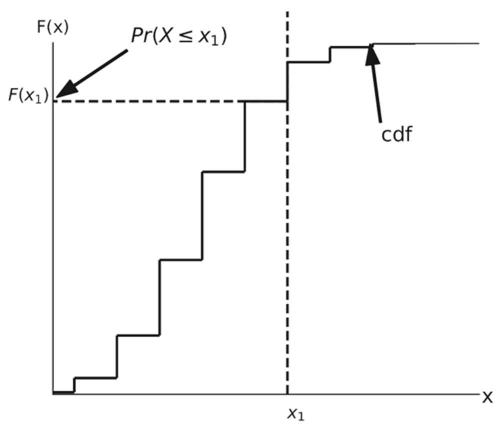


Cumulative distribution function (cdf)

- **Definition**: The **cumulative distribution (density) function** (or **cdf**) for an r.v. X is defined as:

$$F(x) = \begin{cases} \Pr(X \leq x) = \sum_{x_i \leq x} \Pr(x_i) & \text{for discrete r.v. } X \\ \Pr(X \leq x) = \int_{-\infty}^{x} f(t)dt & \text{for continuous r.v. } X \end{cases}$$



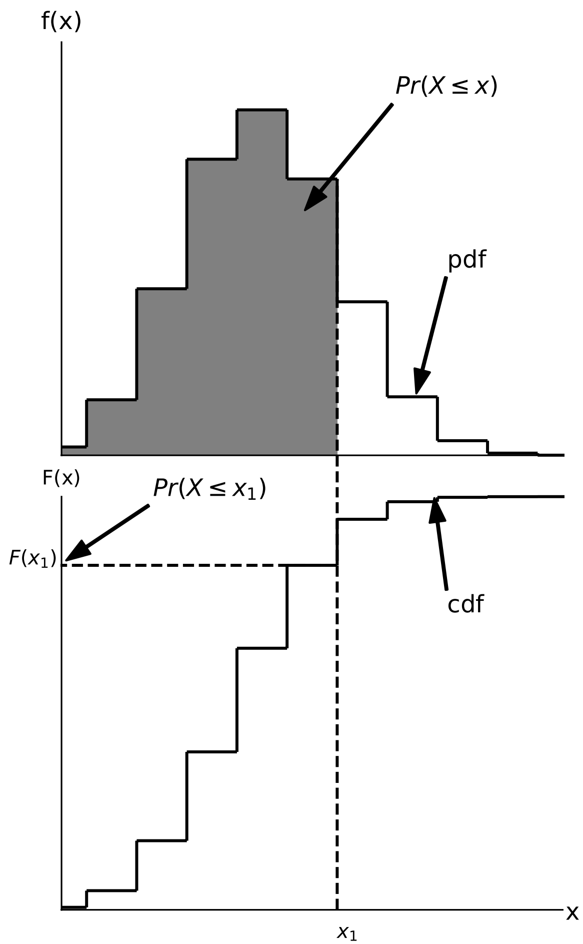


Note:

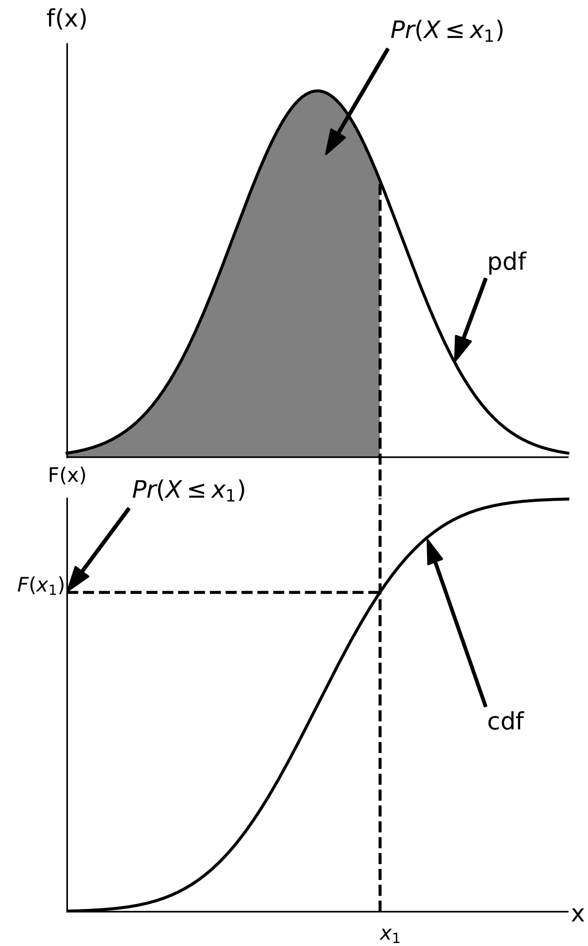
$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

$$\sum_{x \leq \infty} \Pr(x) = 1 \text{ and } \int_{-\infty}^{\infty} f(x)dx = 1$$

The pdf and cdf for discrete (L) and continuous (R) distributions



Discrete pdf (top) and cdf (bottom).

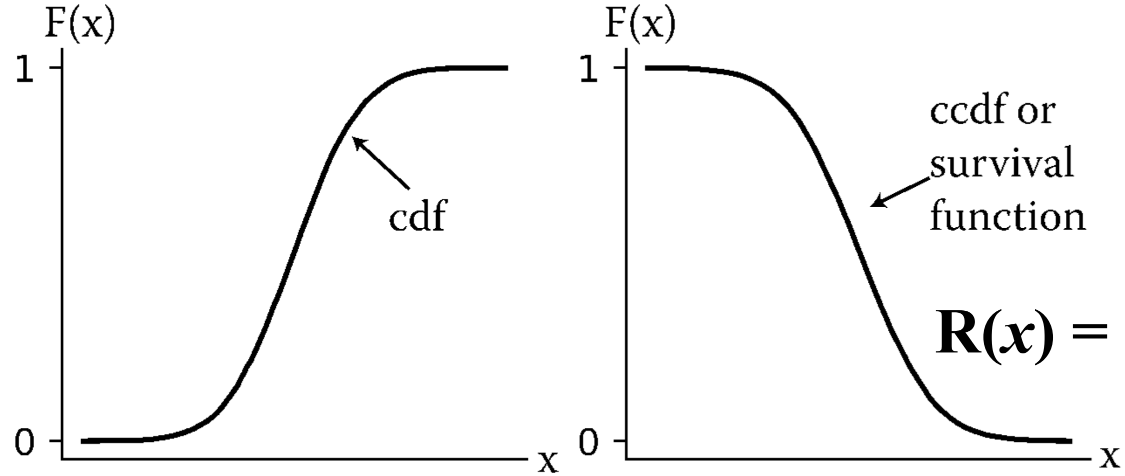


Continuous pdf (top) and cdf (bottom).

Cumulative probability models

- The **complementary cumulative density functions** (or **ccdf**) is also useful especially in instances where a product can fail or survive. This is expressed as:

$$\boldsymbol{R(t) = 1 - F(x)} = \begin{cases} \Pr(X > x) = 1 - \sum_{x_i \le x} \Pr(x_i) & \text{for discrete r.v. } X \\ \int_x^\infty f(t)dt & \text{for continuous r.v. } X \end{cases}$$



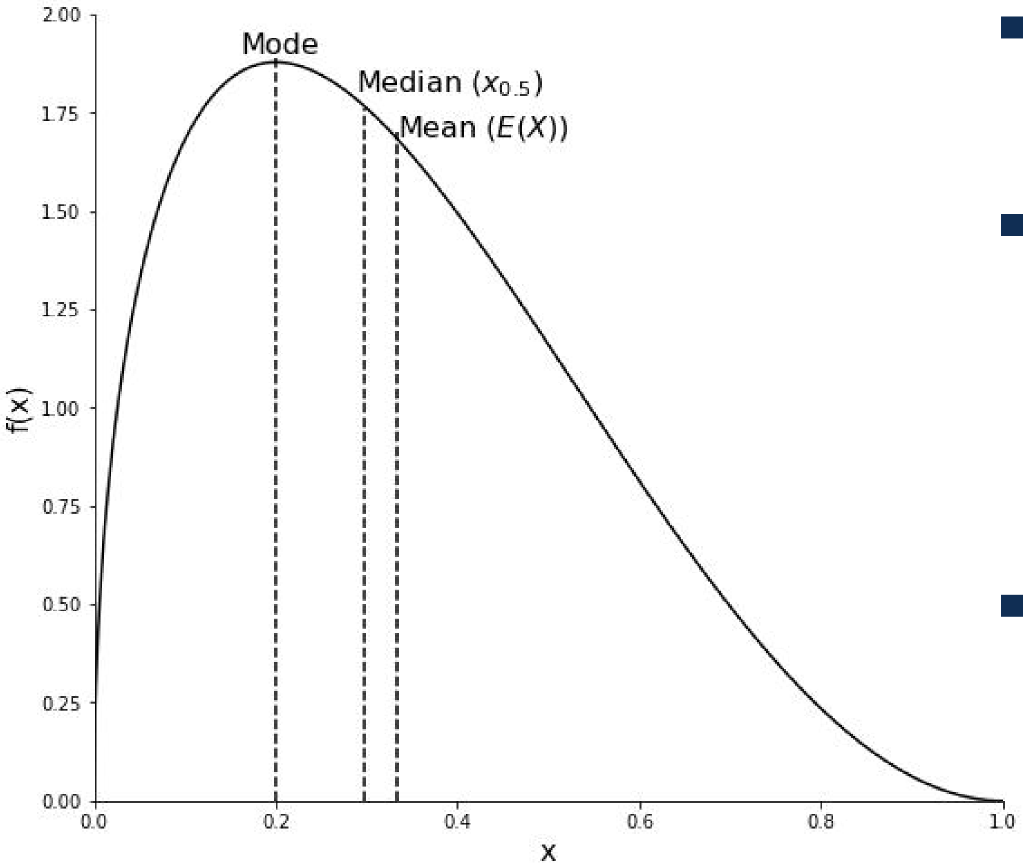
R(x) = 1 – F(x)

Probability models

- Many parametric probability distributions exist – they are families of similarly-shaped distributions which vary based on the defined parameters.

- Many types of parameters in probability distributions:
  - **Location** – shifts the location around x (or t) axis
  - **Scale** (or dispersion) – spread (stretch/shrink)
    - **Rate** (The reciprocal of scale, i.e., 1/scale)
  - **Shape** – defines distribution shape.

Describing distributions: Measures of central tendency



- **Mean**: The expected value of the r.v.: E(X)
- **Median** ($x_{0.5}$): the point where the cdf is equal to 0.5.
  - $x_{0.5} = \mathbf{F^{-1}(0.5)}$
- **Mode**: The highest point of a PDF (the value of the r.v. which has the highest probability of occurrence)

Expected value of X

- For a discrete r.v. X with sample space $\{x_1, x_2, \dots, x_k\}$, and probabilities $\Pr(x_i)$, the expected value of X, E(X) is:

$$E[X] = \sum_{i=1} x_i\, f(x_i)$$

- Similarly for continuous distribution with a pdf of f(x)

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

(Riemann integral)

Expected value: generalized

- The same concepts apply to obtain the expected value of any real-value function of X:
  - For a discrete r.v. X with sample space $\{x_1, x_2, \dots, x_k\}$, if $g(x)$ is a real-valued function of X then expected value of $g(x)$ is defined as:

$$E[g(X)] = \sum_{i=1}^{k} g(x_i) \cdot \Pr(x_i)$$

  - Similarly, for continuous distribution with a pdf of f(x):

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx$$

- And for joint distributions:

$$\mathrm{E}[g(x_1, x_2, \dots, x_n)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n$$

Additional ways to describe distributions

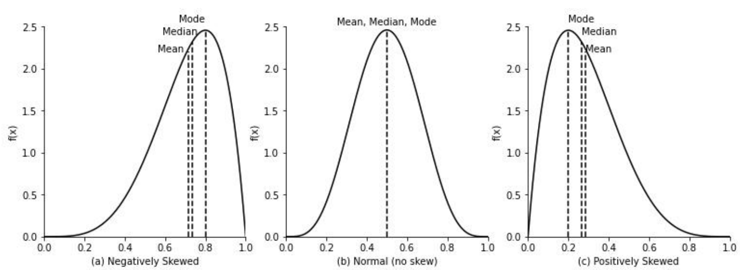
- Some expectation values are given special names and tell us about the shape of the distribution or data.

$\mu_k = E[X^k] \quad k = 1,2, \dots$ $\quad$ k^{th} moment about the origin

$\mu_k = E[(X-\mu_1)^k], \quad k = 2,3, \dots$ $\quad$ k^{th} moment about the mean (central moments)

$\mu'_k = E\left[\left(\frac{X-\mu_1}{\sigma}\right)^k\right], \quad k = 2,3, \dots$ $\quad$ k^{th} standardized moment

- ***We primarily use:***
  - **Mean** (1st raw moment, μ_1), E[X]
  - **Variance** (2nd central moment, μ_2); $\mathrm{Var}(X) = \sigma^2 = E[(X-\mu_1)^2]$
  - **Skewness** (3rd standardized moment, μ'_3); Skew(X) is a measure of the horizontal symmetry (or asymmetry) of the distribution.



  - **Kurtosis** (4th standardized moment, μ'_4) Kurt(X), k, is the measure of whether the distribution is peaked or flat

Example: Calculating expected values

- **Example:** Find E[t] and E[t^2] for t, an exponentially distributed r.v. with $f(x) = \lambda e^{-\lambda t}$
- $E[t] = \int_0^\infty t\lambda e^{-\lambda t} dt$ **Use integration by parts**

$$= -te^{-\lambda t} - \int_0^\infty -e^{-\lambda t} dt$$

$$= -te^{-\lambda t}\Big|_0^\infty - \frac{e^{-\lambda t}}{\lambda}\Big|_0^\infty$$

$$= (0-0) \ - (0 - \frac{1}{\lambda})$$

$$= \frac{1}{\lambda}$$

- Similarly for g(x) = t^2:

$$E[t^2] = \int_0^\infty t^2 \lambda e^{-\lambda t} dt = \frac{2}{\lambda^2}$$

Covariance and correlation

- **<u>Definition</u>**: We define **covariance**, as measure of dependence between two random variables. (how much they vary together)

$$\mathrm{Cov}(X,Y) = \mathrm{E}\{[X-\mathrm{E}(X)][Y-\mathrm{E}(Y)]\}$$
$$= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$$

- Covariance is difficult to interpret, so we normalize it into **correlation**, defined between X_1 and X_2 as:

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(x)\sigma(y)}$$

  - $\pm(0,1)$ correlation coefficient indicates strength of relationship; sign indicates direction of relationship. 0 indicates independence.

Combinations: The binomial coefficient

- A specialized way of counting is "Combinations" defined as combination of ***r*** items that can be selected from ***n*** distinct items without replacement (**order not important**)

$$C_r^n = \binom{n}{r} = \frac{n!}{r!\,(n-r)!} \qquad r = 1,2,\ldots,n$$

- Definition: $1! = 1, \quad 0! = 1$

- **Example:** From 20 light bulbs how many samples of size 4 can be selected?

$$C_4^{20} = \binom{20}{4} = \frac{20!}{4!\,(20-4)!} = 4845$$

Binomial distribution, *X~binom(n,p)*

- Applies to a series of n independent trials where the random variable X may take binary values (e.g., success or failure, on or off, heads or tails) with probability p.

- The probability that outcome #1 occurs exactly x times out of n trials is given by:

$$f(x|n,p) = \begin{cases} \Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & otherwise \end{cases}$$

where,

- p =probability of outcome #1
- n =size of the sample space
- In Excel: binom.dist(x,n, p, false)
- In Matlab: Binopdf(x,n,p)

Binomial CDF

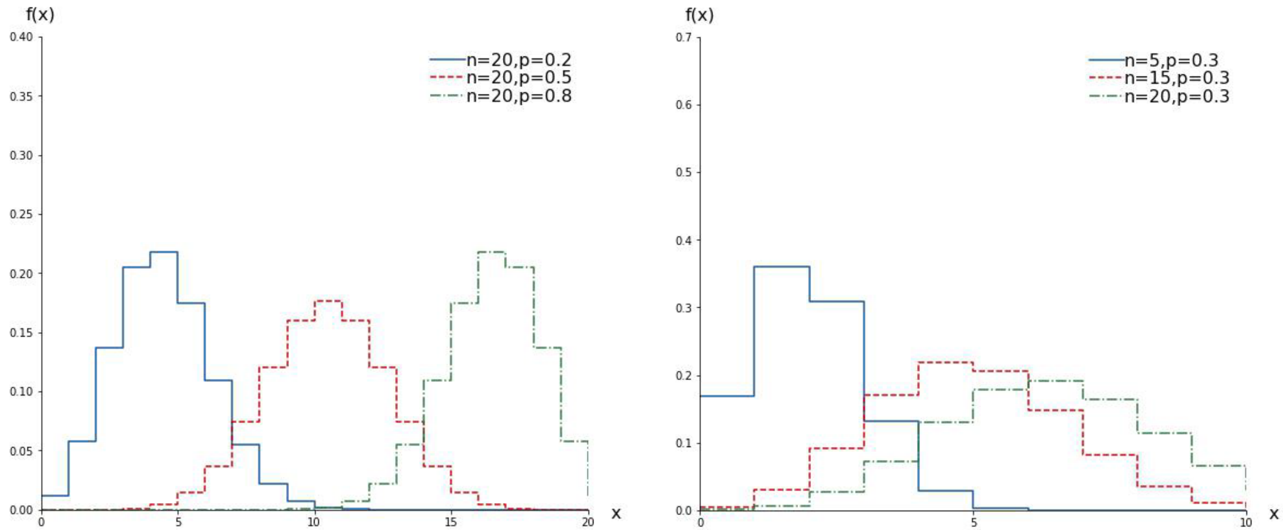
- The **Binomial cdf is**

$$F(x|n,p) = \begin{cases} 0 & x < 0 \\ \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i} & x = 0,1,2,\ldots,n \\ 1 & x \geq n \end{cases}$$

In Excel: binom.dist(x,n, p, true)
In Matlab: binocdf
Excel is always: (false for PDF, true for CDF)

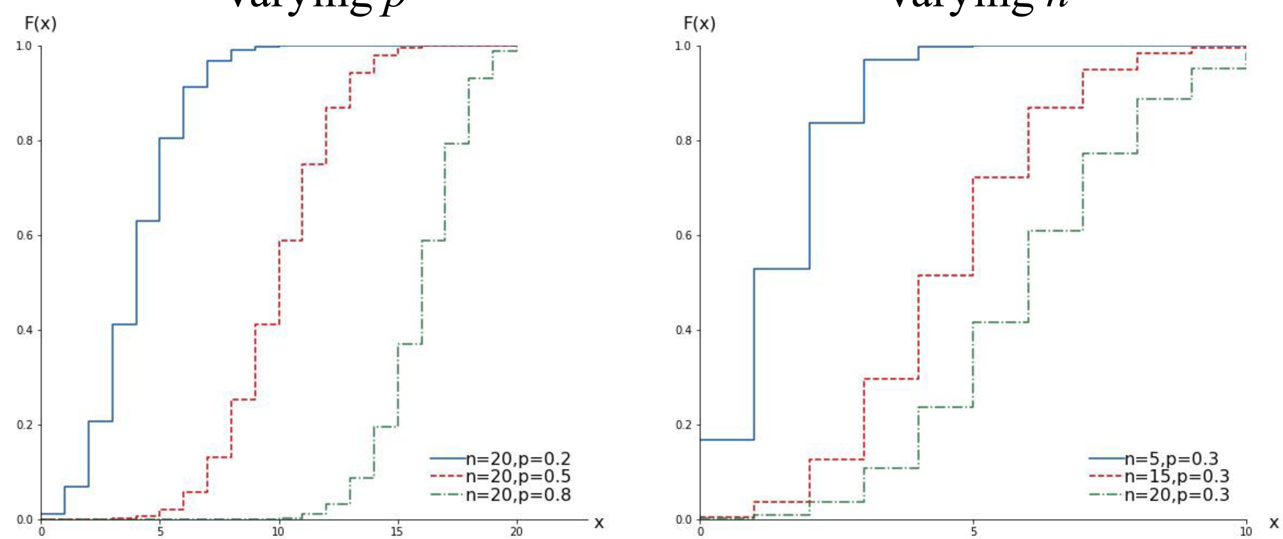
Binomial distribution (cont.)

Binomial pdf:



Varying *p* Varying *n*

Binomial cdf:



Binomial distribution

- Example: Assume p, the probability of a light bulb surviving until the end of a specified mission, is given as: $p = 0.9$. For a set of 4 randomly selected light bulbs, calculate probability of:
- A) Exactly 0, 1, 2, 3, or 4 survivals.
- B) 2 or more survivals.

Binomial distribution

- **Solution**: Assume p =probability of survival (success) of light bulb in a mission. That is $p = 0.9$. For $n = 4$ light bulbs, calculate probability of X=0, 1, 2, 3, or 4 survivals.

A)

x	$\Pr(x)$	$F(x) = \Pr(X \le x)$
0	$\binom{4}{0}(0.9)^0(0.1)^4 = \mathbf{0.0001}$	0.0001
1	$\binom{4}{1}(0.9)^1(0.1)^3 = \mathbf{0.0036}$	0.0001+0.0036=0.0037
2	$\binom{4}{2}(0.9)^2(0.1)^2 = \mathbf{0.0486}$	0.0001+0.0036+0.0486=0.0523
3	$\binom{4}{3}(0.9)^3(0.1)^1 = \mathbf{0.2916}$	0.3439
4	$\binom{4}{4}(0.9)^4(0.1)^0 = \mathbf{0.6561}$	1.0
Total	1.0000	

B) $\mathbf{Pr(X \ge 2)}$ =Pr(2)+Pr(3)+Pr(4) = **0.9963**

$= 1 - \Pr(X<2)$=1-.0037=0.9963

Poisson distribution, *X~poiss(μ)*

- **Definition**: The **Poisson Distribution** expresses probability of a given number of events (X) in a time or space, for events that occurs with a constant rate (or intensity), λ & independently of time since last event.
- **Example**: One can think of X as the number of electronic components that fail per unit of time or number of telephone calls coming in per unit of time.

$$f(x|\mu) = \Pr(X = x) = \begin{cases} \dfrac{\mu^x e^{-\mu}}{x!} & \mu > 0, x = 0,1,2,\ldots \\ 0 & otherwise \end{cases}$$

  - x – The number of times the event takes place
  - $\mu = \lambda t$ is the rate parameter. It's also the expected (mean) number of occurrences in a fixed time period, t.
    - λ – The (constant) event rate or intensity (Units: 1/time)
    - t – The unit of time or space (Units: time)
  - In Excel: poisson.dist(x, μ, false)

Poisson CDF

- The **Poisson cdf is**

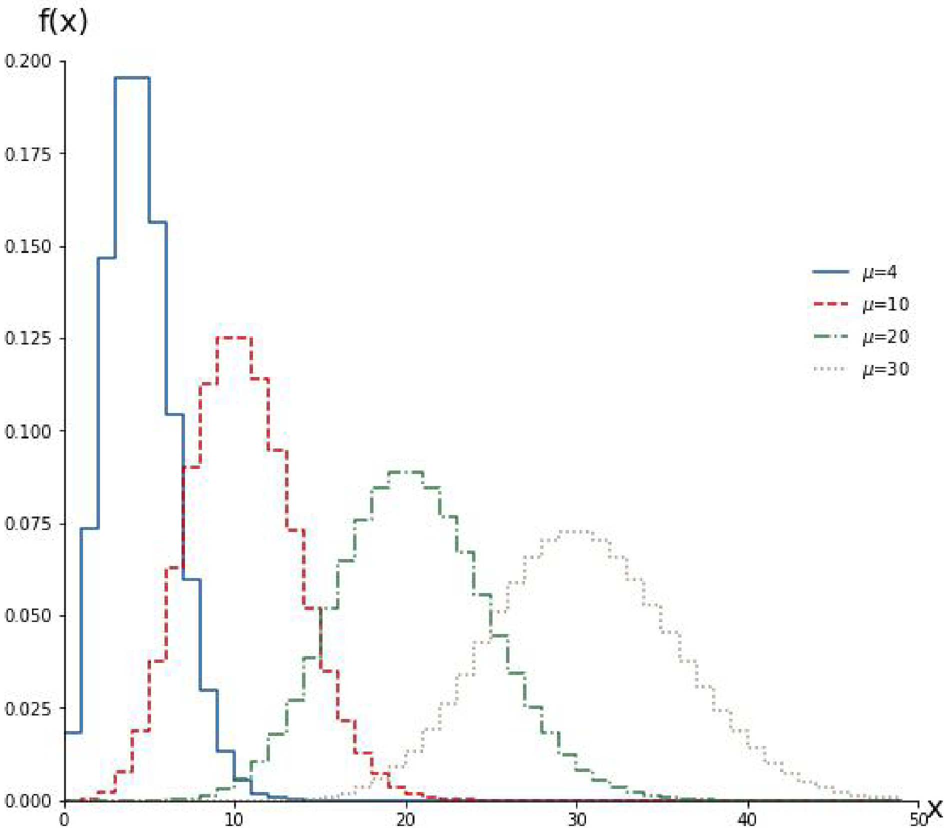
$$F(x|\mu) = \begin{cases} e^{-\mu} \sum_{i=0}^{x} \frac{\mu^i}{i!} & \mu > 0, x = 0,1,2,\ldots \\ 0 & otherwise \end{cases}$$

Matlab function poisscdf(x, μ) to find it.

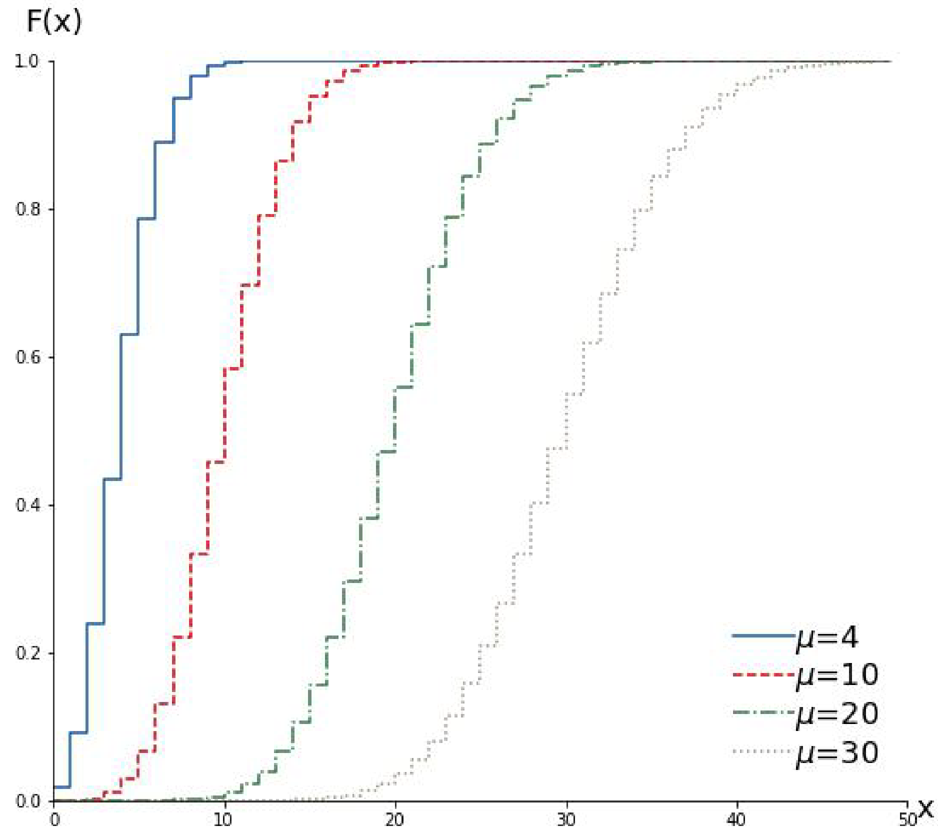
Note: If X is a binomially distributed variable with small p (say $p \leq 0.1$) the Poisson distribution can be used with $\mu = np$.

Poisson distribution (cont.)

pdf



cdf



Poisson distribution: Example 1

- **Example:** Components needed for operation of a machine fail according to a Poisson process with rate of λ = 1 failure/week. What is the probability that exactly 1 component fails during a 4-week period?
  - The expected number of events in 4 weeks is calculates to:

$$\mu = \left(\frac{1}{\text{week}}\right) 4 \text{ weeks} = 4$$

  - The probability for $X = 1$ component failing when $\mu = 4$ is:

$$\Pr(X = 1) = \frac{4^1 e^{-4}}{1!} = 0.0733$$

- **Example part 2:** For this machine, one spare component is available. What is the probability that the machine works for a given 6 weeks?

Poisson distribution: Example 1

- **Example (Part 2)**: For that machine, one spare component is available. What is the probability that the machine works for a given 6 weeks?
  - Here you are solving for X = number of failures

$$\Pr(X \leq 1) = \Pr(X = 0 \ \cup \ X = 1)$$
$$= \Pr(X = 0) + \Pr(X = 1)$$
$$\text{With } \mu = 6$$
$$= e^{-6} + 6e^{-6} = \mathbf{0.0174}$$

Poisson distribution: Example 2 (after class)

Example: Cars typically arrive at a certain stop sign at a rate λ of 3 cars per minute.

- What is the probability that no car arrives in a 10-minute interval?

- What is the probability that more than 20 cars arrive in a 10-minute interval?

Poisson distribution: Example 2

Example: Cars typically arrive at a certain stop sign at a rate λ of 3 cars per minute.

- What is the probability that no car arrives in a 10-minute interval?

$$\mu = \lambda t = \left(3\,\frac{\text{cars}}{\text{minute}}\right) 10 \text{ minutes} = 30 \text{ cars}$$

$$\Pr(x = 0 \text{ cars}|\lambda\text{t} = 30) = \frac{30^0 e^{-30}}{0!} = \boldsymbol{e^{-30}} = \mathbf{9.36 \times 10^{-14}}$$

Excel (poisson.dist(0,30,false))

- What is the probability that more than 20 cars arrive in a 10-minute interval?

$$\Pr(x > 20 \text{ cars}|\lambda\text{t} = 30) = 1 - \Pr(x \leq 20)$$
$$= 1 - \Pr(x = 20) - \Pr(x = 19) - \cdots - \Pr(x = 0)$$

Use Matlab poisscdf(x, μ) or Excel (poisson.dist(20,30,true) to find

$$Pr(X \leq 20) = 0.0352$$

$$1 - Pr(X \leq 20) = \Pr(x > 20 \text{ cars}|\lambda t = 30) = \mathbf{0.9647}$$

Exponential distribution $X \sim \exp(\lambda)$

- **Definition**: The **exponential distribution** is often used to model time-to-failure of unrepairable components and systems because of its mathematical simplicity. The pdf for the exponential distribution is:

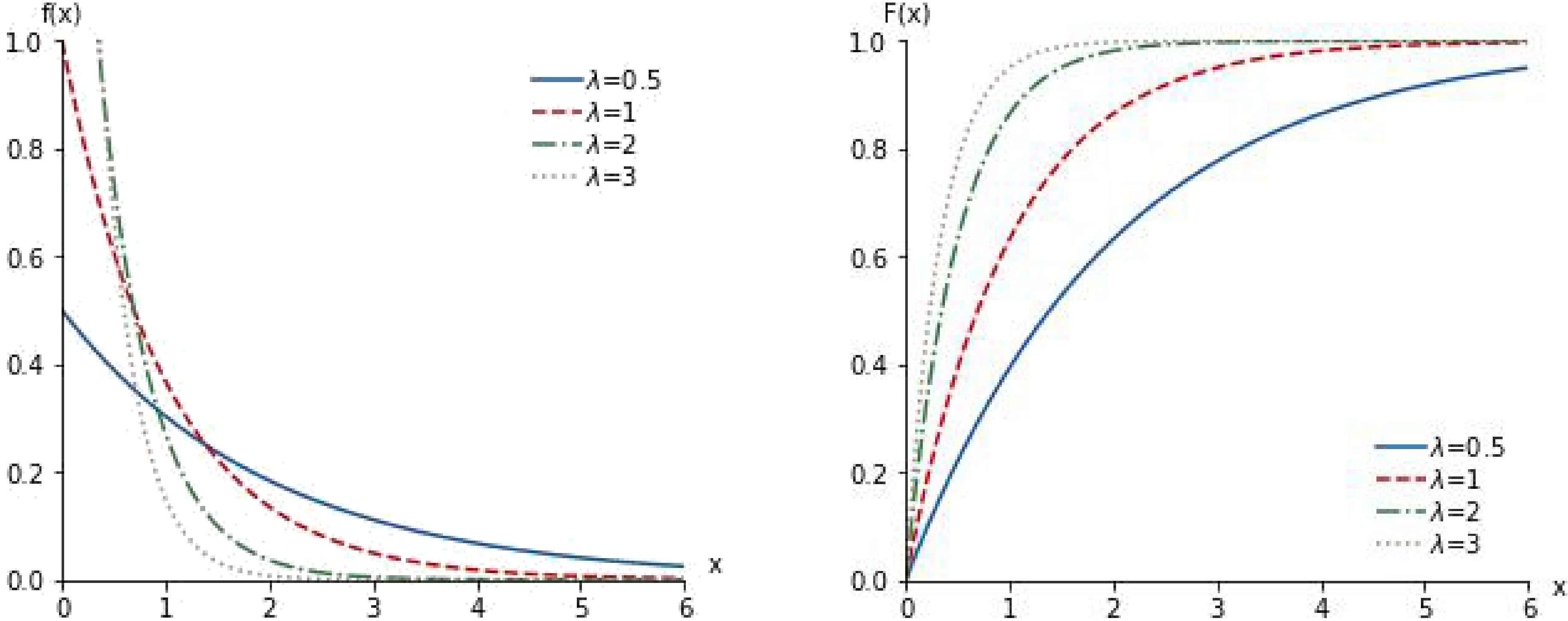
$$f(x|\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

  Where:
  - λ – Rate (1/scale) parameter ($\lambda > 0$);

- The cdf for the exponential distribution is:

$$F(x|\lambda) = \begin{cases} 0 & x \leq 0 \\ \int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x} & x > 0 \end{cases}$$

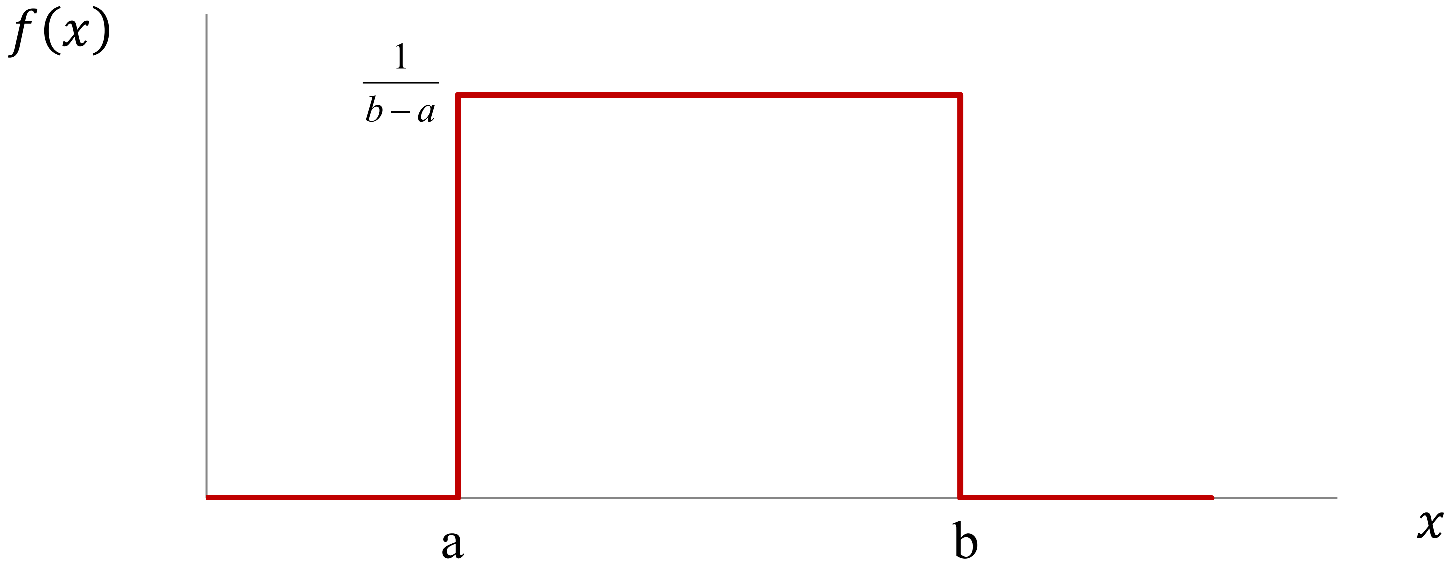
Exponential distribution X~exp(λ)



Uniform distribution *X~unif(a,b)*

- **Definition**: There is a continuous form of the **uniform distribution** sometimes called the **rectangular distribution**.

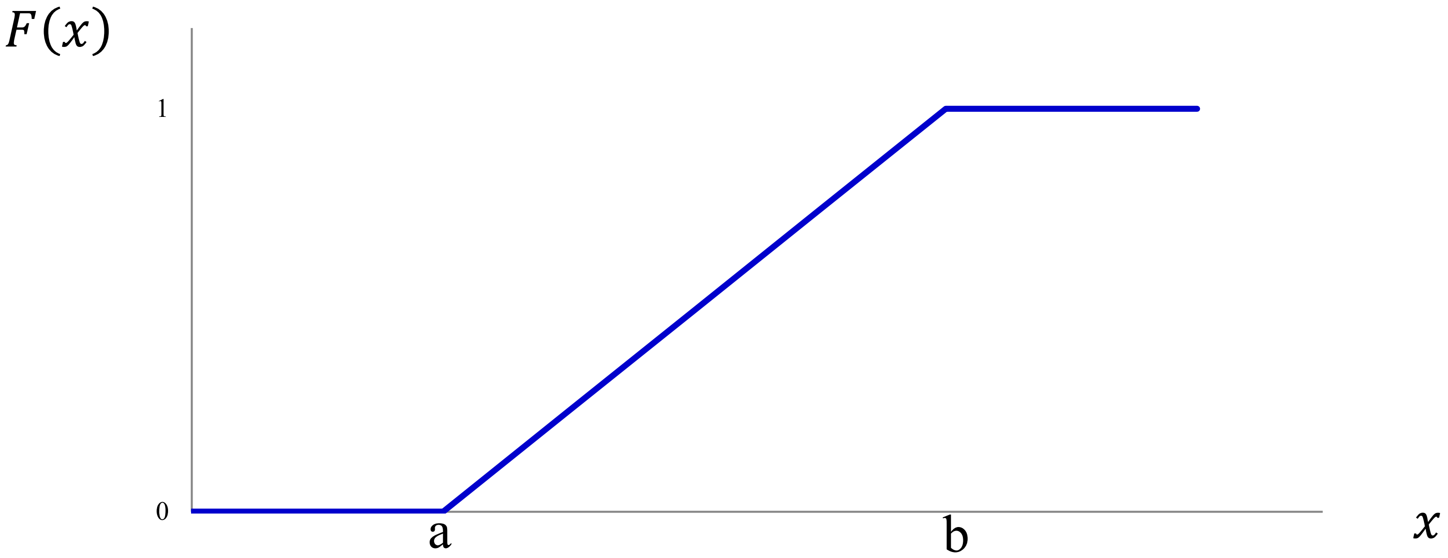
$$f(x|a,b) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$



Continuous uniform distribution

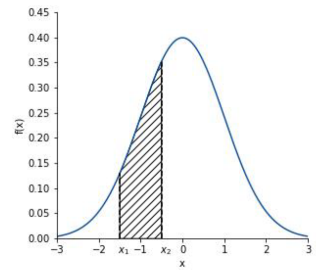
- **<u>Definition</u>**: The **cdf of the continuous uniform distribution** is

$$F(x|a,b) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$



Normal distribution: $X \sim norm(\mu, \sigma)$

- The most well known and widely used pdf. The normal pdf sometimes called Gaussian distribution or Bell curve.
  - **Central Limit Theorem:** Means of non-normal variables are approximately normally distributed
  - The **pdf for a normal distribution** is:



$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty$$

$$= \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$$

Where:

- μ is the mean (location parameter) $-\infty < \mu < \infty$
- σ is the standard deviation (scale parameter) and $\sigma > 0$
- (and σ^2 is the variance)
- In Excel: = norm.dist(x, μ, σ, false)

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\left[-\frac{1}{2}z^2\right]}$$

$\phi(z)$ is the standard normal distribution, where:

$$\boldsymbol{z} = \frac{x-\mu}{\sigma}$$

Normal Distribution

- The **cdf for a normal distribution** is:

$$F(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right]}dt$$

$$= \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right) = \Phi(z)$$

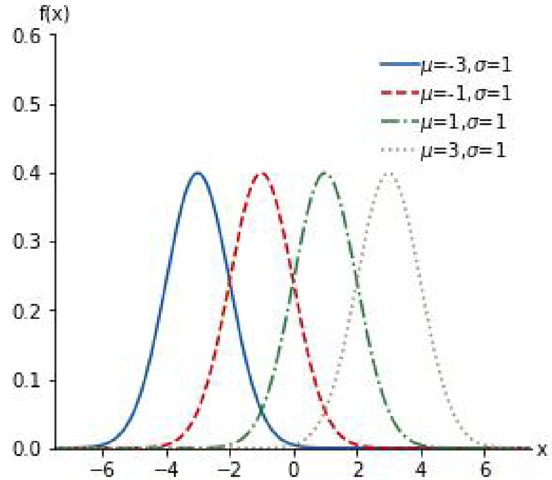
Where Φ is the standard normal CDF with $\mu = 0$ and $\sigma = 1$:

$$\Pr(Z \leq z) = \Phi(z)$$

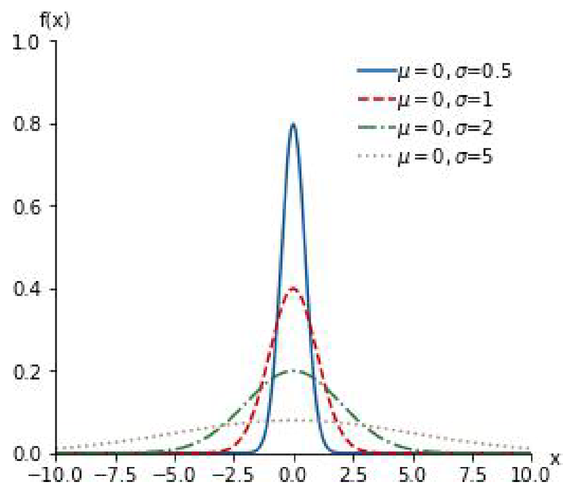
- All normal pdfs and cdfs can be transformed to standard form using the z transform: $Z = \frac{X-\mu}{\sigma} \rightarrow z = \frac{x-\mu}{\sigma}$
- **Appendix A** of the book gives a standard normal CDF table.

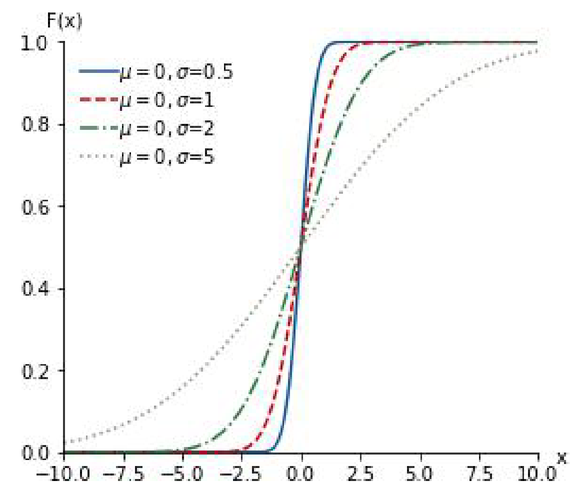
Normal distribution pdf, cdf

Varying μ and constant σ.



Varying σ and constant μ.





Normal distribution example

Example: Given a normal pdf with $\mu = 50$ and $\sigma = 10$. Find the probability that X is between 48 and 62, $\Pr(48 < x < 62)$.

Normal distribution example (cont.)

Example: Given a normal pdf with μ = 50 and σ = 10. Find the probability that *X* is between 48 and 62, $\Pr(48 < x < 62)$.

$$z_1 = \frac{48 - 50}{10} = -0.2 \quad z_2 = \frac{62 - 50}{10} = 1.2$$

$$\Pr(48 < x < 62) = \Pr(-0.2 < z < 1.2)$$

$$\Pr(z = 1.2) - \Pr(z = -0.2) = 0.885 - 0.421 = 0.464$$

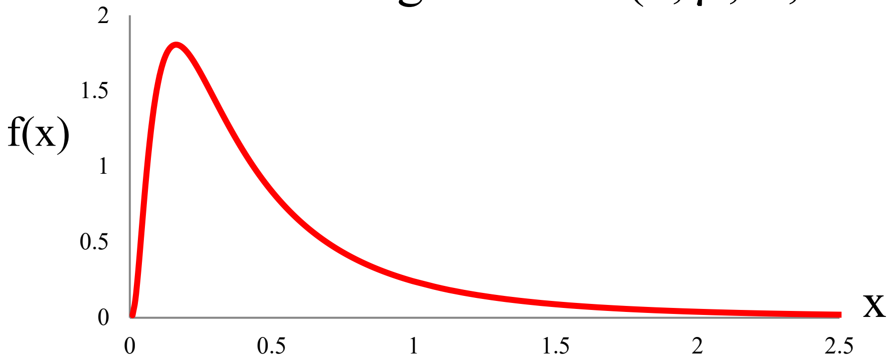
After class: solve using software and by hand

Lognormal distribution $X \sim lognorm(\mu, \sigma)$

- Definition: When the r.v. is defined as always positive it is said to be lognormally distributed if its ln is normally distributed.
  - That is: if x~lognorm, then ln(x) ~ norm.
- In reliability engineering it is used to represent cycles-or-time-to failure in fatigue. Generally used for product of a large number of i.i.d variables.

$$f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \mathrm{e}^{\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right]} = \frac{1}{x\sigma}\phi\left(\frac{\ln x - \mu}{\sigma}\right) \qquad x > 0$$

  - μ is the scale parameter ($-\infty < \mu < \infty$); calculated as E(ln(x)))
  - σ is the shape parameter ($\sigma > 0$); calculated as stdev(ln(x))
- In Excel: =lognorm.dist(x, μ, σ, false); Can also use z-transform



$$\phi(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\left[-\frac{1}{2}z^2\right]}$$

$$\boldsymbol{z} = \frac{\ln(x) - \mu}{\sigma}$$

Lognormal distribution (cont.)

- Thus, the cdf for a lognormal distribution is,

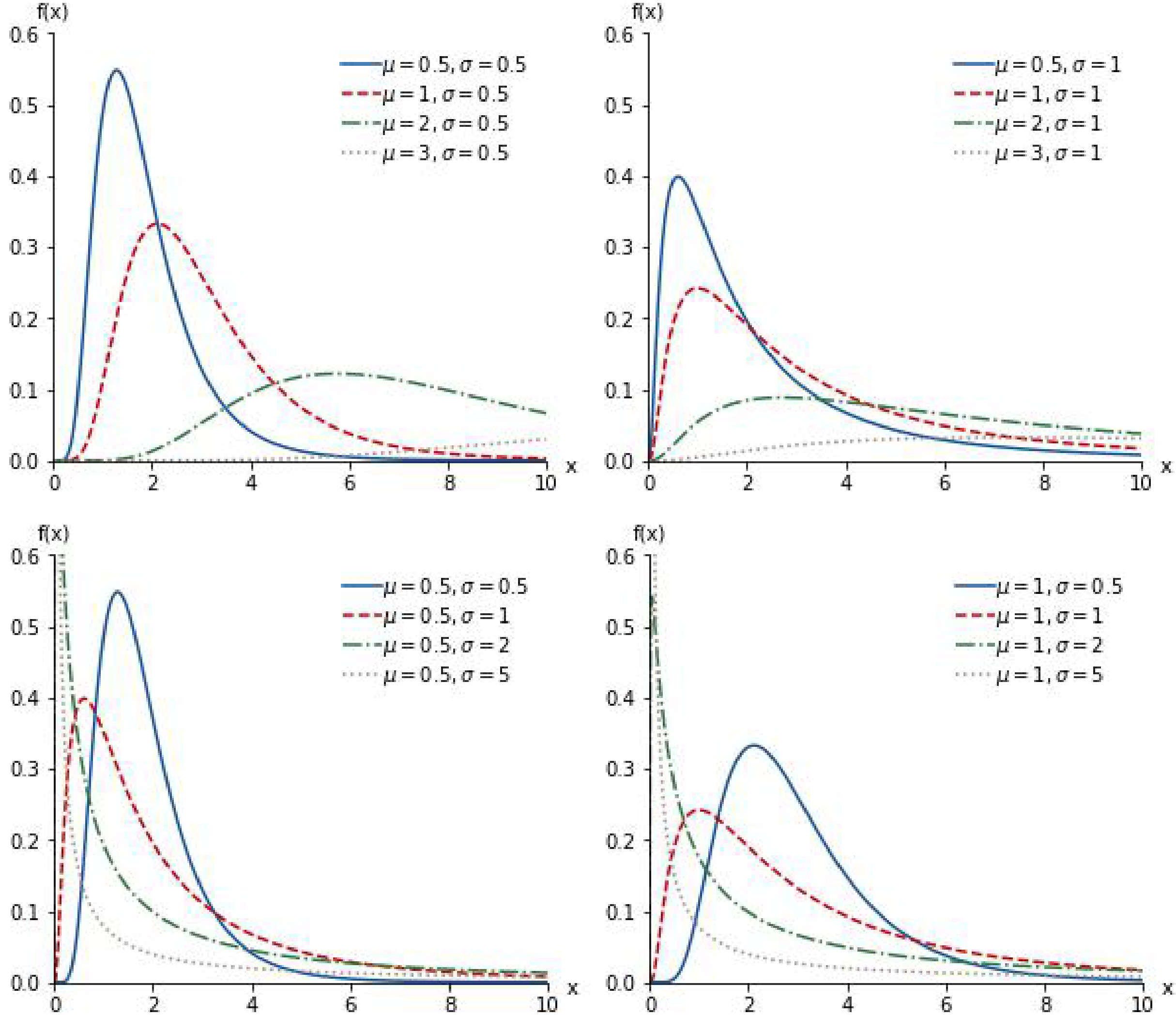
$$F(x|\mu,\sigma) = \int_0^x f(t)dt = \int_0^x \frac{1}{t\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{lnt-\mu}{\sigma}\right)^2\right]} dt$$

$$F(x) = \Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)$$

Lognormal distribution pdf

Varying μ and constant σ.

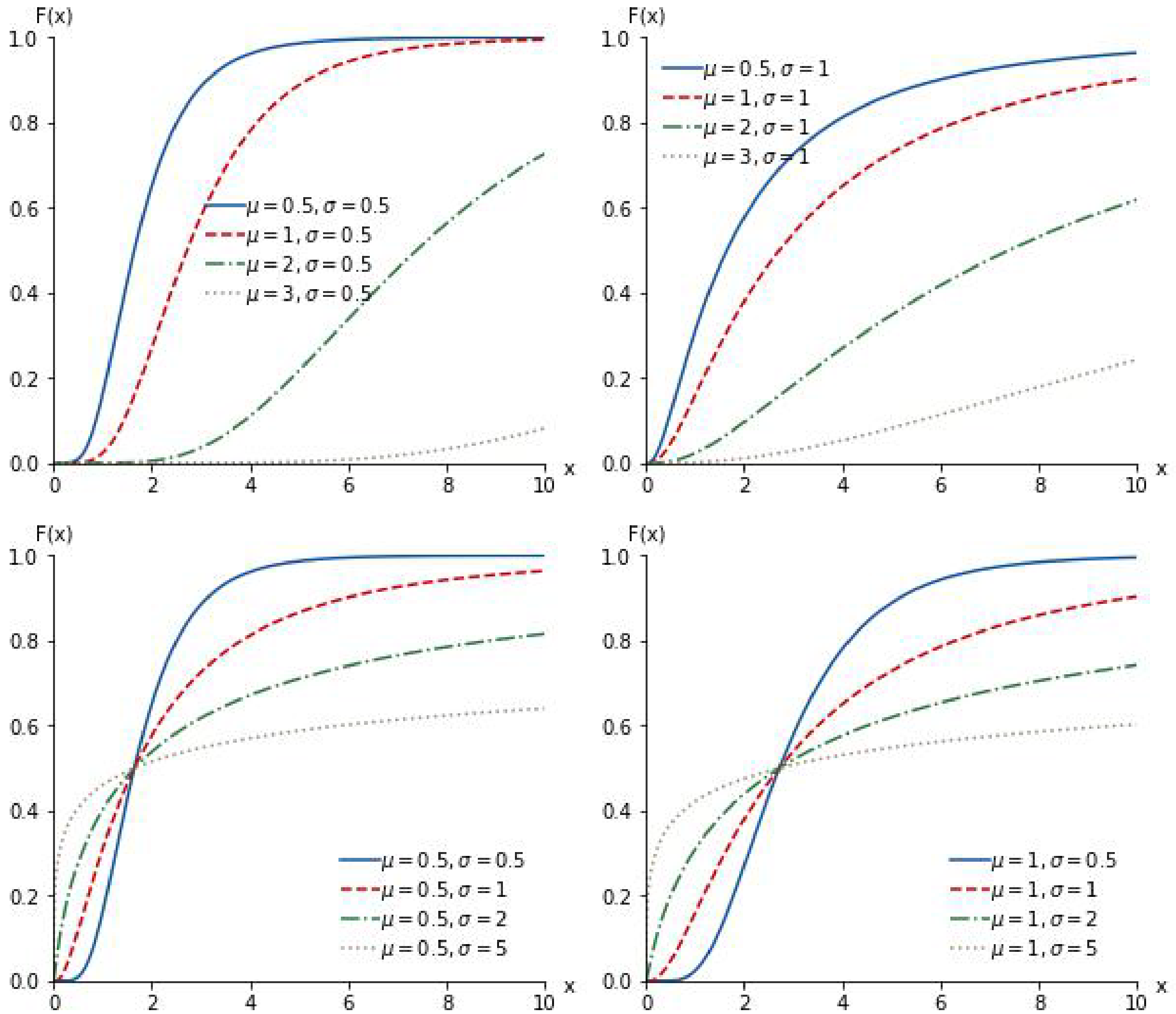
Varying σ and constant μ.



Lognormal distribution cdf

Varying μ and constant σ.

Varying σ and constant μ.



Lognormal distribution

Example: Given $\mu=1$ and $\sigma=1$, find the probability that lognormally distributed r.v. X will fall between 0.5 and 3.5.

Lognormal distribution

Example: Given μ=1, σ=1, find the probability that lognormally distributed *X* will fall between 0.5 and 3.5.

In Excel = lognorm.dist(3.5, 1, 1, true) – lognorm.dist (0.5, 1, 1, true)

$$z_1 = \frac{ln0.5 - 1}{1} = -1.69 \qquad z_2 = \frac{ln3.5 - 1}{1} = 0.25$$

By hand, using Appendix A tables:

$$\Pr(0.5 < X < 3.5) = \Pr(-1.69 < Z < 0.25)$$

$$= \Phi(0.25) - \Phi(-1.69) = 0.59871 - 0.04551 = \mathbf{0.553}$$

(Using software without rounding: **=0.5545**)

Weibull distribution $X \sim weibull(\alpha, \beta)$

- **Definition**: The **Weibull distribution** is a very flexible and popular distribution that can be used in a variety of situations. Primarily for failure or life-based reliability problems.

$$f(x|\alpha,\beta) = \frac{\beta x^{\beta-1}}{\alpha^{\beta}} \mathrm{e}^{\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right]} \quad x \geq 0$$

where,
- α is the scale parameter ($\alpha > 0$)
- β is the shape parameter ($\beta > 0$)

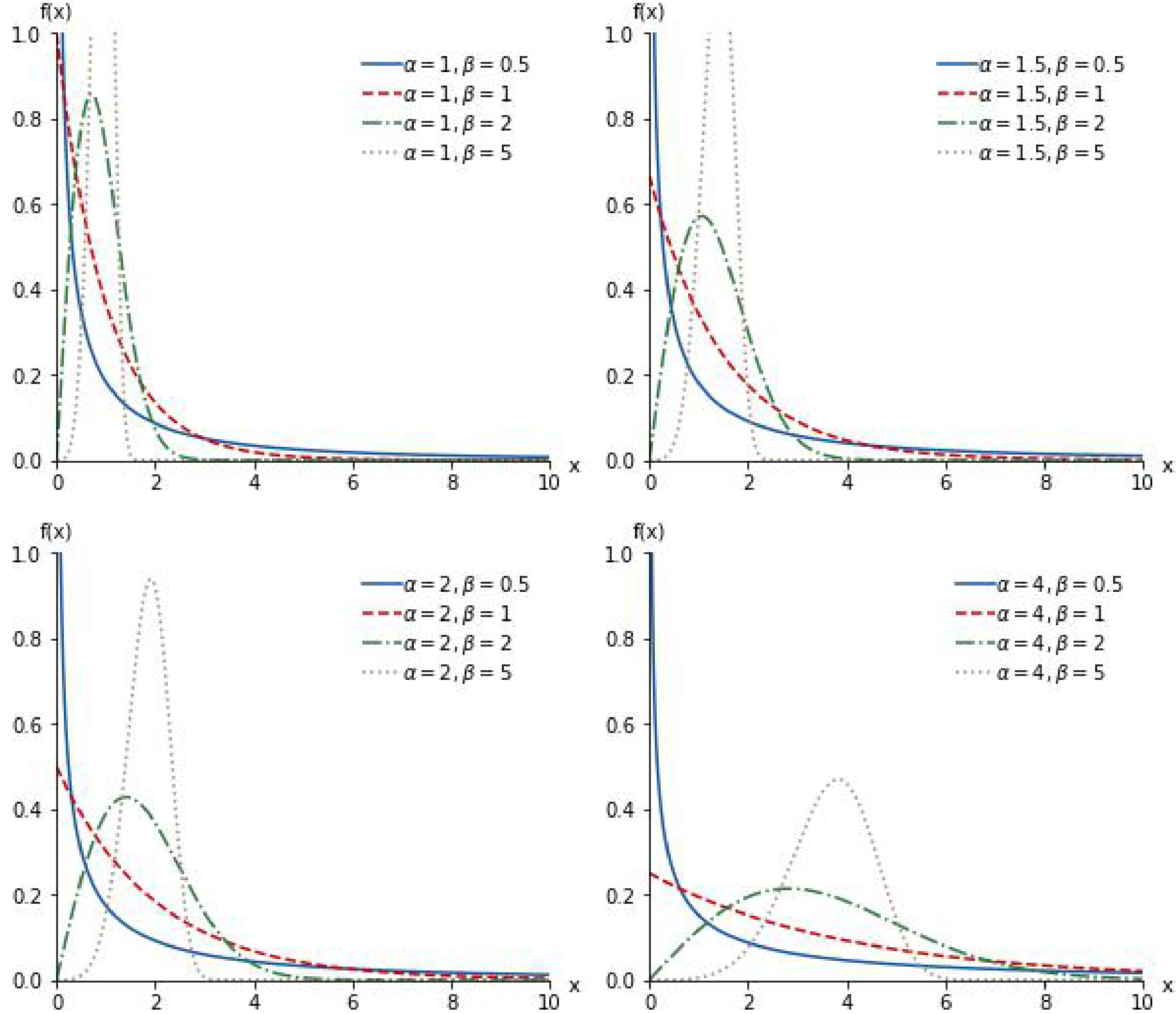
Weibull distribution

Example: For a Weibull distribution with the scale parameter α and shape parameter β, the cdf is

$$F(x|\alpha,\beta) = \begin{cases} 1 - e^{\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right]}, & x, \alpha, \beta > 0 \\ 0 & \text{otherwise} \end{cases}$$

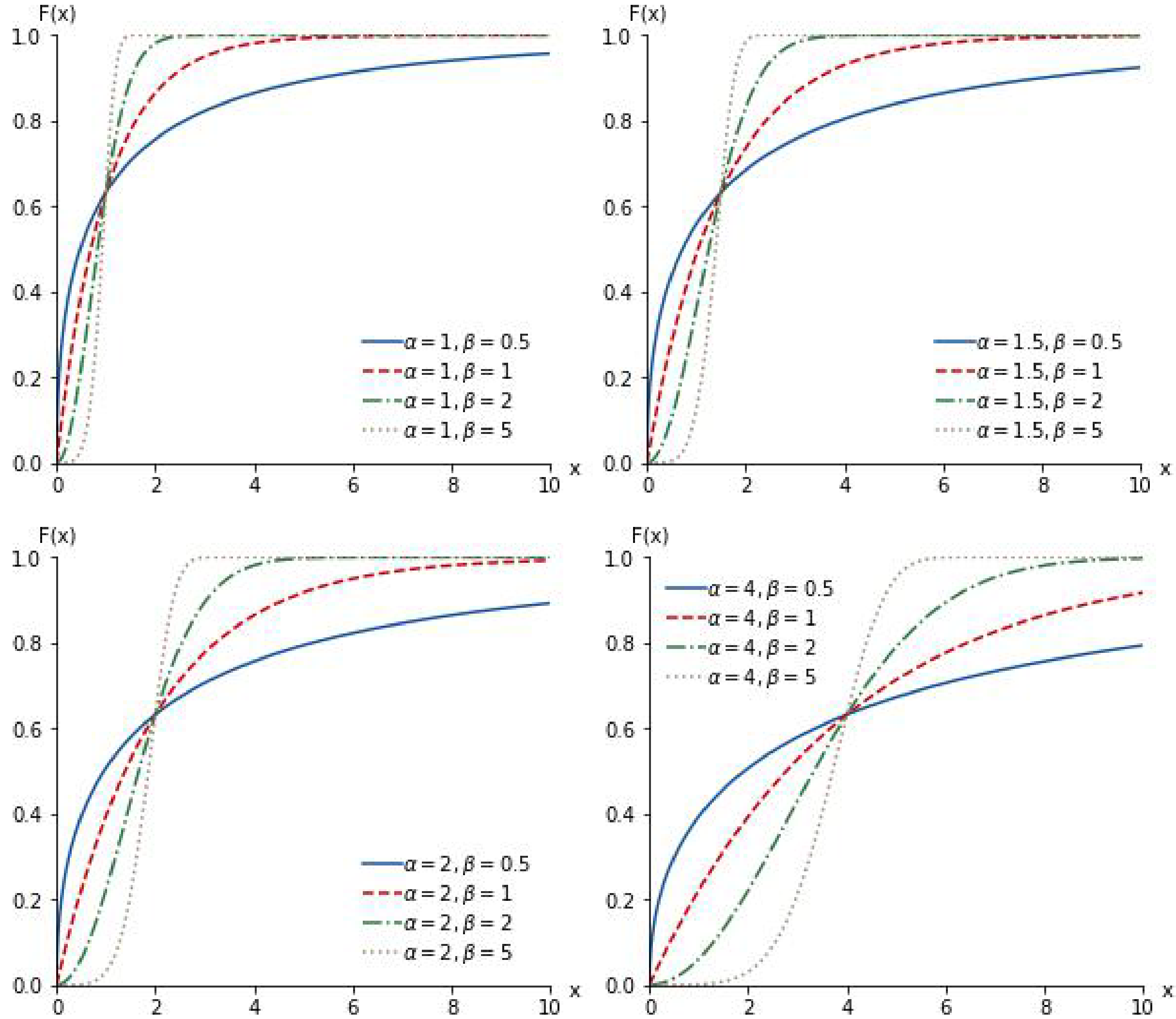
Weibull distribution pdf

Each frame has constant α and varying β.



Weibull distribution cdf

Each frame has constant α and varying β.



Weibull distribution

Example: When the shape parameter $\beta = 1$, determine what other pdf this distribution becomes and what is its governing parameter value?

$$f(x|\alpha,1) = \frac{1x^{1-1}}{\alpha^1}\mathrm{e}^{\left[-\left(\frac{x}{\alpha}\right)^1\right]} = \frac{1}{\alpha}\mathrm{e}^{-\frac{x}{\alpha}}$$

- What does this remind you of? Compare this to

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

- So, the Weibull pdf with $\beta = 1$ becomes an exponential distribution with the parameter $\lambda = \frac{1}{\alpha}$.

Gamma distribution $X \sim gamma(\alpha, \beta)$

- **Definition**: The **gamma distribution** is a generalized form of the exponential distribution. It represents the sum of α independent exponential variables. However, it bears a closer similarity to the Poisson distribution. Two forms depending on the shape parameter:

$$f(x|\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{\left(-\frac{x}{\beta}\right)} & \alpha \text{ is continuous} \\ \dfrac{1}{\beta^{\alpha}(\alpha-1)!} x^{\alpha-1} e^{\left(-\frac{x}{\beta}\right)} & \alpha \text{ is an integer} \end{cases}$$

where,
- α – shape parameter and $\alpha > 0$
- β – scale parameter and $\beta > 0$
- Range – $x \geq 0$

The mean and variance is given as: $\mathrm{E}(X) = \alpha\beta$ and $Var(X) = \alpha\beta^2$

In Excel: Gamma.dist(x, α, β, *false*)

The chi squared distribution is a gamma distribution with $\beta = 2,\ \alpha = df/2$ where df is the number of degrees of freedom.

Gamma function (used in gamma distribution)

- **Definition**: The $\Gamma(\alpha)$ term in the continuous form is known as the **gamma function** which is given as,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{(-x)} dx$$

- Values of the function are listed in the following Table.
  - Alternately, in both Excel and Matlab enter gamma(α)

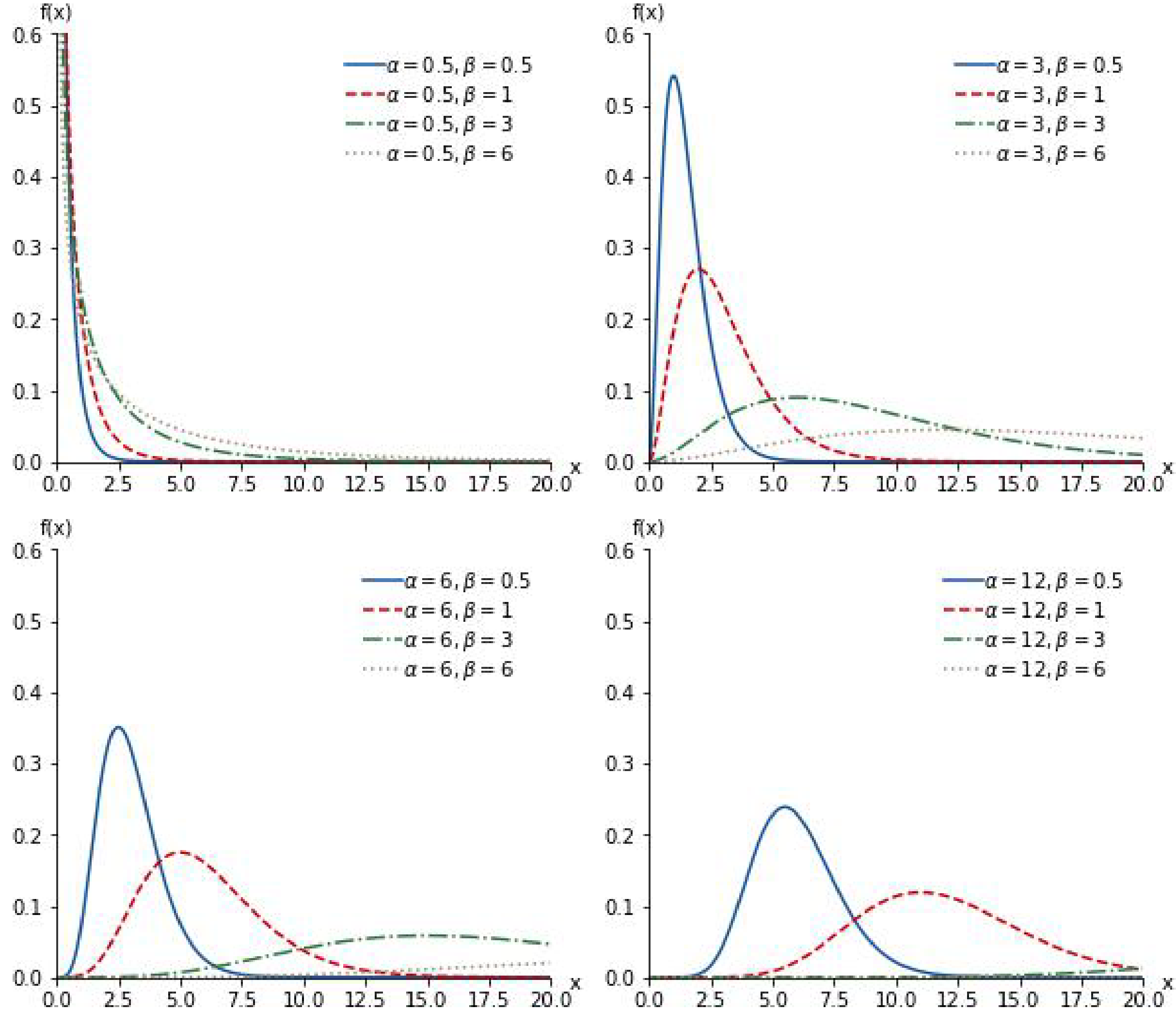
Gamma function table

α	$\Gamma(\alpha)$	α	$\Gamma(\alpha)$	α	$\Gamma(\alpha)$	α	$\Gamma(\alpha)$	α	$\Gamma(\alpha)$
1.0000	1.0000	1.4000	0.8873	1.8000	0.9314	3.0000	2.0000	5.0000	24.0000
1.0200	0.9888	1.4200	0.8864	1.8200	0.9368	3.1000	2.1976	5.1000	27.9318
1.0400	0.9784	1.4400	0.8858	1.8400	0.9426	3.2000	2.4240	5.2000	32.5781
1.0600	0.9687	1.4600	0.8856	1.8600	0.9487	3.3000	2.6834	5.3000	38.0780
1.0800	0.9597	1.4800	0.8857	1.8800	0.9551	3.4000	2.9812	5.4000	44.5988
1.1000	0.9514	1.5000	0.8862	1.9000	0.9618	3.5000	3.3234	5.5000	52.3428
1.1200	0.9436	1.5200	0.8870	1.9200	0.9688	3.6000	3.7170	5.6000	61.5539
1.1400	0.9364	1.5400	0.8882	1.9400	0.9761	3.7000	4.1707	5.7000	72.5276
1.1600	0.9298	1.5600	0.8896	1.9600	0.9837	3.8000	4.6942	5.8000	85.6217
1.1800	0.9237	1.5800	0.8914	1.9800	0.9917	3.9000	5.2993	5.9000	101.2702
1.2000	0.9182	1.6000	0.8935	2.000	1.000	4.0000	6.0000	6.0000	120.0000
1.2200	0.9131	1.6200	0.8959	2.1000	1.0465	4.1000	6.8126	6.1000	142.4519
1.2400	0.9085	1.6400	0.8986	2.2000	1.1018	4.2000	7.7567	6.2000	169.4061
1.2600	0.9044	1.6600	0.9017	2.3000	1.1667	4.3000	8.8553	6.3000	201.8133
1.2800	0.9007	1.6800	0.9050	2.4000	1.2422	4.4000	10.1361	6.4000	240.8338
1.3000	0.8975	1.7000	0.9086	2.5000	1.3293	4.5000	11.6317	6.5000	287.8853
1.3200	0.8946	1.7200	0.9126	2.6000	1.4296	4.6000	13.3813	6.6000	344.7019
1.3400	0.8922	1.7400	0.9168	2.7000	1.5447	4.7000	15.4314	6.7000	413.4075
1.3600	0.8902	1.7600	0.9214	2.8000	1.6765	4.8000	17.8379	6.8000	496.6061
1.3800	0.8885	1.7800	0.9262	2.9000	1.8274	4.9000	20.6674	6.9000	597.4941

For integer values of α, $\Gamma(\alpha+1)=\alpha!$
Note that $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$; Excel and Matlab: gamma(α)

Gamma distribution pdf

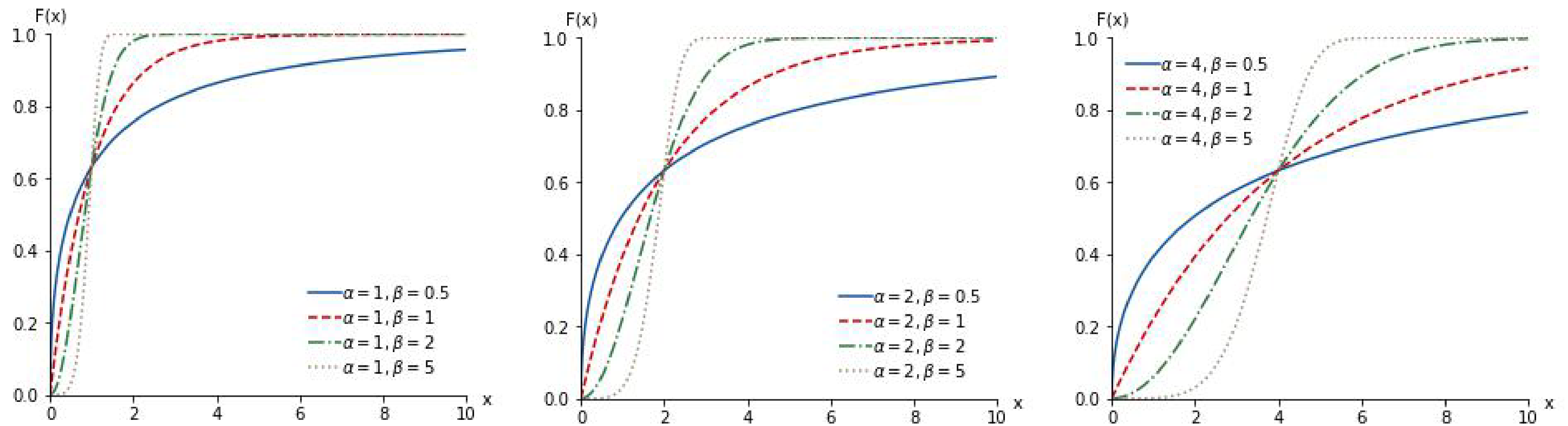
Each frame has constant α and varying β.



Gamma distribution cdf

- The cdf for the gamma distribution is

$$F(x|\alpha,\beta)=\begin{cases}\dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}\displaystyle\int_0^x x^{\alpha-1}\,e^{\left(-\frac{x}{\beta}\right)}dx & x,\alpha,\beta>0 \text{ and is } \alpha \text{continuous}\\ 1-e^{\left(-\frac{x}{\beta}\right)}\displaystyle\sum_{n=0}^{\alpha-1}\frac{x^{\alpha-1}}{\beta^{n}n!} & x,\alpha,\beta>0 \text{ and is } \alpha \text{ an integer}\\ 0 & \text{otherwise}\end{cases}$$



Note: The Chi-Sq. distribution is a special case of the Gamma distribution

Beta distribution $X \sim beta(\alpha, \beta)$

- **Definition**: Use the **beta distribution** for r.v.s that are distributed over the finite interval between 0 and 1.

$$f(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

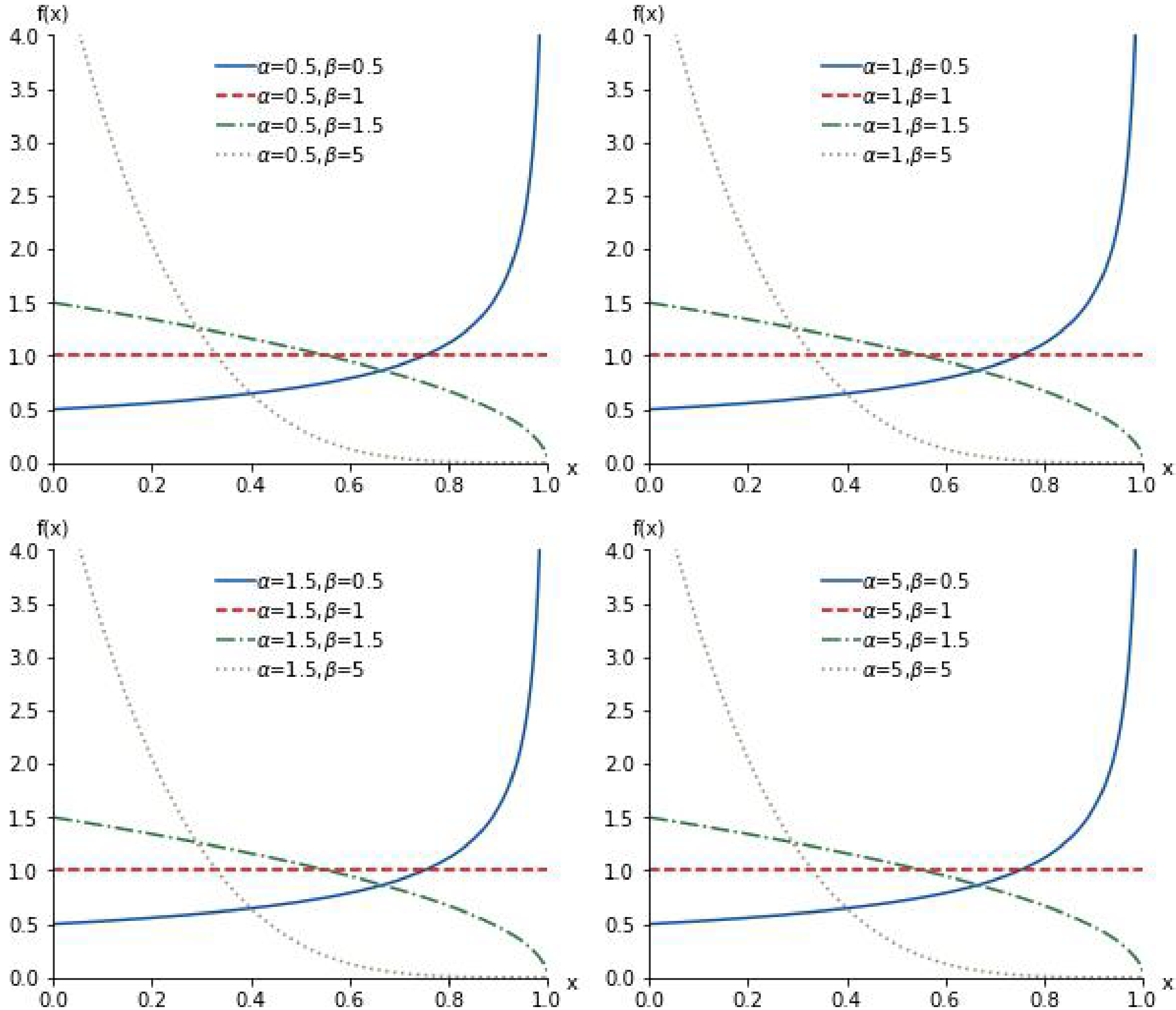
Where:
  - α – shape parameter and $\alpha > 0$
  - β – shape parameter and $\beta > 0$
  - Range – $0 \le x \le 1$
  - Excel: $Beta.Dist(X, \alpha, \beta, false)$

- Like the gamma distribution, it too uses the gamma function. The mean and variance of the beta distribution

$$E(X) = \frac{\alpha}{\alpha+\beta} \text{ and } \mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Beta distribution pdf

Each frame has constant α and varying β.



Beta distribution (cont.)

- The cdf of the beta distribution is

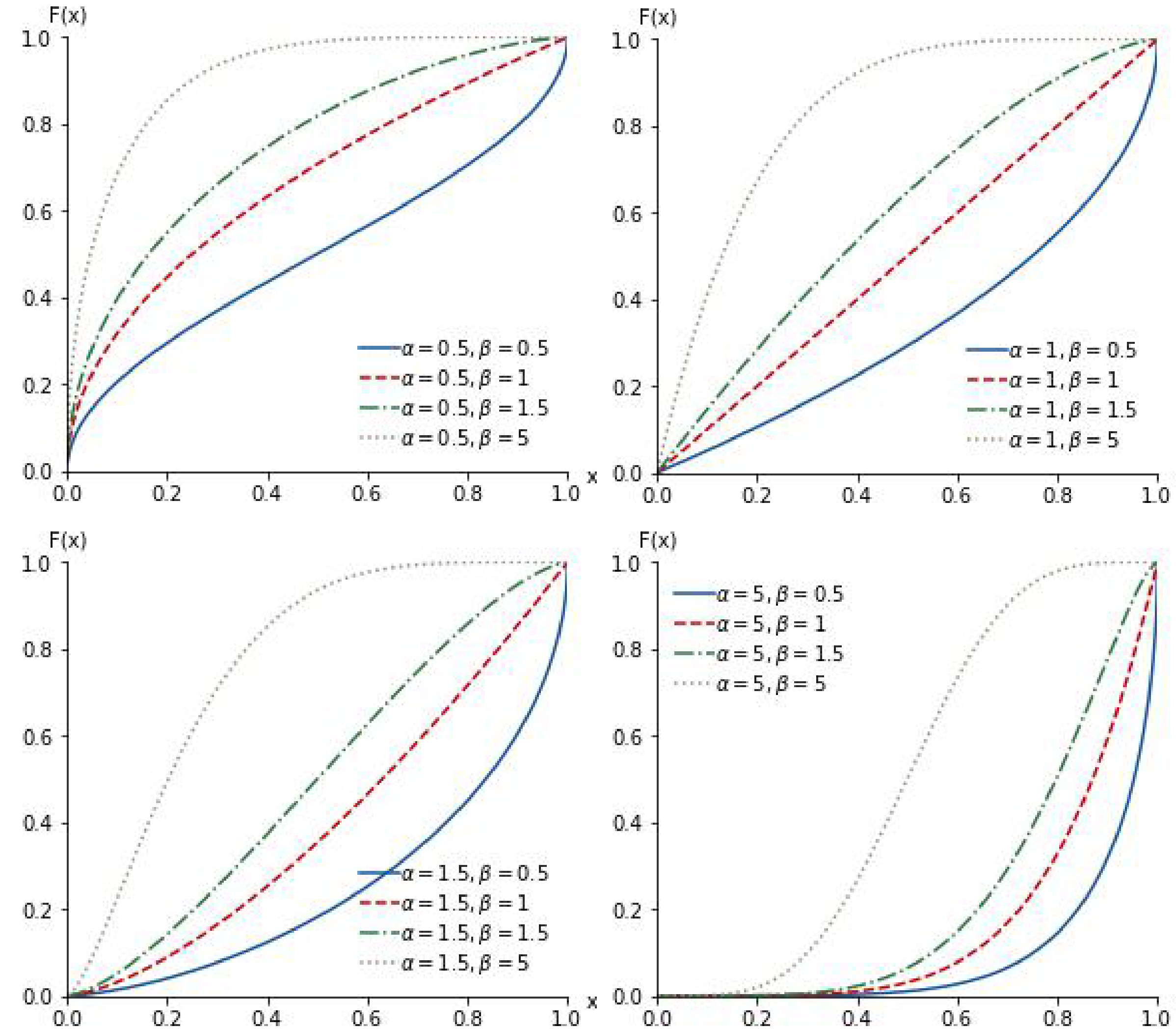
$$F(x|\alpha,\beta) = \begin{cases} 0 & x < 0 \\ \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \displaystyle\int_0^x x^{\alpha-1}(1-x)^{\beta-1}dx & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

Where:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{(-x)} dx$$

Beta distribution cdf

Each frame has constant α and varying β.



Truncated distributions

- **Truncation** arises when the r.v. cannot exist in some range of a distribution.
- A **truncated distribution** is the conditional distribution that results from restricting the domain of another probability distribution.
- The following **general formulas apply to truncated distribution functions**, where $f_0(x)$ and $F_0(x)$ are the pdf and cdf of the non-truncated distribution.

