It Depends

An Overview of Probabilistic Analysis Concepts

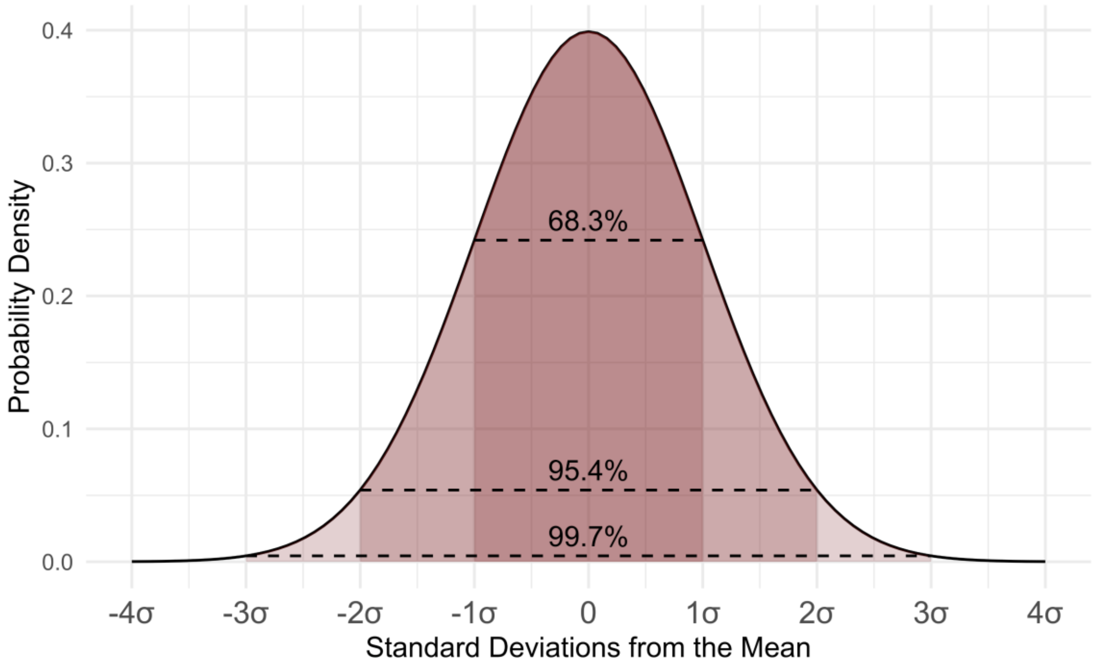
Richard Meshell – Bastion Technologies, Inc.





