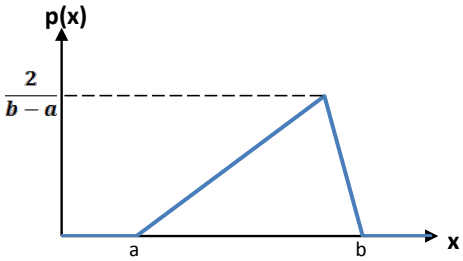
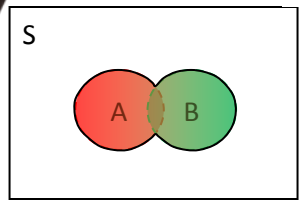
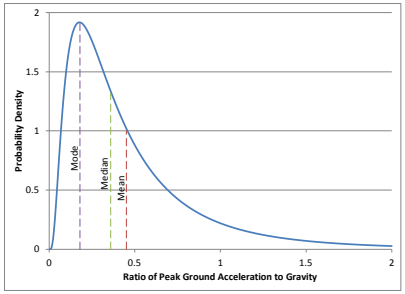
Best Practices in Dam and Levee Safety Risk Analysis

I-1. Basics of Probability and Statistics

29 May 2015







U.S. DEPARTMENT OF THE INTERIOR
BUREAU OF RECLAMATION

RECLAMATION
Managing Water in the West

US Army Corps of Engineers
BUILDING STRONG®

Objectives

- Understand the meaning and application of the probability estimates that are routinely made in risk analysis
- Become familiar with some commonly used definitions and terms
- Become familiar with the basic methods that can be used to support credible risk estimates

Outline

- Key concepts
- Deductive and inductive logic
- Set theory
- Combinatorics
- Probability and statistics
- Bayes Theorem
- Uni-modal Bounds
- Bringing it all together

Key Concepts

- **Events** are the basic building blocks of risk analysis. One way to describe an **event** is as something that could happen, projected into the present as certain (no matter how likely or unlikely in reality).
- e.g. “A spillway release could occur on March 25th” can be projected into the present as “A spillway release occurs on March 25th”

Key Concepts

- The following are all examples of events
  - “Jim orders a pizza”
  - “Jim orders a pizza or a hot dog and Bob does not come into work on Friday”
  - “A flood occurs”
  - “Internal Erosion initiates in the year under consideration”
  - “The maximum reservoir elevation in 2016 is less than 321 feet”

Key Concepts

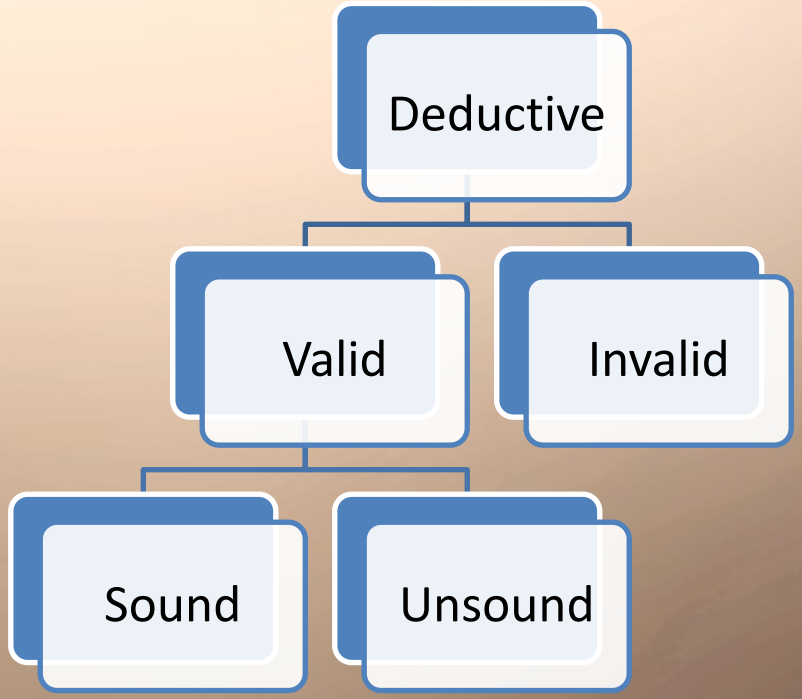
- **Set Theory** provides a framework for the analysis of events and the relationships between events. It is based entirely on logic.
- **Probability Theory** provides a framework for analyzing the *likelihoods* of events and different combinations of events. It is based on set theory and math.
- **Statistics** is the branch of science that deals with the interpretation and analysis of data.

Key Concepts

- **Risk Analysis** combines the key concepts of *Set Theory*, *Probability Theory*, and *Statistics* with the engineering judgment of a team, with the objective of improving understanding of system performance
- Risk estimates must be combined in a way that is technically correct so that the collective impact can be expressed in a credible manner
- Risk estimates must be portrayed in a way that can be readily understood by decision makers so that the appropriate actions may be taken

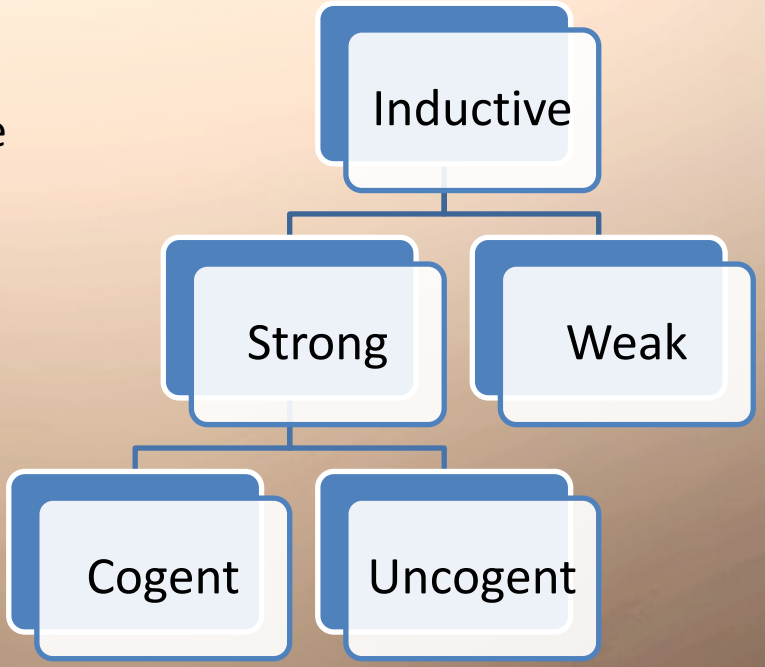
Deductive Arguments

- If the premise is true, then the conclusion is guaranteed to be true
- Valid
  - Cannot have true premises and false conclusion
- Sound
  - Valid and the premises must be true
- Invalid and valid but unsound deductive arguments can have a true conclusion
- Deductive arguments provide proof, but premises can rarely be verified



Inductive Arguments

- Evidence supports conclusions
- No guarantee
  - Confidence increases with more evidence
- Confidence (or degree of belief) can be measured
  - Willingness to take action
  - Probability estimates
- Validity judged by decision maker preferences
- Inductive arguments are supported by evidence, but proof cannot be obtained