PDF:

$$f(x) = \begin{cases} \dfrac{f_0(x)}{F_0(b) - F_0(a)} & for\ x \in (a, b] \\ 0 & otherwise \end{cases}$$

CDF:

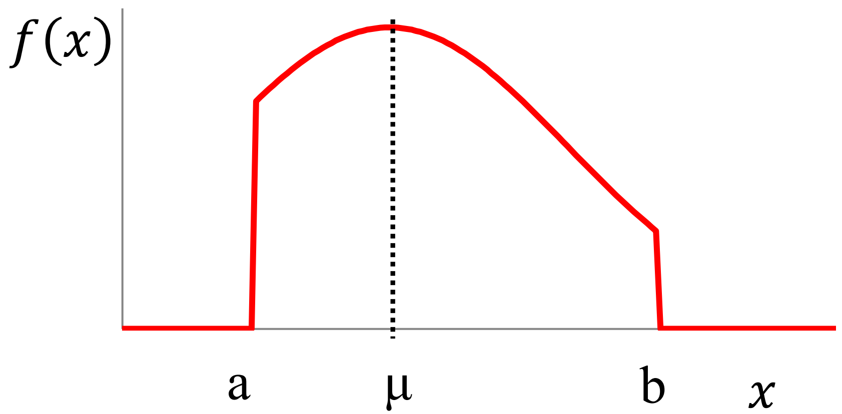
$$F(x) = \begin{cases} 0 & x \leq a \\ \dfrac{\int_a^x f_0(t)dt}{F_0(b) - F_0(a)} & x \in (a, b] \\ 1 & x > b \end{cases}$$

Truncated normal distribution

- $a \leq x \leq b$
- Pdf:

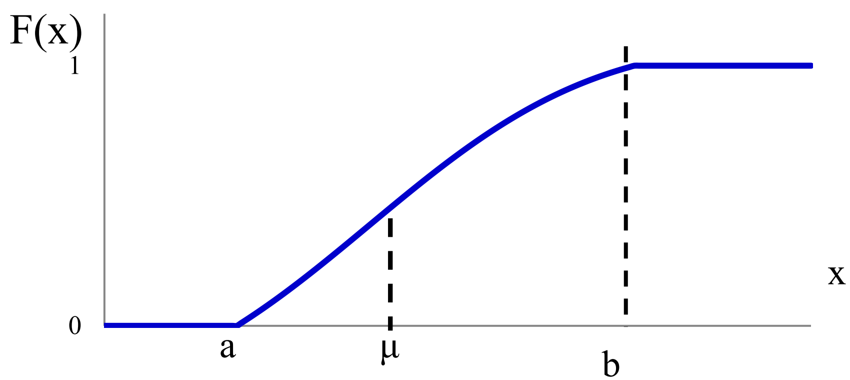
$$f(x) = \frac{f_{norm}(x)}{F_{norm}(x = b_U) - F_{norm}(x = a_L)}$$

$$= \frac{1}{\sigma} \frac{\phi(z_x)}{\Phi(z_b) - \Phi(z_a)}$$



- Cdf:

$$F(x) \begin{cases} 0 & x < a \\ \dfrac{F_{Norm}(x) - F_{Norm}(x = a)}{F_{Norm}(x = b) - F_{Norm}(x = a)} = \dfrac{\Phi(z_x) - \Phi(z_a)}{\Phi(z_b) - \Phi(z_a)} & a \leq x \leq b \\ 1 & x > b \end{cases}$$



Multivariate distributions

- **Definition**: Let's say that several random variables take the following values,

$$(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \text{ or } (x_1, x_2, \dots, x_n)$$

- Then a **joint (multivariate) pdf** is a function $f(x_1, x_2, \dots, x_n)$ that satisfies the following conditions:

  i. $f(x_1, x_2, \dots, x_n) \geq 0$ and $-\infty < x_i < \infty$ for $i = 1,2, \dots, n$

  ii. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = 1$

- **Definition**: The **marginal pdf** of X_i is defined by,

$$f_i(x_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_{i-1} dx_{i+1} \dots dx_n$$

Multivariate distributions

- **Definition**: The **conditional pdf** of X_1 given X_2 is denoted by $g(x_1|x_2)$ and defined as,

$$g(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}, f_2(x_2) \neq 0$$

- Two random variables are **independent** if their pdf follows,

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$$

- Similarly,

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_n(x_n)$$

Reliability Analysis

Module 3: Elements of Component Reliability

Prof. Katrina M. Groth
Mechanical Engineering Department
Center for Risk and Reliability
University of Maryland
kgroth@umd.edu

Objectives for Module 3

- Introduce methods for assessing reliability of components
- Define key terms: Reliability, MTTF, MRL, hazard rate, etc.
- Discuss & use several probability distributions commonly used in reliability engineering

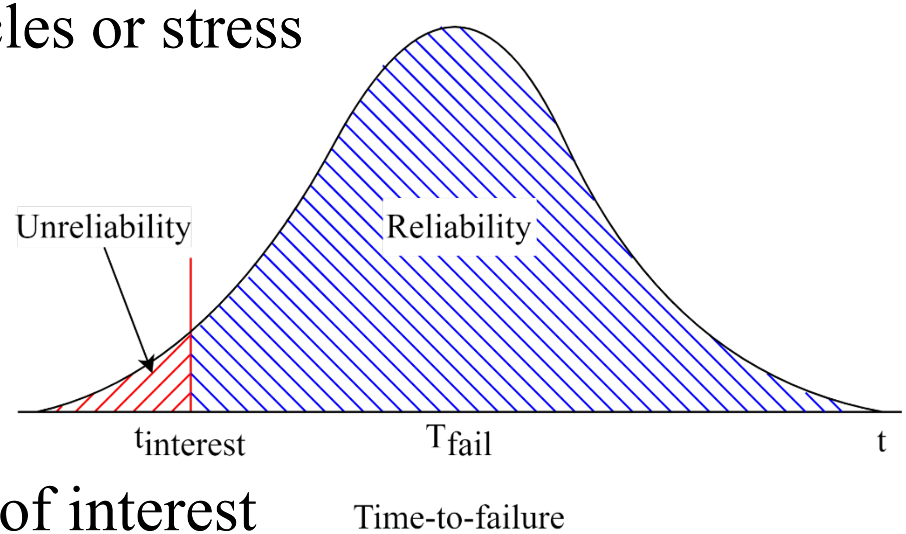
Design for reliability (Robustness)

- Extend item life by controlling or eliminating potential failure modes
  - Designing stronger, more durable items
  - Reducing harmful environmental conditions, uses, etc.
  - Minimizing / controlling loads and stresses
  - Condition monitoring and maintenance programs
- This increases reliability, i.e., the probability of successful achievement of item’s function (mission)

Recall: probabilistic definition of reliability

- **Reliability:** The ability of an item to perform its expected mission under designated operating conditions for a designated period of time, number of cycles or stress

$$\boxed{R(t) = \Pr(T_{fail} > t_{interest})}$$



Where:

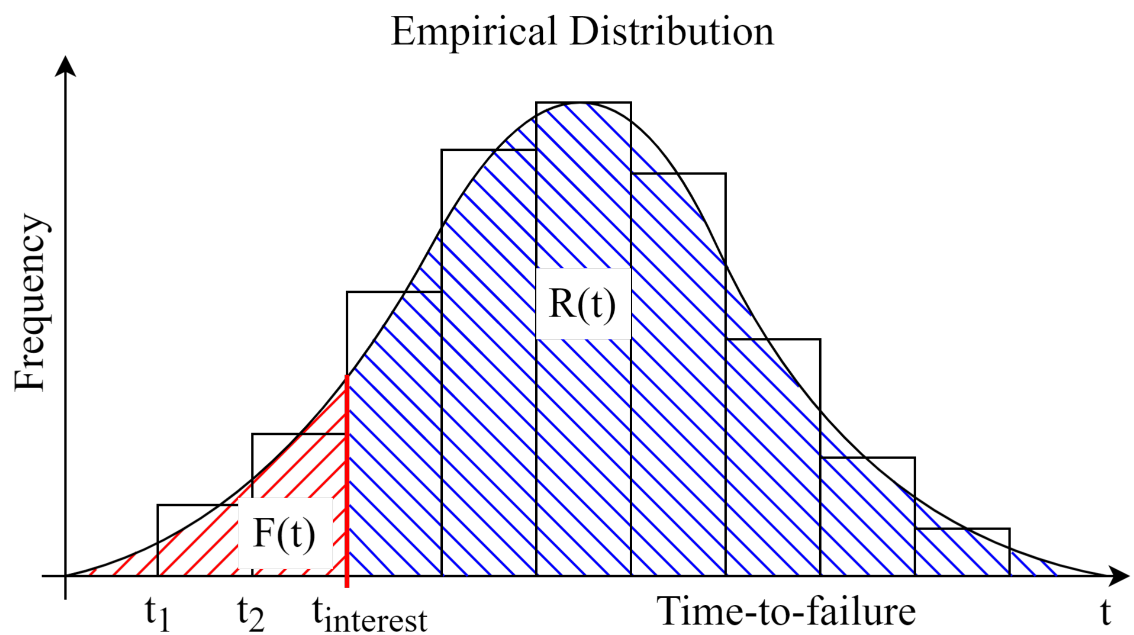
- $t_{interest}$ = mission time or time of interest
- T_{fail} = time-to-failure, cycle-to-failure, stress-to-failure, etc.

Note:

For many component reliability problems: *Time* is an *aggregate* "agent" of the failure; implies conditions are not necessary to model

Intuition: Defining reliability for non-repairable items

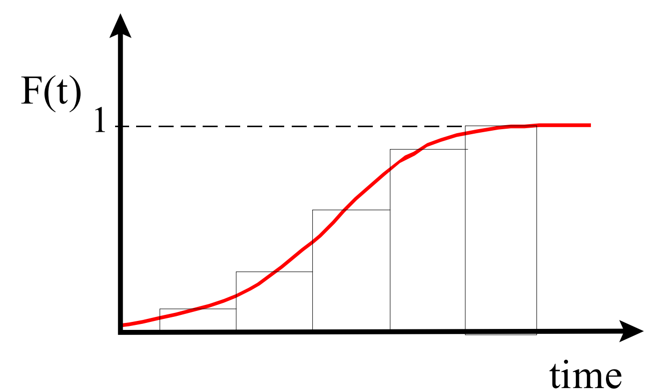
- Let there be many identical items that are subjected to life tests. As time goes by, each item will function for some time and then fail. We plot the frequency of time to failure (of these non-repairable items), we get a continuous time to failure (TTF) distribution.



f(t) = probability density function (pdf) for the *failure times*

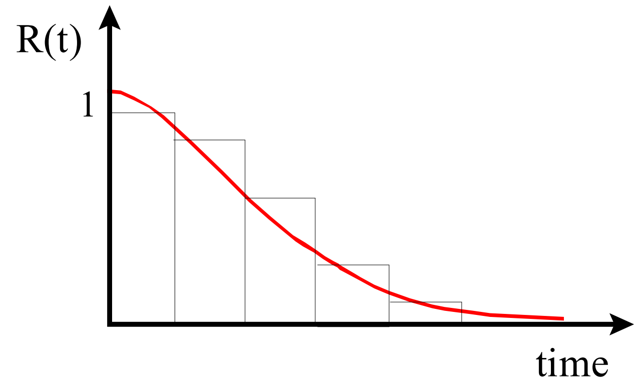
Definition of unreliability

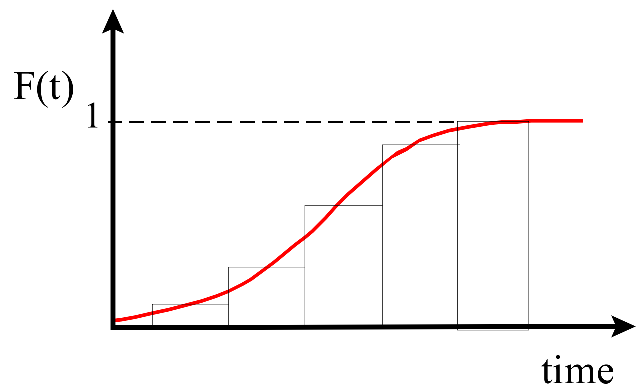
- **Unreliability** of an item: The item fails before the mission is complete. That is, the **failure time** (T_{fail}) of the item is less than the **mission time** (t) of the system.
  - In mathematical notation: the probability that the item fails at or before t. (or "sometime up to mission time t"):
  - $\Pr\left(T_{fail} \leq t\right) = F(t) = \int_0^t f(x)dx$ ← CDF for continuous variable T
  - $\Pr\left(N_{fail} \leq n\right) = F(n) = \sum_{i=1}^{n} f(i)$ ← CDF for discrete variable N (e.g., number of cycles).



Definition of reliability

- The **Reliability** function (or **survival function**) item is therefore:
  - $R(t) = \Pr(T_{\text{fail}} > t) = 1 - F(t)$
    $$= \int_t^\infty f(x)dx$$
  - Since these are probabilities:
  - $0 \leq F(t) \leq 1$
  - $0 \leq R(t) \leq 1$
  - $f(t) \geq 0 \quad -\infty < t < \infty$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$





Sidebar on notation

- For continuous distributions: $\Pr(a > X)$ is the same as $\Pr(a \geq X)$
- For discrete distributions:
  - You must know & define the edge of your set (i.e., "greater than" vs. "greater than or equal to.")

- Be very careful working with a distribution where the difference between "$< a$" and "$\leq a$" matters! (E.g., if the item fails exactly at time $T = t$)
  - Some reliability publications use $R(\mathrm{n}) = \Pr(\mathrm{N} \geq n)$ which results in $R(\mathrm{n}) \neq 1 - \mathrm{F}(\mathrm{n})$

*In the field, notation may not be consistent or readily apparent.
Always check that: $\mathrm{R(t)} + \mathrm{F(t)} = 1$
and adjust accordingly.

Reliability definitions: Conditional reliability

- **Conditional Reliability Function**

If an item has survived to time t, what is the probability that the item will survive for additional time x?

$$R(t+x|t) = \frac{R(t+x)}{R(t)}$$

Reliability definitions: MTTF

- **Mean-Time-To-Failure (MTTF)**
- In a population of items, the items will not all fail at the same time. As such, we get a distribution of time-to-failure.
- The **expected (mean)** time-to-failure is called MTTF, and sometimes the expected life.

We can estimate MTTF from a distribution:

$$\text{MTTF} = \text{E(t)} = \int_0^\infty t f(t) dt \quad \text{or} \quad E(n) = \sum_{i=1}^{\infty} n_i Pr(n_i)$$

We can also prove that $\mathbf{MTTF} = \int_0^\infty \mathbf{R}(\boldsymbol{t})\boldsymbol{dt}$ for cases where $\lim_{t\to\infty} R(t) = 0$

Reliability definitions (cont.)

- **Mean-Time-Between-Failure (MTBF)** – Applies only to repairable components.

When the items are repairable, we have two random variables.

i. Time-between-failures of an item,

ii. Time-to-repair.

This topic will be discussed later when we discuss repairable items (Chapter 7).

Reliability definitions: MRL

- **Mean Residual Life**

If an item has survived up to a certain time, *t*, then the expected remaining time to failure (i.e., expected remaining life) is called the residual mean time to failure, or simply the **mean residual life** (MRL):

$$\mathrm{MRL(t)} = \int_0^{\infty} R(x|t)dx \ = \frac{1}{\mathrm{R}(t)} \int_t^{\infty} \mathrm{R(t')dt'}$$

$$Note: t' = x + t$$

$$\text{If } t = 0, \text{then } MRL = MTTF = \int_0^{\infty} R(t)dt$$

Hazard rate (failure rate) defined

- ***Hazard rate (or failure rate)*** is the instantaneous rate of failure for an item of age *t* over a period of time (Δt) as Δt tends to zero.
- Think of it as: the conditional probability that an item which has survived until time *t* will fail during the following small time interval (Δt)

$$h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{F(t+\Delta t) - F(t)}{R(t)} = \frac{f(t)}{R(t)}$$

(Since $\Pr(t_1 < \mathrm{T} < \mathrm{t}_2 | \mathrm{T} > \mathrm{t}_1) = \frac{F(\mathrm{t}_2) - F(\mathrm{t}_1)}{\mathrm{R}(\mathrm{t}_1)} = \frac{\Delta F(t)}{R(t)}$)

- **Interpretation**: the greater the hazard rate between times t_1 and t_2, the greater the chance of failure in this time interval

Reliability definitions (cont.)

- It can be shown that *h(t)* can be expressed in terms of the reliability function (using definition of derivative of the ln of a function).

$$h(t) = \frac{f(t)}{R(t)} = -\frac{d}{dt}\left[\ln\left(R(t)\right)\right]$$

- As with the pdf and cdf, a **cumulative hazard function *H(t)*** can be obtained from the integral of the hazard function:

$$H(t) = \int_0^t h(x)dx$$

- We can use H(t) to obtain another expression for reliability:

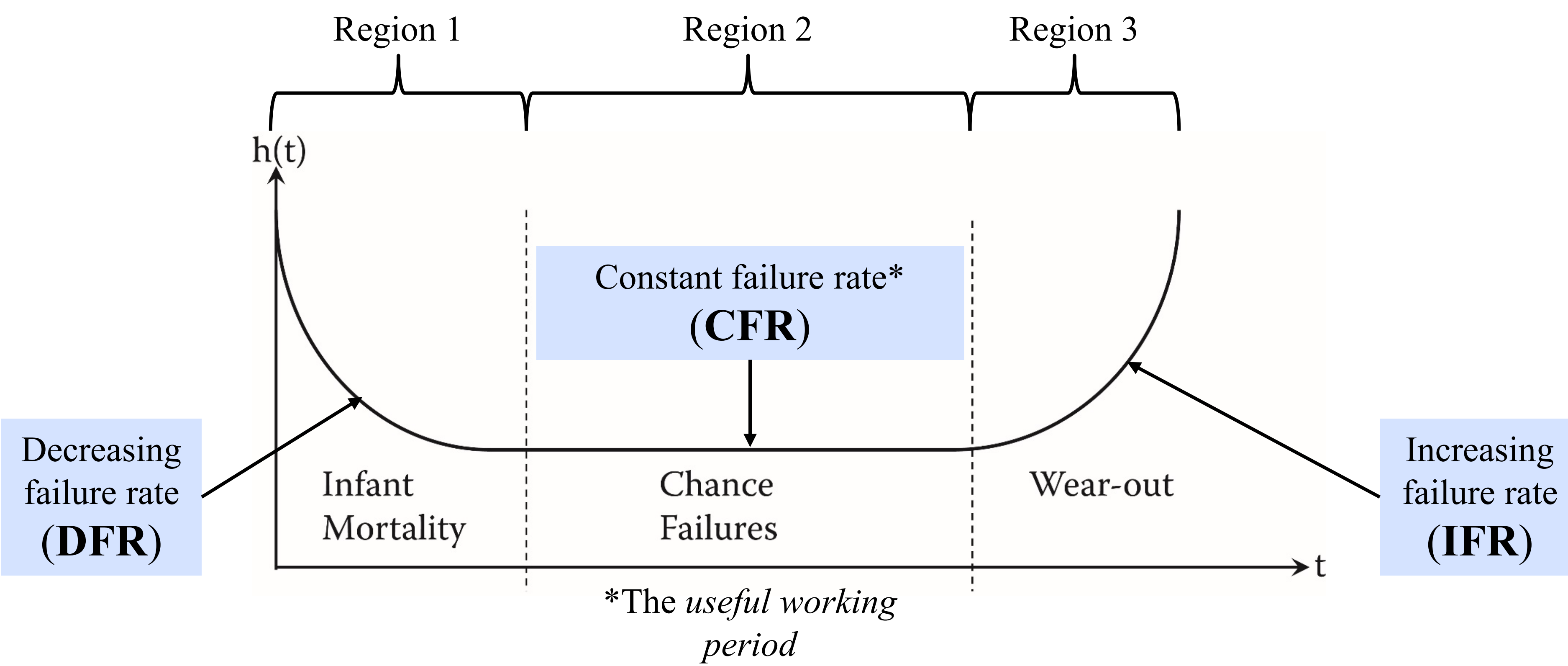
$$R(t) = e^{-\int_0^t h(x)dx} = e^{-H(t)}$$

Reminder: Key relationships

In terms of…	f(t)	F(t)	R(t)	h(t)	H(t)
f(t) =	-	$\frac{dF(t)}{dt}$	$-\frac{dR(t)}{dt}$	$h(t)e^{\left[-\int_0^t h(x)dx\right]}$	$\frac{dH(t)}{dt}e^{[-H(t)]}$
F(t) =	$\int_{-\infty}^{t} f(x)dx$	-	$1 - R(t)$	$1 - e^{\left[-\int_0^t h(x)dx\right]}$	$1 - e^{[-H(t)]}$
R(t) =	$\int_{t}^{\infty} f(x)dx$	$1 - F(t)$	-	$e^{\left[-\int_0^t h(x)dx\right]}$	$e^{[-H(t)]}$
h(t) =	$\frac{f(t)}{\int_t^\infty f(x)dx}$	$\frac{\frac{dF(t)}{dt}}{1-F(t)}$	$-\frac{\frac{dR(t)}{dt}}{R(t)}$	-	$\frac{dH(t)}{dt}$
H(t) =	$-ln\left[\int_t^\infty f(x)dx\right]$	$-ln[1 - F(t)]$	$-ln[R(t)]$	$\int_0^t h(x)dx$	-

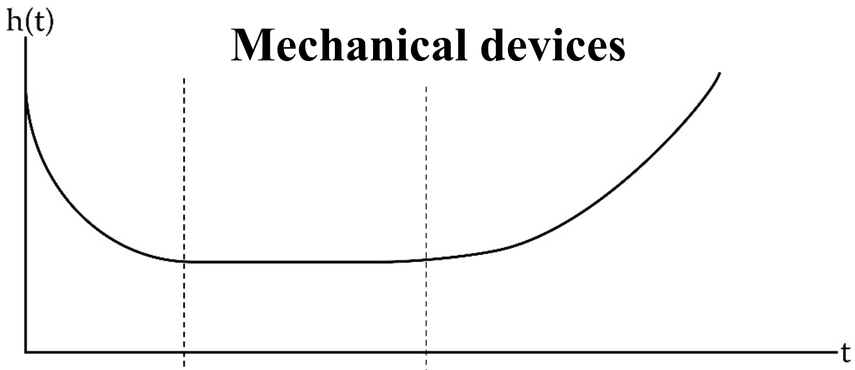
Hazard rate trends: Bathtub curve

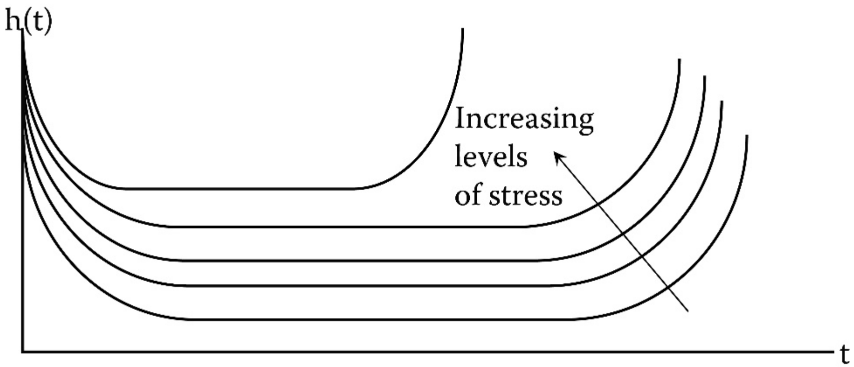
- The hazard rate *h(t)* shows changes in the probability of failure over the lifetime of a component
- In practice, *h(t)* often forms the **bathtub curve**

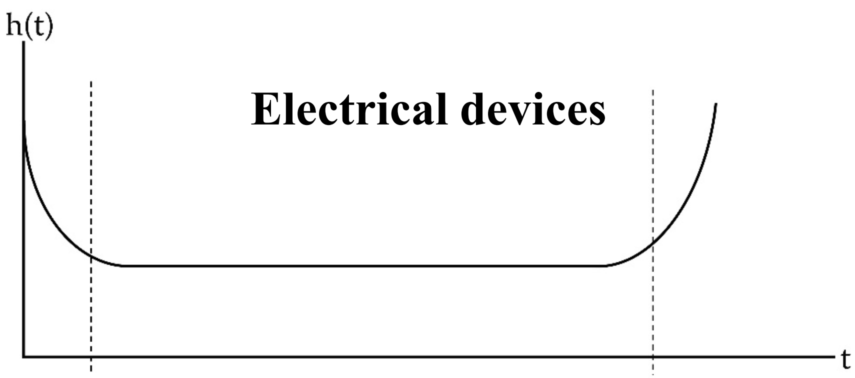


Typical bathtub curves for devices

- The hazard rate *h(t)* can also show differences between component types (e.g., mechanical vs. electrical) and the effects of stress:







Note: electrical devices tend to have longer useful working period (CFR) and shorter burn-in (DFR) and wear-out (IFR) periods.

Increasing the stress level shortens the useful life period.

Bathtub curve (cont.)

- The curve is characterized by three distinct regions:
  - **Region 1 (DFR)** The failure occurs early in the life where the item has a high, but decreasing failure rate. Failures are caused due to initial defects like defective design, poor material, etc.
  - **Region 2 (CFR)** The failures are random, i.e., difficult to predict and failure rate is approximately constant over the time. This is the “useful working period” of the component.
  - **Region 3 (IFR)** As the item reaches the end of life, it starts to deteriorate and wear out. Failures keep increasing and results in a fast-increasing hazard rate.

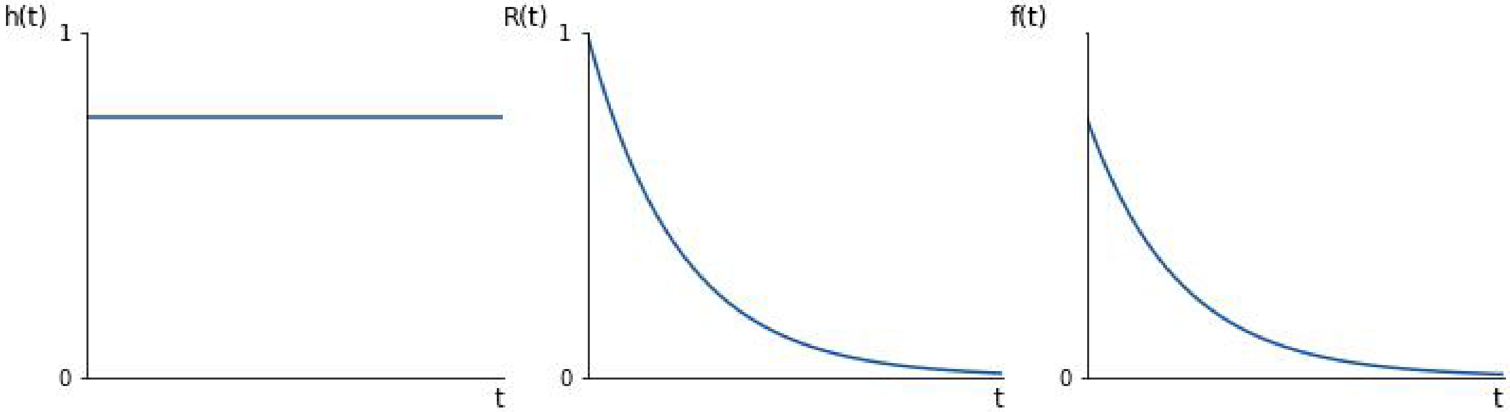
Three distinct hazard rates

- **Case I – Constant failure rate (λ)**
  - That is, $h(t) = \lambda$, and:

$$R(t) = e^{-\int_0^t h(x)dx} = e^{-\int_0^t \lambda dx} = e^{-\lambda \int_0^t dx} = e^{-\lambda t}$$

$$f(t) = h(t) \cdot R(t) = \lambda e^{-\lambda t}$$

This is the **exponential distribution**



So, the random failure region can be modeled by an **exponential distribution model**.

Three distinct hazard rates (cont.)

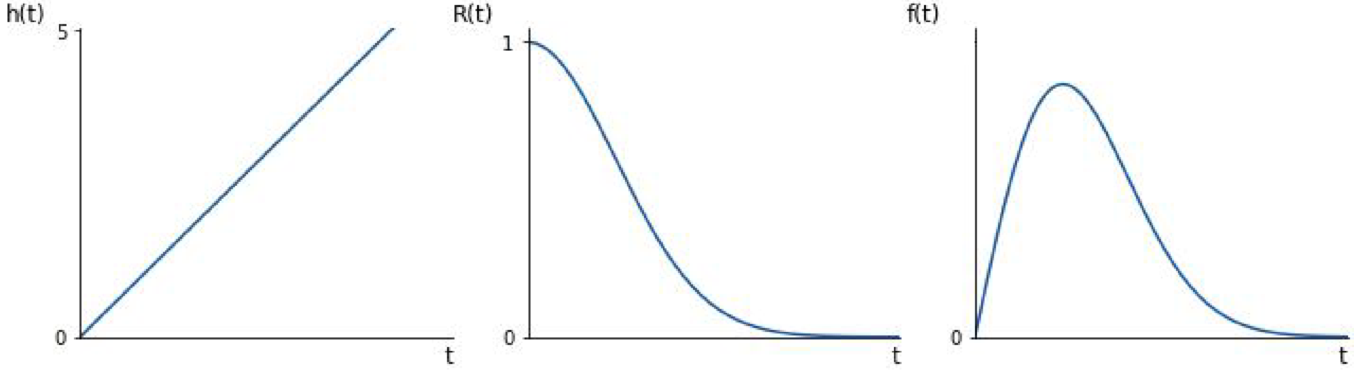
- **Case II – Increasing failure rate**
  - For example, a linearly increasing hazard rate:

$$h(t) = \beta t \text{ (where } \beta \text{ is a constant), and:}$$

$$R(t) = e^{-\int_0^t h(x)dx} = e^{-\int_0^t \beta x dx} = e^{-\beta \int_0^t x dx} = e^{-\frac{1}{2}\beta t^2}$$

$$f(t) = R(t) \cdot h(t) = e^{-\frac{1}{2}\beta t^2} \cdot \beta t = \beta t e^{-\frac{1}{2}\beta t^2}$$

  - This is the Rayleigh distribution. The wear out region in Bathtub curve can be modeled by this distribution. Other distributions may also be used, depending on failure data.



Three distinct hazard rates (cont.)

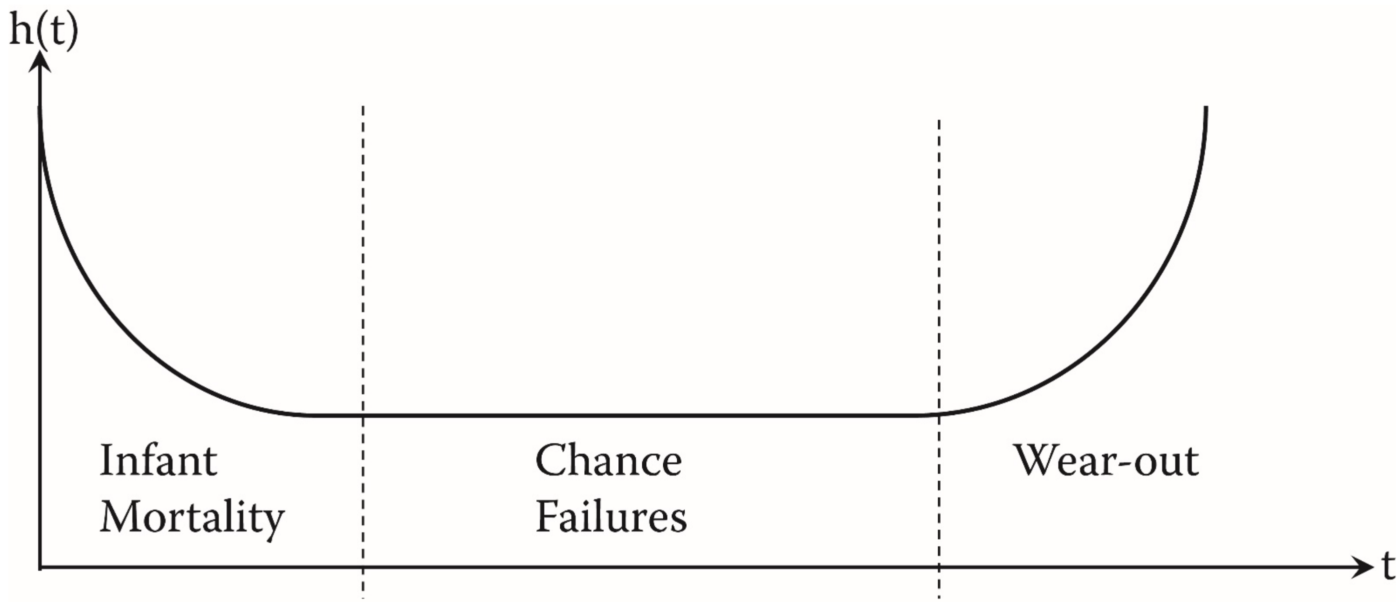
- **Case III – Decreasing failure rate**

This case is similar to Case II, only the hazard rate is a decreasing function of time.

$$h(t) = -\beta t$$

Recapitulation

- $R(t) = \Pr(T > t) = e^{-\int_0^t h(x)dx}$ where *T* is a random variable, represented by a time-to-failure distribution f(t).
- MTTF is the mean of pdf f(t) representing time-to-failure
- $h(t) = \dfrac{f(t)}{R(t)}$



Common distributions in component reliability

- An item's reliability may be represented by many probability distributions; popular distributions in reliability include:
  - Exponential
  - Weibull
  - Gamma
  - Normal
  - Lognormal
- The distributions have their own form of reliability, hazard rate, and MTTF functions.
- Each has merits for different situations.

There is no one-size-fits-all distribution for *any* item

Exponential distribution in reliability modeling

- $T \sim \exp(\lambda)$

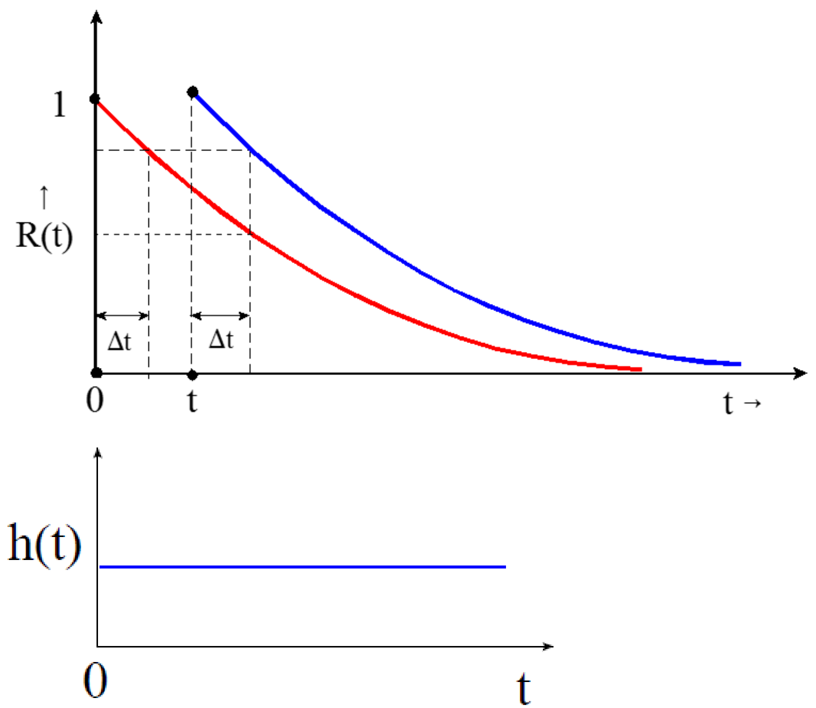
$$f(t) = \lambda e^{-\lambda t}$$

$$F(t) = 1 - e^{-\lambda t}$$

$$R(t) = e^{-\lambda t}$$

$$h(t) = \lambda$$

$$\mathrm{MTTF} = \frac{1}{\lambda}$$



- Exponential distribution has no memory. It is often used for mathematical simplicity.

Uses of the exponential distribution

- Early efforts at collecting reliability data assumed constant failure rates and thus many reliability handbooks provide only CFRs components.
- **Electronic Components**: Some electronic components like capacitors or integrated circuits have been found to follow an exponential distribution.
- **Random Shocks**: Time to the occurrence of random shocks. An example is the failure of a vehicle tire due to puncture from a nail (random shock). The probability of failure in each mile is independent of how many miles the tire has travelled (memoryless). The failure rate when the tire is new is the same as when the tire is old.
- Modeling complex systems (due to need for mathematical simplicity)

In general, many items do not have a constant failure rate, for example due to wear or early failures. Thus, the exponential distribution is often inappropriate to model most life distributions, particularly mechanical components.

Example: Exponential TTF

- **Example:** Consider a component with exponentially distributed time-to-failure, $T \sim \exp(\lambda)$ where $\lambda = 1 \times 10^{-3} \text{hr}^{-1}$. Calculate reliability at 700 hours and calculate MTTF.

Example: Exponential TTF

- **Solution:** Consider a component with exponentially distributed time-to-failure, $T \sim \exp(\lambda)$ where $\lambda = 1 \times 10^{-3} \mathrm{hr}^{-1}$. Calculate reliability at 700 hours and MTTF.
- Calculate R(t = 700hrs) and MTTF

$$\mathrm{R(t = 700) = e^{-\lambda t} = e^{-10^{-3} \cdot 700} = e^{-.7}} = \mathbf{0.4966}$$

$$MTTF = \frac{1}{10^{-3}} = \mathbf{1000\ hrs}$$

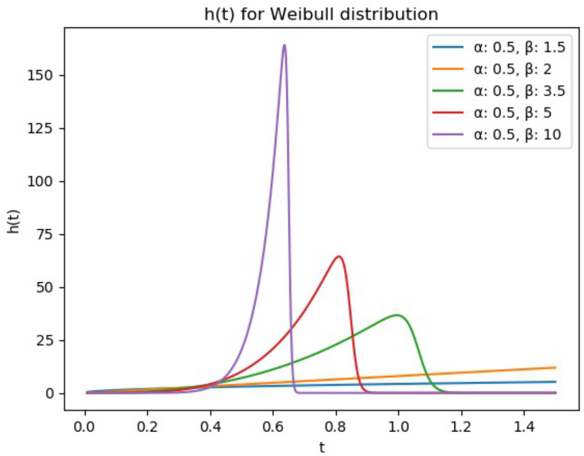
Weibull distribution in reliability modeling

$$f(t) = \frac{\beta t^{\beta-1}}{\alpha^{\beta}} e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$$

$$R(t) = e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$$

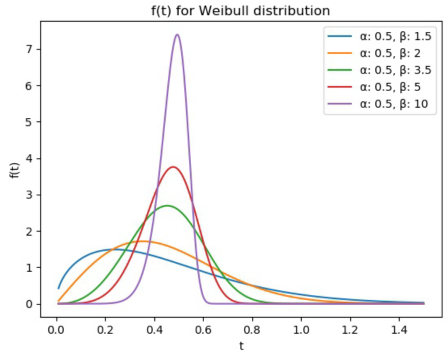
$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{\beta t^{\beta-1}}{\alpha^{\beta}} e^{-\left(\frac{t}{\alpha}\right)\beta}}{e^{-\left(\frac{t}{\alpha}\right)\beta}} = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}$$

$$MTTF = \alpha\Gamma\left(\frac{\beta+1}{\beta}\right)$$

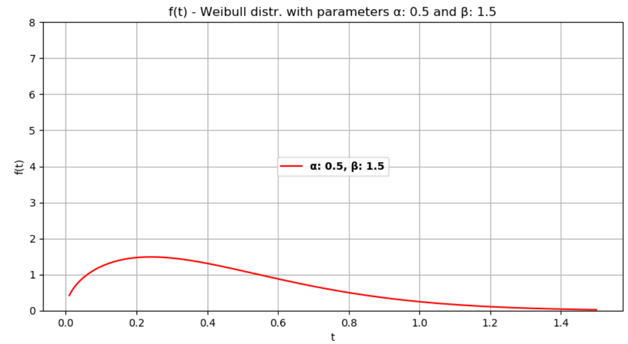


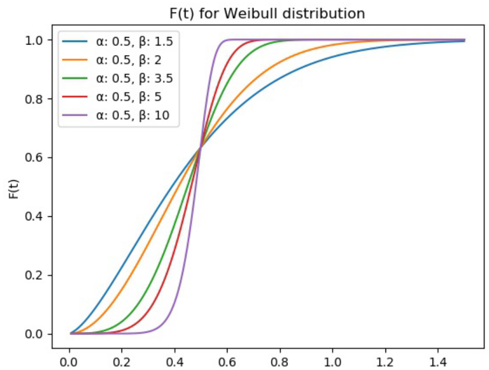
- Where,
  - α = scale parameter
  - β = shape parameter

Weibull distribution (cont.)

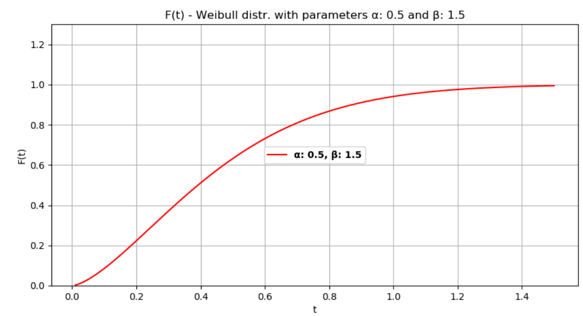


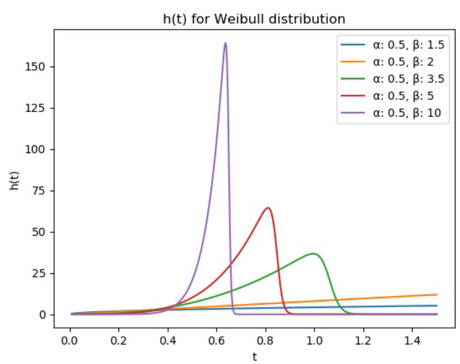
f(t)



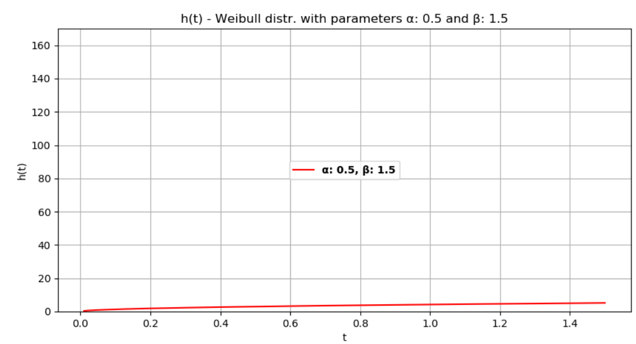


F(t)





h(t)



Uses of the Weibull distribution

- Prevalent in modeling time to failure of many types of basic component (capacitors, ball bearings, motors, transistors, etc.)
- Corrosion caused failure
- Weakest link models

- Also see 3-parameter Weibull distribution used commonly.

Example: Weibull TTF

- **Example:** Lab testing shows that a time-to-failure model of T$\sim weibull(\alpha = 230\ hr, \beta = 2.1)$ is a good fit for a set of newly manufactured roller bearings.

a) Calculate the reliability of a bearing at 240 hours

b) Calculate the MTTF of the bearings.

c) If time: Plot the hazard rate.

Example: Weibull TTF

- **Example Solution:** Lab testing shows that a time-to-failure model of $T \sim weibull(\alpha = 230\ hr, \beta = 2.1)$ is a good fit for a set of newly manufactured roller bearings.

a) Calculate the reliability of a bearing at 240 hours

$$R(t = 240) = e^{-\left(\frac{x}{\alpha}\right)^{\beta}} = e^{-\left(\frac{240}{230}\right)^{2.1}} = \mathbf{0.335}$$

b) Calculate the MTTF of the bearings

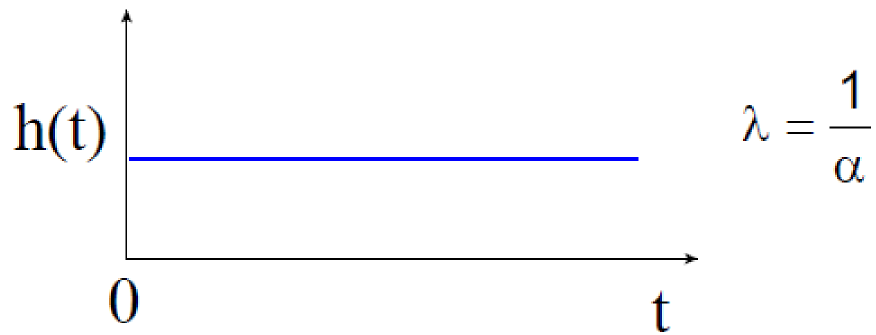
$$MTTF = \alpha\Gamma\left(1 + \frac{1}{\beta}\right) = 230 \cdot \Gamma(1.476) = 230 \cdot 0.886 = \mathbf{203.7\ hrs}$$

Use of Weibull distribution

- **Bathtub Case I** – $\beta = 1$

That is, $h(t) = \frac{1}{\alpha}\left(\frac{t}{\alpha}\right)^{1-1} = \frac{1}{\alpha}$

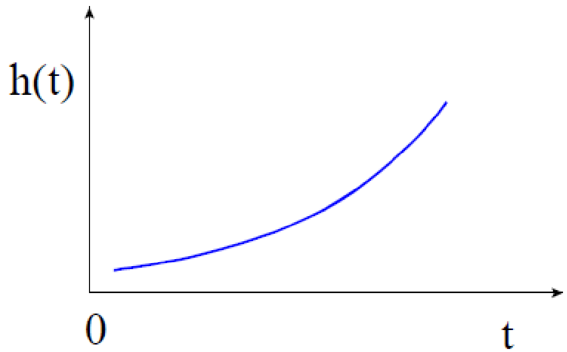
- So it is the same as an exponential distribution with $\lambda = \frac{1}{\alpha}$ and MTTF $= \alpha$.



Use of Weibull distribution (cont.)

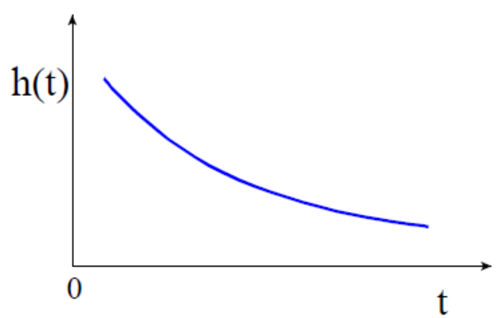
- **Bathtub Case II** – $\beta > 1$

  That is, h(t) is monotonically increasing



- **Bathtub Case III** – $\beta < 1$

  That is, h(t) is monotonically decreasing



Weibull distribution hazard rates

- **$1 < \beta < 2$**: The hazard rate increases less as time increases
- **$\beta = 2$**: The hazard rate increases with a linear relationship to time
- **$\beta > 2$**: The hazard rate increases more as time increases
- **$\beta < 3.447798$**: The distribution is positively skewed (Tail to right)
- **$\beta \approx 3.447798$**: The distribution is approximately symmetrical
- **$3 < \beta < 4$**: The distribution approximates a normal distribution
- **$\beta > 10$**: The distribution approximates a smallest extreme value distribution

Gamma distribution in reliability modeling

$$f(t) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} t^{\alpha-1} e^{-t/\beta} \qquad \text{(For integer } \alpha \text{ recall that: } \Gamma(\alpha) = (\alpha-1)!)$$

$$R(t) = \begin{cases} 1 - \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} \displaystyle\int_0^t x^{\alpha-1} e^{\left(-\frac{x}{\beta}\right)} dx & t, \alpha, \beta \geq 0 \text{ and } \alpha \text{ is continuous} \\ e^{-\frac{t}{\beta}} \displaystyle\sum_{n=0}^{\alpha-1} \frac{t^n}{\beta^n n!} & t, \alpha, \beta \geq 0 \text{ and } \alpha \text{ is an integer} \end{cases}$$

$$h(t) = \frac{f(t)}{R(t)} = \begin{cases} \dfrac{t^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha, \frac{t}{\beta})} e^{\left(-\frac{t}{\beta}\right)} & t, \alpha, \beta > 0 \text{ and } \alpha \text{ is continuous} \\ \dfrac{t^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha) \sum_{n=0}^{\alpha-1} \frac{t^n}{\beta^n n!}} & t, \alpha, \beta > 0 \text{ and } \alpha \text{ is an integer} \end{cases}$$

Where:

- α = shape parameter given by number of events required before failure, maintenance, etc. (α=n)
- β = scale parameter given by mean time to occurrence of one event (sometimes written as $\lambda = 1/\beta$)(and α, β, t ≥ 0)

Uses of gamma distribution

- Time to failure after being subjected to α random Poisson events
- Sum of α independent exponential variables
- Time between maintenance for an item maintained every α uses
- TTF of a system with standby components and independent failures
- Distribution of time between maintenance of items, instruments, and systems after they have been used α times.

Example: Gamma distribution

- **Example**: A generator manufacturer suggests conducting an oil change once every 140 days. (Assume that the time between oil change, T_o, follows an exponential distribution). If there is also a rule in place that says that after 5 oil changes, the oil filter must be replaced,

a) What is the distribution representing the time to replacement of the oil filter (T_R)?

b) What is the mean time to replacement of the oil filter?

c) What is the probability that the generator does not need a new oil filter at its 200-day inspection (that is, what is the reliability at this time)?

Example: Gamma distribution (cont.)

- **Solution**:

a) What is the distribution representing the time to replacement of the oil filter (T_R)?

- For T_R we have a mean time to adjustment of 140 days. Let this be the β value or mean time to occurrence.
- Likewise, α represents the number of required before the occurrence or replacement and this has been identified as 5.
- Therefore, the distribution of T_R is gamma with:

$$\alpha = 5 \text{ event}, \beta = 140 \text{ hr}$$

$$\boldsymbol{T_R \sim gamma(5, 140)}$$

b) What is the mean time to replacement of the oil filter?

- Simply compute the MTTR as,

$$\text{MTTR} = E(T_R) = \alpha\beta$$
$$= 5 \cdot 140$$
$$= \mathbf{700\ days}$$

Example: Gamma distribution (cont.)

- **Solution:**

c) What is the probability that the engine does not need a new oil filter at its 200 day inspection (that is, what is the reliability at 200 days)?

- Calculate the reliability where t = 200 days as,

$$R(t) = \exp\left(-\frac{t}{\beta}\right)\sum_{n=0}^{\alpha-1}\frac{t^n}{\beta^n n!}$$

$$= \exp\left(-\frac{200}{140}\right)\sum_{n=0}^{5-1}\frac{200^n}{140^n n!}$$

$$= e^{\left(-\frac{200}{140}\right)}\left[\frac{(200/140)^0}{0!}+\frac{(200/140)^1}{1!}+\frac{(200/140)^2}{2!}+\frac{(200/140)^3}{3!}+\frac{(200/140)^4}{4!}\right]$$

$$= 0.24\cdot\ (1+1.43+1.02+0.486+0.174)$$

$$= \mathbf{0.985}$$

Normal distribution in reliability modeling

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right]}$$

$$R(t) = 1 - \Phi\left(\frac{t-\mu}{\sigma}\right)$$

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right]}}{1 - \Phi\left(\frac{t-\mu}{\sigma}\right)}$$

Where:

- μ is the mean and σ is the standard deviation

Uses of Normal distribution

- Stress-strength modeling
- Simple repair/inspection tasks with "typical" duration and variance symmetrical about the mean
- Modeling processes that a sum of a large number of independent random variables (see: central limit theorem)
  - Approximation to other distributions

Example: Normal distribution

- **Example:** A component's stress to failure is normally distributed with MSTF of 20 kg/cm^2 and standard deviation of 3 kg/cm^2.

a) Find the reliability at a stress level of 25 kg/cm^2.

b) Find the probability of failure between stress levels 25 and 28 kg/cm^2.

Example: Normal distribution (cont.)

- **Solution:** A component's stress to failure is normally distributed with MSTF of 20 kg/cm^2 and standard deviation of 3 kg/cm^2.

a) Find the reliability at a stress level of 25 kg/cm^2.

$$\mathrm{R}\left(25\frac{kg}{cm^2}\right) = 1 - \Phi\left(\frac{25-20}{3}\right) = 1 - \Phi(1.667) = 0.048$$

b) Find the probability of failure between stress levels 25 and 28 kg/cm^2.

$$F(25 < S < 28) = \Phi(2.667) - \Phi(1.667) = 0.044$$

Example: Normal distribution (cont.)

c) Find the conditional probability of failure before 28kg/cm^2, given the component has survived up to a 25 kg/cm^2 stress.

Example: Normal distribution (cont.)

c) **Solution:** Find the conditional probability of failure before 28kg, given the component has survived under a 25 kg/cm^2 stress.

$$\Pr(S < 28|S > 25) = \frac{\Pr(S < 28 \cap S > 25)}{\Pr(S > 25)} = \frac{\Pr(25 < S < 28)}{\Pr(S > 25)}$$

$$\Pr(S < 28|S > 25) = \frac{0.044}{0.048} = \mathbf{0.92}$$

The conditional survival function is:

$$\Pr(S > x + s|S > s) = \frac{\Pr(S > 28)}{\Pr(S > 25)} = \frac{R(28)}{R(25)} = \frac{0.0038}{0.048} = \mathbf{0.08}$$

Lognormal distribution in reliability modeling

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln t-\mu)^2}$$

$$R(t) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right)$$

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{\sigma t\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{\ln t-\mu}{\sigma}\right)^2\right]}}{1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right)} = \frac{\phi(\frac{\ln t - \mu}{\sigma})}{t\sigma(1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right))}$$

Where the two parameters are:

$$\mu = E(\ln(t)) \text{ and } \sigma^2 = var(\ln(t))$$

- Remember: if $T\sim lognorm(\mu, \sigma^2)$ then $\ln(T) \sim norm(\mu, \sigma^2)$

- Note that you can calculate the mean and variance of the lognormal distribution from the parameters of the lognormal distribution:

$$E(T)\text{=}MTTF = e^{\mu+\frac{\sigma^2}{2}}$$

$$Stdev(T) = \left(e^{\sigma^2} - 1\right)^{\frac{1}{2}} \times MTTF$$

Uses of the lognormal distribution

- The lognormal distribution is good for occurrence of events that may vary by several orders of magnitude, such as the time to finish a repair task.
- It is also good for modeling failure modes or processes that are a result of multiplicative errors (e.g., fatigue cracks)
- Human reliability analysis modeling – time to failure and repair
- Particle sizes in breakage
- Electronic components

Example: Lognormal distribution

- **Example:** The time that it takes an operator to shutdown a system is lognormally distributed with $\mu = 2.0273, \sigma^2 = 0.4608$. Determine probability of shutdown within $t = 20$ seconds (that is, that the operator successfully shuts down the system within 20 seconds).

a) Define the failure events and the success event in mathematical notation.

b) Solve.

Example: Lognormal distribution (cont.)

- **Solution:** The shutdown time is lognormally distributed with $\mu = 2.0273, \sigma^2 = 0.4608$, meaning:
  - $T_{shutdown} \sim \text{lognorm}(2.0273, \sqrt{(0.4608)})$
- Determine probability of shutdown by $t = 20$ seconds.

a) The failure event is $(T_{shutdown} > t)$ and the success event is $(T_{shutdown} \leq t)$.

b) $z_1 = \frac{\ln t - \mu}{\sigma} = \frac{\ln(20) - 2.0273}{\sqrt{0.4608}} = 1.427$

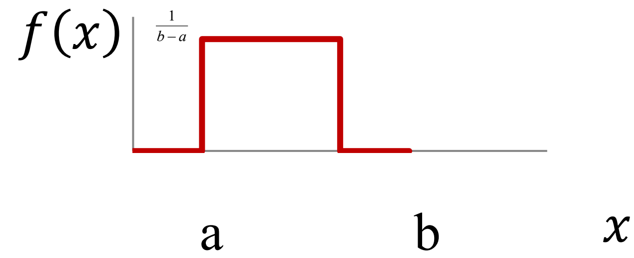
In Excel: $\text{LOGNORM.DIST}(20, 2.0273, \sqrt{0.4608},\ \text{TRUE})$

$$\boxed{\begin{array}{c} \Pr(T > 20) = 1 - F(20) = 0.077 \Rightarrow Unreliability \\ \Pr(T \leq 20) = F(20) = 0.923 \Rightarrow Reliability \end{array}}$$

Continuous uniform distribution in reliability modeling

$$f(t) = \begin{cases} \dfrac{1}{b-a} & a \le t \le b \\ 0 & t < a \text{ or } t > b \end{cases}$$

$$R(t) = \begin{cases} 1 & t < a \\ \dfrac{b-t}{b-a} & a \le t \le b \\ 0 & t > b \end{cases}$$



$$\text{MTTF} = \frac{a+b}{2}$$

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{b-a}}{\frac{b-t}{b-a}} = \frac{1}{b-t}, \qquad a \le t \le b$$

where,

- a = minimum value
- b = maximum value

Uses of uniform distribution in reliability modeling

- When a prior is unknown a uniform distribution can be used as an uninformative prior.
- Random number generation.

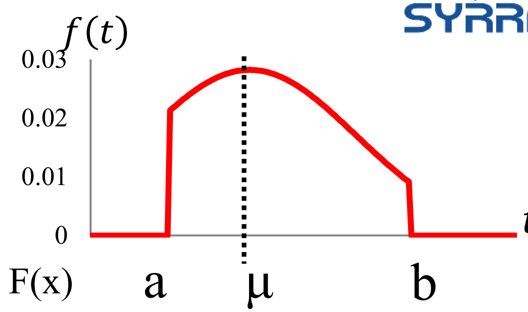
Truncated normal distribution in reliability modeling

$$f(t) = \frac{f_{Norm}(t)}{F_{norm}(t=b) - F_{norm}(t=a)} = \frac{\frac{1}{\sigma}\phi(z_t)}{\Phi(z_b) - \Phi(z_a)}$$

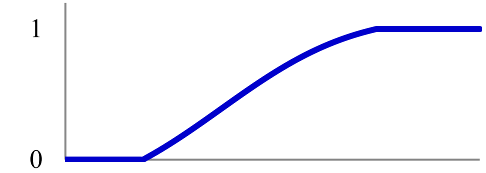
$$R(t) = \begin{cases} 1 & t < a \\ \dfrac{\Phi(z_b) - \Phi(z_t)}{\Phi(z_b) - \Phi(z_a)} & a \le t \le b \\ 0 & t > b \end{cases}$$

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{\sigma}\phi(z_t)[\Phi(z_b) - \Phi(z_t)]}{[\Phi(z_b) - \Phi(z_a)]^2}, \qquad a \le t \le b$$

$$MTTF = \frac{\frac{1}{\sigma}[\phi(z_a) - \phi(z_b)]}{\Phi(z_b) - \Phi(z_a)}$$



F(x)



Where:

- $-\infty < t < \infty, -\infty < \mu_t < \infty, \sigma_t^2 > 0$
- a = minimum value
- b = maximum value

$$z_t = \frac{t-\mu_t}{\sigma_t}$$

$$z_a = \frac{a-\mu_t}{\sigma_t}$$

$$z_b = \frac{b-\mu_t}{\sigma_t}$$

Uses of truncated normal distribution

- Life or time distributions with a known minimum and maximum life or time (e.g. product life spans, repair times)
- Measurements under an inspection threshold (e.g. flaw sizes, pit diameters)

Beta distribution in reliability modeling

$$f(t) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1-t)^{\beta-1} & 0 \le t \le 1 \\ 0 & t < 0 \text{ or } t > 1 \end{cases}$$

$$R(t) = \begin{cases} 1 & t < 0 \\ 1 - \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \displaystyle\int_0^t x^{\alpha-1}(1-x)^{\beta-1}dx & 0 \le t \le 1 \\ 0 & t > 1 \end{cases}$$

$$h(t) = \frac{f(t)}{R(t)} = \frac{\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1-t)^{\beta-1}}{1 - \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^t x^{\alpha-1}(1-x)^{\beta-1}dx}$$
$$= \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha,\beta) - B_t(t|\alpha,\beta)}$$

Where:

α = shape parameter, ($\alpha > 0$)

β = shape parameter, ($\beta > 0$)

Where:

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$$

$$B_t(t|\alpha,\beta) = \int_0^t x^{\alpha-1}\,(1-x)^{\beta-1}dx$$

Uses of the beta distribution

- The beta distribution is a good choice when given a set of percentages of times-to-failure instead of actual times.
- Likelihood ratios and proportion modeling
- Used frequently as a Bayesian conjugate prior for probabilities.

Example: Beta distribution

- **Example:** An energy lab designed a cutting-edge wind turbine that is completely different from any previous design. In the preliminary lab tests of 115 turbines, 92 turbines experienced no failure.

a) Find the expected failure probability, given that they can be modeled with a beta distribution.

b) Given the collected data and using the beta distribution model, what is the probability that the failure rate is below 0.15?

Example: Beta distribution (cont.)

- **Solution:** An energy lab designed a cutting-edge wind turbine that is completely different from any previous design. In the preliminary lab tests of 115 turbines, 92 turbines experienced no failure.

a) Find the expected failure probability, given that they can be modeled with a Beta distribution.

$\alpha = 115 - 92 = 23, \beta = 92.$

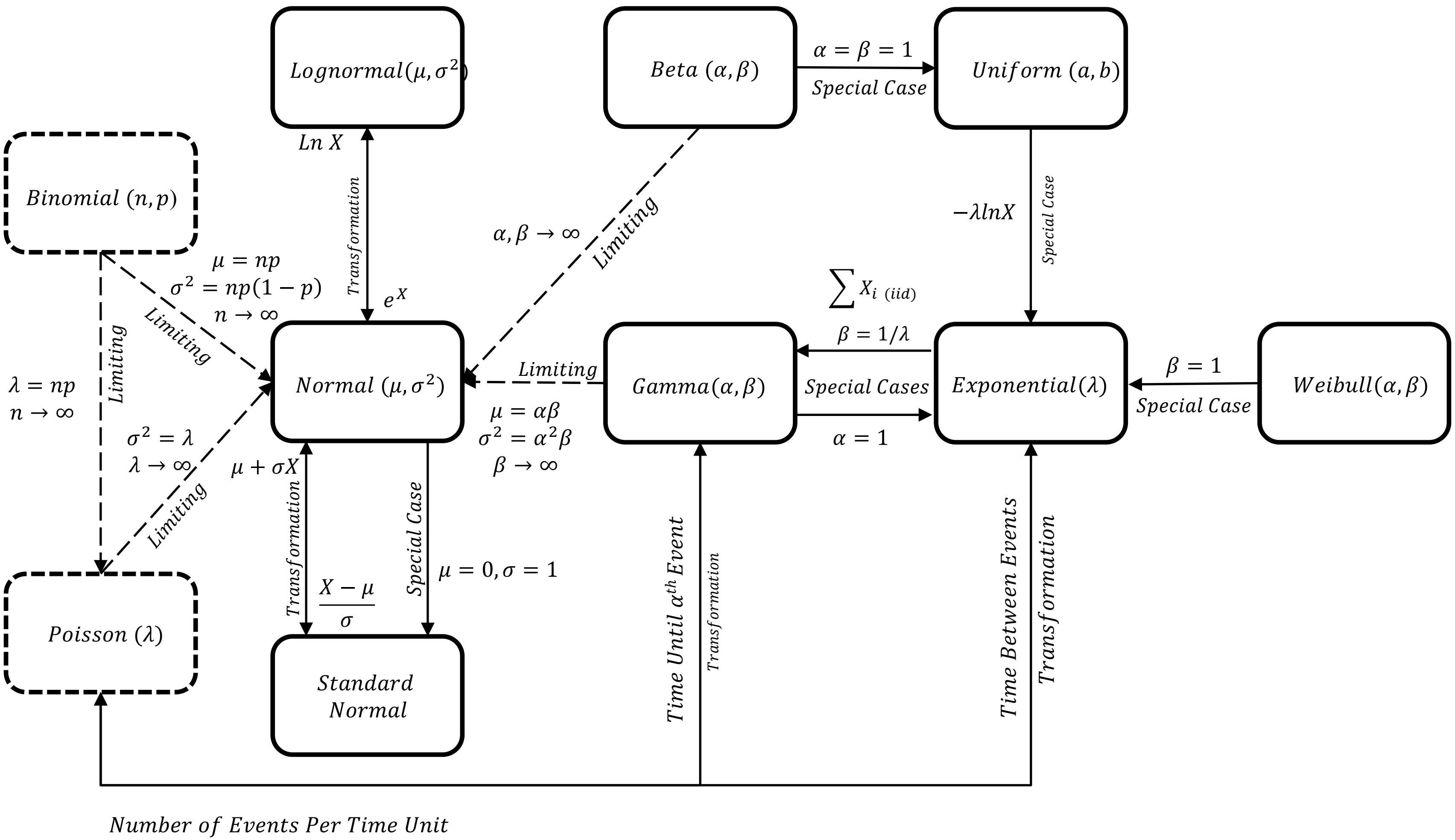
Expected probability of failure, X, is $E(X) = \alpha/(\alpha+\beta) = 0.2$

b) Given the collected data and using the eta distribution model, what is the probability that the failure rate is below 0.15?

$\Pr(x < 0.15) = 0.082$

In Excel: BETA.DIST(0.15, 23, 92, TRUE)=0.082

Relationship between distributions



Reliability Analysis

Module 4: Basic Reliability Math: Statistics

Prof. Katrina M. Groth
Mechanical Engineering Department
Center for Risk and Reliability
University of Maryland
kgroth@umd.edu

Reminder: Key concepts from Modules 2, 3

- Module 2:
  - Laws of probability & calculus of probability
  - Boolean algebra for manipulating events
  - Set theory for defining events
  - Joint vs. marginal vs. conditional probability
  - Common parametric probability distributions & their CDFs, PDFs
- Module 3:
  - Definitions of reliability $R(t)$, and hazard rate $h(t)$, MTTF, MRL
  - Bathtub curve and implications for reliability
  - Component reliability and commonly-used distributions

Objectives for Module 4

- Review descriptive statistics
- Introduce empirical distributions (non-parametric pdfs and cdfs)
- Introduce methods for parameter estimation
  - Point estimates & Interval estimates
- Hypothesis testing (aka statistical tests) for Goodness-of-fit
  - ….are these results possible from my chosen model?
    - Why? The model might not be capable of producing those results.

Data, samples, and populations

- Statistics – process of collecting, analyzing, organizing, and interpreting data about a population of interest.
- Data: Sometimes we have data for the full population, but more often we collect *samples* and want to make decisions about *populations*.
- Important terms:
  - **Sample size:** The number of observations in the sample
  - **Independently and identically distributed (i.i.d.):** An assumption made in classical statistics that all the observations are collected under the same conditions and represent independent events.
  - **Realization**: An observed value of the r.v.
  - **Statistical inference**: Using data to make conclusions about populations
    - Deduction: “Given a population, what will a sample look like?”
    - Induction: “Given a sample, what can be inferred about a population?”
  - Samples are imperfect -- we must make various corrections depending on sample characteristics such as sample size, randomness, non-independence, censoring…

Descriptive statistics from data

- Given a sample of size n, $\{x_1, x_2, \dots, x_n\}$ from the distribution of r.v. X, we can obtain descriptive statistics or sample moments to describe the sample.
- **Sample arithmetic mean:**

$$E(X) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- **Population variance:**

$$Var(X) = s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- **Sample variance:**

$$Var(X) = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

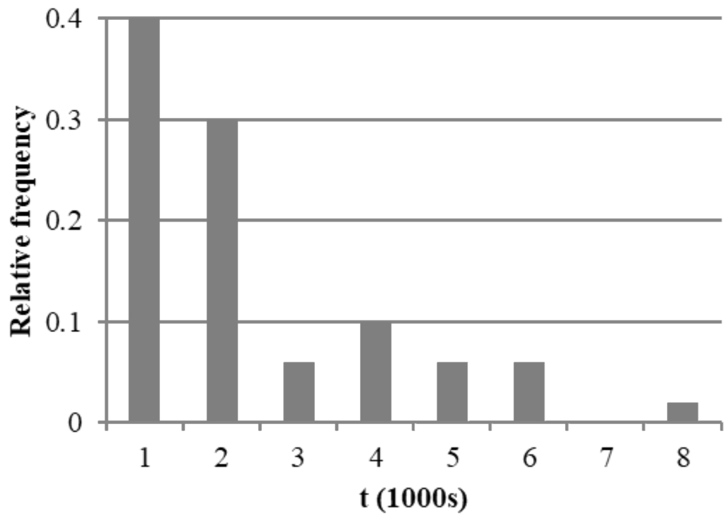
  - Because $\bar{x}$ is estimated from the same sample, the sample variance is biased. This bias is removed by multiplying by $\frac{n}{n-1}$.

Histograms & Empirical distributions

- **Histograms** help visualize data and inform other analysis. They can also display data where no known parametric distribution is a good fit (aka, an **empirical distribution**).
- For example, 50 lightbulb failure times, binned into 8 groups:

Given

t_i (×1000)		Freq.	Relative frequency (n_{t_i}/n)
0	1	20	0.4
1	2	15	0.3
2	3	3	0.06
3	4	5	0.1
4	5	3	0.06
5	6	3	0.06
6	7	0	0
7	8	1	0.02
Total:	**n=**	**50**	**1**



Histogram (Empirical distribution)

100 cables break at the following force, X (lb). What distribution should be used for X? We can use a histogram to visualize the data.

Breaking Strength of Cable (pounds)

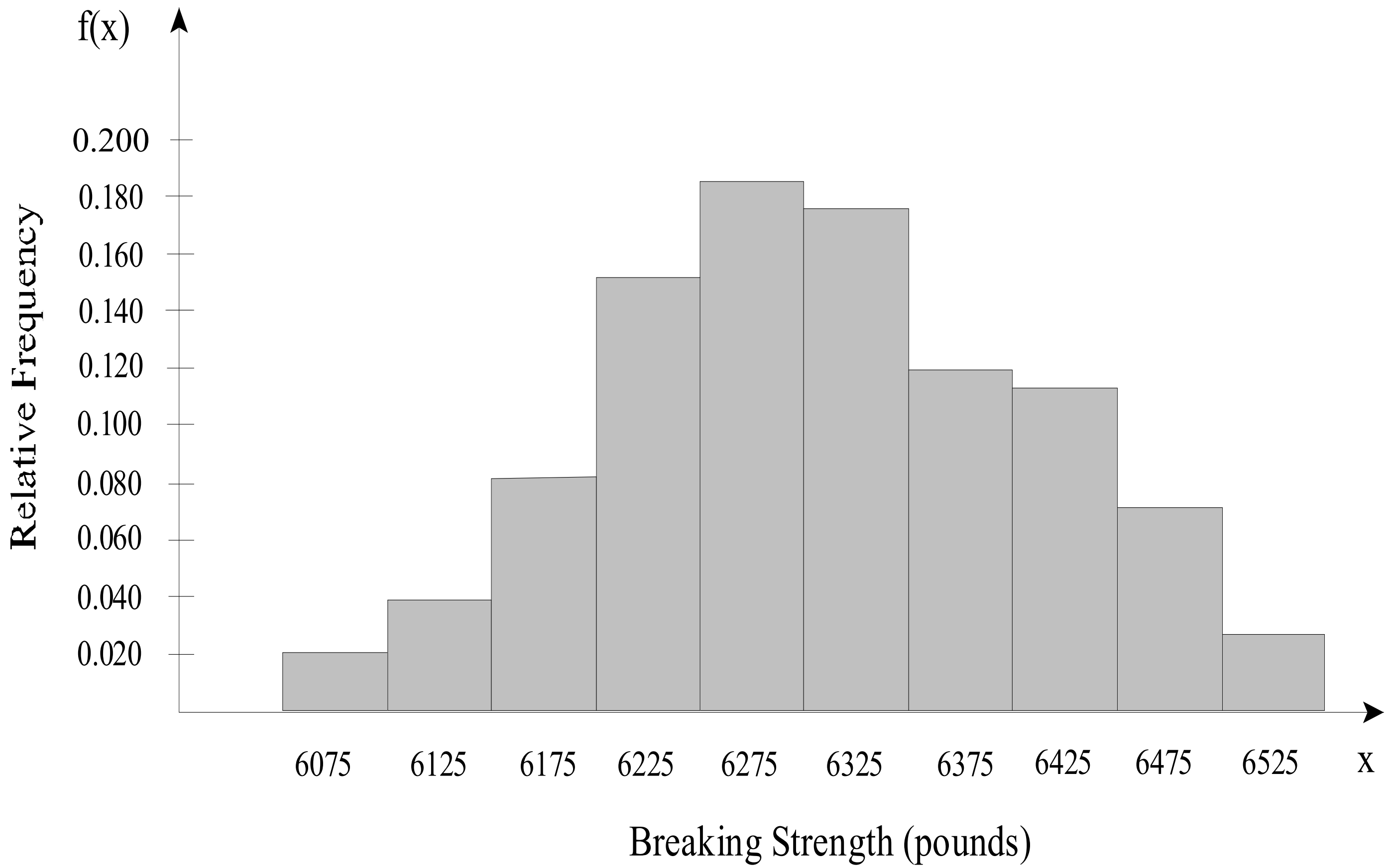
Observations									
6182	6428	6374	6505	6295	6533	6305	6423	6219	6239
6312	6377	6112	6295	6187	6295	6112	6259	6413	6533
6275	6166	6318	6378	6395	6318	6384	6336	6245	6239
6302	6216	6475	6355	6187	6338	6125	6205	6456	6182
6256	6395	6229	6152	6440	6352	6488	6298	6270	6347
6355	6435	6298	6095	6166	6333	6464	6413	6362	6264
6166	6320	6267	6248	6208	6464	6187	6404	6290	6320
6212	6312	6280	6282	6475	6325	6248	6361	6320	6344
6297	6408	6349	6259	6448	6275	6361	6251	6408	6236
6135	6399	6301	6302	6201	6245	6201	6067	6475	6428

Creating a histogram & empirical distribution

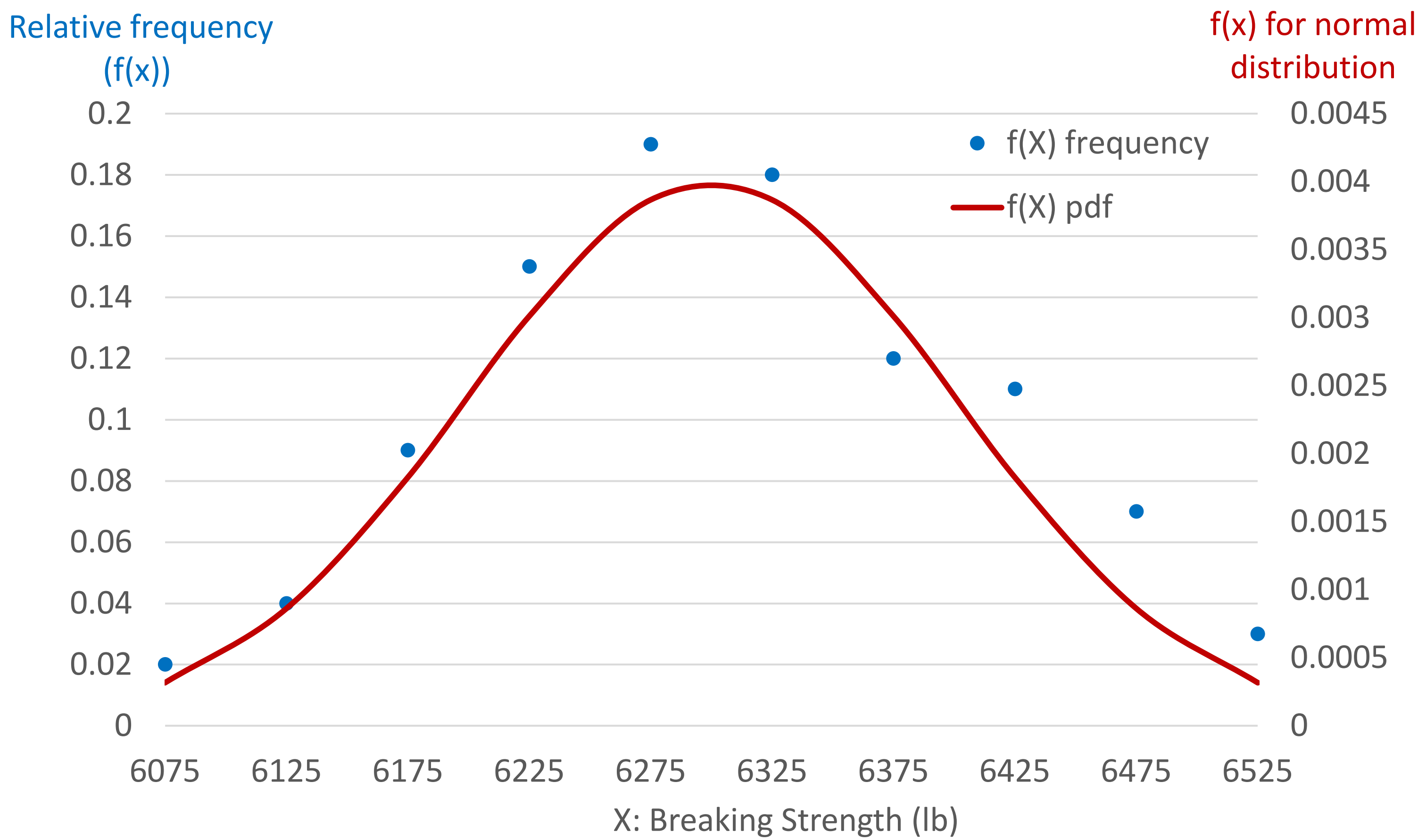
- First, bin the data into equal sized intervals. Then, count the frequencies in each bin and normalize it.

Interval	Midpoint X	Empirical frequency n_e	Relative frequency $f(x) = \frac{n_e}{n}$
$6050-6099$	6075	2	0.020
$6100-6149$	6125	4	0.040
$6150-6199$	6175	9	0.090
$6200-6249$	6225	15	0.150
$6250-6299$	6275	19	0.190
$6300-6349$	6325	18	0.180
$6350-6399$	6375	12	0.120
$6400-6449$	6425	11	0.110
$6450-6499$	6475	7	0.070
$6500-6549$	6525	3	0.030
Totals		100	1.000

Histogram of breaking strengths



Normal curve fit to data



PDF curve for $X \sim norm(\mu = 6300, \sigma = 100)$

Statistical inference

- **Statistical inference:** using sample data to answer questions about the distribution of an r.v.
- For example, how we can use a sample of failure times of an item, t_1, t_2, …, $t_{n,}$ to estimate λ. Therefore, we are inferring from a **specific** sample to a **general** distribution, i.e.,

  Part (sample) $\rightarrow$ Whole (population, model, distribution,)

- This cannot be achieved with **certainty**. Thus, two important aspects of inference are:

  (1) Estimating of distribution (model) parameters & characteristics
  (2) Statistical hypothesis tests

Parameter estimation

- We want to use a set of observations (data) to estimate the parameter(s) of a distribution.
- Desired properties of an estimator: unbiased, consistency, efficient, sufficient
- Let $f(x|\theta)$ denote a distribution (pdf) of r.v. X where θ represents an unknown parameter. Let $x_1, x_2, \ldots, x_n$ denote a sample from $f(x|\theta)$.
  - In frequentist statistics, the parameter set θ is *fixed* (but unknown) – we want to use our *random* sample of data to estimate it.
  - In Bayesian statistics, the parameter set is θ a *random* variable, and we use our *fixed* data (evidence) to update our knowledge of the parameter.

Parameter estimation: Method of moments

- Matching sample moments (e.g., sample mean ($\bar{x}$) and sample variance (s^2) to the distribution moments (e.g., mean ($\hat{\mu}_X$) and variance ($\hat{\sigma}_X^2$)):

$$\bar{x} \Longrightarrow \hat{\mu}_X$$

$$s^2 \Longrightarrow \hat{\sigma}_X^2$$

Example: Sample descriptive stats, method of moments

- **Example**: We want to verify the accuracy of cutting machine. To do this, five samples were cut to different lengths. We recorded the lengths, *X*, and room temperatures, *T*, as:

$$\mathrm{X} = 20.1, 22.13, 19.3, 18.5, 24.3\ \mathrm{mm}$$
$$\mathrm{T} = 20.13, 25.3, 22.4, 19.3, 24.5\ ^{\circ}\mathrm{C}$$

  a) Find the sample mean and variance to estimate the parameters of two normal distributions representing *X* and *T*.
  b) (If time) Use the distributions you found to calculate $Pr(X<19mm)$ and $Pr(T<20.5\ ^{\circ}\mathrm{C})$
  c) (If you finish early, also find the correlation between *X* and *T*).

Example: Sample descriptive stats, method of moments

- **Solution a):** Find the mean

$$\hat{\mu}_x = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}(20.1 + 22.13 + 19.3 + 18.5 + 24.3) = \mathbf{20.87}\boldsymbol{mm}$$

$$\hat{\mu}_T = \bar{T} = \frac{1}{n}\sum_{i=1}^{n} T_i = \frac{1}{5}(20.13 + 25.3 + 22.4 + 19.3 + 24.5) = \mathbf{22.33}℃$$

Example: Sample descriptive stats, method of moments

- **Solution a) (cont.):** Find the Variance:

$$\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{1}{5-1}[(20.1 - 20.9)^2 + (22.13 - 20.9)^2 + (19.3 - 20.9)^2 + (18.5 - 20.9)^2 + (24.3 - 20.9)^2] = \mathbf{5.51}\boldsymbol{mm}$$

$$\hat{\sigma}_T^2 = s_T^2 = \frac{1}{n-1}\sum_{i=1}^{n}(t_i - \bar{T})^2$$

$$= \frac{1}{5-1}[(20.13 - 22.3)^2 + (25.3 - 22.3)^2 + (22.4 - 22.3)^2 + (19.3 - 22.3)^2 + (24.5 - 22.3)^2] = \mathbf{6.89}\ ℃$$

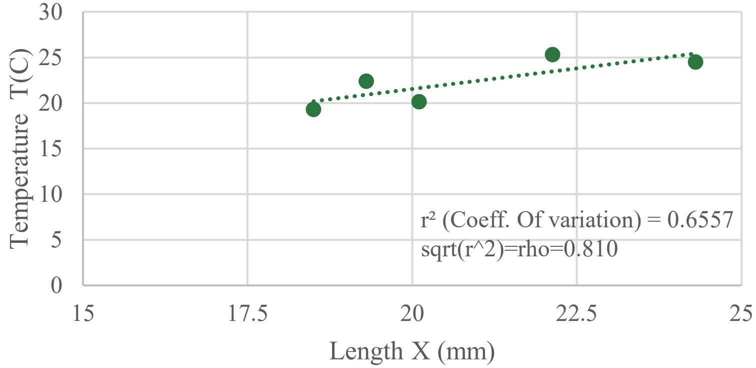
Example: Sample descriptive stats, method of moments

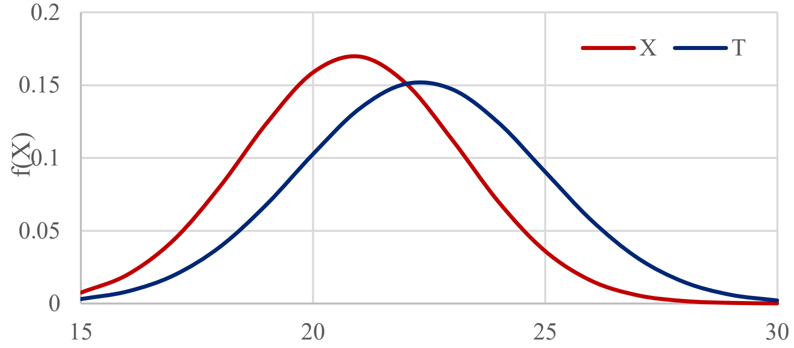
- **Solution b):** Find *Pr(X<19mm)* and *Pr(T<20.5°C)*

$$\mathbf{Pr}(\boldsymbol{X} < \mathbf{19}\boldsymbol{mm}) = Norm.dist\left(20.87, \sqrt{5.51}, true\right) = \mathbf{0.213}$$

$$\mathbf{Pr}(\boldsymbol{T} < \mathbf{20.5}°\mathrm{C}) = Norm.dist\left(22.33, \sqrt{6.89}, true\right) = \mathbf{0.243}$$

- **Solution c):** After class, you can also calculate correlation between these two sets of data.
  - The correlation is 0.810, which indicates a strong positive relationship between sample length and room temperature.





Parameter estimation: Maximum likelihood estimation

- The **likelihood function** of a parameter θ, given the evidence E, (e.g., n sample data points), is given by the probability of the joint occurrence of the data.

$$L(\theta|E) = \prod_{i=1}^{n} f(E_i|\theta)$$

Note: when $f(x|\theta)$ is viewed as a function of (variable) x with θ fixed, it's called a pdf. When it's viewed as a function of (variable) θ with x fixed, it's called a likelihood function, $L(\theta|x)$

  - Likelihood Function L and log-likelihood Λ for complete data:

$$L(\theta|x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

$$\Lambda(\theta|E) = \ln[L(\theta|E)] = \sum_{i=1}^{n} ln[f(x_i|\theta)]$$

Maximum Likelihood Estimate: Definition

- **Maximum Likelihood Estimate (MLE)** of θ is the value of $\hat{\theta}$ such that

$$L(\hat{\theta}|x_1, x_2, \dots, x_n) \geq L(\theta|x_1, x_2, \dots, x_n),$$

for every value of θ.

- The statistic $\hat{\theta}$ is a r.v. called the ML estimator (MLE) of θ.

- We solve for θ by maximizing the likelihood function, i.e.:

$$\left.\frac{\partial L}{\partial \theta}\right|_{\theta=\hat{\theta}} = \left.\frac{\partial lnL}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0$$

Why MLE?

- Useful for comparing distribution fits to the data: distribution with higher likelihood L is a better fit than one with a lower L.

- The uncertainty or confidence intervals of the parameters can be obtained directly.

- One solution that is not affected by the choice of plotting positions (used in linear regression).

Example: MLE

- **a)** For a r.v. X, representing failure times of a widget, you have a set of data of n failure times $(x_1, x_2, \ldots x_n)$. Write the likelihood function of this data. Assume that the likelihood of each data point is given by a normal distribution.

- **b)** Compute the point estimate of the normal parameters (μ, σ) using MLE.

Example: MLE

- **a) Solution:** Write the likelihood function of n failure times, where the likelihood of each data point is given by a normal distribution.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}$$

$$L(x_1, x_2, \ldots, x_n | \mu, \sigma) = \prod_{i=1}^{n} f(x_i | \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]}$$

$$\boldsymbol{L(x_1, x_2, \ldots, x_n | \mu, \sigma) = \frac{1}{\sigma^n (2\pi)^{n/2}} \prod_{i=1}^{n} e^{\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]}}$$

Example: MLE

b) Solution: First, compute the log-likelihood by taking $\ln(L)$ and simplifying it:

$$\ln(L) = -n \cdot \ln\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

- Then differentiate the log-likelihood wrt to μ and σ:

$$\frac{\partial \ln(L)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{\mu n}{\sigma^2} = 0$$

$$\frac{\mu n}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i$$

$$\mu n = \sum_{i=1}^{n} x_i$$

$$\boldsymbol{\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

$$\frac{\partial \ln(L)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

$$\frac{n}{\sigma} = \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$n = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_{i} - \hat{\mu})^2$$

$$\boldsymbol{\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2}$$

- In most cases however, unbiased estimators are used so we multiply the variance estimate by $n/(n-1)$ to get the familiar form with 1/(n-1).

MLE Parameters of normal and lognormal distributions

Continuous, for n data points representing t_i failure times:

Normal Distribution (biased correction)

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} t_i$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(t_i - \hat{\mu})^2$$

Lognormal Distribution (bias corrected)

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} ln t_i$$

$$\hat{\sigma}^2 = \frac{1}{\mathrm{n}-1}\sum_{\mathrm{i}=1}^{n}(\mathrm{ln} t_i - \hat{\mu})^2$$

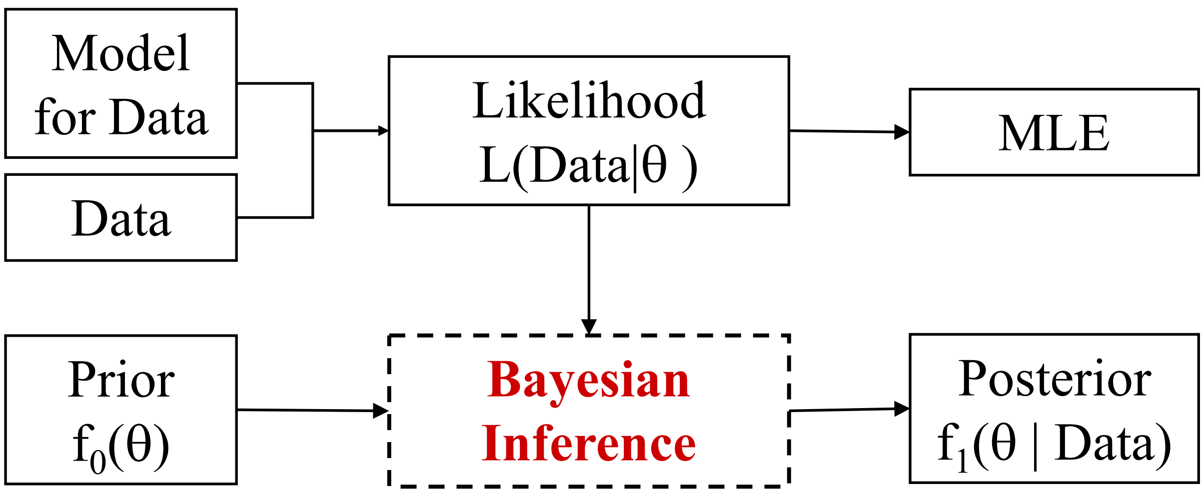
More to be discussed in Ch 5.

Why Bayesian parameter estimation methods?

- To incorporate prior knowledge of the phenomena
  - Combinations of extensive past experience, physical/chemical theory can provide prior information to form a framework for inference and decision making. **--** *it would be shortsighted to ignore this prior information* – even if we do have “enough” data.
- For problems that have insufficient data for (or violate other assumptions inherent in use of) frequentist methods.
- To directly quantify uncertainty

Bayesian parameter estimation

- Bayesian parameter estimation methods are closely related to maximum likelihood methods
- The basic process of Bayesian parameter estimation is to update our prior knowledge of an *unknown* parameter of interest using new data



- Notice use of $L(\theta|x_1, x_2, \ldots, x_n)$ in MLE but $L(x_1, x_2, \ldots, x_n|\theta)$ in Bayesian estimation; MLE treats $L(\theta|E)$ and $L(E|\theta)$ as equivalent, Bayesian est. does not.