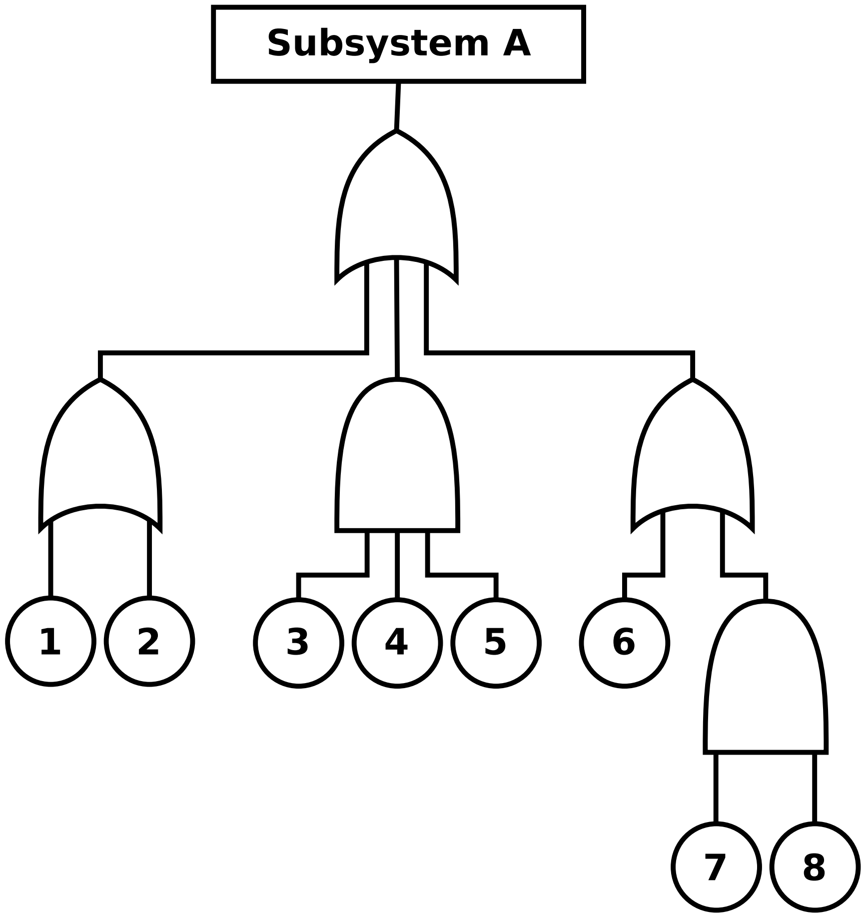
Introduction

- Do you really understand the data that is being used in your probabilistic analysis?
- The PRA analyst has a need to understand the data and the input assumptions being used in order to effectively communicate the risks of the system/mission.
- The decision maker also needs to understand conservatisms and how to interpret uncertainty in the numbers being presented.
- The goal of this presentation is for some familiarization with statistical concepts frequently used in probabilistic analysis



Probabilistic Risk Assessment Background

- PRA was developed over the course of the 20th century with advances in systems engineering but was born out of a need to better understand risk/severity/consequence. In the 50’s and 60’s you had both construction of the first nuclear plants and also some of the first commercial jet airliners. In the 1970’s, you have some of the first academic/technical papers published on the topic. Most famously was Wash-1400, which is seen as the first nuclear PRA in 1975. 1979 was the Three Mile Island accident, which placed a greater emphasis on understanding system complexity.
- NASA began using PRA in the 1980’s, but it wasn’t used in its current capacity until well into the shuttle program.
- PRA is also used in the Oil and Gas Industry to assess hardware and surveillance risks.



Roadmap

- Uncertainties
- Bayesian Statistics
- Common Cause Factors
- Generic Sources

BASTION TECHNOLOGIES

Uncertainty

- **Uncertainty is an inherent property of measured data**
  - The source of the uncertainty can come from plant behavior in unanalyzed conditions, human performance, changes in performance over time, environmental conditions.
  - Ruler Example
- **Generic data is often specified by Mean and Error Factor**
- **Error Factor defines dispersion about the median**
  - Square Root of the $95^{th}/5^{th}$
  - $95^{th}/50^{th}$ or $50^{th}/5^{th}$
- **Sometimes you need to assume values for the Error Factor**
  - Estimating the uncertainty in a single measurement requires judgement on the part of the PRA Analyst
  - Rule of Thumb: Error Factor 5 for generic data, Error Factor 10-15 for expert solicitation
    - "Data Applicability of Heritage and New Hardware for Launch Vehicle Reliability Models", Hassan, Novak, RAMS VIII, 2015-11-3

EPA FAQ Uncertainty

- **Uncertainty is the lack of understanding of the world; while unavoidable it can be reduced through additional investigation or collection of better information.**
- **3 Categories of Uncertainty**
  - Input or parameter uncertainty refers to uncertainties in specific estimates or values used in a mode, such as the average drinking water intake rate
  - Model Uncertainty refers to gaps in the scientific knowledge or theory that is required to make accurate predictions, such as how to two correctly specify exposures to water.
  - Scenario Uncertainty refers to errors, typically of omission, resulting from incorrect or incomplete specification of the risk scenario to be evaluated such as errors in the problem formulation omitting indirect water ingestion exposure pathways. The risk scenario is a set of assumptions for the situation to be evaluated.
- EPA White Paper - PRA

EPA Case Study

- **Case Study #3 and #5**
- **PRA of Exposure to Polychlorinated Biphenyls via Consumption of Fish From a Contaminated Sediment Site**
  - PCBs used as a fire preventative and insulator.
  - Banned for use by the EPA in 1977
  - During the 30yrs prior, estimated that 1.3million pounds were dumped into the Hudson River
- **A deterministic approach informed that consumption of recreationally caught fish provided the highest exposure among relevant exposure pathways, which resulted in cancer risks and noncancer health hazards that exceeded regulatory benchmarks.**
- **Conducted a PRA characterize the variability in risks associated with the fish consumption exposure pathway.**
- **PRA concluded that while the central tendency values agreed, the RME (reasonable maximum exposure) that was determined fell outside the 95th percentile.**