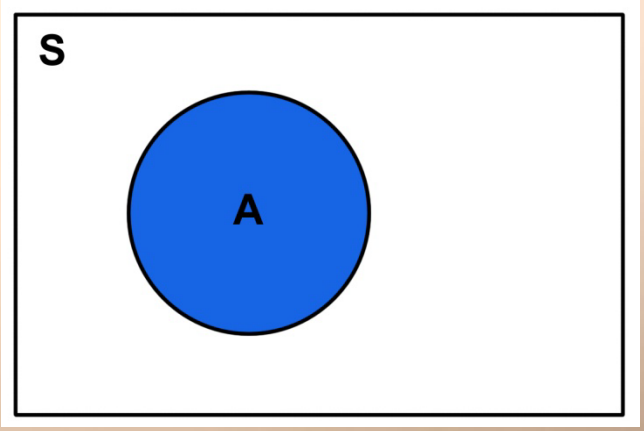


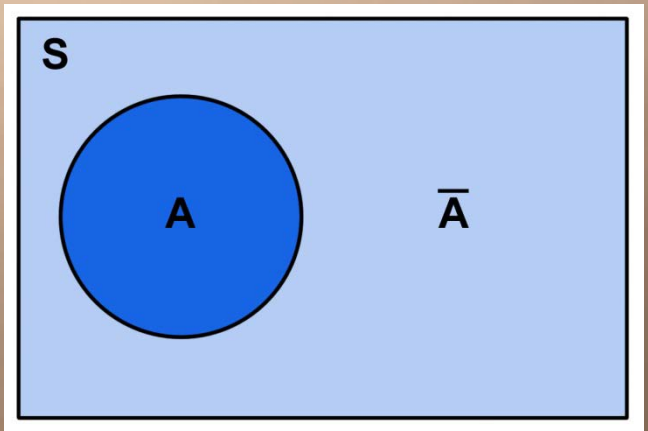
Set Theory

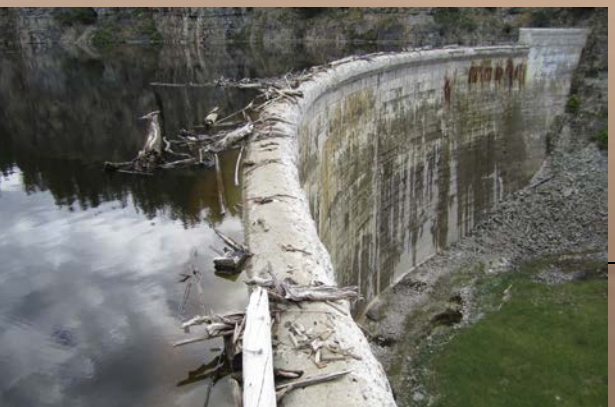
- Framework for evaluating relationships between events
- Set theory concepts are used (implicitly or explicitly) in every risk analysis
- Set theory concepts can be illustrated using Venn diagrams
  - Sample space represented by rectangle
  - Events and their relationships depicted by overlapping circles or other closed shapes

Set Theory Basics

- Event A: “Alpha Dam overtops in 2013”
- Event $\bar{A}$, “Alpha Dam does not overtop in 2013”, is the complement of A

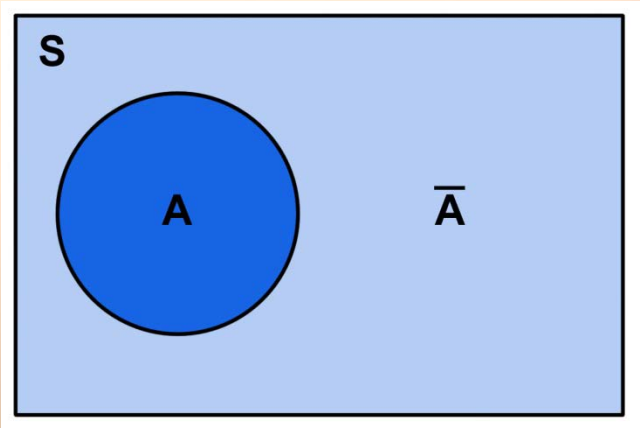




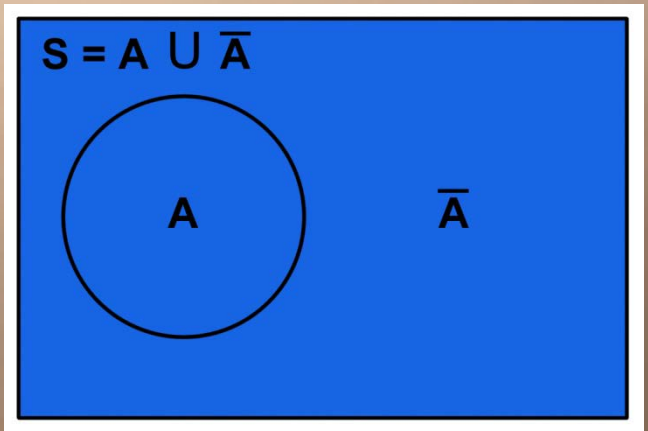


Set Theory Basics

- An event and its complement are <u>M</u>utually <u>E</u>xclusive (cannot both happen)

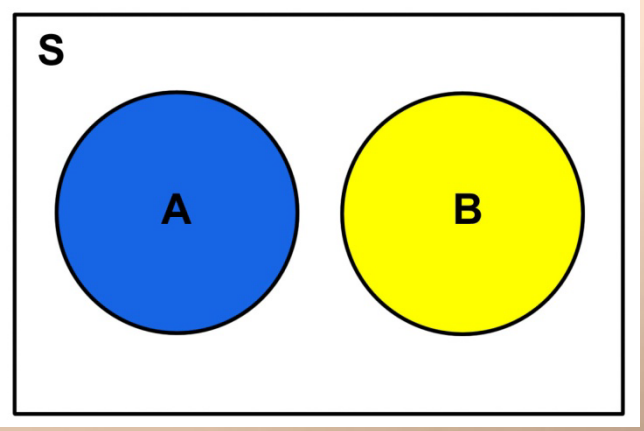


- An event and its complement are <u>C</u>ollectively <u>E</u>xhaustive (their union covers S)

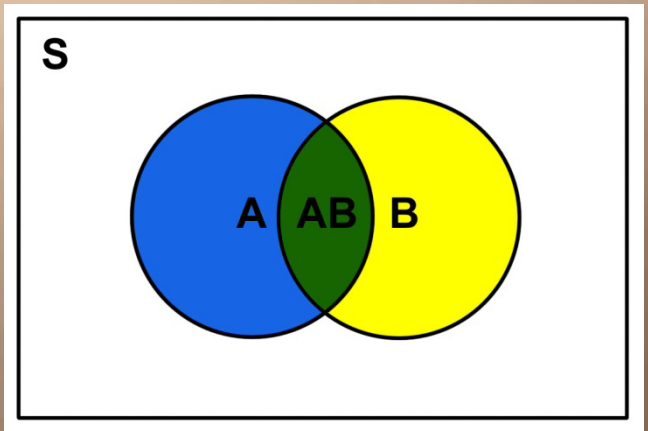


Set Theory Basics

- Event B, “Beta Creek runs dry in 2013”. A, B shown as ME

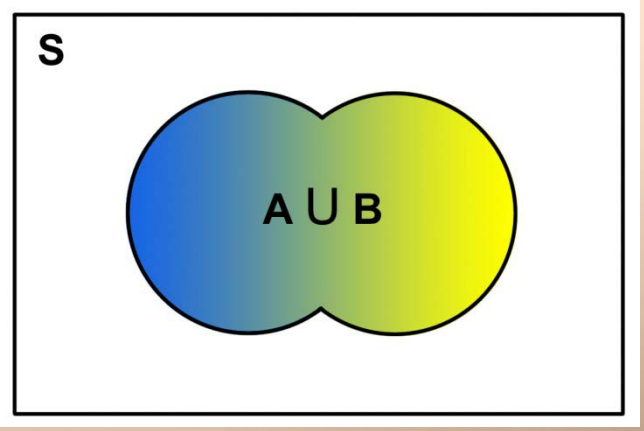


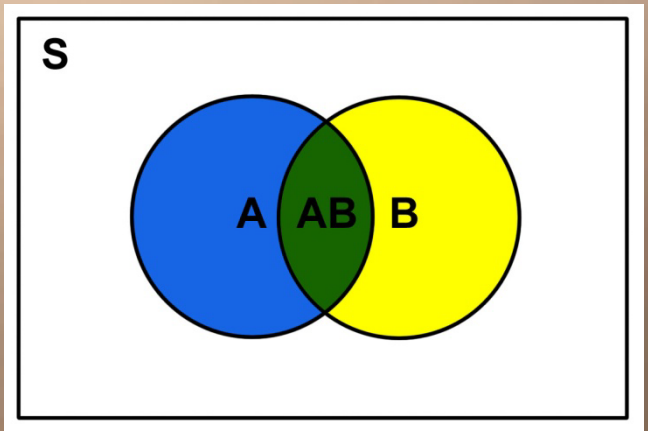
- A ∩ B (also written AB), the intersection event: “Alpha Dam overtops in 2013 **and** Beta Creek runs dry in 2013”



Set Theory Basics

- The event A **U** B, "Alpha Dam overtops in 2013 **or** Beta Creek runs dry in 2013"
- Note that the event A **U** B includes three possibilities: A, B, AB