Bayesian parameter estimation

- The basic process of Bayesian parameter estimation is to update our prior knowledge of an *unknown* parameter of interest using new data (evidence).
- Using the continuous form of Bayes' theorem w/ evidence E:

$$f_1(\theta|E) = \frac{f_0(\theta)L(E|\theta)}{\int_{-\infty}^{\infty} f_0(\theta)L(E|\theta)d\theta}$$

  - **Prior distribution:** $f_0(\theta) or\ \pi_0(\theta)$ pdf representing the distribution of parameter θ estimated from prior information, or the degree of belief of expert judgment, etc.
  - **Likelihood function:** $L(E|\theta)$ or $\Pr(E|\theta)$ representing actual data (*evidence*) about the r.v.
  - **Posterior distribution:** $f_1(\theta|E)$ or $\pi_1(\theta|E)$ or $f_{\Theta|X_1,X_2,X_3,..}(\theta|x_i, x_2, x_x, \dots)$

Bayesian parameter estimation (cont.)

- **Discrete Form**
  - If available data are represented best by a discrete probability function $\Pr(data|\theta)$, then,

$$\Pr(\theta|E) = \frac{\Pr(E|\theta)\Pr(\theta)}{\sum_{i=1}^{\infty}\Pr(E|\theta_i)\Pr(\theta_i)}$$

Prior distribution and likelihood function are both discrete

- **Mixture Form**

$$f_1(\theta|E) = \frac{\Pr(E|\theta)\, f_0(\theta)}{\int_{-\infty}^{\infty}\Pr(E|\theta)\, f_0(\theta)d\theta}$$

Prior distribution is continuous likelihood function is discrete

More to be discussed in Ch 5.

Interval estimation

- Why? Different amounts of data influence our (un)certainty about the parameter
- An attempt to answer: How 'far off' could we be?
- Frequentist approach: Confidence intervals
  - Quantify uncertainty due to sampling error (i.e., limited number of samples)
- Bayesian approach: Credible intervals (aka Bayesian probability interval)
  - Statements about the probability of a parameter based on the evidence

- **Not a characterization of uncertainty due to incorrect model selection or assumptions!**

Parameter estimation: Confidence intervals

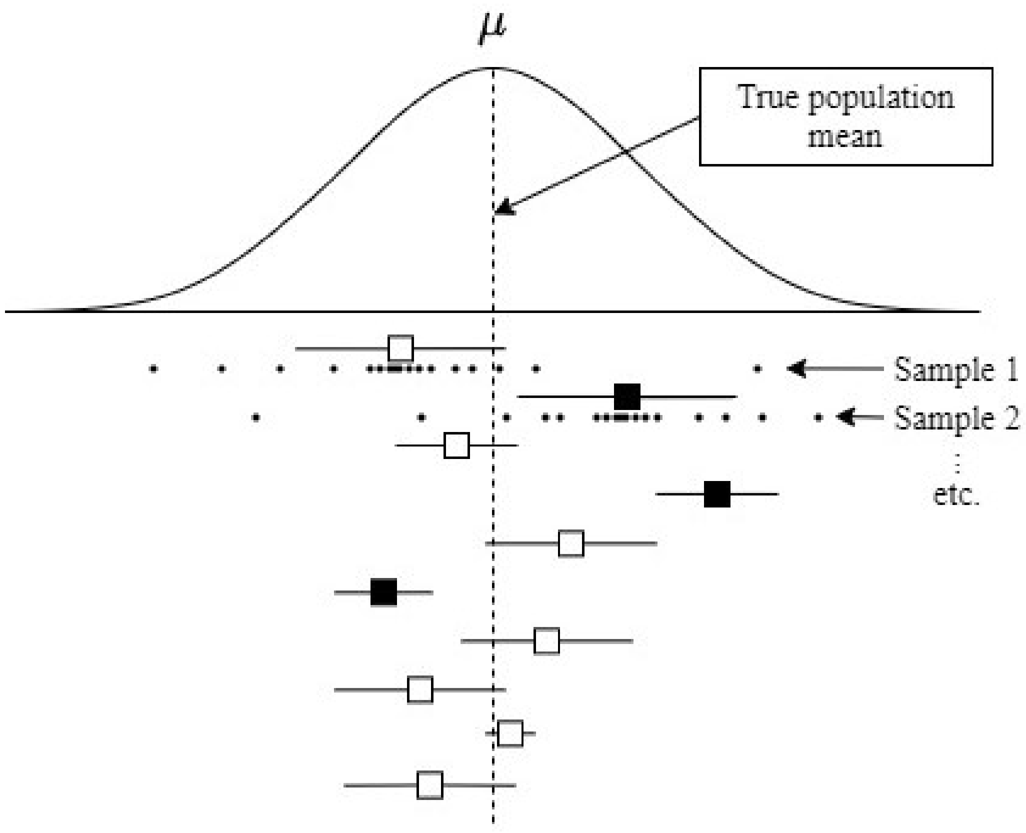
- **Confidence intervals** are a frequentist expression of uncertainty about estimated parameter values.
  - The main purpose is to find an interval with a high probability of containing the true (but unknown) value of a parameter θ
  - Often misinterpreted as a probability that θ is in this interval.
- Consider r.v.s $\theta_l(x_1, \dots, x_n)$ and $\theta_u(x_1, \dots, x_n)$, such that the probability that interval $[\theta_l, \theta_u]$ contains the true value of θ is:

$$\Pr[\theta_l(x_1, \dots, x_n) < \theta < \theta_u(x_1, \dots, x_n)] = 1 - \gamma$$

- Interval $[\theta_l, \theta_u]$ is the **(two-sided) k% confidence interval** for parameter θ with **confidence level** (**confidence coefficient**): $k\% = 100(1-\gamma)\%$.
- **Ex:** For a 90% confidence interval around $R(t)$, $\gamma = 0.1$. Thus, $\Pr(R_L \le R(t) \le R_U) = 1 - \gamma = 0.9$.

Illustration of how confidence intervals are generated

- See a known distribution with mean parameter μ, and 10 rows of samples from the distribution.
- Whiskers represent 70% confidence intervals.
- What percent of the generated intervals contain the true mean? (Hint: box colors)
- Interpretation:
  - In $100(1-\gamma)\%$ of repetitions, μ falls between μ_L and μ_U.
  - Not: the probability that μ is in the given interval.



Confidence intervals express uncertainty due to sample size

- **Example**: If 100 units are tested, consider two situations for exponential parameter estimation:
  - **Case 1: For r = 1 $\boldsymbol{failure}, \boldsymbol{t_0}$ = 10 hrs**
    - $\mathrm{T} = 10 \cdot 100 = 1000$ test hours
    - $\hat{\lambda} = \frac{r}{T} = \frac{1}{1000} = 10^{-3}\mathrm{hr}^{-1}$
  - **Case 2: For r = 10 failures, $\boldsymbol{t_0}$ = 100 hrs**
  - $\mathrm{T} = 100 \cdot 100 = 10{,}000$ test hours
  - $\hat{\lambda} = \frac{\mathrm{r}}{\mathrm{T}} = \frac{10}{10{,}000} = 10^{-3}\mathrm{hr}^{-1}$
  - Both gives you the same $\hat{\boldsymbol{\lambda}}$ estimate, but one has more data. The MLE parameter is the same, but the confidence interval is different for these two datasets.

Normal distribution confidence interval

- **If σ is known:**
  - Construct a confidence interval for μ based on a random sample $\{x_1, \dots, x_n\}$ from a normal distribution: $X \sim norm(\mu, \sigma)$
  - Sample mean is normally distributed: $\bar{x} \sim norm\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Then use the Z transformation

$$\Pr\left(z_{\left(\frac{\gamma}{2}\right)} \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\left(1-\frac{\gamma}{2}\right)}\right) = 1 - \gamma$$

  - Thus:

$$\boxed{\Pr\left(\bar{x} - z_{\left(1-\frac{\gamma}{2}\right)} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\left(1-\frac{\gamma}{2}\right)} \frac{\sigma}{\sqrt{n}}\right) = 1 - \gamma}$$

Confidence interval example

- **Example:** Consider the following sample, taken from a normal distribution with known $\sigma = 3.1$. Find the 90% confidence interval around μ.

5.19, 11.84, 11.54, 8.44, 6.87, 12.96, 13.69, 14.39, 7.83, 10.62

Confidence interval example

Solution:

- $x_i = \{5.19, 11.84, 11.54, 8.44, 6.87, 12.96, 13.69, 14.39, 7.83, 10.62\}$
- $n = 10$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}\cdot(103.37) = 10.34$$

- $\gamma = 0.1 \Longrightarrow \frac{\gamma}{2} = 0.05$
- Using Appendix A, $z_{\left(1-\frac{\gamma}{2}\right)} = 1.65$, so:

$$\Pr\left(\bar{x} - z_{\left(1-\frac{\gamma}{2}\right)}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\left(1-\frac{\gamma}{2}\right)}\frac{\sigma}{\sqrt{n}}\right) = 1-\gamma = 0.9$$

$$\Pr\left(10.34 - 1.65\cdot\left(\frac{3.1}{\sqrt{10}}\right) \le \mu \le 10.34 + 1.65\cdot\left(\frac{3.1}{\sqrt{10}}\right)\right) = 0.9$$

$$\Pr(8.72 \le \mu \le 11.96) = 0.9 \Longrightarrow \boldsymbol{\mu} \in [\mathbf{8.72, 11.96}]$$

Normal distribution confidence interval

- **If σ is unknown:**
  - If σ is estimated from the sample as $\hat{\sigma} = s$, or if sample is small (e.g., $n < 30$), then
  - The r.v. $\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ follows the **Student's *t*-distribution** with *n-1* degrees of freedom, and the **confidence interval around $\boldsymbol{\mu}$** is given:

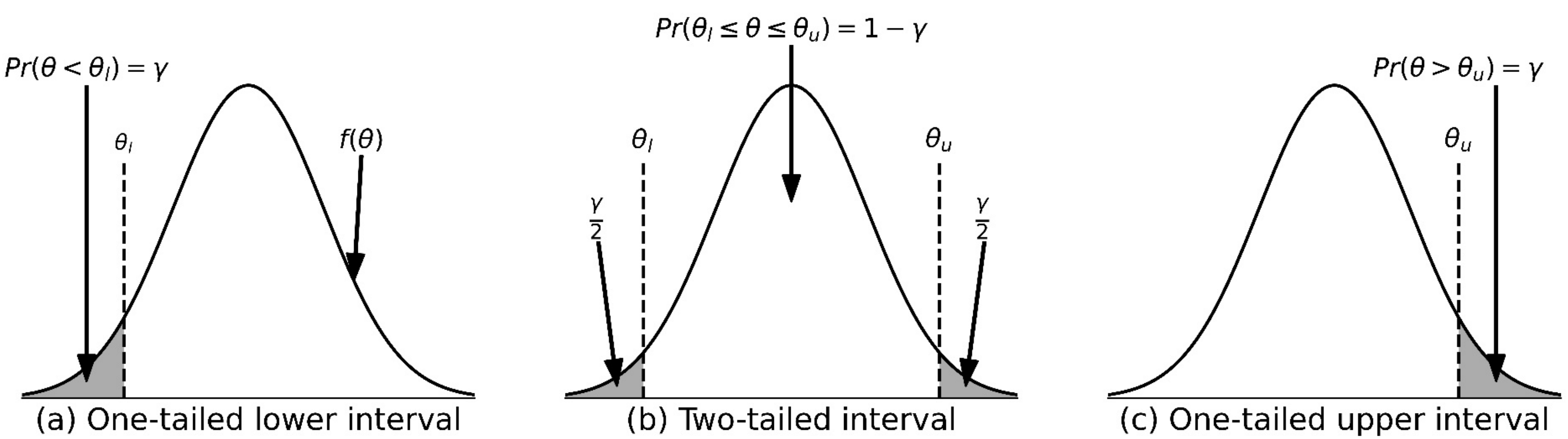
$$\Pr\left(\bar{x} - t_{\left(1-\frac{\gamma}{2}\right)} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\left(1-\frac{\gamma}{2}\right)} \frac{s}{\sqrt{n}}\right) = 1 - \gamma$$

  - Additionally, the **confidence interval around $\boldsymbol{\sigma^2}$** is given:

$$\Pr\left(\frac{(n-1)s^2}{\chi^2_{\left(1-\frac{\gamma}{2}\right)}[n-1]} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\left(\frac{\gamma}{2}\right)}[n-1]}\right) = 1 - \gamma$$

Bayesian credible intervals

- The **credible interval** (Bayesian probability interval) is the Bayesian analog to the confidence interval.
- Using the posterior distribution of the parameter θ:
$$\Pr(\theta_l \le \theta \le \theta_u) = 1 - \gamma$$
- The interval $[\theta_l, \theta_u]$ is the $100(1-\gamma)\%$ credible interval for θ.
- Interpretation:
  - There is a $100(1-\gamma)\%$ probability that the true value of θ is contained within the interval.



Credible interval vs. confidence interval

- Recall that the **confidence interval** is interpreted:
  - $k\% = 100(1-\gamma)\%$ confidence that any generated interval $[\theta_l, \theta_u]$ contains the true value of θ
  - Confidence about the **interval** $[\boldsymbol{\theta_l}, \boldsymbol{\theta_u}]$
- While the **credible interval** is interpreted:
  - Directly gives the probability that the true value of θ falls in the interval $[\theta_l, \theta_u]$
  - The interval has a $k\% = 100(1-\gamma)\%$ probability of containing θ
  - Confidence about the **value of $\boldsymbol{\theta}$**

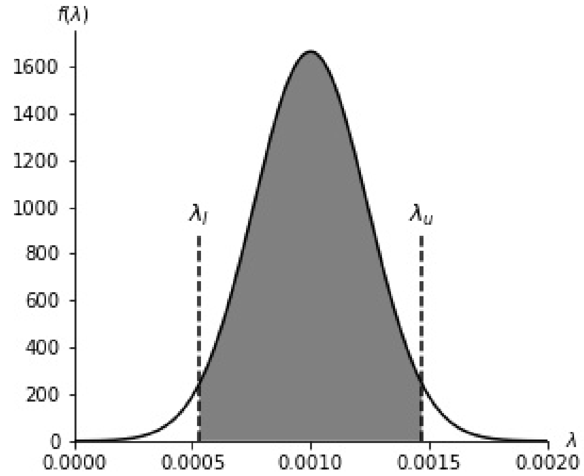
Example: credible intervals

- **Example**: A capacitor has a time to failure, *T*, that can be represented by the exponential distribution with a failure rate λ. The mean value and uncertainty about λ are modeled through a Bayesian inference with the posterior pdf for λ expressed as a normal distribution with a mean of 1×10^{-3} and standard deviation of 2.4×10^{-4}. Find the 95% Bayesian credible interval for the failure rate.

- Which means:
  - T~exp(λ) where $\lambda \sim norm(\mu = 1 \times 10^{-3}, \sigma = 2.4 \times 10^{-4})$

Example: credible intervals

- **Solution:**
- To obtain the 95% credible interval for the normal distribution on λ, we find the 2.5th and 97.5th percents of the normal distribution representing uncertainty on the parameter λ.
- $\lambda_l = \Phi^{-1}(0.025, 1 \times 10^{-3}, 2.4 \times 10^{-4}) = 5.30 \times 10^{-4}$,
- $\lambda_h = \Phi^{-1}(0.975, 1 \times 10^{-3}, 2.4 \times 10^{-4}) = 1.47 \times 10^{-3}$.
- Therefore, $\mathbf{5.30 \times 10^{-4} < \hat{\lambda} < 1.47 \times 10^{-3}}$

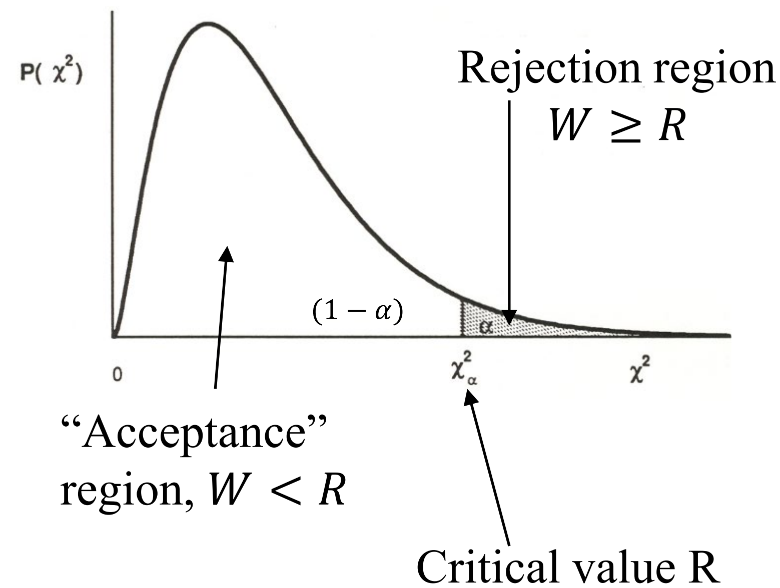


Hypothesis testing and goodness of fit

- *Question*: is observed data consistent with the assumption that it comes from a specific theoretical distribution (i.e., with known form and parameter)?
  - It is necessary to choose properly which distribution appropriately fits a set of data.
  - Even with the obvious choices (e.g., Weibull distribution for failure times), you may need to check for appropriateness of the distribution.
  - Test whether the **hypothesis** that the data originate from a known distribution is true.
- For this, reliability engineers rely on several methods to see which distribution is the best fit:
  1) Chi-Square Goodness-of-fit test
  2) Kolmogorov-Smirnov (K-S) test
  3) Linear Regression Analysis (Least Squares estimation)
- One or more of these methods may be applied to support your choice of distribution.

Generic hypothesis estimation procedure

1. Set **Null Hypothesis** (H_0: $\theta = \theta_0$)
   - Define an appropriate alternative hypothesis (H_1: $\boldsymbol{\theta} \neq \boldsymbol{\theta_0}, \theta = \theta_1, \theta < \theta_0, \theta > \theta_0$, etc.)
2. Define a **significance level**, γ
3. Calculate a **test statistic** (W) using sample data.
4. Calculate **critical value** of the test statistic (R) (i.e., boundary of rejection region)
5. Decision: If the test statistic from the sampled data is in the rejection region, reject H_0; otherwise, don't reject H_0.



Note: we can only *fail to support* the null hypothesis, so we can only reject it rather than positively affirm it as truth. Therefore, the hypothesis is either *rejected* or *not rejected* rather than *accepted* or *not accepted*.

Chi-squared test in frequentist statistics

- Based on a statistic that has an approximate χ^2 distribution.
- Compares observed frequencies (o_i) of data to the expected frequencies (e_i) generated from the hypothesized distributions.
  - Suitable for data sets which are frequencies or counts (not probabilities).
  - Data must be in **mutually exclusive intervals**.
- Checks:
  - The value of the cell "expected" should be ≥ 5 in at least 80% of the cells, and no cell should have an "expected" of less than 1.
    - If not: group some intervals. Only adjacent intervals may be grouped together.
    - This assumption is most likely to be met if the sample size $\geq \#intervals \cdot 5$
- Test statistics χ^2 is a value of an r.v. whose sampling distribution is approximated very closely by the chi-squared distribution with:
  - $v = k - m - 1$ degrees of freedom.
  - (k = # of intervals, m = # of parameters estimated from the data)

Chi-square goodness-of-fit test

- The steps in the **Chi-Square Goodness-of-Fit Test** are:
  a) Choose the hypothesized distribution, H_0 (e.g., an exponential distribution with $\lambda = 2.1$)
  b) Select a significance level (γ)
  c) Calculate the test statistic, W:

$$W = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i}\text{; where} \quad \begin{aligned} o_i &= \text{observed frequency at } i \\ e_i &= \text{expected frequency at } i \end{aligned}$$

  e) Establish a critical value (edge of rejection region) $R = \chi^2_{(1-\gamma)}[df]$
    - df=k-m-1: k = # of intervals, m = # of parameters estimated from the data

      Table in Appendix A
      Excel: chisq.inv(1-γ, df)
      Matlab: chi2inv(1-γ, df)

  e) If $W > R$ reject H_0; otherwise do not reject.

Chi-square goodness-of-fit test

- **Example**: Consider the data below for X, the number of replacement parts which need to be ordered each week by a repair facility. Test the hypothesis that these come from a Poisson distribution. Use a significance level of $\gamma = 10\%$.

Number of Failed Parts/Week (x)	Observed Frequency (o_i)
0	18
1	18
2	8
3	5
4	2
5	1

Hints:

- Calculate the Poisson parameter first
- Don't get the total number of weeks confused with the total number of parts ordered.

Chi-square goodness-of-fit test

- **Solution:** We are told to test the fit to a Poisson distribution, but we're not given a parameter, so first the parameter μ has to be estimated:

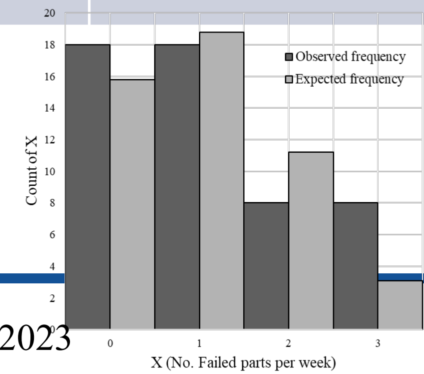
$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n\ weeks} \cdot \sum_{i=1}^{n} x_i\ parts = \frac{62}{52} = 1.192\,\frac{\text{parts}}{\text{week}}$$

- For this test, use k (number of intervals) = 4 [because the last three rows should be combined].
  - The grouping is a good idea when e_i has a very low value or differs by a large factor (5-10) different from other e_i's calculated.

Chi-square goodness-of-fit test

- **Solution:** Test the hypothesis that these data come from a Poisson distribution

Number of Failed Parts/Week (x_i)	Observed Frequency (o_i)	Hypothesized or Expected Frequency (e_i)	$W = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i}$
0	18	$52 \times \Pr(X = 0) = 15.8$	0.31
1	18	$52 \times \Pr(X = 1) = 18.8$	0.036
2	8	$52 \times \Pr(X = 2) = 11.2$	0.923
3	5	$52 \times \Pr(X = 3) = 4.5$	
4	2 } 8	$52 \times \Pr(X = 4) = 1.3$ } 6.1	0.588
5	1	$52 \times \Pr(X = 5) = 0.3$	
$\sum x_i = 62$ parts	n= 52 weeks	Check: sum = 52 weeks	$\mathbf{W = 1.86}$



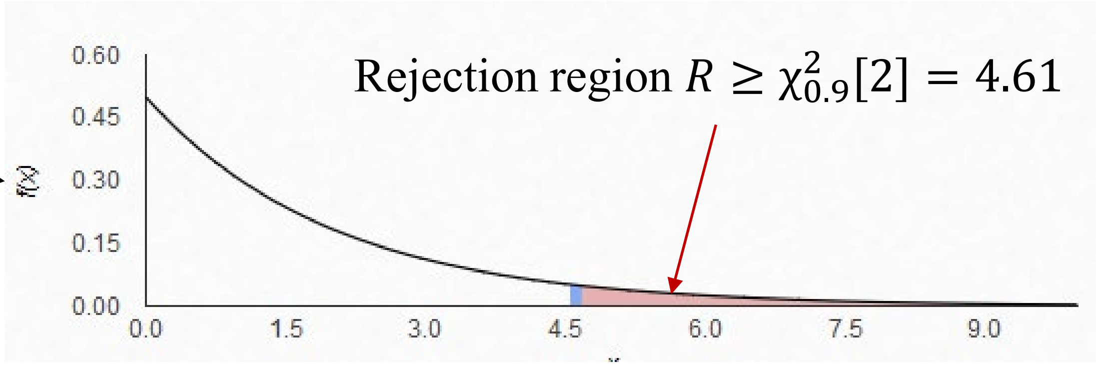
Chi-square goodness-of-fit test

- **Solution:** With k = 4 bins and m=1 (because 1 parameter, $\hat{\mu}$, is estimated from the data).

$$Test\ stat: \chi^2_{0.9}[k-m-1] = \chi^2_{0.9}[4-1-1] = \chi^2_{0.9}[2] = 4.61$$
$$Reject \geq (\chi^2_{0.9}[2] = 4.61)$$

- Our $W = 1.86$ so $W < R$
- So the hypothesis **is not rejected.**

pdf for $\chi^2_{0.9}[2]$



K-S (Kolmogorov-Smirnov) Test

- χ^2 requires substantial amounts of data to provide frequencies
- K-S test uses each individual data point (not frequencies) and is effective for small samples.
- Suppose for an r.v. T, we observe an **ordered sample** of data $t_1 \leq t_2 \leq t_3 \leq \cdots \leq t_n$

- The **sample cdf** $S_n(t)$ is defined for the ordered sample as:

$$S_n(t) = \begin{cases} 0 & -\infty < t < t_{i=1}, \\ \dfrac{i}{n} & t_i \leq t < t_{i+1}, i = 1, 2, \dots, n-1, \\ 1 & t_n \leq t < \infty. \end{cases}$$

The K-S Procedure

- **K-S Test Procedure:**
  1) Choose a hypothesized cumulative distribution $F(t)$ for sample.
  2) Select a significance level γ for the test
  3) Define the rejection region $R > D_n(\gamma)$. Critical values of the test statistic $D_n(\gamma)$ can be obtained from Appendix A.
  4) If $D > D_n(\gamma)$, reject the hypothesized distribution and conclude that $F(t)$ does not fit the data. Otherwise, do not reject.
- The K-S Test is performed similar to the χ^2 test:
  - H_0: T has cdf $F(t)$
    H_1: T does not have cdf $F(t)$
  - The K-S test statistic, W, is defined:

$$W = \max[|F(t_i) - S_n(t_i)|, |F(t_i) - S_n(t_{i-1})|]$$

  - The critical value of test statistic is $R = D_n(\gamma)$
    - n: number of observations.
    - $D_n(\gamma)$: See Appendix A

Example: K-S test

- **Example:** Are the failure times 8, 20, 34, 46, 63, 86, 111, 141, 186 and 266 hrs adequately modeled by an exponential distribution with parameter $\lambda = 0.01/\text{hr}$. Use $\gamma = 0.05$.
  - H_0: The r.v. T has cdf $F(t) = 1 - e^{-\lambda t} = 1 - e^{-0.01t}$
  - H_1: The r.v. T does not have cdf $F(t) = 1 - e^{-0.01t}$
- The calculated rejection region is:

$$R > D_{10}(0.05) = 0.409$$

  - $n = 10$ (number of observations)
  - $\gamma = 0.05$ (chosen significance level)

Example: K-S test solution

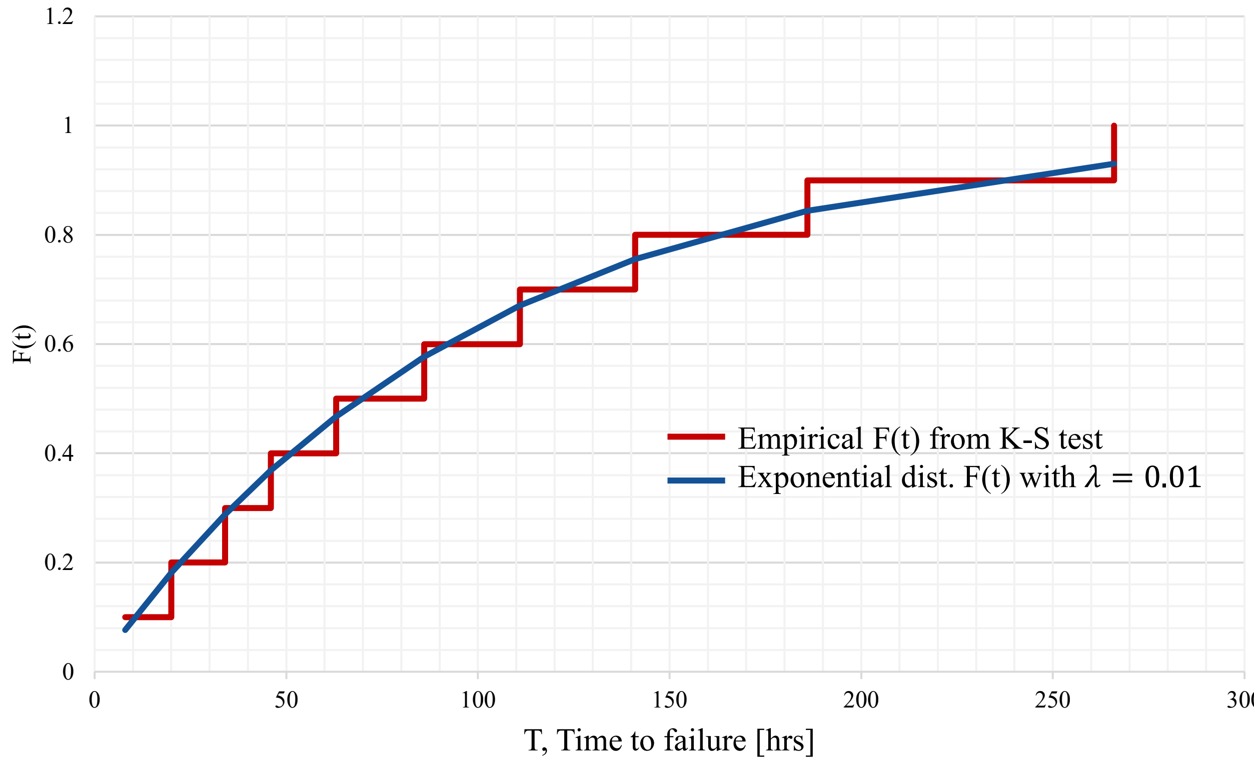
i	t_i	$S_n(t_i)$	$S_n(t_{i-1})$	$F(t_i)$	$F(t_i) - S_n(t_i)$	$F(t_i) - S_n(t_{i-1})$
1	8	0.1	0.0	0.077	0.023	0.077
2	20	0.2	0.1	0.181	0.019	0.081
3	34	0.3	0.2	0.288	0.012	0.088
4	46	0.4	0.3	0.369	0.031	0.069
5	63	0.5	0.4	0.467	0.033	0.067
6	86	0.6	0.5	0.577	0.023	0.077
7	111	0.7	0.6	0.670	0.030	0.070
8	141	0.8	0.7	0.756	0.044	0.056
9	186	0.9	0.8	0.844	0.056	0.044
10	266	1.0	0.9	0.930	0.070	0.030

$$W = \max[|F(t_3) - S_n(t_2)|] = \mathbf{0.088}$$

$W = 0.088 < 0.41$ thus we **don't reject $\mathbf{H_0}$** and conclude that the exponential distribution, T~exp($\lambda = 0.01$), is an acceptable model.

Example: K-S test solution shows good visual fit

- **Example:** By plotting the Empirical and fitted CDF we also see that the solution shows good visual fit to the exponential distribution.



Regression analysis

- **Definition**: In **Regression Analysis**, a correlation is made between a given dependent variable Y, and j number of independent variables $X_1, X_2, \dots, X_j$ (also called explanatory variables).
  - Explanatory variables $X_1, \dots, X_j$ do not have to be statistically independent
- Such a relation can be expressed as:

$$E[Y|X_1, \dots, X_j] = \beta_0 + \beta_1 x_1 + \ \dots \beta_j x_j$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \varepsilon$$

- This is the regression model where $\beta_1, \beta_2, \dots, \beta_k$ are regression parameters (or coefficients) and ε is the random model error (which can be expressed in any form of pdf distribution)
  - The normal distribution is commonly used to describe ε, with $E(\varepsilon) = 0$ and finite σ^2.
- An example of such a correlation would be the dependency between a crack's length and two factors: the number of cycles elapsed and the tensile force on the material.

Regression: Least squares method

- Consider a very simple example of linear regression case with dependent variable (*Y*) and one explanatory factor (*X*):

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- Take some data containing n pairs of observations:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n).$$

- The parameters are then estimated via MLE of the estimation, or minimizing the **sums of squares**:

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x)^2$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = 0, \qquad \frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = 0$$

Regression analysis

- The resulting **least square point estimates** of the regression parameters are:

$$\boldsymbol{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}} \quad and \quad \boldsymbol{\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

- Where:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad and \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

- The variance of Y may also be obtained as a result (assuming that random error is a normal distribution of mean zero):

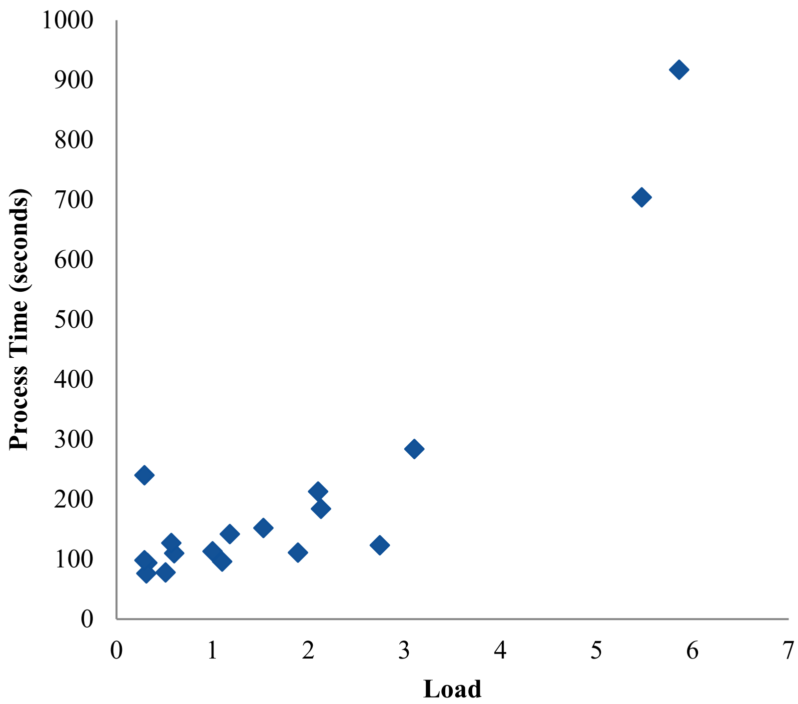
$$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} \text{ where } \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Example: Regression analysis

- **Example:** The following table and figure shows how long it takes for a computer to run a certain program as a function of the system load.

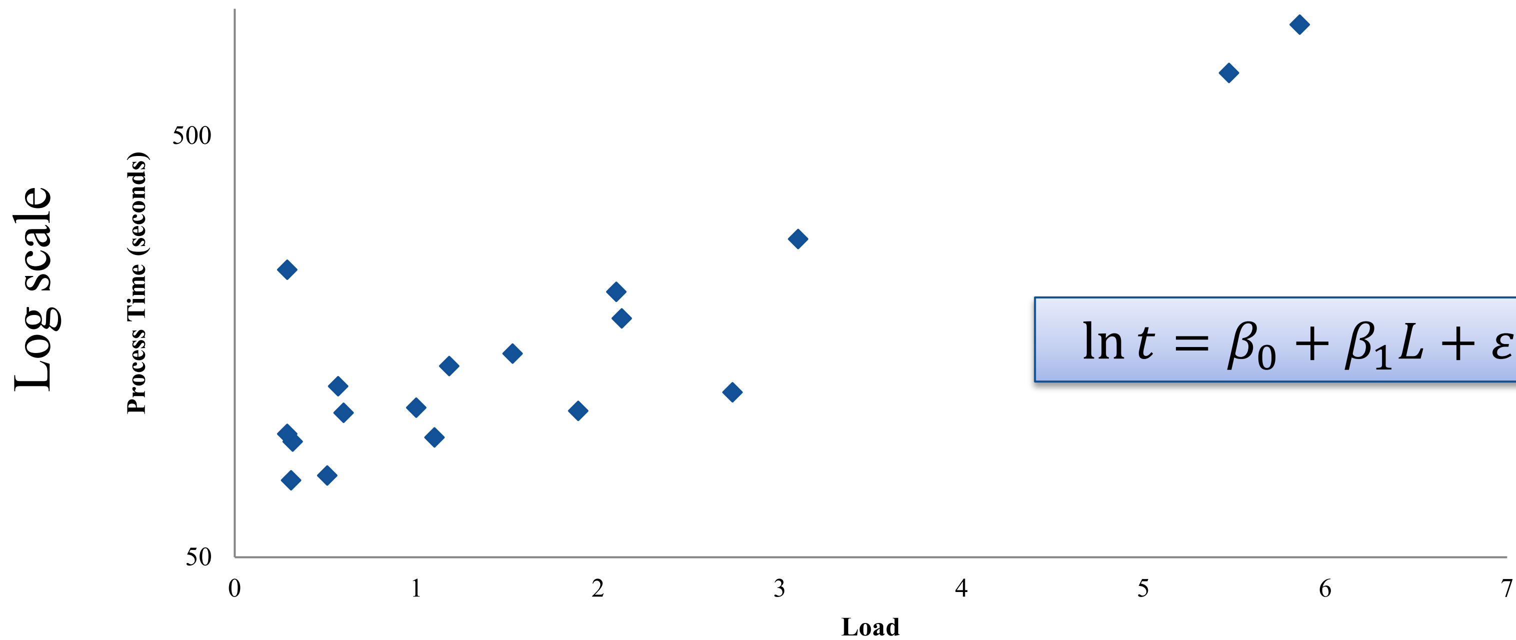
Process time, t_i	Load, L_i	Process time, t_i	Load, L_i
123	2.74	110	0.6
704	5.47	213	2.1
184	2.13	284	3.1
113	1	917	5.86
94	0.32	142	1.18
76	0.31	127	0.57
78	0.51	96	1.1
98	0.29	111	1.89
240	0.29	152	1.53

$n = 18$



Regression analysis

- **Example (cont):** Assume that this can be expressed semilog-linearly:



$$\ln t = \beta_0 + \beta_1 L + \varepsilon$$

- Where L is the system load and t is the processing time in seconds.
- Use linear regression analysis to find the point estimates for β_0, β_1, and σ. Assume the random error follows a normal distribution with a mean of 0.

Regression analysis (cont.)

- **Example (cont.):** Following the equations gets:

$$\ln(t) = \beta_0 + \beta_1 L + \varepsilon \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(L_i - \bar{L})(\ln(t_i) - \ln(\bar{t}))}{\sum_{i=1}^{n}(L_i - \bar{L})^2} = 0.367$$

$$\bar{L} = \frac{1}{n}\sum_{i=1}^{n} L_i = 1.72 \qquad \hat{\beta}_0 = \ln(\bar{t}) - \hat{\beta}_1 \bar{L} = 4.44$$

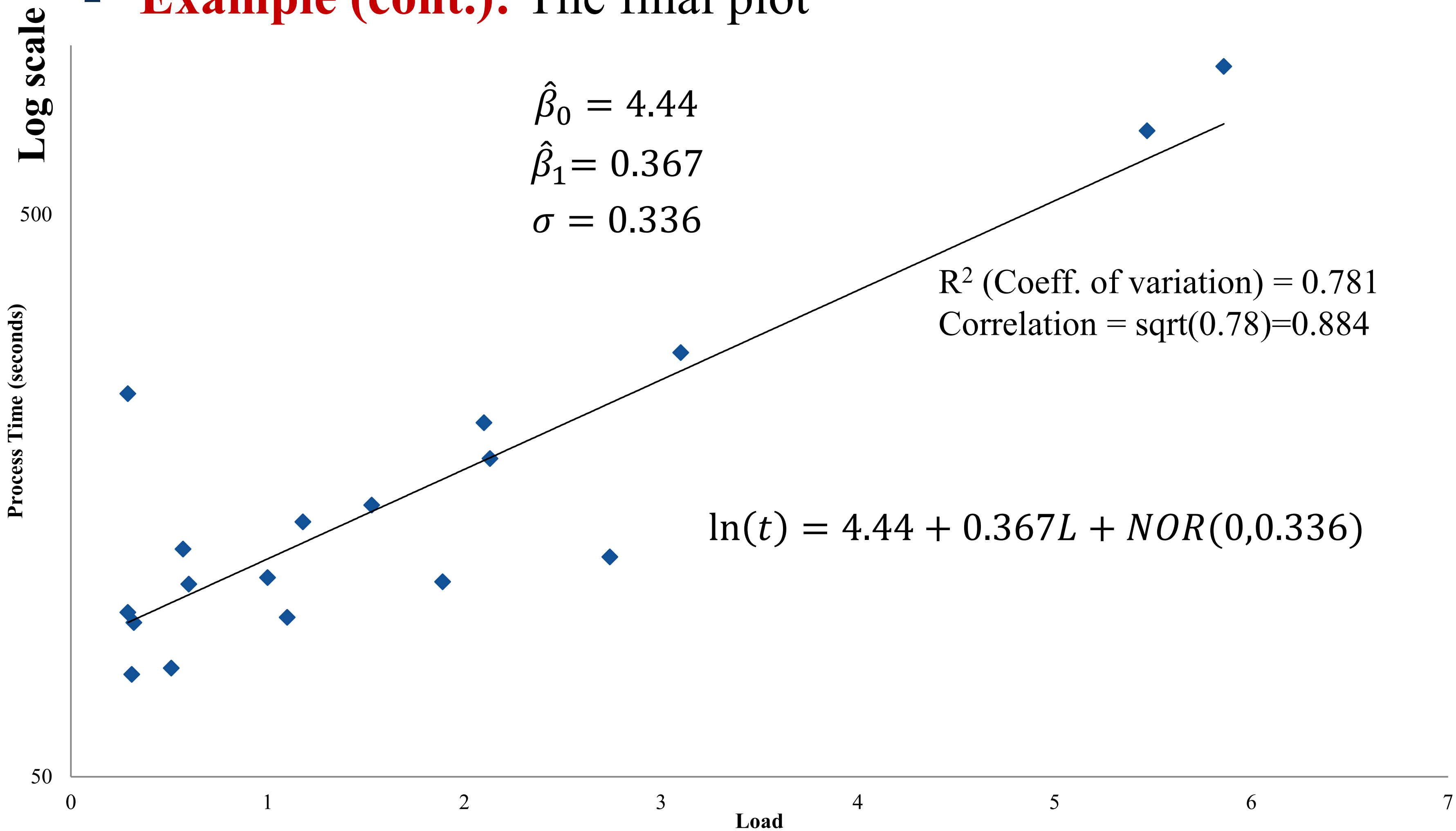
$$\ln(\bar{t}) = \frac{1}{n}\sum_{i-1}^{n} \ln(t_i) = 5.07$$

$$\ln(\hat{t}_i) = \hat{\beta}_0 + \hat{\beta}_1 L_1 = 4.44 + 0.367 L_i$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(\ln(t_i) - \ln(\hat{t}_i))^2}{n-2} = 0.113, \sigma = 0.336$$

Regression analysis (cont.)

- **Example (cont.):** The final plot



Reliability Analysis

Module 5A: Reliability Data Analysis & Model Selection

Prof. Katrina M. Groth

Mechanical Engineering Department

Center for Risk and Reliability

University of Maryland

kgroth@umd.edu

Objectives for Module 5

- To date, we've talked about how to define reliability and several related quantities, how to use parametric probability distributions, and how to calculate descriptive statistics.
- **Now we discuss:**
  - Types of reliability data and data sources
  - Non-parametric reliability modeling procedures
  - Model selection: how to use observed data to select an appropriate probability model for a set of data
  - Parameter estimation (w/uncertainty) from complete and censored data.

Key Assumptions in Module 5

- We have empirical data & we have verified the data quality.
- Data come from: **iid (Identical & independently distributed) exchangeable observations**
  - Practical interpretation in reliability: the elements of sample are obtained independently and under the same conditions.
- We're uncertain about
  - If a component might fail, when a specific component will fail, etc.
- …so we are trying to select distributions & estimate (unknown) parameters to predict these things.
- We are ignoring: Causes of failures.

Data → models → probabilities & descriptive stats

Some Excel, Matlab & R scripts available

- https://crr.umd.edu/computational-scripts. It can also be found under Research > Selected Publications > Computational Scripts.

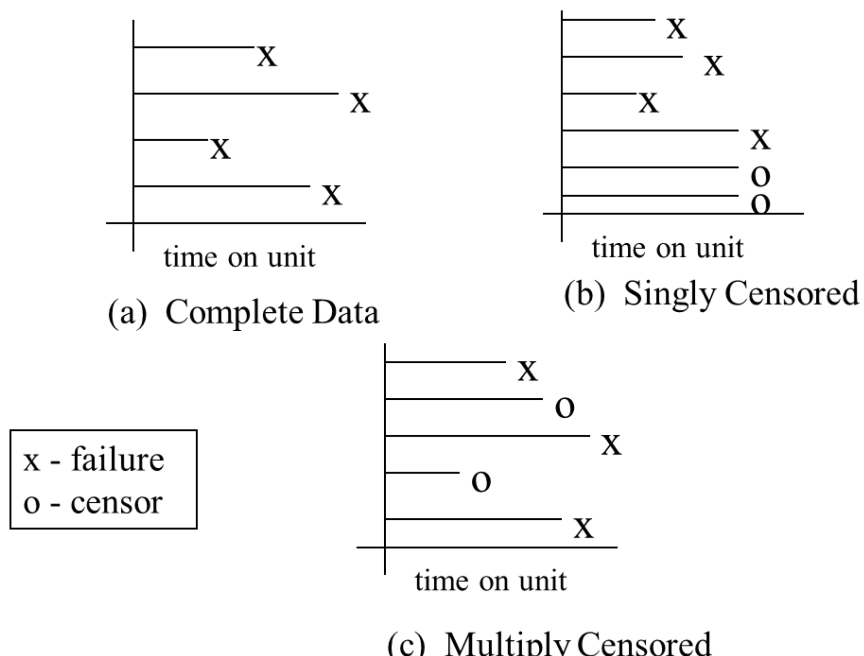
- Also see RARE2011 software
  - If you run into an error opening it in Excel – Be sure you have allowed add-ins.
    - Excel -> Options -> "Add-ins" then select "analysis toolpak VBA" and you'll see the option "go" -- press "go" it should enable.

Types of failure data

- Reliability data sources:
  - Operational or field data
    - Production and/or customer returns
    - Surveillance, maintenance, and field service
  - Generic databases
  - Testing
    - Reliability tests (prototype, production)
    - Environmental tests
    - Reliability growth tests
- Reliability data characteristics:
  - Grouped vs. ungrouped
  - Large samples vs. small samples
  - Complete vs. censored
  - With or without replacement

Complete vs. censored data

- Data are **complete** when t_i (the exact time/cycle of the specified failure mode) is available for all items $i = 1, \dots, n$
  - E.g., if a test is run until all items fail, and all failure times are recorded
- **Censored data** mean there is any item *i* with an unknown exact failure time, or failure via a different failure mode



(a) Complete Data

(b) Singly Censored

(c) Multiply Censored

Types of censored data

- Consider n items in a test
- Data are **right censored** if a failure time is not known, but is known to be greater than a given value
  - Common in reliability: only r ($r \leq n$) items fail in test duration t_{test}
  - The $n - r$ non-failed items thus have failure times $t \geq t_{test}$
- Data are **interval censored** if a failure time is not known, but is known to fall within an interval
  - This type of data may come from, e.g., periodic inspection. If an item fails in between inspections, only the interval will be known
- Data are **left censored** if a failure time is only known to be less than a given value

Type I and Type II censoring

- Data is **Type I** (time right-singly-censored) if the test is terminated at a **non-random** time t_{test}
  - Place n items on a **time-terminated test**, which will be terminated after a predetermined time t_{test} has elapsed.
  - Items that fail with $t \leq t_{test}$ will have known, specific failure times. Items that do not fail are thus right censored.
- Data is **Type II** censored if the test is terminated at a **non-random** number of observed failures
  - Place n items on a **failure-terminated test**, which will be terminated after a predetermined number of failures r is observed.
  - Only the r smallest times to failure ($t_1 \leq \ldots \leq t_r$) out of n sample times to failure are known

Generic failure data sources

- Skim **Appendix B** in the textbook for failure data related to various mechanical and electrical components.
- See comprehensive list in **Chapter 5.2**, including:
  - NSWC
  - OREDA
  - Ignition
  - RIAC
  - IEEE Std. 500-1984
- Influencing factors are ofen used to adjust generic data
  - Environment, design and manufacturing, operational factors
- After class: spend a few minutes looking at one relevant to your industry.

Generic approach for identifying a distribution from data

- Multiple methods can help you identify candidate distributions:
  1. Construct a histogram of the data (e.g., failure or repair times)
  2. Compute descriptive statistics of the data
  3. Analyze empirical statistics of failure rate (non-parametric)
  4. Use prior knowledge of failure process, or properties of the theoretical distribution
  5. Construct a probability plot (& implement linear regression/least squares)

Selecting probability distributions for your failure data

- There are ways in which we can establish a distribution directly from data…or check fit against a known distribution form:

  - Non-parametric

  (No distribution is assumed)

  - Parametric

  (A distribution is assumed)

Probability Plotting

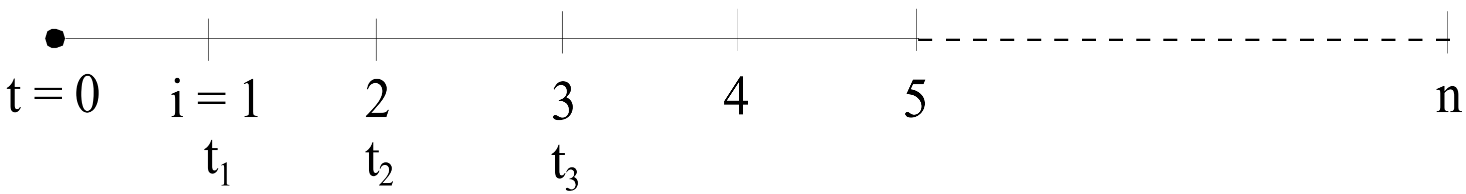
Goodness of fit tests

Nonparametric procedures for reliability functions

- We can estimate reliability parameters directly from data using **non-parametric (empirical, distribution-free)** approaches
- Nonparametric approaches attempt to directly estimate the reliability characteristics (e.g., $f(t), R(t), h(t)$) from a sample.
  - Useful for exploratory/preliminary analysis
  - Used in probability plotting
  - Key assumption: i.i.d. data
- We use various corrections/estimators depending on how much data we have & if the data are complete or censored
  - e.g., Blom, Kimball, Nelson-Aalen estimators, mean plotting position, Kaplan-Meier etc.

Non-parametric estimation

- Ordered data for *n* times to failure: $t_1 \leq t_2 \leq \ \ldots \ \leq t_n$



- Recall that

$$h(t) = \frac{\text{probability of failure in } (t + \Delta t) \text{ given surival to t}}{\text{time interval}}$$

for an interval $\Delta t = t_{i+1} - t_i$

That is:

$$h(t) = \frac{\text{\# failures between t and t+}\Delta t}{\text{\# of units surviving past t}} * \frac{1}{\Delta t}$$

Non-parametric estimation (cont.)

- $\mathrm{h(t)} = \dfrac{\text{\# failures between t and t+}\Delta\text{t}}{\text{\# of units surviving past t}} * \dfrac{1}{\Delta t}$

$$\boxed{\hat{\mathrm{h}}(\mathrm{t}_i) = \frac{1}{(\mathrm{n}-\mathrm{i})(\mathrm{t}_{\mathrm{i}+1}-\mathrm{t}_{\mathrm{i}})}} \qquad i = 1,2,\ldots,n-1$$

For small samples (~n<25), to make an unbiased estimate of h(t), corrections are necessary

$$\hat{h}(t_i) = \frac{1}{(n-i+0.625)(t_{i+1}-t_i)}, \qquad i = 1,\ldots,n-1$$

Note: 0.625 and 0.25 come from statistical analyses by Kimball to minimizing bias for the Weibull distribution with small samples.

Non-parametric estimation (cont.)

Estimators for R(t) and f(t) are equally straightforward:

$$R(t_i) = \frac{n - i}{n}$$

$$f(t) = h(t) * R(t)$$

- And the unbiased estimators of R(t) and f(t) are the **Blom (Kimball) estimators or plotting positions**

$$\hat{R}(t_i) = \frac{n - i + 0.625}{n + 0.25}, \qquad i = 1, \dots, n$$

$$\hat{f}(t_i) = \frac{1}{(n + 0.25)(t_{i+1} - t_i)}, \qquad i = 1, \dots, n - 1$$

- These are designed to de-bias small ($n \lesssim 25$) samples and estimate $h(t), R(t)$, and $f(t)$.
- These are recommended for use with small samples but can be used for larger samples.

Non-parametric reliability estimation (cont.)

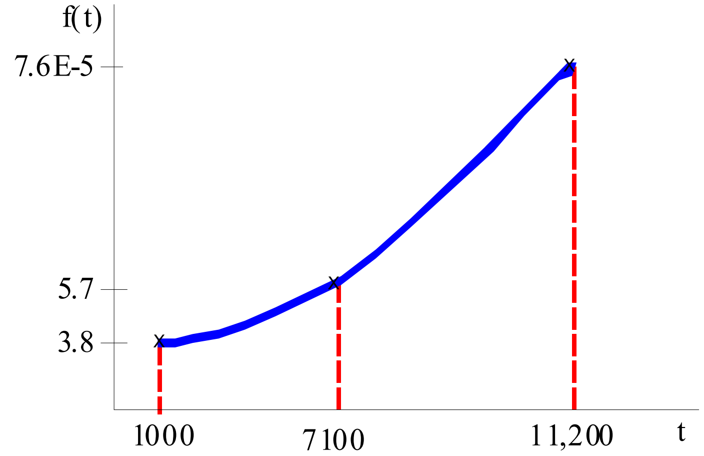
- **Example:** We observe 4 component failure times (in hrs): (1000, 7100, 11200, 14300). Use the Kimball plotting positions to obtain a non-parametric estimate of $\hat{f}(t)$, hazard rate $\hat{h}(t)$, and Reliability $\hat{R}(t)$.

i	$\mathbf{t_i}$ (hrs)	$\hat{\mathbf{f}}(\mathbf{t_i})$	$\hat{\boldsymbol{R}}(\boldsymbol{t_i})$	$\hat{\boldsymbol{h}}(\boldsymbol{t})$
1	1,000			
2	7,100			
3	11,200			
4	14,300			

Non-parametric estimation (cont.)

- **Example:** we observe 4 component failure times (hrs.): (1000, 7100, 11200, 14300)

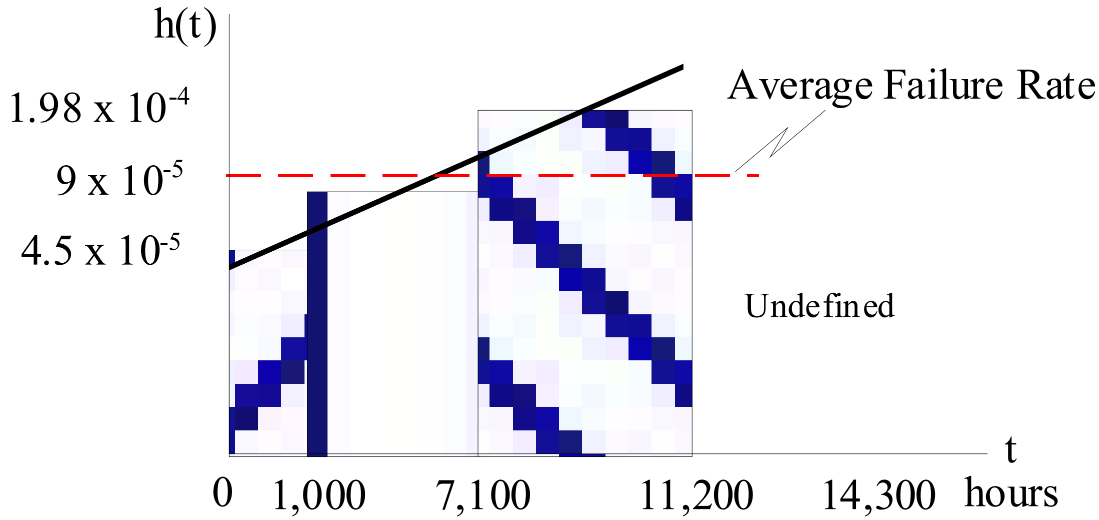
$\mathbf{i}$	$\mathbf{t_i}$ **(hrs)**	$\hat{\mathbf{f}}(\mathbf{t_i}) = \dfrac{\mathbf{1}}{(\mathbf{n}+\mathbf{0.25})(\mathbf{t_{i+1}}-\mathbf{t_i})}$
1	1,000	$\dfrac{1}{(4.25)(6100)} = 3.88 \times 10^{-5}\text{hr}^{-1}$
2	7,100	$\dfrac{1}{(4.25)(4100)} = 5.74 \times 10^{-5}\text{hr}^{-1}$
3	11,200	$\dfrac{1}{(4.25)(3100)} = 7.59 \times 10^{-5}\text{hr}^{-1}$
4	14,300	$\dfrac{1}{(4.25)(?)} = \text{undefined}$

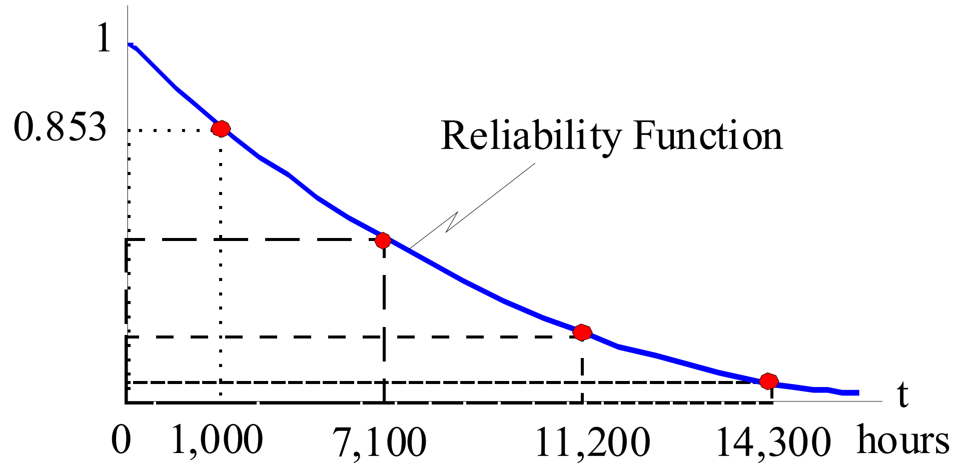


Non-parametric reliability estimation (cont.)

i	$t_i\ (hour)$	$\hat{R}(t_i) = \frac{(n-i+0.625)}{(n+0.25)}$	$\hat{h}(t) = \frac{1}{(n-i+0.625)(t_{i+1}-t_i)}$
1	1,000	$\frac{(4-1+0.625)}{4+0.25)} = 0.853$	$\frac{1}{(4-1+0.625)(7100-1000)} = 4.52 \times 10^{-5}$
2	7,100	$\frac{(4-2+0.625)}{(4+0.25)} = 0.617$	$\frac{1}{(4-2+0.625)(11200-7100)} = 9.29 \times 10^{-5}$
3	11,200	$\frac{(4-3+0.625)}{(4+0.25)} = 0.382$	$\frac{1}{(4-3+0.625)(14300-11200)} = 1.99 \times 10^{-4}$
4	14,300	$\frac{(4-4+0.625)}{(4+0.25)} = 0.147$	$\frac{1}{(4-4+0.625)(?-14300)} = undefined$

Non-parametric reliability estimation (cont.)





So the hazard rate is slightly increasing and the reliability is decreasing exponentially

Nonparametric procedure for large or grouped samples

- The **Nelson-Aalen Nonparametric Estimators** are designed for use with large or grouped samples
- Key assumption: Times to failure are grouped into equal increments Δt (required)

$$\hat{h}(t_i) = \frac{N_f(t_i)}{N_s(t_i)\Delta t} \qquad \hat{R}(t_i) = \frac{N_s(t_i)}{N} \qquad \hat{f}(t_i) = \frac{N_f(t_i)}{N\Delta t}$$

- $N_f(t_i) =$ # of failures observed in interval $(t_i, t_i + \Delta t)$
- $N_s(t_i) =$ # of surviving components in the interval starting at t_i
- t_i is usually the *lower endpoint* of interval Δt, but this can differ between practitioners, so be careful

Non-parametric reliability estimation example

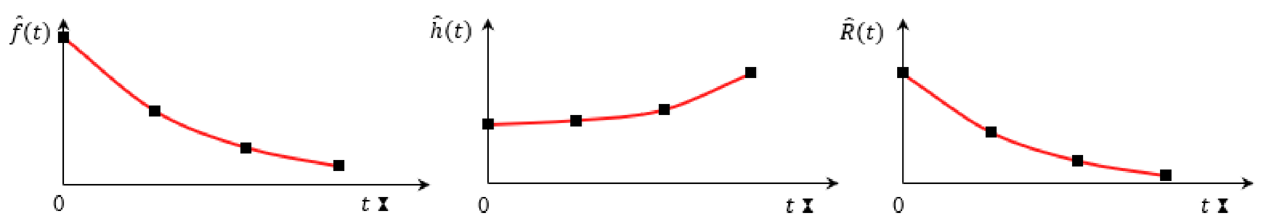
- **Example:** Given $n = 150$ observed failure times (hrs), estimate $\hat{f}(t), \hat{h}(t)$, and $\hat{R}(t)$. Use t_i as the lower endpoint of the interval in this example.

i	Interval	$N_f(t_i)$
1	$0 < t < 1000$	80
2	$1000 < t < 2000$	40
3	$2000 < t < 3000$	20
4	$3000 < t < 4000$	10

Non-parametric reliability estimation example

- **Solution:** Given $n = 150$ observed failure times, estimate $\hat{f}(t), \hat{h}(t)$, and $\hat{R}(t)$. Use t_i as the lower endpoint of the interval in this example.

i	Interval	$N_f(t_i)$	$N_S(t_i)$	$\hat{f}(t_i) = \frac{N_f(t_i)}{N\Delta t}$	$\hat{h}(t_i) = \frac{N_f(t_i)}{N_S(t_i)\Delta t}$	$\hat{R}(t_i) = \frac{N_S(t_i)}{N}$
1	$0 < t < 1000$	80	150	$\frac{80}{150 \times 1000} = 5.33 \times 10^{-4}$	$\frac{80}{150 \times 1000} = 5.33 \times 10^{-4}$	$\frac{150}{150} = 1.00$
2	$1000 < t < 2000$	40	70	$\frac{40}{150 \times 1000} = 2.67 \times 10^{-4}$	$\frac{40}{70 \times 1000} = 5.71 \times 10^{-4}$	$\frac{70}{150} = 0.47$
3	$2000 < t < 3000$	20	30	$\frac{20}{150 \times 1000} = 1.33 \times 10^{-4}$	$\frac{20}{30 \times 1000} = 6.67 \times 10^{-4}$	$\frac{30}{150} = 0.20$
4	$3000 < t < 4000$	10	10	$\frac{10}{150 \times 1000} = 6.67 \times 10^{-5}$	$\frac{10}{10 \times 1000} = 1.00 \times 10^{-3}$	$\frac{10}{150} = 0.07$



Parameter estimation

- We can also use data to estimate parameters of the underlying probability distribution

Parameter Estimation

- Probability Plotting(Linear regression w/ least squares estimates)
- Maximum Likelihood Estimation
- Bayesian Estimation

Probability plotting

- Observed data may be plotted on coordinates (previously: probability papers) such that the resulting life cdf falls on a straight line.
  - To visually assess fit
  - And then to use linear regression (least squares) to estimate parameters of the distribution.
- Probability papers for many types of distributions exist (several have been uploaded on ELMS; more commonly, we use software).

Probability plotting

- General procedure:
  - List your data in Excel or Matlab
    - *make sure that list is ordered; if it's not, sort it!
  - Find $\hat{R}(t_i)\ or\ \hat{F}(t_i)$ for each time or interval (Use the non-parametric estimators)
  - Make the known expression for R(t) or F(t) linear by taking ln of both sides (as many times as needed)
  - Plot values
  - Find trendline (least squares fit) & match to parameters

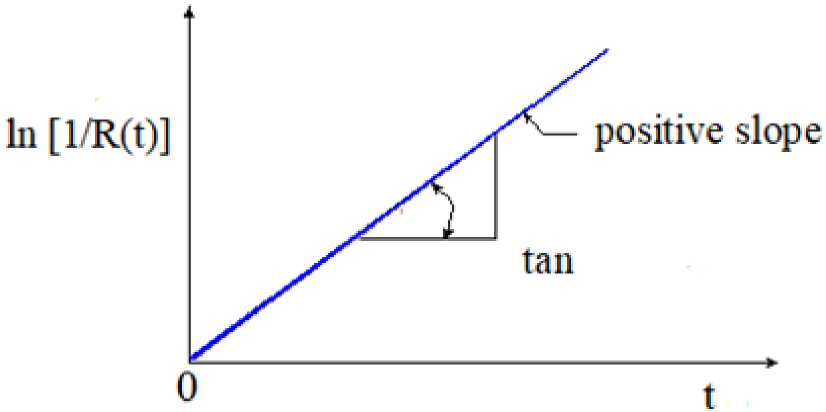
Exponential probability plotting

1. Order the n failure times $t_1 \le t_2 \le \ \dots \le t_i \le \ \dots \ \le t_n$
2. Calculate the reliability estimator (using Kimball Plotting Position or Nelson-Aalen approach depending on sample size)
3. Recall: for exp. dist: $R(t) = e^{-\lambda t}$ or $\frac{1}{R(t)} = e^{\lambda t}$

   By taking logarithms of both sides we get a linear equation: $ln\left[\frac{1}{R(t)}\right] = \lambda t$
4. Plot ln $\left(\frac{1}{R(t)}\right)$ vs. t. A straight line is a good fit and suggests that the exponential distribution is an adequate model.
5. Get the trendline:

   Set y-intercept = 0

   And the Slope = λ



Example

- The following 20 failure times (in days) were recorded for an electrical component: 51.1, 41.6, 12.9, 13.8, 22.8, 14.8, 18.5, 14.3, 27.1, 29.7, 32, 39.5, 41.3, 4.2, 3.3, 61.7, 92.2, 106.6, 148.8, 198.1 days.
- Use probability plotting to determine whether the data come from an Exponential distribution.
- Find the MTTF from this distribution.

Example – Solution

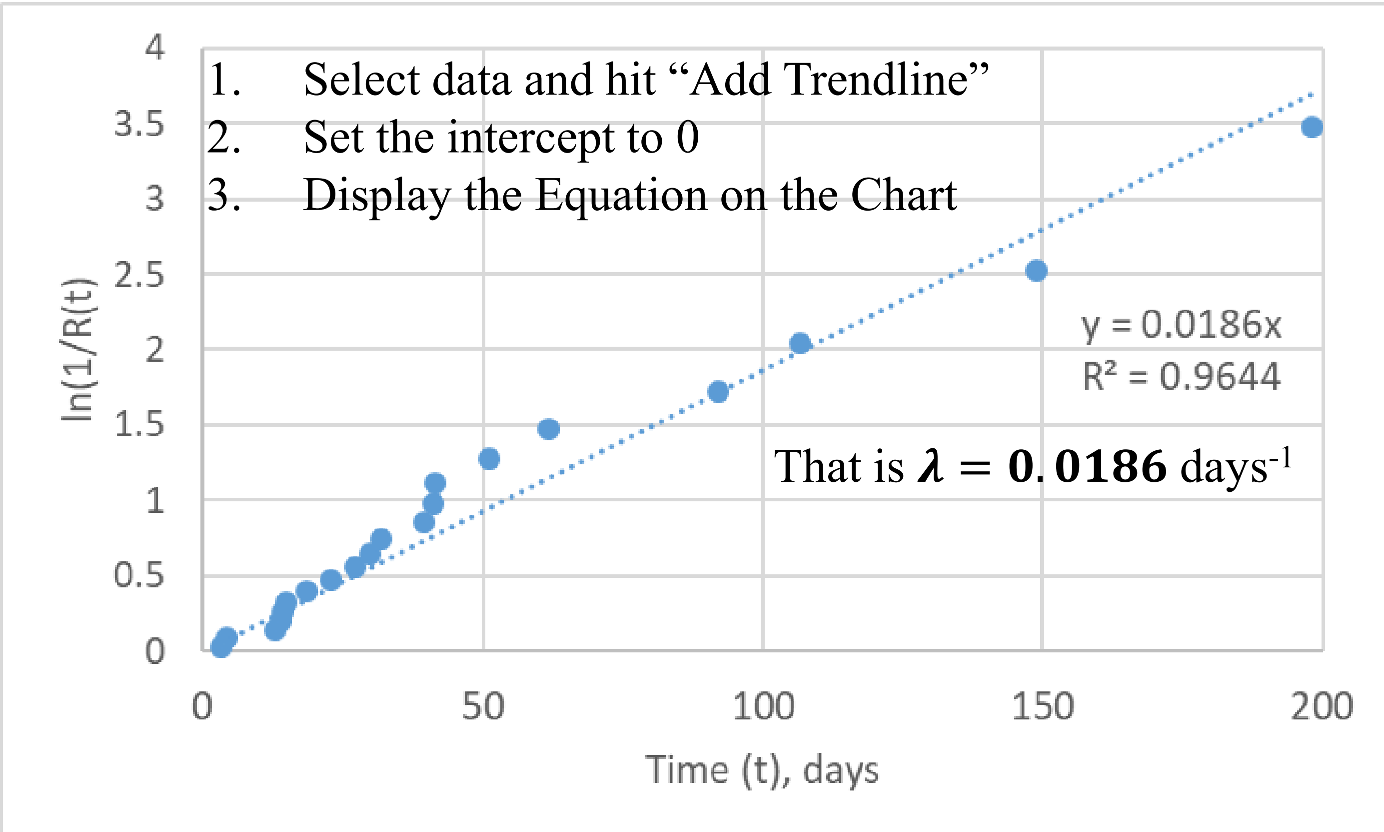
i	t_i	$R(t_i)$	$1/R(t_i)$	$\ln(1/R(t_i))$
1	3.3	0.969	1.032	0.031
2	4.2	0.920	1.087	0.084
3	12.9	0.870	1.149	0.139
4	13.8	0.821	1.218	0.197
5	14.3	0.772	1.296	0.259
6	14.8	0.722	1.385	0.325
7	18.5	0.673	1.486	0.396
8	22.8	0.623	1.604	0.472
9	27.1	0.574	1.742	0.555
10	29.7	0.525	1.906	0.645
11	32	0.475	2.104	0.744
12	39.5	0.426	2.348	0.853
13	41.3	0.377	2.656	0.977
14	41.6	0.327	3.057	1.117
15	51.1	0.278	3.600	1.281
16	61.7	0.228	4.378	1.477
17	92.2	0.179	5.586	1.720
18	106.6	0.130	7.714	2.043
19	148.8	0.080	12.462	2.523
20	198.1	0.031	32.400	3.478

With a relatively small, ungrouped sample – use the estimators to create $\hat{R}(t_i)$:

$$\hat{R}(t_i) = \frac{n - i + 0.625}{n + 0.25}$$

$$\frac{1}{\hat{R}(t_i)} = \frac{n + 0.25}{n - i + 0.625}$$

Example- Solution



MTTF= $\mathbf{53.76}$ days (compare to the mean of the data sample of 48.15)

Note: if you do this in RARE, it doesn't set intercept to 0. RARE gives $\boldsymbol{\lambda} = \mathbf{0.0179}$ days^{-1} but has a small intercept (0.101) which RARE neglects.

Exponential plotting (cont.)

- We may observe two or more straight lines. This happens when, e.g., there may be initially a particular failure mode and another failure mode become dominant.

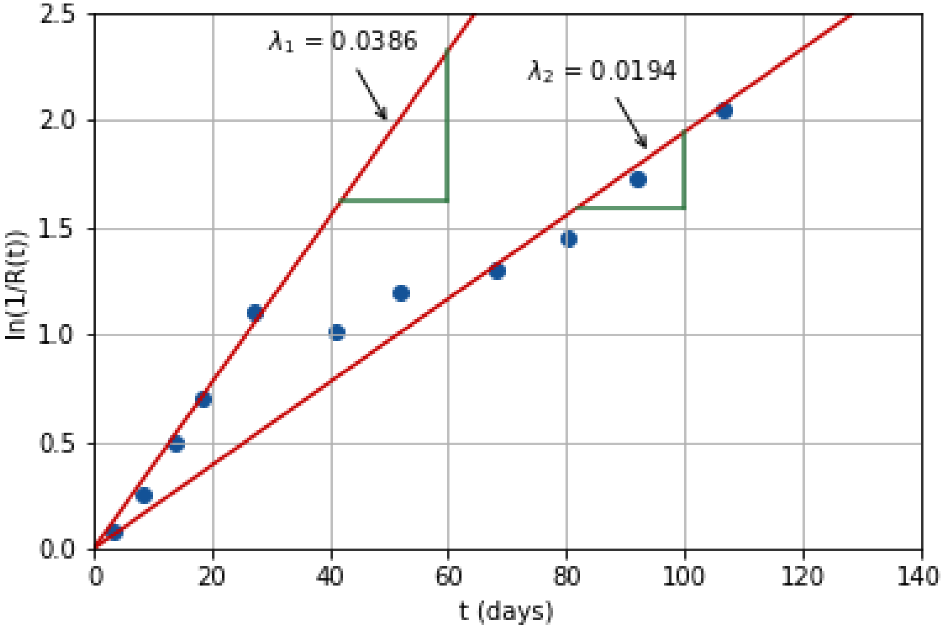
$$\lambda_E = \lambda_1 + \lambda_2$$

Therefore,

$$R = R_1 \cdot R_2$$

$$= e^{-\lambda_1 t} e^{-\lambda_2 t}$$

$$= e^{-\lambda_E t}$$



Weibull probability plotting (cont.)

- The goal is to calculate the shape parameter β and the scale parameter α.
- Weibull $\mathrm{R(t)} = \mathrm{e}^{-\left(\frac{\mathrm{t}}{\alpha}\right)^{\beta}}$

$$\ln\left(\ln\left[\frac{1}{\mathrm{R(t)}}\right]\right) = \beta \cdot \ln(\mathrm{t}) - \beta \cdot \ln(\alpha)$$

****note double ln here**

- Plot $\mathrm{Ln}\left(\mathrm{Ln}\left[\frac{1}{R(t)}\right]\right)$ vs. $\ln(t)$
  - If data falls on a straight line, Weibull is a good fit
  - Add linear trendline.
  - Slope = β
  - Y-int = $-\beta \times \ln(\alpha)$
  - Solve for β, α

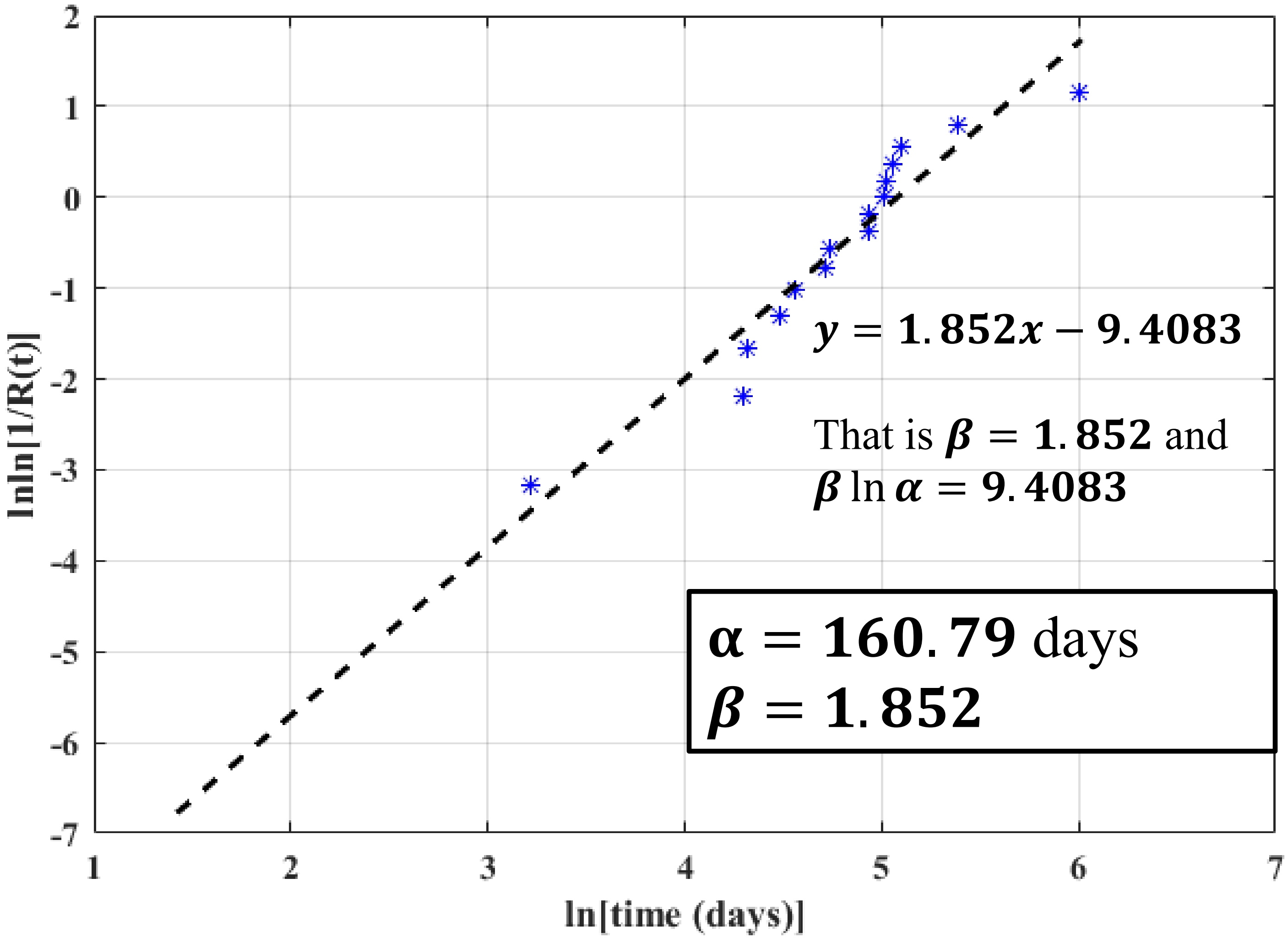
Example

- The following failure times were obtained from testing 15 units until each had failed: 25.1, 73.9, 75.5, 88.5, 95.5, 112.2, 113.6, 138.5, 139.8, 150.3, 151.9, 156.8, 164.5, 218, 403.1 days. Determine whether the data represent the Weibull distribution. If the data are a reasonable fit, find the shape and scale parameters.

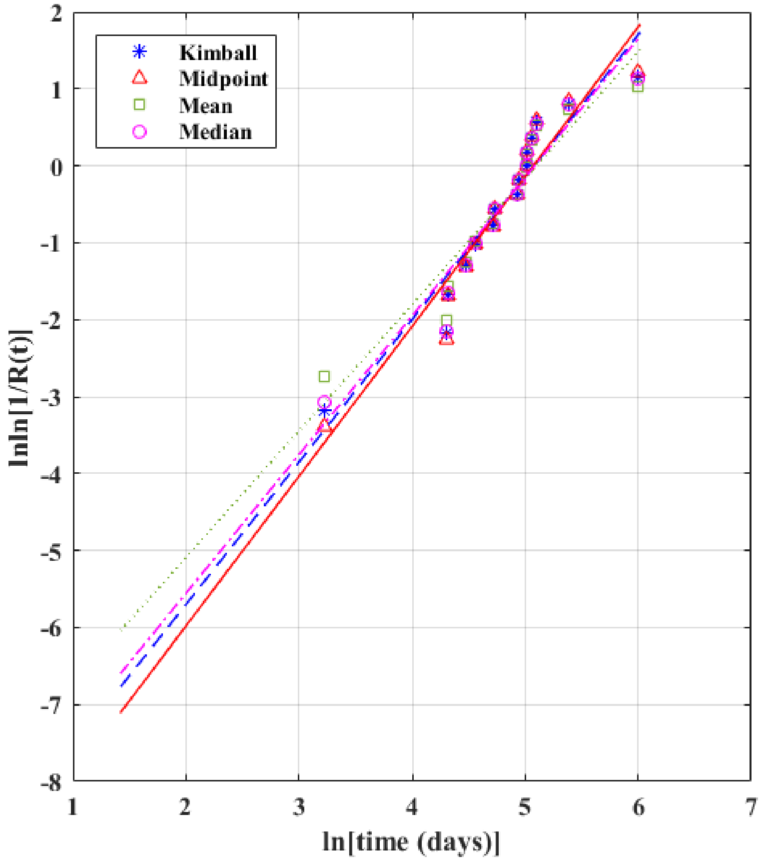
Example – Solution

i	ti	Ln[ti]	R(ti)	1/R(ti)	LnLn[1/R(ti)]
1	25.1	3.2229	0.9590	1.0427	-3.1737
2	73.9	4.3027	0.8934	1.1193	-2.1833
3	75.5	4.3241	0.8279	1.2079	-1.6665
4	88.5	4.4830	0.7623	1.3118	-1.3041
5	95.5	4.5591	0.6967	1.4353	-1.0179
6	112.2	4.7203	0.6311	1.5844	-0.7761
7	113.6	4.7327	0.5656	1.7681	-0.5623
8	138.5	4.9309	0.5000	2.0000	-0.3665
9	139.8	4.9402	0.4344	2.3019	-0.1818
10	150.3	5.0126	0.3689	2.7111	-0.0026
11	151.9	5.0232	0.3033	3.2973	0.1766
12	156.8	5.0550	0.2377	4.2069	0.3624
13	164.5	5.1029	0.1721	5.8095	0.5650
14	218	5.3845	0.1066	9.3846	0.8061
15	403.1	5.9992	0.0410	24.4000	1.1615

Example – Solution



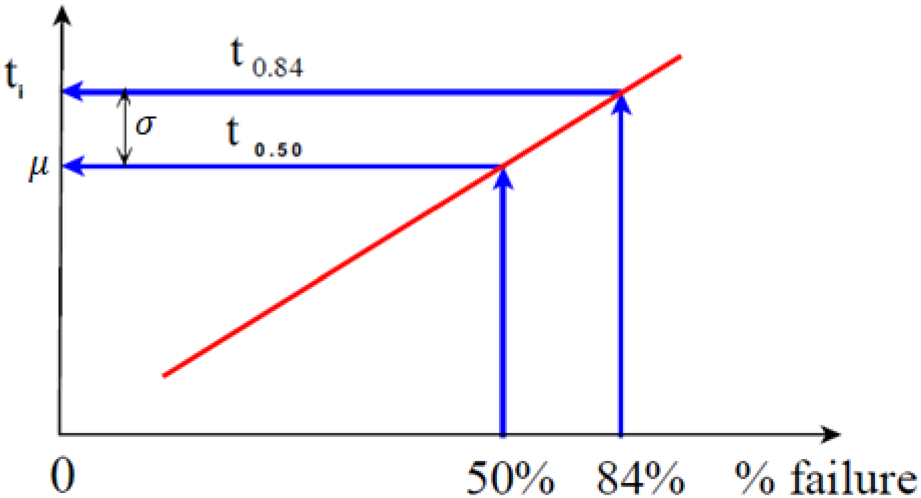
Alternative plotting positions exist – see textbook for details

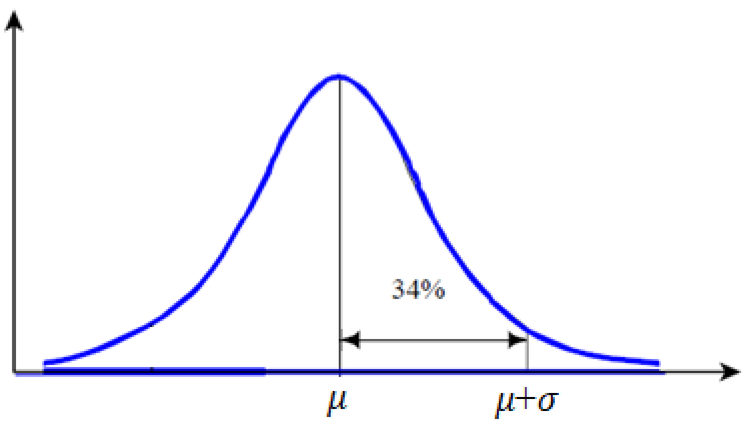


Midpoint:	$R(t) = \dfrac{n - i + 0.5}{n}$
Mean:	$R(t) = \dfrac{n - i + 1}{n + 1}$
Median:	$R(t) = \dfrac{n - i + 0.7}{n + 0.4}$

Normal distribution probability plotting

- We plot $\Phi^{-1}(F(t))$ against t for the normal distribution
  - Can also plot $\Phi^{-1}(F(t))$ against $\ln(t)$ for the lognormal distribution
- $F(t)$ is constructed using appropriate non-parametric method
  - E.g., $F(t) = \frac{i-0.375}{n+0.25}$ if using the Kimball estimators.
- $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution





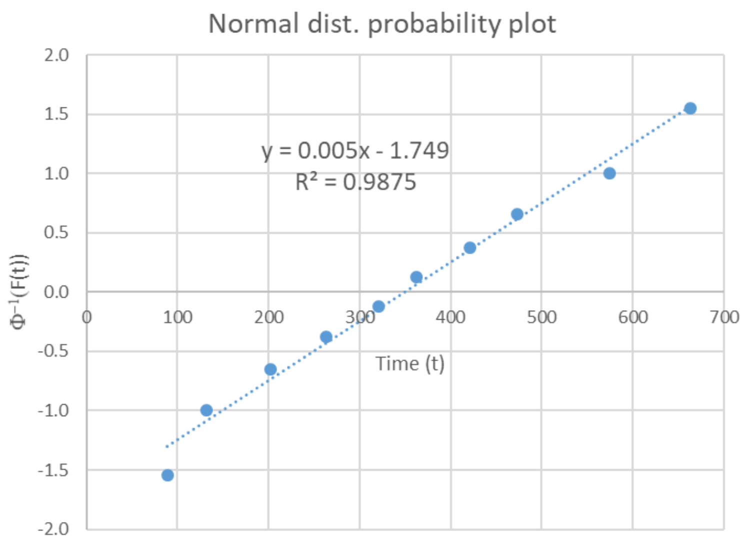
Normal distribution

Normal plot of the example

- **Example:** Normal dist. $F = \Phi(\frac{t-\mu}{\sigma})$, Linearizes as: $\Phi^{-1}(F) = \frac{t}{\sigma} - \frac{\mu}{\sigma}$

=Norm.S.inv(F)

i	1	2	3	4	5	6	7	8	9	10
t_i (days)	89	132	202	263	321	362	421	473	575	663
$F(t)=\frac{i-0.375}{n+0.25}$	0.061	0.159	0.256	0.354	0.451	0.549	0.646	0.744	0.841	0.939
$\Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right)$	-1.55	-1.00	-0.66	-0.38	-0.12	0.12	0.38	0.66	1.00	1.55

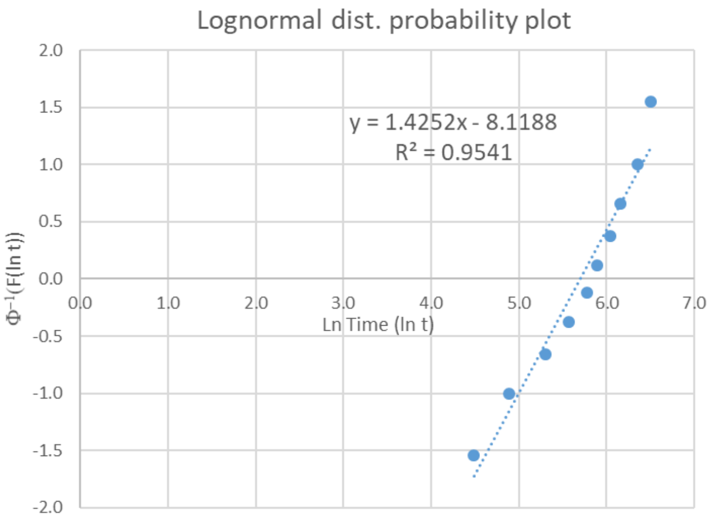


$t_{0.50} = \mu$	$\mu = 349.8$
$t_{0.84} = \mu + \sigma$	$\sigma = 200$

Lognormal plot of the example

- **Example: Lognormal distribution** $F = \Phi(\frac{\ln t - \mu}{\sigma})$, linearize as $\Phi^{-1}(F) = \frac{\ln t}{\sigma} - \frac{\mu}{\sigma}$

i	1	2	3	4	5	6	7	8	9	10
$F(t) = \frac{i-0.375}{n+0.25}$	0.061	0.159	0.256	0.354	0.451	0.549	0.646	0.744	0.841	0.939
$\Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right)$	-1.55	-1.00	-0.66	-0.38	-0.12	0.12	0.37	0.66	1.00	1.55
t_i (days)	89	132	202	263	321	362	421	473	575	663
lnt_i	4.5	4.9	5.3	5.6	5.8	5.9	6.0	6.2	6.4	6.5



Lognormal parameters	
$\ln(t_{0.5})$	5.70
$\ln(t_{0.84})$	6.39
μ	5.70
σ	0.70

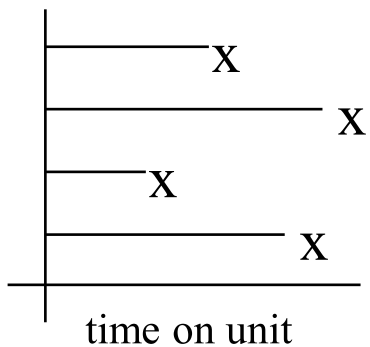
Conclusion: In this case, the Normal distribution fits better than the Lognormal distribution (i.e., higher R^2 value for the normal plot.)