EPA Case Study

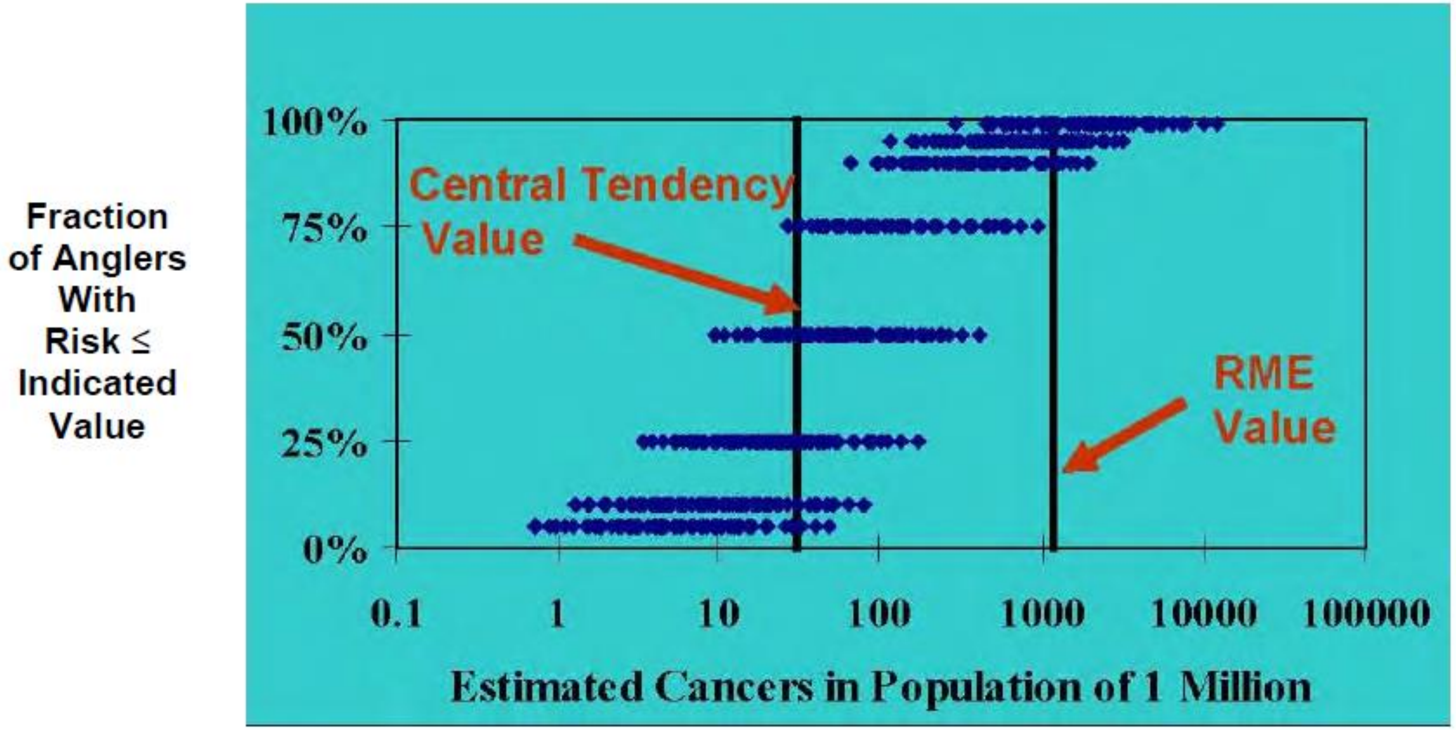