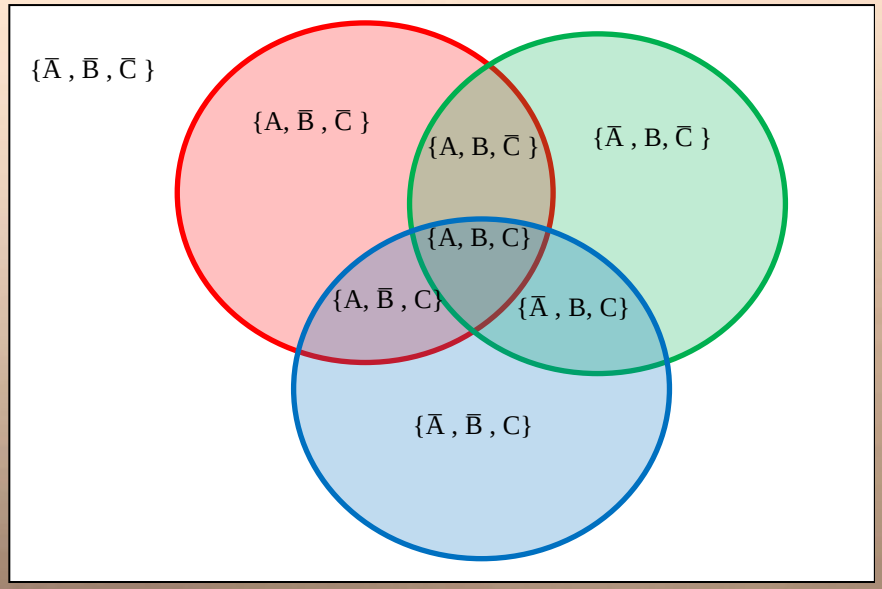


Combinatorics

- Enumeration, combination, and permutation of events
- Applicable to risk analysis
  - Describe and understand event combinations
- Concepts illustrated using Venn diagrams

Combinatorics

- Eight permutations might be obtained from three potential failure modes



Permutation	System Response for Potential Failure Mode A	System Response for Potential Failure Mode B	System Response for Potential Failure Mode C
1	No Breach	No Breach	No Breach
2	Breach	No Breach	No Breach
3	No Breach	Breach	No Breach
4	No Breach	No Breach	Breach
5	Breach	Breach	No Breach
6	Breach	No Breach	Breach
7	No Breach	Breach	Breach
8	Breach	Breach	Breach

Physical (frequency) probability

- Assumes stable frequency over numerous trials
- Strictly speaking, requires observation of the event in experiments or trials
- E.g. after rolling a standard six sided die many times, a risk analyst might estimate that the probability of rolling a two is about 1/6
- Similarly, if the RWS has exceeded 344 ft. three times in the last decade, then the probability of exceeding it next year is probably about 0.3

Degree of belief (subjective) probability

- Rational weighting of evidence as manifested by willingness to take action
- E.g. a risk analyst might still estimate the probability of rolling a two is about 1/6 based on:
  - Assumption – Die is fair
  - Observation – The geometry of the cube
  - Judgment – Past experience with similar dice
- Similarly, might conclude that a concrete arch dam is likely to crack under a 0.6g PHA based on the observed condition of the dam, an assumed depth of damage, and prior experience with low-strength concrete

Probability Axioms

- Probabilities are non-negative real numbers

$$P(A) \geq 0$$

- Probability of the certain event is 1.0

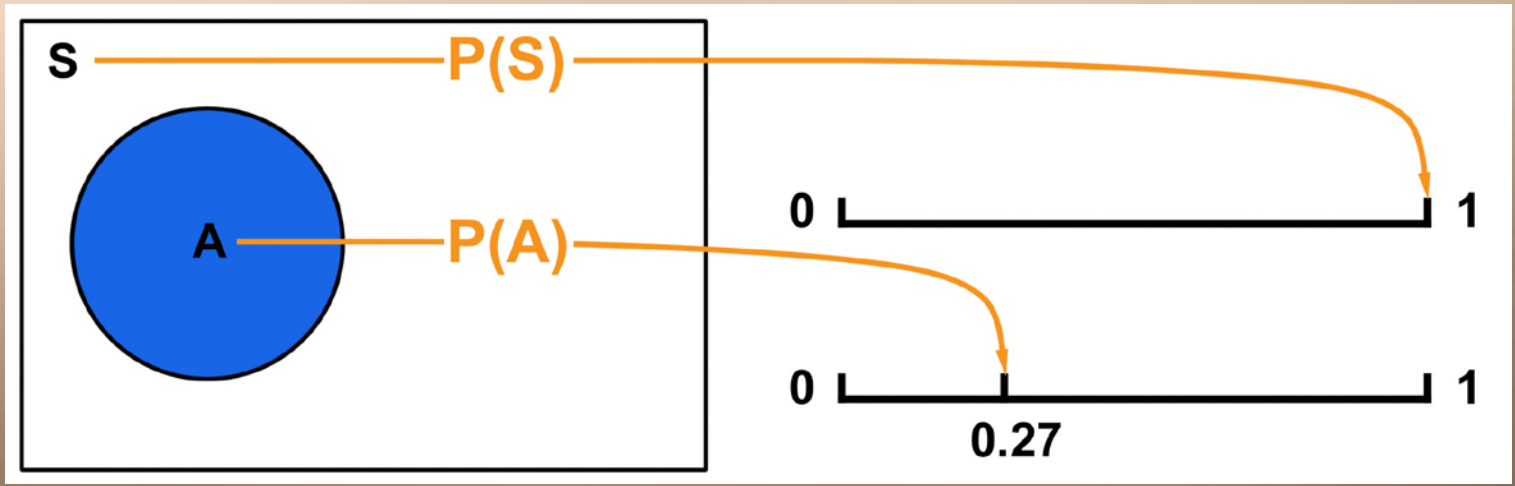
$$P(S) = 1$$

- Probability of the union of two mutually exclusive events is equal to the sum of their probabilities

$$P(A\ U\ B) = P(A) + P(B)$$

Probability Theory basics

- Probability theory introduces the concept of size or area to the event space. How?
- P(event) can be thought of as a function that maps events in event space to the interval [0,1]

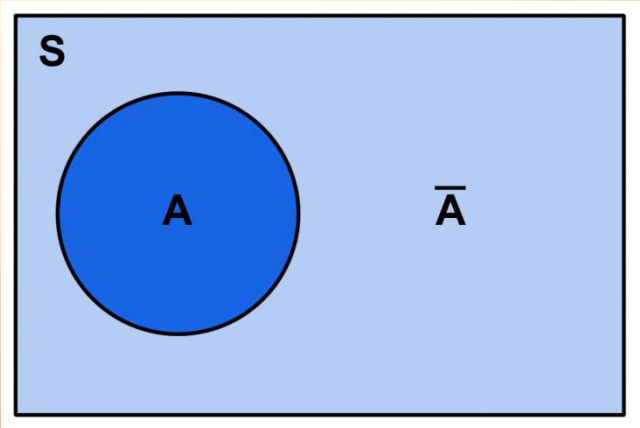


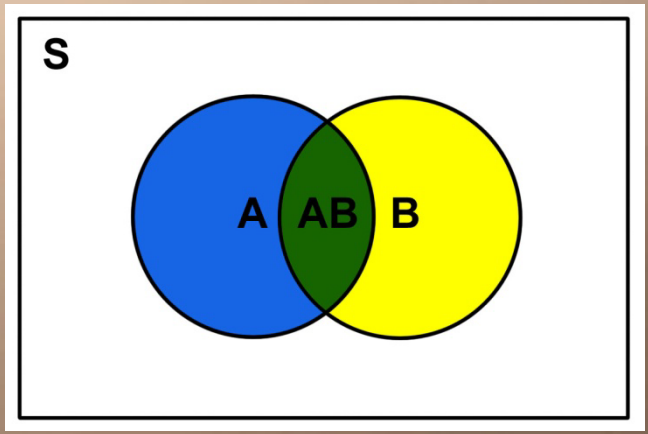
Probability Theory basics

Formulas implicitly/explicitly used in risk analysis:

- $P(\bar{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(AB)$
- $P(A \cap B) = P(A \mid B) * P(B)$

  "Intersection Formula" or "Multiplication Rule"