Parameter estimation

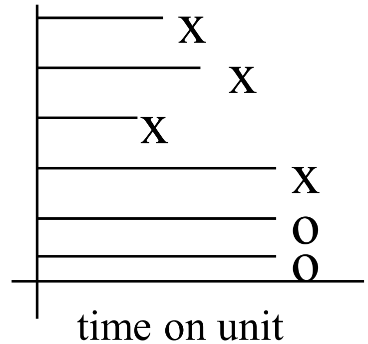
- Recall: In Module 4, you learned how to dervice the MLE estimators for many distributions.
- Now we'll present **point estimates** & **interval estimates** for the parameters of common reliability distributions

Reminder! Reliability data are often censored

- What happens now?

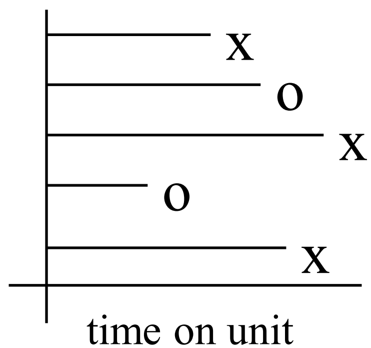


(a) Complete Data



(b) Singly Censored

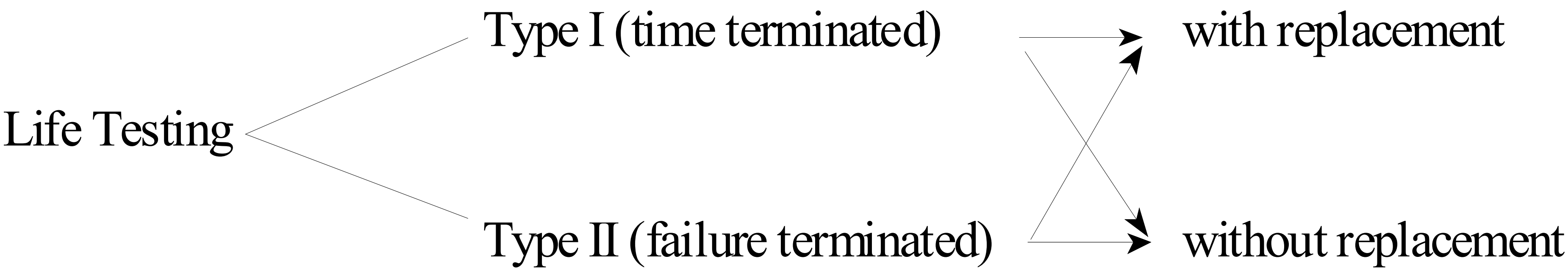
x - failure
o - censor



(c) Multiply Censored

Types of Life Testing Data

- Life testing is done to get failure data for reliability estimation methods:



- **The result is that reliability data are almost always censored.**

Recall: MLE

- **Likelihood Function for known failure times:**

$$L(\theta|x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

$$\Lambda(\theta|E) = \ln\{L(\theta|E)\} = \sum_{i=1}^{n} \ln[f(x_i\,|\theta)]$$

- **Maximum Likelihood (ML)** estimate of θ is the value of $\hat{\theta}$ such that

$$L(\hat{\theta}|x_1, x_2, \dots, x_n) \geq L(\theta|x_1, x_2, \dots, x_n)$$

for every value of θ. Statistic $\hat{\theta}$ is a r.v. called the ML estimator (MLE) of θ.

Solve for θ

$$\left.\frac{\partial L}{\partial \theta}\right|_{\theta-\hat{\theta}} = \left.\frac{\partial lnL}{\partial \theta}\right|_{\theta-\hat{\theta}} = 0$$

Higher L →better fit.

Likelihood functions for different types of reliability data

Type of Observation	Likelihood Function	Example Description
Exact lifetimes	$L_i(\theta \| t_i) = f(t_i \| \theta)$	Failure time is known.
Left censored	$L_i(\theta \| t_i) = F(t_i \| \theta)$	Component failed before time t_i.
Right censored	$L_i(\theta \| t_i) = 1 - F(t_i \| \theta) = R(t_i \| \theta)$	Component survived to time t_i.
Interval censored	$L_i(\theta \| t_i) = F\left(t_i^{RI} \| \theta\right) - F\left(t_i^{LI} \| \theta\right)$	Component failed between t_i^{LI} and t_i^{RI}.
Left truncated	$L_i(\theta \| t_i) = \frac{f(t_i \| \theta)}{R(t_L \| \theta)}$	Component failed at time t_i where observations are truncated before t_L.
Right truncated	$L_i(\theta \| t_i) = \frac{f(t_i \| \theta)}{F(t_U \| \theta)}$	Component failed at time t_i where observations are truncated after t_U.
Interval truncated	$L_i(\theta \| t_i) = \frac{f(t_i \| \theta)}{F(t_U \| \theta) - F(t_L \| \theta)}$	Component failed at time t_i where observations are truncated before t_L and after t_U.

Example: Likelihood function creation

- Assume we have a sample of size D
  - Some are known, exact failure times ($\delta_i = 1$)
  - Some are right-censored times ($\delta_i = 0$)
- The likelihood function is constructed:

$$L(\theta|D) = c\prod_{i}\left\{[f(t_i|\theta)]^{\delta_i} \times [1 - F(t_i|\theta)]^{1-\delta_i}\right\}$$

- Where:
  - c = combinatorial constant
  - $f(t_i|\theta)$ = likelihood function for exact data points
  - $1 - F(t_i|\theta)$ = likelihood function for right-censored data points

MLE parameters of various distributions

- For *complete data:* point estimates of the MLE estimates for the parameters of many relevant distributions are known.
  - Now we'll discuss confidence intervals on those parameters, too.
- *For censored data:* you must maximize the likelihood function for this data; then use the Fisher information matrix (or established, derived functional relationships) to come up with values needed to estimate confidence intervals.

MLE parameters of exponential dist for complete data

- **Exponential Distribution: n failures at times t_i**

$$\mathrm{L} = \prod_{i=1}^{n} \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} t_i}$$

Solve for n failures

$$\Lambda = \ln(L) = n \ln \lambda - \lambda \sum_{i=1}^{n} t_i$$

$$\left.\frac{\partial \Lambda}{\partial \lambda}\right|_{\lambda=\hat{\lambda}} = \frac{n}{\hat{\lambda}} - \sum_{i=1}^{n} t_i = 0 \rightarrow$$

$$\boldsymbol{\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i}}$$

Exponential distribution MLEs for censored data

- **Type I with replacement:**
  - n components are placed under test.
  - t_{end} time at which the test is terminated.
  - TTT accumulated component test hours (**total time on test**)
  - r failures have been observed (up to t_o)

$$TTT = nt_{end}$$

$$\hat{\lambda} = \frac{r}{TTT}$$

$$\widehat{MTTF} = \frac{TTT}{r}$$

And the number of units actually used in the test (n') is: $n' = n + r$

Exponential distribution MLEs for censored data

- **Type I without replacement:**

$$TTT = \sum_{i=1}^{r} t_i + (n-r)t_{end}$$

Accumulated time on test of r failed components.

Accumulated time on test of the non-failing components.

$$\hat{\lambda} = \frac{r}{TTT}$$

Exponential distribution MLEs for censored data

- **Type II with replacement:**
  - n components placed on test.
  - t_r the time after which test is terminated when the rth failure has occurred. So rth failure time is specified by t_r and is a random variable.

$$TTT = nt_r$$

$$\hat{\lambda} = \frac{r}{TTT}$$

  - Total units put on test (n') is:

$$n' = n + r - 1$$

Exponential distribution MLEs for censored data

- **Type II without replacement:**

$$TTT = \sum_{i=1}^{r} t_i + (n - r)t_r$$

and

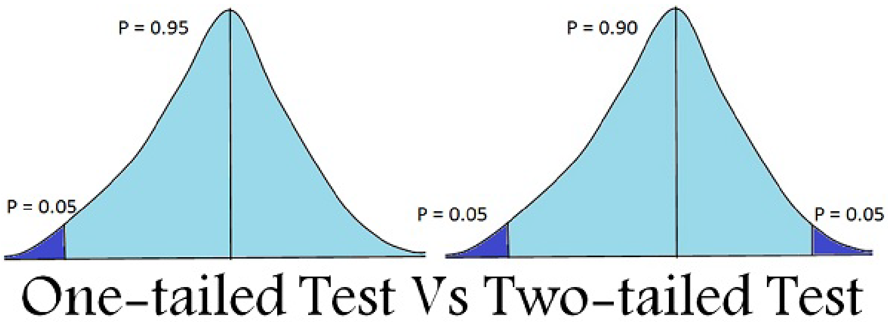
$$n' = n$$

Confidence intervals express uncertainty due to sample size

- **Example**: If 100 units are tested, consider two situations for exponential parameter estimation:
  - **Case 1: For r = 1** $\boldsymbol{failure, t_0 = 10}$. hrs
    - $\mathrm{T} = 10 \times 100 = 1000$
    - $\hat{\lambda} = \frac{r}{T} = \frac{1}{1000} = 10^{-3}\mathrm{hr}^{-1}$
  - **Case 2: For r = 10 failures,** $\boldsymbol{t_0 = 100\ hrs}$
  - $\mathrm{T} = 100 \times 100 = 10{,}000$
  - $\hat{\lambda} = \frac{\mathrm{r}}{\mathrm{T}} = \frac{10}{10{,}000} = 10^{-3}\mathrm{hr}^{-1}$
  - Both gives you the same $\boldsymbol{\hat{\lambda}}$ estimate, but one has more data. The MLE parameter is the same, but the confidence interval is different for these two datasets.

Reminders: confidence interval

- The $1-\alpha$ confidence interval for a parameter θ is the interval such that:
  - $\Pr\left(\hat{\theta}_{lower} \leq \theta \leq \hat{\theta}_{upper}\right) = 1-\alpha$
  - e.g, for a 90% confidence interval, $\alpha = 0.1$. $\Pr\{R_L \leq R(t_0) \leq R_U\} = 1 -$ α, and thus in 100(1 - α)% of repetitions of that test, the population parameter falls between R_L and R_U.
  - Used to quantify uncertainty due to sampling error (i.e., limited number of samples),
    - Not uncertainty due to incorrect model selection or assumptions!



Exponential dist.: Confidence intervals

	Type I (Time Terminated Test for complete data)			
	One-Sided Confidence Limits		**Two-Sided Confidence Limits**	
Parameter	**Lower Limit**	**Upper Limit**	**Lower Limit**	**Upper Limit**
λ	0	$\frac{\chi^2_{(1-\gamma)}[2r+2]}{2TTT}$	$\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}[2r]}{2TTT}$	$\frac{\chi^2_{\left(1-\frac{\gamma}{2}\right)}[2r+2]}{2TTT}$
MTTF	$\frac{2TTT}{\chi^2_{(1-\gamma)}[2r+2]}$	∞	$\frac{2TTT}{\chi^2_{\left(1-\frac{\gamma}{2}\right)}[2r+2]}$	$\frac{2TTT}{\chi^2_{\left(\frac{\gamma}{2}\right)}[2r]}$
$R(t)$	$e^{-\left[\frac{\chi^2_{(1-\gamma)}[2r+2]}{2Ttt} t_E\right.}$	1	$e^{-\left[\frac{\chi^2_{\left(1-\frac{\gamma}{2}\right)}[2r+2]}{2TTT} t_e\right.}$	$e^{-\left[\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}[2r]}{2TTT} t_{end}\right]}$
	Type II (Failure Terminated Test for complete data)			
	One-Sided Confidence Limits		**Two-Sided Confidence Limits**	
Parameter	**Lower Limit**	**Upper Limit**	**Lower Limit**	**Upper Limit**
λ	0	$\frac{\chi^2_{(1-\gamma)}[2r]}{2TTT}$	$\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}[2r]}{2TTT}$	$\frac{\chi^2_{\left(1-\frac{\gamma}{2}\right)}[2r]}{2TTT}$
MTTF	$\frac{2TTT}{\chi^2_{(1-\gamma)}[2r]}$	∞	$\frac{2TTT}{\chi^2_{\left(1-\frac{\gamma}{2}\right)}[2r]}$	$\frac{2TTT}{\chi^2_{\left(\frac{\gamma}{2}\right)}[2r]}$
$R(t)$	$e^{-\left[\frac{\chi^2_{(1-\gamma)}[2r]}{2TTT} t_{end}\right.}$	1	$e^{-\left[\frac{\chi^2_{\left(1-\frac{\alpha}{2}\right)}[2r]}{2TTT} t_{end}\right.}$	$e^{-\left[\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}[2r]}{2TTT} t_{end}\right]}$

- Where $\chi^2_\gamma[x]$ is a chi-square distribution value, which has two parameters
  - Degree of freedom (x)
  - Some confidence level (γ)

- Note: uncensored data can be treated as a special case of a Type II (failure terminated) test.

Exponential confidence intervals

- These are type 1 data, so the confidence interval is calculated as:

$$\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}(2r)}{2TTT} \leq \hat{\lambda} \leq \frac{\chi^2_{\left(1-\frac{\gamma}{2}\right)}(2r+2)}{2TTT},\qquad \text{For a 90\% confidence interval, } 1-\alpha = 0.9,\quad \alpha = 0.1$$

- **Case 1: For r = 1, TTT = 1000, the 90% confidence interval is:**
  - $\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}(2)}{2(1000)} \leq \hat{\lambda} \leq \frac{\chi^2_{\left(1-\frac{\gamma}{2}\right)}(4)}{2(1000)} \Rightarrow \frac{0.1026}{2000} \leq \hat{\lambda} \leq \frac{9.49}{2000}$
  - $\mathbf{5.13 \times 10^{-5} \leq \hat{\lambda} \leq 4.75 \times 10^{-3}}$

- **Case 2: For r = 10, TTT = 10000, the 90% confidence interval is:**
- $\frac{\chi^2_{\left(\frac{\gamma}{2}\right)}(20)}{2(10{,}000)} \leq \hat{\lambda} \leq \frac{\chi^2_{\left(1-\frac{\gamma}{2}\right)}(22)}{2(10{,}000)} \Rightarrow \frac{10.85}{20{,}000} \leq \hat{\lambda} \leq \frac{33.92}{20{,}000}$
  - $\mathbf{5.43 \times 10^{-4} \leq \hat{\lambda} \leq 1.70 \times 10^{-3}}$

Example: Exponential confidence intervals

- 25 units are placed on test for 500 hours. Eight failures occur at times 75, 115, 192, 258, 312, 389, 410, 496 hours. Failed units are replaced.

- **Find:**
  A. The MLE of λ
  B. The two-sided 90% confidence limits on λ.

Example: Exponential confidence intervals

- **Solution:**
  A. This is time-terminated test, i.e., (**Type I**), with n=25 and $t_{end} = 500\ hrs,$ therefore,
- TTT = 25 * 500 hrs = 12,500 component-hrs.

$$\hat{\lambda} = \frac{number\ of\ failures}{total\ component\ time} = \frac{8}{12{,}500} = 6.4\ \times 10^{-4} hr^{-1}$$

Example: Exponential two-sided confidence interval

- **Solution B:**
- For a type I test, exponential distribution the two-sided confidence interval expression is:
  - $$\frac{\chi^2_{\gamma/2}(2\text{r})}{2\text{TTT}} \leq \hat{\lambda} \leq \frac{\chi^2_{1-\gamma/2}(2\text{r}+2)}{2\text{TTT}}$$

Here, $\gamma = 0.1, \frac{\gamma}{2} = 0.05, TTT = 12{,}500, r = 8$

$$\frac{\chi^2_{0.05}(16)}{2 * 12500} \leq \hat{\lambda} \leq \frac{\chi^2_{0.95}(18)}{2 * 12500}$$

$$\frac{7.96}{25{,}000} \leq \hat{\lambda} \leq \frac{28.87}{25{,}000}$$

$$\boxed{3.18 \times 10^{-4} \leq \hat{\lambda} \leq 1.15 \times 10^{-3} hr^{-1}}$$

Exponential distribution: Right censored data summary

Type I data is time terminated and type II data is failure terminated. r is number of failures and n is the number of units being observed , t_i is the time to failure of a failed unit. t_r is the time after which test is terminated (for type I data this is the specified test time; for type II data, this is the time, when the r[th] failure occurs)

Case	MLE	Total time, TTT	Confidence interval
Type II w/ replacement	$\hat{\lambda} = r/TTT$	$TTT = nt_r$ Where	$\frac{\chi^2_{\frac{\gamma}{2}}(2r)}{2TTT} \le \lambda \le \frac{\chi^2_{1-\frac{\gamma}{2}}(2r)}{2TTT}$ (Exact)
Type II w/o replacement	$\hat{\lambda} = r/TTT$	$TTT = \sum_{i=1}^{r} t_i + (n-r)t_r$	
Type I, w/ replacement	$\hat{\lambda} = r/TTT$	$TTT = n\, t_r$	$\frac{\chi^2_{\frac{\gamma}{2}}(2r)}{2TTT} \le \lambda \le \frac{\chi^2_{1-\frac{\gamma}{2}}(2r+2)}{2TTT}$ (Approximate)
Type I, w/o replacement	$\hat{\lambda} = r/TTT$	$TTT = \sum_{i=1}^{r} t_i + (n-r)t_r$	

Note: uncensored data can be treated as a special case of a Type II (failure terminated) test.

Example: Exponential two-sided confidence interval

- **Example:**
  - A plant had 50 instrument failures in a year among a total of 5613 such instruments.

A. Find 95% confidence limits on λ

B. Find 95% confidence limits on R (8760 hrs.),

C. Find the point estimate and 95% percentile estimate of the time at which R = 0.8 for each instrument.

Example: Exponential two-sided confidence interval

- **Solution A) Type I test**
  - $TTT = 5613\ units\ \times 8760\ hours = 4.9 \times 10^7$ component-hours, therefore,

$$\hat{\lambda} = \frac{50\ failures}{4.9 \times 10^7} = 1.0 \times 10^{-6} hr^{-1}$$

$\gamma = 1 - 0.95 = 0.05$, therefore,

$$\frac{\chi^2_{\frac{\gamma}{2}}(2 \times 50)}{2\mathrm{TTT}} \leq \lambda \leq \frac{\chi^2_{1-\frac{\gamma}{2}}(2 \times 50 + 2)}{2\mathrm{TTT}}$$

$$\frac{\chi^2_{0.025}(100)}{2\mathrm{TTT}} \leq \lambda \leq \frac{\chi^2_{0.975}(102)}{2\mathrm{TTT}}$$

$$\frac{74.55}{2TTT} \leq \hat{\lambda} \leq \frac{131.54}{2TTT}$$

With T = 4.9E7 hrs.

$$\boldsymbol{7.6E - 7 \leq \lambda \leq 1.3E - 6hr^{-1}}$$

Example: Exponential two-sided confidence interval

Solution B)

$$\hat{R}(8760) = e^{-\hat{\lambda}t} = e^{-1.0E-6(8760)} = 0.991$$

$$\exp(-\frac{\chi^2_{0.975}(102)}{2TTT} * 8760) \leq R \leq \exp(-\frac{\chi^2_{0.025}(100)}{2TTT} 8760)$$

$$\exp\left(-\frac{131.54 * 8760}{2(365)(24)(5613)}\right) \leq \hat{\mathrm{R}} \leq \exp\left(-\frac{74.53 * 8760}{2(365)(24)(5613)}\right)$$

$$\mathbf{0.9884 \leq \hat{R} \leq 0.9934}$$

Example: Exponential two-sided confidence interval

Solution C)

- To find reliable life at R = 0.8

$$R(t) = e^{-\lambda t} \rightarrow lnR = -\lambda t \rightarrow t = \frac{lnR}{-\lambda}$$

$$\hat{t}_{0.8} = \frac{-\ln(0.8)}{\hat{\lambda}} = 233{,}144\ hours\ \cong 25.5\ years$$

$$\frac{-\ln(0.8)}{1.3E-6} \leq t_{0.8} \leq \frac{-\ln(0.8)}{7.5E-7}$$

$$171{,}649 \leq t_{0.8} \leq 297{,}525\ hours$$
$$19.59 \leq t_{0.8} \leq 33.96\ years$$

MLE Parameters of various distributions

- For some distributions and data types (e.g., exponential with complete or right censored data), we have specific known relationship forms for certain data types for certain distributions.
  - E.g., exponential with complete or right censored data
  - E.g., others we will cover shortly.
- Other data types and parameters require deriving confidence intervals from the Fisher Information Matrix;

MLE: parameter uncertainty estimation

- Where parameter values have been estimated using the MLE process, the uncertainty of each parameter is quantified using the observed **Fisher Information Matrix ($\boldsymbol{J}(\boldsymbol{\theta})$)** as follows:
- For example, in the case of a distribution with *n* parameters, I is given by:
- $I(\widehat{\boldsymbol{\theta}}) = \begin{bmatrix} -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\, D\,)}{\partial \theta_1^2} & -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|D)}{\partial \theta_1 \partial \theta_2} & \cdots & -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\, D\,)}{\partial \theta_1 \partial \theta_p} \\ -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\, D\,)}{\partial \theta_2 \partial \theta_1} & -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\, D\,)}{\partial \theta_2^2} & \cdots & -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\, D\,)}{\partial \theta_2 \partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\, D\,)}{\partial \theta_p \partial \theta_1} & -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\, D\,)}{\partial \theta_p \partial \theta_2} & \cdots & -\frac{\partial^2 \Lambda(\boldsymbol{\theta}|\, D\,)}{\partial \theta_p^2} \end{bmatrix}_{\theta_i = \widehat{\theta_i}}$
- $where\ \Lambda = \ln(\boldsymbol{\theta}|x_i)$

MLE: Parameter uncertainty estimation

- The inverse of the Fisher Information Matrix gives the covariance matrix which has the estimated variance of each parameter as follows:

$$Var(\hat{\theta}) = [I(\hat{\theta})]^{-1} = \begin{bmatrix} var(\theta_1) & cov(\theta_1, \theta_2) & \cdots & cov(\theta_1, \theta_n) \\ cov(\theta_2, \theta_1) & var(\theta_2) & \cdots & cov(\theta_2, \theta_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\theta_n, \theta_1) & cov(\theta_n, \theta_2) & \cdots & var(\theta_n) \end{bmatrix}$$

- Using these values, the desired confidence intervals of each parameters can be found.

t-Distribution table

- Student's *t*-distribution (the *t*-distribution): a distribution that arises when estimating the mean of a normally distributed variable when sample size is small and (population) standard deviation is unknown.
- Note: t distribution is symmetrical, e.g., $t_{1-\frac{\gamma}{2}}(df)$=- $t_{\frac{\gamma}{2}}(df)$
- Lookup table in Appendix A.
- Excel: Use t.inv ($\frac{\gamma}{2}, df$) or t.inv.2t(γ, df)

MLE parameters of various distributions

- **Normal distribution: Complete data**
  - We know that for a sample of size n

$$\hat{\mu} = \frac{\sum_{i=1}^{n} t_i}{n} \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(t_i - \hat{\mu})^2}{n-1}$$

confidence interval for mean (when $\sigma$is unknown as estimated as s)

$$\hat{\mu} - \frac{s}{\sqrt{n}} * t_{\frac{\gamma}{2}}(n-1) \leq \mu \leq \hat{\mu} + \frac{s}{\sqrt{n}} * t_{\gamma/2}(n-1)$$

Where:

$(1-\gamma)$ = confidence level,

$t \rightarrow$ one-tailed t-distribution. (Or use t_{γ} with two-tailed distribution.)

$df = n - 1$ = degrees of freedom

MLE parameters of various distributions

- **Normal distribution: right censored data**
- For a sample of size n where m components fail (n > m)

$$\hat{\mu} = \frac{\sum_{i=1}^{m} t_i}{n} \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{m} (t_i - \hat{\mu})^2}{n-1}$$

  - The confidence interval for mean (when σ is unknown and estimated as s)

$$\hat{\mu} - \frac{s}{\sqrt{n}} * t_{\frac{\gamma}{2}}(m-1) \le \mu \le \hat{\mu} + \frac{s}{\sqrt{n}} * t_{\frac{\gamma}{2}}(m-1)$$

where

$(1-\gamma)$ = confidence level,

$t \rightarrow$ one-tailed t-distribution (Or use t_γ with two-tailed t distribution)

$with\ m-1 = degrees\ of\ freedom$

MLE parameters of various distributions

- **Normal distribution (cont.): complete data and right censored**
  - Confidence limits for variance with complete data is:

$$\frac{(n-1)s^2}{\chi^2_{1-\frac{\gamma}{2}}[n-1]} \le \sigma^2 \le \frac{(n-1)\hat{s}^2}{\chi^2_{\frac{\gamma}{2}}[n-1]}$$

  - With right censored data, it is:

$$\frac{(n-1)s^2}{\chi^2_{1-\frac{\gamma}{2}}[m-1]} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{\frac{\gamma}{2}}[m-1]}$$

when m failures occur in n observations

MLE parameters of various distributions

- **Lognormal distribution: Complete data**

Note: numerical methods are required for dealing with *incomplete data.* So only complete data are presented for lognormal dist. See textbook.

$$\hat{\mu} = \frac{\sum_{i=1}^{n} lnt_i}{n}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(lnt_i - \hat{\mu}_t)^2}{n-1}$$

MLE parameters of various distributions

- **Lognormal distribution: complete data**
  - Two-sided confidence limits on $\hat{\mu}$ **t → one-tailed t-distribution**

$$\hat{\mu} - \frac{\hat{\sigma}}{\sqrt{n}} * t_{\frac{\gamma}{2}}(n-1) < \hat{\mu} < \hat{\mu} + \frac{\hat{\sigma}}{\sqrt{n}} * t_{\frac{\gamma}{2}}(n-1)$$

  - One sided confidence limits on $\hat{\mu}_t$

$$0 \leq \hat{\mu} < \hat{\mu} + \frac{\hat{\sigma}}{\sqrt{n}} * t_{\gamma}\ (n-1)$$

$$\hat{\mu} - \frac{\hat{\sigma}}{\sqrt{n}} * t_{\gamma}(n-1) < \hat{\mu} < \infty$$

  - And corresponding on σ_t^2:

$$\frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\frac{\gamma}{2}}(n-1)} \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{\chi^2_{\frac{\gamma}{2}}(n-1)}$$

$$0\ \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{\chi^2_{\gamma}(n-1)}$$

$$\frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\gamma}(n-1)} \leq \sigma^2 \leq \infty$$

Example: Lognormal MLE & confidence intervals

- **Example:** Consider the follow time-to-failure values t. Assuming the data are from a lognormal distribution, find:
- A) point estimates of the parameters
- B) the 90% confidence interval on the parameters
- C) The MTTF

$\boldsymbol{Lnt_i}$	4.3	4.7	5.3	5.7	5.9	6.0	6.2
$\boldsymbol{t_i}$	75	115	192	312	389	410	496

Solution: Lognormal MLE & confidence intervals

$\boldsymbol{Lnt_i}$	4.3	4.7	5.3	5.7	5.9	6.0	6.2
$\boldsymbol{t_i}$	75	115	192	312	389	410	496

- **Solution A)** $\hat{\mu}_t = \Sigma \frac{lnt_i}{n} = 5.46 \quad \hat{\sigma}_t = 0.71$
- **Solution B) For $\boldsymbol{\mu_t}$:**

$$\hat{\mu} - \frac{\hat{\sigma}}{\sqrt{n}} * t_{\frac{\gamma}{2}}(n-1) < \mu < \hat{\mu} + \frac{\hat{\sigma}}{\sqrt{n}} * t_{\frac{\gamma}{2}}(n-1)$$

n=7

$t_{\left(\frac{\gamma}{2}\right)}(7-1) = t_{0.05}(6) = 1.943$

$$5.46 - \frac{0.71}{\sqrt{7}} * 1.943 \leq \mu \leq 5.46 + \frac{0.71}{\sqrt{7}} * 1.943$$

$$\mathbf{4.93 \leq \boldsymbol{\mu} \leq 5.98}$$

Solution: Lognormal MLE & confidence intervals

- **Solution B: for $\boldsymbol{\sigma^2}$**

$$\frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\frac{\gamma}{2}}(n-1)} \le \sigma^2 \le \frac{(n-1)\hat{\sigma}^2}{\chi^2_{\frac{\gamma}{2}}(n-1)}$$

$$\frac{(6)0.71^2}{\chi^2_{0.95}(6)} \le \sigma^2 \le \frac{(6)0.71^2}{\chi^2_{0.05}(6)}$$

$$\frac{(6)0.71^2}{12.59} \le \sigma^2 \le \frac{(6)0.71^2}{1.64}$$

$$\mathbf{0.240} \le \boldsymbol{\sigma^2} \le \mathbf{1.844}$$

- **Solution C:**

$$\mathbf{MTTF} = \widehat{\boldsymbol{E}(\boldsymbol{t})} = \exp\left(\hat{\mu} + \frac{\hat{\sigma}^2}{2}\right) = \exp\left(5.46 + \frac{0.71^2}{2}\right) = \mathbf{302.49}$$

MLE parameter confidence intervals - Binomial

- **Binomial distribution**
  - Recall the binomial distribution:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where

$$C_x^n = \binom{n}{x} = \frac{n!}{(n-x)!\,x!}$$

$p \rightarrow$ probability of failure

$q \rightarrow$ 1-p probability of success

$n \rightarrow$ number of trials

$x \rightarrow$ number of failures out of n trials

MLE parameter confidence intervals - Binomial

MLE for complete data: $\hat{p} = \frac{x}{n}$

Confidence limits on p can be found from the **Clopper-Pearson procedure:**

$$p_{Lower} = \left\{1 + \frac{(n-x+1)}{x} F_{1-\frac{\gamma}{2}}(2n-2x+2; 2x)\right\}^{-1}$$

$$p_{Upper} = \left\{1 + \frac{n-x}{(x+1)F_{1-\frac{\gamma}{2}}(2x+2; 2n-2x)}\right\}^{-1}$$

where $F_{1-\gamma/2}(f_1; f_2)$ is the F distribution with f_1 and f_2 degrees of freedom to the right and left, respectively, for $(1 - \gamma/2)$ confidence level.

See Tables in Appendix. Or in Excel: F.INV($1 - \gamma/2, f_1, f_2$)

MLE parameter confidence intervals (cont.)

- **Example:** An emergency pump is in standby mode. There have been 563 start tests for the pump, and 3 failures have been observed.
  A. Estimate the probability of failure on demand, p_{fod} and
  B. Find the 90% confidence interval for the probability of failure on demand.

MLE parameter confidence intervals (cont.)

- **Solution:**
  A. Number of trials, n = 563, and 3 failures, therefore:

$$\hat{\boldsymbol{p}}_{\boldsymbol{fod}} = \frac{x}{n} = \frac{3}{563} = \mathbf{0.0053}$$

MLE parameter confidence intervals (cont.)

- **Solution:**
  B. For confidence level 90%, $(1-\gamma) = .9,$ thus $\frac{\gamma}{2} = 0.05, 1 - \frac{\gamma}{2} = 0.95$
- Given x=3 failures in n=563 tests:

$$p_L = \left\{1 + \frac{(563 - 3 + 1)}{3} * F_{(0.95)}\ (1122; 6)\right\}^{-1}$$

$$= \frac{1}{1 + \frac{561}{3} * 3.67} = 0.00145$$

$$p_U = \left\{1 + \frac{563-3}{(3+1)F_{0.95}(8;1120)}\right\}^{-1} = \frac{1}{1+\frac{560}{4*1.94}} = 0.0137$$

$$\boxed{\begin{array}{c} \hat{\boldsymbol{p}} = \mathbf{0.0053} \\ \mathbf{0.00145} \le \boldsymbol{p} \le \mathbf{0.0137}, \end{array}}$$

MLE parameter confidence intervals (cont.)

- **Weibull distribution**
  - When there is only complete failure and/or right censored data the point estimates can be solved using the following expressions.
  - Note that **numerical methods** are needed to solve $\hat{\beta}$ then substitute to find $\hat{\alpha}$. *To try this out: Use the SOLVER function in in Excel.*

$$\hat{\beta} = \left[\frac{\Sigma(t_i)^{\hat{\beta}} ln(t_i) + (n-r)(t_r)^{\hat{\beta}} ln(t_r)}{\Sigma(t_i)^{\hat{\beta}} + (n-r)(t_r)^{\hat{\beta}}} - \frac{1}{r}\sum ln(t_i)\right]^{-1}$$

Where t_i are complete data and t_r are right censored; r is the number of complete data points.

$$\hat{\alpha} = \left[\frac{\Sigma(t_i)^{\hat{\beta}}}{n} + (n-r)(t_r)^{\hat{\beta}}\right]^{\frac{1}{\hat{\beta}}}$$

MLE parameter confidence intervals (cont.)

- **Weibull distribution (cont.)**
  - Are derived from Fisher information Matrix and require numerical methods to solve.

$$\hat{\beta}\exp\left(-Z_{1-(\gamma/2)}\frac{\sqrt{\mathrm{var}(\hat{\beta})}}{\hat{\beta}}\right)\le\beta\le\hat{\beta}\exp\left(Z_{1-(\gamma/2)}\frac{\sqrt{\mathrm{var}(\hat{\beta})}}{\hat{\beta}}\right),$$

$$\hat{\alpha}\exp\left(-Z_{1-(\gamma/2)}\frac{\sqrt{\mathrm{var}(\hat{\alpha})}}{\hat{\alpha}}\right)\le\alpha\le\hat{\alpha}\exp\left(Z_{1-(\gamma/2)}\frac{\sqrt{\mathrm{var}(\hat{\alpha})}}{\hat{\alpha}}\right),$$

$$I(\alpha,\beta)=\begin{bmatrix}\frac{\beta^2}{\alpha^2} & \frac{\Gamma'(2)}{-\alpha}\\ \frac{\Gamma'(2)}{-\alpha} & \frac{1+\Gamma''(2)}{\beta^2}\end{bmatrix}=\begin{bmatrix}\frac{\beta^2}{\alpha^2} & \frac{1-\gamma}{\alpha}\\ \frac{1-\gamma}{\alpha} & \frac{\frac{\pi^2}{6}+(1-\gamma^2)}{\beta^2}\end{bmatrix}\cong\begin{bmatrix}\frac{\beta^2}{\alpha^2} & \frac{0.422784}{-\alpha}\\ \frac{0.422784}{-\alpha} & \frac{1.823680}{\beta^2}\end{bmatrix}$$

MLE parameter confidence intervals (cont.)

- **Example:** 5 components are put on a test with the following failure times: 535, 613, 976, 1031, 1875 hours
  - $\hat{\beta}$ is found by numerically solving:

$$\boxed{\hat{\beta} = \left[\frac{\Sigma\left(t_i^F\right)^{\hat{\beta}} ln\left(t_i^F\right)}{\Sigma\left(t_i^F\right)^{\hat{\beta}}} - 6.8118\right]^{-1} = 2.275}$$

  - $\hat{\alpha}$ is found by solving:

$$\boxed{\hat{\alpha} = \left[\frac{\Sigma\left(t_i^F\right)^{\hat{\beta}}}{n_F}\right]^{\frac{1}{\hat{\beta}}} = 1140}$$

$$\boxed{cov\left(\hat{\alpha}, \hat{\beta}\right) = \begin{bmatrix} 278386 & 293 \\ 293 & 3.1463 \end{bmatrix}}$$

MLE Parameters of various distributions

Gamma Distribution

$$L(\alpha, \beta | t_i \dots) = \frac{1}{\beta^{\alpha n}\Gamma(\alpha)^n} \prod_{i=1}^{n} t_i^{\alpha-1} \exp\left(-\frac{t_i}{\beta}\right)$$

$$\ln L = \alpha n \ln\left(\frac{1}{\beta}\right) - n\ln[\Gamma(\alpha)] + (\alpha - 1)\sum_{i=1}^{n} \ln(t_i) - \frac{1}{\beta}\sum_{i=1}^{n} t_i$$

$$\left.\frac{\partial \ln L}{\partial \alpha}\right|_{\alpha=\hat{\alpha}} = n\ln\left(\frac{1}{\beta}\right) - n\Psi(\alpha) + \sum_{i=1}^{n} \ln(t_i) = 0$$

$$\left.\frac{\partial \ln L}{\partial \beta}\right|_{\beta=\hat{\beta}} = \alpha\beta n - \sum_{i=1}^{n} t_i = 0$$

where $\Psi(\alpha) = \frac{d}{dx}\ln[\Gamma(\alpha)]$ digamma function

- Solve by using numerical methods on both equations simultaneously

MLE Parameters of various distributions

Continuous Uniform Distribution

$$\boldsymbol{\hat{a} = \min(t_i, \dots, t_n)}$$
$$\boldsymbol{\hat{b} = \max(t_i, \dots, t_n)}$$

Beta Distribution

$$L(\alpha, \beta; \mathrm{t_i} \dots) = \frac{\Gamma(\alpha+\beta)n}{\Gamma(\alpha)\Gamma(\beta)} \prod_{i=1}^{n} t_i^{\alpha-1}(1-t_i)^{\beta-1}$$

$$lnL = n\{\ln[\Gamma(\alpha+\beta)] - \ln[\Gamma(\beta)]\} + (\alpha-1)\sum_{i=1}^{n} lnt_i + (\beta-1)\sum_{i=1}^{n} \ln(1-t_i)$$

$$\boldsymbol{\left.\frac{\partial lnL}{\partial \alpha}\right|_{\alpha=\hat{\alpha}} = \Psi(\alpha) - \Psi(\alpha+\beta) - \frac{1}{n}\sum_{i=1}^{n} \ln t_i = 0}$$

$$\boldsymbol{\left.\frac{\partial lnL}{\partial \beta}\right|_{\beta=\hat{\beta}} = \Psi(\beta) - \Psi(\alpha+\beta) - \frac{1}{n}\sum_{i=1}^{n} \ln(1-t_i) = 0}$$

where $\Psi(\alpha) = \frac{d}{d\alpha}\ln[\Gamma(\alpha)]$ is the digamma function

- Solve by using numerical methods on both equations simultaneously

MLE Parameters of various distributions

Truncated Normal Distribution

- First find point estimates for $z_a = \frac{a_L - \mu}{\sigma}$ and $z_b = \frac{b_U - \mu}{\sigma}$

$H_1(z_1, z_b) = \frac{\hat{\mu} - a_L}{b_U - a_L}$ $\quad\quad$ $\hat{\mu} = \frac{1}{n}\sum_0^n x_i$

$H_2(z_a, z_b) = \frac{\sigma^2}{(b_U - a_L)^2}$ $\quad\quad$ $\hat{\sigma}^2 = \frac{1}{n-1}\sum_0^n (x_i - \mu)^2$

- Solving for H_1 and H_2 simultaneously gives:
- $\boldsymbol{\hat{\sigma} = \frac{b_U - a_L}{\hat{z}_b - \hat{z}_a}}$ $\quad\quad$ $\boldsymbol{\hat{\mu} = a_L - \hat{\sigma}}$

MLE Parameters of various distributions

Multivariate Normal Distribution

$$\widehat{\vec{\boldsymbol{\mu}}} = \frac{1}{n}\sum_{t=1}^{n} \vec{x}_t$$

$$\hat{\Sigma}_{ij} = \frac{1}{n-1}\sum_{t=1}^{n}(x_{i,t} - \hat{\mu}_i)(x_{j,t} - \hat{\mu}_j)$$

$$\vec{x}_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{d,t} \end{bmatrix}, t = 1,2,\dots,n$$

Bivariate Normal Distribution

$$\hat{\mu}_{x_1} = \frac{1}{n}\sum_{i=1}^{n} x_{i,1};\ \hat{\sigma}_{x_1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{i,1} - \hat{\mu}_{x_1}\right)^2$$

$$\hat{\mu}_{x_2} = \frac{1}{n}\sum_{i=1}^{n} x_{i,2};\ \hat{\sigma}_{x_2}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{i,2} - \hat{\mu}_{x_2}\right)^2$$

$$\hat{\rho} = \frac{1}{n\hat{\sigma}_{x_1}\hat{\sigma}_{x_2}}\sum_{i=1}^{n}(x_{i,1} - \hat{\mu}_{x_1})(x_{i,2} - \hat{\mu}_{x_2})$$

Nonparametric methods for censored data

- Nonparametric (Empirical) reliability estimates with censored data - Apply to right censored data.
  - Relates to the nonparametric methods we learned earlier: Kimball, Kaplan Meier Plotting methods we discussed earlier.
  - There are ways to do this by relating the sample to the binomial distribution – see textbook.
- Start with sorted (ordered) data, and note the units that actually failed vs. censored data.
  - Right censored data are shown by a + sign).
  - 150, 340+, 560, 800, 1130+, 1720, 2470+, 4210+, 5230, 6890

Rank Adjustment Method

- The *rank adjustment method* is the most accurate method for plotting censored failure data.
  - For *n* units tested, ordered from $t_{i=1} \leq t_{i=2} \leq \cdots \leq t_{i=n}$where *m* units have survived including and beyond the ith unit..
- We calculated a rank adjustment (or order) for each data point:

$i_{t_i} = j_{t_{i-1}} + \frac{(n+1)-j_{t_{i-1}}}{1+m}$ (Call i_{t_i} the *initial rank*)

- Note: $\mathrm{i}_{\mathrm{t}_1} = 1$;
  - The *adjusted rank,* j_{t_i}, for non-censored units can be calculated as:

$$\mathrm{j}_{\mathrm{t}_\mathrm{i}} = \mathrm{j}_{\mathrm{t}_{\mathrm{i}-1}} + \frac{(\mathrm{n}+1) - \mathrm{j}_{\mathrm{t}_{\mathrm{i}-1}}}{\mathrm{n} - \mathrm{i}_{\mathrm{t}_\mathrm{i}} + 2}$$

where (*i*) is the initial rank (order) and (*j*) is the adjusted rank. We then use the non-parametric estimators to plot.

Rank adjustment method for censored data plots (cont.)

- **Example:** Given the following failure times for n = 10 components, 150, 340+, 560, 800, 1130+, 1720, 2470+, 4210+, 5230, 6890

A) Use the rank adjustment method to create a plot of the Weibull distribution for the failure times.

B) Assuming the Weibull is a suitable fit, use the plot to estimate the parameters of the Weibull distribution.

Rank adjustment method for censored data plots (cont.)

- **Solution A)**

Weibull plotting positions

i	t_i (hrs)	$j_{t_i} = j_{t_{i-1}} + \frac{(n+1) - j_{t_{i-1}}}{2 + n - i_{t_i}}$	$F(t_i) = \left(\frac{j_{t_i} - 0.375}{n + 0.25}\right)$	R(t) =1-F(t)	Ln(t)	$\ln(\ln(\frac{1}{R(t)}$
1	150	1	$\left(\frac{1-0.375}{10+0.25}\right) = .061$	0.939	5.011	-2.77
2	340+	-	-			
3	560	$1 + \frac{(10+1) - 1}{2 + 10 - 3} = 2.111$	$\left(\frac{2.111-0.375}{10+0.25}\right) = 0.169$	0.831	6.328	-1.68
4	800	$2.111 + \frac{(10+1) - 2.111}{2 + 10 - 4} = 3.222$	.2778	0.722	6.685	-1.12
5	1130+	-	-			
6	1720	$3.222 + \frac{(10+1) - 3.222}{2 + 10 - 6} = 4.519$	0.4042	0.596	7.450	-0.66
7	2470+	-	-			
8	4210+	-	-			
9	5230	6.679	0.615	0.385	8.562	-0.05
10	6890	8.84	0.8258	0.174	8.838	0.56

Rank adjustment method for censored data plots (cont.)

Solution B)

Least Squares regression equation:
y=0.815x-6.777
R^2=0.98

Weibull parameters:
α=4084.8
$\beta = 0.815$
(Reminder, slope=β)

(You can also compare to the MLE estimate of parameters: α=3926.86; $\beta = 0.97$)

Kaplan-Meier method (aka product-limit method)

- Applicable for right censored data. And grouped data (see textbook)
- Each term in the expression below is **Conditional Probability of Surviving** past time t. The product is the **Unconditional Surviving Probability (aka the survival function):**

$$\hat{R}(t) = \Pi_{t_j \leq t} \left(\frac{n_j - 1}{n_j} \right)$$

where j = reverse rank.

- For $0 \leq t \leq t_i \quad R(t) = 1$

A measure of uncertainty of the estimated reliability is:

$$Var\left[\hat{R}(t)\right] = \Sigma_{t_j \leq t} \frac{1}{n_j(n_j - 1)}$$

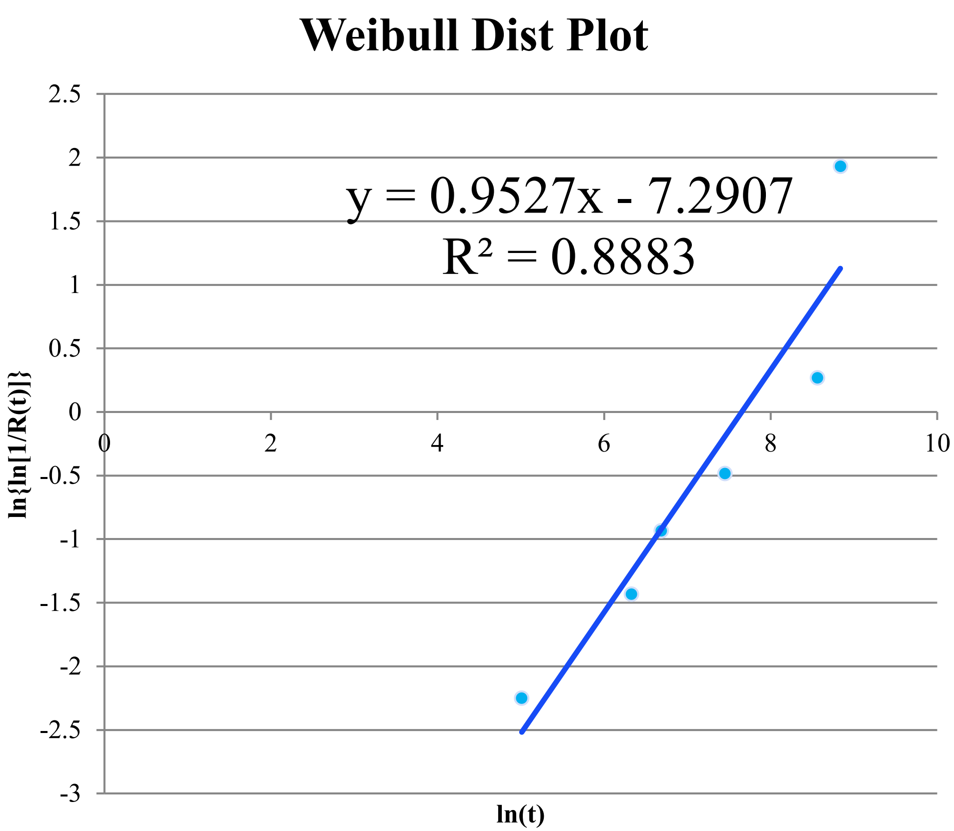
Kaplan-Meier method (cont.)

- **Example using same data as before:**

Weibull plotting positions

i	t_i	j	$\frac{n_j-1}{n_j}$	$\hat{R}(t_i)=\Pi_{t_j \le t}\left(\frac{n_j-1}{n_j}\right)$	ln(t)	Ln(ln(1/R(t))
1	150	10	0.900	0.900 x 1.000 = 0.900	5.01	-2.25
2	340+	9	-	-		
3	560	8	0.875	0.875 x 0.900 = 0.788	6.33	-1.43
4	800	7	0.857	0.857 x 0.788 = 0.675	6.68	-0.93
5	1130+	6	-	-		
6	1720	5	0.800	0.800 x 0.675 = 0.540	7.45	-0.48
7	2470+	4	-	-		
8	4210+	3	-	-		
9	5230	2	0.500	0.500 x 0.540 = 0.270	8.56	0.27
10	6890	1	0.000	0.000 x 0.270 = 0.000	8.84	1.93

Kaplan-Meier plot



- $\beta = 0.953, \alpha = 2105$

Kaplan-Meier method (cont.)

- **Note:** If two or more failure occurs at time t_j then

$$\hat{R}(t) = \Pi_{t_j \leq t}\left(\frac{n_j - d_j}{n_j}\right)$$

where d = number of failure in the jth time ranking

Reliability Analysis

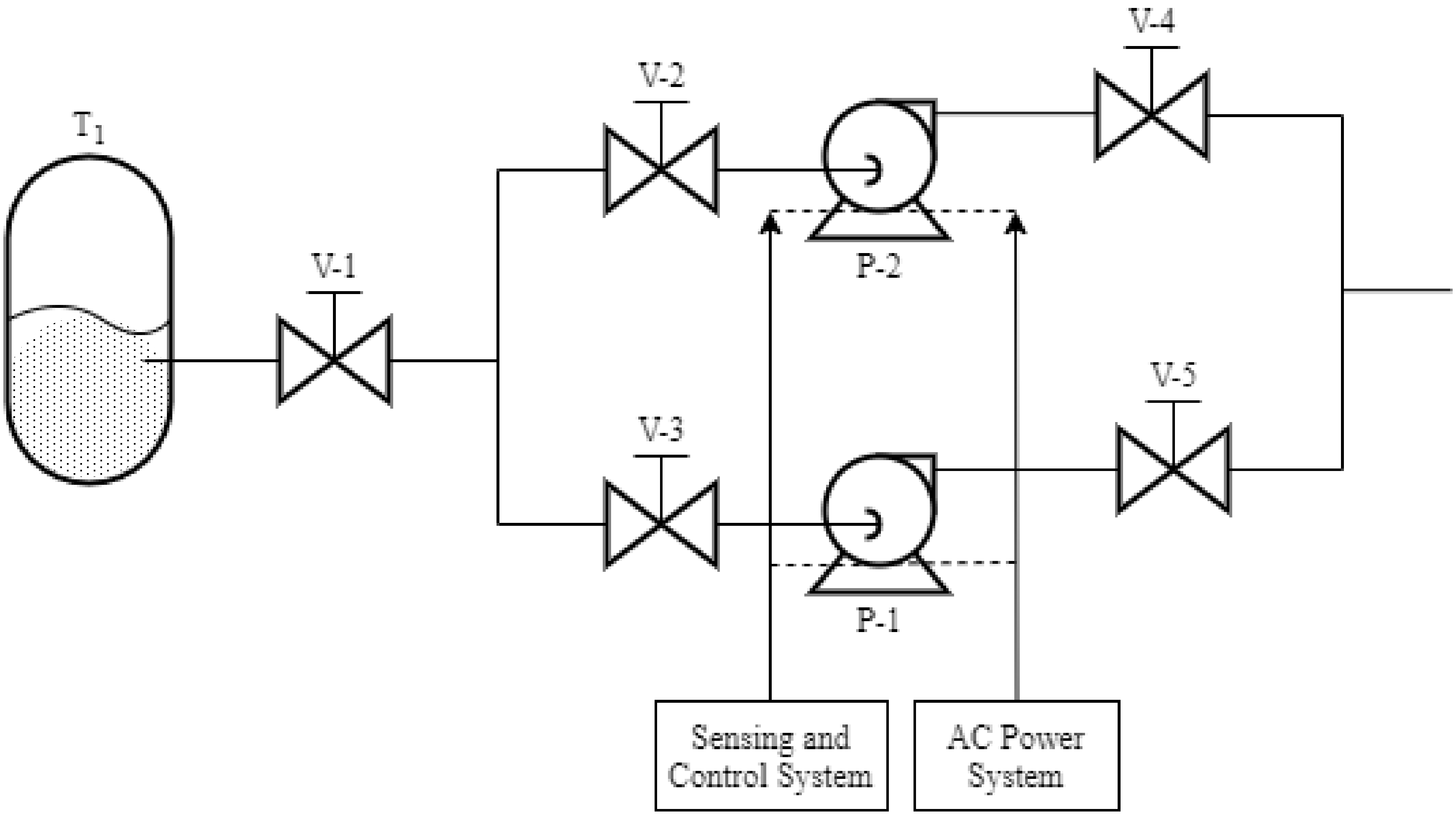
Module 6A and 6B: System Reliability Analysis

Prof. Katrina M. Groth
Mechanical Engineering Department
Center for Risk and Reliability
University of Maryland
kgroth@umd.edu

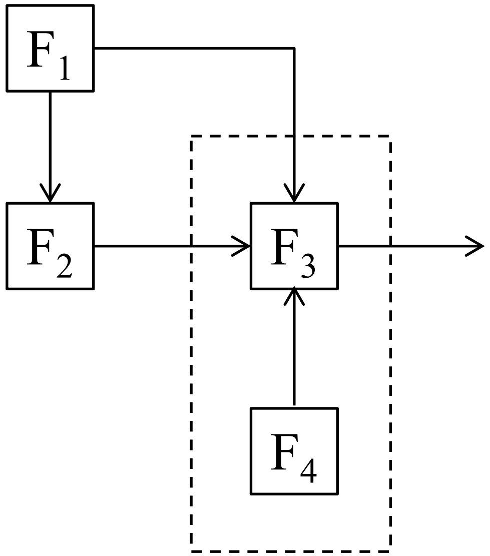
System reliability analysis: Outline

- We will talk about $R_s = f\left(R_{c_{i=1}}(t), R_{c_2}(t), \ldots, R_{c_n}(t)\right)$
- Reliability Block Diagrams & Types of Systems
  1. Parallel, Series, *k*-out-of-*n*, Standby and Shared load systems
  2. Complex systems
- Complex system evaluation methods
- Fault trees/Success trees
- Event trees & Event Sequence Diagrams
- Failure Mode and Effect Analysis (and Failure Mode and Effect Criticality Analysis)

System reliability analysis: An example of a pumping system



Functional block diagram for the pumping system



Function	Functional Description	Components Involved
F_1	AC Power	AC
F_2	Sensing & Control	S
F_3	Pumping	V-2, V-3, V-4, V-5, P-1, P-2
F_4	Source	T-1, V-1

Failure modes of some relevant components (reminder: Chapter 1 defines failure modes)

Component Type	Failure Modes
All Valves	1 – Fail to open 2 – Fail to close 3 – Spurious closure 4 – Plug 5 – Leak 6 – Partially closed
All Pumps	1 – Fail to start 2 – Fail to run 3 – Leak 4 – High internal circulation
Tank	1 – Rupture 2 – Frozen 3 – Leak
AC Power	1 – Lost
Controller	1 – Failed 2 – Biased

Notation

- **Be very careful with your notation.**
  - Overbar always indicates negation
- **Typically, we use:**
  - $\overline{A}$ or R_A for the event that the corresponding item *works*
  - A or F_A for the event that the corresponding item *fails*
  - But this gets flipped in some contexts and problems. Pay attention to the problem statement to determine which is which!
  - Make sure your notation & problem is set up correctly: know what is success and what is failure. Define your own notation if needed.
  - In general: λ and similar parameters usually refer to *failure rate*, but this is not always the case (especially as we progress in this class).

Reminders & Notes for Systems problems

- There are two aspects to system models: qualitative & quantitative.
  - Qualitative: The system structure functions describe the system (failure) logic as a function of its components.
  - Quantitative: The probabilities & mathematics used in the quantification of the system models.
- We use Boolean algebra to manipulate events & sets.
- We use probabilities to talk about uncertainty of events.
- In general: Do the Boolean algebra first, then the probabilities. Don't start quantifying too early.
- Most of the quantification in this chapter assumes independent events.
  - You **must** verify independence before assuming it!

Types of Systems

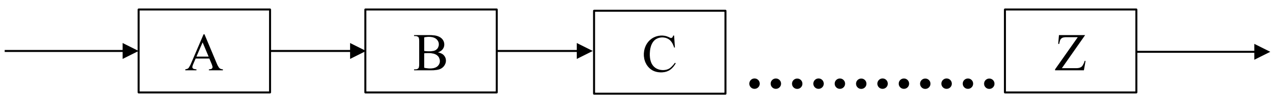
- Series
- Parallel
- k-out-of-n (combination of above)
- Standby
- Shared Load
- Complex systems—combinations of above

Reliability block diagrams (RBDs)

- Success-oriented method that can provide a good visualization of simpler systems
- Show the functional configuration of the system
  - Which component(s) are required for the system to work
  - Which component(s) failure will cause the system to fail
- RBDs often correspond to the physical configuration of the system
  - **However**, there may be instances when the RBD does not model the physical configuration

RBDs: Series

- **Series System:**



  - When ***all*** of the items ***work***, system ***works***, that is:

$$R_s(t) = \Pr(\bar{A} \cap \bar{B} \cap \bar{C} \cap \cdots \cap \bar{Z})$$
$$R_s(t) = \Pr(R_A \cap R_B \cap R_C \cap \cdots \cap R_Z)$$

  - If these events are **independent** ($A \perp B \perp C \ldots$), this becomes:

$$R_s(t) = R_A(t) \cdot R_B(t) \cdot R_C(t) \cdot \ldots \cdot R_z(t)$$
$$R_s(t) = \prod_{i=1}^{n} R_i(t)$$

  - If they're not independent, use the chain rule of probability

RBDs: Series

- **MTTF of the System (under certain assumptions*):**
  - If each component has a **constant failure rate** (i.e., the times to failure, *t*, are exponentially distributed: $R_i(t) = e^{-\lambda_i t}$ for each component *i*) and they are independent.
  - The reliability of the system is:

$$R_s(t) = \prod_{i=1}^{n} e^{-\lambda_i t} = e^{-\sum_{i=1}^{n} \lambda_i t} = e^{-\lambda_s t}$$

  - Therefore: $\lambda_s = \sum_{i=1}^{n} \lambda_i$

$$MTTF_s = \frac{1}{\lambda_s} = \frac{1}{\sum_{i=1}^{n} \lambda_i}$$

  - ***If these assumptions aren't met, you would calculate MTTF using the formulas in Ch. 3, or another model we'll learn.**

$$MTTF = E(t) = \int_0^{\infty} x f_s(x) dx = (if \ \lim tf(t) \to 0) \int_0^{\infty} R_s(x) dx$$

RBDs: Series

- **For n independent & identically distributed (i.i.d) units** with constant failure rate (CFR):

$$R_s(t) = e^{-n\lambda t}$$

$$\lambda_S = n\lambda$$

$$MTTF_S = \frac{1}{n\lambda}$$

Example: Series system

- **Example:**
- You have four-component series system, where the components are iid with constant failure rate. If $R_s(100 \text{ hrs})$ is 0.95, find the individual component $MTTF_i$ and the $MTTF_s$ of the system.

Example: Series system

- **Solution:** Four component series system where the components are iid with CFR. If R_s(100 hrs) is 0.95, find the individual component MTTF.

$$n = 4, \lambda_i = const., R_s(100) = 0.95$$

Find $MTTF_i$

$R_s = e^{-n\lambda_i t}$

$\ln(0.95) = -n\lambda_i t = -400\lambda_i$

$-.05129 = -400\lambda_i$

$\lambda_i = 1.28 \times 10^{-4}$

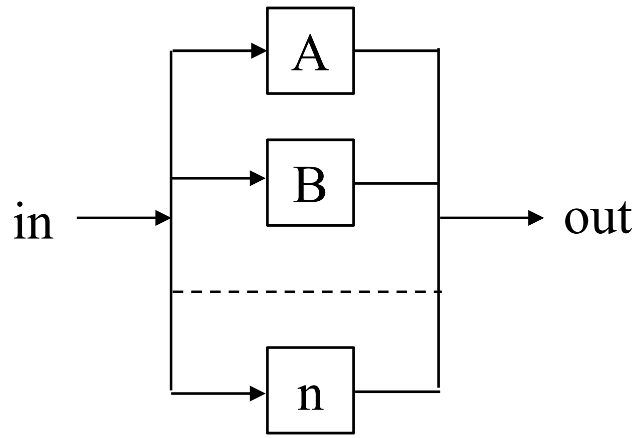
$\boldsymbol{MTTF_i = \frac{1}{\lambda_i} = 7798.29}$

For comparison:

$\boldsymbol{MTTF_S = \frac{1}{4 \cdot \lambda_i} = 1949.6 hrs}$

RBDs: Parallel

- **Parallel System**



- System fails when **all** the blocks fail:

$$F_s(t) = \Pr(F_A \cap F_B \cap F_C \dots)$$
$$R_S(\mathrm{t}) = 1 - \Pr(F_A \cap F_B \cap F_C \dots)$$

RBDs: Parallel

- *If* these events are **independent**, $(A \perp B \perp C \ldots)$, this becomes:

$$F_S(t) = F_A(t) \cdot F_B(t) \cdot F_C(t) \ldots = \prod_{i=1}^{n} F_i(t)$$

$$where\ F_i(t) = Unreliability\ of\ one\ component$$

$$R_S(t) = 1 - \prod_{i=1}^{n} F_i(t)$$

$$R_s(t) = 1 - \prod_{i=1}^{n} [1 - R_i(t)]$$

- As before, if they are not independent, use the chain rule.

RBDs: Parallel

- **MTTF of the System (under certain assumptions):**
  - If each component has a constant failure rate, (i.e., the times to failure are exponentially distributed; i.e., $R_i(t) = e^{-\lambda_i t}$), and they are independent.
  - **The hazard rate of a parallel system is not constant, so $\boldsymbol{MTTF_s \neq \frac{1}{\lambda_s}}$**

Dervice from $h_s(t) = \frac{f(t)}{R(t)} = \frac{-\frac{dR(t)}{dt}}{R(t)}$

$$MTTF_s = \int_0^{\infty} R_s(t)dt$$

  - For a parallel system of two components:

$$MTTF_s = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2}$$

(See textbook for binomial expansion to derive for additional items).

RBDs: Parallel

- Special case: When all n components are i.i.d with CFR λ_i

$$R_s(t) = 1 - (1 - e^{-\lambda_i t})^n$$

$$MTTF_i = \frac{1}{\lambda_i}$$

$$\boxed{MTTF_s = MTTF_i \left(1 + \frac{1}{2} + \frac{1}{3} \ldots + \frac{1}{n}\right)}$$

Example: Parallel system

- **Example:**
- Two parallel identical and independent components have CFR. It is desired that $R_s(1000hr) = 0.95$. Find the component $MTTF_i$ and the system $MTTF_s$.

Example: Parallel system

- **Solution:** Two parallel identical and independent components have CFR. It is desired that $R_s(1000hr) = 0.95$. Find the component $MTTF_i$ and the system $MTTF_s$.
- $R_s(1000hr) = 0.95 = 1 - (1 - e^{-\lambda_i t})^2$

$0.05 = (1 - e^{-\lambda_i t})^2$

$\ln(0.05) = -2.99 = 2\ln(1 - e^{-\lambda_i t})$

$e^{(-2.99/2)} = 0.223 = 1 - e^{-\lambda_i t}$

$0.776 = e^{-\lambda_i t}$

$\lambda_i = 2.531e^{-4} \Rightarrow \boldsymbol{MTTF_i = 3951.1hr}$

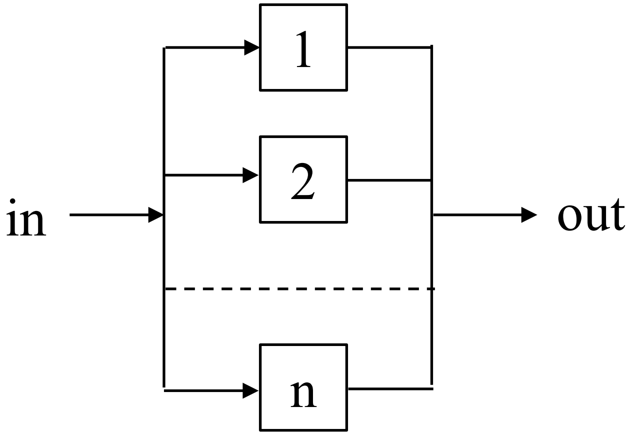
$\boldsymbol{MTTF_s} = MTTF_i\left(1 + \frac{1}{2} + \frac{1}{3} \dots + \frac{1}{n}\right) = 3951.1 \cdot (1 + 0.5) = \boldsymbol{5926hr}$

RBDs: *k*-out-of-*n* redundant system

- ***k*-out-of-*n* system**
  - If any of the **k** blocks **out of n** iid blocks **works** so that **the system works**, then:

$$R_s(t) = \sum_{x=k}^{n} \binom{n}{x} [R_i(t)]^x [1 - R_i(t)]^{n-x}$$

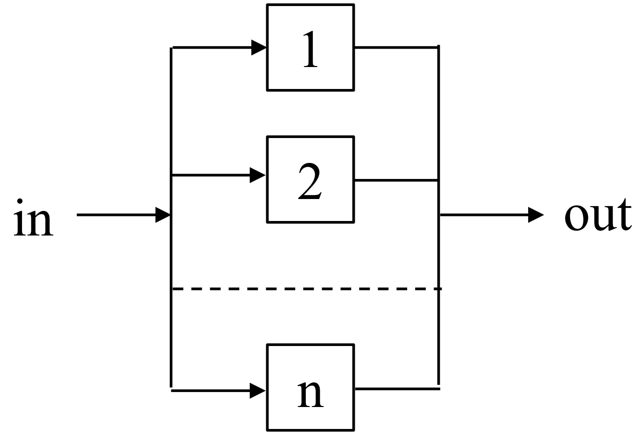
$$= 1 - \sum_{x=0}^{k-1} \binom{n}{x} [R_i(t)]^x [1 - R_i(t)]^{n-x}$$



  - Recall: $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

Example (k-out-of-n redundant system)

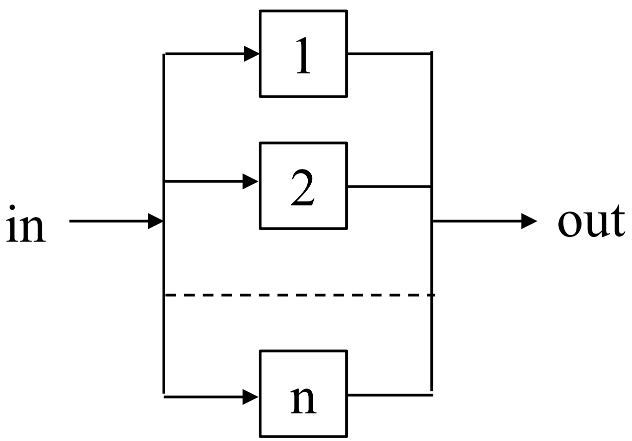
- **Example:** Calculate R_S the reliability of a system for a 2-out-of-3 for success case. Assume that the 3 components are i.i.d and each has reliability R.



Example (k-out-of-n redundant system): Solution

- **Solution:** Calculate R_S the reliability of a system for a 2-out-of-3 for success case. Assume that the 3 components are iid and each has reliability R.

  Given n = 3, k = 2



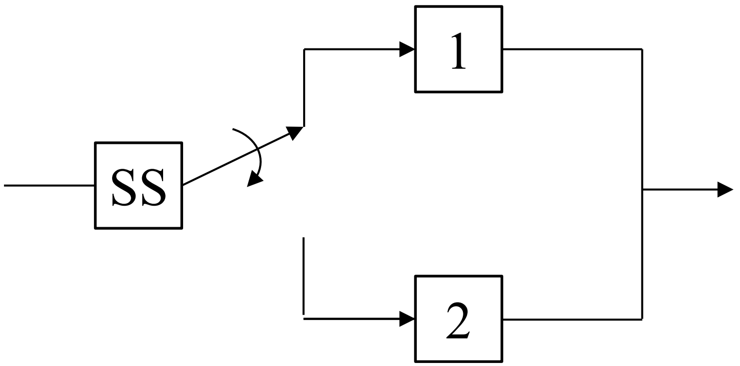
$$R_s(t) = \sum_{x=k}^{n} \binom{n}{x} [R_i(t)]^x [1 - R_i(t)]^{n-x}$$

$$R_s = \frac{3!}{2! \times 1!} (R)^2 (1 - R) + \frac{3!}{3! \times 0!} (R)^3 (1 - R)^0$$

$$= \boldsymbol{3(R)^2(1 - R) + R^3 = 3R^2 - 2R^3}$$

RBDs: Standby redundant system

- **Standby Redundant System**
  - For simplicity consider the two unit standby system pictured below.



  - Operation can be categorized as:

1. Block 1 operates until it fails
2. Sensing/switch recognizes Block 1 failure and switches to Block 2
3. Block 2 starts to operate (if it has not failed while on standby)
4. Block 2 ultimately fails

After Steps (1)-(4)→**System Fails**

RBDs: Standby system

- **Case I: Two-unit system. Imperfect switching, Standby failures can occur.** Different failure rates for during operation and during standby:

$$R \rightarrow in\ operation \rightarrow \lambda \qquad\qquad R' \longrightarrow in\ standby \longrightarrow \lambda'$$

$$R_s(t) = R_1(t) + \int_0^t \left\{ \underbrace{f_1(t_1)dt_1}_{A} \cdot \underbrace{R_{ss}(t_1)}_{B} \cdot \underbrace{R_2{}'(t_1)}_{C} \cdot \underbrace{R_2(t-t_1)}_{D} \right\}$$

  - $R_1(t)$→Reliability of Item 1 (Probability that it works until mission time t)
  - A → Probability that Item 1 fails at a time $t_1, t_1 < t$ (in a small time interval dt1)
  - B → Reliability of sensing and switching device at t_1(Probability that it works at t_1)
  - C → Probability that Item 2 does not fail while in standby mode (Reliability of Item 2 at time t_1)
  - D → Probability that Item 2 works until mission time t (Reliability of Item 2 from t_1 to t).

RBDs: Standby system

- For **independent components, with exponential dist. of TTF** with $\lambda_1, \lambda_2, \lambda'_2, \lambda_{ss}$

$$R_{sys}(t) = R_1(t) + \int_0^t f_1(t_1) R_{ss}(t_1) R_2'(t_1) R_2(t - t_1) dt_1$$

$$R_{sys}(t) = e^{-\lambda_1 t} + \int_0^t \lambda_1 e^{-\lambda_1 t_1} e^{-\lambda_{ss} t_1} e^{-\lambda_2' t_1} \mathrm{e}^{-\lambda_2 (t - t_1)} dt_1$$

$$R_{sys}(t) = e^{-\lambda_1 t} + \lambda_1 e^{-\lambda_2 t} \int_0^t e^{-\lambda_1 t_1} e^{-\lambda_{ss} t_1} e^{-\lambda_2' t_1} e^{-\lambda_2 t_1} dt_1$$

$$R_{sys}(t) = e^{-\lambda_1 t} + \lambda_1 e^{-\lambda_2 t} \left\{ \frac{-\mathrm{e}^{-[\lambda_1 + \lambda_{ss} + \lambda_2' - \lambda_2] t_1}}{\lambda_1 + \lambda_{ss} + \lambda_2' - \lambda_2} \Bigg|_0^t \right\}$$

$$\boxed{R_{sys}(t) = e^{-\lambda_1 t} + \left[\frac{\lambda_1 e^{-\lambda_2 t}}{\lambda_1 + \lambda_{ss} + \lambda_2' - \lambda_2} \right] \left(1 - e^{-(\lambda_1 + \lambda_{ss} + \lambda_2' - \lambda_2) t}\right)}$$

RBDs: Standby redundant

- **Case II: Perfect switch, no standby failure and iid units with CFR.**
  - This can be reduced to the so-called "***shock model***." In this case, "***shocks***" occur at a constant rate and the system fails after all blocks are failed. Shocks cause failure, the n^{th} shock causes the system to fail.

$$R_s(t) = 1 - \int_0^t \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} dx \qquad [i.e., T_{fail} \sim dgamma\left(n, \frac{1}{\lambda}\right)]$$

  - When n is an integer, this becomes:

$$R_s(t) = e^{-\lambda t}\left[1 + \lambda t + \frac{(\lambda t)^2}{2!} + \frac{(\lambda t)^3}{3!} + \cdots + \frac{(\lambda t)^{n-1}}{(n-1)!}\right]$$

$$MTTF_s = \frac{n}{\lambda}$$

  - Recall: The $Gamma(\alpha, \frac{1}{\lambda})$ distribution is the convolution (the sum) of α exponentially (λ) distributed variables.

Example: Standby system

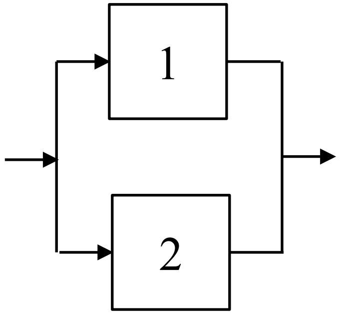
- **Example:** In a two-block standby system with perfect switch and no standby failures, if failure rate of each block is $\lambda = 0.001/hr$, which is the rate of occurrence of shocks, then find the MTTF of the system.

Example: Standby system

- **Solution:** In a two-block standby system with perfect switch and no standby failures, if failure rate of each block is $\lambda_i = 0.001/hr$, which is the rate of occurrence of shocks, then find the MTTF of the system.

$$MTTF_s = \frac{n}{\lambda_i} = \frac{2}{0.001} = 2000\ hours$$

RBDs: Shared load



- **Shared Load System**
  - Initially both items share the load, with identical times to failure distribution being $f_h(t)$. When one item fails, other item operates at a higher stress (i.e., full load) and has increased failure rate, with time to failure distribution $f_f(t)$.

$$R_s(t) = \underbrace{[R_h(t)]^2}_{A} + 2\int_0^t \left\{ \underbrace{f_h(t_1)dt_1}_{B} \cdot \underbrace{R_h(t_1)}_{C} \cdot \underbrace{R_f(t-t_1)}_{D} \right\}$$

  - Probability that:
    - A → Both items remain operational in half load until mission time *t*
    - B → One item fails in half load at $t_1, t_1 < t$
    - C → The other item works at half load until $t_1, t_1 < t$
    - D → The other item works at full load after t_1 until mission time *t*

RBDs: Shared load

- Example of Shared Load System
  - If both items have **exponential time-to-failure** model with failure rates:

$\lambda_h \rightarrow$ half load failure rate $\qquad R_h(t) = e^{-\lambda_h t}$

$\lambda_f \longrightarrow$ full load failure rate $\qquad R_f(t) = e^{-\lambda_f t}$

$$R_{sys}(t) = e^{-2\lambda_h t} + 2\int_0^t \lambda_h e^{-\lambda_h t_1} e^{-\lambda_h t_1} e^{-\lambda_f (t-t_1)} dt_1$$

$$R_{sys}(t) = e^{-2\lambda_h t} + 2\lambda_h e^{-\lambda_f t} \int_0^t \mathrm{e}^{-(2\lambda_h - \lambda_f)t_1} dt_1$$

$$R_{sys}(t) = e^{-2\lambda_h t} + 2\lambda_h e^{-\lambda_f t} \left\{ \frac{\mathrm{e}^{-(2\lambda_h - \lambda_f)t_1}}{-(2\lambda_h - \lambda_f)} \right\} \Bigg|_0^t$$

$$\boxed{R_{sys}(t) = e^{-2\lambda_h t} + \left[\frac{2\lambda_h e^{-\lambda_f t}}{(2\lambda_h - \lambda_f)} \right] \left[1 - e^{-(2\lambda_h - \lambda_f)t} \right]}$$

$$if\ 2\lambda_h - \lambda_f > 0, then\ MTTF = \int_0^\infty R_s(t)dt$$

Example: Shared load

- **Example:** For a two-unit shared load system with $\lambda_h = 0.002, \mathrm{hr}^{-1}$ $\lambda_f = 0.003\ hr^{-1}$
  a) Find the Reliability at t=500 hr
  b) Find $MTTF_s$

Example Shared load: Solution

a) $R_s(t) = e^{-2\lambda_h t} + \left[\frac{2\lambda_h e^{-\lambda_f t}}{(2\lambda_h - \lambda_f)}\right][1 - e^{-(2\lambda_h - \lambda_f)t}]$

- With t=500, $\lambda_h = 0.002, \lambda_f = 0.003$: **R(t=500)**=0.135+(.893)(.393)=**0.487**

b) MTTF

- Recall: $R_s(t) = e^{-2\lambda_h t} + \left[\frac{2\lambda_h e^{-\lambda_f t}}{(2\lambda_h - \lambda_f)}\right][1 - e^{-(2\lambda_h - \lambda_f)t}]$
- Since $(2\lambda_h - \lambda_f > 0)$, we can use $MTTF = \int_0^\infty R_s(t)dt$:

$$MTTF = \int_0^\infty \left\{ e^{-2\lambda_h t} + \left[\frac{2\lambda_h e^{e^{-\lambda_f t}}}{(2\lambda_h - \lambda_f)}\right][1 - e^{-(2\lambda_h - \lambda_f)t}] \right\} dt$$

$$= \frac{-1}{2\lambda_h} e^{-2\lambda_h t} - \frac{2\lambda_h e^{-\lambda_f t}}{\lambda_f(2\lambda_h - \lambda_f)} + \frac{2\lambda_h e^{-2\lambda_h t}}{2\lambda_h(2\lambda_h - \lambda_f)}\Bigg|_0^\infty \quad Since\ e^{-\lambda_f t} e^{-(2\lambda_h - \lambda_f)t} = e^{-2\lambda_h t}$$

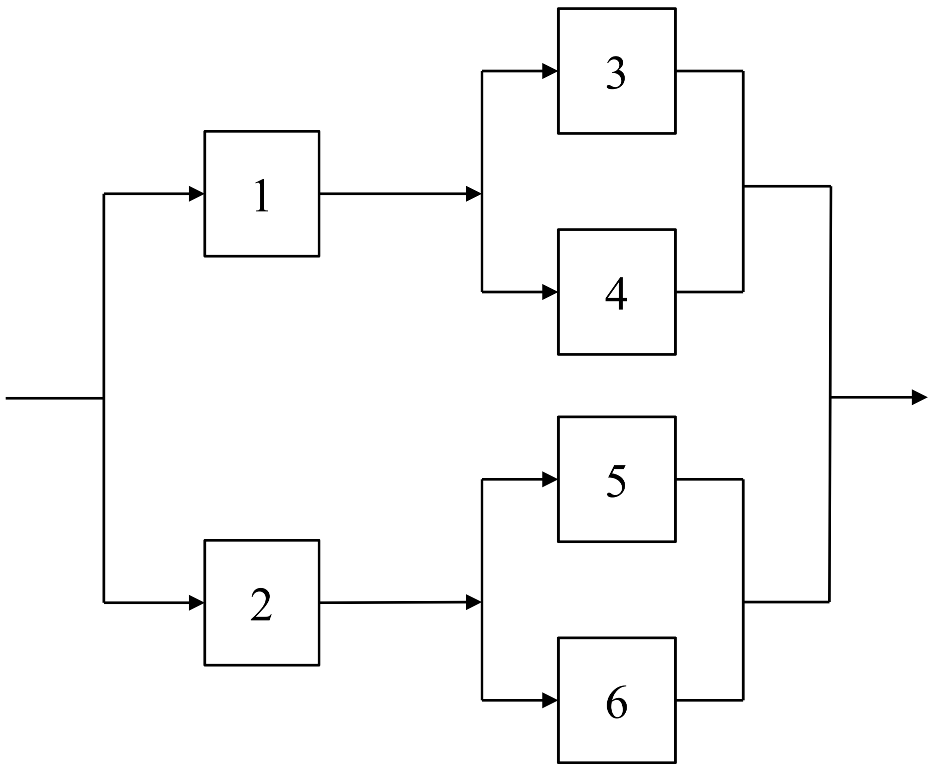
$$= \frac{1}{2\lambda_h} + \frac{2\lambda_h}{\lambda_f(2\lambda_h - \lambda_f)} - \frac{1}{2\lambda_h - \lambda_f}$$

$$MTTF = \frac{1}{2(0.002)} + \frac{2(0.002)}{0.003 + (2(0.002) - 0.003)} - \frac{1}{(2(0.002 - 0.003)}$$

$$= 250 + 1{,}333.33 - 1000 = 583.33\ hrs$$

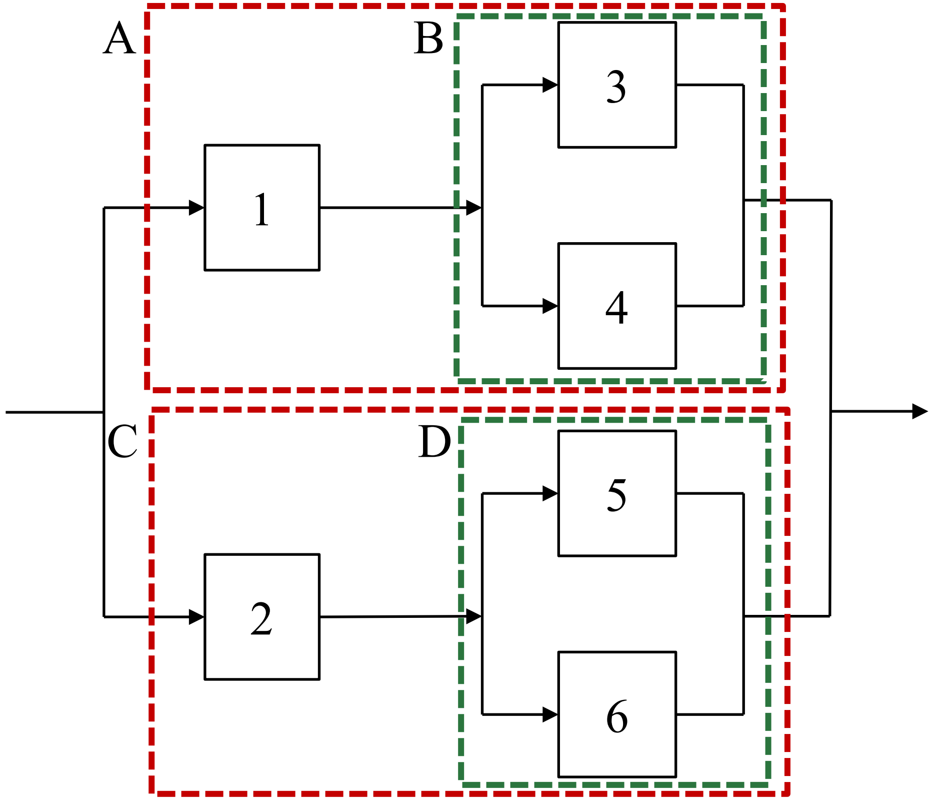
Complex Systems

- Putting it together: Reliability of Parallel-Series Systems.
  - There are some systems which are neither series nor parallel systems, but are some combination of these patterns.



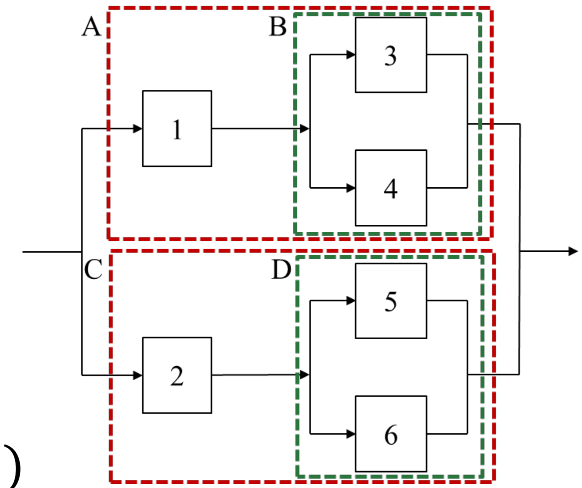
Complex systems: Series, parallel reduction

- **Method 1: Decomposition (Series, parallel reduction)**
  - Decomposing the system into a combination of series, parallel blocks.
  - For the example here: Blocks A, B, C, D.