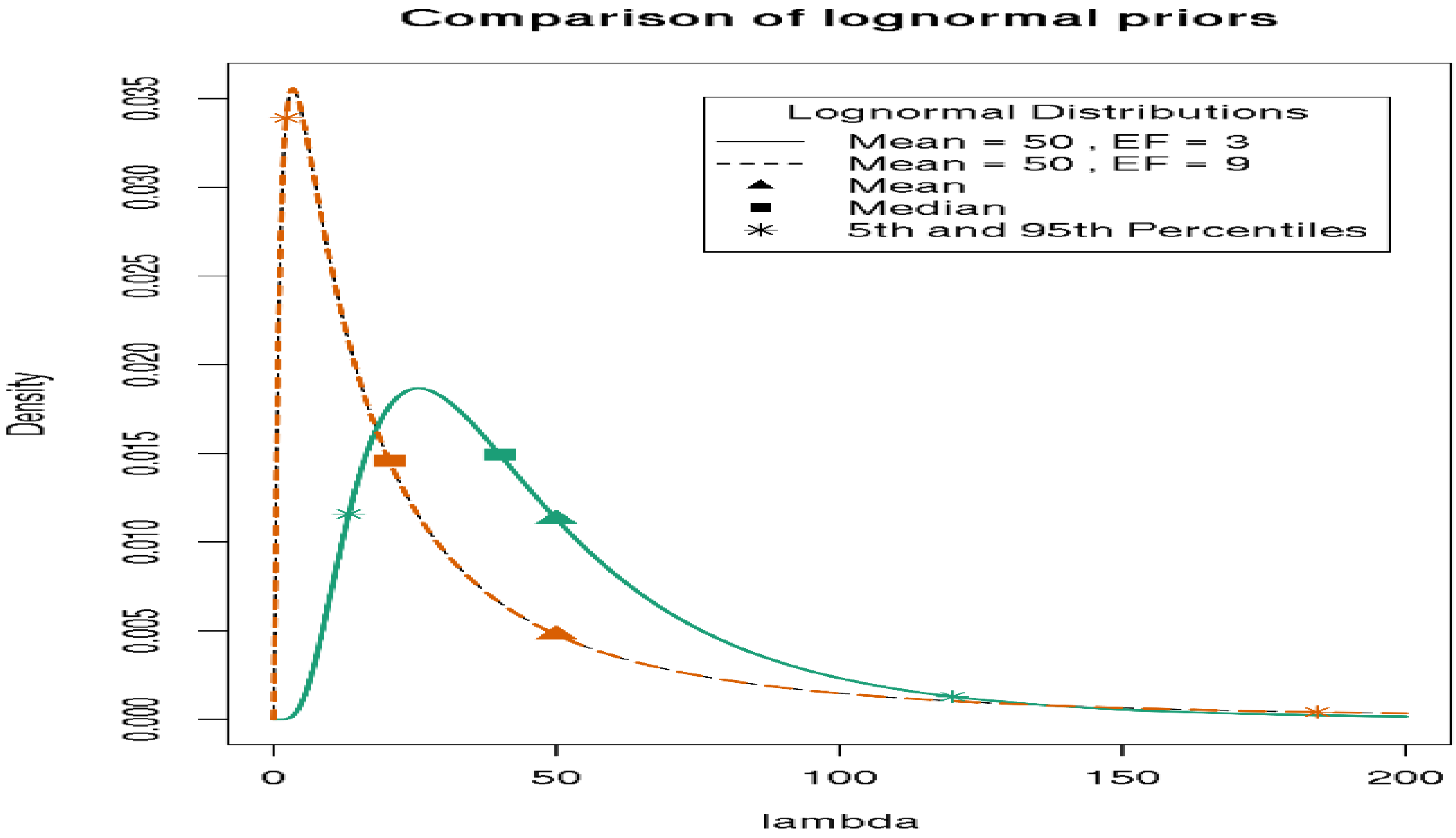