Note: (A | B) is *not* an event

Probability Theory basics

- Mutually exclusive events
  - Both events cannot occur at the same time
  - Floods and earthquakes usually treated as mutually exclusive events when developing PFMs (e.g. rarely consider a flood pool for EQ PFM)
  - For mutually exclusive events

$$P(A\ U\ B) = P(A) + P(B)$$

- Statistically independent events
  - Occurrence of an event(s) does not affect the probability for the other event(s)
  - Ideally, PFMs are developed in a manner that results in a statistically independent set of PFMs
  - For statistically independent events

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

Probability Theory basics

- Statistically dependent (general case)
  - Occurrence or nonoccurrence of an event(s) affects the occurrence probability for the other event(s)
  - E.g. overtopping breach and subsequent filling of a leveed area might affect the probability for initiation of internal erosion along the levee
  - For statistically dependent events

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

    - Again, this is referred to as the 'multiplication rule'

Conditional Probability

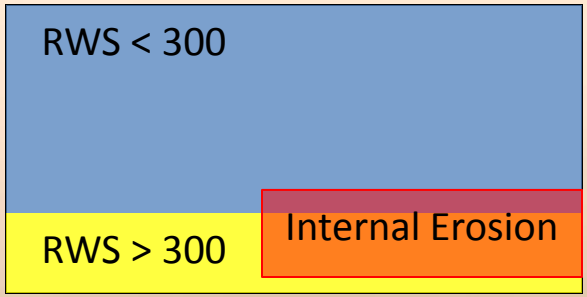
- Probability of an event, given that another event has already occurred

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

- Note that for SI events A and B, P(B|A) = P(B)
- Conditional probabilities are often directly estimated (rather than calculated) for use in event trees
- The conditioning operation (|) transforms the sample space

Conditional Probability

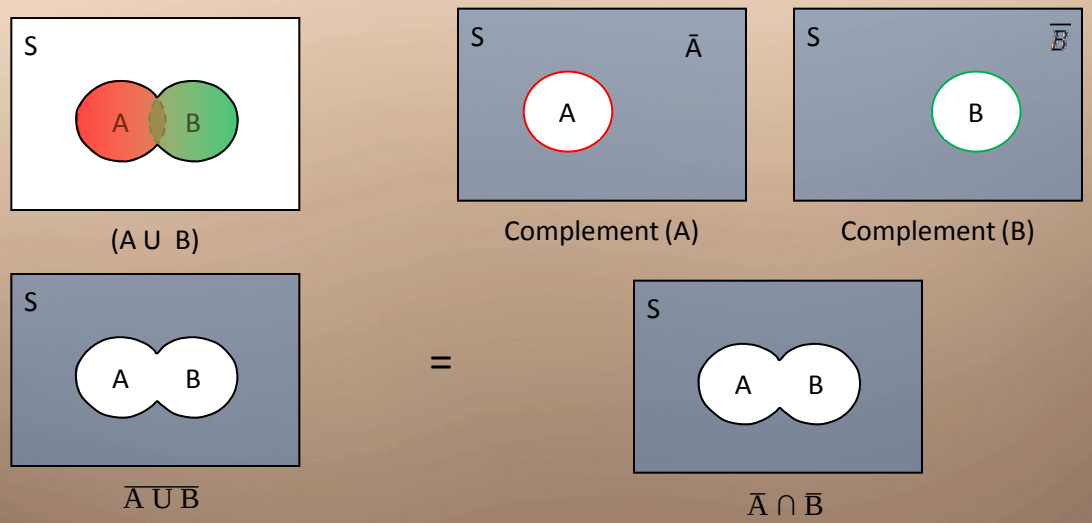
- Probability that internal erosion initiates next year



- Probability that internal erosion initiates next year given that the maximum reservoir level exceeds 300’

de Morgan's Rule

- In set theory,
  - The complement of the union of two events is equal to the intersection of the complements
  - Rule can be generalized to *n* events



de Morgan's Rule

- In practice, this can simplify calculation of the probability of the union of multiple events
  - Total probability of system failure given 'n' potential failure modes is equal to one minus the probability of no failure

$$\mathrm{P(Fail)} = 1 - \mathrm{P(No\ Fail)}$$ (in general)

$$= 1 - \prod_{i=1}^{n} [1 - \mathrm{P(Fail_i)}]$$ (for *n* SI PFMs)

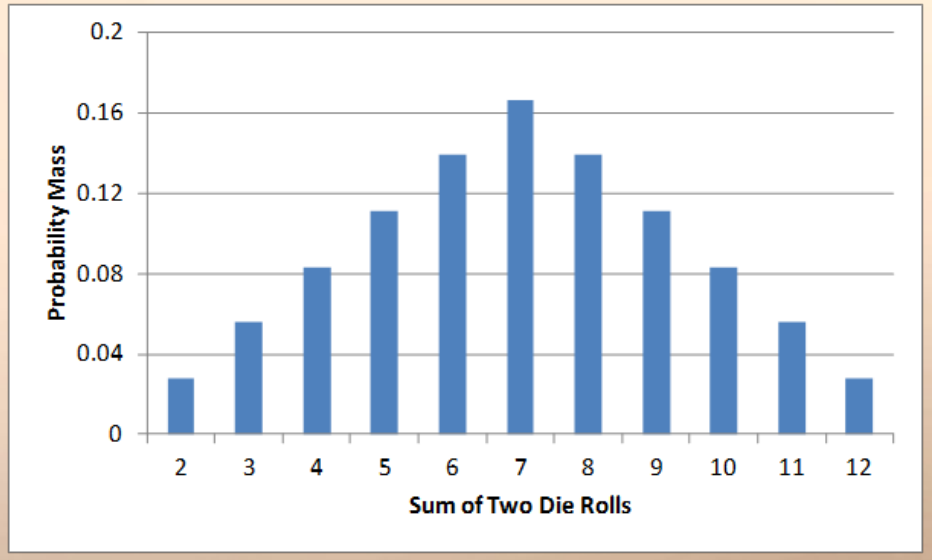
Random Variables

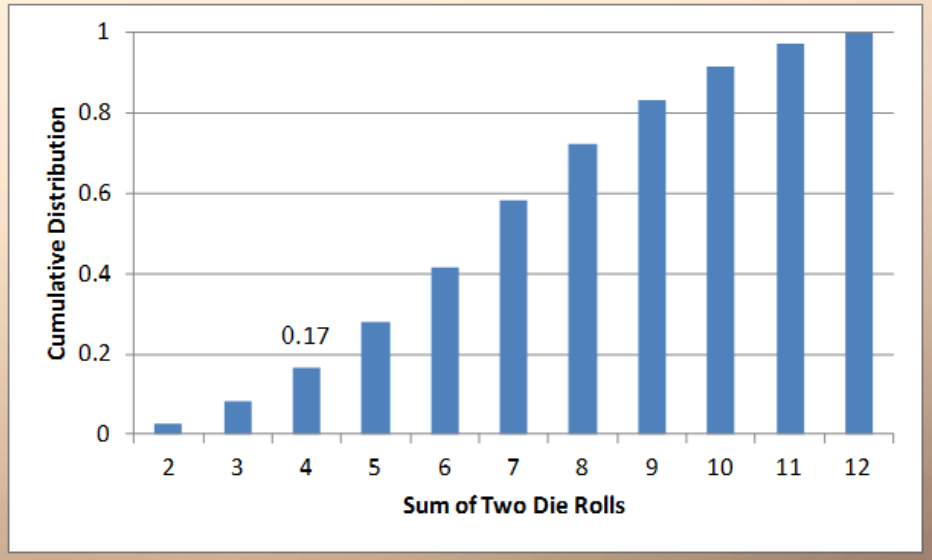
- A random variable is used to represent a quantity whose value cannot be known with certainty
  - Maximum annual peak ground acceleration
  - Number of fatalities resulting from a levee failure
  - Virtually all numeric variables considered in a risk analysis, including the probability estimates themselves
- Underlying process does not have to be random
  - Presence of a fault in the foundation (state of nature)
- Each value (or range of values) of a random variable is associated with a probability

Discrete Random Variable

- Finite number of possible values
  - Number of monoliths that slide in and EQ
  - Number of spillway gates that don't operate in a flood
  - Number of rockfall impacts at a given site and in a given year, with a total volume of 1000 cu. yd. or greater
- Described by probability mass function (PMF) or cumulative distribution function (CDF)