Complex systems: Series, parallel reduction

- **Example**
  - For the example at right: Blocks are A, B, C, D



$A \parallel C$

$R_S = 1 - (1 - R_A)(1 - R_C) = R_A + R_C - R_A R_C$

$R_A = R_1 R_B = R_1(R_3 + R_4 - R_3 R_4)$

$R_C = R_2 R_D = R_2(R_5 + R_6 - R_5 R_6)$

$$\boldsymbol{R_S = R_1(R_3 + R_4 - R_3R_4) + R_2(R_5 + R_6 - R_5R_6) - R_1R_2(R_3 + R_4 - R_3R_4)(R_5 + R_6 - R_5R_6)}$$

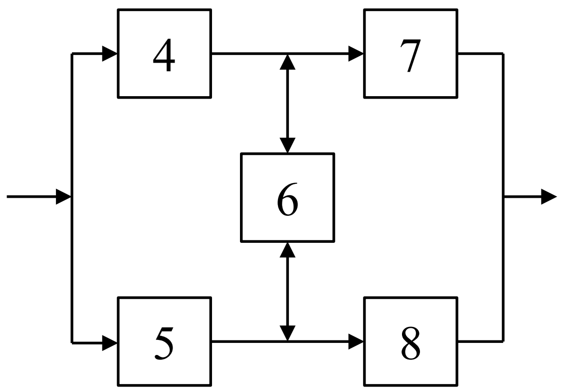
- If $R_i(t) = e^{(-\lambda_i t)}$:

$$R_s(t) = e^{-\lambda_1 t}\left(e^{-\lambda_3 t} + e^{-\lambda_4 t} - e^{-(\lambda_3+\lambda_4)t}\right) + e^{-\lambda_2 t}\left(e^{-\lambda_5 t} + e^{-\lambda_6 t} - e^{-(\lambda_5+\lambda_6)t}\right) - e^{-(\lambda_1+\lambda_2)t}\left[e^{-\lambda_3 t} + e^{-\lambda_4 t} - e^{-(\lambda_3+\lambda_4)t}\right]\left[e^{-\lambda_5 t} + e^{-\lambda_6 t} - e^{-(\lambda_5+\lambda_6)t}\right]$$

$$MTTF = \int_0^\infty R_s(t)dt$$

Complex systems: Decomposition

- **Method 1: Decomposition (Non-Series Parallel Systems)**
  - The system is analyzed by reducing it into parallel-series systems. Conditional probability is used to modify it to parallel-series system.
    - Recall: the law of total probability $\Pr(X) = \Pr(X \cap Y) + \Pr(X \cap \overline{Y})$
    - And the definition: $\Pr(X \cap Y) = \Pr(X|Y)\Pr(Y)$

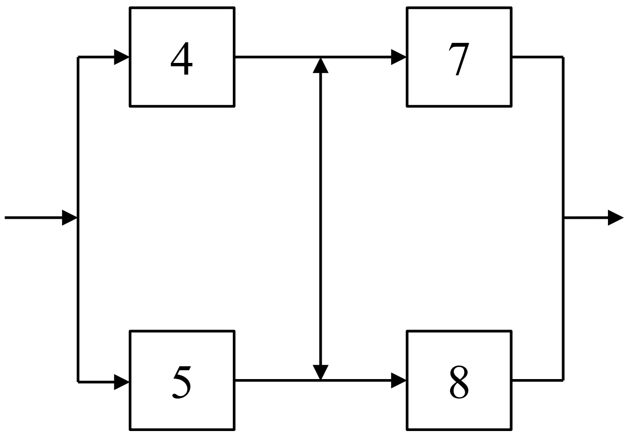


$$R_s(t) = \underbrace{R_s(t|item\ 6\ works)}_{\boldsymbol{A}} \cdot\ R_6(t) + \underbrace{R_s(t|item\ 6\ fails)}_{\boldsymbol{B}} \cdot [1 - R_6(t)]$$

- Item 6 works => $\overline{6}$ or R_6
- Item 6 fail => 6 or F_6

Decomposition

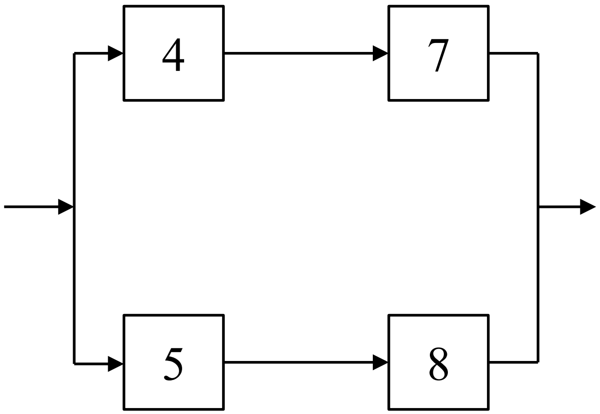
- Quantity A (item 6 works) represents:



- $R_A = R_s(t|\bar{6}) = (R_4 + R_5 - R_4R_5) \cdot (R_7 + R_8 - R_7R_8)$

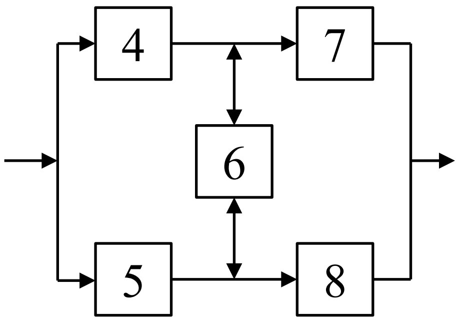
Decomposition (cont.)

- Quantity B (item 6 fails) represents:



- $R_B = R_s(t|6) = R_4R_7 + R_5R_8 - R_4R_5R_7R_8$

Decomposition (cont.)



- Putting it all together:

$$R_s(t) = \underbrace{R_s(t|\bar{6})}_{A} \cdot\ R_6(t) + \underbrace{R_s(t|6)}_{B} \cdot [1 - R_6(t)]$$

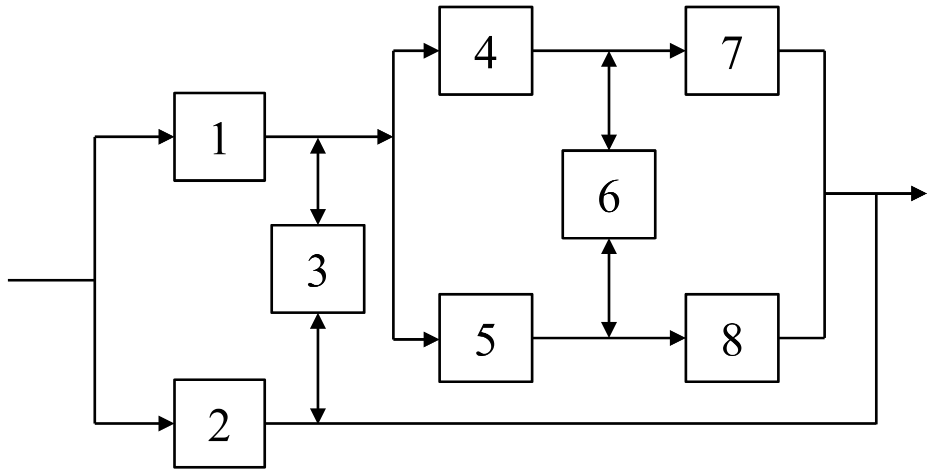
$$\boldsymbol{R_s(t) = (R_4 + R_5 - R_4R_5)(R_7 + R_8 - R_7R_8)R_6 + (R_4R_7 + R_5R_8 - R_4R_5R_7R_8)[1 - R_6]}$$

- If all $R_i(t = 50) = 0.9, R_A = 0.9801$ and $R_B = 0.9639$

$$\boldsymbol{R_S(t = 50) = 0.9801 \cdot 0.9 + 0.9639 \cdot 0.1 = 0.9785}$$

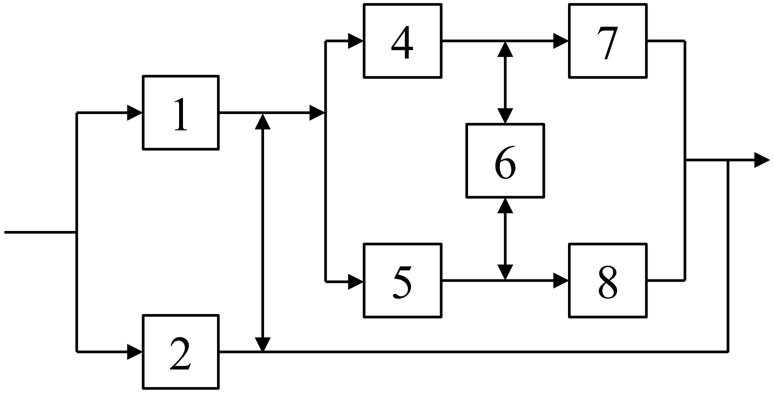
We can do this for increasingly large systems

- **Example (after class):** Conduct decomposition on the RBD below.
  - Hint: Use $R_s(t) = \underbrace{R_s(t|item\ 3\ works)}_{C} \cdot\ R_3(t) + \underbrace{R_s(t|item\ 3\ fails)}_{D} \cdot [1 - R_3(t)]$
  - Hint: Use your solution from our previous example involving 45678

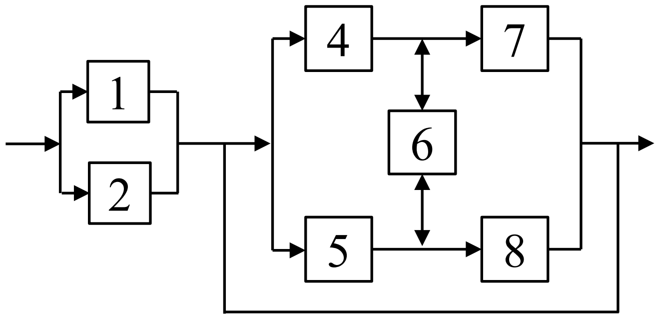


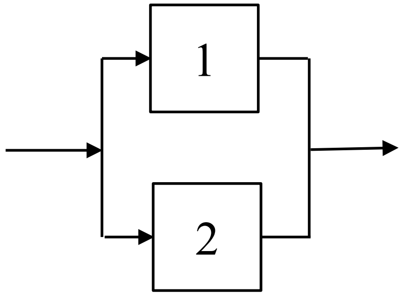
Example Solution: (after class)

- Quantity C (unit 3 good) represents:



- When we redraw, it become evident that this is a simple 2-unit parallel system

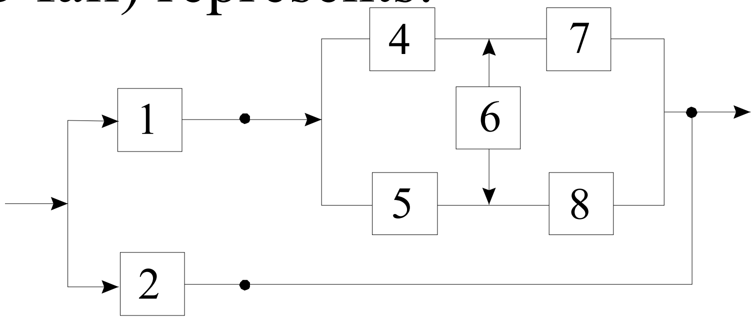




- $R_s(t|\bar{3}) = R_1 + R_2 - R_1 R_2$

Example Solution: (after class)

- Quantity D (item 3 fail) represents:



- We can use the decomposition of S45678 from before.
- Thus, we have:

$$R_D = R_s(t|3) = R_1 R_{s45678} + R_2 - R_1 R_2 R_{s45678}$$

$=R_1((R_4 + R_5 - R_4R_5)(R_7 + R_8 - R_7R_8)R_6 + (R_4R_7 + R_5R_8 - R_4R_5R_7R_8)[1 - R_6]) + R_2 - R_1R_2[(R_4 + R_5 - R_4R_5)(R_7 + R_8 - R_7R_8)R_6 + (R_4R_7 + R_5R_8 - R_4R_5R_7R_8)[1 - R_6]]$

And if we put it all together w/previous slides:

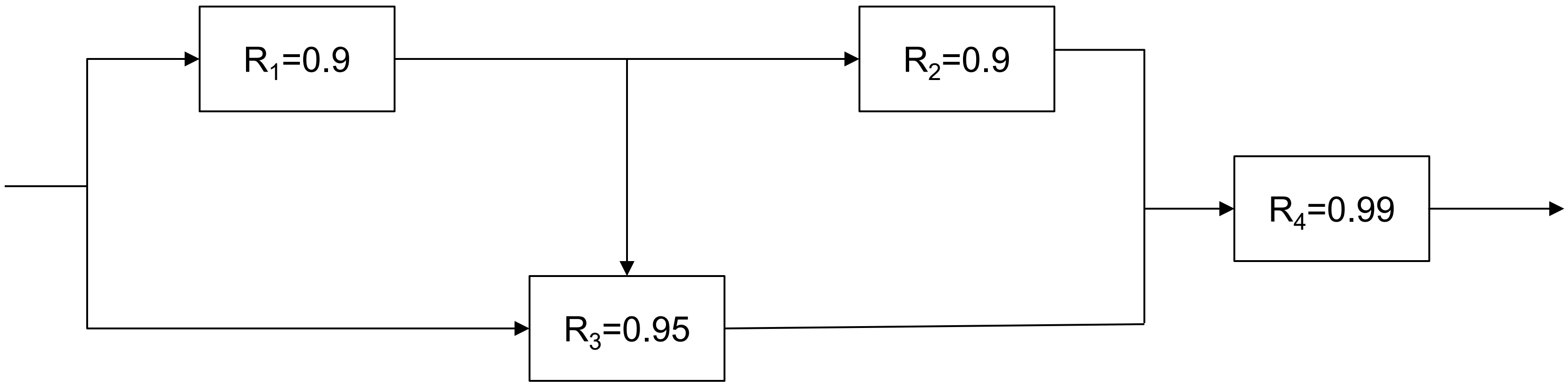
$$\boldsymbol{R_s(t) = [R_1 + R_2 - R_1R_2]R_3 + R_D \cdot [1 - R_3]}$$

Using $R_i(50) = 0.9$ again for all components:

$$\boldsymbol{R_s(50) = 0.99 \cdot 0.9 + (0.9 \cdot 0.9785 + 0.9 - 0.9 \cdot 0.9 \cdot 0.9785) \cdot 0.1 = 0.9899}$$

Extra example (for after class)

- Determine the reliability of the following system using the decomposition method:



Extra example (for after class)

- Solution: Split the system into:
  - $Rs = (R_s | R_3 = 1) * R_3 + (R_s | \bar{R}_3)(1 - R_3)$.

$$R_S = R_3 R_4 + (1 - R_3)(R_1)(R_2)(R_4)$$
$$\boldsymbol{R_S = R_3 R_4 + R_1 R_2 R_4 - R_1 R_2 R_3 R_4}$$
$$= 0.9806$$
$$\boldsymbol{R_S = 0.981}$$

- (Another method of solving this is to realize that the link from R_1 to R_3 is essentially irrelevant, and the two parts of the system can be viewed as ($R_1 R_2$ in parallel with R_3) then in series with R_4.

Complex systems: Path sets & cut sets

- **Complex Systems: Methods 2 & 3: Path sets & Cut sets**
  - **Path sets:** The set of paths (successes) that form a connection from input to output—aka guarantee system operation.
  - **Cut sets:** The set of unit (failures) which interrupt all possible connections between input and output—aka guarantee system failure.
  - **Minimal path- or cut-set:** The minimum number to guarantee connection (or disconnection)
  - Be very careful about your notation. When in doubt, define it at the top of your problem.
  - If X denotes component failure, you should have:

$$X \text{ in a cut set, but } \overline{X} \text{ in a path set}$$

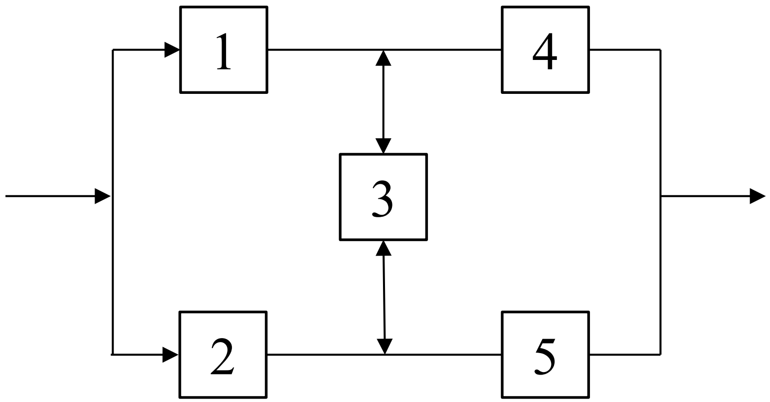
Notes on path sets & cut sets

- Quantification with the cut or path sets uses the *minimal* cut sets of *minimal* path sets.
- **The *order* of the sets** denotes how many terms are in that set.
  - E.g., a cut set which contains two elements (A, B) has order 2.
- *A lot of qualitative insights can be gained from the cut sets before you start quantifying – e.g., whether your system has any single point failures (cut sets of order 1), how many cut sets a certain component appears in, etc.*

Example: Path sets & cut sets

- **Complex Systems: Methods 2 & 3: [Minimal] Path sets & Cut sets**

Minimal Path Sets   Minimal Cut Sets



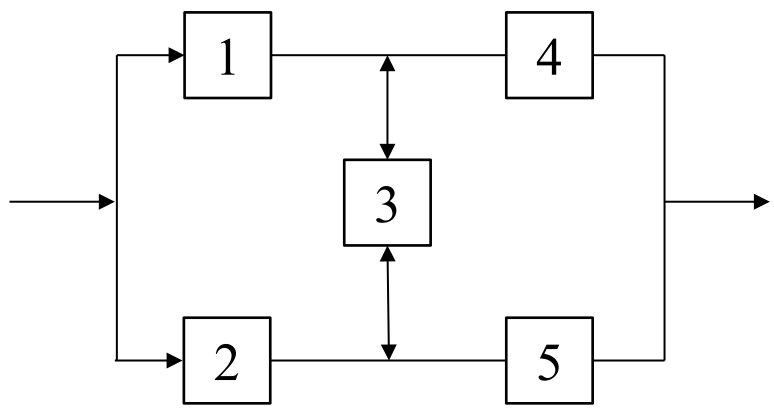
Example: Path sets & cut sets: Solution

- For the system the minimal path sets are:

$$P_1 = \{\overline{1}, \overline{4}\}, P_2 = \{\overline{2}, \overline{5}\}, P_3 = \{\overline{1}, \overline{3}, \overline{5}\}, P_4 = \{\overline{2}, \overline{3}, \overline{4}\}$$

- And the minimal cut sets are:

$$C_1 = \{1,2\}, C_2 = \{4,5\}, C_3 = \{2,3,4\}, C_4 = \{1,3,5\}$$



Path sets: Quantification

- **For path sets**: The system fails when none of the (minimal) path sets, P_i, works.
- Notation:
  - $Pr(\overline{P_i}) = Pr_F(P_i)$ = Probability that min. path i (P_i) is not successful (not available)
  - $Pr(P_i) = Pr_S(P_i)$= Probability that min. path i (P_i) is successful (available)

$$\boxed{\begin{gathered}\boldsymbol{R_s = \mathbf{Pr}(P_1 \cup P_2 \cup P_3 \cup \cdots \cup P_m)} \\ R_s = 1 - Pr(\overline{P_1} \cap \overline{P_2} \cap \overline{P_3} \cap \overline{P_4})\end{gathered}}$$

  - **If all min. path sets are mutually exclusive (which usually isn't true):**

$$R_s = \Pr(P_1) + \Pr(P_2) + \Pr(P_3) + \Pr(P_4)$$

  - **If they're not, we can treat this as an upper bound on reliability:**

$$R_s \leq \Pr(P_1) + \Pr(P_2) + \Pr(P_3) + \Pr(P_4)$$

  - **If the min. path sets are independent, this becomes:**

$$R_s = 1 - \Pr(\overline{P_1})\Pr(\overline{P_2})\Pr(\overline{P_3})\Pr(\overline{P_4})$$

$$R_S = 1 - [1 - \Pr(P_1)][1 - \Pr(P_2)][1 - \Pr(P_3)][1 - \Pr(P_4)]$$

  - **If path sets are not independent, this is an upper bound on reliability.**

Cut sets: Quantification

- **For cut sets**: The system fails if one of the minimal cut sets occurs
- Notation
  - $\Pr(C_i)$ = Probability that min. cut set i (C_i) occurs
  - Therefore,

$$\boxed{\begin{aligned} F_s(t) &= \Pr(C_1 \cup C_2 \cup C_3 \cup C_4) \\ R_s(t) &= 1 - \Pr(C_1 \cup C_2 \cup C_3 \cup C_4) \end{aligned}}$$

  - If the cut sets are mutually exclusive (…not usually true):

$$R_s(t) = 1 - (\Pr(C_1) + \Pr(C_2) + \Pr(C_3) + \Pr(C_4))$$

  - If they're not:

$$R_s(t) \geq 1 - (\Pr(C_1) + \Pr(C_2) + \Pr(C_3) + \Pr(C_4))$$

Rare event approximation

- The **rare event approximation** allows us to eliminate some terms to significantly simplify the math.
- Recall that probability of the union of two events (given by addition law of probability) is:

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

$$\text{If } A \perp B \text{, this becomes} = \Pr(A) + \Pr(B) - \Pr(A)\Pr(B)$$

- Or in compact form: $\Pr(A \cup B) = (1 - [1 - \Pr(A)][1 - \Pr(B)])$
- If we have *rare events*, the term $\Pr(A)\Pr(B)$ is very small.

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \underbrace{\Pr(A)\Pr(B)}_{\substack{eliminate\ this \\ term\ for\ rare \\ events.}}$$

$$\boxed{\Pr(A \cup B) \approx \Pr(A) + \Pr(B)}$$

- **General rule for "rare":** $\Pr(E_i) < \frac{1}{50n}\ where\ n = \#\ of\ events$

System reliability analysis (cont.)

- When **rare event approximation** is applied (or if we assume the error from assuming mutual exclusivity of sets is small):

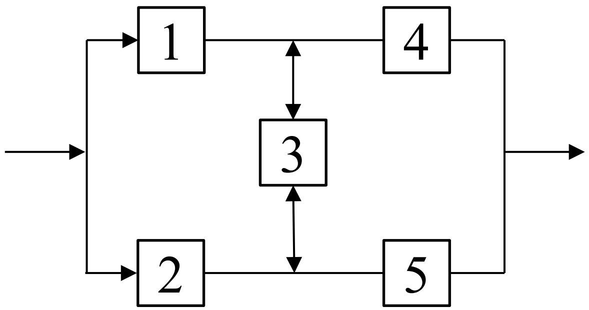
$$R_s^{cutset} < R_s < R_s^{pathset}$$

- Therefore, cut sets give the lower bound of reliability, whereas path sets give the upper bound of reliability.

Example Part 2: Path sets and cut sets

- **Example:** Use both the cut set and path set methods to find the upper and lower bounds for reliability for the system at time *t*, for the system with minimal path sets $P_1 = \{\overline{1}, \overline{4}\}, P_2 = \{\overline{2}, \overline{5}\}, P_3 = \{\overline{1}, \overline{3}, \overline{5}\}, P_4 = \{\overline{2}, \overline{3}, \overline{4}\}$ and minimal cut sets $C_1 = \{1, 2\}, C_2 = \{4, 5\}, C_3 = \{2, 3, 4\}, C_4 = \{1, 3, 5\}$

Assume the units are independent of each other, $1 \perp 2 \perp 3 \perp 4 \perp 5$



Example Part 2: Path sets and cut sets

- **Solution:** Using the cut sets $C_1 = \{1,2\}, C_2 = \{4,5\}, C_3 = \{2,3,4\}, C_4 = \{1,3,5\}$

$$R_s(t) = 1 - \Pr(C_1 \cup C_2 \cup C_3 \cup C_4)$$

$$R_s(t) \geq 1 - \left[\underbrace{Pr_f(C_1)}_{(1-R_1)(1-R_2)}\right] + \left[\underbrace{Pr_f(C_2)}_{(1-R_4)(1-R_5)}\right] + \left[\underbrace{Pr_f(C_3)}_{(1-R_2)(1-R_3)(1-R_4)}\right] + \left[\underbrace{Pr_f(C_4)}_{(1-R_1)(1-R_3)(1-R_5)}\right]$$

$$R_s(t) \geq 1 - \begin{bmatrix} (1-R_1)(1-R_2) + \\ (1-R_4)(1-R_5) + \\ (1-R_2)(1-R_3)(1-R_4) + \\ (1-R_1)(1-R_3)(1-R_5) \end{bmatrix}$$

Example Part 2: Path sets & cut sets

- **Solution:** Using the path sets$P_1 = \{\overline{1}, \overline{4}\}, P_2 = \{\overline{2}, \overline{5}\},\ \ P_3 = \{\overline{1}, \overline{3}, \overline{5}\}, P_4 = \{\overline{2}, \overline{3}, \overline{4}\}$

$$R_s \leq 1 - Pr_f(\overline{P_1})Pr_f(\overline{P_2})Pr_f(\overline{P_3})Pr_f(\overline{P_4})$$

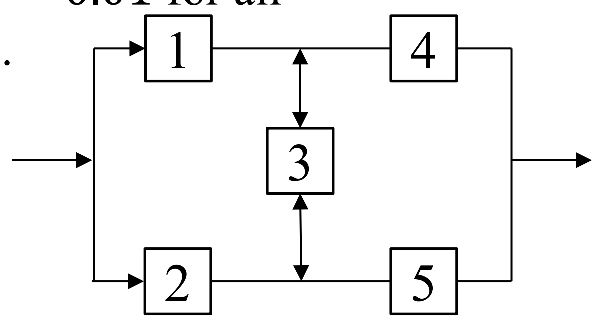
$$R_s \leq 1 - [1 - Pr_s(P_1)][1 - Pr_s(P_2)][1 - Pr_s(P_3)][1 - Pr_s(P_4)]$$

$$\boldsymbol{R_s(t) \leq 1 - [1 - R_1(t)R_4(t)][1 - R_2(t)R_5(t)][1 - R_1(t)R_3(t)R_5(t)]}$$
$$\boldsymbol{[1 - R_2(t)R_3(t)R_4(t)]}$$

Truth Tables

Method 4:

- List every possible combination of component states & system failures.
  - **Pros:** Mutually exclusive!
  - **Cons:** This gets long. 2^n combinations for a system with n components and binary (2) states.
- A truth table for our example is shown at right with $\Pr(F_i) = 0.01$ for all components.



Seq. Number	1	2	3	4	5	Combination	System Operation	Probability
1	S	S	S	S	S	$R_1R_2R_3R_4R_5$	S	0.59019
2	F	S	S	S	S	$F_1R_2R_3R_4R_5$	S	0.06531
3	S	F	S	S	S	$R_1F_2R_3R_4R_5$	S	0.06561
4	S	S	F	S	S	$R_1R_2F_3R_4R_5$	S	0.06561
5	S	S	S	F	S	$R_1R_2R_3F_4R_5$	S	0.06561
6	S	S	S	S	F	$R_1R_2R_3R_4F_5$	S	0.06561
7	F	F	S	S	S	$F_1F_2R_3R_4R_5$	F	0.00729
8	F	S	F	S	S	$F_1R_2F_3R_4R_5$	S	0.00729
9	F	S	S	F	S	$F_1R_2R_3F_4R_5$	S	0.00729
10	F	S	S	S	F	$F_1R_2R_3R_4F_5$	S	0.00729
11	S	F	F	S	S	$R_1F_2F_3R_4R_5$	S	0.00729
12	S	F	S	F	S	$R_1F_2R_3F_4R_5$	S	0.00729
13	S	F	S	S	F	$R_1F_2R_3R_4F_5$	S	0.00729
14	S	S	F	F	S	$R_1R_2F_3F_4R_5$	S	0.00729
15	S	S	F	S	F	$R_1R_2F_3R_4F_5$	S	0.00729
16	S	S	S	F	F	$R_1R_2R_3F_4F_5$	F	0.00729
17	F	F	F	S	S	$F_1F_2F_3R_4R_5$	F	8.10E-04
18	F	F	S	F	S	$F_1F_2R_3F_4R_5$	F	8.10E-04
19	F	F	S	S	F	$F_1F_2R_3R_4F_5$	F	8.10E-04
20	F	S	F	F	S	$F_1R_2F_3F_4R_5$	S	8.10E-04
21	F	S	F	S	F	$F_1R_2F_3R_4F_5$	F	8.10E-04
22	F	S	S	F	F	$F_1R_2R_3F_4F_5$	F	8.10E-04
23	S	F	F	F	S	$R_1F_2F_3F_4R_5$	F	8.10E-04
24	S	F	F	S	F	$R_1F_2F_3R_4F_5$	S	8.10E-04
25	S	F	S	F	F	$R_1F_2R_3F_4F_5$	F	8.10E-04
26	S	S	F	F	F	$R_1R_2F_3F_4F_5$	F	8.10E-04
27	F	F	F	F	S	$F_1F_2F_3F_4R_5$	F	9.00E-05
28	F	F	F	S	F	$F_1F_2F_3R_4F_5$	F	9.00E-05
29	F	F	S	F	F	$F_1F_2R_3F_4F_5$	F	9.00E-05
30	F	S	F	F	F	$F_1R_2F_3F_4F_5$	F	9.00E-05
31	S	F	F	F	F	$R_1F_2F_3F_4F_5$	F	9.00E-05
32	F	F	F	F	F	$F_1F_2F_3F_4F_5$	F	1.00E-05

$$\sum Pr = \Pr(S) + \Pr(F) = 1.00$$
$$\Pr(F) = 2.15 \times 10^{-2}$$

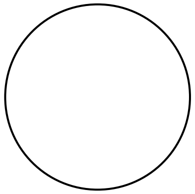
Logic trees

- **Fault Trees (FT) & Success Trees, Event Trees (ET)**
  - RBDs become very complex to use when number of units increases; this motivates the use of other types of models like FTs and ETs.
  - FT Analysis (FTA) is a deductive method – deducing how component failures contribute to a top event.
  - Approach: Top-down decomposition of the logic of a system (failure or success)
    - Many equivalent trees are possible
  - Only includes important events – not all postulated failures included.
    - What is important is a function of industry procedures and assumptions, system design, available data, analysis choices…
  - The FT is a qualitative model of the system logic, but we use quantitative algorithms to evaluate it (e.g., cut set method).

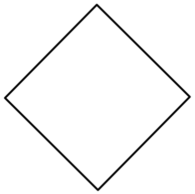
Fault tree analysis steps

1. Define the system to be analyzed & its boundaries
2. Define the top event
3. Construct the fault tree
4. Perform qualitative evaluation of fault tree (logic evaluation)
5. Perform quantitative evaluation of the FT (compute probability of top event)

Logic trees: event symbols

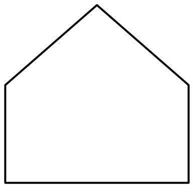


Basic Event



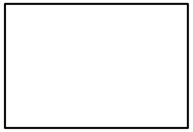
Undeveloped Event

Not developed further because it is of insufficient consequence, or because information is not available.



External Event

External but normally expected to occur.



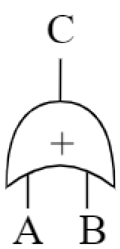
Intermediate Event

Provided for explanation only.

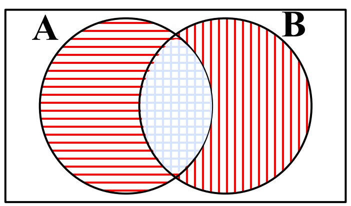
Logic tree analysis: Gates (1)

- Top-down deductive decomposition of a failure into basic causes of failure using Boolean Logic

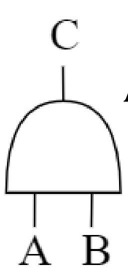
Gates (Logic Representation) in Logic Trees



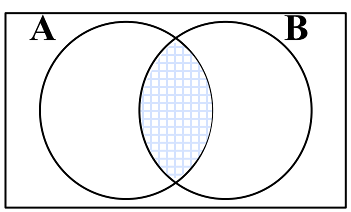
OR gate



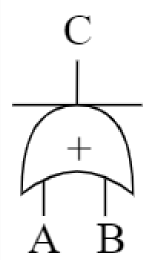
Union Operation
$C = A \cup B$
$C = A + B$



AND gate

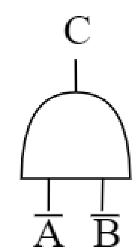


Intersection Operation
$C = A \cap B$
$C = A \cdot B$

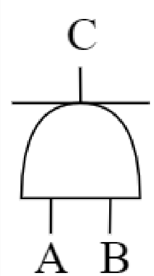


NOT OR gate
NOR

$C = \overline{A \cup B}$
$C = \bar{A} \cap \bar{B}$
$C = \bar{A} \cdot \bar{B}$

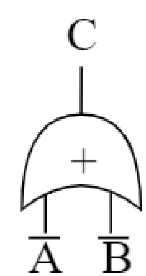


Input 1	Input 2	Output
		1
	1	
1		
1	1	



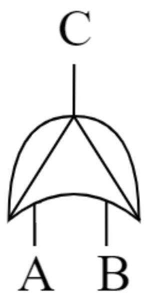
NOT AND gate
NAND

$C = \overline{A \cap B}$
$C = \bar{A} \cup \bar{B}$
$C = \bar{A} + \bar{B}$

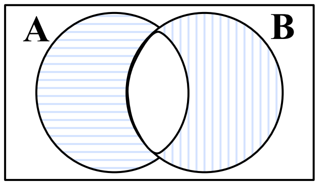


Input 1	Input 2	Output
		1
	1	1
1		1
1	1	

Logic tree analysis: Gates (2)



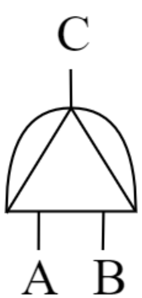
Exclusive OR gate



$C = (A \cap \bar{B}) \cup (\bar{A} \cap B)$

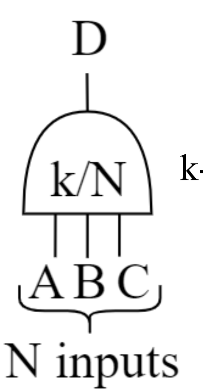
$C = (A \cdot \bar{B}) + (\bar{A} \cdot B)$

Input 1	Input 2	Output
	1	1
1		1
1	1	



Priority AND gate

Output C occurs if all of the inputs occur in a specific order (the order usually indicated by a conditioning event): A occurs first, and then B occur



k-out-of-N gate

If any k combination of the inputs occur, then the output will occur

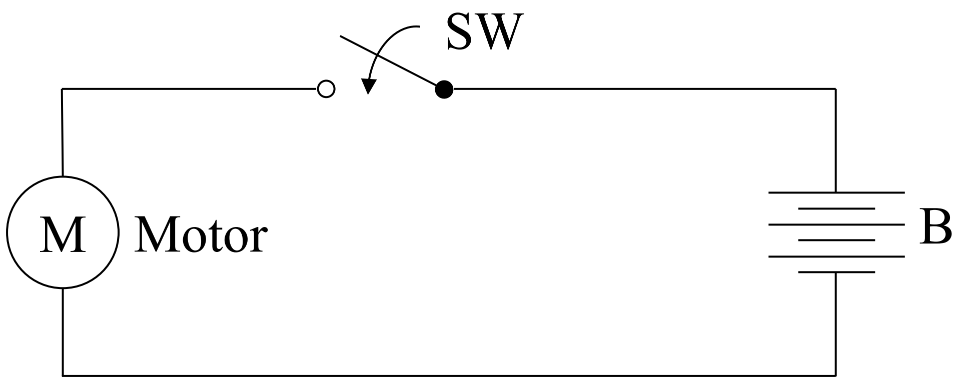
For a k=2, N=3 2-out-of-3 gate:

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

$D = (A \cdot B) + (A \cdot C) + (B \cdot C)$

Example: Fault tree analysis

- **Example**: A motor system, wherein a motor is powered by a battery, connected by a switch. Assume independence of the components.

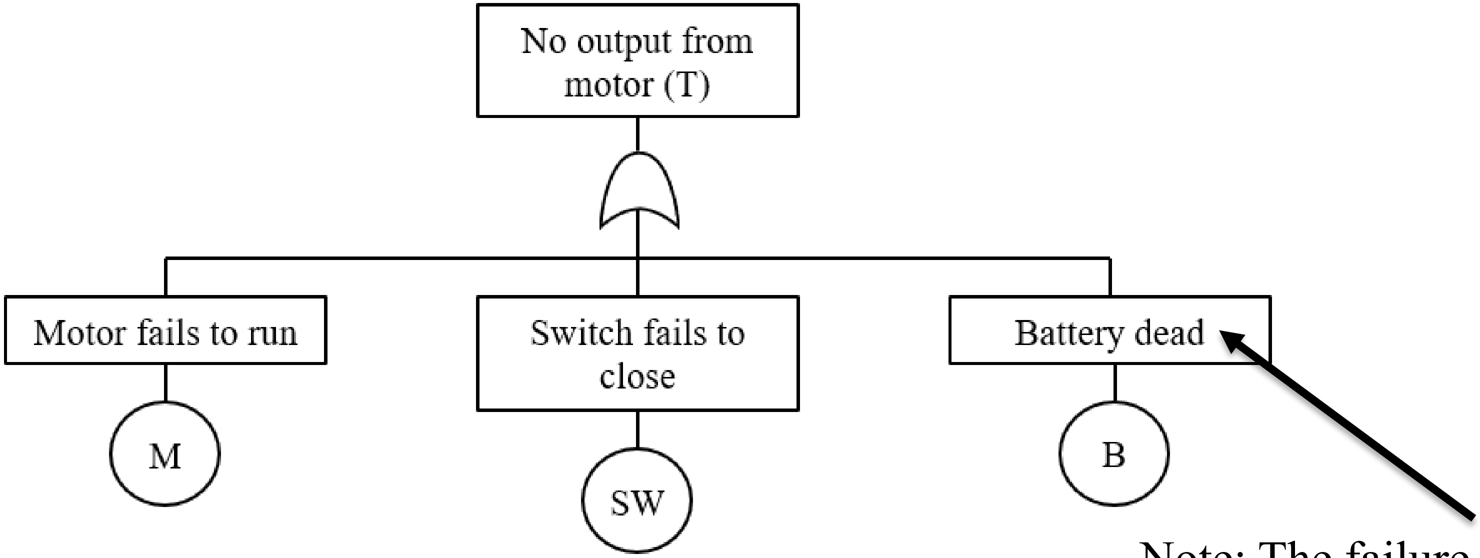


The system can be represented as a series model in a block diagram form

- **Part 1:** Draw a fault tree for the event "No output from motor system."
- **Part 2:** Use the successive substitution method to get the cut sets. Use both the exact method and the cut set method (with rare event approximation) to determine the expression for the top event probability.

Example: Fault tree analysis

- **Solution (1)**:
  A fault tree for the event "No output from motor system" will be:



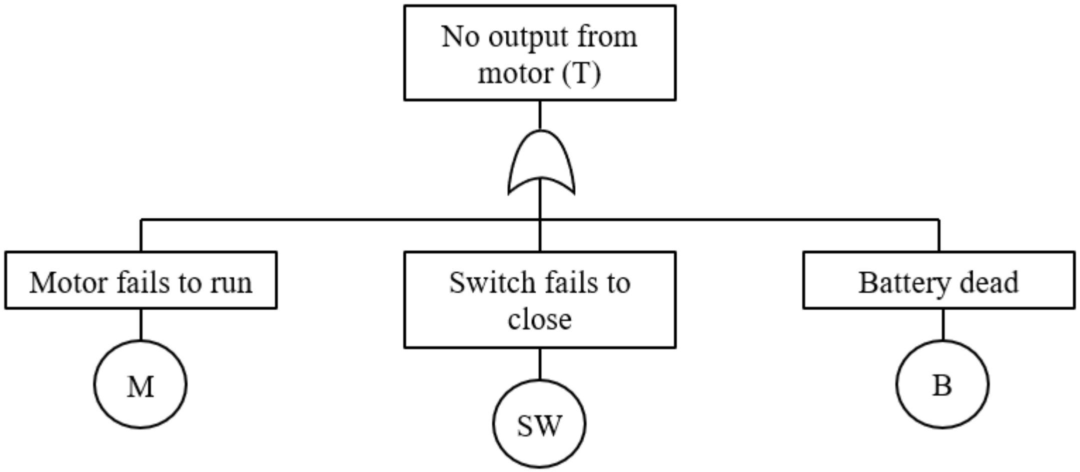
Note: The failure modes matter here!

FT successive substitution

- **Successive Substitution Method**: Boolean Reduction Process of a FT.
  - Reminder: Boolean reduction table on page 31.
  - **Step 1:** Top-down Boolean substitution of the gates.
  - **Step 2:** Boolean reduction of the final expression to reach minimal cut sets.
  - **Step 3:** Quantification.

FTA example (cont.): Qualitative solution

- **Solution (2):** Now we can evaluate the logic of this tree using Boolean substitution:



$$System\ failure\ logic, F_s = T = B \cup M \cup SW$$

Note: This Boolean expression represents the cut sets of the fault tree.

FTA example: Quantitative evaluation

- **Solution (3):** Next, we focus on the quantification aspects of this FT:

$$For\ T: B \cup M \cup SW$$

$$\boldsymbol{Unreliability, F_s} = \mathbf{Pr}(\boldsymbol{T}) = \mathbf{Pr}(\boldsymbol{B} \cup \boldsymbol{M} \cup \boldsymbol{SW})$$

To calculate this probability, we use one of the methods discussed previously: cut sets, exact expression, truth table, etc.

FTA example: Quantitative evaluation

- **Solution (4):**

$$F_S = \Pr(T) = \Pr(B \cup M \cup SW)$$

- **The exact calculation is:**

$$\Pr(T) \\ = \Pr(B) + \Pr(M) + \Pr(SW) - \Pr(B, M) - \Pr(B, SW) - \Pr(M, SW) \\ + \Pr(B, M, SW)$$

- **Since the components are independent this becomes:**

$$\Pr(T) \\ = \Pr(B) + \Pr(M) + \Pr(SW) - \Pr(B)\Pr(M) - \Pr(B)\Pr(SW) \\ - \Pr(M)\Pr(Sw) + \Pr(M)\Pr(B)\Pr(Sw).$$

Or, written in more compact form:

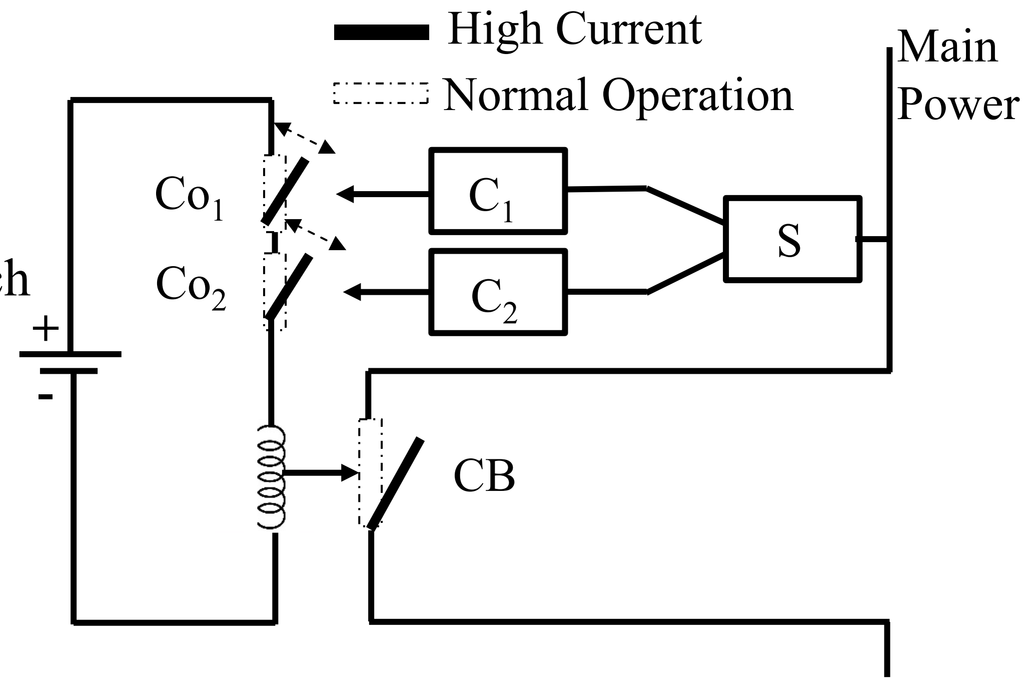
$$\Pr(T) = 1 - (1 - \Pr(B))(1 - \Pr(M))(1 - \Pr(SW))$$

Or, an approximation can be obtained with the cut set method with the rare event approximation:

$$\mathbf{Pr}(\boldsymbol{T}) \cong \mathbf{Pr}(\boldsymbol{B}) + \mathbf{Pr}(\boldsymbol{M}) + \mathbf{Pr}(\boldsymbol{SW})$$

FTA: CB system example

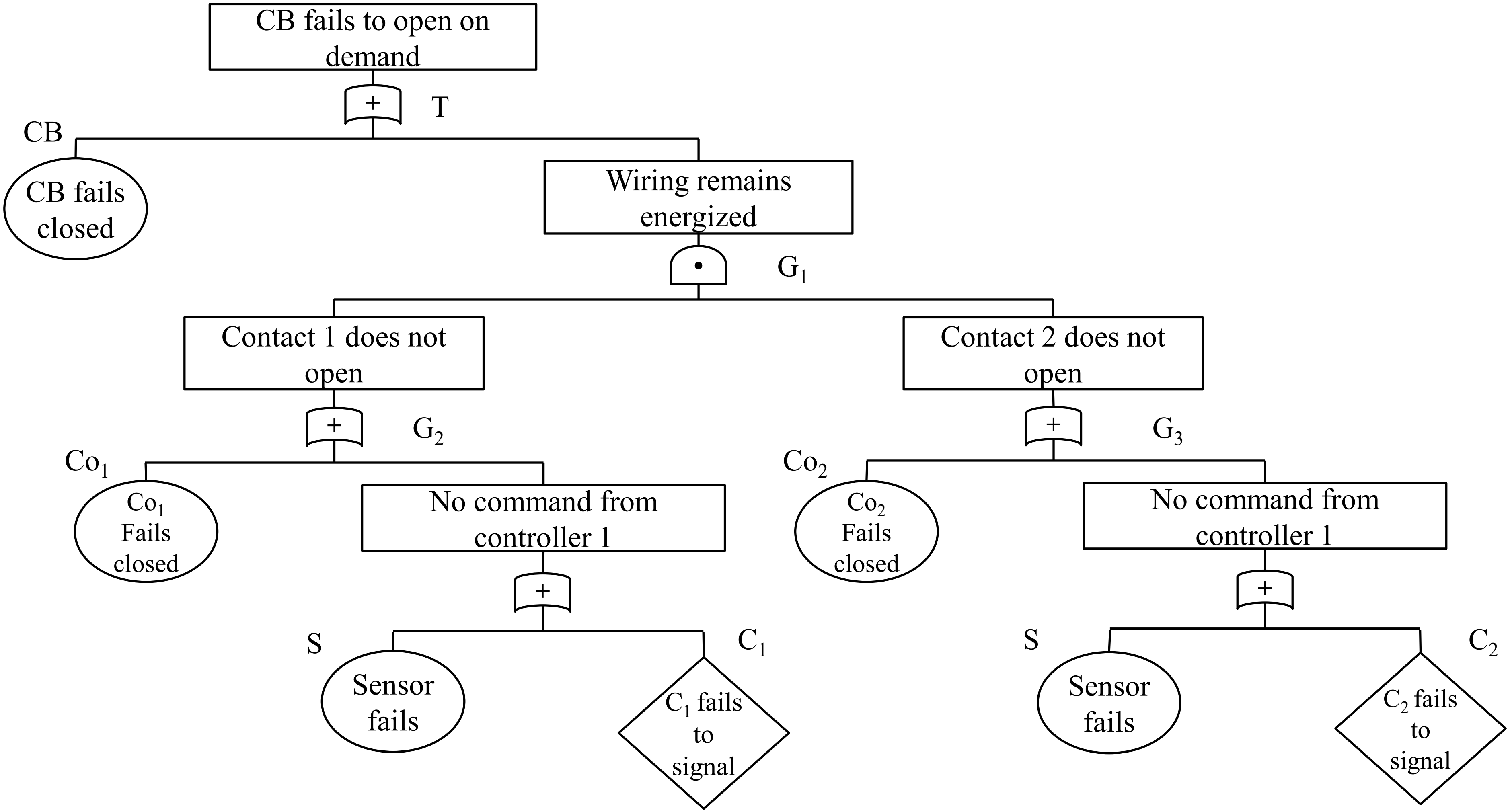
- **Example:** Circuit breaker System
  - C_1 and C_2 = Controller
  - S = Sensor
  - Co_1 and Co_2 = Contacts
  - CB = Circuit Breaker Switch



- **Functions:**
  - If the current gets too high in the main power line, the circuit breaker switch (CB) should open to prevent a short circuit.
  - Sensor S detects high current. If S indicates high current, the controllers issue a command to open contacts (Co_1, Co_2). If any contact is open, the power to the coil is removed and CB opens.

CB system example

- **CB Fault tree**



CB system example (cont.)

- Boolean Substitution Process:

$$T = CB \cup G_1 = CB + G_1$$
$$G_1 = G_2 \cdot G_3$$
$$G_2 = S + Co_1 + C_1$$
$$G_3 = S + Co_2 + C_2$$
$$G_1 = (S + Co_1 + C_1) \cdot (S + Co_2 + C_2)$$
$$T = \mathrm{CB} + \underbrace{(\mathrm{S} \cdot \mathrm{S})}_{\mathrm{S}} + \mathrm{S} \cdot \mathrm{Co}_2 + \mathrm{S} \cdot \mathrm{C}_2 + \mathrm{Co}_1 \cdot \mathrm{S} + \mathrm{Co}_1 \cdot \mathrm{Co}_2 + \mathrm{Co}_1 \cdot \mathrm{C}_2 + \mathrm{C}_1 \cdot \mathrm{S} + \mathrm{C}_1 \cdot \mathrm{Co}_2 + \mathrm{C}_1 \cdot \mathrm{C}_2$$

- Applying Boolean reduction on this expression gives the minimal cut sets:

$$T = \mathrm{CB} + \underbrace{(\mathrm{S} \cdot \mathrm{S})}_{\mathrm{S}} + \cancel{\mathrm{S} \cdot \mathrm{Co}_2} + \cancel{\mathrm{S} \cdot \mathrm{C}_2} + \cancel{\mathrm{Co}_1 \cdot \mathrm{S}} + \mathrm{Co}_1 \cdot \mathrm{Co}_2 + \mathrm{Co}_1 \cdot \mathrm{C}_2 + \cancel{\mathrm{C}_1 \cdot \mathrm{S}} + \mathrm{C}_1 \cdot \mathrm{Co}_2 + \mathrm{C}_1 \cdot \mathrm{C}_2$$

$$\mathbf{T = CB + S + Co_1 \cdot Co_2 + Co_1 \cdot C_2 + C_1 \cdot Co_2 + C_1 \cdot C_2}$$

Fault tree analysis (cont.)

- Quantifying via minimal cut set approach (and assuming indep.) gives:

$$\mathbf{Pr(T)} \approx \mathbf{Pr(CB)} + \mathbf{Pr(S)} + \mathbf{Pr}(\mathbf{Co_1}) \cdot \mathbf{Pr}(\mathbf{Co_2}) + \mathbf{Pr}(\mathbf{Co_1}) \cdot \mathbf{Pr}(\mathbf{C_2}) + \mathbf{Pr}(\mathbf{C_1}) \cdot \mathbf{Pr}(\mathbf{Co_2}) + \mathbf{Pr}(\mathbf{C_1}) \cdot \mathbf{Pr}(\mathbf{C_2})$$

- If we assume that unreliability for each component is 0.01, then:

$$\mathbf{Pr(T)} \approx \mathbf{2} \times \mathbf{.01} + \mathbf{4} \times (\mathbf{.01} \times \mathbf{.01}) \approx \mathbf{0.0204}$$

Combinatorial approach (aka Truth Table) for the same FT

- A more accurate quantification approach
- Finds all combinations (mutually exclusive) that cause top event to occur
  - 2^6 combinations exist = 64

$$C_1^6 + C_2^6 + C_3^6 + C_4^6 + C_5^6 + C_6^6 = 2^6$$

- Each of these elements are the (non-minimal) cut sets of the Systems (you can test to verify)

Local Failure	System State	Local Failure	System State	Local Failure	System State	Local Failure	System State
CB^1	F	CB S	F	S Co_1	F	Co_1 Co_2	F
S	F	CB Co_1	F	S C_1	F	Co_1 C_2	F
Co_1	S	CB C_1	F	S Co_2	F	C_1 Co_2	F
C_1	S	CB Co_2	F	S C_2	F	C_1 C_2	F
Co_2	S	CB C_2	F	Co_1 C_1	S	Co_2 C_2	S
C_2	S						

$$CB^1 = CB \cdot \overline{S} \cdot \overline{Co_1} \cdot \overline{Co_2} \cdot \overline{C_1} \cdot \overline{C_2}$$

$$\Pr(Mutually\ Exclusive\ Set\ 1) = \Pr(CB) \cdot \Pr(\overline{S}) \cdot \Pr(\overline{Co_1}) \cdot \Pr(\overline{Co_2}) \cdot \Pr(\overline{C_1}) \cdot \Pr(\overline{C_2})$$

$$= 0.01 \times 0.99^5\ (if\ all\ failure\ probabilities\ are\ 0.01)$$

$$= 0.0095099$$

- Accounting for all 64 rows, you should get a value of: $\Pr(F_S) = 0.020292$

FT vs. BDD vs. Truth Table

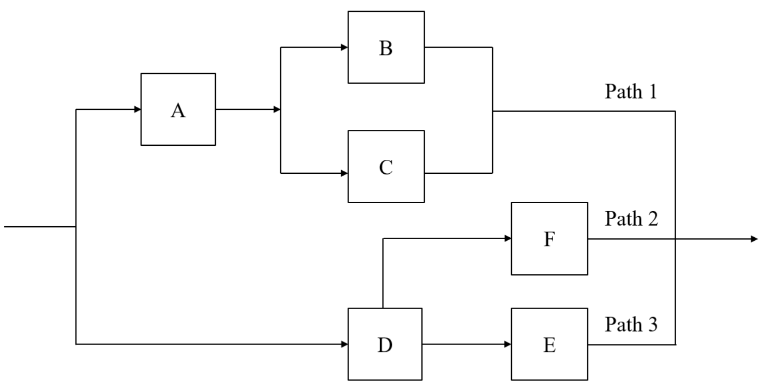
- All methods list combinations of events that cause system failure.

FT Boolean reduction	BDD	Truth Table
AD, BCD, AEF, BCEF	AD, $\bar{A}BCD$, $A\bar{D}EF$, $\bar{A}\bar{D}BCEF$	$A\bar{B}\bar{C}D\bar{E}\bar{F}, AB\bar{C}D\bar{E}\bar{F}, A\bar{B}CD\bar{E}\bar{F}, A\bar{B}\bar{C}DE\bar{F}, A\bar{B}\bar{C}D\bar{E}F,$ $A\bar{B}\bar{C}\bar{D}EF, \bar{A}BCD\bar{E}\bar{F}, ABCD\bar{E}\bar{F}, AB\bar{C}DE\bar{F},$ $A\bar{B}CDE\bar{F}, \bar{A}BCDE\bar{F}, AB\bar{C}D\bar{E}F, A\bar{B}CD\bar{E}F,$ $\bar{A}BCD\bar{E}F, AB\bar{C}\bar{D}EF, A\bar{B}C\bar{D}EF, \bar{A}BC\bar{D}EF, A\bar{B}\bar{C}DEF,$ $ABCDE\bar{F}, ABCD\bar{E}F, ABC\bar{D}EF, AB\bar{C}DEF,$ $A\bar{B}CDEF, \bar{A}BCDEF, ABCDEF$

Minimal cut sets

Mutually exclusive minimal cut sets

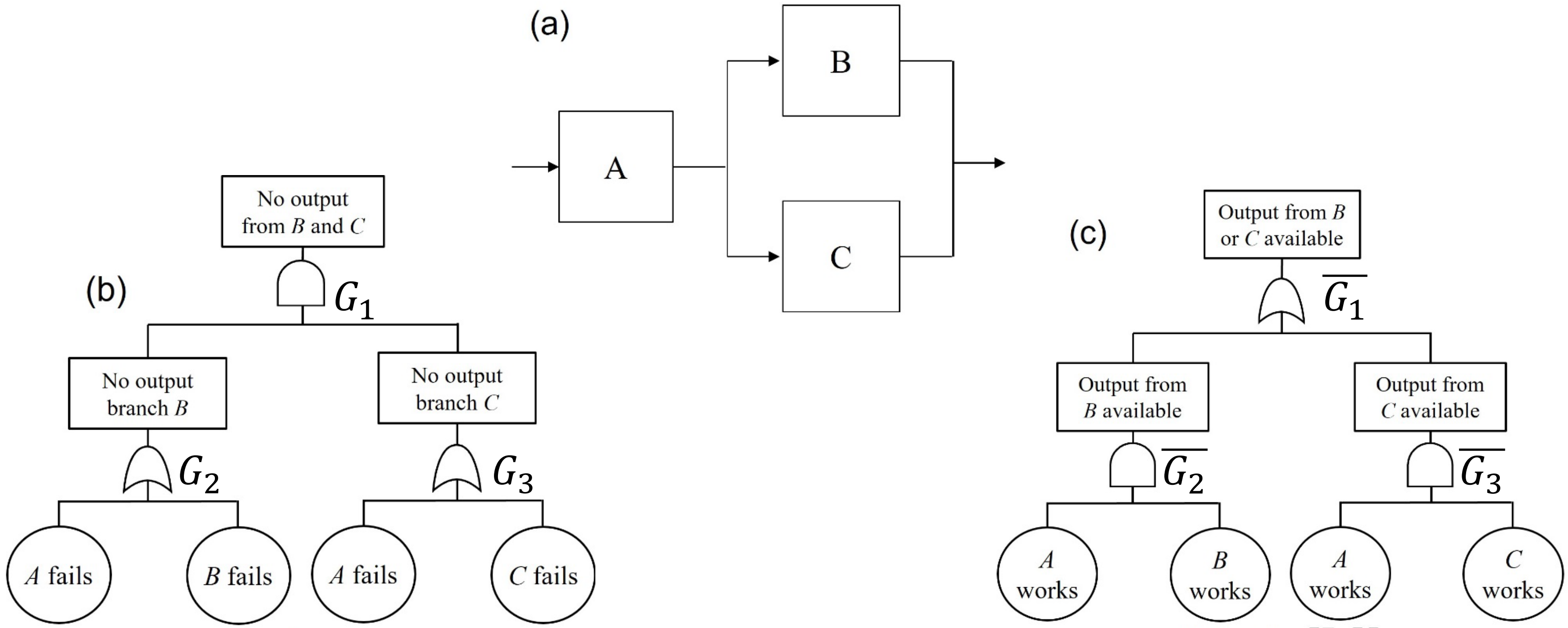
All cut sets (mutually exclusive)



Results correspond to RBD at right

Comparison of success trees and fault trees

- Consider the system:



Cut sets: A, BC

Path sets: $\bar{A}\bar{B}, \bar{A}\bar{C}$

i. Fault tree, cut sets are:

$$T = G_1 = G_2 \cdot G_3$$

$$T = (A + B) \cdot (A + C)$$

$$= A + B \cdot C$$

ii. For equivalent Success tree, path sets are:

$$\bar{T} = \overline{G_1} = \overline{G_2} + \overline{G_3}$$

$$\bar{T} = \overline{A + B} + \overline{A + C} = \bar{A} \cdot \bar{B} + \bar{A} \cdot \bar{C}$$

$$= \bar{A} \cdot \bar{B} + \bar{A} \cdot \bar{C}$$

Comparison of success tree and fault tree (cont.)

- You can use Boolean Laws to find path sets of the Success Tree from cut sets of Fault Tree:

$$\bar{T} = \left(\overline{A + B \cdot C}\right) = \bar{A} \cdot \overline{B \cdot C} = \bar{A} \cdot (\bar{B} + \bar{C})$$

$$\bar{T} = \bar{A} \cdot \bar{B} + \bar{A} \cdot \bar{C} \text{ same as (ii)}$$

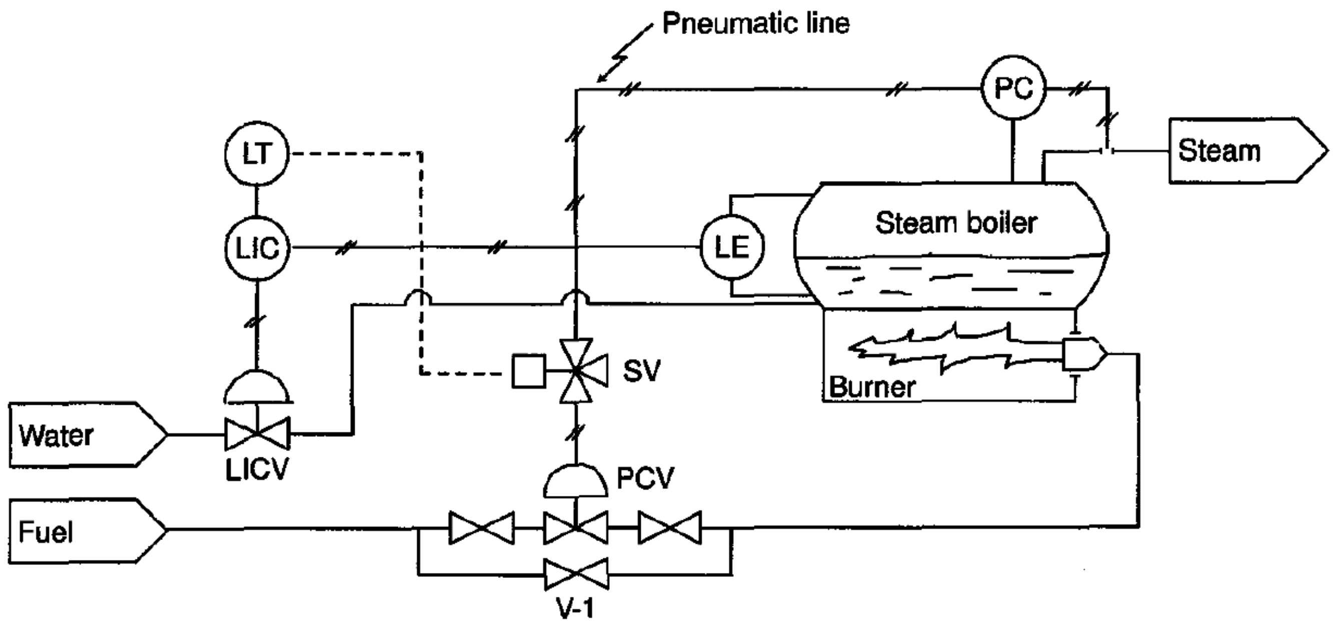
- So success trees are the complement (logical inverse) of fault trees (in most of the cases).

$$fault\ tree \rightarrow success\ tree$$
$$OR\ gate \rightarrow AND\ gate$$
$$AND\ gate \rightarrow OR\ gate$$
$$\overline{A} \cdot B + \overline{B} \cdot A \rightarrow A \cdot B + \overline{A} \cdot \overline{B}$$

(exclusive or)

Boiler example (page 1 of 3)

- This steam boiler system supplies steam to a process system at a specified pressure.



- A critical situation occurs if the boiler is boiled dry. In this case the pressure in the vessel will increase very rapidly and the vessel may explode

a. Construct a fault tree where the Top event is the critical situation mentioned above. (Only use components given a variable name on next 2 pages – but annotate tree with the relevant failure mode)

b. Determine all minimal cut sets in the fault tree.

Boiler example (page 2 of 3)

- **Description of system provided to you by boiler engineer (2 slides)**

Water from a feedwater system (W) is led to the boiler through pipe with a regulator valve called a level indicator controller valve (LICV). Fuel is led to the burner chamber through a pipe with a regulator valve called a pressure controller valve (PCV). The valve PCV is installed in parallel with a bypass valve (V-1) together with two isolation valves to facilitate inspection and maintenance of the PCV during normal operation.

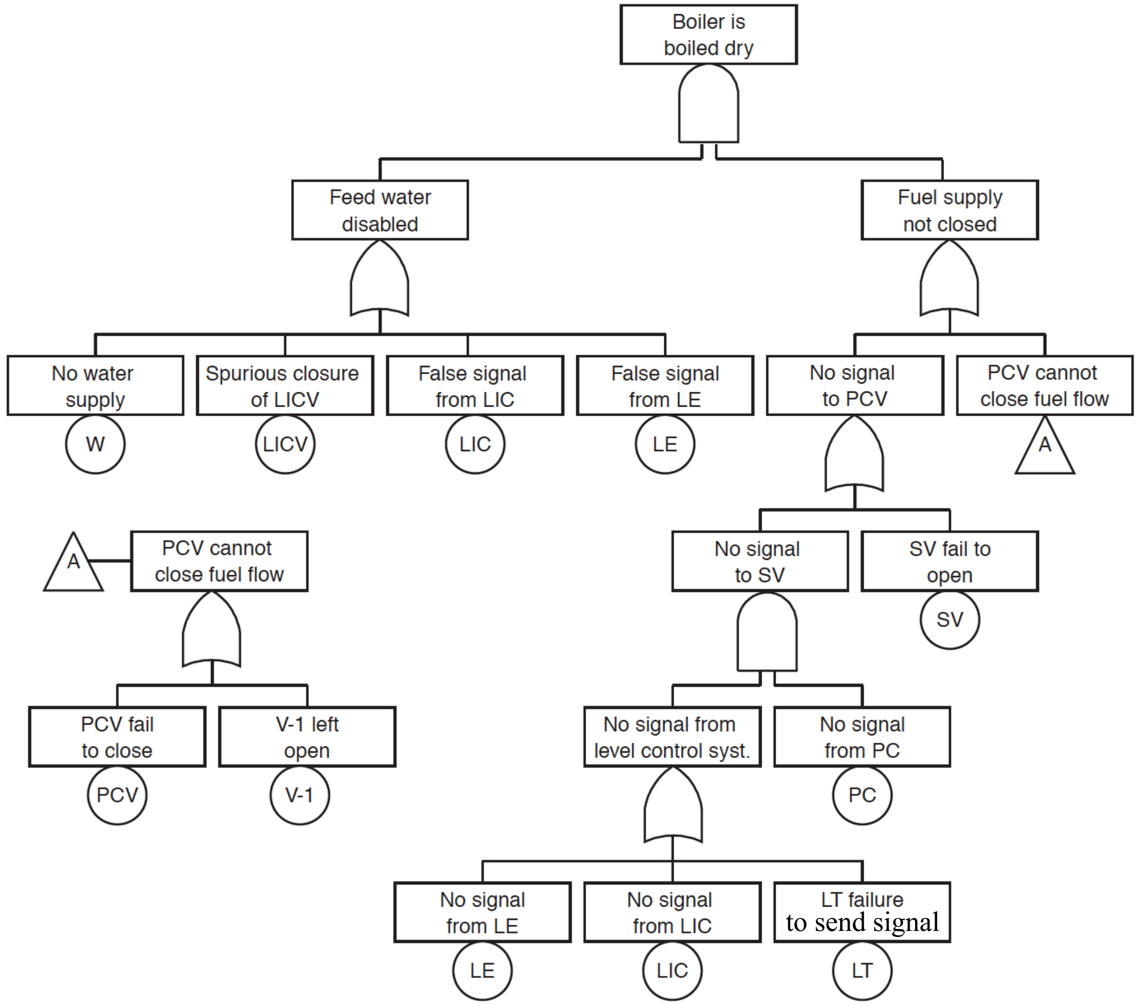
The level of the water in the boiler is monitored by a level emitter (LE). The water level is maintained in an interval between a specified low level and a specified high level by a pneumatic control circuit connected to the water regulator valve LICV. The level indicator controller (LIC) translates the pneumatic "signal" from LE to a pneumatic "signal" controlling the valve LICV.

Boiler example (page 3 of 3)

It is very important that the water level does not fall below the specified low level. When the water level approaches the low level, a pneumatic "signal" is passed from the level indicator controller LIC to the level transmitter (LT). The LT translates the pneumatic signal into an electrical signal which is sent to the solenoid valve (SV). The solenoid valve again controls the valve PCV on the fuel inlet pipeline. This circuit is thus installed to cut off the fuel supply in case the water level comes below the specified low level.

The pressure in the boiler and in the steam outlet pipeline is monitored by a pressure controller PC which is connected to the solenoid valve SV, and thereby to the valve PCV on the fuel inlet pipeline. This circuit is thus installed to cut off the fuel supply in case the pressure in the boiler increases above a specified high pressure

Boiler example: Solution (page 1 of 2)

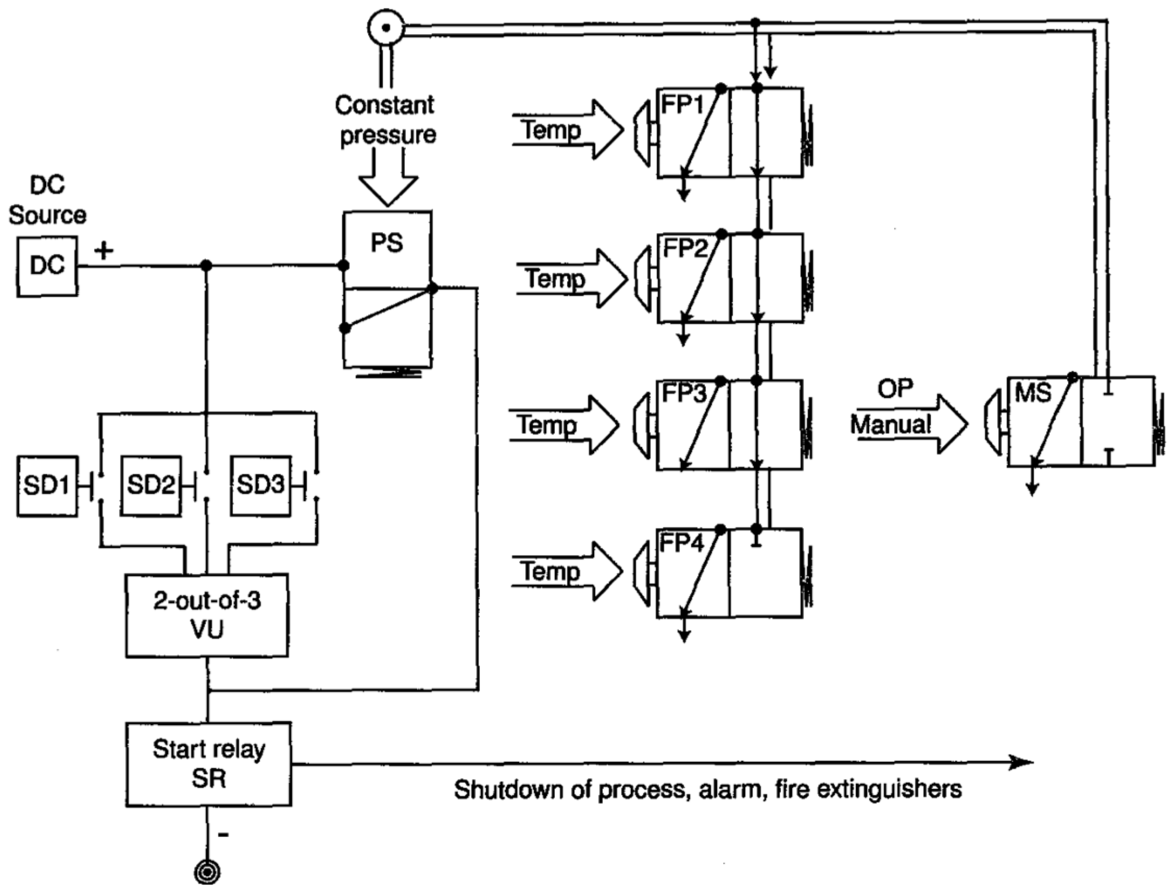


Boiler example: Solution (page 2 of 2)

{W, SV}	{W, PCV}	{W, V-1}
{LICV, SV}	{LICV, PCV}	{LICV, V-1}
{LIC, SV}	{LIC, PCV}	{LIC, V-1}
{LE, SV}	{LE, PCV}	{LE, V-1}
{LIC, PC}	{LE, PC}	{W, PC, LT}
{LICV, PC, LT}		

Another example* (SKIP slides 82-88/do after class)

- Consider the following fire detector system:



*Source: System Reliability Theory: Models, Statistical Methods, and Applications, Second Edition, Wiley, 2004, M. Rausand, A. Hoyland

Another example

- The fire detector system is divided into two parts, heat detection and smoke detection. In addition, there is an alarm button that can be operated manually.
- *Heat Detection*:

In the production room there is a closed, pneumatic pipe circuit with four identical fuse plugs, FP_1, FP_2, FP_3, and FP_4. These plugs let air out of the circuit if they are exposed to temperatures higher than 72C. The pneumatic system has a pressure of 3 bars and is connected to a pressure switch PS. If one or more of the plugs are activated, the switch will be activated and give an electrical signal to the start relay for the alarm and shutdown system. In order to have an electrical signal, the direct current (DC) source must be intact.

Another example

- *Smoke Detection*:

The smoke detection system consists of three optical smoke detectors, SD_1, SD_2, and SD_3; all are independent and have their own batteries. These detectors are very sensitive and can give warning of fire at an early stage. In order to avoid false alarms, the three smoke detectors are connected via a logical 2-out-of-3 voting unit, VU.

This means that at least two detectors must give fire signal before the fire alarm is activated. If at least two of the three detectors are activated, the 2-out-of-3 voting unit will give an electric signal to the start relay, SR, for the alarm and shutdown system. Again the DC voltage source must be intact to obtain an electrical signal.

Another example

- *Manual Activation*:

Together with the pneumatic pipe circuit with the four fuse plugs, there is also a manual switch, MS, that can be turned to relieve the pressure in the pipe circuit. If the operator, OP, who should be continually present, notices a fire, he can activate this switch. When the switch is activated, the pressure in the pipe circuit is relieved and the pressure switch, PS, is activated and gives an electric signal to the start relay, SR. Again the DC source must be intact.

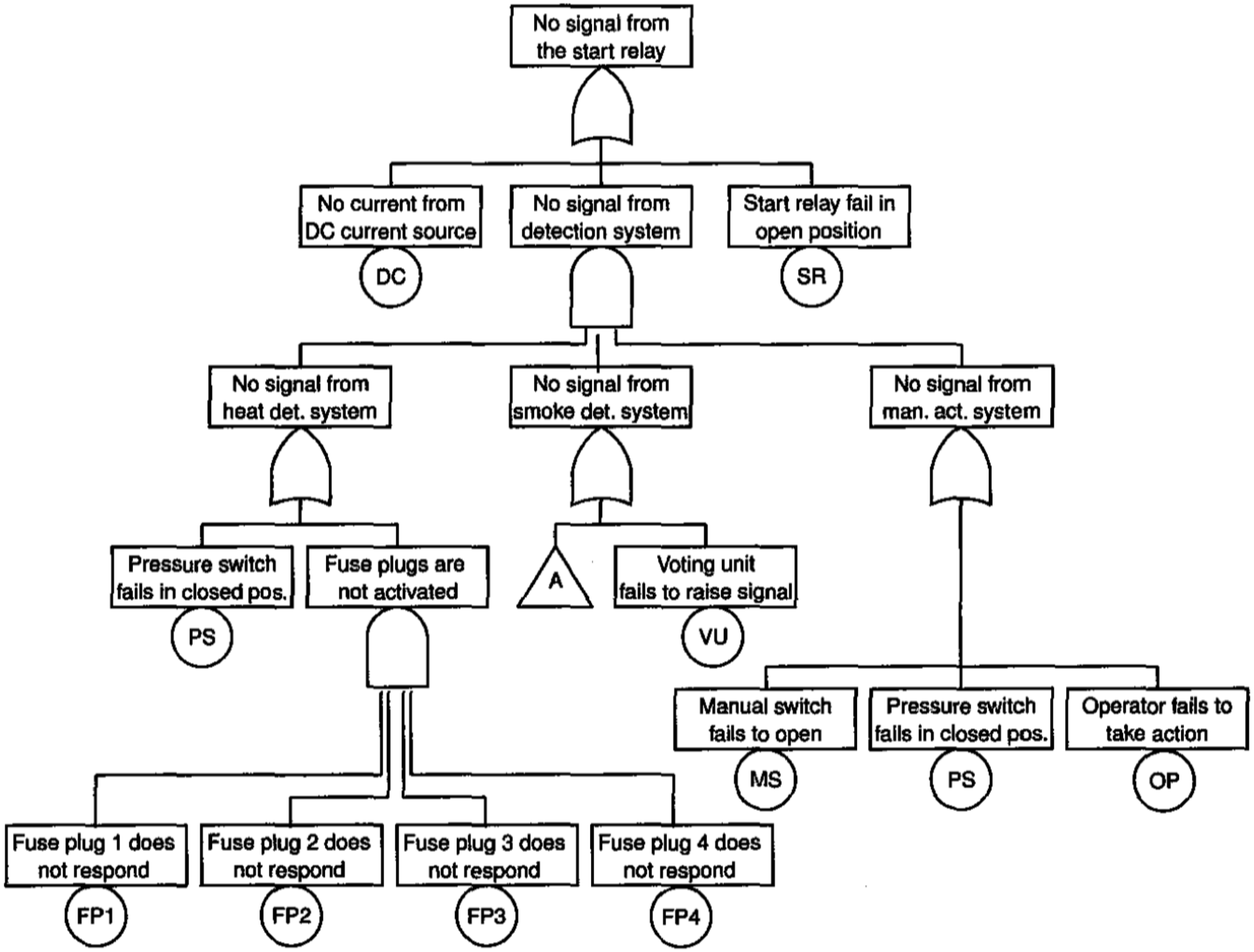
- *The Start Relay*:

When the start relay, SR, receives an electrical signal from the detection systems, it is activated and gives a signal to (i) Shut down the process, and (ii) Activate the alarm and the fire extinguishers.

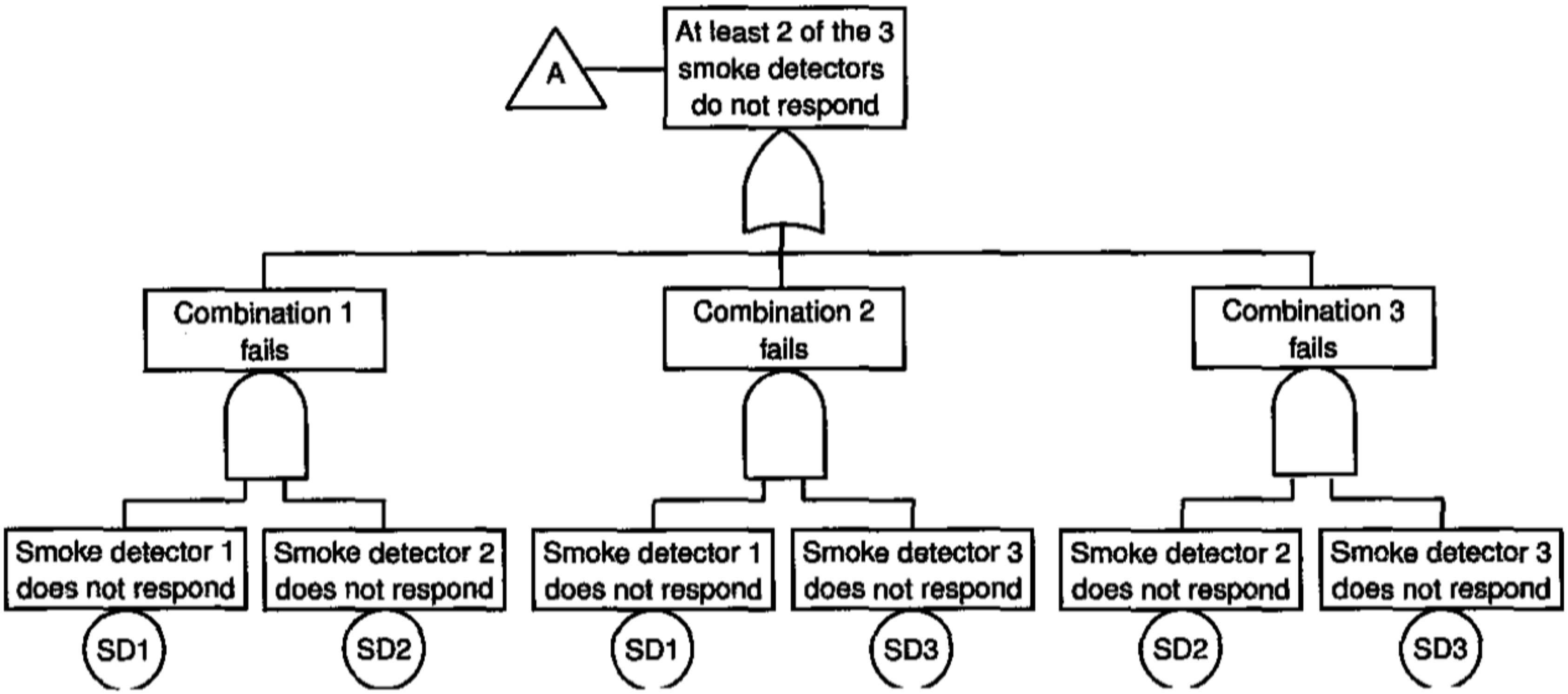
Another example

- Assume now that a fire starts
- The fire detector system should detect and give warning about the fire
- Let the TOP event be: *no signals from the start relay SR when a fire condition is present*
- Develop the fault tree and find the cut sets

Another example: Fault tree



Another example: Fault tree



Minimal cut-sets for this example are (14 in total):
DC + SR+PSVU+PSSD1SD3+PSSD2SD3+PSSD1SD2+FP1FP2FP3FP4MSVU+
FP1FP2FP3FP4OPVU+FP1FP2FP3FP4MSSD1SD2+FP1FP2FP3FP4MSSD2SD3+
FP1FP2FP3FP4MSSD1SD3+FP1FP2FP3FP4OPSD1SD2+FP1FP2FP3FP4OPSD1SD3+
FP1FP2FP3FP4OPSD2SD3

Event tree method

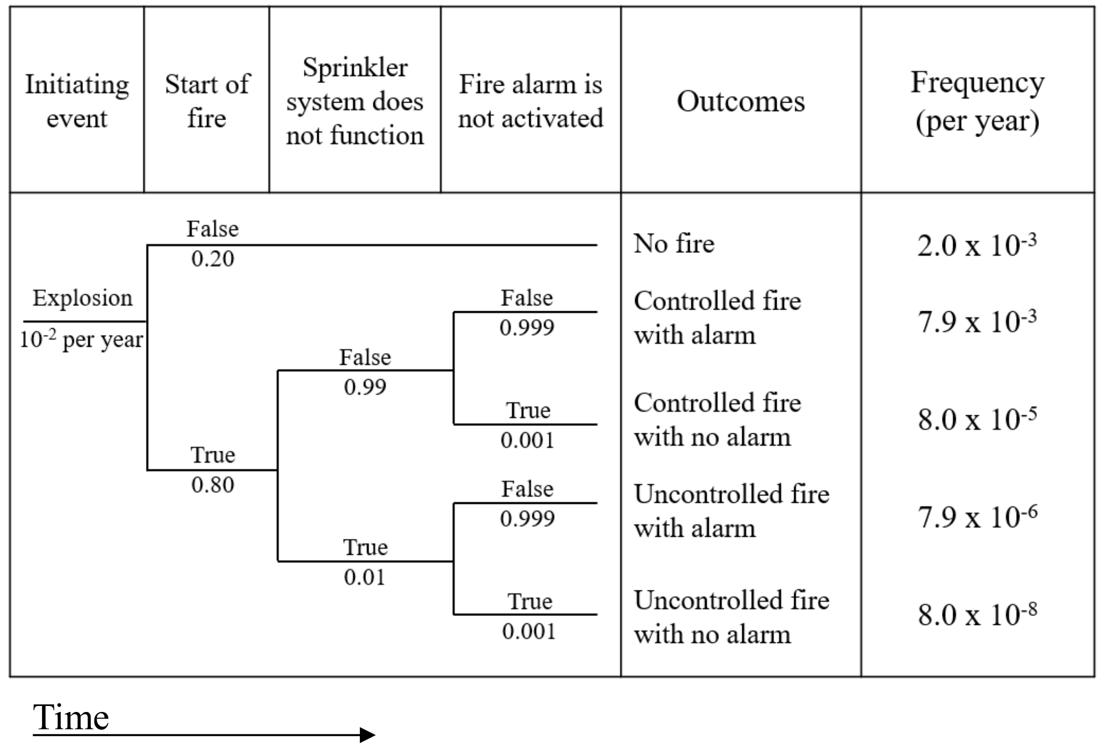
If successful operation of a system depends on an ***approximately*** chronological but discrete operation of some of its units or subsystems, then an event tree is a useful logical model for the system.