Figure A-1. Monte Carlo Cancer Summary Based on a One-Dimensional Probabilistic Risk Analysis of Exposure to Polychlorinated Biphenyls. The estimated cancer rate was calculated based on the consumption of fish from a contaminated sediment site. Source: USEPA 2000b.

Same Mean, Difference Error Factors



Uncertainties

- <u>Takeaway</u>: Uncertainty is much part of the PRA process as system modeling. It answers the question "how well is your data understood by your model”
- Question the level at which you believe you understand your data.
- This understanding propagates through risk models and ultimately affects your results and decisions

Bayesian Statistics

- Formulated by Thomas Bayes, a 17th century statistician, it wasn’t until 2 years after his death in 1763 that “An Essay towards solving a Problem in the Doctrine of Chances” was presented to the Royal Society by Richard Price.
- Pierre-Simon Laplace independently discovered Bayes Theorem in 1774, to which formulated the equation and description “the probability of a cause (given an event) is proportional to the probability of the event (given its cause).”

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

- Laplace eventually learned of Bayes’ Theorem, further confirming his theory
- Laplace's friend Bouvard used his method to calculate the masses of Jupiter and Saturn from a wide variety of observations

Bayesian Statistics

- **Highly reliable components produced in small quantities, such as in space applications, do not have enough operating time and failure history to yield useful confidence bounds using classical data analysis methods.**
- **Bayesian inference is a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available.**
- **Bayesian inference derives the posterior probability as a consequence of two antecedents: a prior probability and a "likelihood function" derived from a statistical model for the observed data**

BASTION TECHNOLOGIES

Most Common Distributions in PRA

- **Discrete**
  - Poisson
    - **Used to predict # of occurrence of events governed by a constant failure rate**
  - Binomial
    - **Commonly used model for Fail To Start or failure to change state (demand)**
- **Continuous**
  - Exponential
    - **Used to model the random time to failure of operating components under assumption of a constant failure rate**
  - Lognormal
    - **Used to express the uncertainty in many PRA data parameters such as failure rates, initiating event frequencies, basic event probabilities**
  - Beta
    - **Used to express uncertainty in common cause failure parameters (e.g. beta factors)**

Prior Distributions

- **Priors heavily influence the posterior distributions**
- **Informative Priors**
  - Expresses specific, definite information about a variable
  - This is often existing knowledge of the component (surveillance on existing equipment, etc.)
- **Non-Informative/Weak Priors/Minimally Informative**
  - Expresses vague or general information about a variable
  - Jeffries and Constrained Non-Informative (generalized Jeffries) are used most often in PRA
  - You might want to use these if the situation begin evaluated is changing with time, then over-reliance on old information in formulating the prior can lead to priors that excessively dominate new data.
  - Constrained Non-Informative Priors, Atwood, INL 1994
  - Minimally Informative Prior Distributions for PSA, Kelly, Youngblood, Vedros, INL 2010