Discrete Random Variable





Continuous Random Variable

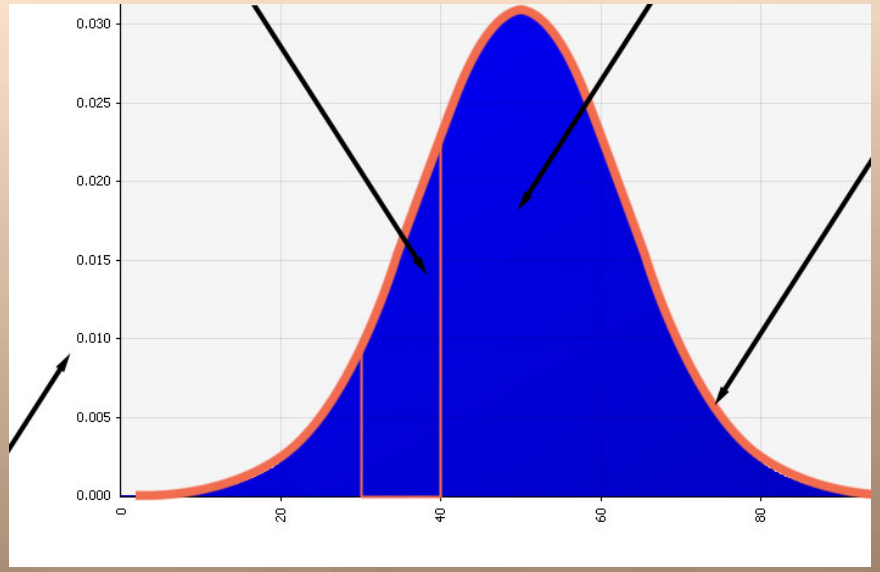
- Infinite number of possible values
  - Peak ground acceleration
  - Friction angle for a soil
- Described by probability density function (PDF) or cumulative distribution function (CDF)
  - Probability of a specific value is equal to zero
  - Partitions, ranges, or intervals used to describe values and their associated probabilities

Key characteristics of a PDF

- Consider a distribution describing the potential deformation of a 300-foot high embankment dam resulting from liquefaction of the hydraulic fill core

The area of the slice between 30 and 40 represents the Probability that 30 to 40 feet of deformation occurs

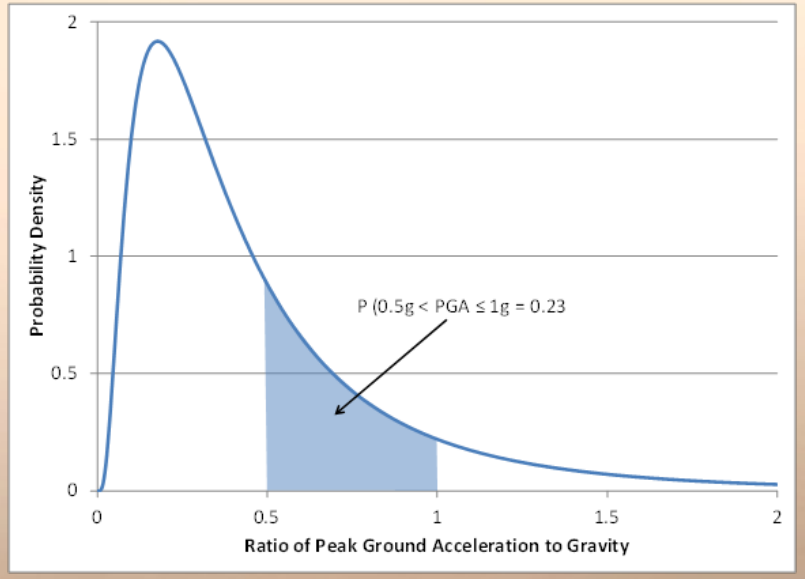
The "interior" portion of the distribution; area of the interior equals 1.0

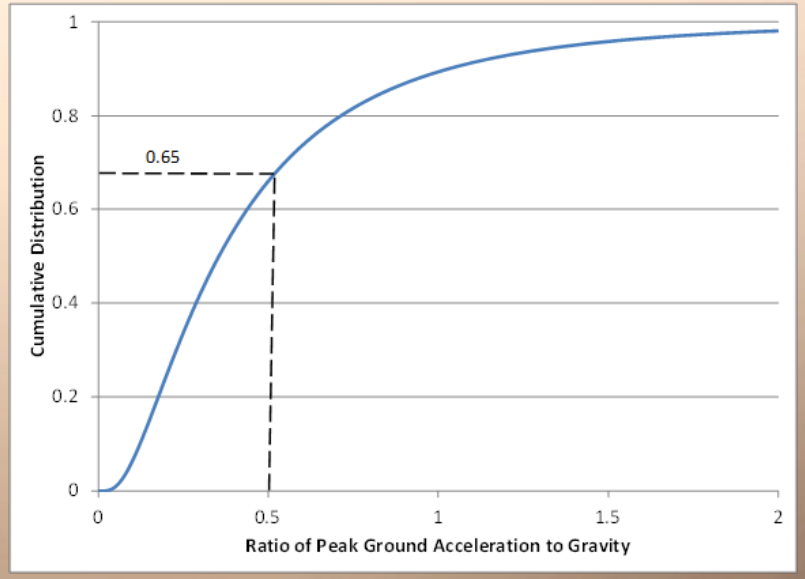


"Exterior" portion of the distribution; this curve is known as the probability density function or *PDF*

The numbers along this axis are a density (not a probability), to ensure that the area under the curve is equal to 1.0

PDF versus CDF

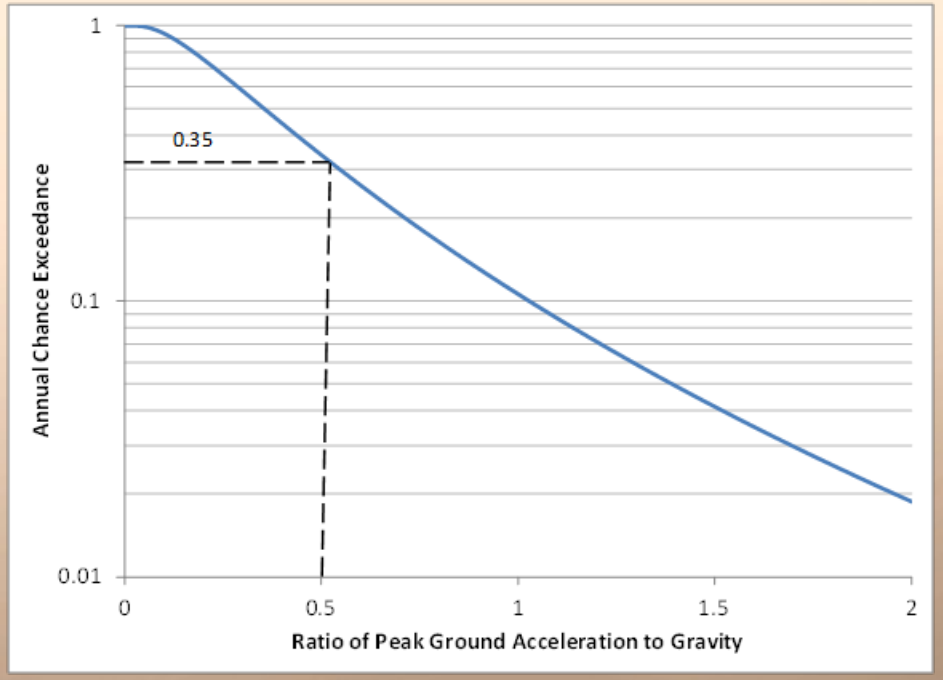




$$P(a < X \leq b) = \int_a^b f_X(x)dx$$

$$P(-\infty < X \leq b) = \int_{-\infty}^b f_X(x)dx$$

Complementary Cumulative Distribution (Exceedance Curve)