Key Features

- Sequence of events
- Initiating event (with frequency)
- Up/down branching
  - In this class:
  - Up = Success
  - Down = Failure
- (Conditional) probabilities

or independent events

- Mutually exclusive sequences
- End state descriptions



Event tree quantification

A Initiating event	B Start of fire	C Sprinkler system does not function	D Fire alarm is not activated	Outcomes	Frequency (per year)
Explosion 10^{-2} per year	False 0.20			No fire	2.0×10^{-3}
	True 0.80	False 0.99	False 0.999	Controlled fire with alarm	7.9×10^{-3}
			True 0.001	Controlled fire with no alarm	8.0×10^{-5}
		True 0.01	False 0.999	Uncontrolled fire with alarm	7.9×10^{-6}
			True 0.001	Uncontrolled fire with no alarm	8.0×10^{-8}

$$\boldsymbol{\Sigma = 0.01}$$

- Note: typically written with marginal notation but should be interpreted as conditional when needed!

$$= f(A)\Pr(\bar{B}|A)$$

$$= f(A)\Pr(B|A)\Pr\left(\overline{C}\middle|B,A\right)\Pr\left(\overline{D}\middle|\overline{C},B,A\right)$$

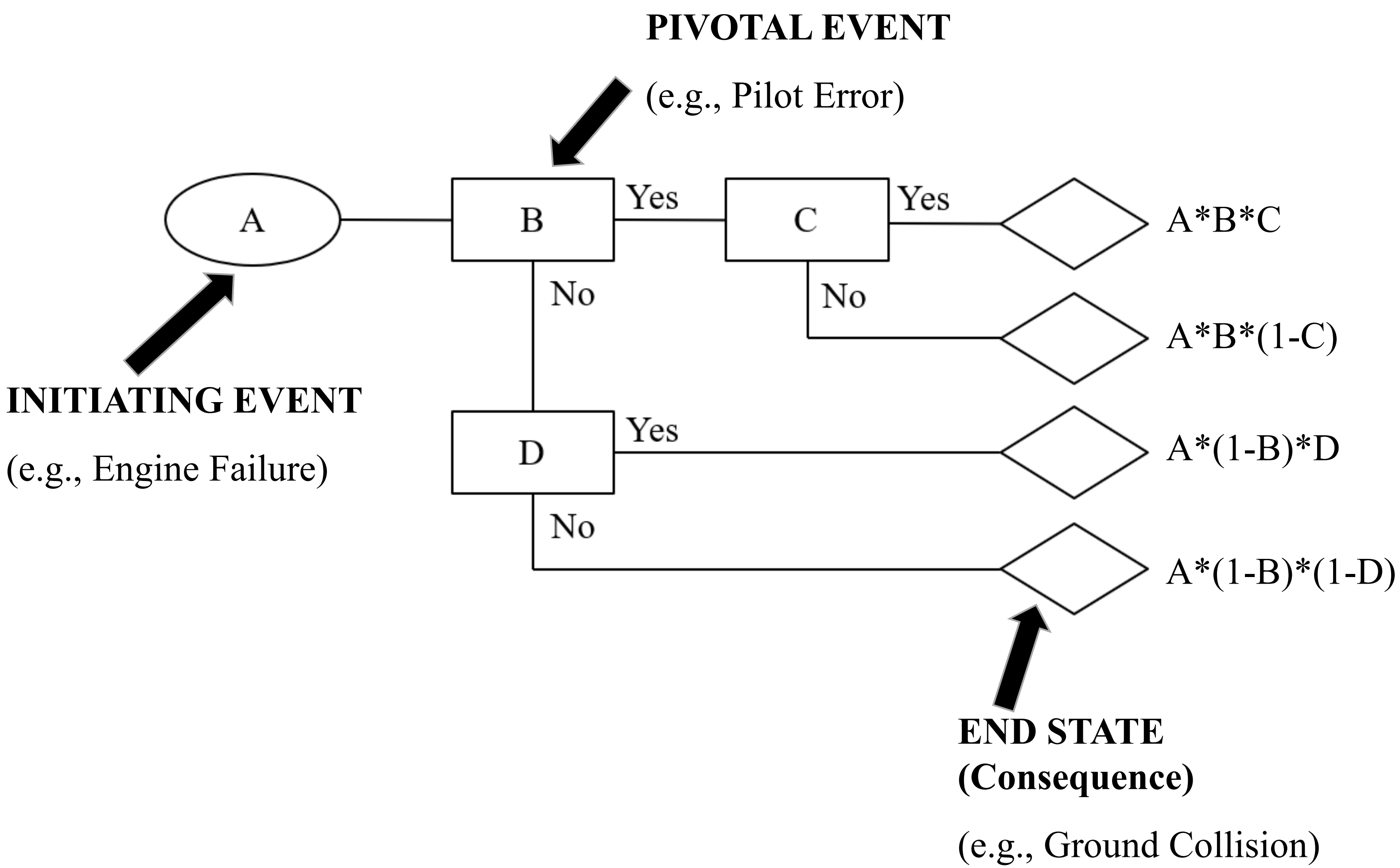
$$= f(A)\Pr(B|A)\Pr\left(\overline{C}\middle|B,A\right)\Pr\left(D\middle|\overline{C},\ B,A\right)$$

$$= f(A)\Pr(B|A)\Pr(C|B,A)\Pr\left(\overline{D}\middle|C,B,A\right)$$

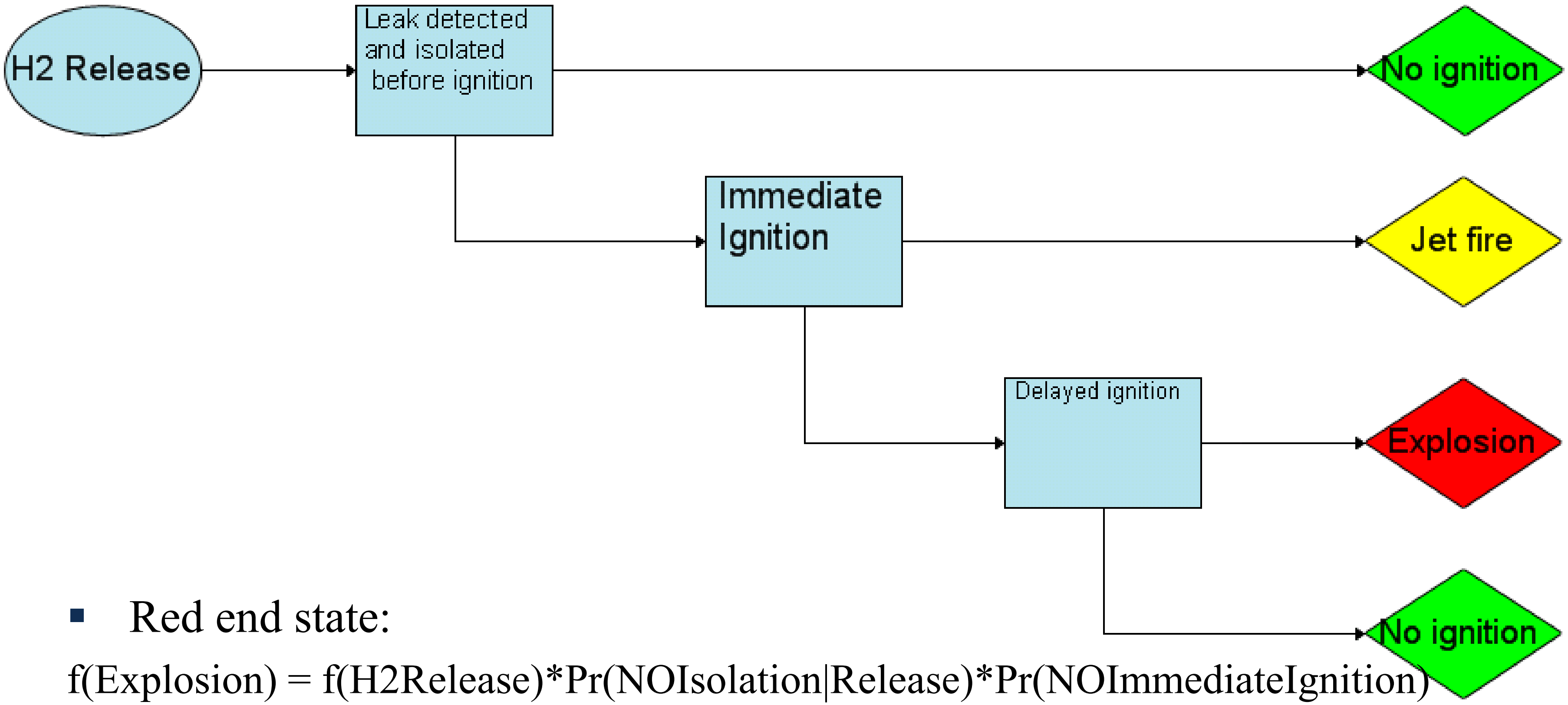
$$= f(A)\Pr(B|A)\Pr(C|B,A)\Pr(D|C,B,A)$$

Adapted from: M. Rausand, A. Hoyland, *System Reliability Theory: Models, Statistical Methods, and Applications*, Second Edition, Wiley, 2004.

Event sequence diagram/event tree



ESD: Hydrogen release from H2 dispenser

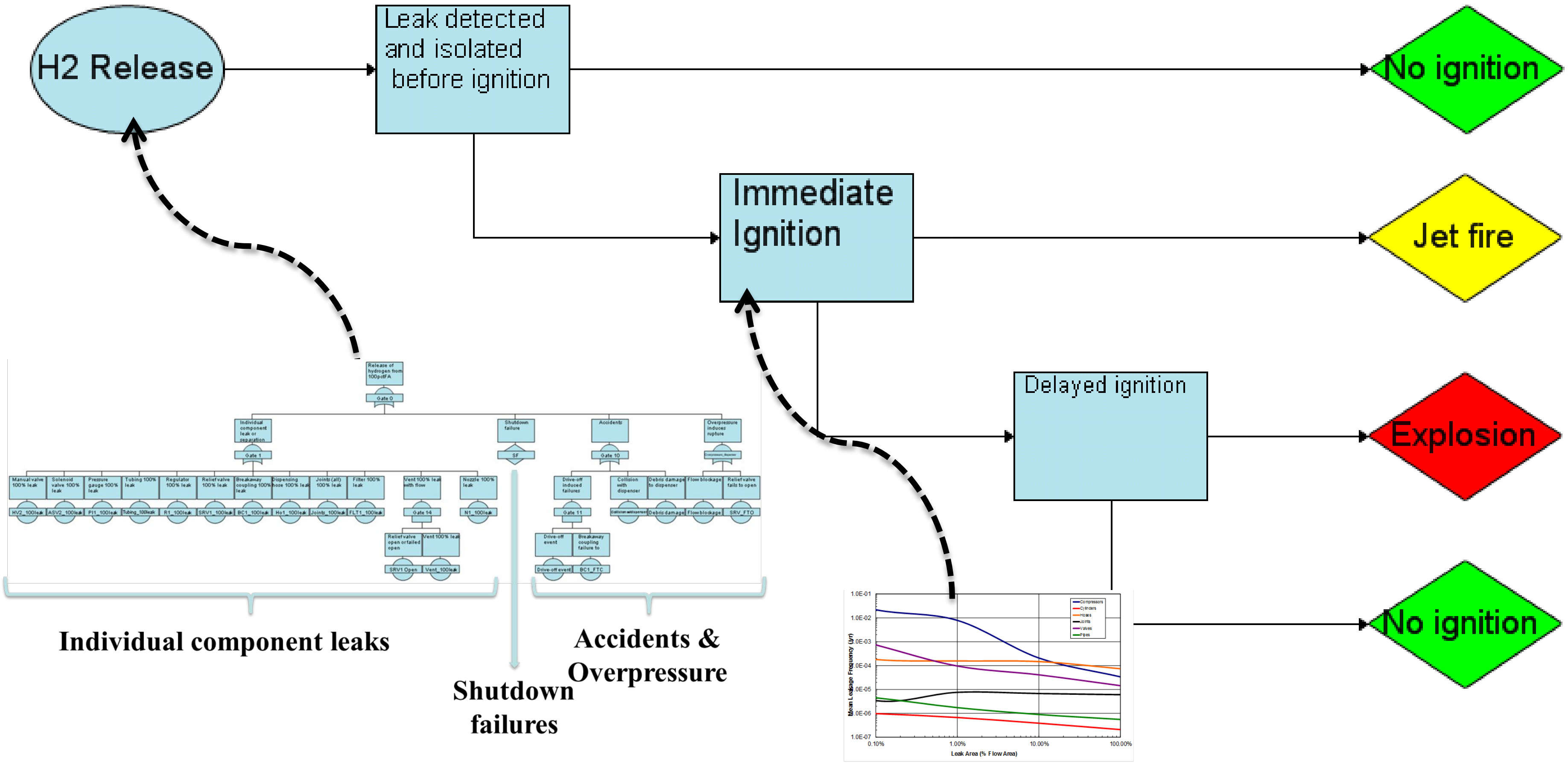


- Red end state:

f(Explosion) = f(H2Release)*Pr(NOIsolation|Release)*Pr(NOImmediateIgnition)*Pr(Delayedignition)

Adapted from: Groth, Katrina M., and Ethan S. Hecht. 2017. "HyRAM: A Methodology and Toolkit for Quantitative Risk Assessment of Hydrogen Systems." *International Journal of Hydrogen Energy*, International Journal of Hydrogen Energy, 42 (11): 7485–7493. doi:10.1016/j.ijhydene.2016.07.002.

Using event tree + fault tree for system reliability

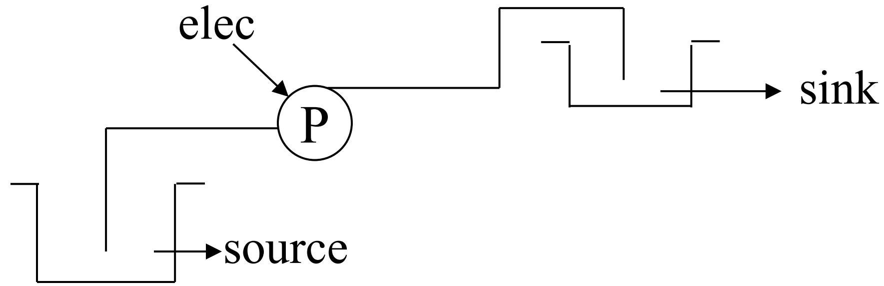


Connected to FT(s) (Or BNs)…

Or Direct use of data

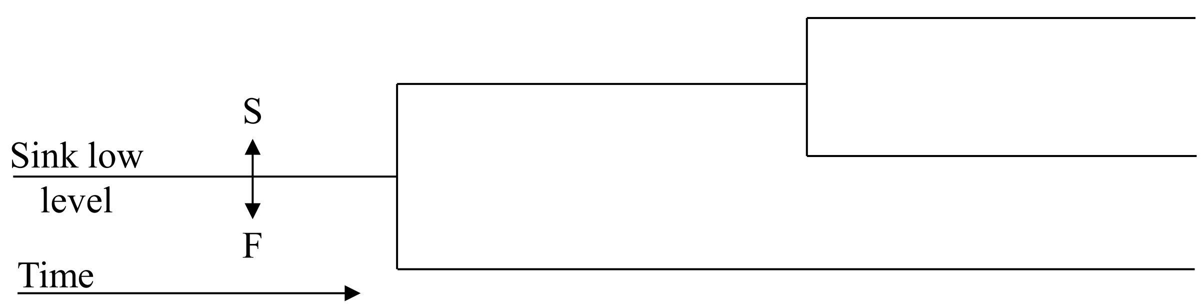
Event tree for pumping system reliability

- **Example: Pumping system designed to fill a sink from a source tank.**



I = frequency of "sink low" event
Elec = Electric power fails off
P = Pump fails off

If the sink is low, the system should pump water from the source. An ET for this would be:

I (Sink Low)	Elec Power (Elec)	Pump (P)	End State
			$I \cdot \overline{Elec} \cdot \overline{P}$: System Functions (S1)
			$I \cdot \overline{Elec} \cdot P$: System Fails (F1)
			$I \cdot Elec$: System Fails (F2)

Event tree for pumping system reliability

- Since the event tree sequences are mutually exclusive, we can calculate probability of system failure as:

$$
\begin{aligned}
\Pr(system\ failure) &= \Pr(F_1 + F_2) \\
&= \Pr\left(I \cdot \overline{Elec} \cdot P + I \cdot Elec\right) \\
&= \Pr\left(I \cdot \overline{Elec} \cdot P\right) + \Pr(I \cdot Elec)
\end{aligned}
$$

If *I* happens with a frequency 10 times/mo, *Elec* power fails off with probability 0.1, and *P* fails off with probability 0.05,

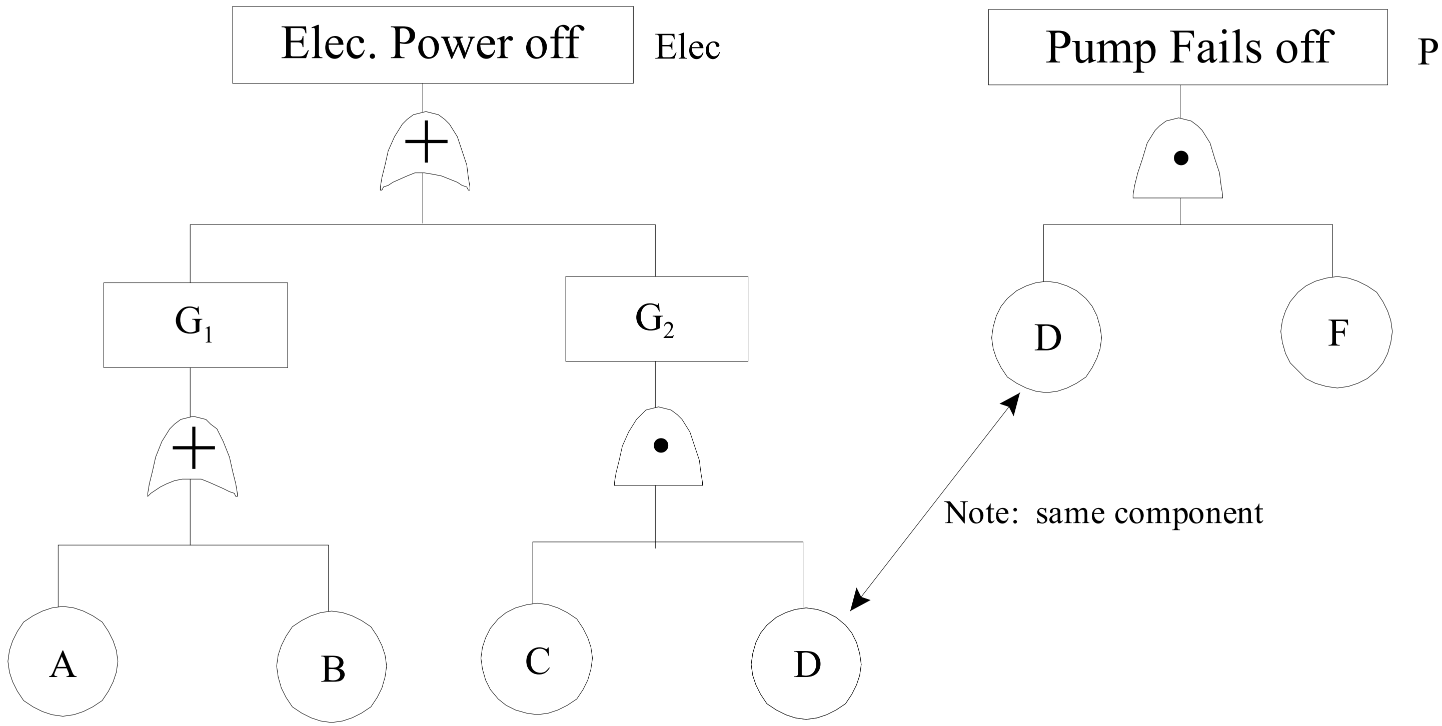
f(S1)=8.55/month

f(F1)=0.45/month

f(F2)=1.0/month

Event tree + fault tree method

- Now, assume *Elec* and *P* can be represented by a set of fault trees showing how they could fail:



Event tree + fault tree method

- First start by solving the FTs

$$Elec = G_1 + G_2$$

$$G_1 = A + B, \qquad G_2 = C \cdot D$$

$$\boldsymbol{Elec = A + B + C \cdot D}$$

$$\overline{Elec} = \overline{A + B + C \cdot D} = (\overline{A} \cdot \overline{B}) \cdot (\overline{C \cdot D}) = \overline{A} \cdot \overline{B} \cdot (\overline{C} + \overline{D})$$

$$\boldsymbol{\overline{Elec} = \overline{A} \cdot \overline{B} \cdot \overline{C} + \overline{A} \cdot \overline{B} \cdot \overline{D}}$$

$$\boldsymbol{P = D \cdot F}$$

$$\boldsymbol{\overline{P} = \overline{D} + \overline{F}}$$

Event tree + fault tree method

- To find the ET sequence F_1, which is $I \cdot \overline{Elec} \cdot P,$ start inserting the logic from your FTs

$$I \cdot \overline{Elec} \cdot P = I \cdot \left(\overline{A} \cdot \overline{B} \cdot \overline{C} + \overline{A} \cdot \overline{B} \cdot \overline{D}\right) \cdot D \cdot F$$

$$= I \cdot \overline{A} \cdot \overline{B} \cdot \overline{C} \cdot D \cdot F + I \cdot \overline{A} \cdot \overline{B} \cdot \overline{D} \cdot D \cdot F$$

- Since $D \cdot \overline{D}$ is a null set:

$$\boldsymbol{I \cdot \overline{Elec} \cdot P = I \cdot \overline{A} \cdot \overline{B} \cdot \overline{C} \cdot D \cdot F}$$

Event tree + fault tree method

- In the same way, we find the expression for the bottom ET sequence, F_2 :

$$\boldsymbol{I \cdot Elec = I \cdot A + I \cdot B + I \cdot C \cdot D}$$

- And the final ET sequence, S_1 (the one representing system success).

$$I \cdot \overline{Elec} \cdot \overline{P} = I \cdot \left(\overline{A} \cdot \overline{B} \cdot \overline{C} + \overline{A} \cdot \overline{B} \cdot \overline{D}\right) \cdot \left(\overline{D} + \overline{F}\right)$$
$$= I \cdot \overline{A} \cdot \overline{B} \cdot \overline{C} \cdot \overline{D} + I \cdot \overline{A} \cdot \overline{B} \cdot \overline{C} \cdot \overline{F} + I \cdot \overline{A} \cdot \overline{B} \cdot \overline{D} + I \cdot \overline{A} \cdot \overline{B} \cdot \overline{D} \cdot \overline{F}$$
$$= \boldsymbol{I \cdot \overline{A} \cdot \overline{B} \cdot \overline{C} \cdot \overline{F} + I \cdot \overline{A} \cdot \overline{B} \cdot \overline{D}}$$

Event tree + fault tree method

- Assume the following probabilities $\Pr(\bullet)$ and frequency $f(I)$, and independent events, find system failure probability (use the rare event approximation on the FT cut sets)

Event	Failure Probability or Freq.	Success Probability	Event	Failure Probability or Frequ.	Success Probability
I	10 per month		C	0.02	0.98
A	0.01	0.99	D	0.05	0.95
B	0.01	0.99	F	0.01	0.99

- Since the event tree sequences are mutually exclusive:

$$\Pr(system\ failure) = \Pr(I \cdot Elec) + \Pr(I \cdot \overline{Elec} \cdot P)$$

$$= \underbrace{f(I)\Pr(A) + f(I)\Pr(B) + f(I)\Pr(C)\Pr(D)}_{Not\ mutually\ exclusive\ items, but\ rare\ event\ applies} + \underbrace{f(I)\Pr(\overline{A})\Pr(\overline{B})\Pr(\overline{C})\Pr(D)\Pr(F)}_{Rare\ event\ doesn't\ apply}$$

Event tree + fault tree method

- Frequency of System Failure with rare event approximation is:

$F_1 + F_2$

$= 10(0.01) + 10(0.01) + 10(0.02)(0.05) + 10(0.99)(0.99)(0.98)(0.05)(0.01)$

$= 10(0.021 + 0.00048) = \mathbf{0.2148}\ \boldsymbol{failures/month}$

- The Frequency of System Failure without rare event approximation uses the addition law of probability:

$= f(I)\{(1-(1-P(A))(1-P(B))(1-P(CD))) + P(\overline{ABC}DF)\}$

$= 10\ \{(1-(1-.01)(1-.01)(1-.02 * .05) + (0.048)\}$

$=10\ \{(1-(.99)(.99)(.999)) + (0.00048)\}$

$=10(0.02088+0.00048)$

$= \mathbf{0.2136}\ \boldsymbol{failures/month}$

Event tree + fault tree method

- Frequency of System Success is:

$$I \cdot \overline{elec} \cdot \overline{P} = I \cdot \left(\overline{A} \cdot \overline{B} \cdot \overline{C} + \overline{A} \cdot \overline{B} \cdot \overline{D}\right) \cdot \left(\overline{D} + \overline{F}\right)$$

$$= I \cdot \overline{A} \cdot \overline{B} \cdot \overline{C} \cdot \overline{D} + I \cdot \overline{A} \cdot \overline{B} \cdot \overline{C} \cdot \overline{F} + I \cdot \overline{A} \cdot \overline{B} \cdot \overline{D} + I \cdot \overline{A} \cdot \overline{B} \cdot \overline{D} \cdot \overline{F}$$

$$= I \cdot \overline{A} \cdot \overline{B} \cdot \overline{C} \cdot \overline{F} + I \cdot \overline{A} \cdot \overline{B} \cdot \overline{D}$$

For success quantification, the rare event approximation doesn't apply because the probabilities are too high. The two cut sets for this sequence aren't mutually exclusive, either. So, we calculate it using the addition law of probability:

$$R_s = 1 - \prod (1 - R_i)$$

So, assuming independence between the two parts of the sequence:

$$= 10\{1 - [(1 - 0.99 \times 0.99 \times 0.98 \times 0.99)(1 - 0.99 \times 0.99 \times 0.95)]\} \cong 9.9662$$

$$= 10\{1 - [(1 - 0.9508)(1 - 0.931095)]\}$$

$$= 10\{1 - [(0.0491)(0.0689)]\} = 10 \cdot 9.9662 \cong 9.9662$$

Event tree + fault tree method

- Exact calculation for frequency of System Success is:

$$I \cdot \overline{elec} \cdot \overline{P} = I \cdot \overline{A} \cdot \overline{B} \cdot \overline{C} \cdot \overline{F} + I \cdot \overline{A} \cdot \overline{B} \cdot \overline{D}$$

It should be noted that these two parts of the sequence aren't independent (even though the units A, B, C, D are independent of each other) – this is because of the element D which appears in both FTs), so the exact answer would be:

$$\mathrm{I} \cdot (\Pr(\bar{A}, \bar{B}, \bar{C}, \bar{F}) + \Pr(\bar{A}, \bar{B}, \bar{D}) \; - \Pr(\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{F}))$$

$= 10((.99 \times .98 \times .99 \times .99) + (.99 \times .99 \times .95) - (.99 \times .99 \times .98 \times .95 \times .99))$

$= 10((0.9509) + (.9311) - (.9037))=$**9.786**

Reliability Analysis

Module 6C: Failure Modes and Effects Analysis (FMEA/FMECA)

Probability Class				
	High	M	H	H
	Medium	L	M	H
	Low	L	L	M
		Minor	Moderate	Critical
		Severity Class		

Prof. Katrina M. Groth

Mechanical Engineering Department

Center for Risk and Reliability

University of Maryland

kgroth@umd.edu

#	Failure Mode	Effect	Severity	Likelihood
ASV1	External Leak	H2 accumulation above leak	3 - Critical	4 - Frequent
Tubing	External Leak	H2 accumulation above leak	3 - Critical	4 - Frequent
	Rupture/separation	Large H2 release if HV2 and N1 also fail	4 - Catastrophic	2 - Occasional
F1	Flow blockage	Potential overpressure at filter induces filter separation	2 - Marginal	3 - Reasonably probable
	Fluid contamination	Contaminated H2	2 - Marginal	3 - Reasonably probable
	External Leak	Accumulation of H2 above F1	3 - Critical	4 - Frequent
R1	External Leak	Accumulation of H2 in building	3 - Critical	4 - Frequent

Why FMEA?

- Pros
  - Valuable qualitative insights
  - Facilitates participation of multiple types of expertise
  - Comprehensive
  - Scalability to different design stages
  - Early insight into potential problems
  - Feedback process to address problems
  - Enables initial reliability analysis on low-maturity systems.
- Cons
  - "Not probabilistic" in the textbook. Meaning: Simplistic quantification results in high uncertainties, subjectivity, limited "big picture" insights
  - Limited insight into system-level failures
  - Can only address one failure at a time
  - High time, effort, expertise required

Anything else?

Types

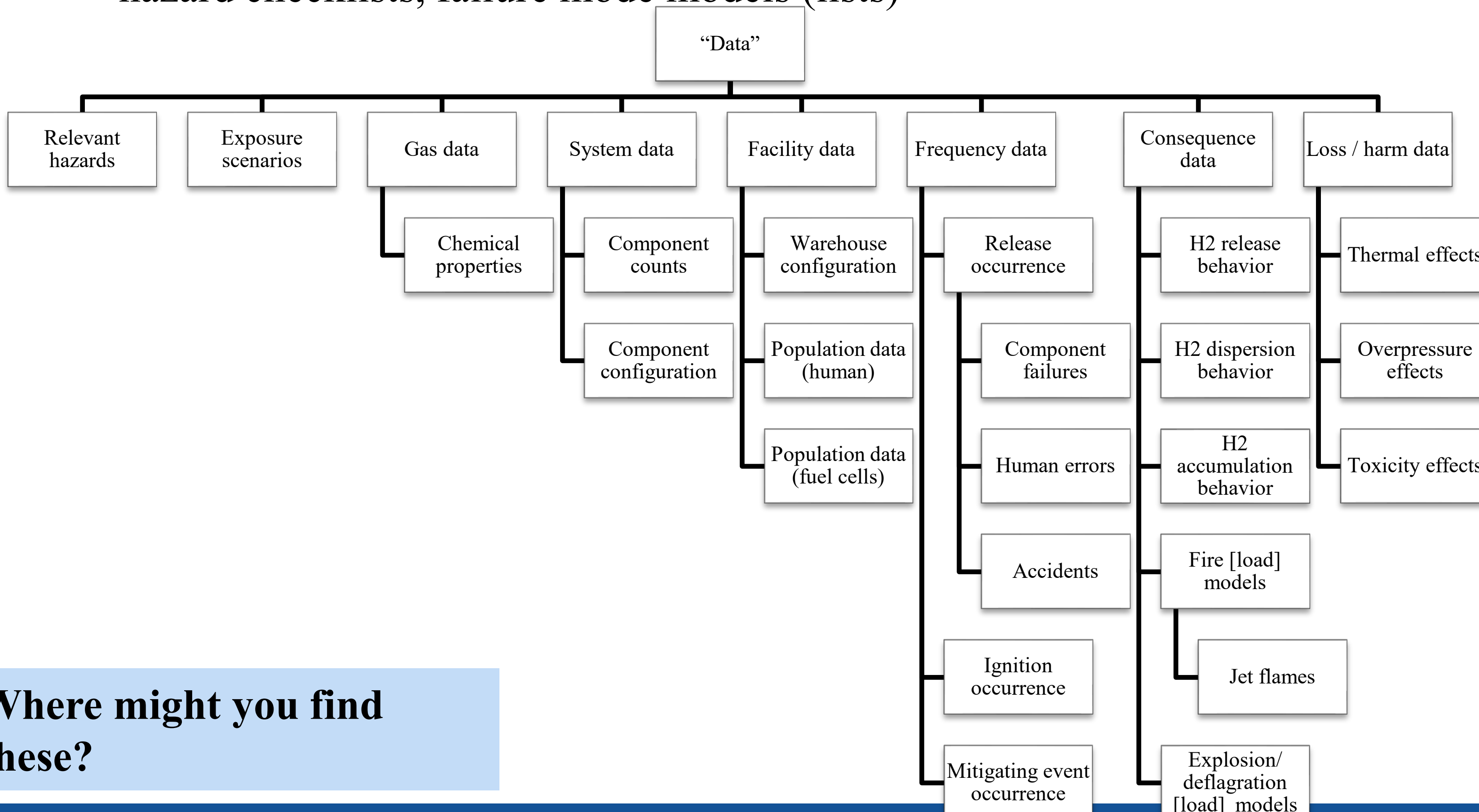
- Types of FMEA: See text or published procedures
  - Design FMEA
  - Process FMEA
  - Concept FMEA
- Published FMEA procedures:
  - **MIL-STD-1629A**. Procedures for Performing Failure Mode, Effects and Criticality Analysis. 1980.
  - SAE RP J1739
    - Ford FMEA Handbook- based on SAE RP J1739 http://www.quality.ford.com/cpar/fmea/
  - **SAE ARP5580**. Aerospace Recommended Practice-Recommended Failure Modes and Effects Analysis Practices for Non-Automobile Application. SAE International. Updated May 2012.
  - **IACS Rec No. 138**. Recommendation for the FMEA process for diesel engine control systems. Dec 2014. http://www.iacs.org.uk/media/2644/rec_no_138_pdf2553.pdf
  - **Other well-documented procedures exist**. Others you work with?

Planning an FMEA (1)

- **<u>Define goal</u>**. Options from SAE ARP5580 include:
  - Enhancing system safety by uncovering failure modes that result in hazardous conditions
  - Assessing the mission related effects of critical and/or undetectable failures
  - Influencing the design engineer to select a design with a high probability of operation success
  - Assisting the design engineer to select a design with a high probability of operation success
  - Providing data for development of effective maintenance support
- **<u>Define method and ground rules</u>**
  - Terminology, assumptions, worksheet format, end effects categories, severity definitions, boundary conditions, failure criteria, level of detail
    - An FMEA standard will define many of these
- **<u>Assemble the team</u>**
  - 2-5 core team members plus access to experts from risk analysis, design, manufacturing, operations, maintenance, etc.
  - Experienced members + newer engineers; diverse perspectives

Planning an FMEA (2)

- **<u>Assemble the information basis</u>**
  - System diagrams, system descriptions, system breakdowns, data sources, hazard checklists, failure mode models (lists)



Where might you find these?

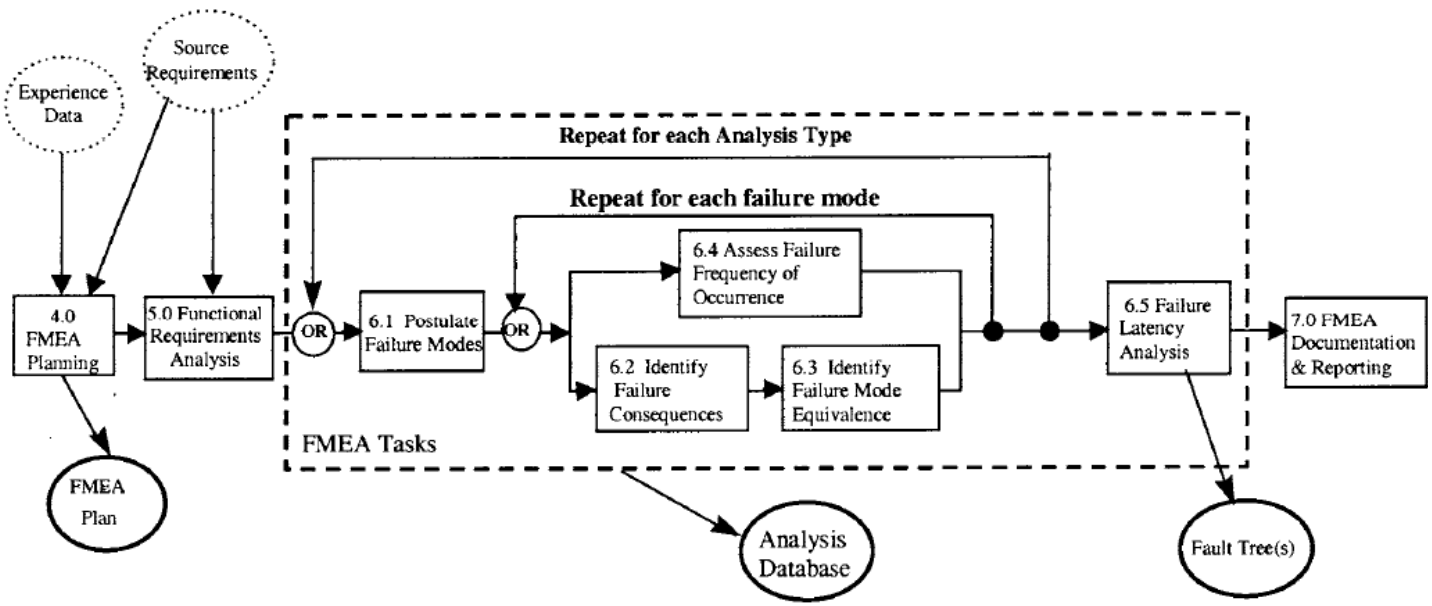
(A variant of this figure has subsequently been published in: R. Moradi, & KM Groth. 2019. "Hydrogen Storage and Delivery: Review of the State of the Art Technologies and Risk and Reliability Analysis." *International Journal of Hydrogen Energy*, 44 (23): 12254–12269. doi:10.1016/j.ijhydene.2019.03.041.

General Procedure for FMEA/FMECA (after planning…)

- **Define the system to be analyzed**
  - System boundaries
  - Internal and interface functions
  - Failure definitions
- **Construct a block diagram of the system**
  - Structural (hardware)
  - Functional block diagram
  - Reliability block diagram (RBD)
- **Complete FMEA worksheet**
  - Identify failure modes and effects
  - Assign severity and likelihood
  - Identify compensation provisions, design corrections
- **Document the analysis (!)**

FMEA Methodology: SAE ARP 5580

- For an FMEA for the LNG Locomotive/Tender System, SAE ARP 5580 is recommended
  - "*Aerospace Recommended Practice—Recommended Failure Modes and Effects Analysis Practices for Non-Automobile Application*." (May 2012 v.)
  - Based on level of available design details [*conceptual design phase*], a "Product Design Hardware" "functional analysis" approach was used (see Table 2 in ARP5580)



FMEA Worksheets- SAE ARP 5580

SYRRA

Product Design Hardware Functional Failure Mode and Effects Analysis

System: LNG Tender Supplying Fuel to a Locomotive diesel/LNG Internal Combustion Enginee

Operating Mode: Operation

Date: 12/9/2017

Analysts: [redacted] K. Groth

Version: Draft

Affiliation: University of Maryland

							Failure Effects							
Unit Indenture Level	**Assembly Indenture Level**	**Item/Functional Identification**	**Function**	**Failure Mode ID**	**Failure Modes and Causes**	**Failure Mode Model**	**Local Effects**	**Next Higher Level**	**End Effects**	**Severity Class**	**Item Failure Rate**	**Failure Mode Distribution Ratio**	**Failure Mode Rate**	**Probability Class**
30	37	Shutoff Valve	Controls flow to fuel into [39]	37.02	Valve operates spuriously due to control issues, short, etc.	Premature operation	GNG flow stopped prematurely	Low flow of GNG to engine	Locomotive performance is compromised	1	1.37	0.001	0.00	Low
30	37	Shutoff Valve	Controls flow to fuel into [39]	37.03	Leakage from valve due to seal failure, mechanical damage, etc	Failure to meet functional specs	Leakage		Potential release of GNG	2	1.37	0.668	0.92	Low

Worksheet Format: MIL-STD

- MIL-STD 1629, Task 101

System:
Indenture Level:
Reference Drawing:
Mission:

Data:
Sheet Number:
Compiled by:
Approved by:

ID	Item/ Functional Identification	Function	Failure Modes and Causes	Mission Phase Operational Mode	Failure Probability and Data Source	Failure Effects			Failure Detection Method	Compensating Provisions	Severity Class	Remarks
						Local	Next Higher	End				

- MIL-STD 1629, Task 102

System:
Indenture Level:
Reference Drawing:
Mission:

Data:
Sheet Number:
Compiled by:
Approved by:

ID	Item/ Functional Identification	Function	Failure Modes and Causes	Mission Phase Operational Mode	Severity Class	Failure Probability and Data Source	Failure Effect Prop.	Failure Mode Ratio	Failure Rate, 1/hour	Mission Duration hour	Failure Mode Criticality	Item Criticality $C = \sum C_m$	Remarks

FMEA + Criticality Analysis = FMECA

- **FMEA (MIL-STD 1629, Method 101)**
  - Standard: **Severity classification of Failure Mode Occurrence** on a 1 to 4 scale
  - Variation: **Risk Priority Number (RPN)** = Occurrence * Severity * Detection
    - **All above estimates evaluated on a <u>relative</u> 1 to 5 scale**
- **FMECA (MIL-STD 1629, Method 102)**
  - Qualitative: **Severity of Failure Mode Occurrence** on a 1 to 5 relative scale
  - Quantitative: Criticality Number =Part Failure Rate * Failure Mode Ratio * Probability of Function Loss * Operating Time.
    - **All above estimates are obtained through generic (field) failure data**

Failure mode criticality number

- A numerical value used to rank each potential failure mode based on its likelihood of occurrence and the consequence of its effect.

$$C_m = \lambda t \alpha \beta$$

- Where,
  - λ = *Part Failure Rate* (estimated by an appropriate failure data analysis or calculated from MIL-HDBK-217)
  - t = *the estimated mission time* of the unit (system)
  - α = *Failure Mode Ratio* as the fraction of the part failure rate related to a particular failure mode (estimated by an appropriate failure data analysis or calculated from MIL-HDBK-217)
  - β = *Failure Effect Probability*

Failure Effect	$\boldsymbol{\beta}$
Actual loss	1.0
Probable loss	0.1 - 1.0
Possible loss	0.0 – 0.1
No loss	0.0

Failure mode criticality number (cont.)

- The unit criticality number is the sum of criticality numbers for all individual failure modes of that unit: $C_{unit} = \sum C_{fm}$
- Notes:
  - Use the RBDs to evaluate the failure effect probability β for non-series systems.
  - The notion of the failure mode ratio assumes that **independence** of individual failure modes ($\sum \alpha_i = 1$).

Failure mode by criticality number (cont.)

- **Example**
  - A 1 kV varistor has a generic failure rate of $\lambda_p = 1 \times 10^{-6}/hour$ and the "short-circuit" failure mode ratio is $\alpha = 0.8$. The "short-circuit" failure mode results into the probable loss of a High Voltage protection circuit with $\beta = 0.001$. For all other failure modes, $\beta = 0.01$.
  - Determine the criticality number for the 10 month (7200 hour) period of the system operating time.

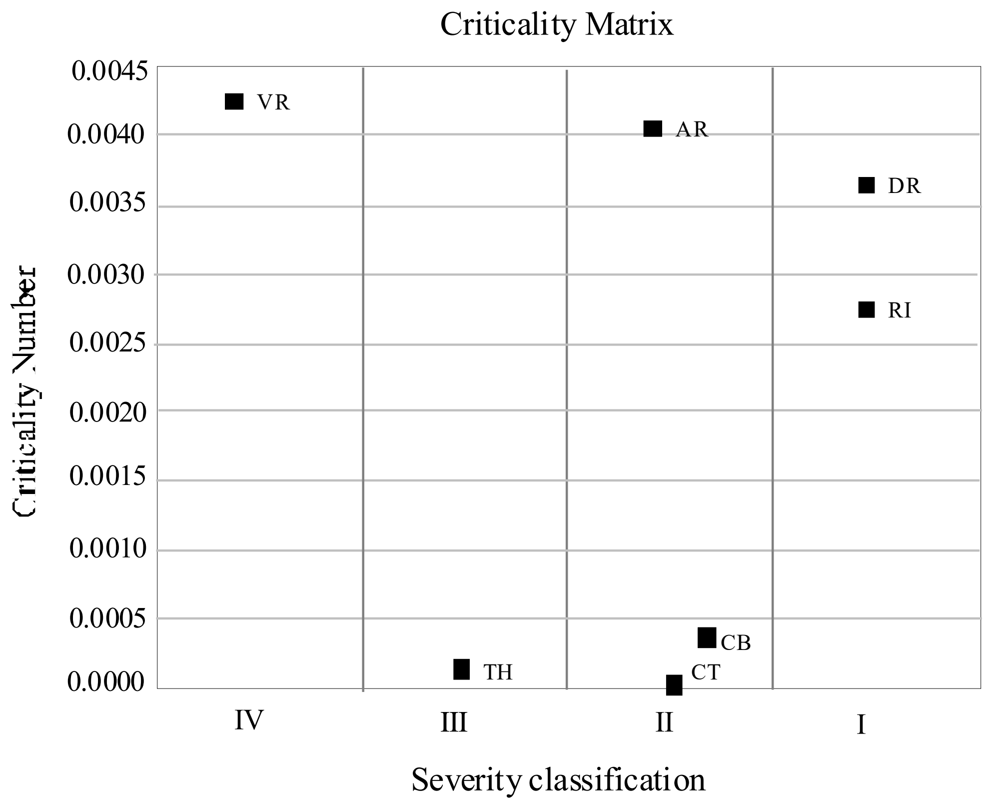
Failure mode by criticality number (cont.)

- **Example**
  - A 1 kV varistor has a generic failure rate of $\lambda_p = 1 \times 10^{-6}/hour$ and the "short-circuit" failure mode ratio is $\alpha = 0.8$. The "short-circuit" failure mode results into the probable loss of a High Voltage protection circuit with $\beta = 0.001$. For all other failure modes, $\beta = 0.01$.
  - Determine the criticality number for the 10 month (7200 hour) period of the system operating time.
- **Solution**
  - For short circuit: $C_m = (1 \times 10^{-6})(7200)(0.8)(0.001) = 5.76 \times 10^{-6}$
  - For the other failure modes: $C_m = (1 \times 10^{-6})(7200)(0.2)(0.01) = 1.44 \times 10^{-5}$

$$C_{varistor} = \sum C_m = 2.02 \times 10^{-5}$$

Criticality matrix

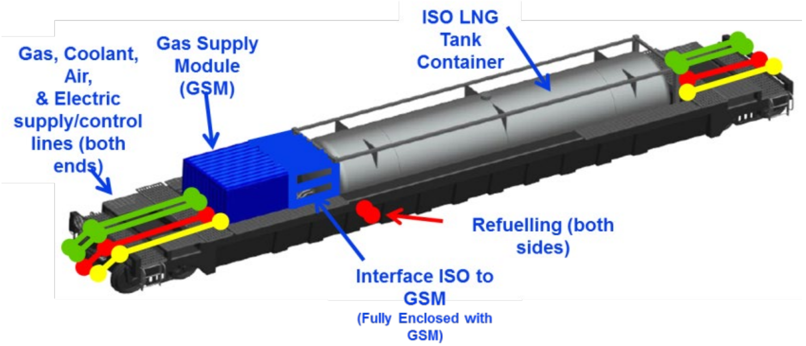
- A visual method to compare (and prioritize) the failures with respect to their severity and criticality (may also involved red/yellow/green coloring to support visualization.



FMEA- What are some attributes of success?

- Repeatability and traceability
  - Defined objectives and scope
  - Clear definitions of failure modes, consequences, the system, and criteria (or data used) to assign severity and likelihood
  - Final analysis reflects actual product (document the system)
  - Documentation of all of the above (and below)
- Diverse, representative team
  - Experts from all aspects of the system
    - Hardware, process, design, operations, maintenance
  - Experienced risk analysts
  - And beyond the team…honest reviewers
- To get there…follow a *rigorous* or *standardized* approach
  - Level of detail matched to type of technology, maturity, goals

Inspired in part by a true analysis



Case Study: Modified FMEA for LNG rail

System overview

- System elements:
  - LNG Tender (Intended to fuel an LNG/diesel dual fuel locomotive)
- Non-system elements: (Relevant for interfaces)
  - Track system
  - Human operators
  - Interface system
- Operating modes
  - Line-haul operation ("Operation phase")
  - Refueling
  - Maintenance
  - Storage

FMEA Plan for today's exercise

- **Goals**
  - Enhancing system safety by uncovering failure modes that result in hazardous conditions
  - Influencing the design to mitigate the impact of failure on the final product
- **Define method and ground rules**
  - Defined in next few slides + handout
- **Assemble the team**
  - Members from risk analysis, design, manufacturing, operations, maintenance, etc.
    - **Reflect: what role do you fill on your team?**
- **Assemble the information basis**
  - Illustrative example attached.
    - System diagrams, system descriptions, system breakdowns, data sources, hazard checklists, failure mode models (lists).

Basis and Assumptions

- Analysis focused on the Line-haul operation phase
  - System is expected to spend a majority of time in this phase.
  - Operation phase includes: mainline travel and siding. We are not addressing switching or classification.
- System definition was based entirely on readily available public information
  - Based on Canadian Patent published in 2013.
  - No detailed system information was otherwise available.
- Focused on the Tender system
- The only hazard considered is release of natural gas (in any form) from any part of the system. (Defines 'failure')

Hazard identification

- Hazard: "A condition or physical situation with a potential for harm" (SFPE) [or loss]
  - What *could* go wrong?
  - …And which ones are you including in the risk analysis?

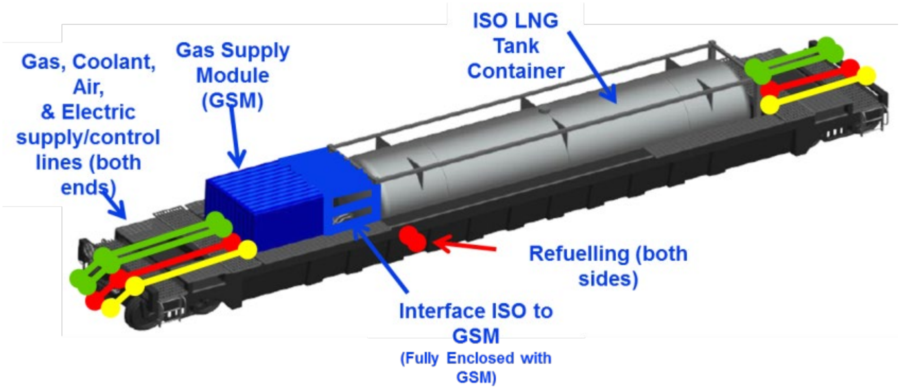
What are the hazards?	How do they manifest?
• Mechanical • Thermal • Chemical • Electrical • Biological • Radiation • Digital • …	• Pressure? Impacts? • Fire? Freezing? • Corrosion? Oxidation? • Toxicity? Tenability? • Bacteria, virus, plant? • ..

Which of these apply for LNG/GNG?

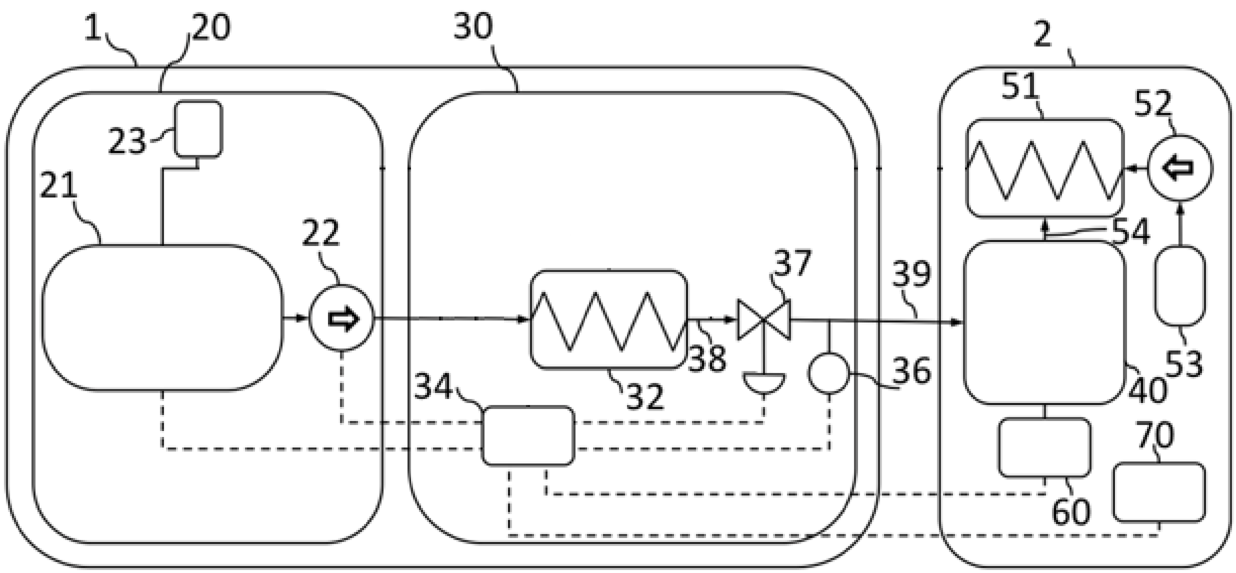
Hazards for the LH2 tanker

- Mechanical
  - Effects of overpressure (direct or indirect)
  - Impact from debris/projectiles
- Thermal
  - Heat flux (from various types of fires and smoke)
  - Freezing from exposure to cryogenic fluids
- Chemical
  - Tenability (asphyxiation (From NG or from smoke))

System drawing



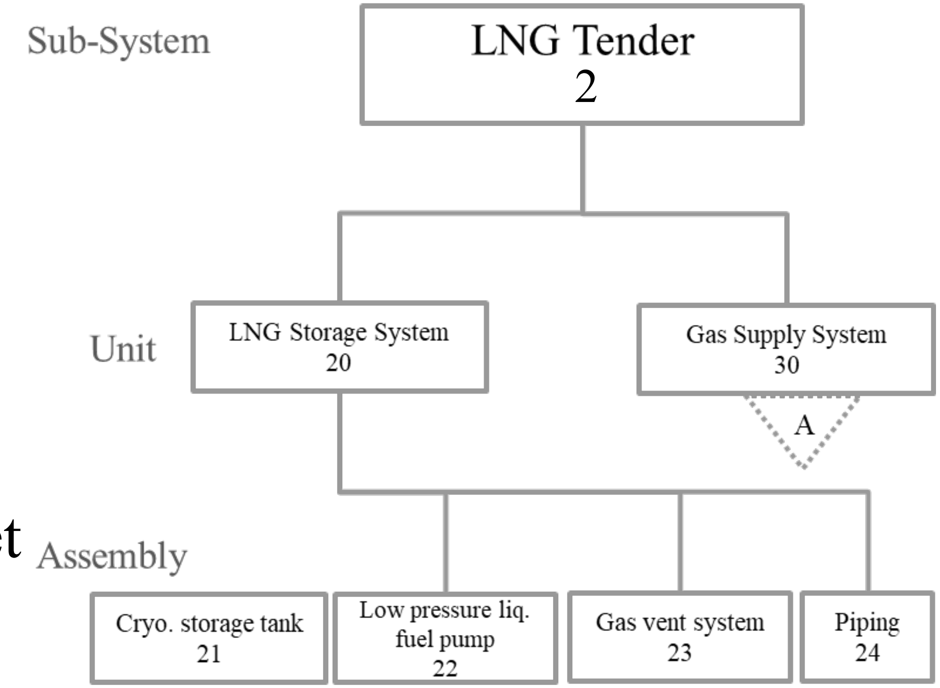
- See handout for a P&ID
- The drawing has been modified and simplified from the original to create a more simplified example for discussion purposes



System diagram source: Canadian Patent Document 2762697- Figure 1 – "Preferred embodiment." Full description of systems, parts, and interfaces provided in patent document

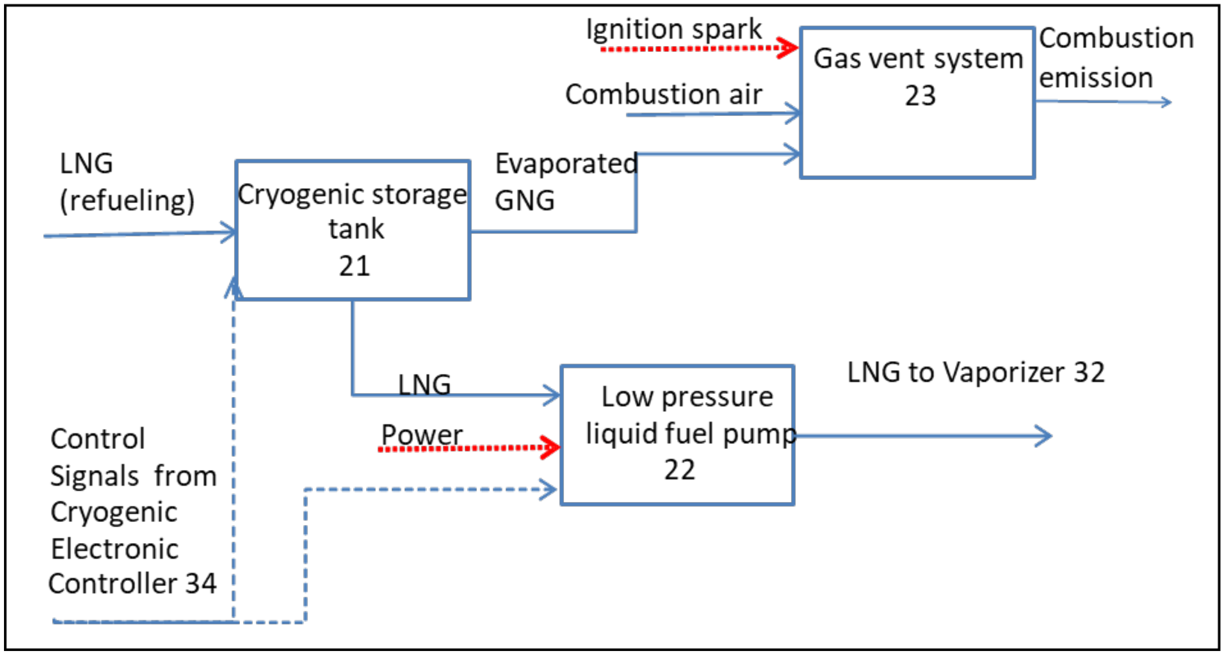
Break into 3 teams

- **Team 21** Cryogenic storage tank
- **Team 22** Low pressure liquid storage pumps
- **Team 23** Gas vent system
- Over the next several slides: you will fill in the elements of the ARP5580 FMEA worksheet for your component.
- Assemble results from each group to get the “final” FMEA for the class.



***Notice hierarchical numbering system in an FMEA**

LNG storage system block diagram



Identification of failure modes

- Systematic analysis of each assembly level component.
- Examination of assembly inputs and outputs

Failure Mode Models
Premature operation Failure to operate at prescribed time Failure to cease operation at prescribed time Failure to meet functional specifications Failure conditions caused by the operational environment

- Focused on failures that could lead to a release of GNG or LNG
- Note (Beyond current scope: Credible failure scenarios could also be identified from the Reliability Information Analysis Center (RIAC) Non-Electric Parts Reliability Database (NPRD) and Failure Mode Distribution (FMD) and previously published FMEAs of LNG vehicles and facilities.)

Identification of Failure Effects

- Failure Effects analyzed by identifying the consequence of each failure mode on operation of the assembly operation as well as the next higher indenture level
- The end effect on the system should also be identified
  - Consider specifically whether LNG or GNG could potentially be released in an uncontrolled manner

Severity Class	Criteria: Severity of Effect
1. Minor	No potential release of LNG or GNG (e.g., from failure of a component that does not process LNG or GNG)
2. Moderate	Potential leak or small-scale release of LNG or GNG (e.g., from a leaking seal, breach of line carrying vented GNG)
3. Critical	Potential for catastrophic release of LNG or GNG (e.g., from a breach of a line carrying LNG, from a rupture of storage tank, from failure of a tank relief valve)

Probability characterization

- To characterize the probability that a given failure mode occurs, it is common to use a failure rate as an approximation
- Failure Rate Source Priority
  - Field Data from exact equipment in exact environment
  - Failure rates from similar systems
  - Tables of generic component failure rates
- (RIAC) Non-Electric Parts Reliability Database (NPRD) was used as a standardized approach for estimating failure rates for the assemblies in this analysis
- Assumed to be constant over lifetime of component

Failure rate calculation

- Failure Rate: $\lambda = \alpha\beta t\lambda_p$ where
  - λ_p is the failure rate for all failure modes for a specific component
  - α is the fraction of component failure corresponding to the failure mode
    - (Ignore this part for the class exercise example; data not provided; just calculate it as $\frac{1}{\#\ failure\ modes\ you've\ identified}$).
    - $\beta, t\colon treat\ as\ 1\ for\ this\ exercise$
- $\lambda_p(per\ million\ hrs) = \frac{Number\ of\ failures}{Total\ hours\ (million)}$
- If no failures were reported, a Jeffrey's prior = 0.5 failures (half of an event) was used
- For failure modes involving an accident:

Railroad Accident Failure Modes and Frequencies

	Human Factors	Track and Infrastructure Defects	Rolling Stock Defects	Miscellaneous Causes	Total
Cars Derailed per Million Car-Miles	0.055	0.151	0.128	0.055	0.389

Probability classes

- The calculated failure rates were grouped into ranges to identify a qualitative characterization
- An order of magnitude scale was used
- A qualitative class was chosen due to uncertainty of failure rates from applying rates for similar equipment in different environments

Probability Class	**Criteria: Failure Rate**
High	$\lambda > 10.0$ per Million Hours or Million Track Miles
Medium	$\lambda =$ between 1 and 10 per Million Hours or Million Track Miles
Low	$\lambda < 1.0$ per Million Hours or Million Track Miles

Risk Priority

- Simple 3x3 Matrix
- Used to prioritize the failure events

Probability Class	High	M	H	H
	Medium	L	M	H
	Low	L	L	M
		Minor	Moderate	Critical
		Severity Class		

FMEA Conclusions

- Four of the failure modes were scored as high risk in the risk priority matrix in (LaFleur et al 2017):
  - **21.01—An over pressurization of the LNG storage tank due to failure of the relief valve,** either by failure to open or failure to vent at the rate of methane boil-off. Because the relief valve penetrates both inner and outer tanks, it is a single point failure mechanism that could lead to the uncontrolled release of the LNG tank contents.
  - **21.02—A leak of LNG due to a failure of a fitting or outlet in the LNG tank**. The fitting or outlet failure could be due to mechanical, installation or material defect or mechanical damage. Detailed specifications for the ports and outlets on the tank have not been designed; however, they also represent single points of failure through the double-walled tank that could lead to a release of the bulk of the LNG contents.
  - **21.07—This failure mode involved the embrittlement or cracking of the outer tank due to leakage or failure of the inner tank**. The outer tank is typically made of carbon steel and is not rated for cryogenic storage. Although this is a compound failure mode, requiring a failure of two components, it could result in a release of the tank contents and should be targeted for engineered safety in the design process.
  - **22.04—This failure mode involves a liquid LNG fuel pump scenario that experiences cavitation**. Cavitation is especially dangerous as it involves localized areas of low pressure which could cause boiling of the LNG and ultimately lead to rupture of the pump resulting in uncontrolled release of LNG or an explosion.

LaFleur, C. B.; Muna, A. B.; Groth, K. M.; St. Pierre, M. & Shurland, M. “Failure analysis of LNG rail locomotives.” *Proceedings of the 2017 Joint Rail Conference (JRC2017),* The American Society of Mechanical Engineers (ASME), 2017

Reliability Analysis

Module 6D: Bayesian Networks

Prof. Katrina M. Groth
Mechanical Engineering Department
Center for Risk and Reliability
University of Maryland
kgroth@umd.edu

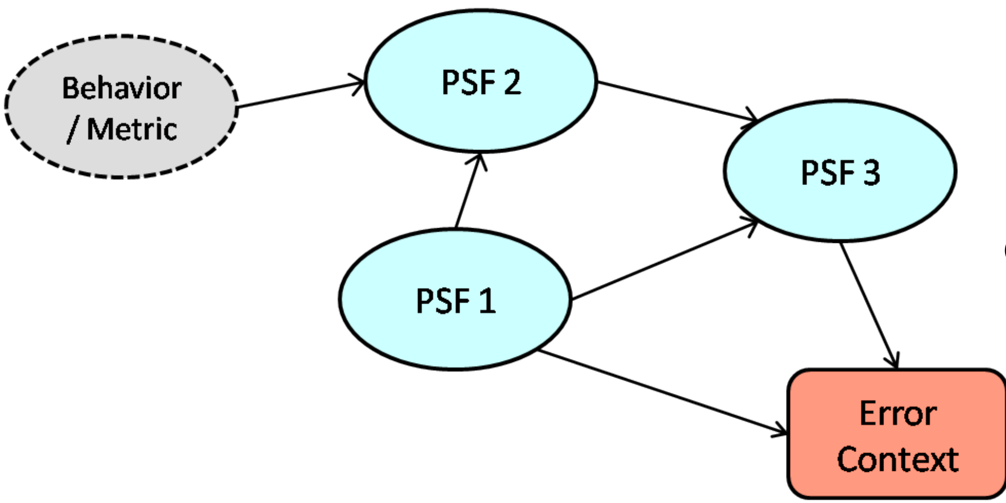
Opening thought

Probability is not really about numbers;
it is about the structure of reasoning.

Glenn Shafer as quoted in Judea Pearl, *Probabilistic Reasoning in Intelligent Systems*, 1988

Preview: Bayesian Networks

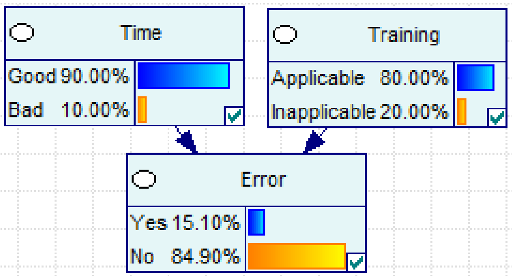
- In this module, we use probability to encode a knowledge base and conduct **probabilistic reasoning (with uncertainty)**
- …With a tool called a **Bayesian Network (BN)**



$P(EC \cap PSF1 \cap PSF2 \cap PSF3 \cap BM)$

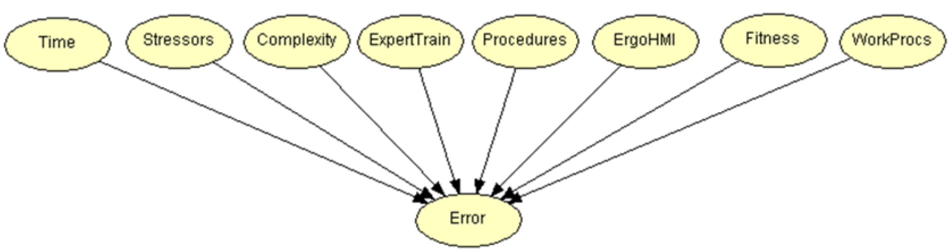
	Parent	$\boldsymbol{Pr(a)}$	$\boldsymbol{Pr(\overline{a})}$
Child	$\Pr(b)$	$\Pr(b\|a)$	$\Pr(b\|\overline{a})$
	$\Pr(\overline{b})$	$\Pr(\overline{b}\|a)$	$\Pr(\overline{b}\|\overline{a})$

Parent	$Pr(a)$	$Pr(\overline{a})$
$Pr(b)$	$Pr(b\|a)$	$Pr(b\|\overline{a})$
$Pr(\overline{b})$	$Pr(\overline{b}\|a)$	$Pr(\overline{b}\|\overline{a})$



Outline

- What & Why of BNs
- Building a BN
- Inference with BNs
- HRA Example
- A few BNs for PRA and safety
- Wrap-up



Terminology, abbreviations, notation

- Joint distribution: $Pr(A \cap B) = Pr(A, B)$
- Marginal (unconditional) distribution: $Pr(A)$
- Conditional distribution: $Pr(A|B)$

- BN: Bayesian (Belief) Network
- HCL: Hybrid causal Logic
- HRA: Human Reliability Analysis
- PRA: Probabilistic Risk Assessment
- PSF: Performance Shaping Factor

Caution

- "Bayesian" "Bayesian updating" is a general term
- Bayesian Networks are not in the textbook
- Bayesian parameter estimation ≠ Bayesian Network
  - Bayesian parameter estimation is used extensively in reliability
  - Bayesian Networks are prevalent in computer science, artificial intelligence, medicine, etc.; gaining popularity in probabilistic risk assessment (PRA) (especially human reliability analysis, HRA)

Fundamental goal (ENRE, and also BNs)

- Enable decision makers to make better decisions under uncertainty
- Uncertainty
  - Due to imperfect understanding of the system
  - Due to incomplete knowledge about the state of the system (current or future)
  - Due to "inherent randomness" in the system behavior
  - Due to changing conditions after data is collected.

  - And….How do we deal with uncertainty? (Probability)

Some BN application areas (way out of date by now)

- Medicine
  - CHILDE: Congenital heart disease diagnosis
  - MUNIN: Preliminary Diagnosis of neuromuscular diseases
  - SWAN: System for insulin adjustment for diabetics
  - PATHFINDER: Diagnosis of breast cancer
- Business and Management
  - Market forecasting in oil industry
  - Finance-Fraud/Uncollectible debt collection
  - Modeling impact of organizational change
- Engineering and Science
  - Diagnosis of faults in waste-water treatment process
  - Failure mode and effect analysis with BBN's
  - BaNTERA: Pipeline third-party excavation damage risk assessment
  - Many applications in human reliability analysis

How are they used in our field?

- To build a defensible probability distribution for hard-to-quantify problems (e.g., HRA, aging, software)
  - To break problems down into quantifiable (or elicitable) chunks
  - To add additional levels of detail and traceability
  - To address dependency
- To enable use of *some data* for problems where the alternative is no data (e.g., HRA)
- To enable *appropriate* use of experts (appropriate experts, appropriate probability elicitation)
- To provide causal understanding, not just statistics

Bayesian Network: A tool & a model

- A model which…
  - Explicitly encodes relevant variables & dependencies
  - …In terms of a simplified probability distribution
  - Permits multiple types of data/information to be used in a single reasoning framework
- A tool for **reasoning under uncertainty**
  - Conducting inference (reasoning from cause to effect) and diagnosis (reasoning from effect to cause)
  - About uncertain states, with limited information, under changing conditions

What do BNs do for us?

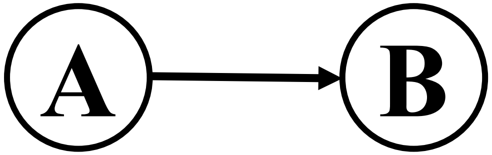
- Provide a framework for reasoning about uncertain events
- BN offers a way to assess probabilities based on partial information
  - Analyst is not required to assess the state of unknown variables
- Produces results that are:
  - Reproducible
  - Supportive by a broad base of knowledge
- Expandable in scope and depth

Definition of a BN

Probability theory + Graph theory

Aka a type of **Probabilistic Graphical Model (PGM)**

- A BN is characterized by 3 elements:
  - A **directed acyclic graph (DAG)** with nodes (Vertices) representing random variables
  - Arcs (Edges) which represents probabilistic influence
  - Each node has an associated probability distribution (usually discrete) – used to generate a joint probability of the nodes.



$\Pr(A)$ $\Pr(B|A)$

Recall: $P(A \cap B) = \Pr(B|A)\Pr(A))$

BNs: Modeling aspects

- **Building the model**: Aka, Learning, Elicitation, and Parameterization
  - Identify the nodes & define states
  - Structure the DAG
  - Assigning Conditional Probability Distributions (CPDs)
- **Using the model**: Inference, Belief propagation
  - The process of computing the [updated] probability distribution, given the evidence
  - Occurs via inference algorithms, e.g., exact inference, sampling

(Bayesian) Reasoning with a BN

- The fully quantified model represents the entirety of the *prior* information available to the analyst
- The analyst *sets evidence* about the state of one or more variables
- The model updates the probability of the rest of the network (the *posterior*)

So why are these Bayesian?

- Judea Pearl says:

"Bayes means:

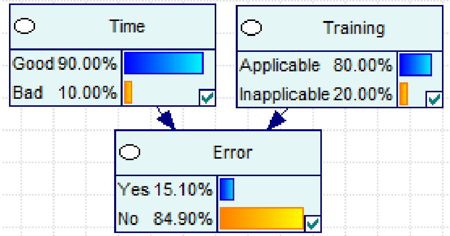
1. Using knowledge we possess prior to obtaining data,
2. Encoding such knowledge in the language of probabilities
3. Combining those probabilities with data and
4. Accepting the combined results as a basis for decision making and performance evaluation."

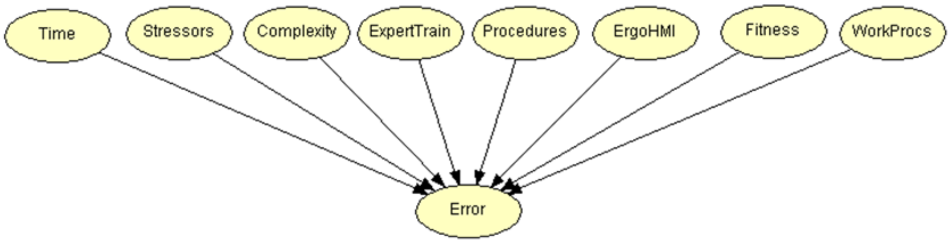
Judea Pearl, "Bayesianism and causality, or, why I am only a half-Bayesian" *Foundations of Bayesianism*, **2001**, *24*, 19-34

Outline

- What is a BN?
- Building a BN
- Inference with BNs
- HRA Example
- A few BNs for PRA and safety
- Wrap-up

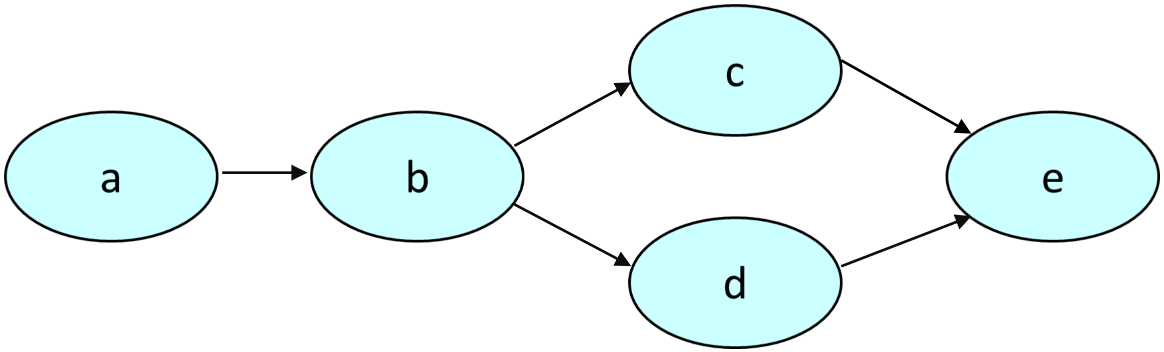
Parent	$Pr(a)$	$Pr(\bar{a})$
$Pr(b)$	$Pr(b\|a)$	$Pr(b\|\bar{a})$
$Pr(\bar{b})$	$Pr(\bar{b}\|a)$	$Pr(\bar{b}\|\bar{a})$





BN pieces

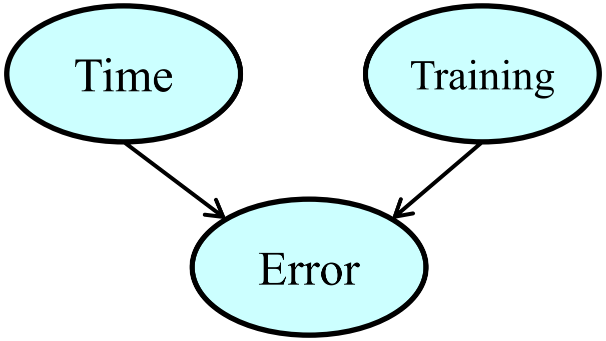
- BN encodes
  - Relevant variables and their states
  - (In)dependency among variables
  - The simplified joint probability distribution of the system



$$Pr(a,b,c,d,e) = Pr(e|a,b,c,d) \cdot Pr(d|a,b,c) \cdot Pr(c|a,b) \cdot Pr(b|a) \cdot Pr(a)$$
$$= Pr(e|c,d) \cdot Pr(d|b) \cdot Pr(c|b) \cdot Pr(b|a) \cdot Pr(a)$$

BN language

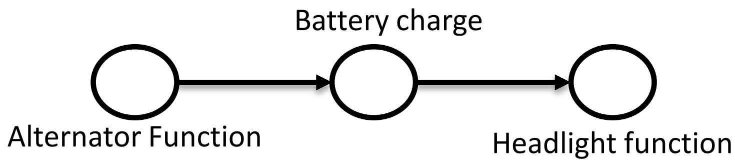
- Consider the following simple net



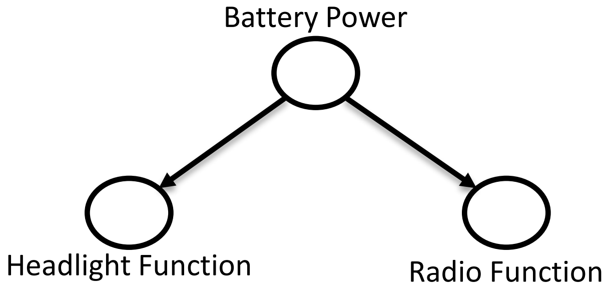
- Nodes *Time* and *Training* are **parent nodes** for node *Error; Error* is their **child node**.
- *Time* and *training* are also (they have no parents)
- *Time* and *training* are <u>conditionally</u> independent

Types of structure in a BN

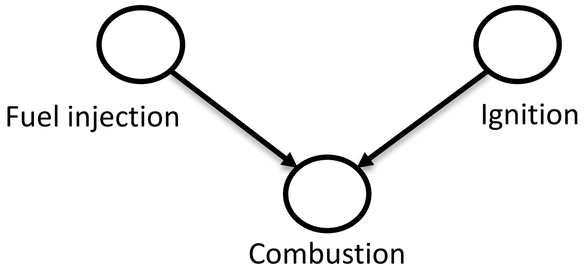
1. Indirect connection



2. Common Cause (Parent)



3. Common Effect (Child)

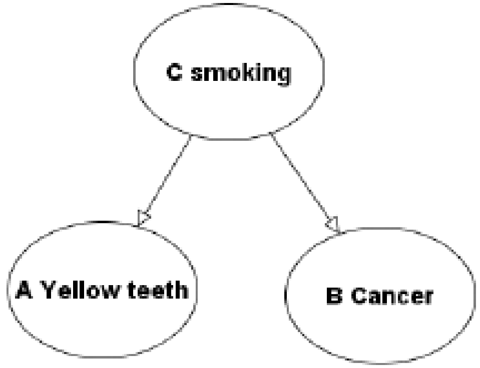


Example: BN causal structures

- **Example**: Draw the causal BN model in which nodes A, B, and C correspond respectively to "yellow teeth", "Cancer", "smoking".

Example: BN causal structures

- **Solution**: Draw the causal BN model in which nodes A, B, and C correspond respectively to "yellow teeth", "Cancer", "smoking".



- Notice:
  - Direction of arcs from C to A, B reflects causality
  - Lack of arc between A, B shows lack of direct causal relationship

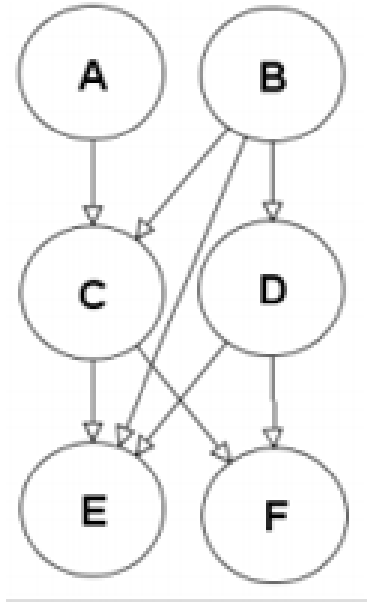
Underlying formulas

- Remember: $0 \le Pr \le 1$ and $\Sigma P(universe) = 1$

Law of Total Probability	$$\boldsymbol{Pr(a_i) = \sum_j Pr\,(a_i \cap b_j)}$$ **Marginalizes** out variables
Chain Rule (of probability)	$Pr(X_n \cap X_{n-1} \cap \cdots \cap X_2 \cap X_1) =$ $P(X_n \vert X_{n-1}, \ldots X_2, X_1) \cdot P(X_{n-1} \vert \ldots X_2, X_1) \cdot P(X_2 \vert X_1) \cdot P(X_1)$ **Factorizes** a joint probability into conditional probabilities
Chain Rule (of BNs) (The above, with conditional independence)	If A and B are independent… $Pr(A \vert B) = Pr(A)$ $and\ thus$ $Pr(A \cap B) = Pr(A) \cdot Pr(B)$ $$Pr(X_1, X_2, \ldots X_n) = \prod_i Pr(X_i \vert Par_G(X_i))$$
Bayes' Theorem	$$\Pr(X \vert E) = \frac{\Pr(E \vert X)\Pr(X)}{\Pr(E)}$$ Allows forward and backward **propagation of evidence**

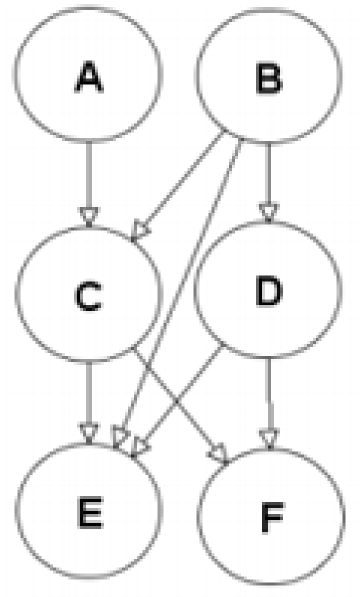
Example: BN causal structures

- **Example**: For the following BN models involving 6 variables A,B,C,D,E,F, write the expression for the full simplified joint probability distribution Pr(A, B, C, D, E, F)



Example: BN causal structures

- **Solution**: For the following BN models involving 6 variables A,B,C,D,E,F, write the expression for the full simplified joint probability distribution Pr(A, B, C, D, E, F)



$$\boldsymbol{Pr(A, B, C, D, E, F)}$$
$$= \boldsymbol{Pr(A)Pr(B)Pr(C|A, B)Pr(D|B)Pr(E|B, C, D)Pr(F|C, D)}$$

To find the probability of error = “yes”

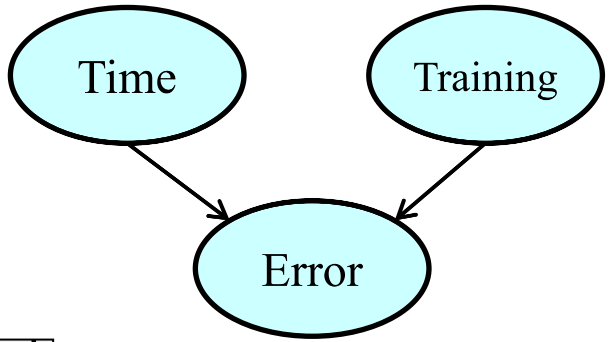
- Computational Steps:
  1. List all the combinations of the states of its parents
  2. Assign the probabilities to all states of these combinations.
  3. For each combination of states assign the conditional probability* of the states of the child given the states of its parents
  4. Compute the marginal probability of the child.

* The conditional probabilities are interpreted as the degree of influence of various states of the parents on the states of error.