BASTION TECHNOLOGIES

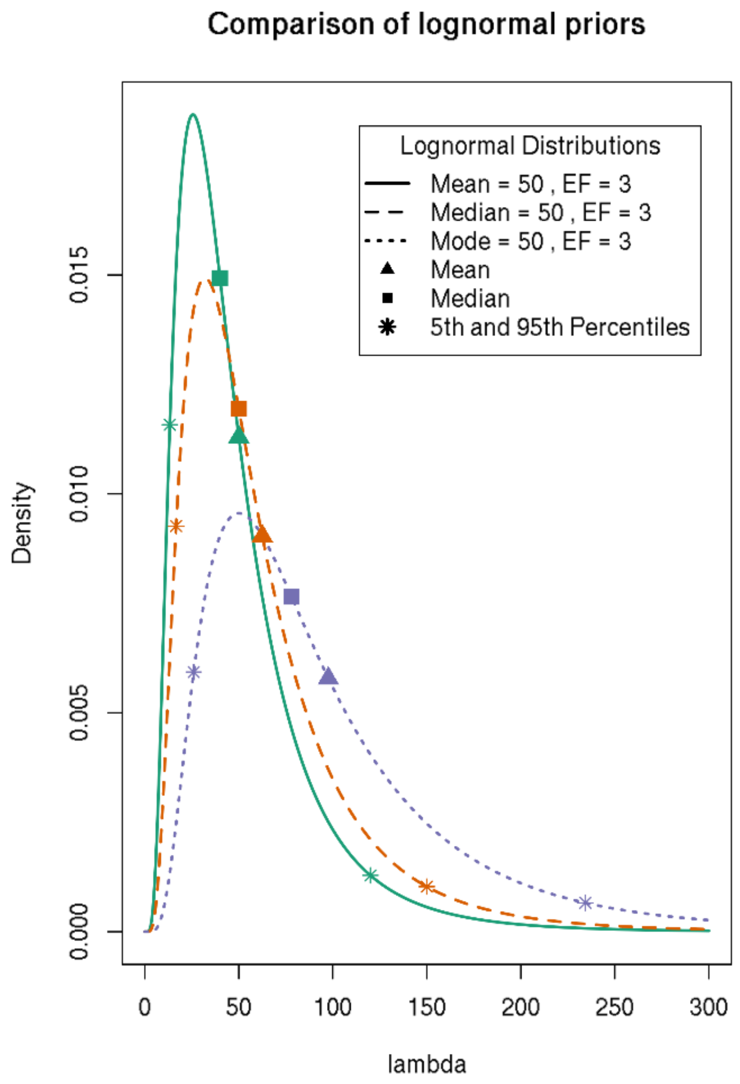
Mean, Median, Mode

- **Mean**
  - First Moment of your data set (Center of Mass), common measure of central tendency

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- **Median**
  - Middle value when data is sorted. The median provides information on where most of the data lies, thus is not sensitive to extreme values
  - The median will vary more from sample to sample than the mean
- **Mode**
  - The mode is the most frequently occurring value in the dataset. The mode is not influenced by extreme values or outliers
  - There may be more than one mode
  - Not typically used outside of Pareto Plotting

Prior Assumptions



- **Reliability predictions are reported as point estimates, and the PRA needs to estimate uncertainty**
- **This data is usually assumed to represent the mean failure rate, with an error factor applied using heuristic guidelines modeled with a lognormal distribution**
- **This guideline is generally true, but can lead to non-conservatisms in the uncertainty estimation.**
- **Illustrated here, an assumption of whether your data represents the mean/median/mode will greatly affect the shape of your distribution**

Bayesian Statistics Summary

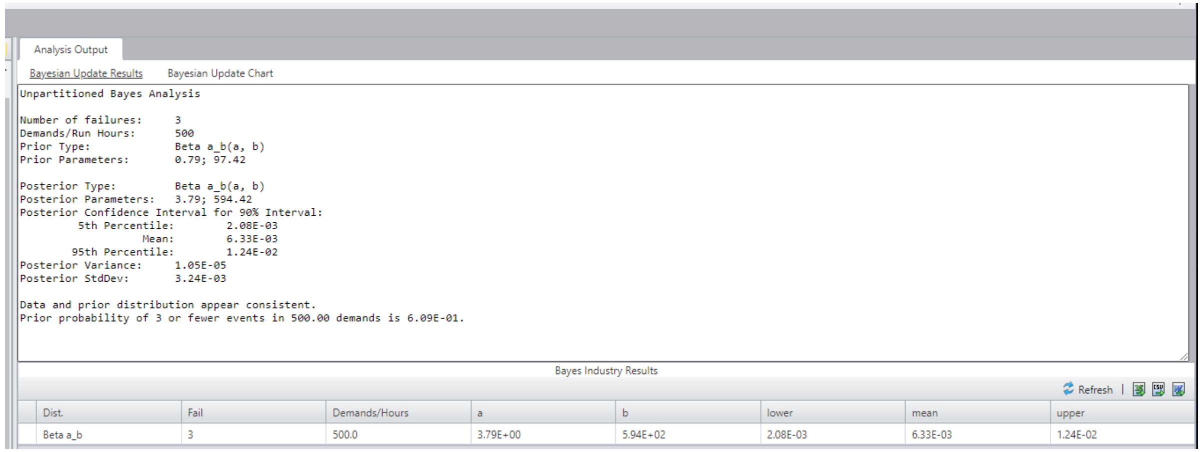
BASTION TECHNOLOGIES

- **<u>Takeaways</u>:**
  - Bayesian updating can give you a better idea of the nature of your component data given limited testing
  - In general, your system/component data represents the mean. However, plot your priors to make sure they make sense
  - You might assume the median during expert elicitation

- **Resources**
  - Uncertainty Estimation Cheat Sheet, Britton, Al Hassan, Ring NASA 2017