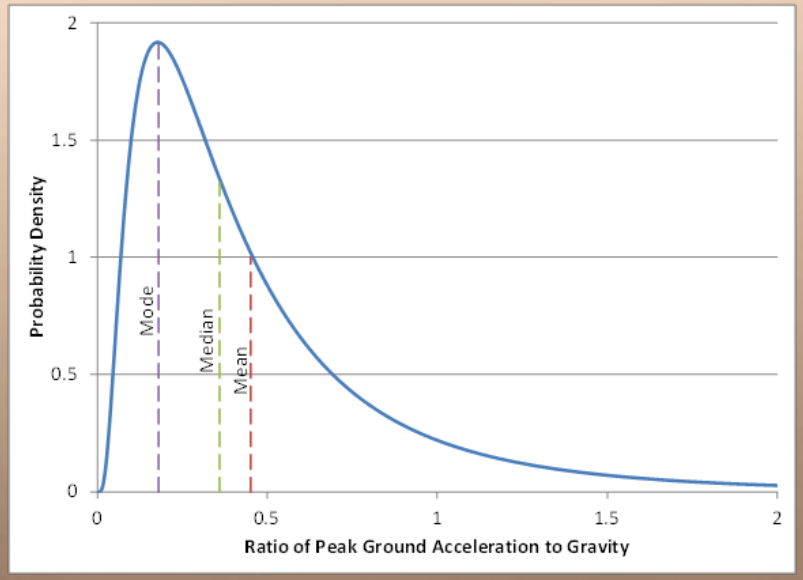


$$P(a < X \leq \infty) = \int_{a}^{\infty} f_X(x)dx$$

$$P(a < X \leq \infty) = 1 - P(-\infty < X \leq a)$$

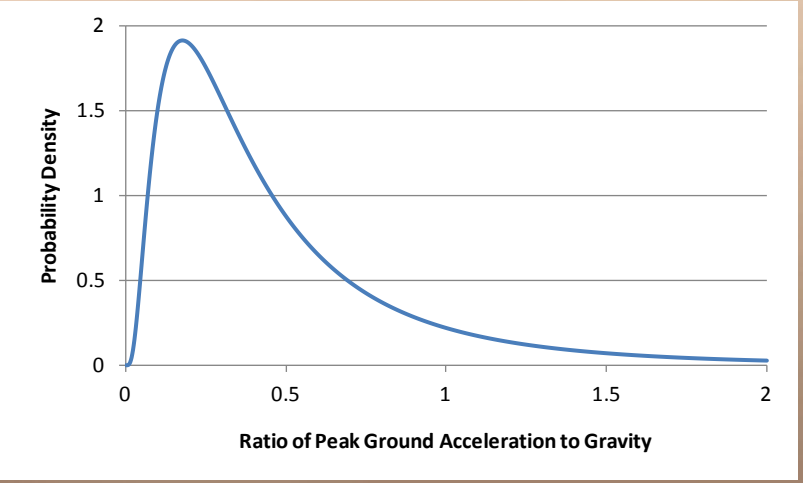
Point Estimators

- Mean – Expected value (centroid)
- Median – 50th percentile (equal area)
- Mode – Most frequent (maximum probability density)

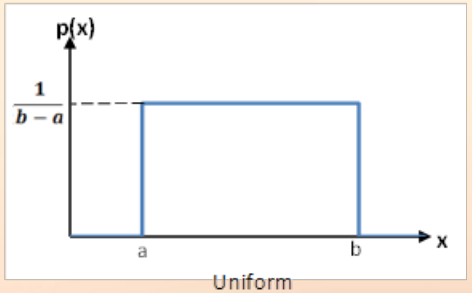


Point Estimators

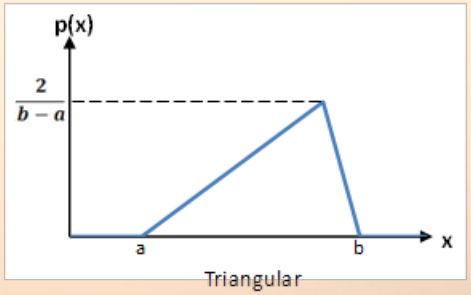
- Mean – First moment, expected value
- Variance – Second moment, central tendency
- Skew – Third moment, symmetry



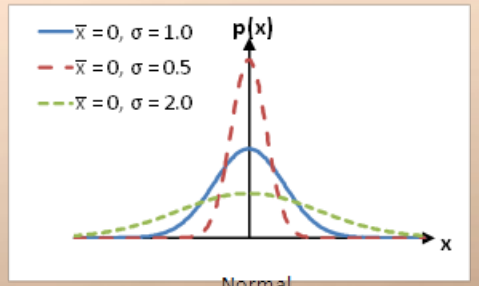
Common Probability Distributions



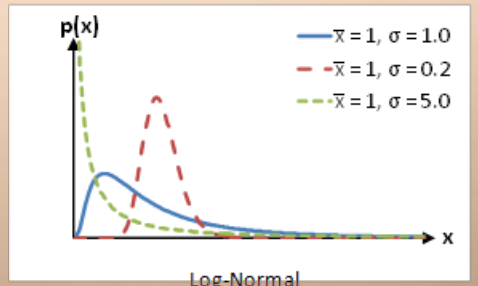
Uniform



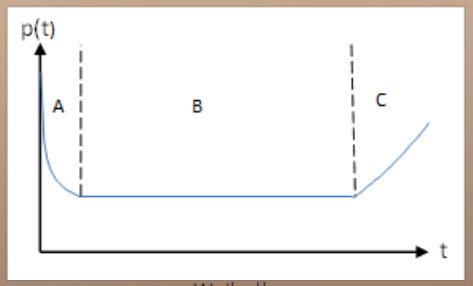
Triangular



Normal



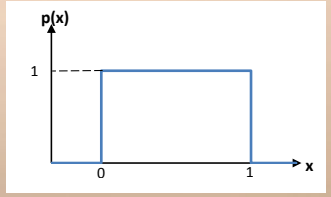
Log-Normal



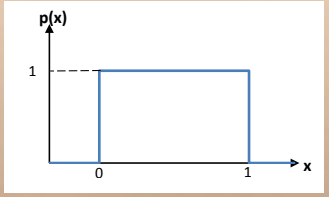
Weibull

Central Limit Theorem

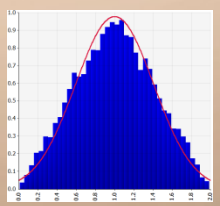
- Sum of distributions trends toward a normal distribution
- Product of distributions trends toward a log normal distribution

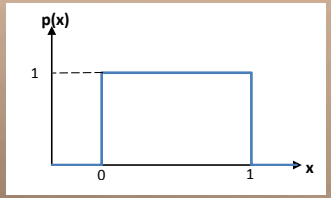


+

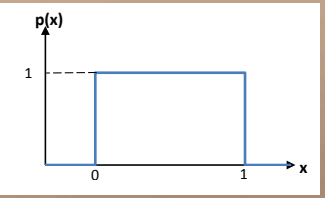


=

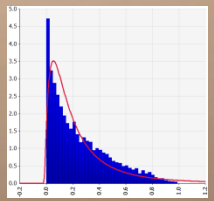




x



=



Bayes Theorem

- Provides a way of updating a probability estimate to account for new evidence
- Can be used to inform subjective probability estimates so that new evidence is not given too much or too little weight
- Care must be taken to ensure that the original estimate does not already consider the “new” information (e.g., when updating an internal erosion estimate for newly available low plasticity data even though the original estimate already considers a range of PIs)

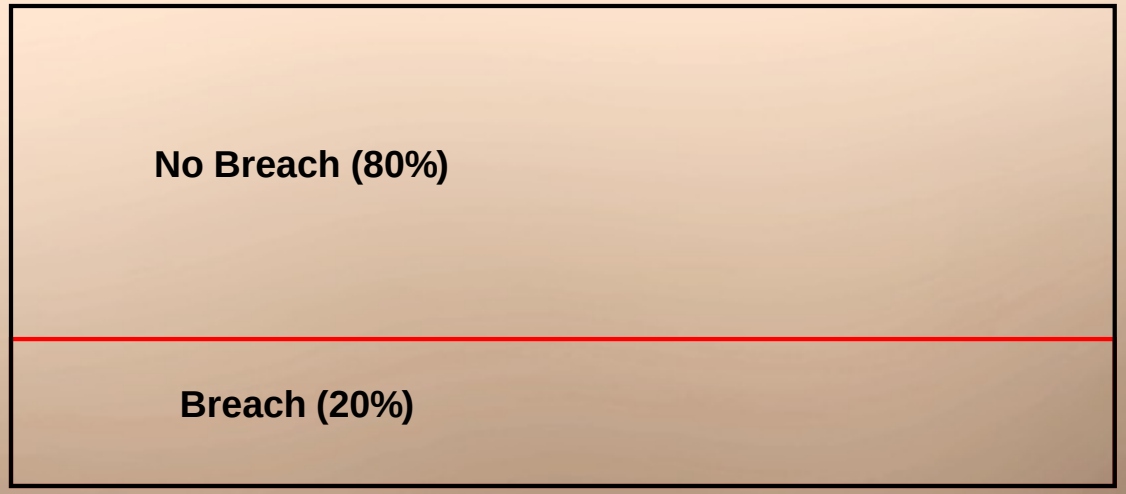
Bayes Theorem

- P(x) : Prior probability for an event absent evidence
- P(O|x) : Conditional probability of an observation given the event
- P(O) : Probability of the observation
- P(x|O) : Posterior probability of the event given the observation

$$P(x|O) = \frac{P(x)P(O|x)}{P(O)}$$

Prior Probability

- P(breach) = 0.2



Conditional Probability of the Observation

- $$P(seepage \mid breach) = \frac{0.15}{0.15 + 0.05} = 0.75$$

No Breach and No Observed Seepage (55%)	**No Breach and Observed Seepage (25%)**
Breach and No Observed Seepage (5%)	**Breach and Observed Seepage (15%)**