Mathematical formalism

- Assume binary states for all nodes (for now)
  - Time

Good	0.9
Bad	0.1



  - Training

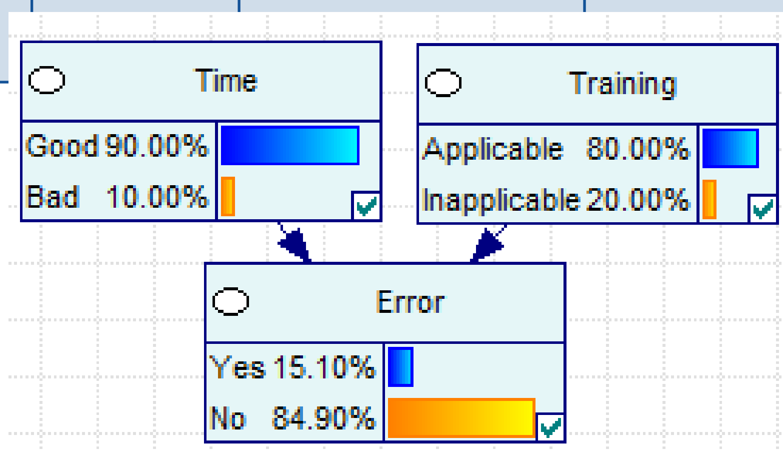
Applicable	0.8
Inapplicable	0.2

  - Error

Time	Good		Bad	
Training	Applicable	Inapplicable	Applicable	Inapplicable
Yes	0.1	0.25	0.3	0.5
No	0.9	0.75	0.7	0.5

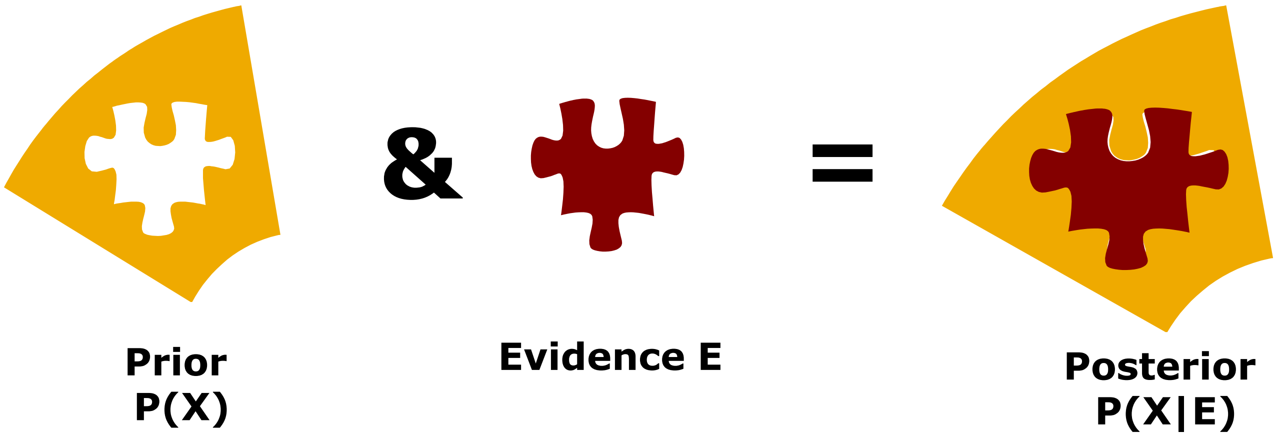
To find the probability of error = “yes”

Time	**Training**	**Probability of Combination**	**Conditional Probability of Error = YES**	**Unconditional Probability of Z = z**
Good	Applic.	$= .9 \times .8$	0.1	$p_1 = .9 \times .8 \times .1 = 0.072$
Good	Inapplic.	$= .9 \times .2$	0.25	$p_2 = .9 \times .2 \times .25 = 0.045$
Bad	Applic.	$= .1 \times .8$	0.3	$p_3 = .1 \times .8 \times .3 = 0.024$
Bad	Iapplic.	$= .1 \times .2$	0.5	$p_4 = .1 \times .2 \times .5 = 0.010$
				$Pr = \sum p_1 = 0.151$



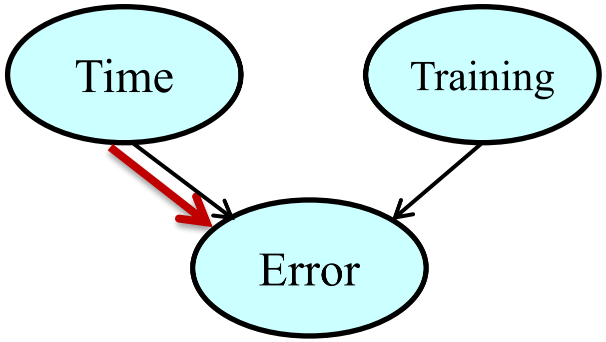
Inference in a BN

- The fully quantified model represents the entirety of the prior information available to the analyst
- The analyst makes an observation about the state of one or more variables
- We calculate the posterior probability of the rest of the network

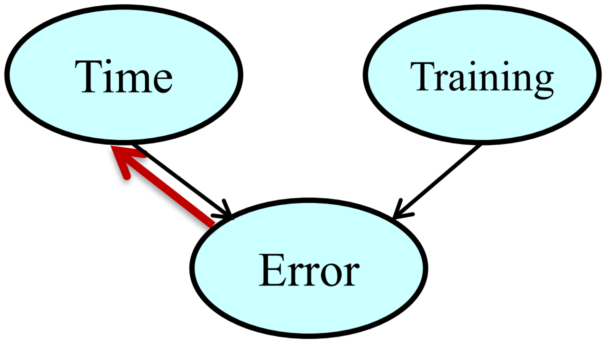


Types of reasoning/inference

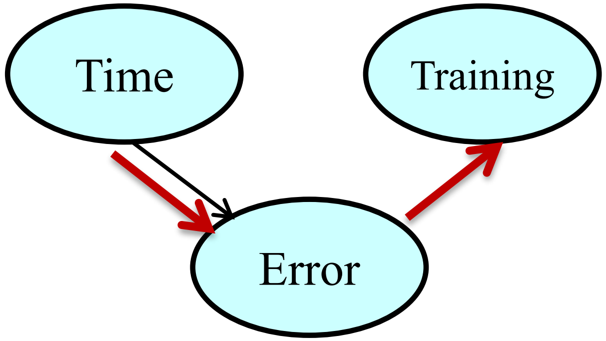
- **Causal**: (Forward propagation; Induction)



- **Evidential**: (Backward propagation; Diagnosis)



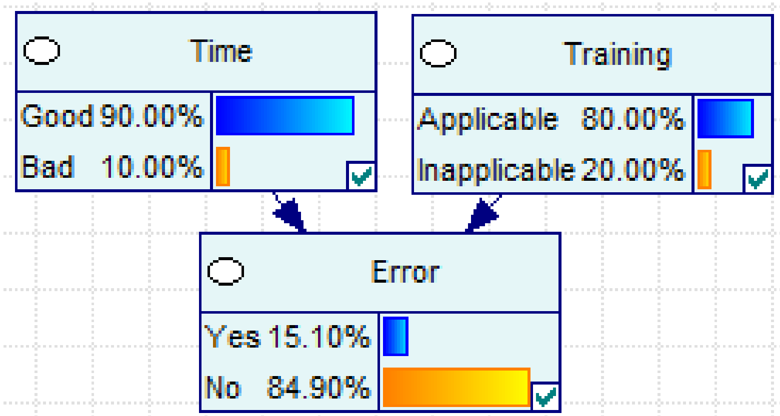
- **Intercausal**: (Across nodes)



Forward reasoning

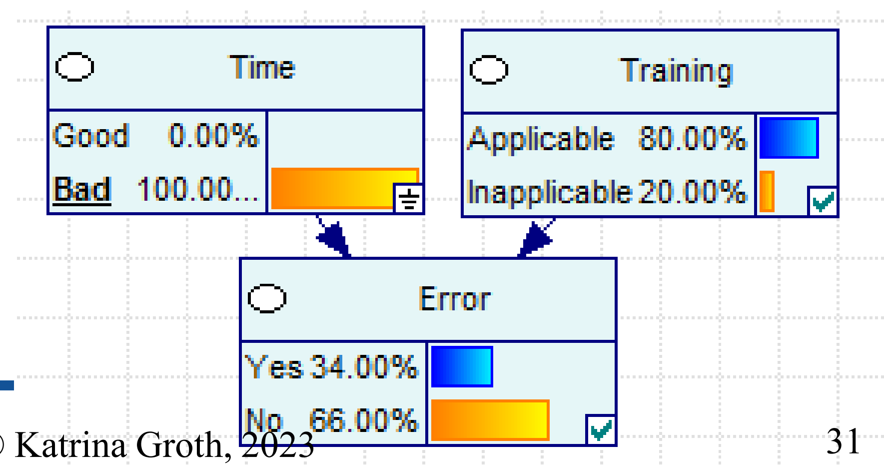
- Observing *Time* = *Bad* changes belief about error (Pr(Yes) goes from .151 to .34)

Prior: (Before)



Evidence: **Observation:** Time = Bad

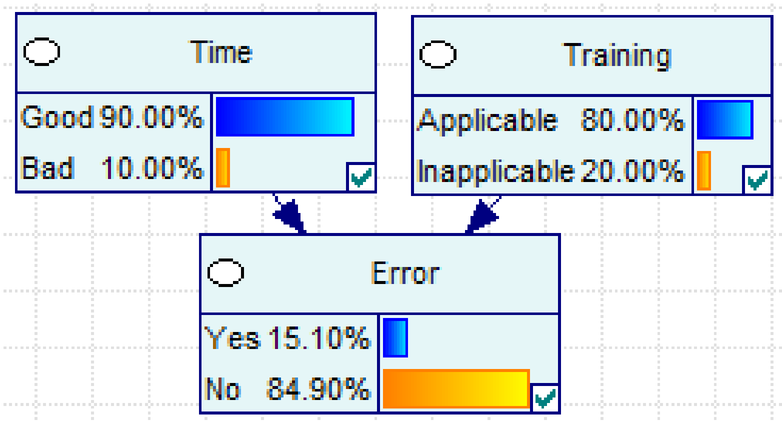
Posterior: (After)



Backward reasoning

- Observing *Error = yes* changes belief about both time and training

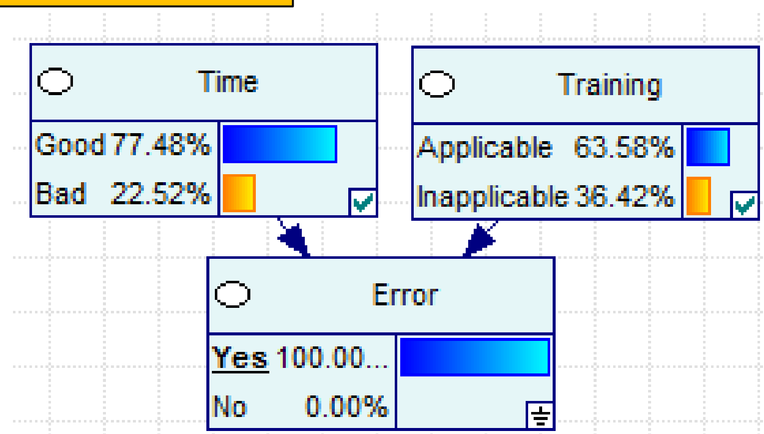
Prior:
(Before)



Evidence:

Observation: Error = yes

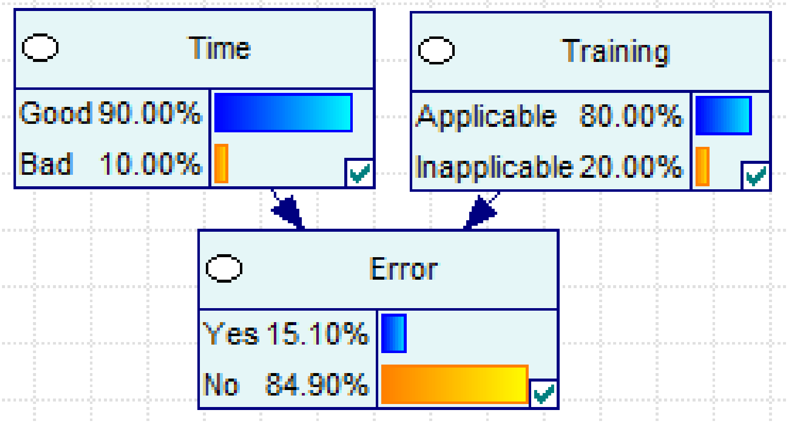
Posterior:
(After)



Intercausal reasoning (both)

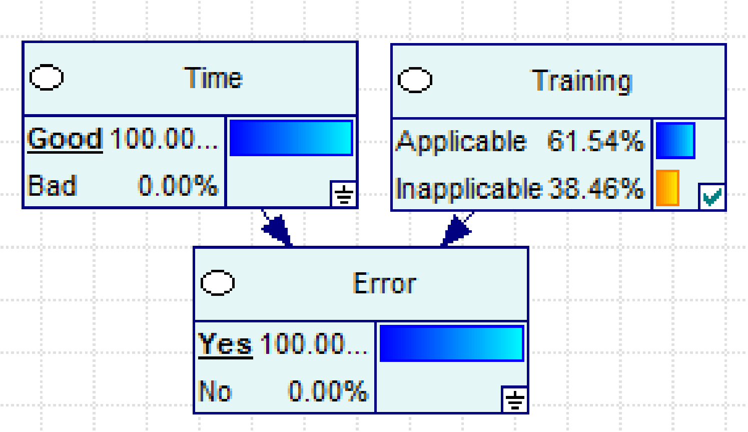
- Observing *Error = yes and Time = Good* changes belief about training

Prior:
(Before)

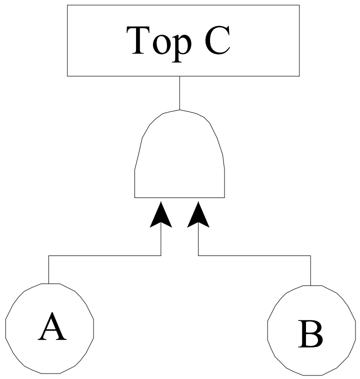


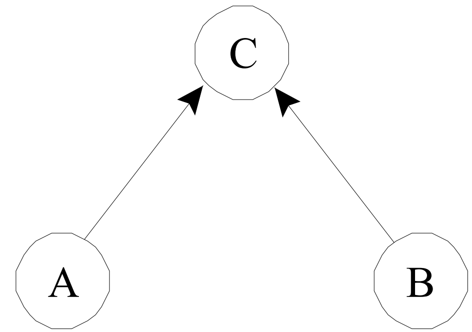
Evidence: **Observations:** Error = yes; Time = Good

Posterior:
(After)



How to implement in PRA- Option 1: Replace Fault/Event Trees with BNs



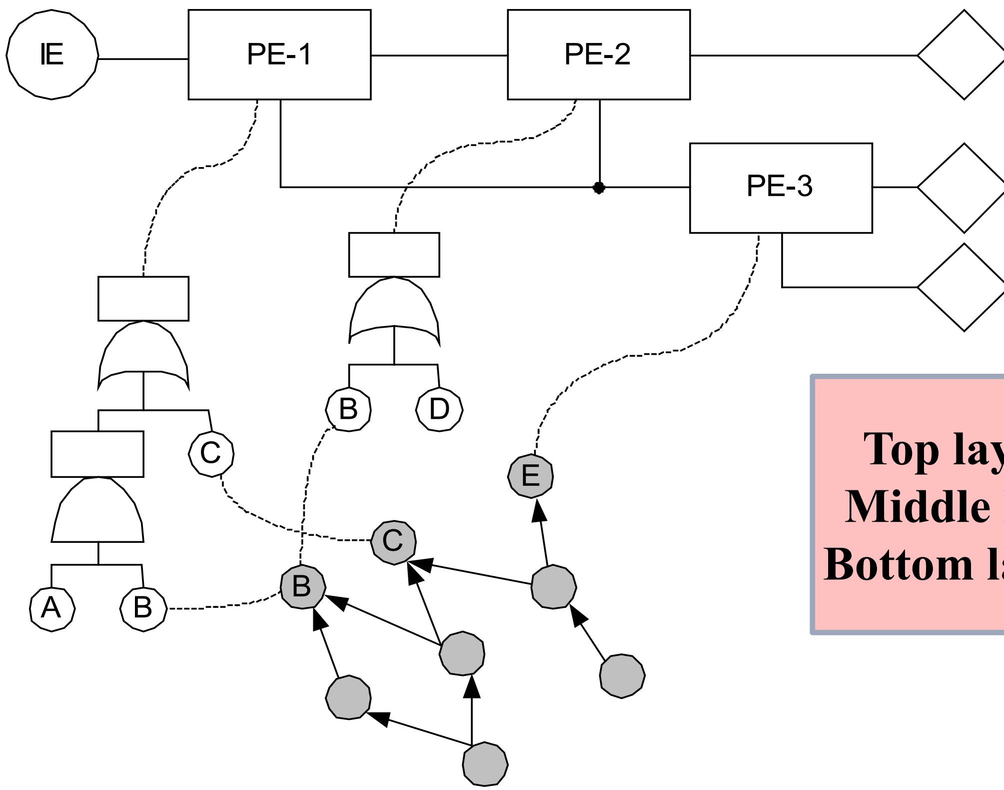


$$\Pr(C = 1|A = 0, B = 0) = 0$$
$$\Pr(C = 1|A = 0, B = 1) = 0$$
$$\Pr(C = 1|A = 1, B = 0) = 0$$
$$\Pr(C = 1|A = 1. B = 1) = 1$$
$$\Pr(C = 1|A = 1, B = 1) = 1$$

$$\Pr(C) = \Pr(A)\Pr(B)$$

$$\Pr(C)$$
$$= \Pr(C|A,B)\Pr(A,B) + \Pr\left(C\middle|\overline{A},B\right)\Pr\left(\overline{A},B\right)$$
$$+ \Pr\left(C\middle|A,\overline{B}\right)\Pr\left(A,\overline{B}\right) + \Pr\left(C\middle|\overline{A},\overline{B}\right)\Pr\left(\overline{A},\overline{B}\right)$$
$$= \Pr(A,B)$$

How to implement in PRA-the better option: HCL/Trilith: Adds BNs to the PRA Framework



Top layer: ESD
Middle layer: FT
Bottom layer: BBN

Groth, Katrina; Wang, Chengdong & Mosleh, Ali. Hybrid causal methodology and software platform for probabilistic risk assessment and safety monitoring of socio-technical systems. *Reliability Engineering and System Safety,* **2010**, *95*, 1276-1285

Example HRA method: SPAR-H

1. Assess context in terms of PSFs (Performance Shaping Factors
   - Available time
   - Stress/stressors
   - Complexity
   - Experience/training
   - Procedures
   - Ergonomics/HMI
   - Fitness for duty
   - Work processes
2. Calculate HEP (Human Error Probability)

$$HEP = NHEP \cdot \prod_{i=1}^{8} PSF_i$$

Where NHEP = 0.01 for diagnosis tasks and 0.001 for action tasks

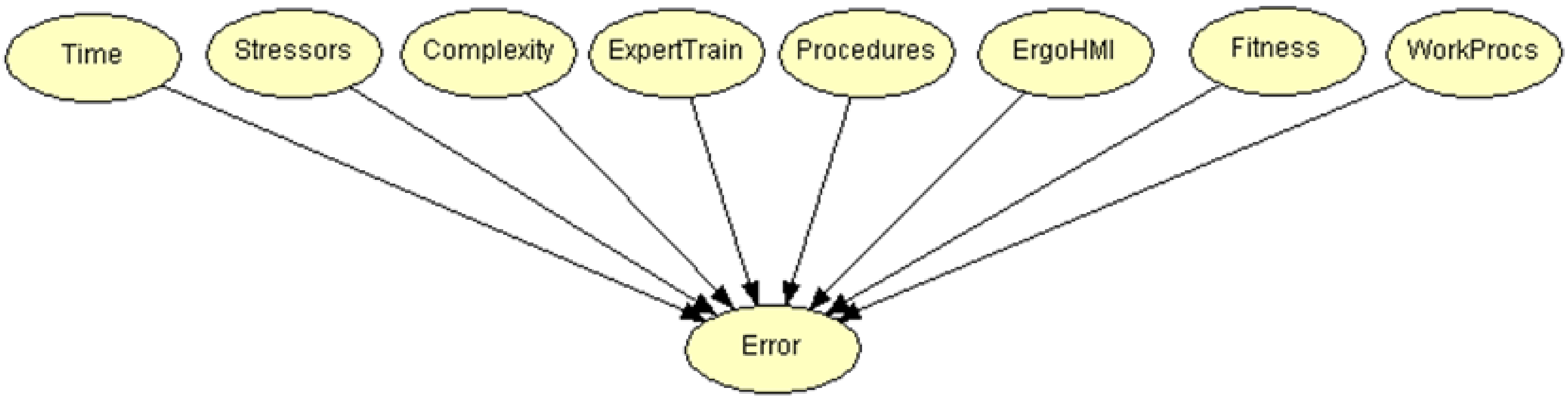
A. Evaluate PSFs for the Action Portion of the Task, If Any.

PSFs	PSF Levels	Multiplier for Action
Available Time	Inadequate time	P(failure) = 1.0
	Time available is ≈ the time required	10
	Nominal time	1
	Time available ≥ 5x the time required	0.1
	Time available is ≥ 50x the time required	0.01
	Insufficient Information	1
Stress/ Stressors	Extreme	5
	High	2
	Nominal	1
	Insufficient Information	1
Complexity	Highly complex	5
	Moderately complex	2
	Nominal	1
	Insufficient Information	1
Experience/ Training	Low	3
	Nominal	1
	High	0.5
	Insufficient Information	1
Procedures	Not available	50
	Incomplete	20
	Available, but poor	5
	Nominal	1
	Insufficient Information	1
Ergonomics/ HMI	Missing/Misleading	50
	Poor	10
	Nominal	1
	Good	0.5
	Insufficient Information	1
Fitness for Duty	Unfit	P(failure) = 1.0
	Degraded Fitness	5
	Nominal	1
	Insufficient Information	1
Work Processes	Poor	5
	Nominal	1
	Good	0.5
	Insufficient Information	1

Challenges for SPAR-H

- Poor handling of uncertainty
  - Is “unknown” really equivalent to “nominal”?
- Poor credibility (HRA in general, not just SPAR-H)
  - In PRA, data is used to build credibility/confidence;
  - Very few HRA methods use data, and those that do use very little data
  - Use of expert-elicited probabilities
    - System expert ! = probability expert
- Traceability
  - Tenuous link between inputs and outputs
    - Subjective—your “high stress” might be my “low stress” situation

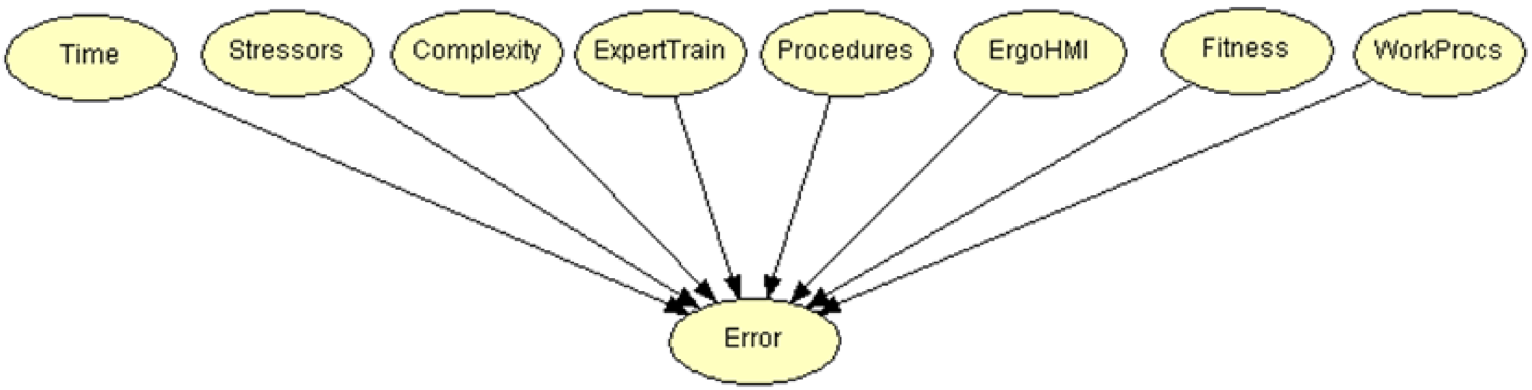
SPAR-H BN: Structure



- SPAR-H method:
  - 8 PSFs affect error probability
  - PSFs act independently on error (margin independence)
  - Interdependency among PSFs is acknowledged, but not modeled.

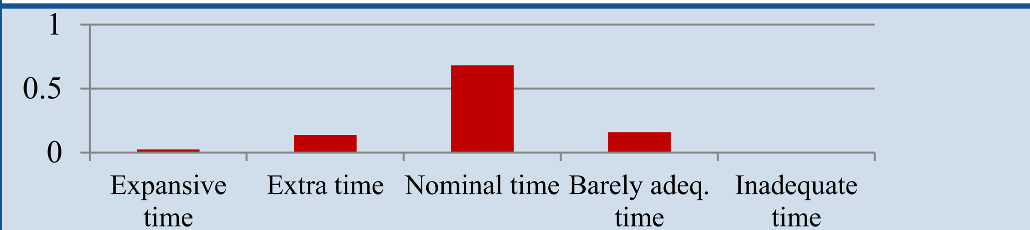
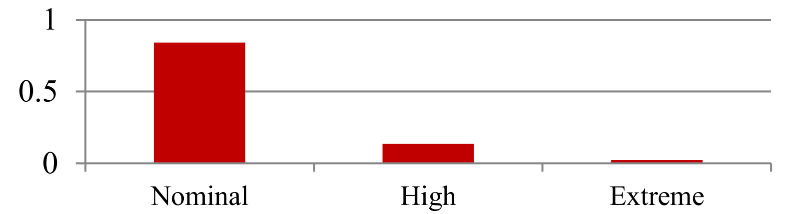
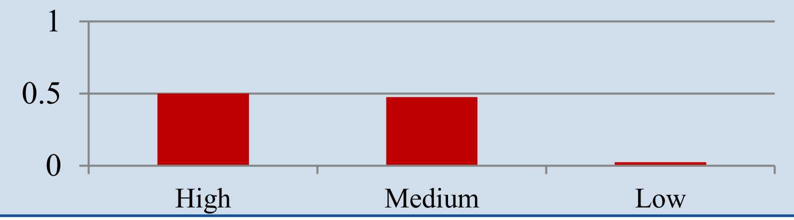
Quantification

- Factorizing the joint distribution allows us to specify different parts of the model using different sources of information (including data)



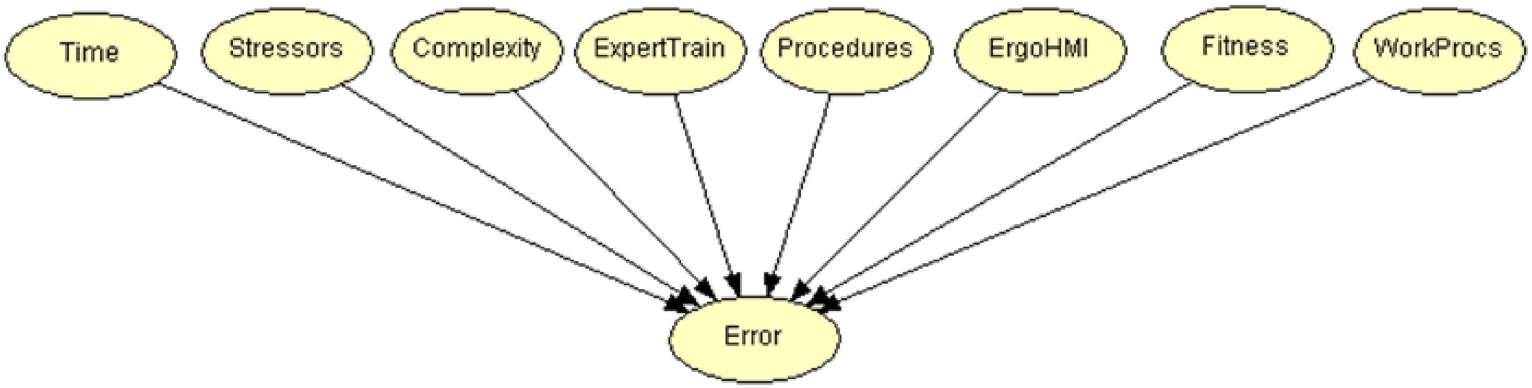
$$
\begin{aligned}
&P(Error) \\
&= \sum_{PSFs} [P(Error|Time, Stress, Complexity, ExpertTrain, Procedures, ErgoHMI, Fitness, WorkProcs) \times P(Time) \\
&\times P(Stress) \times P(Complexity) \times P(ExpertTrain) \times P(Procedures) \times P(ErgoHMI) \times P(Fitness) \\
&\times P(WorkProcs)]
\end{aligned}
$$

Quantification: P(PSFs)

PSF	Source	Probability Distribution
P(Time) 5 States	NUREG/CR-6949	
P(Stress) 3 States	NUREG/CR-6949	
P(ExpertTrain) 3 States	Curve fit (Available from plant data)	

- Similar NUREG/CR-6949 values for P(Complexity), P(Procedures), P(ErgoHMI), P(Fitness), P(WorkProcs)
- Next steps: Adding simulator data to this model (ask me after class)

HRA: BN version of SPAR-H



Prior: SPAR-H

Data: simulator

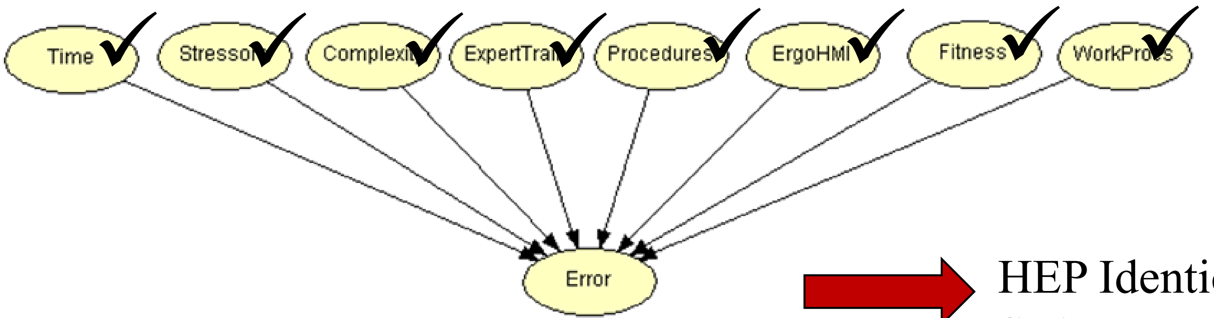
$$P(Error) = \sum_{PSFs} P(Error|Time, Str, Compl, Expert, Procs, HMI, Fit, WPs) *$$

$$\underbrace{\ldots P(Time) * P(Stress) * P(Complexity) * P(Expert) * P(Procs) * P(HMI) * P(Fit) * P(WPs)}_{Priors: Experts; Industry\ data}$$

*Groth, Katrina M. & Swiler, Laura P. Bridging the gap between HRA research and HRA practice: A Bayesian Network version of SPAR-H. Reliability Engineering and System Safety, **2013**, 115, 33-42.*

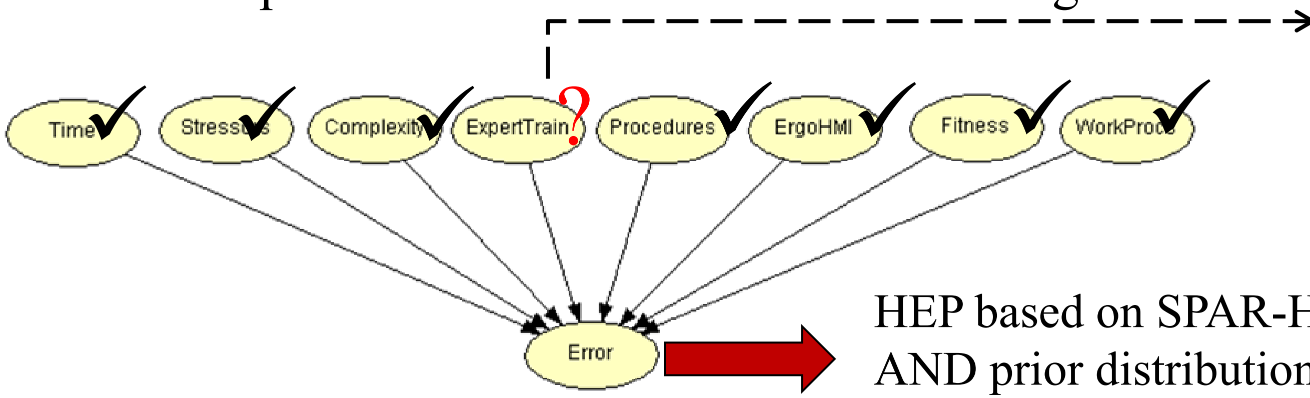
BN benefits: Addresses uncertainty

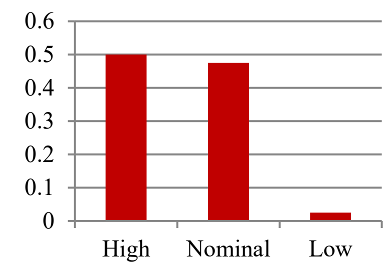
- Certainty cases



HEP Identical to SPAR-H model

- "Insufficient information" cases:
  - Uses prior distribution rather than assuming nominal

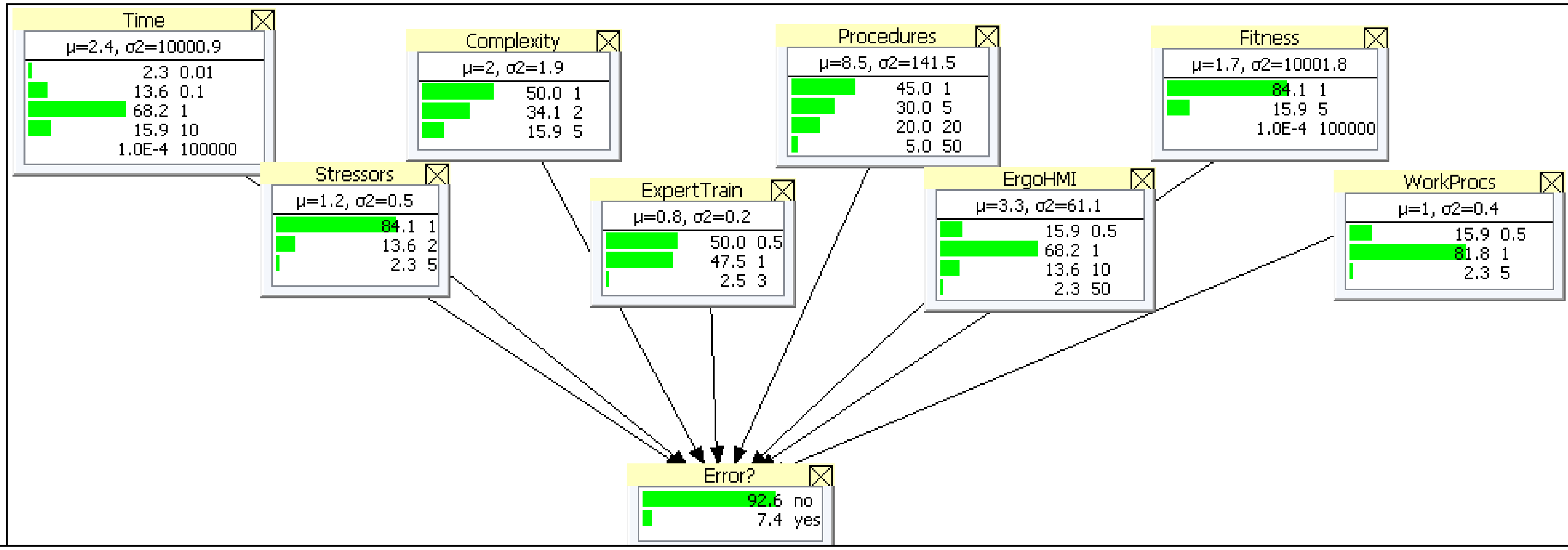




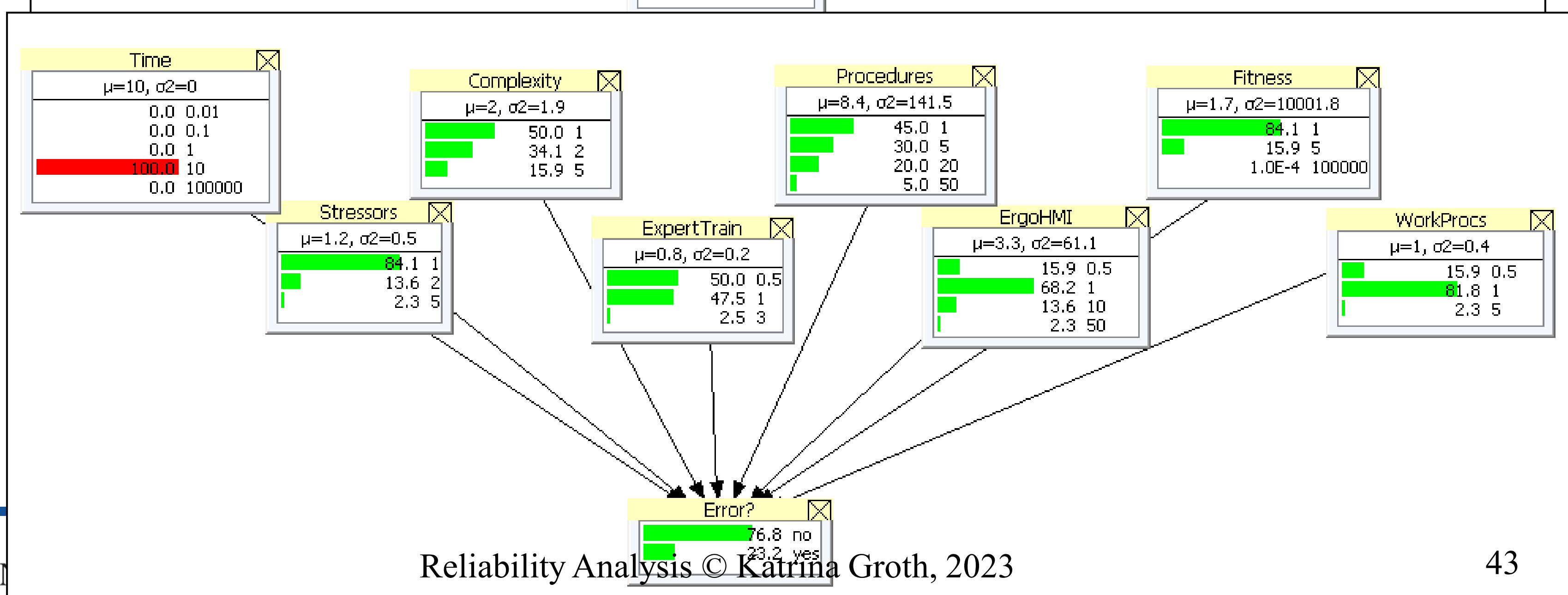
HEP based on SPAR-H model AND prior distributions of unknowns

causal reasoning (just like SPAR-H)

Prior

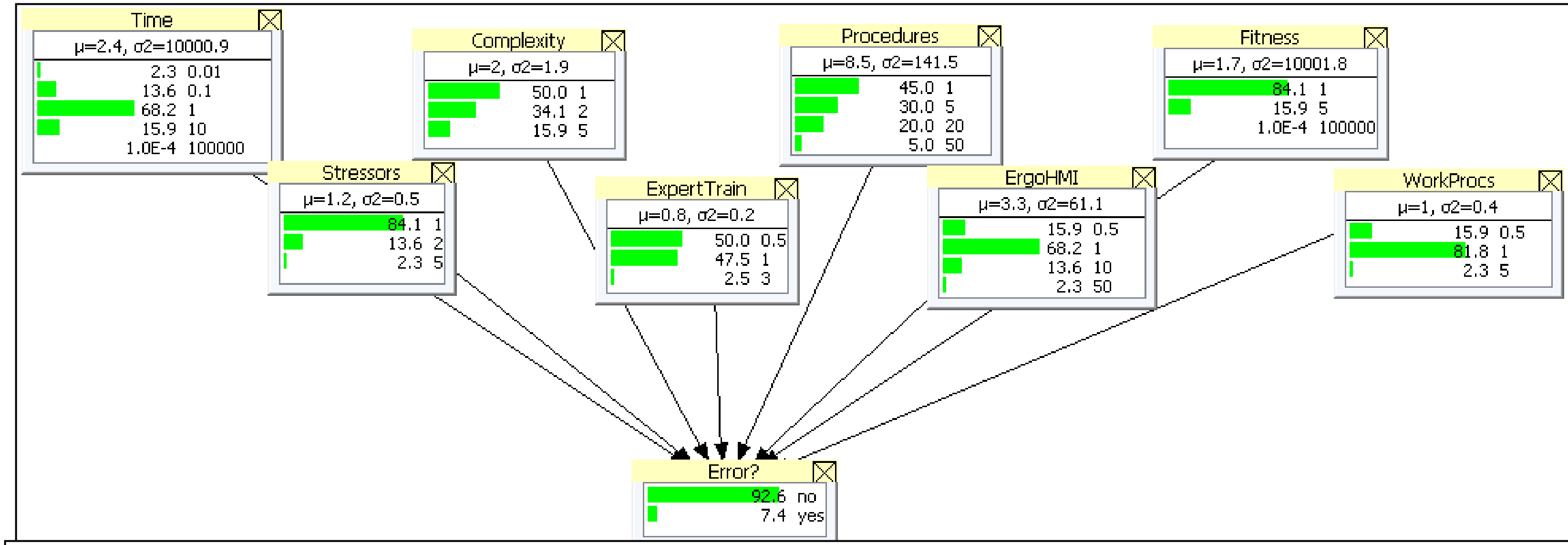


Posterior

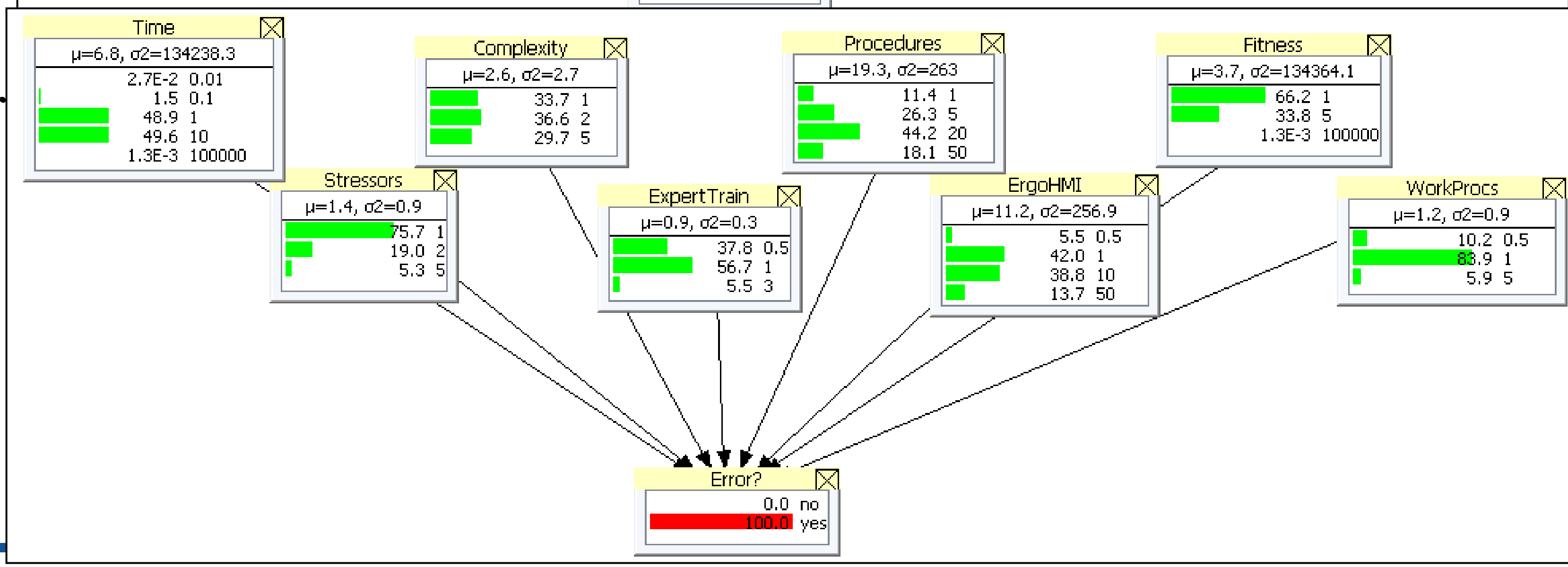


Evidential reasoning (new, powerful!)

Prior

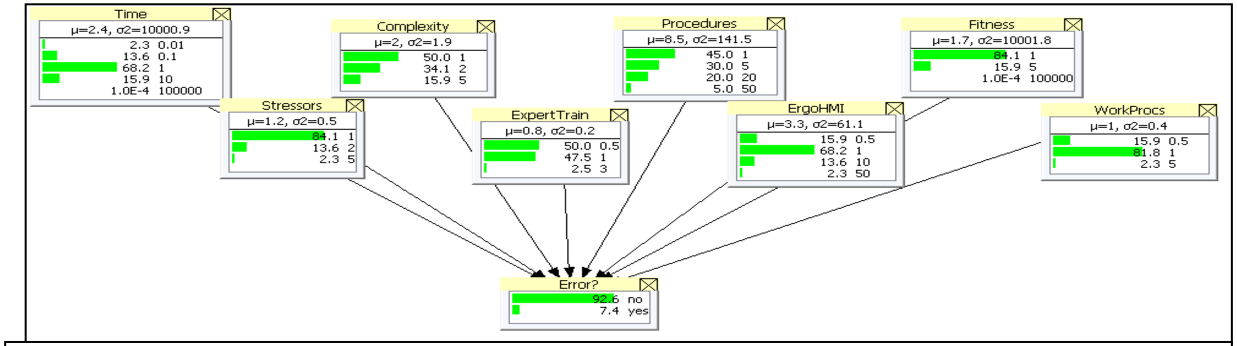


Posterior

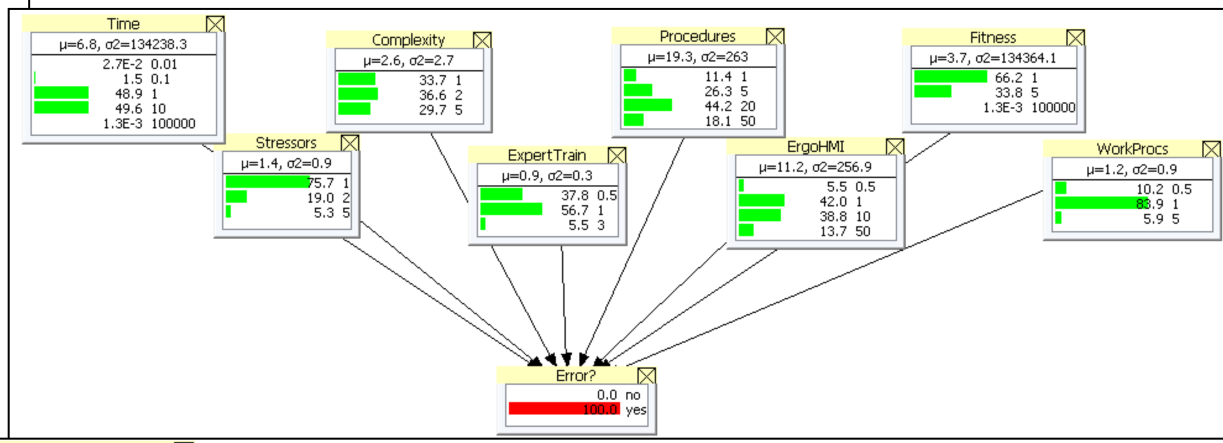


Intercausal reasoning (explaining away)

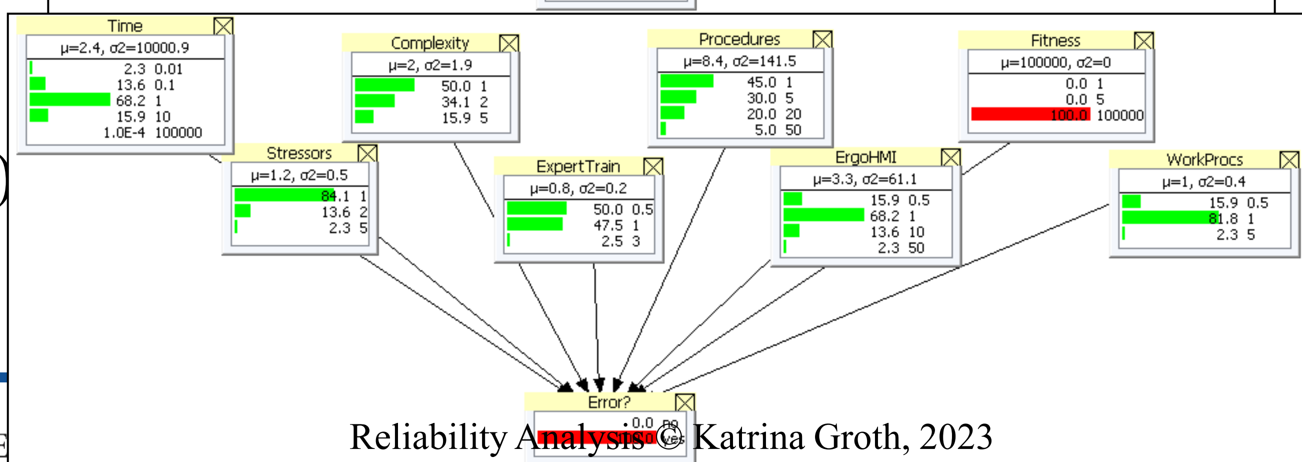
Prior



Posterior (1)

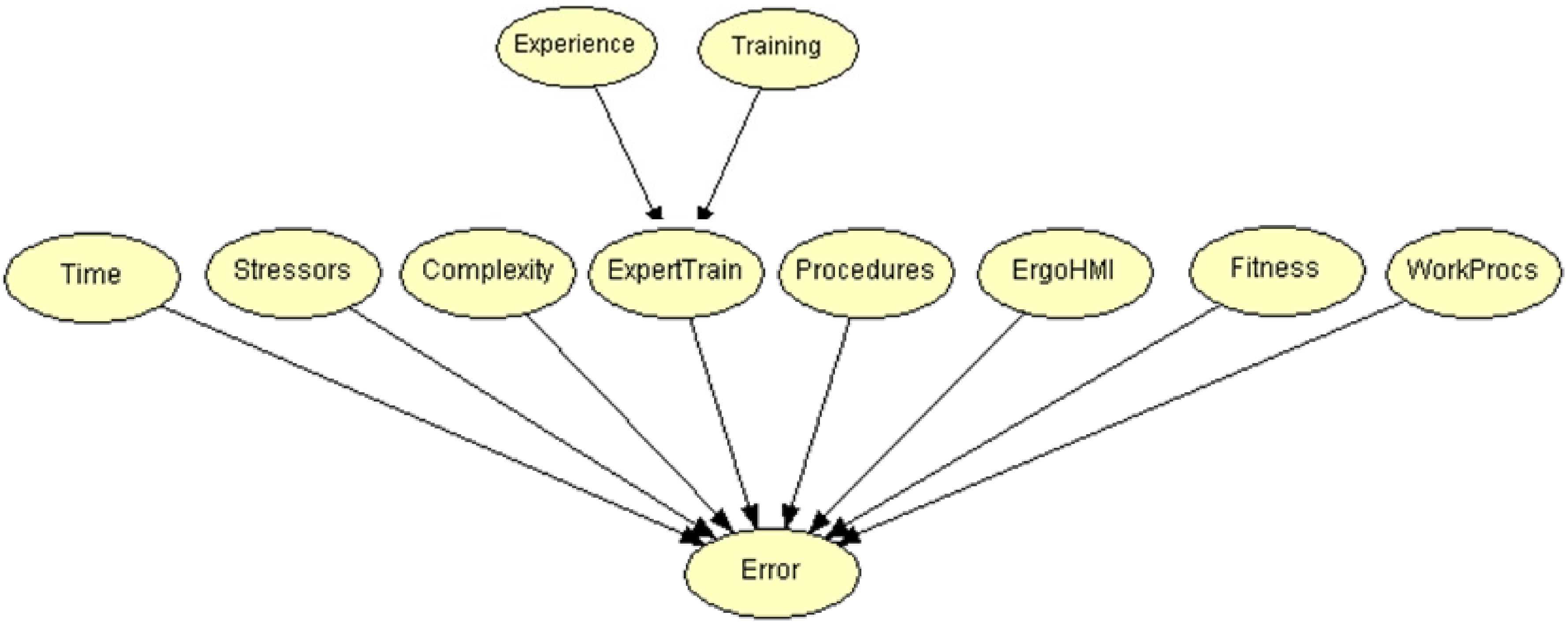


Posterior (2)



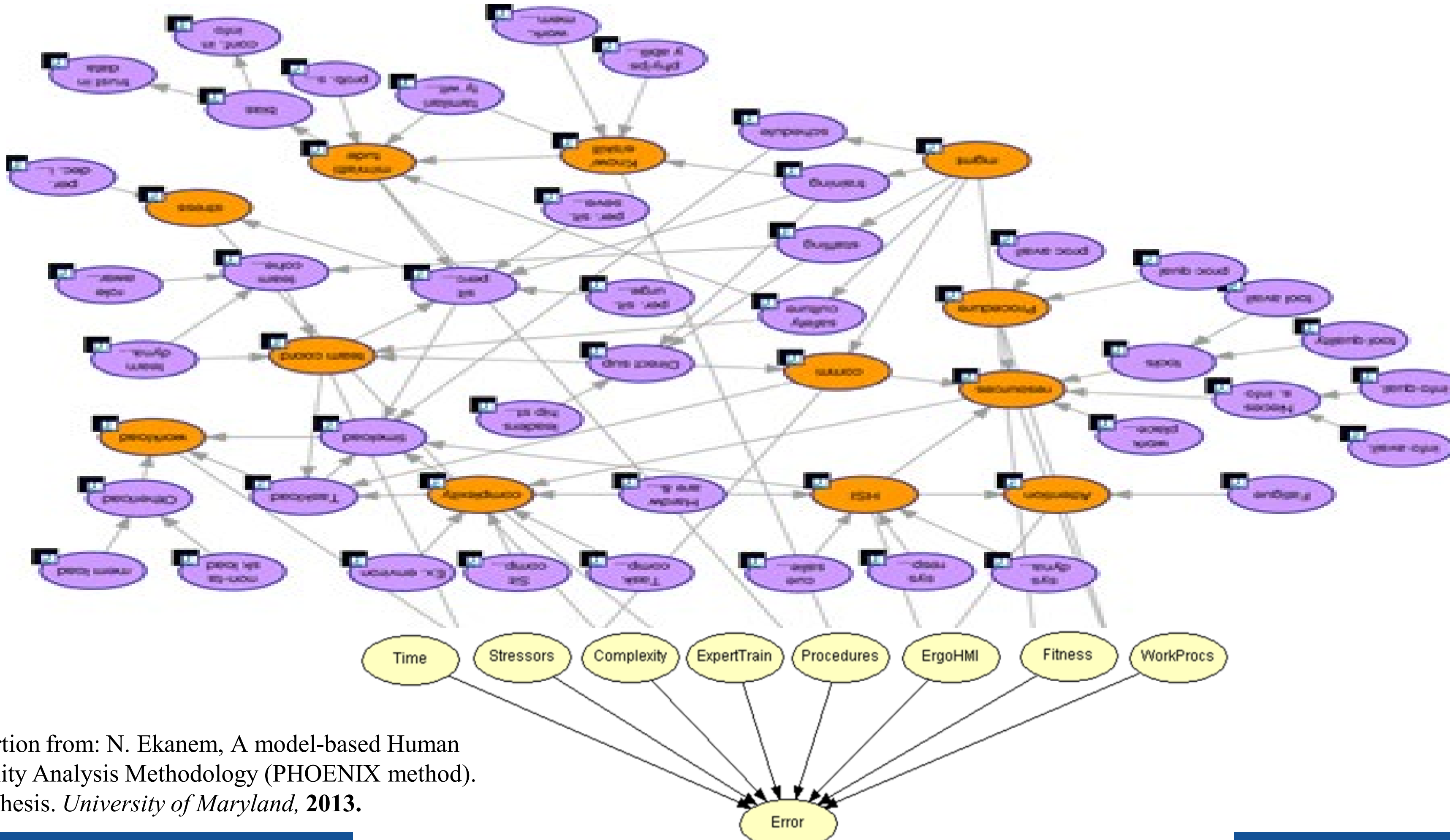
Next steps: Extending the model

- Expanded model using additional Performance Influencing Factors (PIFs)



Even deeper: Internal algorithm

- Linking more detailed questions (purple) to the PSFs



Top portion from: N. Ekanem, A model-based Human Reliability Analysis Methodology (PHOENIX method). Ph.D. Thesis. *University of Maryland,* **2013.**

BN benefits for HRA

- Enhance HRA technical basis
  - Opportunity to build model with multiple types of information/data (existing HRA methods, Halden, cognitive, operation experience, etc.)
  - Results are reproducible
- Expandable in scope and depth
  - Supports better PSF assignments by plant analysts
- Enhance usability
  - Allows analysis with partial information
    - Analyst is not required to assess the state of unknown variables
  - Seamless integration with software
    - Users see structured list of questions instead of complicated model

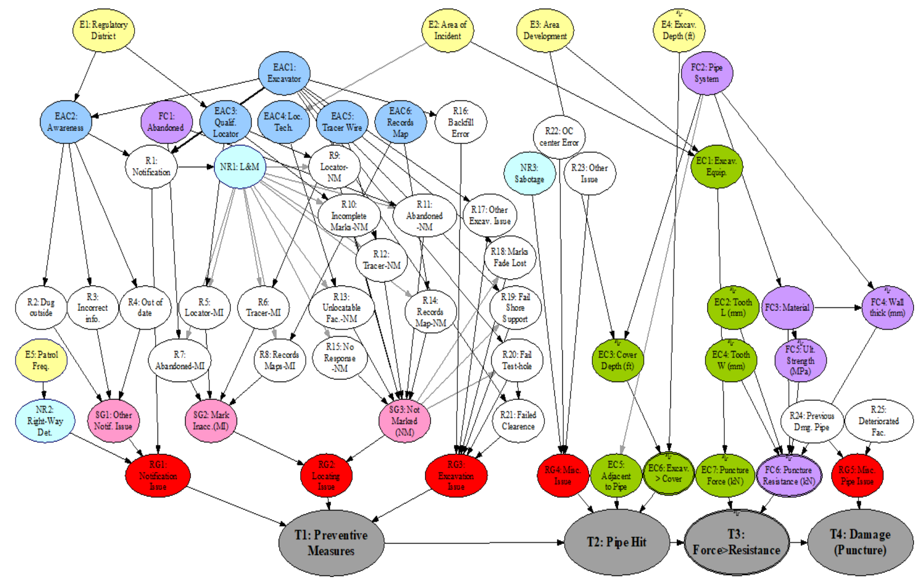
Implications for HRA

- Using HRA data adds credibility
  1. It's possible to use HRA data to update existing HRA methods
  2. It is inconsistent with PRA practice to NOT update HRA methods
  3. You don't need perfect HRA data to do the updating, if you separate out the probabilities of the PSFs from the probabilities of error, given PSFs.
- Expanding causal details adds traceability
  1. Adding plant-specific details makes it easier to assign PSF states (reduces subjectivity of PSF assignments)
  2. Also adds value to users—more detailed identification of ways to prevent human errors

BaNTERA: a Bayesian network for third-party excavation risk assessment

Decision support tool for third-party damage (TPD) risk

- Extensive causal structure including excavation process, context, and failures
- Enables heterogenous data fusion:
  - Customer-specific dig-in data
  - PHMSA historical incidents
  - CGA DIRT database
  - GTI models and experts

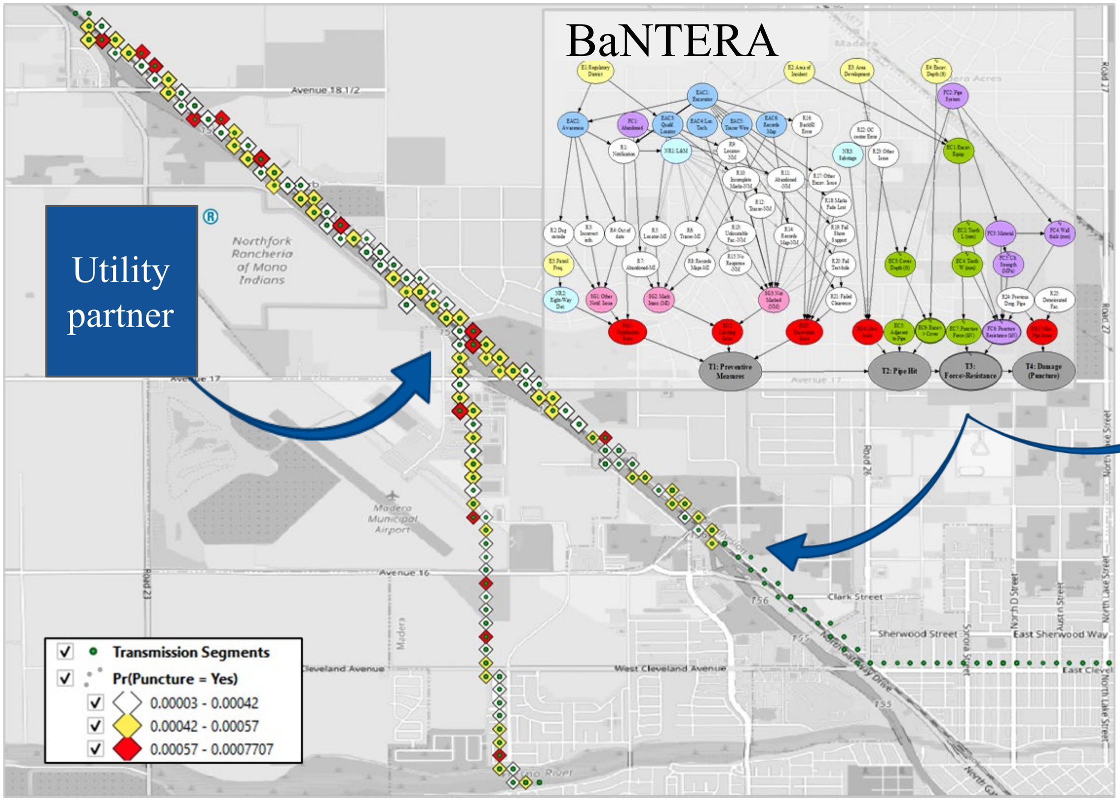


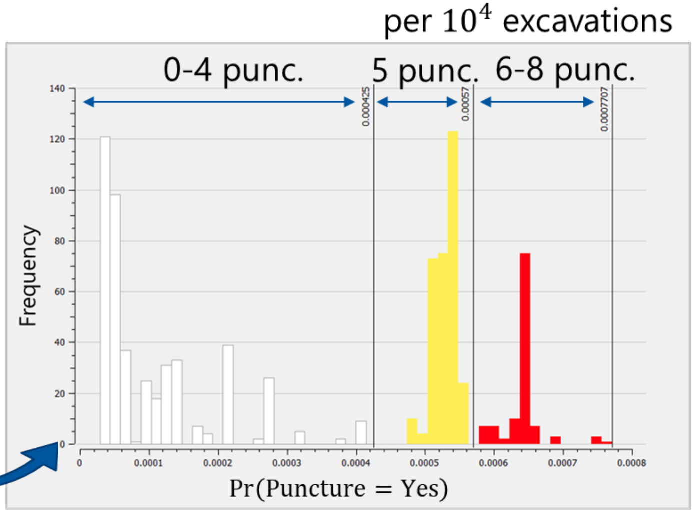
64 variables, 101 causal relationships

- First-of-kind model for integrating information about TPD causes and probabilities.
- Usable for quantifying risk – *and* identifying how to reduce it

A. Ruiz-Tagle, A.D. Lewis, C. Schell, E. Lever, & K. M. Groth. “BaNTERA: A Bayesian Network for Third-Party Excavation Risk Assessment” *Reliability Engineering & System Safety*, July 2022, 223.

BaNTERA applications: GIS-level pipeline risk assessment





Validated against nationwide hits per 1,000 notifications

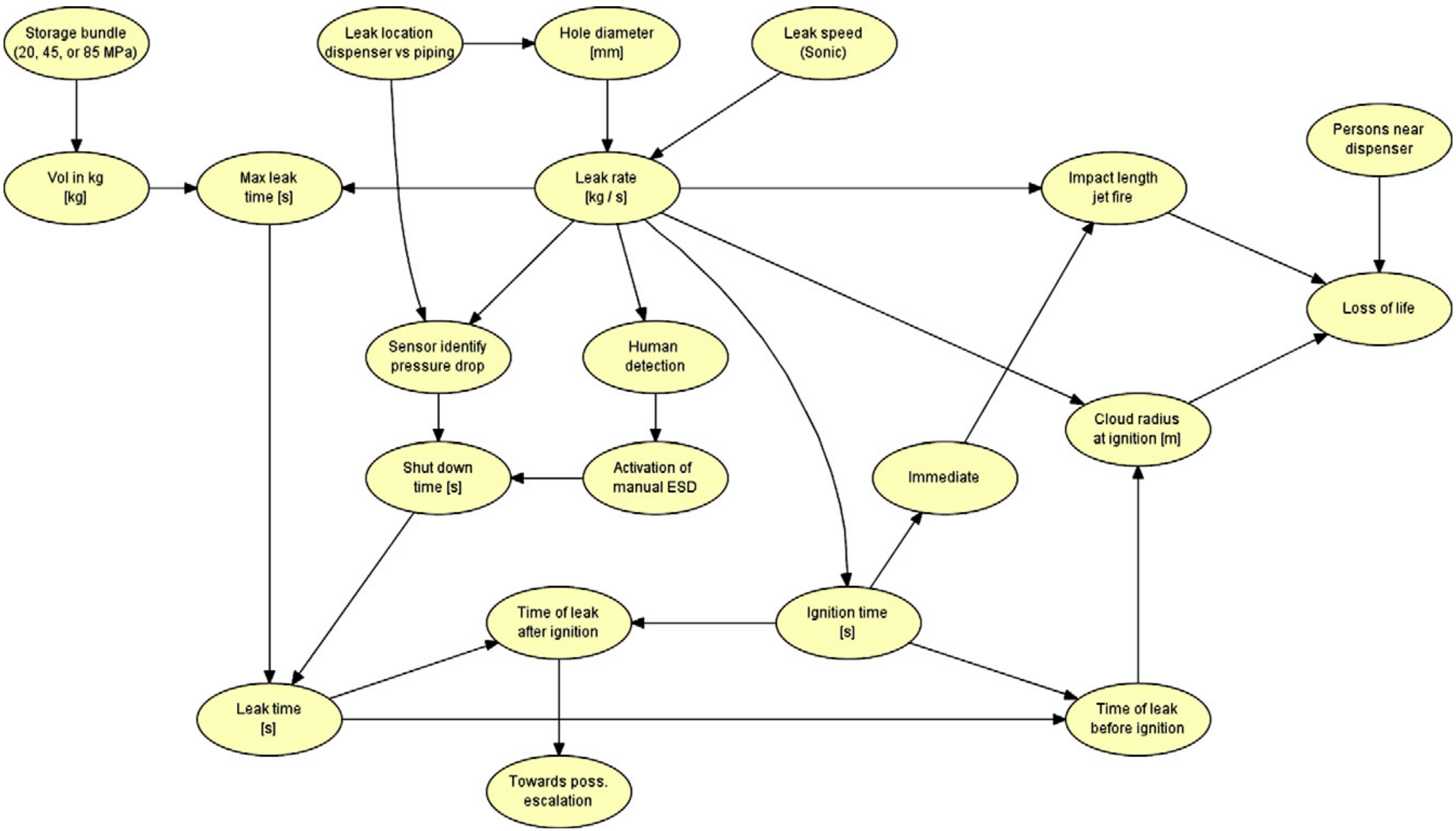
- Historical: 2.712
- BaNTERA: 2.518

BaNTERA not only assess the likelihood of TPD to future natural gas pipelines, but also quantifies and explains its uncertainty

BaNTERA demonstration videos

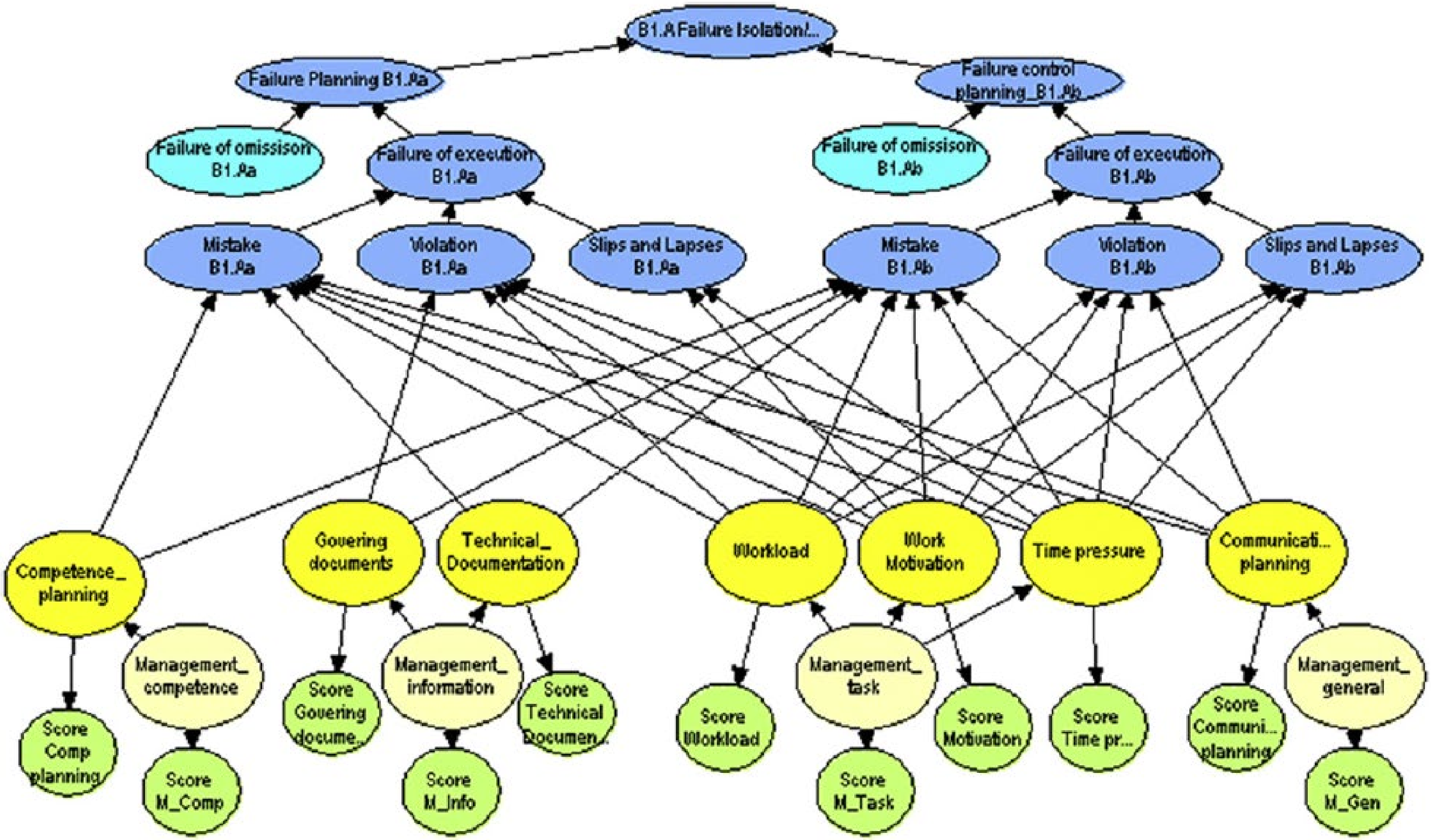
Video 1: Introduction to BaNTERA	https://umd.box.com/s/2oz1vzmakcpb8gtk6k5xwpnreotgntqk Explains the basics of BaNTERA
Video 2: Predictive Reasoning with BaNTERA	https://umd.box.com/s/eaifm5a3z7h7vf7dfoe0n1i2aergt07j Video shows using BaNTERA to calculate the probability of puncture for a known One Call ticket
Video 3: Diagnostic and Explanatory Reasoning with BaNTERA	https://umd.box.com/s/ye7l9mwvteuv2245qv19rrncs0pj9wlp Diagnosing root causes of a pipe hit

PRA: Hydrogen dispensing



Haugom, G. P. & Friis-Hansen, P. Risk modelling of a hydrogen refuelling station using Bayesian network. *International Journal of Hydrogen Energy*, **2011**, *36*, 2389-2397.

PRA: Offshore oil maintenance errors



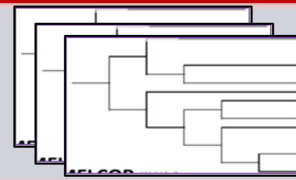
Vinnem, J. E.; Bye, R.; Gran, B. A.; Kongsvik, T.; Nyheim, O. M.; Okstad, E. H.; Seljelid, J. & Vatn, J. Risk modelling of maintenance work on major process equipment on offshore petroleum installations. *Journal of Loss Prevention in the Process Industries,* **2012***, 25*, 274-292

BN-Based “Smart SAMGs”

Generate spectrum of accident scenarios

Goal: Identify potential accident scenarios

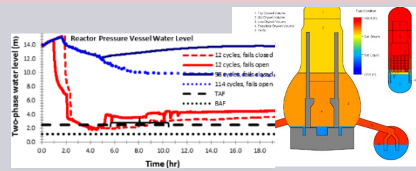
Tool: DDET/ADAPT simulation scheduler



Simulate reactor physics for each scenario

Goal: Predict range of plant parameters for known system faults

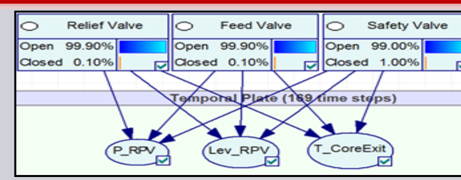
Tool: MELCOR



Encode results in a generic knowledge base

Goal: Build a map between known parameters and known faults

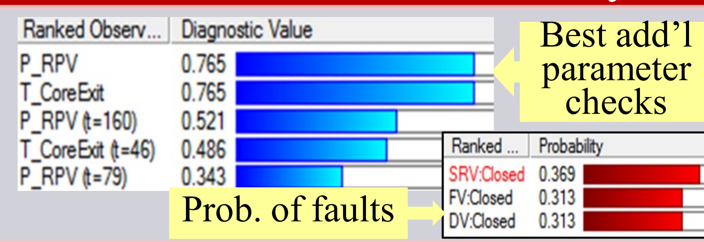
Tool: Bayesian Networks



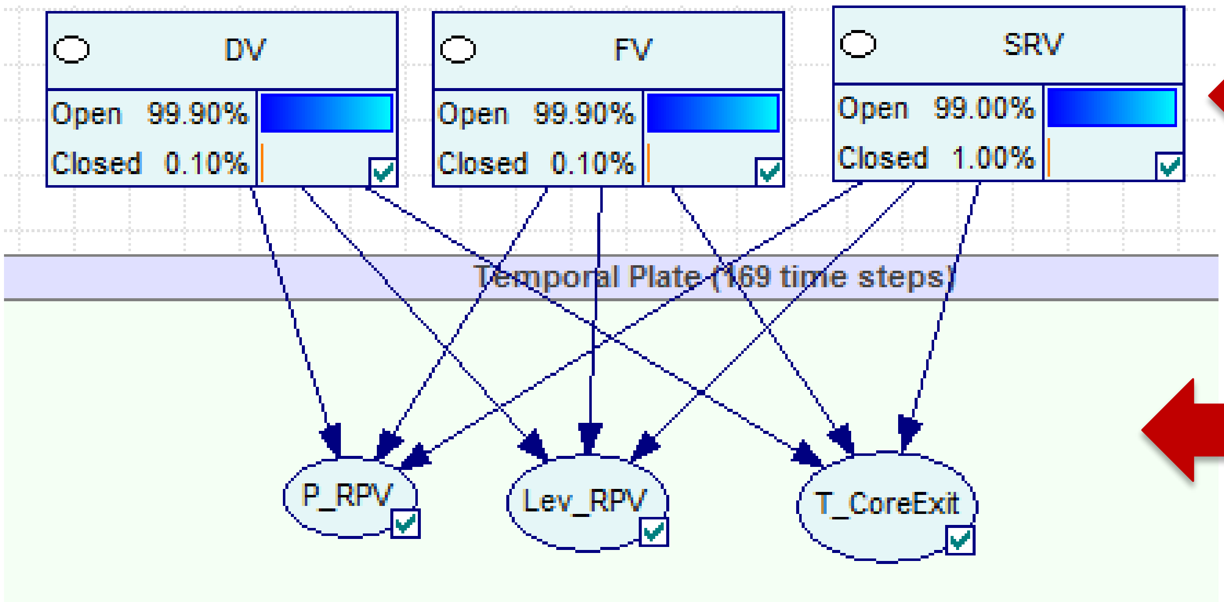
Enable queries for specific parameters, faults, under uncertainty

Goal: Enable users to diagnose specific faults, identify key indicators, ask “what-if”

Tool: Probabilistic queries, differential diagnosis, value of information



“Smart procedures”: Building a probabilistic map between plant conditions and plant parameters

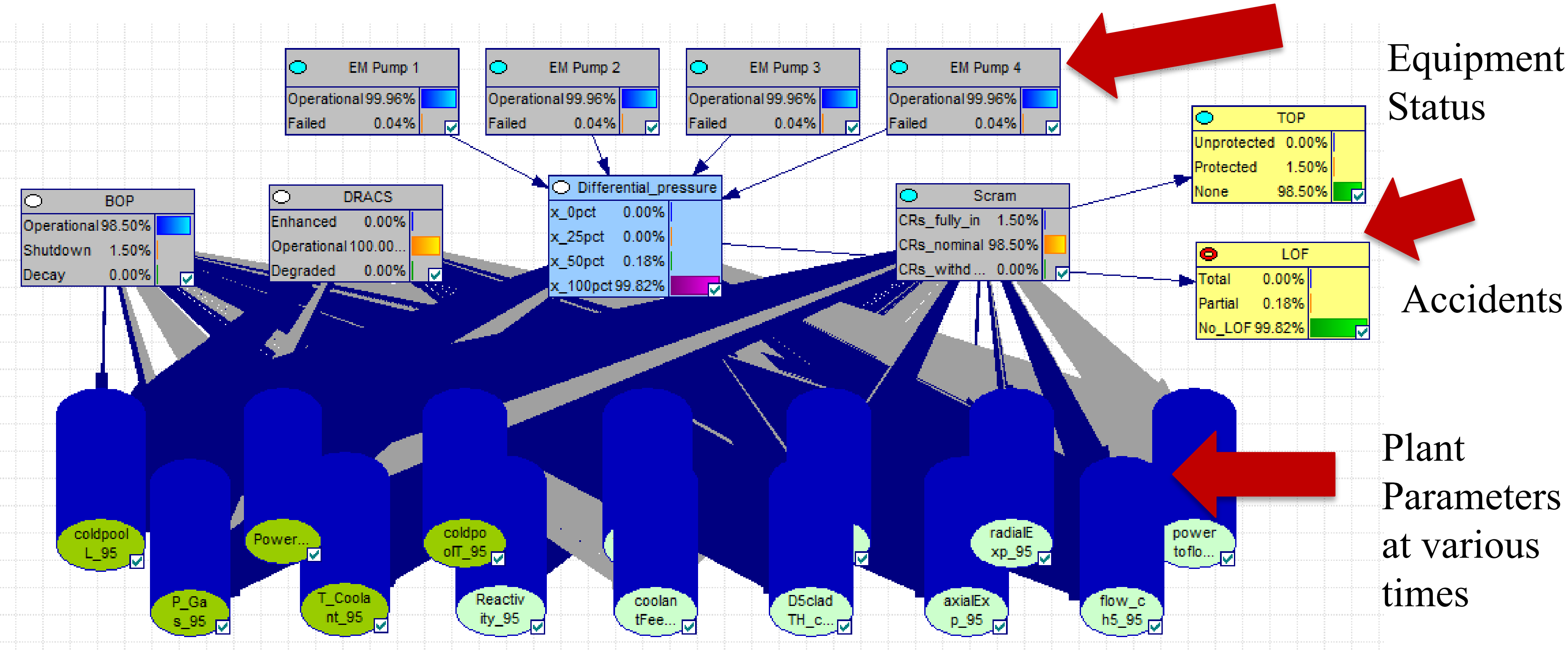


Equipment status (disease)

Indicators to check (tests)

Groth, K. M.; Denman, M. R.; Cardoni, J. N. & Wheeler, T. A. ``Smart Procedures'': Using dynamic PRA to develop dynamic, context-specific severe accident management guidelines (SAMGs) Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 12), 2014.

Proof-of-concept model for SMART Procedures



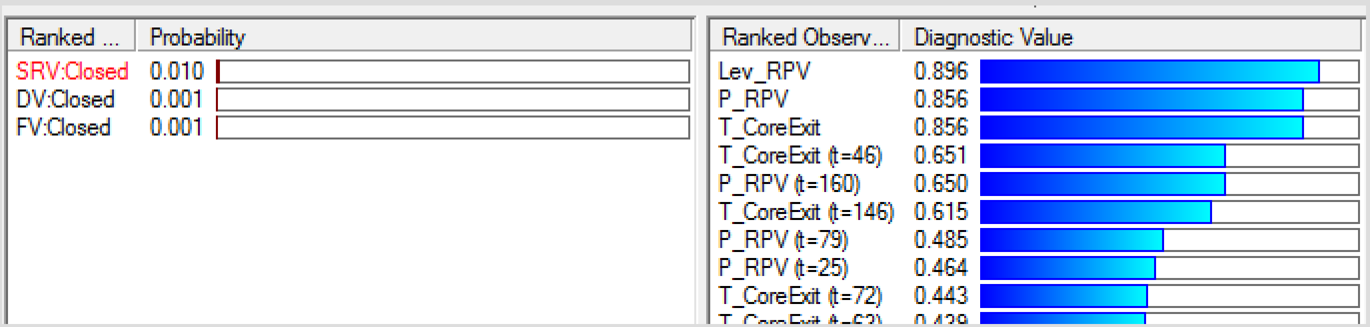
BN-based tool can be used to provide insight into instruments are most essential for diagnosis of specific accidents. This information can provide insight into, reactor design e.g., which instruments need to be accident hardened.

K. M. Groth, M. R. Denman, T. B. Jones, M. C. Darling, and G. F. Luger, "Building and using dynamic risk-informed diagnosis procedures for complex system accidents," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 234(1), pp. 193–207, Feb. 2020.

Groth, K. M.; Denman, M. R.; Jones, T.; Darling, M. & Luger, G. Proof-of-concept accident diagnostic support for sodium fast reactors Proceedings of the European Society for Reliability Annual Meeting (ESREL 2015), 2015.

Smart SAMG diagnosis

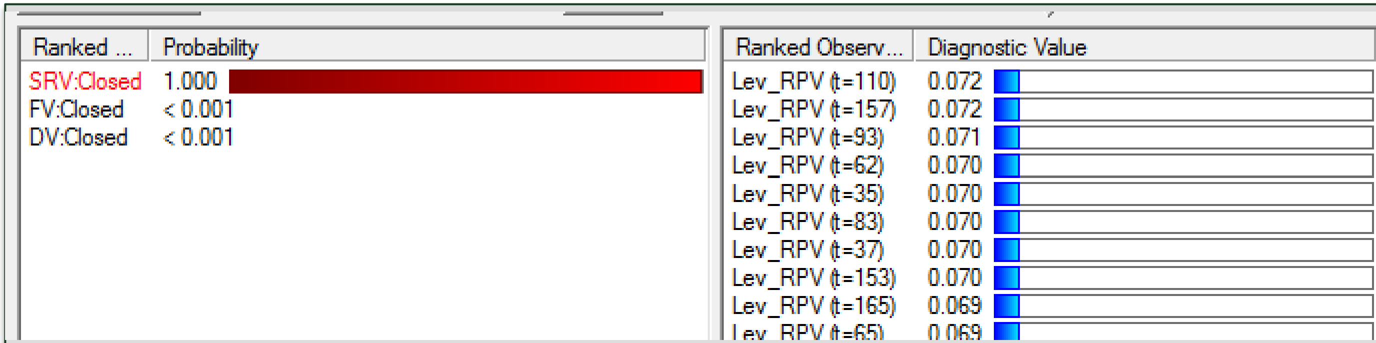
Prior (Unknown accident)



Suggests checking RPV level (t0), RPV pressure (t0), Core Exit temp (t0)

$$Observation\text{:}\, RPV\ Level(time\ 0) = low$$

Posterior (Condition-specific)



Suggests checking RPV level (110, t157, t93)

A single key observation dramatically changes belief about ECCS status and value of additional tests

Implemented in GeNIe: http://genie.sis.pitt.edu/

Diagnostic values of tests

For FV failure

Ranked Observ...	Diagnostic Value
T_CoreExit (t=46)	0.319
Lev_RPV	0.316
T_CoreExit (t=146)	0.232
P_RPV (t=160)	0.224
P_RPV (t=128)	0.217
T_CoreExit (t=108)	0.202
Lev_RPV (t=157)	0.200
P_RPV (t=58)	0.197
Lev_RPV (t=68)	0.195
T_CoreExit	0.191
P_RPV	0.191
Lev_RPV (t=159)	0.188
Lev_RPV (t=151)	0.187
T_CoreExit (t=44)	0.184
P_RPV (t=79)	0.184
Lev_RPV (t=46)	0.182
T_CoreExit (t=123)	0.176
P_RPV (t=101)	0.175
T_CoreExit (t=72)	0.174
T_CoreExit (t=63)	0.171
Lev_RPV (t=106)	0.167

Suggested checks: Core exit temp (t46), RPV level(t0)

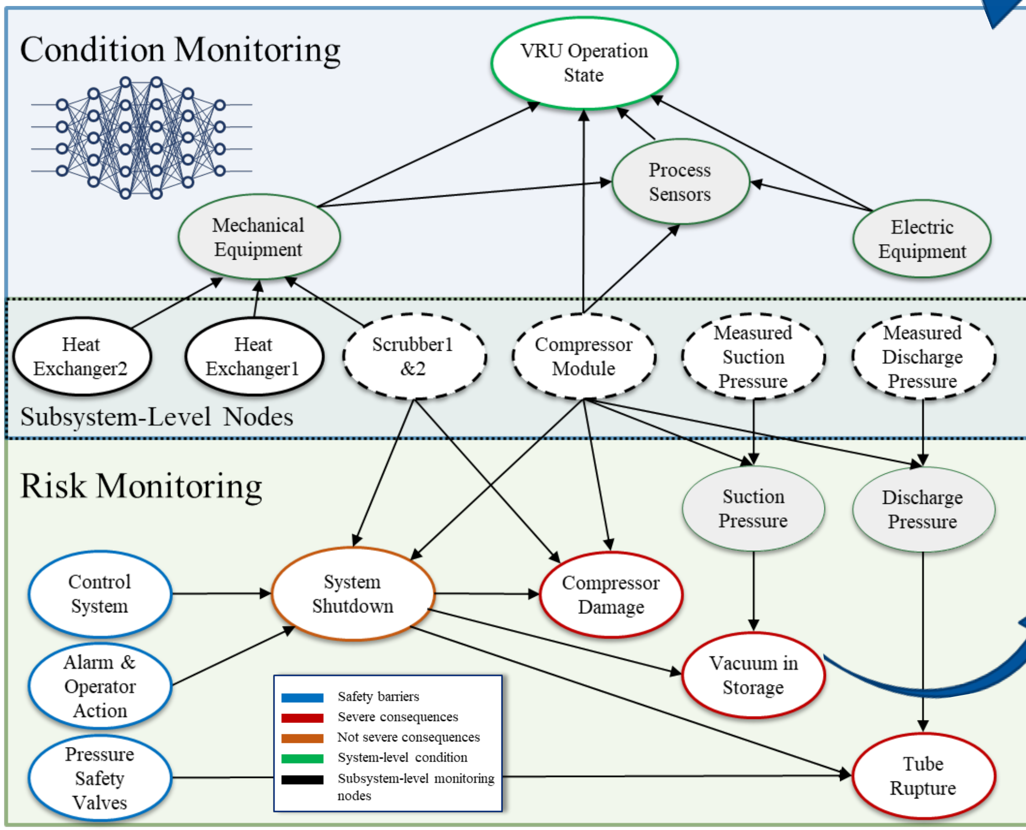
For SRV failure

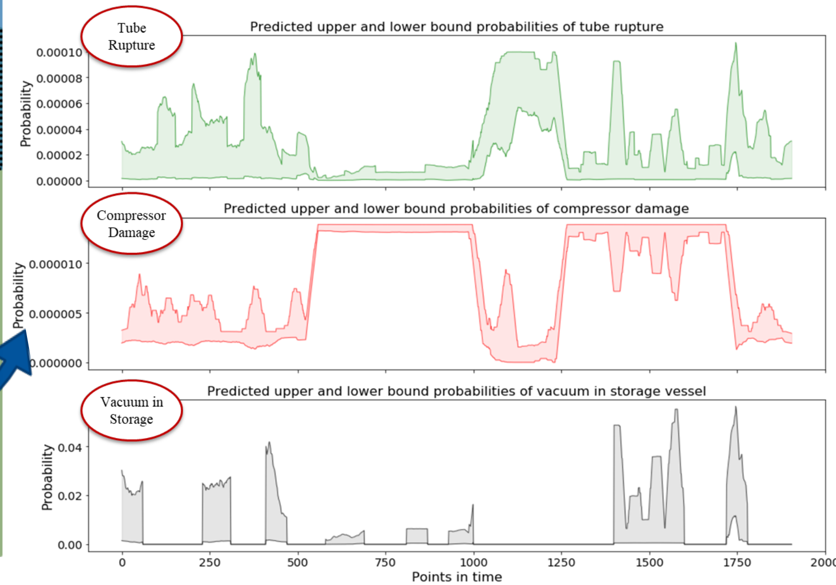
Ranked Observ...	Diagnostic Value
Lev_RPV	0.896
P_RPV	0.856
T_CoreExit	0.856
T_CoreExit (t=46)	0.651
P_RPV (t=160)	0.650
T_CoreExit (t=146)	0.615
P_RPV (t=79)	0.485
P_RPV (t=25)	0.464
T_CoreExit (t=72)	0.443
T_CoreExit (t=63)	0.439
Lev_RPV (t=160)	0.433
Lev_RPV (t=61)	0.421
T_CoreExit (t=44)	0.414
P_RPV (t=58)	0.406
Lev_RPV (t=156)	0.406
T_CoreExit (t=123)	0.382
T_CoreExit (t=108)	0.372
Lev_RPV (t=161)	0.364
T_CoreExit (t=98)	0.361
P_RPV (t=128)	0.359
T_CoreExit (t=70)	0.358

Suggested checks: RPV Press(t0), RPV level(t_0)

Different tests provide greater diagnostic power for different diseases (and some provide little value for either disease)

Developed predictive algorithms for risk & condition monitoring of an oil & gas vapor recovery unit

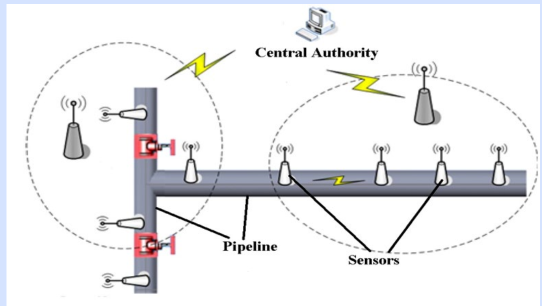


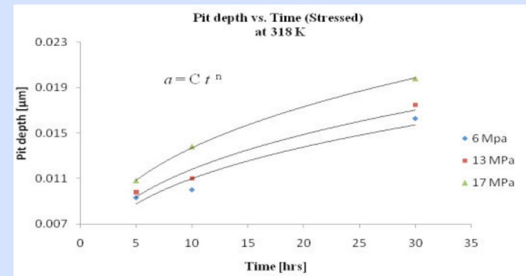


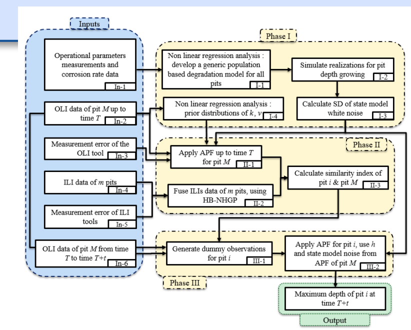
Our new methods can dynamically monitor the risk of complex engineering systems

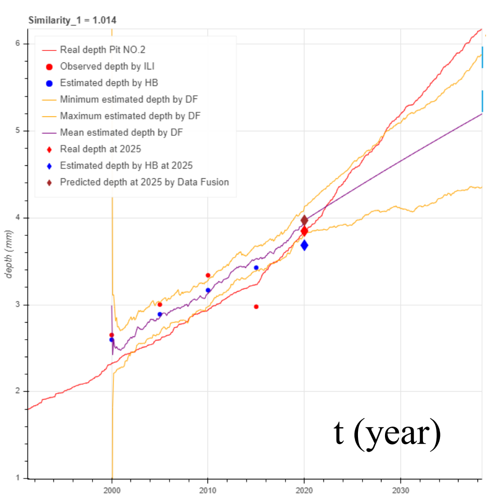
Moradi, Ramin, Sergio Cofre-Martel, Enrique Lopez Droguett, Mohammad Modarres, and Katrina M. Groth. "Integration of deep learning and Bayesian networks for condition and operation risk monitoring of complex engineering systems." Reliability Engineering & System Safety 222 (2022).