Bayesian Updating Tools

- **NRC Reliability Calculator is a great tool to do quick updating**
- **https://rads.inl.gov/calculator/Default.aspx**


Analysis Output

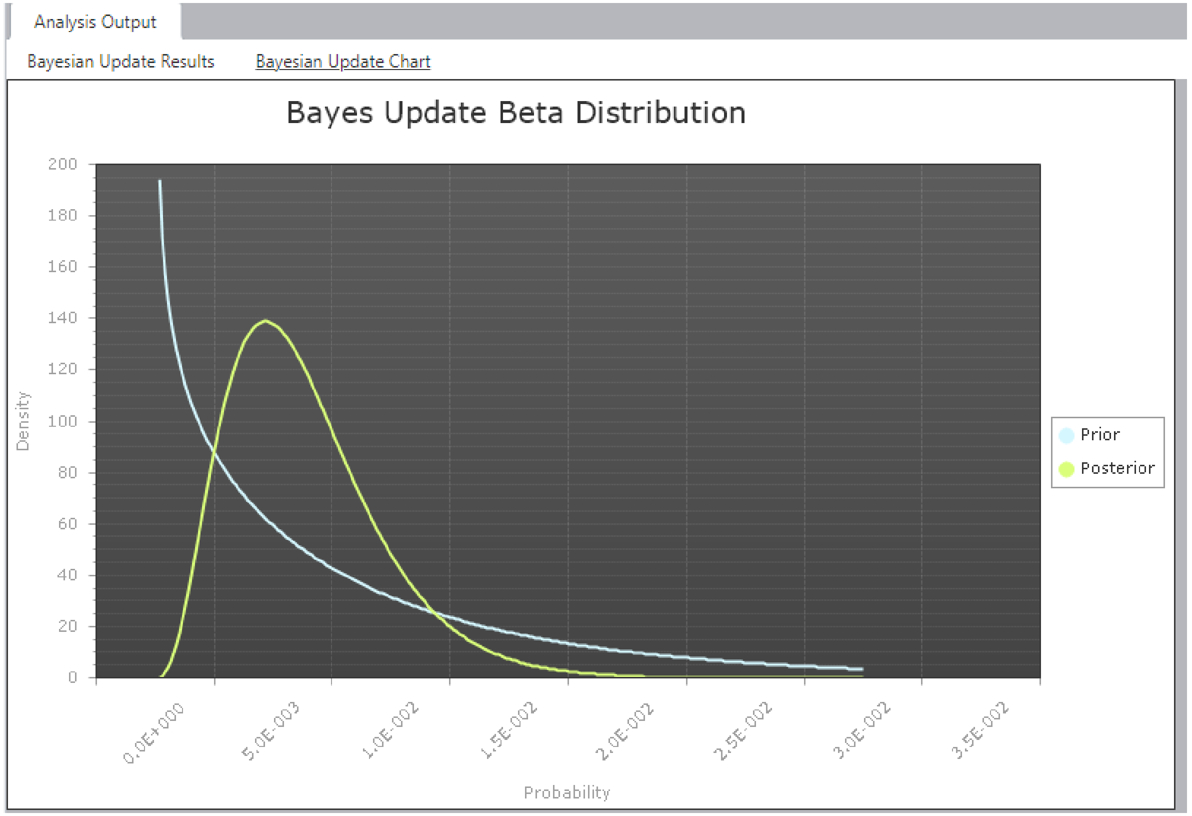
Bayesian Update Results   Bayesian Update Chart

```
Unpartitioned Bayes Analysis

Number of failures:     3
Demands/Run Hours:      500
Prior Type:             Beta a_b(a, b)
Prior Parameters:       0.79; 97.42

Posterior Type:         Beta a_b(a, b)
Posterior Parameters:   3.79; 594.42
Posterior Confidence Interval for 90% Interval:
         5th Percentile:        2.08E-03
                   Mean:        6.33E-03
        95th Percentile:        1.24E-02
Posterior Variance:     1.05E-05
Posterior StdDev:       3.24E-03

Data and prior distribution appear consistent.
Prior probability of 3 or fewer events in 500.00 demands is 6.09E-01.
```