Probability of the Observation

- P(seepage) = 0.15 + 0.25 = 0.4

No Breach and No Observed Seepage (55%)	**No Breach and Observed Seepage (25%)**
Breach and No Observed Seepage (5%)	**Breach and Observed Seepage (15%)**

Posterior Probability

- Initial estimate for probability of breach is 0.2
- If seepage has been observed, the probability of breach should logically increase
- An updated estimate for probability of breach can be estimated using Bayes theorem

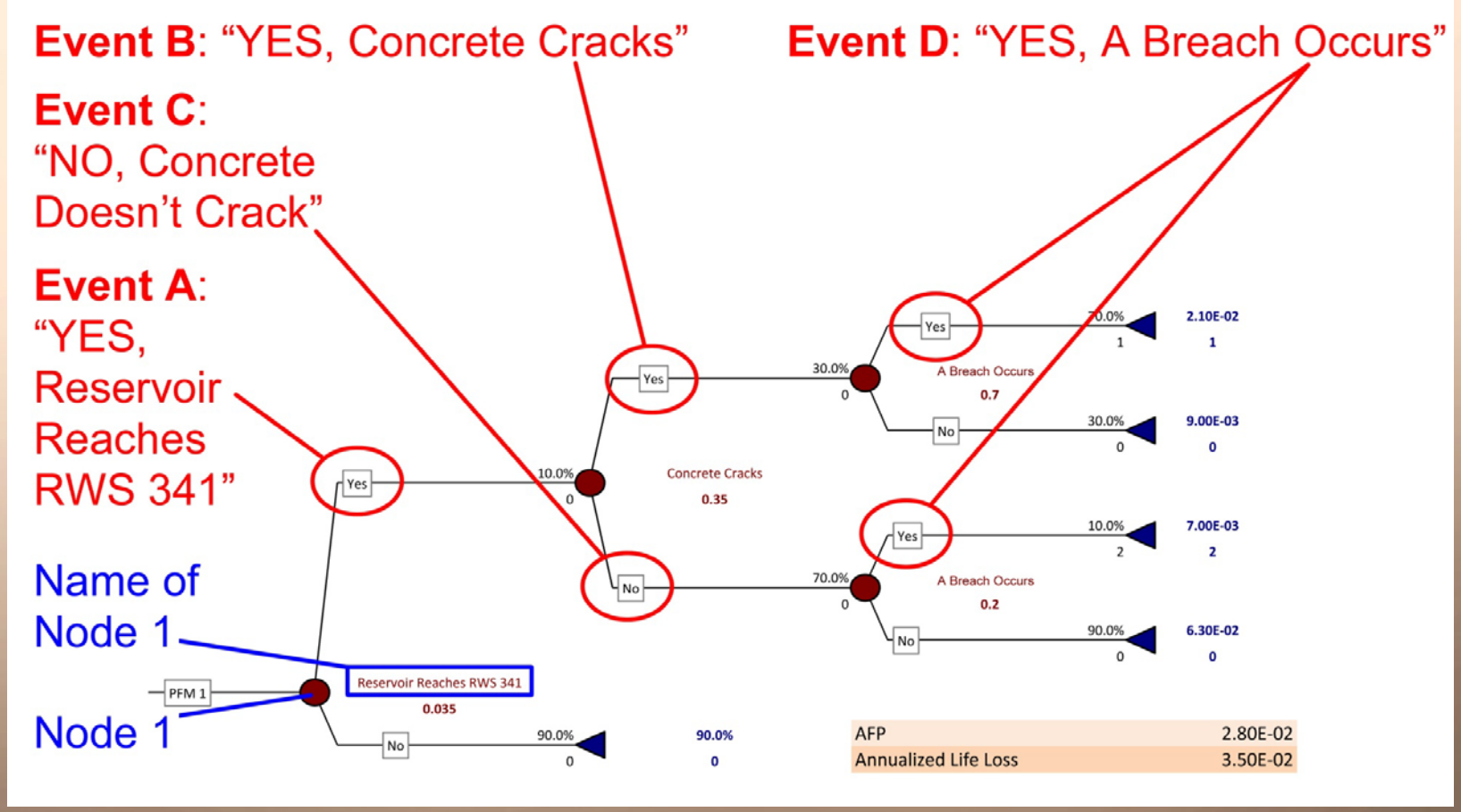
$$P(breach \mid seepage) = \frac{0.2 * 0.75}{0.4} = 0.375$$

Uni-Modal Bounds

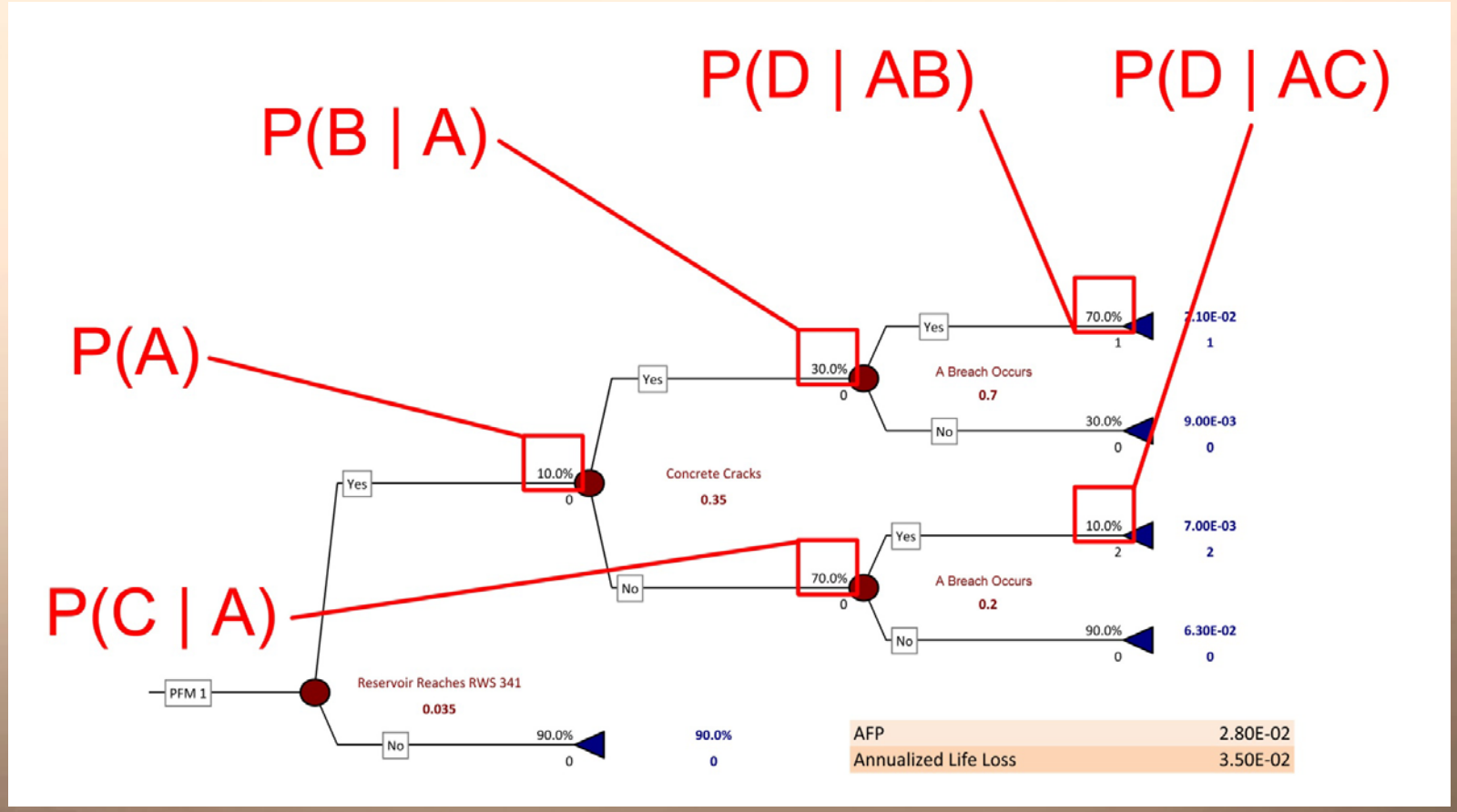
- Probability for the union of *n* positively correlated events lies between an upper and lower bound
  - Lower bound, perfectly correlated
  - Upper bound, perfectly uncorrelated (typically assumed in practice)