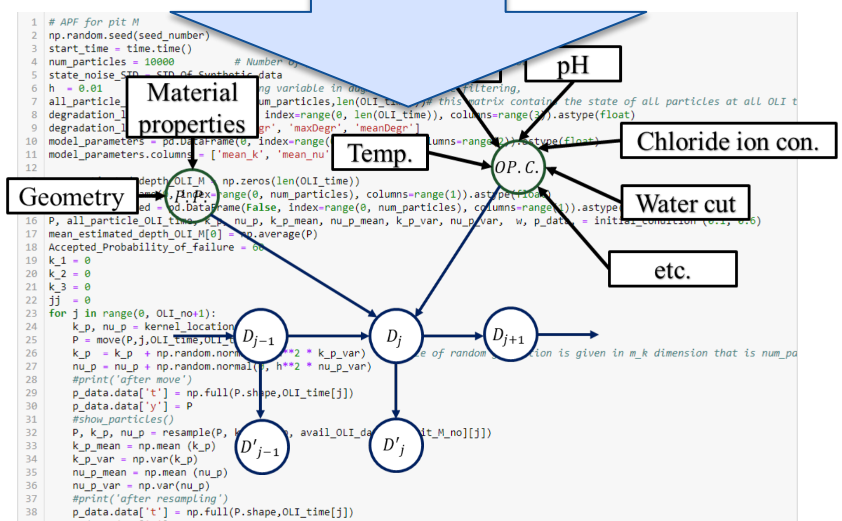
Developed a hybrid (data + physics) corrosion degradation model for pipeline Prognostics and Health Management







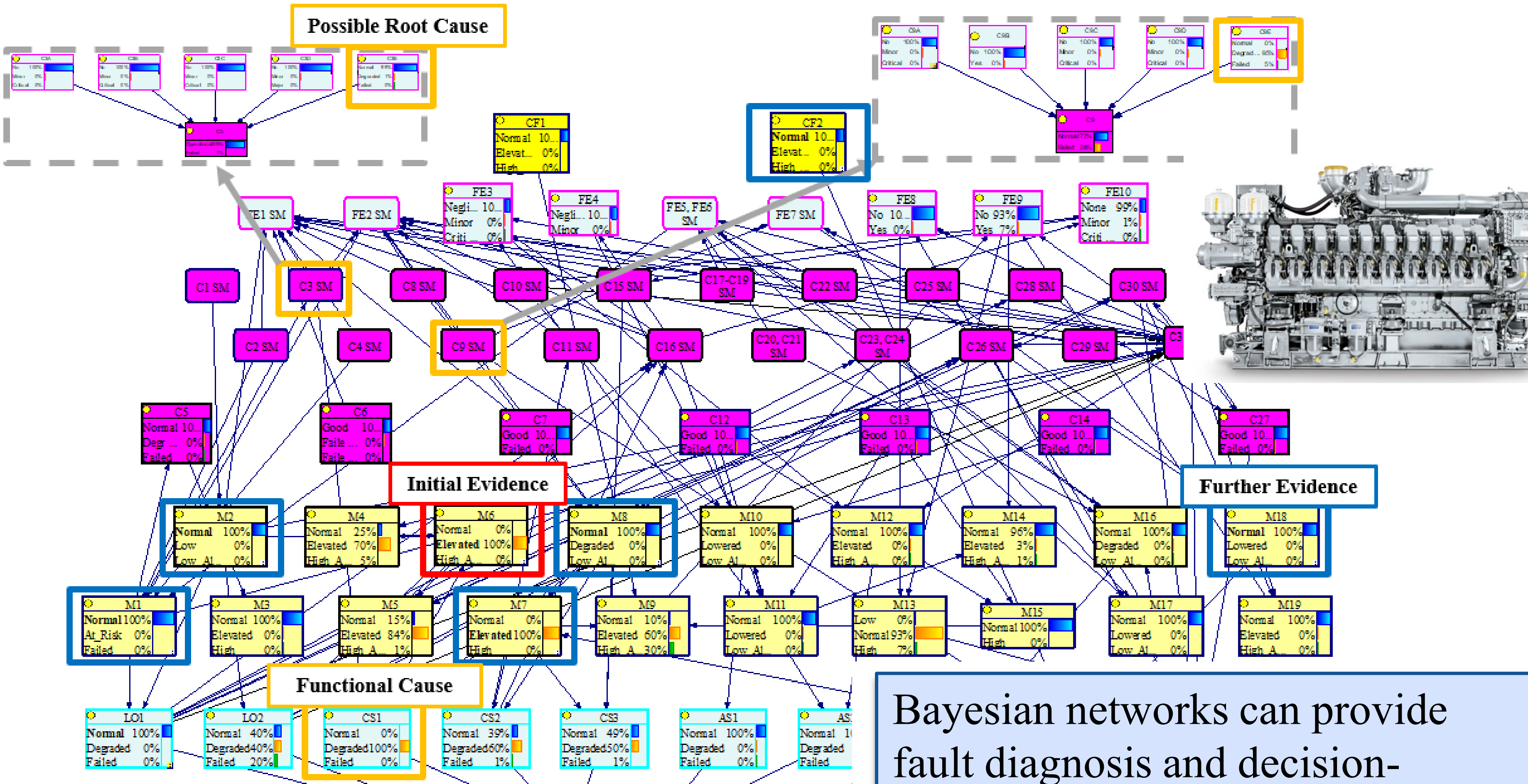




Application to maintenance policy planning

- Heidary, R. & Groth, K. M. "A hybrid population-based degradation model for pipeline pitting corrosion." *Reliability Engineering & System Safety*, **2021**, *214*
- Heidary, R. & Groth, K. M. "A hybrid model of internal pitting corrosion degradation under changing operational conditions for pipeline integrity management" *Structural Health Monitoring*, **2020**, *19*.

Applications: day-to-day operations – ship engine fault detection and diagnosis



Bayesian networks can provide fault diagnosis and decision-making support on daily operations

Reynolds, Steven (advisor: Katrina M. Groth). "Application of a Bayesian Network Based Failure Detection and Diagnosis Framework on Maritime Diesel Engines." M.S. Thesis, University of Maryland (2022).

New: ReMUSCLE: Resilient Maintenance of Maritime Unmanned Surface Vehicles

DDSL

Objectives

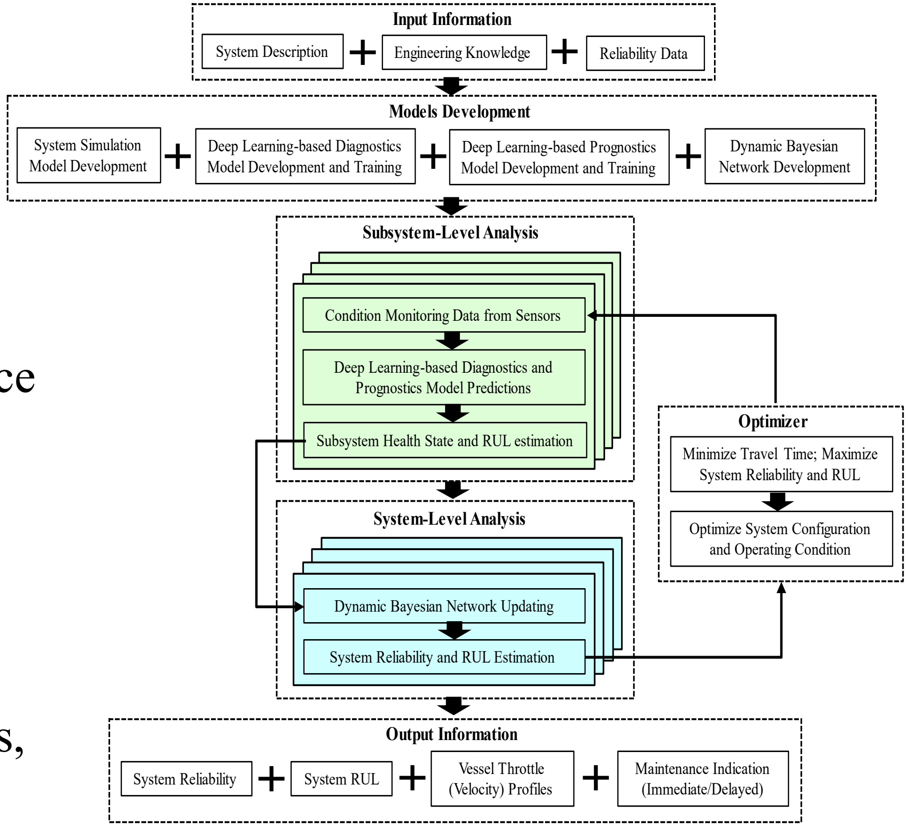
- Making the USV system self-aware of its health state
- Anticipate, respond and maintain its operations under hazardous conditions

Approach

- Integration of PHM, PRA, and Performance Optimization
- Data fusion:
  - Sensor data, maintenance log, failure data from physics-based system simulations, reliability databases
- Use of Deep Learning, Bayesian Networks, and Optimization Techniques

Outcome

- An AI-based decision support tool for predictive maintenance of USVs

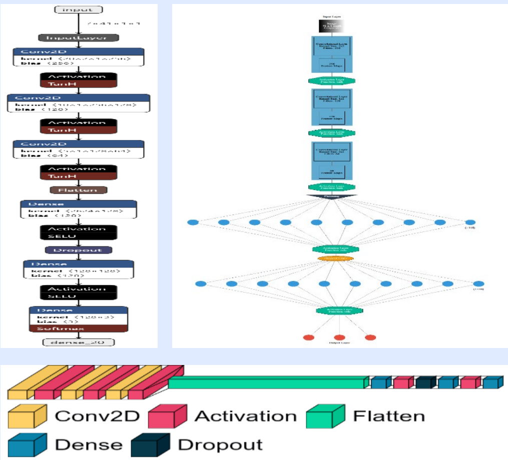


A **unified framework** for system **health monitoring**, **reliability evaluation** and reliability-based **performance optimization** for USVs

Hazra, I., Chatterjee, A., Southgate, J., Weiner, M. J., Groth, K. M., & Azarm, S. (2023). A reliability-based optimization framework for planning operational profiles for unmanned systems. *Journal of Mechanical Design*, 146(5), p.51704, May 2024.

Fusing Deep Learning with Bayesian Network

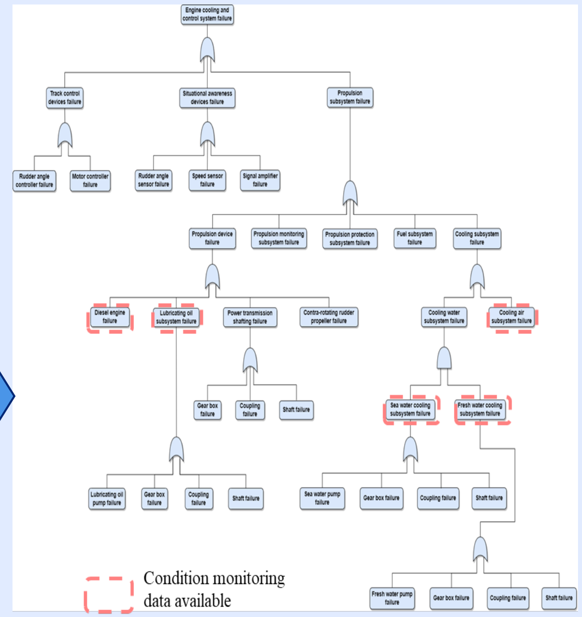
Deep Learning Models for Diagnosis



Diagnostic model	Training accuracy	Testing accuracy
Diesel engine	94.61%	97.49%
Fresh water cooling subsystem	88.36%	88.84%
Sea water cooling subsystem	96.08%	98.34%
Lubricating oil subsystem	83.33%	87.06%

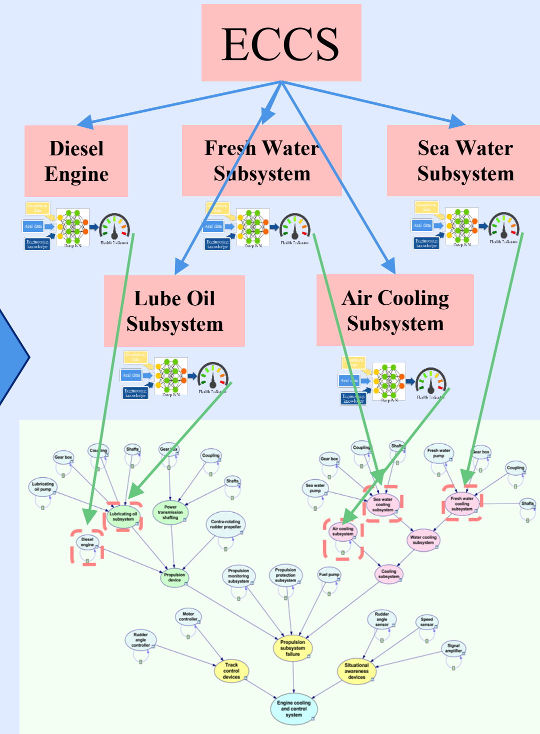
Deep Learning-Based Diagnostic Models are Developed for Subsystems (Engine, Cooling Subsystems, Lube Oil Subsystem) Health State Classification

Fault Tree for the Engine Cooling and Control System



- *We developed a comprehensive fault tree for the ECC system based on the system failure logic and component failure data*
- *Failure data are obtained from the non-electronics parts reliability data (NPRD) handbooks*

Dynamic Bayesian Network Development and Integration with Diagnosis Models



Fusion of new information from condition monitoring data obtained from sensors

t = 0

	Prob.
Working	1
Failed	0

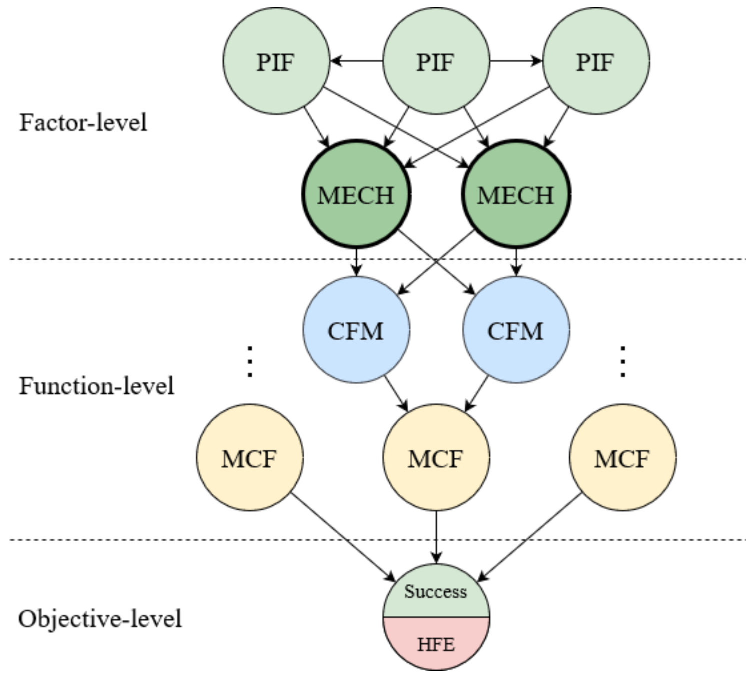
t >= 1

(Self) [t-1]	Working	Failed
Working	1-P(Failed)	0
Failed	P(Failed)	1

Hazra, I., Chatterjee, A., Southgate, J., Weiner, M. J., Groth, K. M., & Azarm, S. (2023). A reliability-based optimization framework for planning operational profiles for unmanned systems. *Journal of Mechanical Design*, 146(5), p.51704, May 2024.

Background: HRA Constructs

- Underlying causal layers of human performance can be captured using Bayesian networks
  - Causal dependence between each layer
  - Probabilities of child node states modified by parent node states



PIF **Performance Influencing Factor**: Multi-dimensional characteristics of the performance context.

MECH **Human Failure Mechanism**: Cognitive/physical process leading to the failure mode.

CFM **Crew Failure Mode**: Failure pathways that define failure to achieve a function.

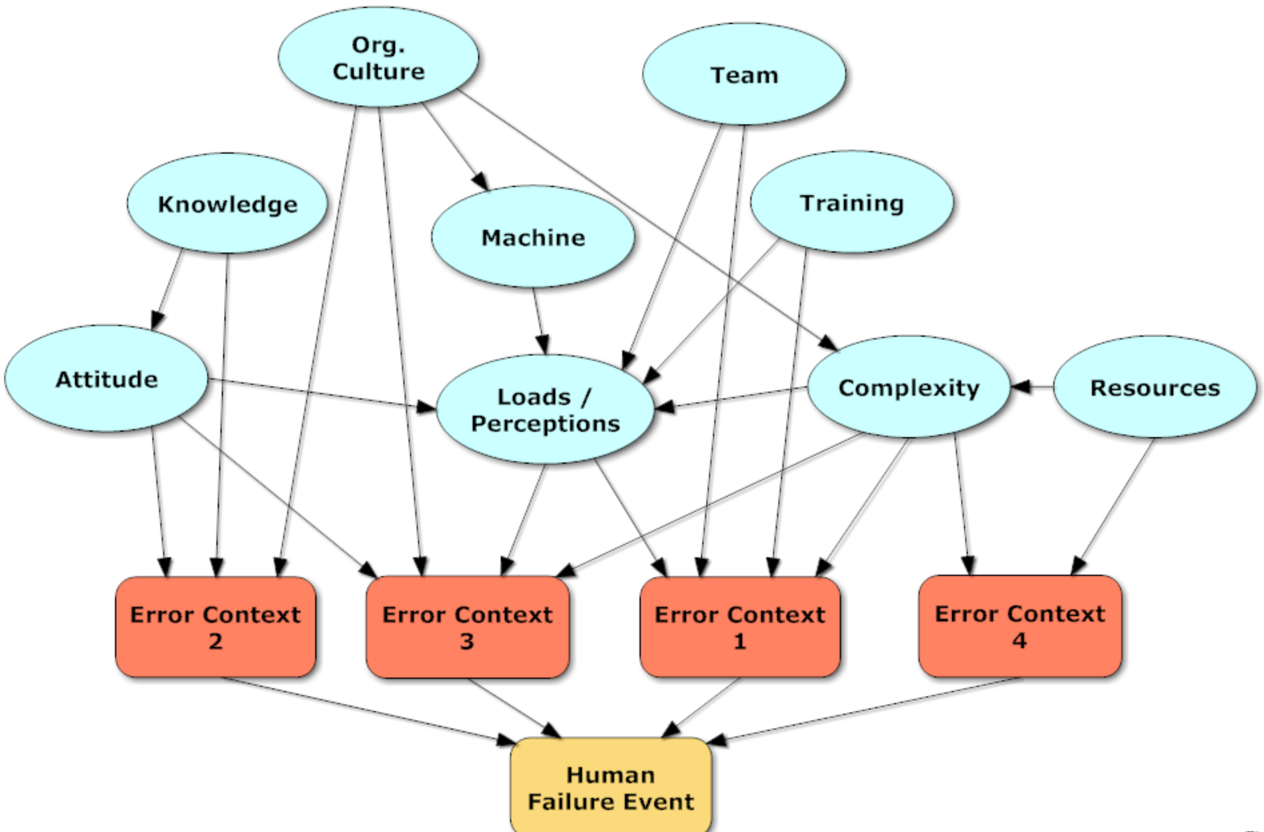
MCF **Major Crew Function**: High-level actions taken during a scenario.

HFE **Human Failure Event**: Highest-level failure considered; includes *at least* one *failed* Major Crew Function.

The HFE is the culmination of the failure process; there are multiple different mechanisms that might bring about an HFE and multiple modes that describe how it occurs.

Adapted from Paglioni, V. P., and Groth, K. M. 2022. "Dependency Definitions for Quantitative Human Reliability Analysis." *Reliability Engineering & System Safety*, 220.

Nuclear HRA: Data-informed quantification model



Training

LTA	0.37
Adequate	0.63

Org. Culture

LTA	0.48
Adequate	0.52

Resources

LTA	0.40
Adequate	0.60

Team

LTA	0.46
Adequate	0.54

Knowledge

LTA	0.53
Adequate	0.47

		Org. Culture	
		LTA	Adeq.
Machine	LTA	0.36	0.62
	Adequate	0.64	0.38

		Knowledge	
		LTA	Adeq.
Attitude	LTA	0.47	0.87

	Org. Culture	LTA		Adeq.	
	Resources	LTA	Adeq.	LTA	Adeq.
Complexity	LTA	0.62	0.50	0.57	0.52
	Adequate	0.38	0.50	0.43	0.48

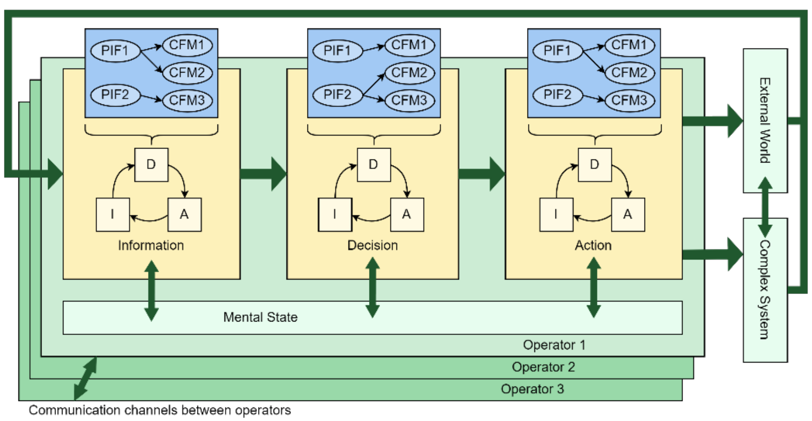
$$P(HPE) = \Sigma_{PSFs} P(HFE|EC1, EC2, EC3, EC4) \times P(EC1|PSFs) \times P(EC2|PSFs) \times P(EC3|PSFs) \times P(EC4|PSFs) \times P(PSFs)$$

$$Marginal: \Pr(Err) = 1.88E-03$$

Groth, Katrina M. & Mosleh, Ali. Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability,* **2012**, 226, 361-379.

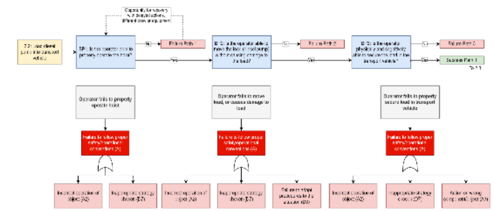
Expanding the causal logic foundations of human reliability analysis

- Human reliability analysis (HRA) asks the questions: *how, why,* and *when* does human failure occur?
- Need for cognitively realistic, yet quantifiable, HRA models that capture the full range of contexts of operation
- Other models lack psychological realism or an analyst-friendly methodology.



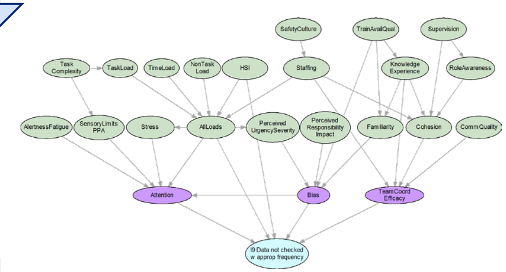
Task 1

Identify causal chains from contextual factors to failure events; develop cognitive literature-based Bayesian network (BN) structure



Task 2

Leverage existing databases and expert elicitation resources to quantify the BN nodes



Task 3

Demonstrate applicability through case study simulation to validate model structures

Goal: Development of **strong theoretical basis of HRA**, compatible with current models.

A. Al-Douri, C. Levine, and K. M. Groth, "Identifying Human Failure Events (HFEs) for External Hazard Probabilistic Risk Assessment," Reliability Engineering & System Safety, vol. 235, 2023.

C. S. Levine, A. Al-Douri, and K. M. Groth, "Causal Pathways Leading to Human Failure Events in Information-Gathering System Response Activities," in NPIC&HMIT 2023, Knoxville, TN, Jul. 2023.

Research Objectives and Anticipated Results

Objective: Development of **comprehensive, causal logic-based HRA models**, quantified and compatible with current HRA data and methods.

Objective 1
Identify & model causal pathways leading from contextual factors to potential failure events

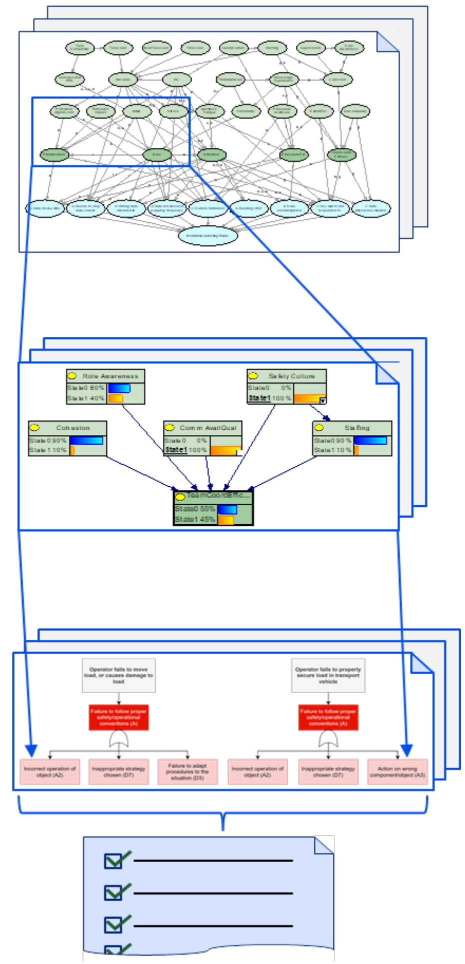
Result
Comprehensive BN HRA models for info, decision, and action phases of human performance based in causal logic

Objective 2
Leverage existing HRA databases, expert knowledge, and data resources to quantify BN nodes

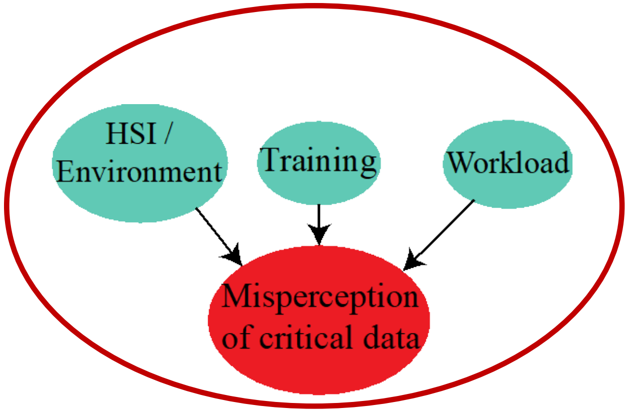
Anticipated Result
Fully quantified, data-driven BN structures capable of prospective, retrospective analysis of crew failure probability

Objective 3
Conduct validation of causal models: demonstrate usability, applicability to breadth of scenarios

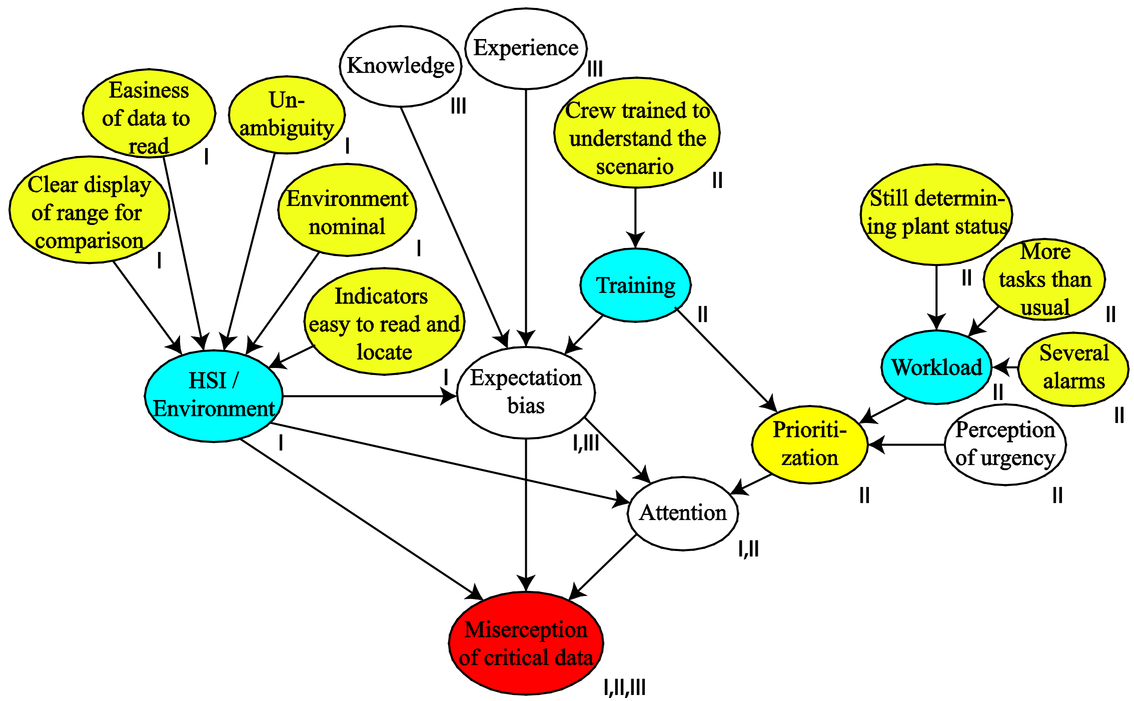
Anticipated Result
Validated model through case study simulation for external actions HRA use case



IDHEAS HRA model with extended causal details



Assessment questions
And
Cognitive literature

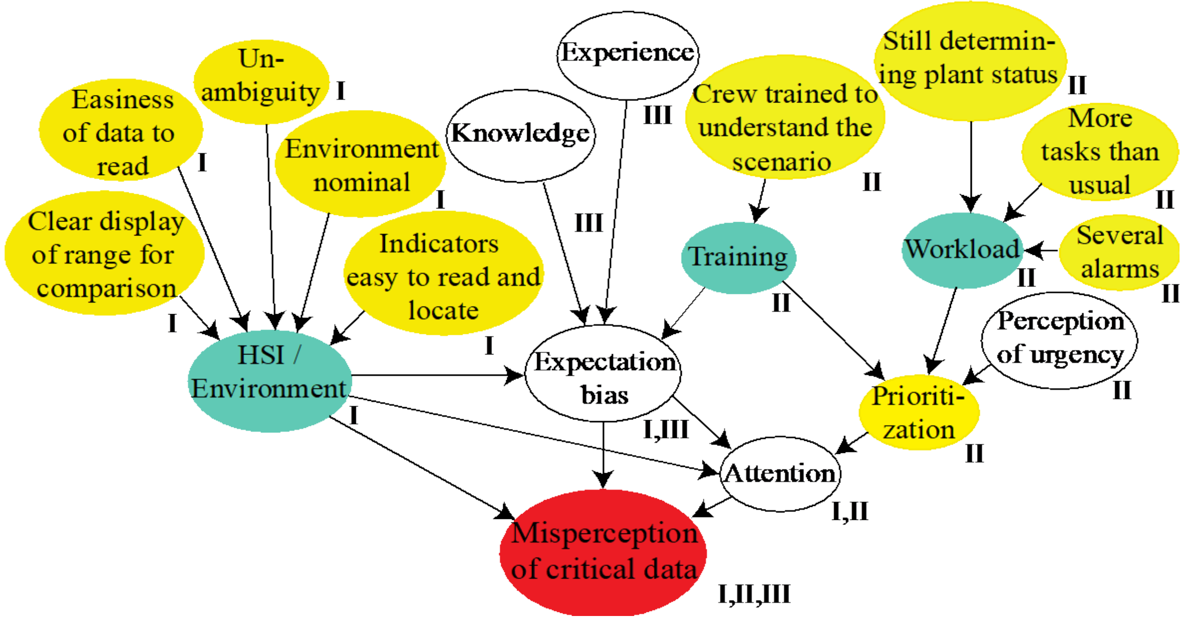


Zwirglmaier, K.; Straub, D. & Groth, K. M.
Framework for a Bayesian Network Version of IDHEAS
Proceedings of the European Society for Reliability Annual Meeting (ESREL 2015), **2015**

Zwirglmaier, K.; Straub, D. & Groth, K. M. "Capturing cognitive causal paths in human reliability analysis with Bayesian network models" *Reliability Engineering & System Safety*, **2017**, *158*, 117-129

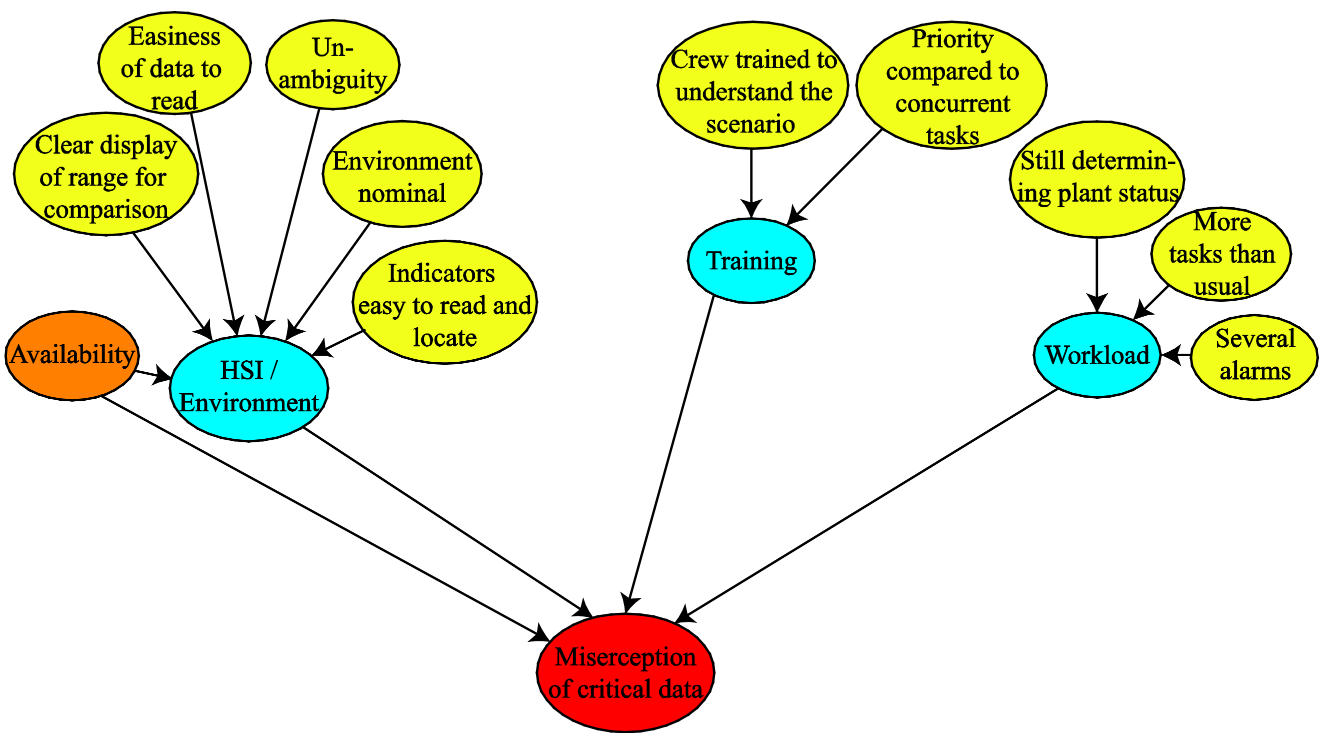
Capturing causal paths in HRA & using node reduction to simplify the structure

- Expanded BN version of "critical data misperceived"
  - Target node (red)
  - PSFs (cyan)
  - Adapted from the IDHEAS report to specify the PSFs (yellow)
  - Nodes introduced to illustrate the macro-cognitive path (white)
- This structure explicitly illustrates the causal paths from psychology literature & uses this in quantification of human error



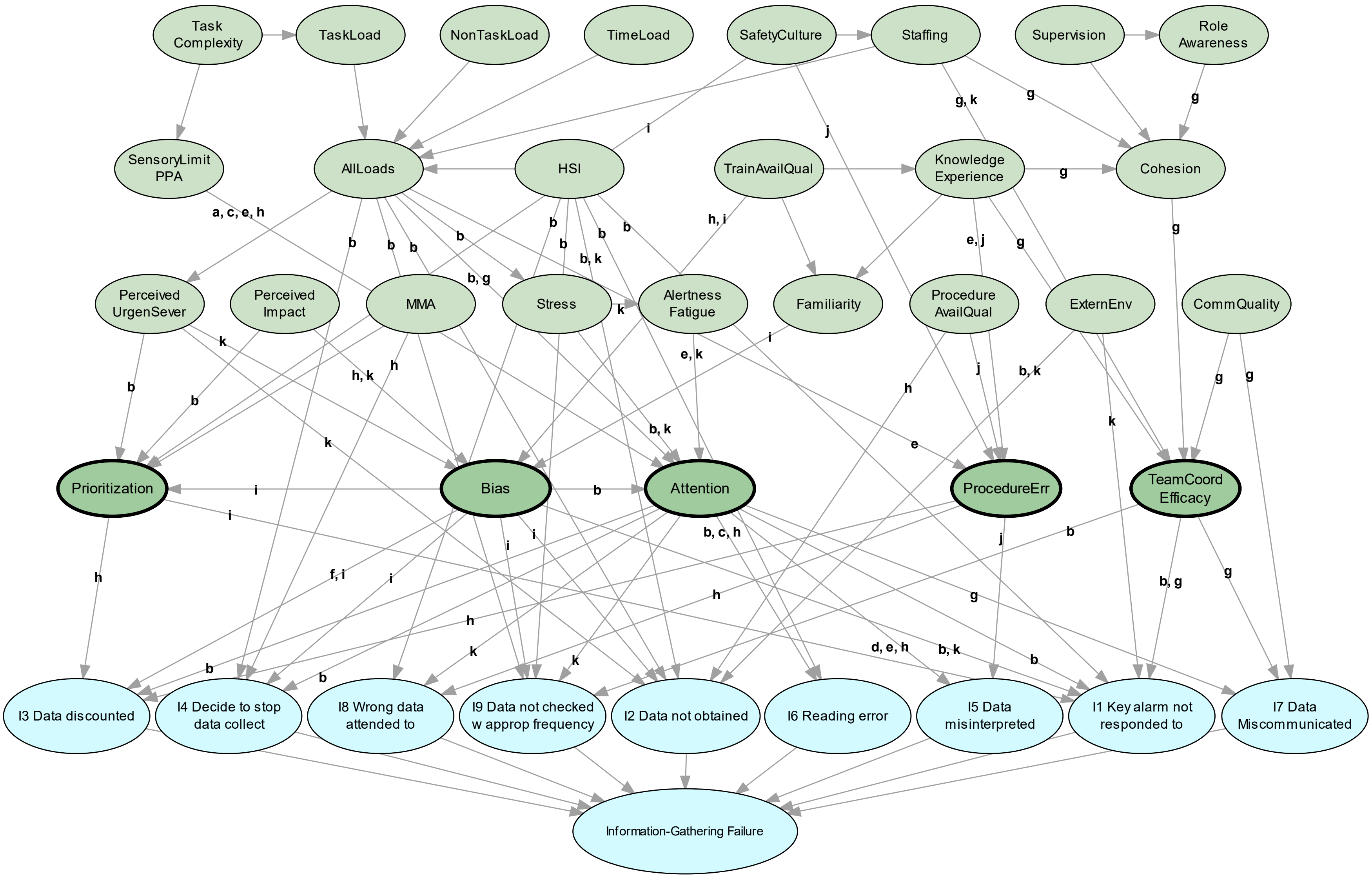
Formal reduction of BN structure

- Some PSFs nodes ate not explicitly captured in IDHEAS (e.g., experience, attention, prioritization)
  - Their quantification is not in the current scope of IDHEAS
  - It is important to keep these relations in mind when quantifying the model
- Since it is not within the scope of IDHEAS to quantify these nodes we remove them from the network following the algorithm by Shacter (1986 &1988)



Zwirglmaier, K.; Straub, D. & Groth, K. M. "Capturing cognitive causal paths in human reliability analysis with Bayesian network models" *Reliability Engineering & System Safety,* **2017**, *158*, 117-129

Preliminary Results: Information-Gathering Bayesian network structure



C. S. Levine, A. Al-Douri, and K. M. Groth, "Causal Pathways Leading to Human Failure Events in Information-Gathering System Response Activities," in 13th Nuclear Plant Instrumentation, Control & Human-Machine Interface Technologies (NPIC&HMIT 2023), Knoxville, TN, Jul. 2023.

If you want to get more complicated

- Continuous BNs
- Dynamic BNs
- Inference algorithms
- Value of information
- Bayesian updating the probabilities in the BN

Software packages

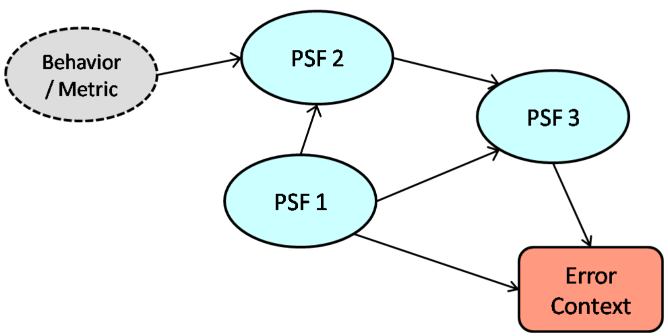
- Tools with graphical user interfaces
  - GeNIe (http://genie.sis.pitt.edu/)
  - Hugin (http://www.hugin.com)
  - Netica (http://www.norsys.com/)
  - MSBNx (http://research.microsoft.com/en-us/um/redmond/groups/adapt/msbnx/)
- Other flexible tools
  - Bayes Net Toolbox for Matlab (https://code.google.com/p/bnt/)
- Designed for Risk Analysts
  - Trilith (University of Maryland, contact Ali Mosleh or Katrina Groth)
    - Integration of BNs with ET/FT
  - AgenaRisk (Commercial package)
    - (BNs only)

Key benefits

- **Completeness**: Includes all relevant variables, not just easily observable variables or variables where data is plentiful. Allows variables to be interdependent.
- **Documentation**: Explicitly represents all variables and relationships deemed relevant to the problem space.
- **Simplification**: Decomposes problem into manageable pieces; simplifies acquisition of probability distribution.
  - It's easier to gather data about $p(d|b)$ than about $p(d|a, b, c \dots)$
- **Credibility**: The BN allows analyst to assemble information from multiple sources into a single model.
  - Populating the model with the most credible information (or expert)
- **Modifiability**: Analysts can update conditionally independent sections of the model without changing the entire model. Model is expandable in scope and depth.
- **Insight**: Enables analysts to make predictions without perfect information; enables understanding of cause-and-effect behavior, performing "what-if" analyses.

Summary

- BNs are a tool for:
  - Encoding a knowledge base (via a series of conditional probabilities)
  - Performing a probabilistic reasoning (induction, deduction) with the knowledge base
- Benefits
  - Completeness & Insight: Includes all variables, not just those with data
  - Simplicity: Decomposes a large problem into manageable pieces
  - Credibility: Models built with info. & data from multiple sources



$$= P(EC|PSF1, PSF3) \cdot P(PSF3|PSF1, PSF2) \cdot P(PSF2|PSF1, BM) \cdot P(PSF1) \cdot P(BM)$$

$P(EC \cap PSF1 \cap PSF2 \cap PSF3 \cap BM)$

References (1)

- Applied:
  - Norman Fenton and Martin Neil (2013) *Risk Assessment and Decision Analysis with Bayesian Networks.* CRC Press.
  - Finn V. Jensen and Thomas D. Nielsen (2007) *Bayesian Networks and Decision Graphs.* Springer.
  - Langseth, H. & Portinale, L. Bayesian networks in reliability. *Reliability Engineering and System Safety,* **2007***, 92*, 92-108.
  - M. Druzdzel and L. Van der Gaag, Building probabilistic networks: Where do the numbers come from," *IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no. 4, pp. 481-486, 2000
- Theoretical:
  - Judea Pearl (1998) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufman
  - Judea Pearl, “Bayesianism and causality, or, why I am only a half-Bayesian” *Foundations of Bayesianism*, 2001, 24, 19-34 .

References (2)

- Example models in this presentation

  - Darling, M. C.; Luger, G. F.; Jones, T. B.; Denman, M. R. & Groth, K. M. "Intelligent Modeling for Nuclear Power Plant Accident Management." *International Journal on Artificial Intelligence Tools*, In press; accepted October 2017.
  - Ekanem, N. A model-based Human Reliability Analysis Methodology (PHOENIX method). Ph. D. Thesis. *University of Maryland,* **2013**
  - Groth, Katrina M. & Swiler, Laura P. Bridging the gap between HRA research and HRA practice: A Bayesian Network version of SPAR-H. *Reliability Engineering and System Safety,* **2013***, 115*, 33-42.
  - Groth, Katrina M. & Mosleh, Ali. Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, **2012**, 226, 361-379.
  - Groth, Katrina; Wang, Chengdong & Mosleh, Ali. Hybrid causal methodology and software platform for probabilistic risk assessment and safety monitoring of socio-technical systems. *Reliability Engineering and System Safety,* **2010***, 95*, 1276-1285
  - Groth, K. M.; Denman, M. R.; Cardoni, J. N. & Wheeler, T. A. ``Smart Procedures'': Using dynamic PRA to develop dynamic, context-specific severe accident management guidelines (SAMGs) *Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 12)*, 2014.
  - K. M. Groth, M. R. Denman, T. B. Jones, M. C. Darling, and G. F. Luger, "Building and using dynamic risk-informed diagnosis procedures for complex system accidents," Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 234(1), pp. 193–207, Feb. 2020.
  - Groth, K. M.; Denman, M. R.; Jones, T.; Darling, M. & Luger, G. "Proof-of-concept accident diagnostic support for sodium fast reactors." *Proceedings of the European Society for Reliability Annual Meeting (ESREL 2015),* 2015.
  - Hazra, I., Chatterjee, A., Southgate, J., Weiner, M. J., Groth, K. M., & Azarm, S. (2023). A reliability-based optimization framework for planning operational profiles for unmanned systems. *Journal of Mechanical Design*, 146(5), p.51704, May 2024.
  - Heidary, R. & Groth, K. M. "A hybrid population-based degradation model for pipeline pitting corrosion." *Reliability Engineering & System Safety,* **2021***, 214*.
  - Heidary, R. & Groth, K. M. "A hybrid model of internal pitting corrosion degradation under changing operational conditions for pipeline integrity management" *Structural Health Monitoring,* **2020***, 19*.
  - C. S. Levine, A. Al-Douri, and K. M. Groth, "Causal Pathways Leading to Human Failure Events in Information-Gathering System Response Activities," in 13th Nuclear Plant Instrumentation, Control & Human-Machine Interface Technologies (NPIC&HMIT 2023), Knoxville, TN, Jul. 2023.
  - Moradi, Ramin, Sergio Cofre-Martel, Enrique Lopez Droguett, Mohammad Modarres, and Katrina M. Groth. "Integration of deep learning and Bayesian networks for condition and operation risk monitoring of complex engineering systems." Reliability Engineering & System Safety 222 (2022).
  - Reynolds, Steven,"Application of a Bayesian Network Based Failure Detection and Diagnosis Framework on Maritime Diesel Engines." M.S. Thesis, University of Maryland (2022).
  - A. Ruiz-Tagle, A.D. Lewis, C. Schell, E. Lever, & K. M. Groth. "BaNTERA: A Bayesian Network for Third-Party Excavation Risk Assessment" *Reliability Engineering & System Safety*, July 2022, 223.
  - Haugom, G. P. & Friis-Hansen, P. Risk modelling of a hydrogen refuelling station using Bayesian network. *International Journal of Hydrogen Energy,* **2011***, 36*, 2389-2397.
  - Vinnem, J. E.; Bye, R.; Gran, B. A.; Kongsvik, T.; Nyheim, O. M.; Okstad, E. H.; Seljelid, J. & Vatn, J. Risk modelling of maintenance work on major process equipment on offshore petroleum installations. *Journal of Loss Prevention in the Process Industries,* **2012***, 25*, 274-292
  - Zwirglmaier, K.; Straub, D. & Groth, K. M. "Capturing cognitive causal paths in human reliability analysis with Bayesian network models" Reliability Engineering & System Safety, 2017, 158, 117-129
  - Zwirglmaier, K.; Straub, D. & Groth, K. M. "Framework for a Bayesian Network Version of IDHEAS." Proceedings of the European Society for Reliability Annual Meeting (ESREL 2015), 2015.

ENRE 447/602 Reliability Analysis

Module 9: Risk Analysis

Katrina M. Groth
Associate Professor
Mechanical Engineering Department
Center for Risk and Reliability
University of Maryland
kgroth@umd.edu

Module 9 Objectives

- Understand how the previous modules combine to form a cohesive process for **quantitative risk assessment (QRA)**
- Understand the connection between risk analysis, risk assessment, risk management, and risk communication

Risk assessment methods

Type	Example Methods
Qualitative to Semi-quantitative	▪ FMEA (Failure Modes & Effects Analysis) ▪ FMECA (Failure Modes, Effects, & Criticality Analysis) ▪ PHA (Process Hazard Analysis) ▪ HAZOP (Hazards & Operability Analysis)

Quantitative	▪ QRA (Quantitative Risk Assessment) or PRA (Probabilistic Risk Assessment), including:
	  ▪ Fault trees
	  ▪ Event trees
	  ▪ Bayesian networks
	  ▪ Simulation
	  ▪ Hazard models

- Categories of risk analysis applications:
  - **Safety**: estimate potential harms due to natural or anthropogenic causes
  - **Security**: estimating access and harm due to intentional malicious actions
  - **Health**: estimate potential disease, injury, and mortality
  - **Financial:** estimating potential monetary losses
  - **Environmental:** estimate losses due to noise, contamination, pollution, etc.

Definition: QRA

- **Risk** is defined as a triplet:
  - What can go wrong? (**Scenario** – S_i)
  - How likely is it to happen? (**Probability** – P_i)
  - If it does happen, what are the consequences? (**Consequence** – C_i)
- Thus, **Quantitative Risk Assessment (QRA)** QRA has three main pieces:
  1. Identifying the scenarios S_i that comprise sequences of events, root causes, and outcomes of challenges to the system.
  2. Determining the probability P_i or frequency of occurrence F_i for each event E_i in each scenario S_i.
  3. Evaluating the consequences C_i of each scenario S_i occurrence.
- The **total expected risk value** R is the expected consequence summed over possible scenarios:

$$R = \sum_{S_i} P_i \times C_i$$

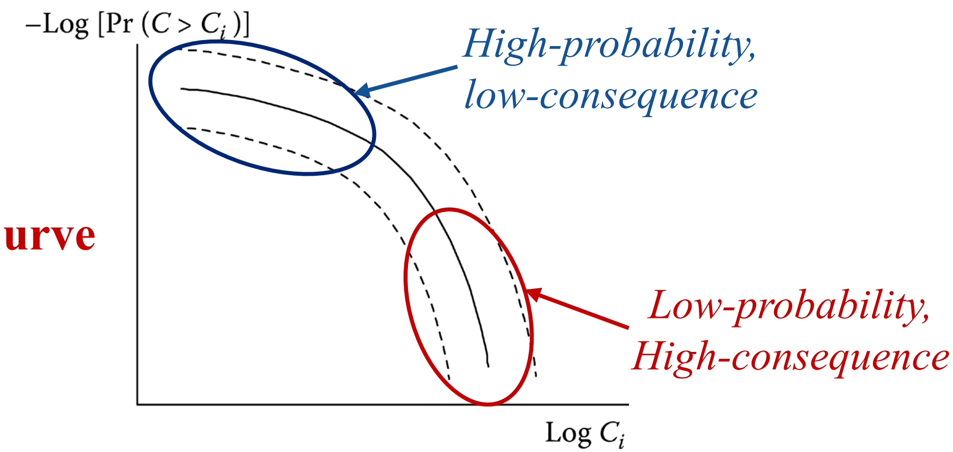
QRA results interpretation

- The total risk value R can mask qualitative aspects of the risk.
- **Example:** Scenario 1 occurs with a frequency $f_1 = 0.01\ \mathrm{yr}^{-1}$ and a \$1,000,000 loss. Scenario 2 occurs with $f_2 = 1\ \mathrm{yr}^{-1}$ and a \$10,000 loss. Thus, the total risk values are equal:

$$R_1 = f_1 \times C_1 = 0.01\ \mathrm{yr}^{-1} \times \$1{,}000{,}000 = \$10{,}000/\mathrm{yr}$$
$$R_2 = f_2 \times C_2 = 1\ \mathrm{yr}^{-1} \times \$10{,}000 = \$10{,}000/\mathrm{yr}$$

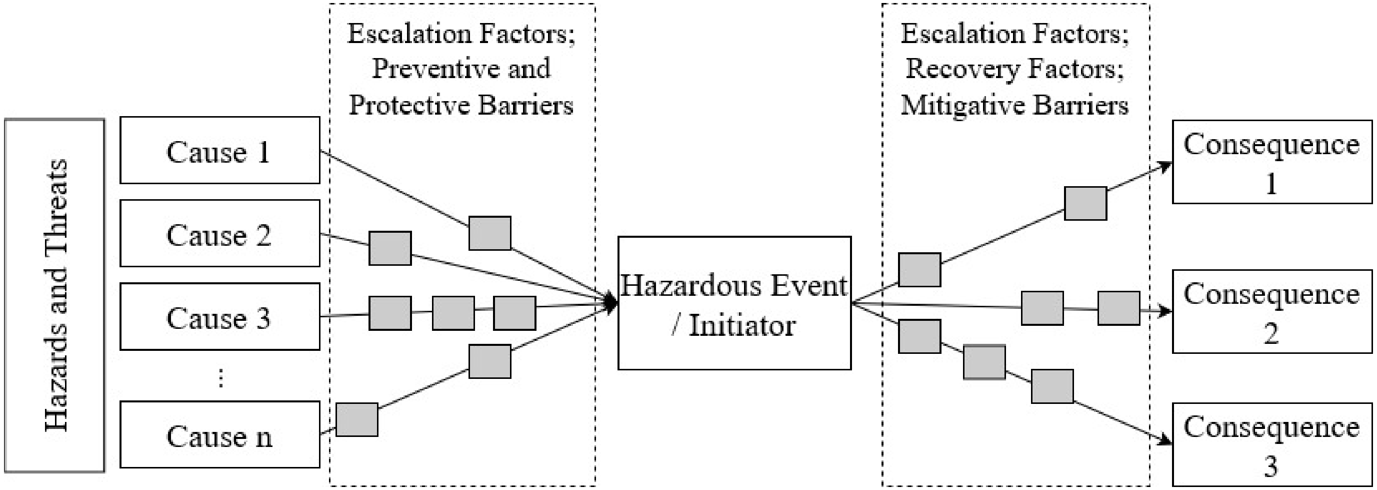
- QRA results can also be interpreted via a **risk profile** that plots probability against consequences, which can differentiate the low-P, high-C events from high-P, low-C events and thus provide a basis for decision.

Farmer's Curve



QRA Process

- The process of QRA is a function of the scenarios, probabilities/frequencies, and consequences in a system.
- QRA attempts to identify all possible scenarios that lead to losses, and for each scenario:
  - Probabilities or frequencies of each event
  - Description and amount of consequences
- The **bowtie model** visualizes the key elements of QRA



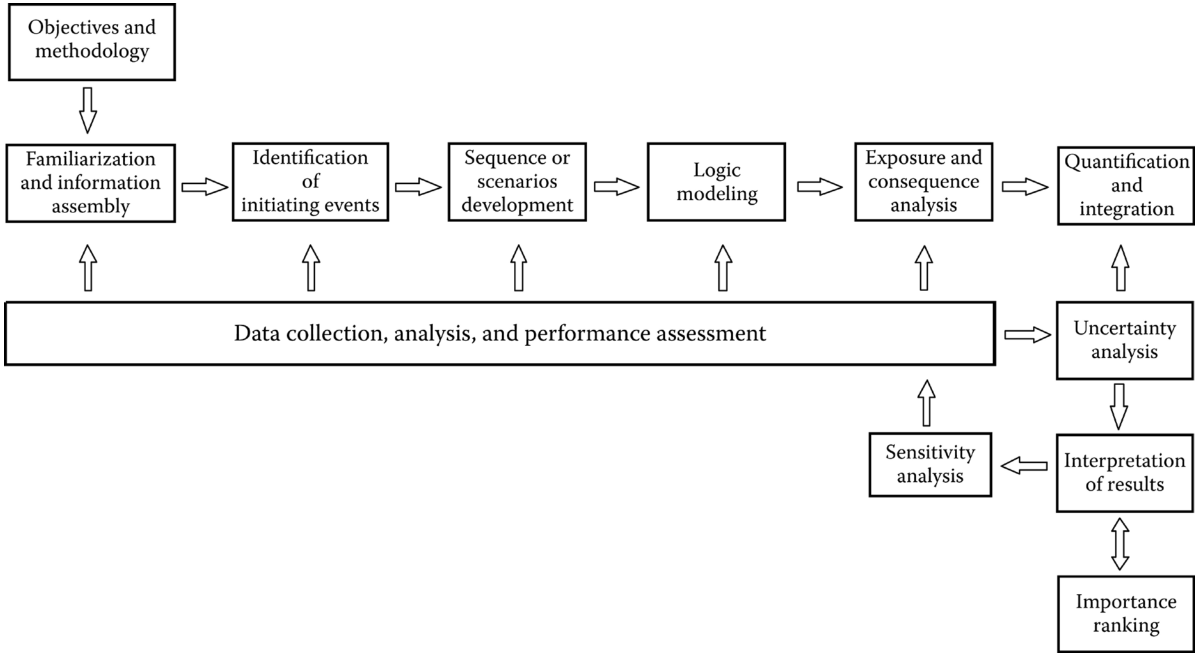
Steps in QRA

QRA requires characterizing several key aspects of a system and scenario:

1. **Initiating event**: identify root causes, preventive & protective barriers, escalating factors
2. **Hazard exposure**: form the chain of events for each scenario S_i that lead to exposure of the hazard and, if not mitigated, consequence C_i.
3. **Hazard identification**: survey the process/system to identify hazards.
4. **Barrier identification**: identify the barriers that contain, prevent, or minimize exposure to the hazard.
5. **Barrier challenges**: identify the mechanisms by which barriers may be challenged and/or degrade.
   - **Barrier strength degradation:** reduced thickness, material property changes
   - **Barrier stress:** internal forces or pressure, penetration/distortion from external forces
   - **Degradation conditions:** process malfunction, poor design, natural phenomena

PRA process

- **Probabilistic risk assessment (PRA)** is the most well-known QRA method.
  - *The primary value of a PRA is to highlight the system design and operational deficiencies and support subsequent risk management efforts to identify and optimize resources that can be invested on improving the design and operation of the system.*



PRA process (1)

1. **Objectives & Methodology**
   - Define/review the method, scope, rules and objectives of the analysis
2. **Familiarization & Information Assembly**
   - Identify major barriers, structures, systems, human interventions, subsystem interactions
   - Study past failures, dependent events, near-misses, abnormalities
3. **Initiating Events**
   - Identify the events that, if not responded to, could result in hazard exposure
4. **Scenario Development**
   - Create scenario event trees that encompass all potential exposure paths
5. **Causal Logic Modeling**
   - Create event and fault trees
6. **Data Collection, Analysis, and Assessment**
   - Gather data from generic sources, past events, expert judgment, etc.
7. **Consequence Analysis**
   - Characterize the range of possible effects from hazard exposure
8. **Quantification & Integration**
   - Form a single, coherent model from the logic trees; find minimal cut sets and frequencies

PRA process (2)

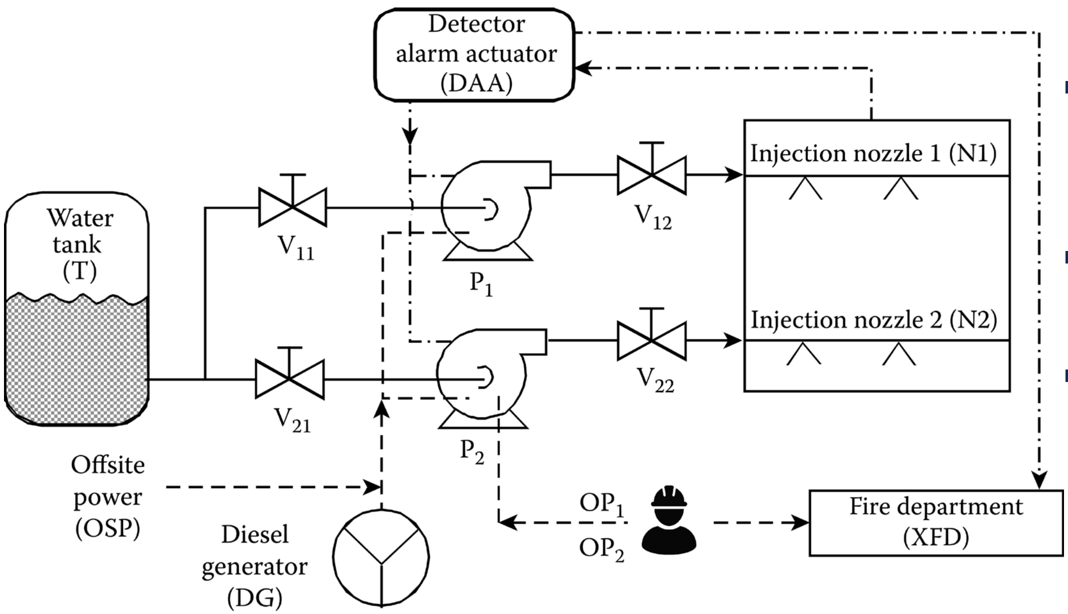
9. **Uncertainty Analysis**
   - Incorporate the uncertainties from all facets of the PRA and inform risk managers
10. **Sensitivity Analysis**
    - Determine the significance of model/parameter choice
    - Identify PRA elements that might be sensitive to the final risk results
    - Vary the value of sensitive items to propagate the changes through to the final result
    - Rank the elements that are most sensitive
11. **Risk Ranking and Importance Analysis**
    - Rank the scenarios and system elements corresponding to their risk/safety significance
    - Importance Measures indicate the absolute or relative (to other system elements) importance of specific system elements
12. **Interpretation of Results**
    - Determine the accuracy of logic models and scenario structures, assumptions, and scope of PRA
    - Identify system elements for which better information is necessary to reduce uncertainties
    - Revise the PRA and reinterpret results until stable and accurate results are obtained

Strengths of PRA

- PRA is the most rigorous formal QRA approach with many strengths:
  1. Integrated, systematic explanation of design and operational features
  2. Insight into the root causes of failure and effectiveness of barriers
  3. Incorporation of causal factors, system interactions, and interfaces
  4. Incorporation of operating experience and relevant data
  5. Explicit consideration of uncertainty
  6. Analysis and comparison of competing risks
  7. Formal sensitivity evaluation of assumptions and data
  8. Absolute and relative importances of systems & components
  9. Consistent and transparent framework for data & information fusion
  10. Documented process for exploring priorities & encouraging discourse

Example PRA: Fire protection system (1/6)

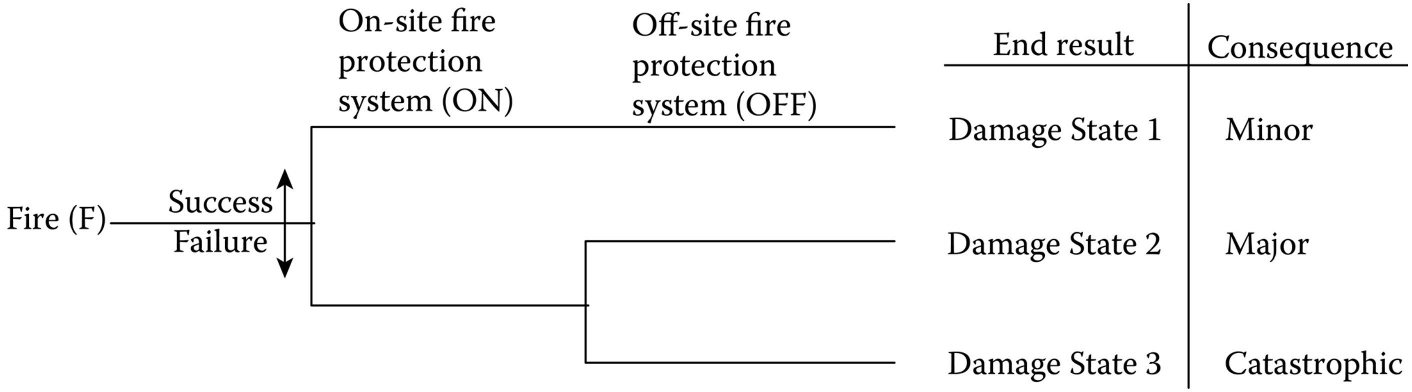
- Assess the potential risk and financial loss from possible failure scenarios of a fire protection system at a power plant
  - Independent nozzle trains – each can control any fire; nozzle N1 is primary.
  - Signal from DAA automatically starts pump P_1 and alerts fire department
  - Operator OP_1 can start the second injection path manually.
    - If this does not occur, operator OP_2 calls for fire department
    - Damage is greater if fire department required (i.e., nozzles unable to extinguish)



- If normal off-site power (OSP) not available, on-site diesel generator (DG) provides power to pumps.
- Battery power for pumps always available.
- Valves are normally open, but manually shut when pumps are being repaired.

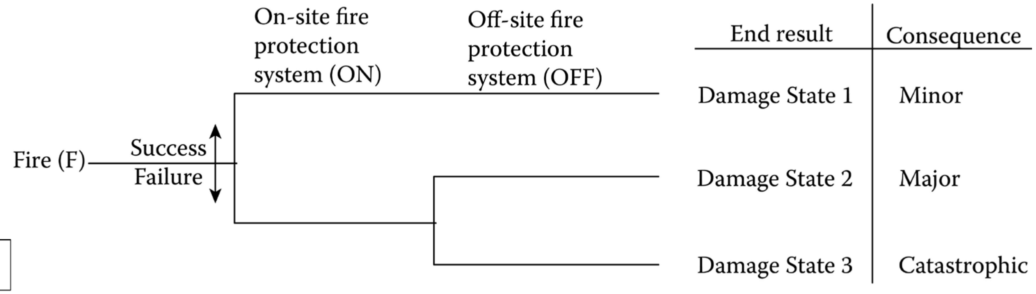
Example PRA: Fire protection system (2/6)

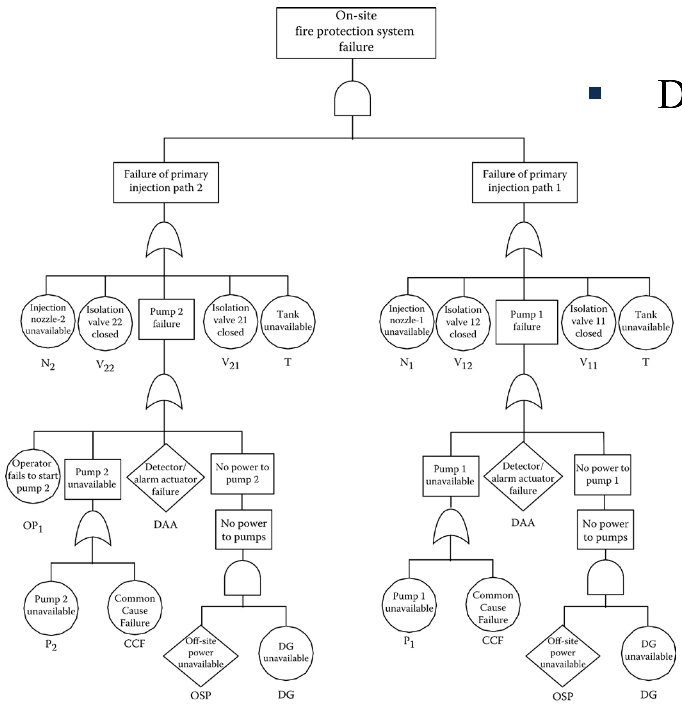
- **Identify initiating events:**
  - Find all events that could cause a sustained fire in the main building
    - Equipment malfunction, human error, facility conditions
  - Estimate the frequency of each event
  - If multiple events lead to the same magnitude of fire (e.g., same consequence), sum the individual frequencies
  - Assume fire frequency: $f_F = 7.1 \times 10^{-4}\ \text{yr}^{-1}$ (only initiating event)
- **Scenario development:**
  - Model the cause-effect relationships between the fire and following events
  - **Recall:** two types of protective measures (onsite pumps/tanks, offsite fire dept.)



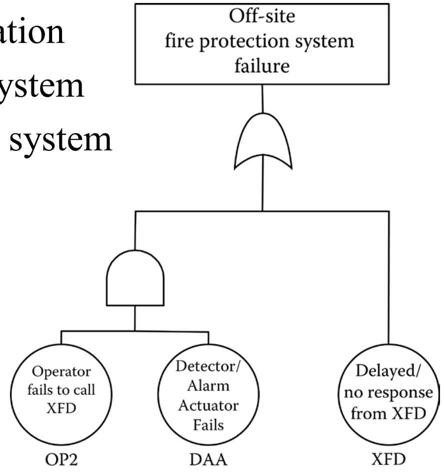
Example PRA: Fire protection system (3/6)

- **Causal logic model development:**
  - Identify the failures (equipment or human) that lead to failure of pivotal events





- Describe all basic events that lead to failure in **fault trees**
  - Note that there are physical dependencies between components that are not accounted for in the fault tree
  - Dependencies will be considered in quantification
  - Left: fault tree for the on-site fire protection system
  - Right: fault tree for the off-site fire protection system

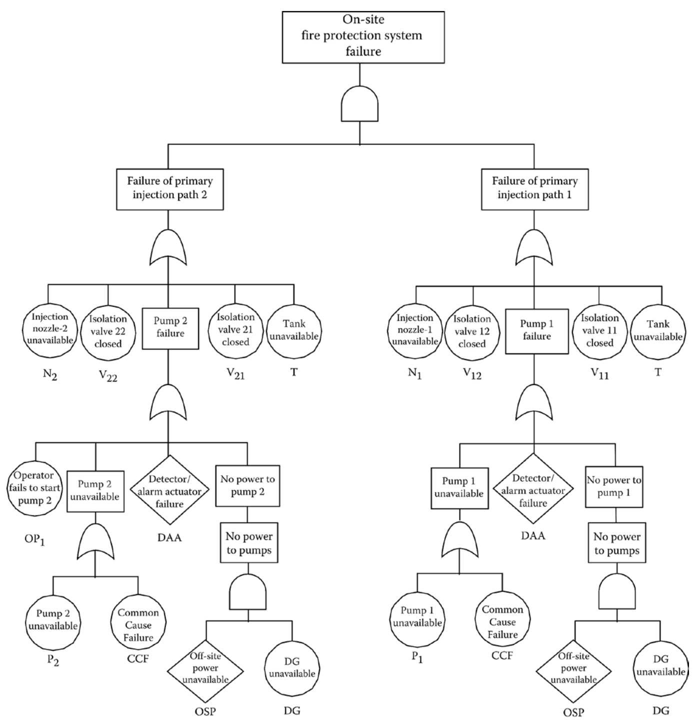


Example PRA: Fire protection system (4/6)

- **Failure data analysis:**
  - Find the basic event probabilities
  - Assume CCF b/w valves and nozzles, ≥10 hrs operation
- **Quantification and Interpretation:**
  - Find the cut sets and probabilities of the fault trees

Cut Set No.	Cut Set	Probability
1	XFD	**1×10^{-4}**
2	$OP_2 \cdot DAA$	**1×10^{-7}**

$$\text{Total Pr(OFF)} = \sum_i C_i \approx 1 \times 10^{-4}$$

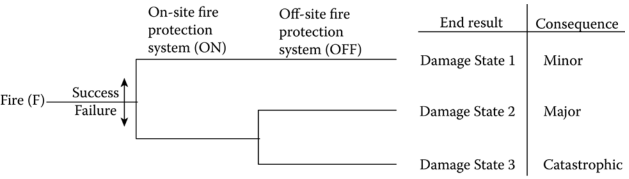


Cut Set #	Cut Set	Probability/(% of Total)	Cut Set #	Cut Set	Probability/(% of Total)
1	T	1.0×10^{-5} (0.36%)	13	$V_{21} \cdot V_{12}$	1.8×10^{-5} (0.65%)
2	DAA	1.0×10^{-4} (3.62%)	14	$V_{21} \cdot P_1$	6.7×10^{-5} (2.43%)
3	OSP·DG	6.1×10^{-6} (0.22%)	15	$V_{21} \cdot V_{11}$	1.8×10^{-5} (0.65%)
4	$N_2 \cdot N_1$	1.0×10^{-10} (~0%)	16	$OP_1 \cdot N_1$	1.0×10^{-7} (~0%)
5	$N_2 \cdot V_{12}$	4.2×10^{-8} (~0%)	17	$OP_1 \cdot V_{12}$	4.2×10^{-5} (1.52%)
6	$N_2 \cdot P_1$	1.6×10^{-7} (0.01%)	18	$OP_1 \cdot P_1$	1.6×10^{-4} (5.80%)
7	$N_2 \cdot V_{11}$	4.2×10^{-8} (~0%)	19	$OP_1 \cdot V_{11}$	4.2×10^{-5} (1.52%)
8	$V_{22} \cdot N_1$	4.2×10^{-8} (~0%)	20	$P_2 \cdot N_1$	1.6×10^{-7} (0.01%)
9	$V_{22} \cdot V_{12}$	1.8×10^{-5} (0.65%)	21	$P_2 \cdot V_{12}$	6.7×10^{-5} (2.43%)
10	$V_{22} \cdot P_1$	6.7×10^{-5} (2.43%)	22	$P_2 \cdot P_1$	2.6×10^{-4} (9.42%)
11	$V_{22} \cdot V_{11}$	1.8×10^{-5} (0.65%)	23	$P_2 \cdot V_{11}$	6.7×10^{-5} (2.43%)
12	$V_{21} \cdot N_1$	4.2×10^{-8} (~ 0%)	24	CCF	1.8×10^{-3} (65.2%)

$$\textbf{Total } Pr(ON) = \sum_i C_i \approx 2.8 \times 10^{-3}$$

Example PRA: Fire protection system (5/6)

- With the cut sets from the fault trees, we find the cut sets of each scenario:



Scenario 1 $= F \cdot \overline{ON}$

Scenario 2 $= F \cdot ON \cdot \overline{OFF}$

Scenario 3 $= F \cdot ON \cdot OFF$

Scenario No.	Cut Sets	Frequency	Comment
1	$F \cdot \overline{ON}$	$7.1\times10^{-4}(1-2.8\times10^{-3}) \approx 7.1\times10^{-4}$	Since the probability can be directly evaluated for $\overline{ON}$ without the need to generate cut sets, only the probability is calculated
2	$F \cdot DAA \cdot \overline{XFD} \cdot \overline{OP_2}$	7.1×10^{-8}	1. Only cut sets that have a contribution greater than 1% of total are shown. 2. Cut set $F \cdot DAA \cdot \overline{XFD} \cdot \overline{DAA}$ is eliminated since $DAA \cdot \overline{DAA} = \phi$
	$F \cdot V_{22} \cdot P_1 \cdot \overline{XFD} \cdot \overline{OP_2}$	4.8×10^{-8}	
	$F \cdot V_{21} \cdot P_1 \cdot \overline{XFD} \cdot \overline{OP_2}$	4.8×10^{-8}	
	$F \cdot V_{22} \cdot P_1 \cdot \overline{XFD} \cdot \overline{DAA}$	4.8×10^{-8}	
	$F \cdot V_{21} \cdot P_1 \cdot \overline{XFD} \cdot \overline{DAA}$	4.8×10^{-8}	
	$F \cdot OP_1 \cdot V_{12} \cdot \overline{XFD} \cdot \overline{OP_2}$	3.0×10^{-9}	
	$F \cdot OP_1 \cdot V_{12} \cdot \overline{XFD} \cdot \overline{DAA}$	3.0×10^{-9}	
	$F \cdot OP_1 \cdot P_1 \cdot \overline{XFD} \cdot \overline{OP_2}$	1.1×10^{-7}	
	$F \cdot OP_1 \cdot P_1 \cdot \overline{XFD} \cdot \overline{DAA}$	1.1×10^{-7}	
	$F \cdot OP_1 \cdot V_{11} \cdot \overline{XFD} \cdot \overline{OP_2}$	3.0×10^{-9}	
	$F \cdot OP_1 \cdot V_{11} \cdot \overline{XFD} \cdot \overline{DAA}$	3.0×10^{-9}	
	$F \cdot P_2 \cdot V_{12} \cdot \overline{XFD} \cdot \overline{OP_2}$	4.8×10^{-8}	
	$F \cdot P_2 \cdot V_{12} \cdot \overline{XFD} \cdot \overline{DAA}$	4.8×10^{-8}	
	$F \cdot P_2 \cdot P_1 \cdot \overline{XFD} \cdot \overline{OP_2}$	1.8×10^{-7}	
	$F \cdot P_2 \cdot P_1 \cdot \overline{XFD} \cdot \overline{DAA}$	1.8×10^{-7}	
	$F \cdot P_2 \cdot V_{11} \cdot \overline{XFD} \cdot \overline{OP_2}$	4.8×10^{-8}	
	$F \cdot P_2 \cdot V_{11} \cdot \overline{XFD} \cdot \overline{DAA}$	4.8×10^{-8}	
	$F \cdot CCF \cdot \overline{XFD} \cdot \overline{OP_2}$	1.3×10^{-6}	
	$F \cdot CCF \cdot \overline{XFD} \cdot \overline{DAA}$	1.3×10^{-6}	
	2 Total	$\sum C_i = 3.7\times10^{-6}$	
3	$F \cdot DAA \cdot OP_2$	7.1×10^{-11}	1. Only cut sets that have a contribution greater than 1% of total are shown.
	$F \cdot DAA \cdot XFD$	7.1×10^{-12}	
	$F \cdot V_{22} \cdot P_1 \cdot XFD$	4.8×10^{-12}	
	$F \cdot V_{21} \cdot P_1 \cdot XFD$	4.8×10^{-12}	
	$F \cdot OP_1 \cdot V_{12} \cdot XFD$	3.0×10^{-12}	
	$F \cdot OP_1 \cdot P_1 \cdot XFD$	1.1×10^{-11}	
	$F \cdot OP_1 \cdot V_{11} \cdot XFD$	3.0×10^{-12}	
	$F \cdot P_2 \cdot V_{12} \cdot XFD$	4.8×10^{-12}	
	$F \cdot P_2 \cdot P_1 \cdot XFD$	1.8×10^{-11}	
	$F \cdot P_2 \cdot V_{11} \cdot XFD$	4.8×10^{-12}	
	$F \cdot CCF \cdot XFD$	1.3×10^{-10}	
	3 Total	$\sum C_i = 2.6\times10^{-10}$	

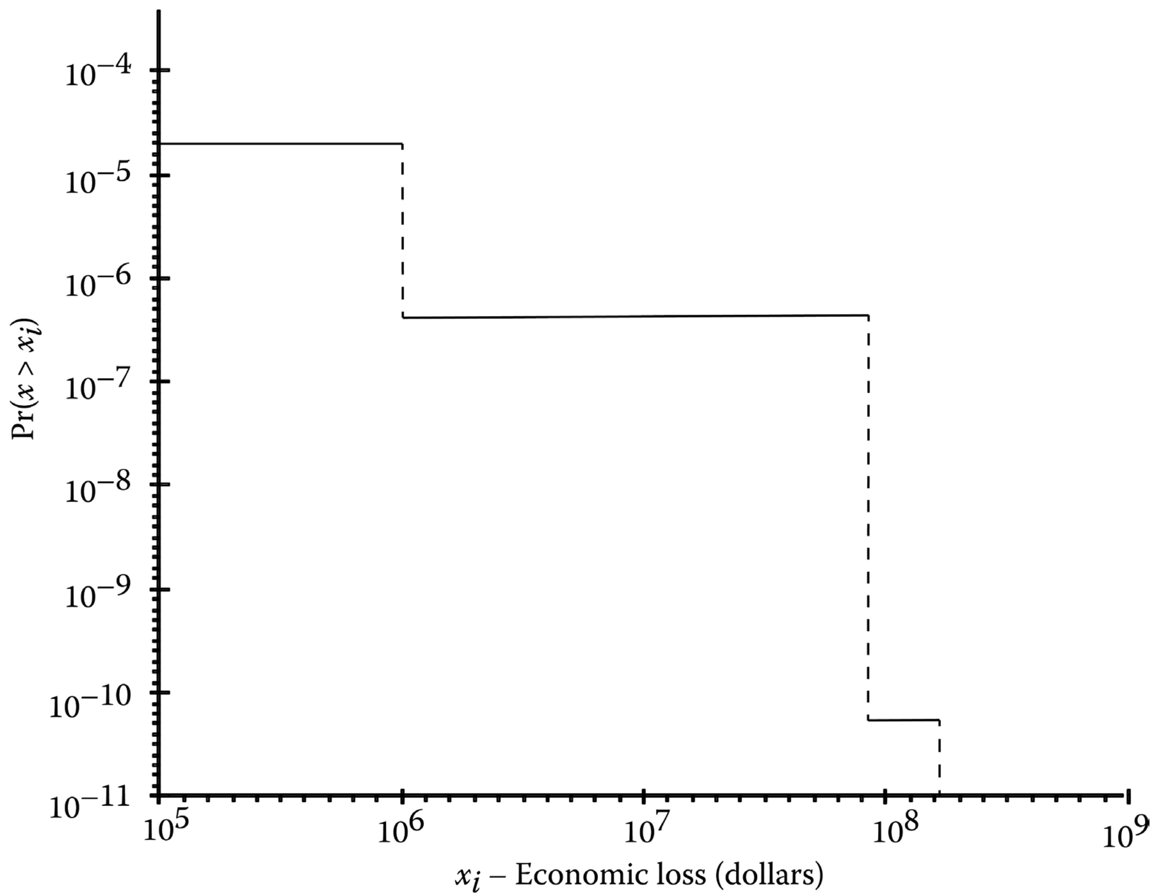
- **Consequences of each scenario:**

Scenario #	Economic Consequence
1	\$1,000,000
2	\$92,000,000
3	\$210,000,000

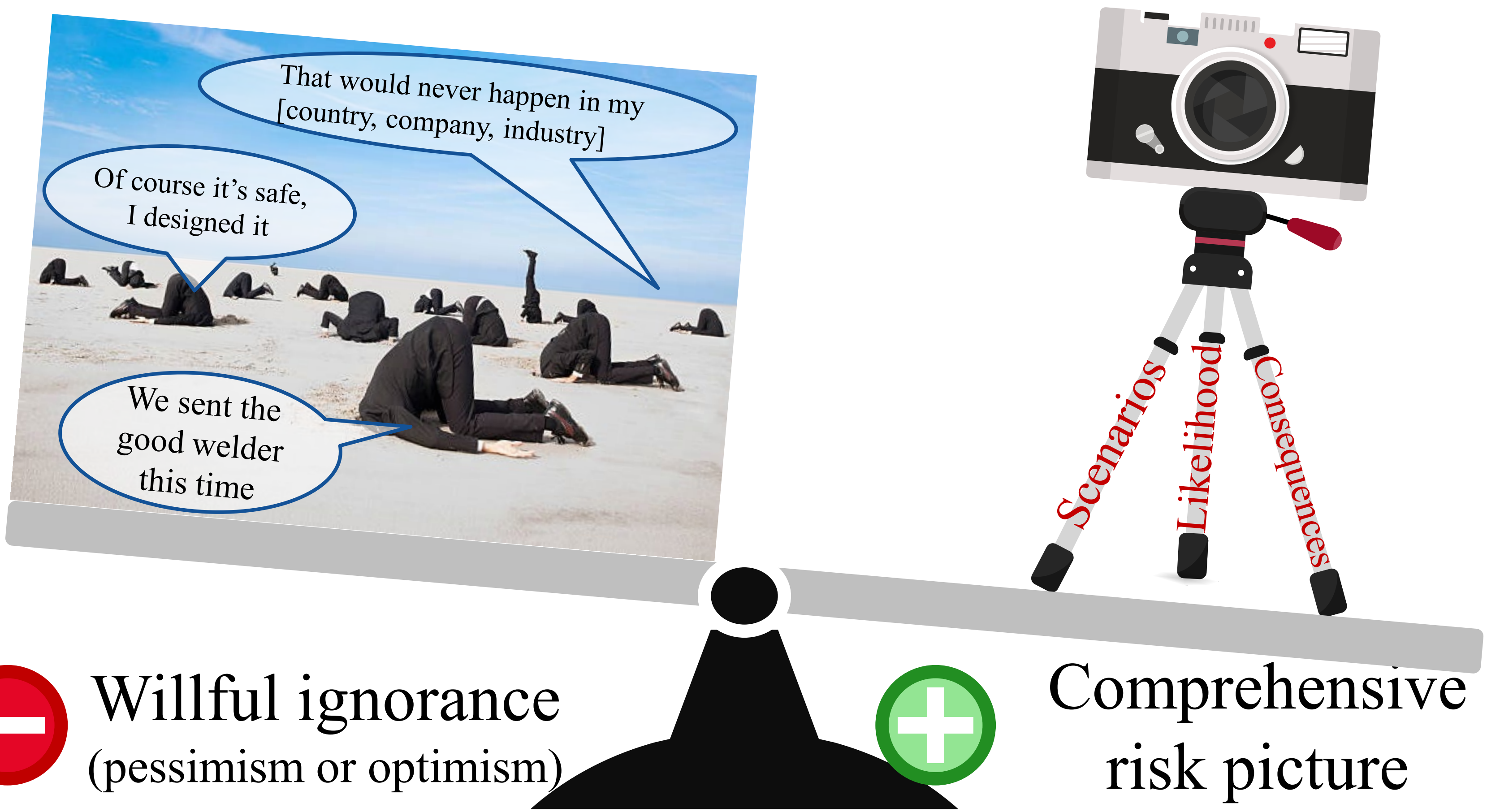
- **Risk associated with each scenario:**

Scenario #	Economic Consequence
1	$(7.1\times10^{-4})(\$1{,}000{,}000) = \$710.00\ \text{yr}^{-1}$
2	$(3.7\times10^{-6})(\$92{,}000{,}000) = \$340.00\ \text{yr}^{-1}$
3	$(2.6\times10^{-10})(\$210{,}000{,}000) = \$0.05\ \text{yr}^{-1}$

Example PRA: Fire protection system (6/6)



Principles



Willful ignorance
(pessimism or optimism)

Comprehensive risk picture

Revisit Quiz 1: Re-Write your failure scenario for the event you discussed on day 1 now that you have more knowledge from this class.

- Your scenario **must include**:
  - Event/accident name, date, description (1-2 sentences), and photo
  - System failure mode, operating environment, consequence(s)
  - At least one of: hardware failure mode, failure mechanism, environment, and human failure
- Example: Air France 447 Airbus A330 (July 1, 2009)

Component & failure mode | Component failure mechanism

Flow blockage of **pitot tubes,** caused by **freezing,** led to

Subsystem failure mode: **erroneous airspeed indication** resulting in **loss of control,** — System failure mode

loss of aircraft, fatality of all 228 passengers and crew, — Consequences

with pilot failure to diagnose the erroneous output due to inadequate information, and time load. — Human failure mode & PIFs

Occurred at **high altitude, low temperature, in stormy weather.** — Environment

Image: Brazilian Navy/ Getty Images

Revisiting Quiz 1

- Could you develop an event tree for the scenario? What about a fault tree? Which events would be placed where?
- Where could you get data for the events that happened in your failure scenario?

Closing thoughts for this course

The future of PRA

Engineering a safer world together