Bayes Industry Results

Refresh

Dist.	Fail	Demands/Hours	a	b	lower	mean	upper
Beta a_b	3	500.0	3.79E+00	5.94E+02	2.08E-03	6.33E-03	1.24E-02



Consideration of Common Cause

- **A Common Cause Failure is a failure that meets 4 criteria (NUREG-5485)**
  - Two or more individual components fail or are degraded, including failures during demand, in-service testing, or deficiencies that would have resulted in a failure if a demand signal had been received
  - Components fail within a selected period of time such that success of the PRA mission would be uncertain
  - Component failures result from a single shared cause and coupling mechanism
  - A component failure occurs within the established component boundary
- **More succinctly, it can be thought of as a failure mechanism that defeats redundancy, a special class of dependent failure**
- **Very often, common cause failures are risk drivers in the PRA cutsets**

Examples of Common Cause Failures

- **B-747 4-engine shutdown from volcanic plume (1989)**
- **Fukushima Daiichi (tsunami wiping out redundant systems)**
- **Three Mile Island Accident**
  - Redundant AFWS Pumps Valved out
  - Operator Switched off redundant SI Pumps

Common Cause Modeling

- **Beta Factor**
  - The item failure rate λ is split into an independent part, λ_i, and a dependent part, λ_c, such that:
    - **λ= λ_i+ λ_c**
  - The beta factor is the fraction of all item failures that are common cause failures. Can also be interpreted as the conditional probability that the failure is a CCF
  - Systems with 2 components
- **MGL Method**
  - Mathematical extension of the Beta Factor to treat multiple levels of CCF (i.e., failures involving CCF of at least 2, 3, 4…etc components)
  - Appropriate for systems with a lot of redundancy

$$\beta = \frac{\sum_{k=2}^{m} kN_k}{\sum_{k=1}^{m} kN_k} \qquad \Upsilon = \frac{\sum_{k=3}^{m} kN_k}{\sum_{k=2}^{m} kN_k} \qquad \Delta = \frac{\sum_{k=4}^{m} kN_k}{\sum_{k=3}^{m} kN_k}$$

- where N_k is the number of <u>events</u> involving the failure of exactly k components. Therefore, kN_k is the number of failed <u>components</u>

- **Alpha Factor**
  - Simple expressions and point estimators are not always attainable for MGL
  - The fraction of all failure events that occur in a group of *m* items and involve exactly *k* items due to common cause.

$$\alpha_{k:m} = \frac{\binom{m}{k} \cdot Q_{k:m}}{\sum_{j=1}^{m} \binom{m}{j} \cdot Q_{j:m}}$$

Common Cause Summary

- **Takeaway: Common cause failures are failures that defeat redundancy.**
- **CCF analysis is an additional insight in the fault tree, and can tell you if your system is more/less resilient to CCF**
- **Resources**
  - Common Cause Failures, Rousand
  - PRA Basics for Regulatory Applications P-105, Calley, Knudsen, INL 2016
  - NUREG-5485, “Guidelines on Modeling Common-Cause Failures in PRA”

Conservative Numbers

- **“Risk Is This Or Less”**
- **PRA is a great tool to have early in the design process, however limited component data and system information can misinform how well the engineered system is understood.**
- **Generic component data can be used in lieu of testing data to provide insight to system models.**
  – As testing data becomes available, Bayesian Updating can be used.
- **Relative risk insights can provide insight to how design changes affect the risk**
  – Compare risk impacts of competing changes

BASTION TECHNOLOGIES

Generic Numbers

- **Open Source Generic Data Sources**
- **Nonelectronic Parts Reliability Data (NPRD) – 2011**
- **MIL-HDBK-217 for electrical components**
- **NUREG/CR-6928 – Initiating Event & Component Data**
  - NRC Industry Average Parameter Estimation - 2020
- **Nuclear Reactor Operating Experience Data (NROD)**
  - Registration Required
- **Several paid options for component reliability data**

Conclusion

- **A better understanding of PRA techniques and tools can give analysts and decision makers insights to system risks results.**
- **Bayesian Statistics, Uncertainty Analysis, and Common Cause Modeling are all methodologies that are integral to the risk modeling, but are difficult to present and report in a concise manner.**
  - Car Buyers

Backup