$$\max_{i} P(E_i) \leq P(E) \leq 1 - \prod_{i=1}^{n} [1 - P(E_i)]$$

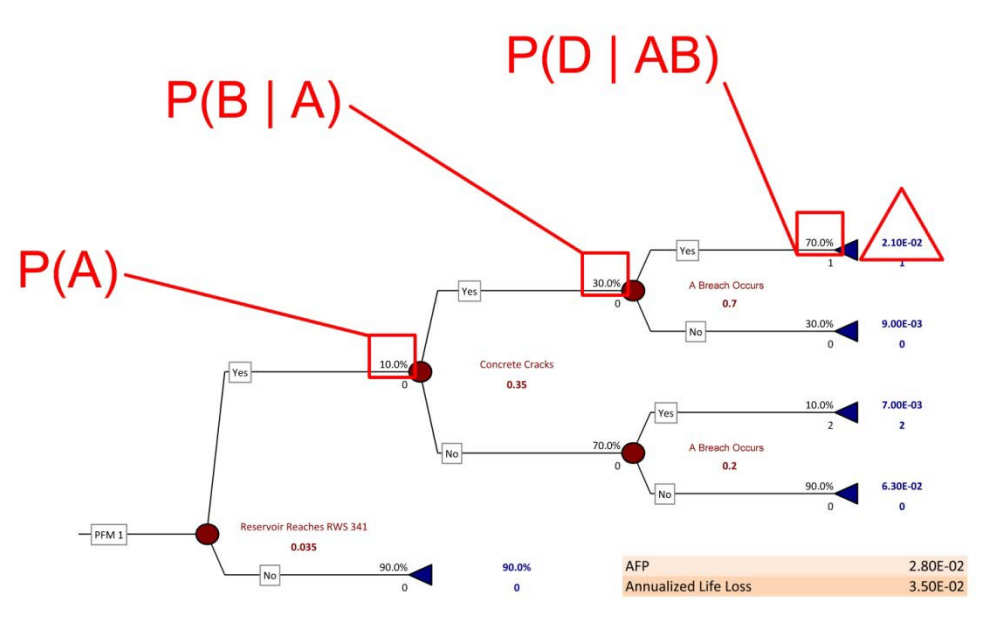
Probability, statistics, and event trees



Probability, statistics, and event trees



Probability, statistics, and event trees

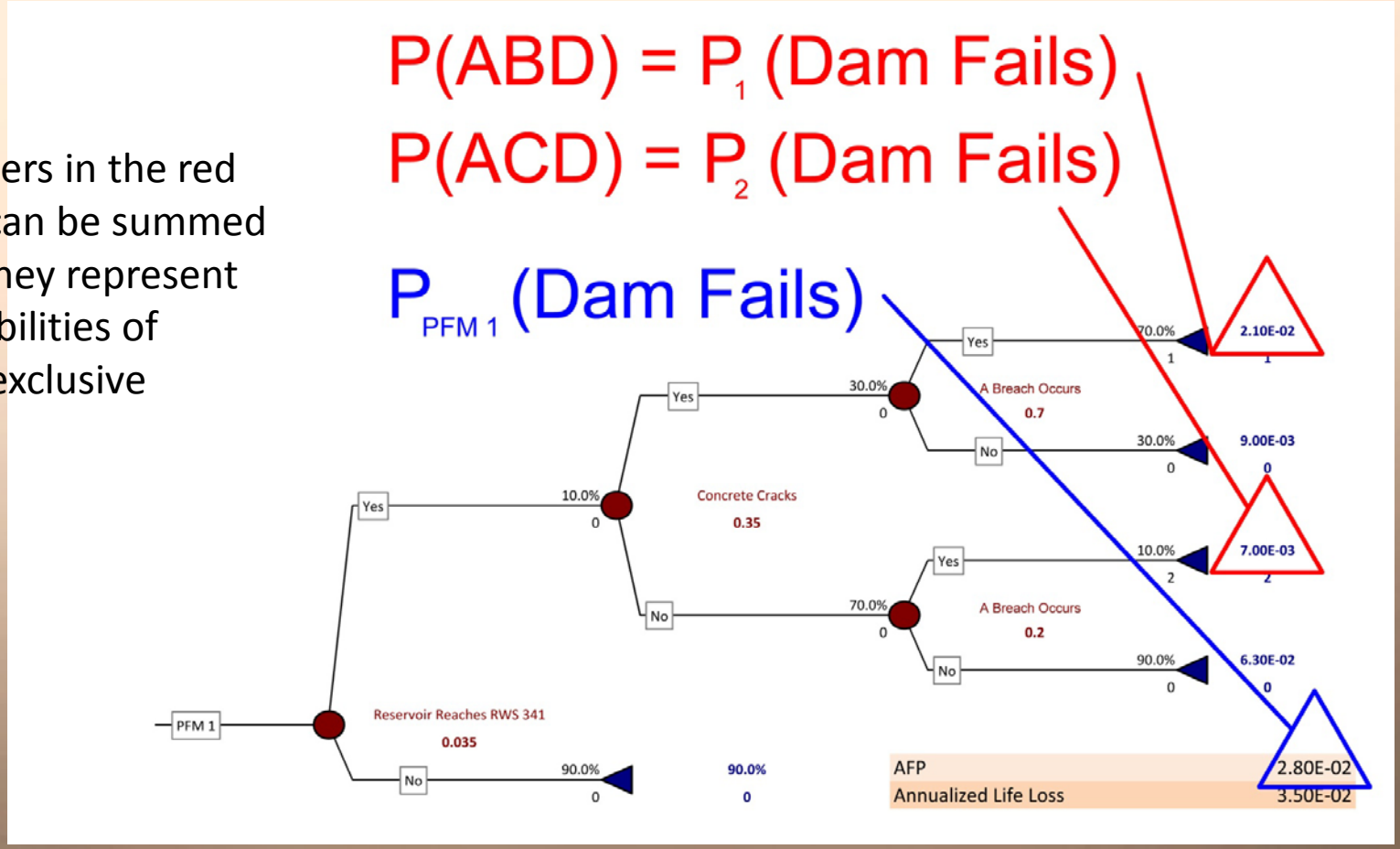
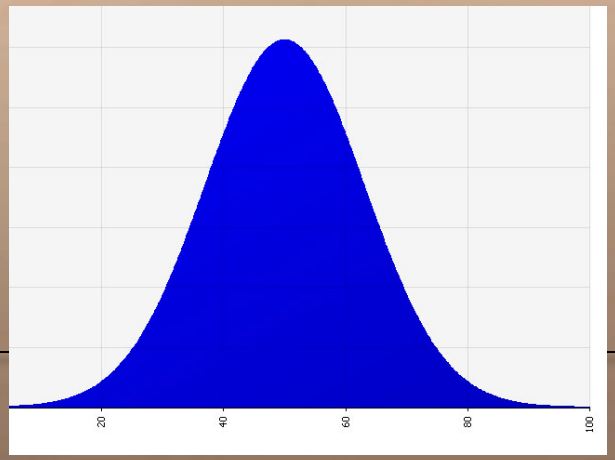


- Probability estimates are multiplied along the top branch to obtain the number shown in the red triangle

- By the intersection formula, the product P(A) * P(B | A) * P(D | AB) is equal to P(ABD)

Probability, statistics, and event trees

The numbers in the red triangles can be summed because they represent the probabilities of mutually exclusive events



Questions