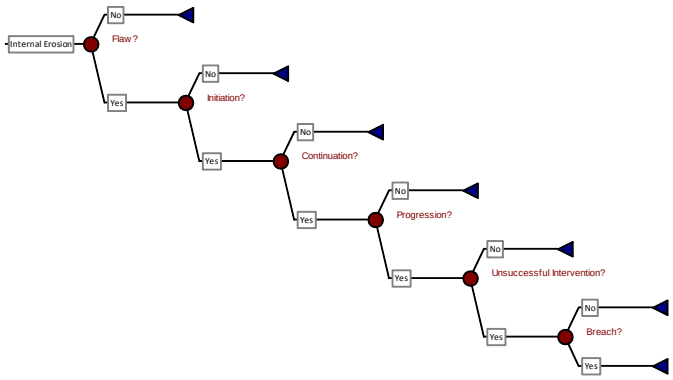
Best Practices in Dam and Levee Safety Risk Analysis

I-5. Event Trees

June 23, 2015



RECLAMATION
Managing Water in the West

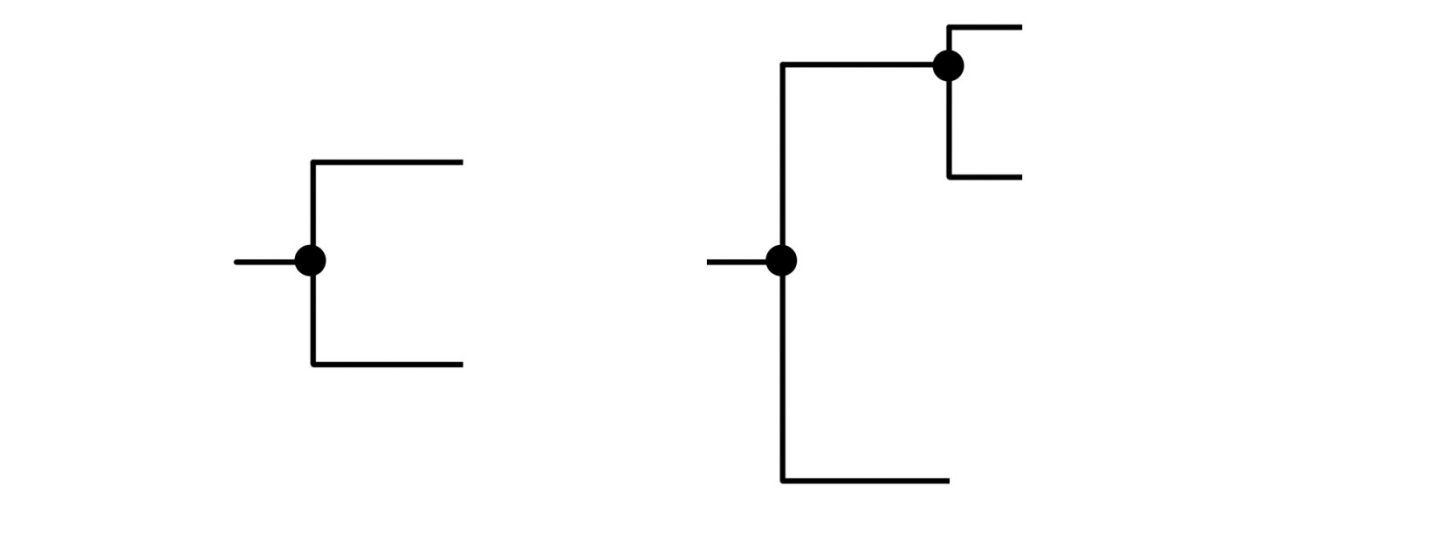
US Army Corps of Engineers
BUILDING STRONG®

Outline

- Introduction and basic concepts
  - What is an event tree?
  - What do I need to construct an event tree?
  - What is the theoretical basis for an event tree?
- Special topics
  - Partitioning the range of loadings
  - Background calculations
  - Annualized Life Loss and event trees
  - Common Cause Adjustment
  - Monte Carlo simulation

What is an event tree?

- A graph object consisting of any number of branches and vertices (or “nodes”)



Are event trees helpful in a risk analysis?

- Event trees can help improve understanding of the causal structure of the potential failure mode (PFM) being considered
- The benefits of using an event tree include:
  - Improved understanding of the PFMs by team members (i.e. those who are estimating risks)
  - Improved understanding of the PFMs by report authors (i.e. those who are interpreting the risks)
  - Improved understanding of the PFMs by by decision makers (i.e. those who must act based on reported risks)

Are event trees required in a risk analysis?

- No! While convenient, the use of an event tree is by no means required
  - The phrase “event tree approach” is sometimes used interchangeably with the phrase “risk analysis”.
  - However, the rules of probability that govern the risk analysis process hold true regardless of whether an event tree is constructed
  - There is **nothing** that an event tree can do that cannot be done using a spreadsheet alone

What do I need to construct an event tree?

- A potential failure mode description
- Probability estimates for the component events of the PFM
- A software program for building event trees
  - BOR uses Precision Tree, a component of the Palisade Decision Suite that functions as an add-on to Microsoft Excel
  - Additional software may be necessary to perform a Monte Carlo simulation. BOR uses the @risk component of the Palisade Decision Suite

Is there a theoretical basis for event tree calculations?

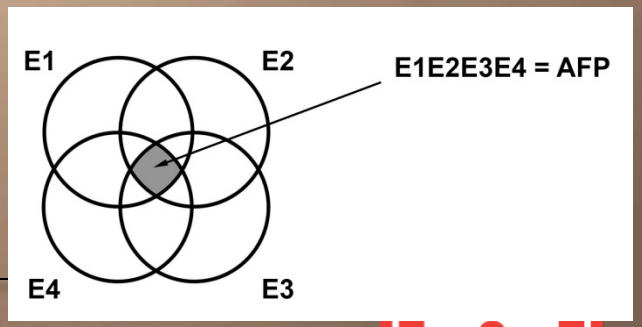
- The internal logic of an event tree is almost entirely based on:
  - The intersection formula (multiplication rule)
    - $P(AB) = P(A) * P(B|A)$ *note: AB is short for $A \cap B$*
    - $P(ABC) = P(A) * P(B|A) * P(C|AB)$
  - The total probability theorem
    - Given ME and CE events A, B, and C, $P(E) = P(E|A)P(A) + P(E|B)P(B) + P(E|C)P(C)$
    - The conditioning events could be e.g. load ranges

Example 1

- Consider the following PFM description:
  - In the year under consideration, an earthquake with a peak horizontal acceleration in excess of 0.6g occurs at the site, triggering foundation liquefaction. Embankment slope instability occurs upstream, resulting in crest loss in excess of the available freeboard and a breach of the reservoir
- The PFM can be discretized as follows:
  - Event E1: 0.6g PHA is exceeded at the site in a given year
  - Event E2: Extensive foundation liquefaction is triggered
  - Event E3: Upstream slope instability occurs
  - Event E4: A breach occurs

Example 1

- For the component events of the PFM, the following probabilities have been estimated:
  - $P(E1) = 0.001$
  - $P(E2|E1) = 0.5$
  - $P(E3|E1 \cap E2) = 0.1$
  - $P(E4|E1 \cap E2 \cap E3) = 0.2$
- What is the Annual Failure Probability (AFP)?
- Hint: The AFP is equal to the intersection probability of the events that comprise the PFM

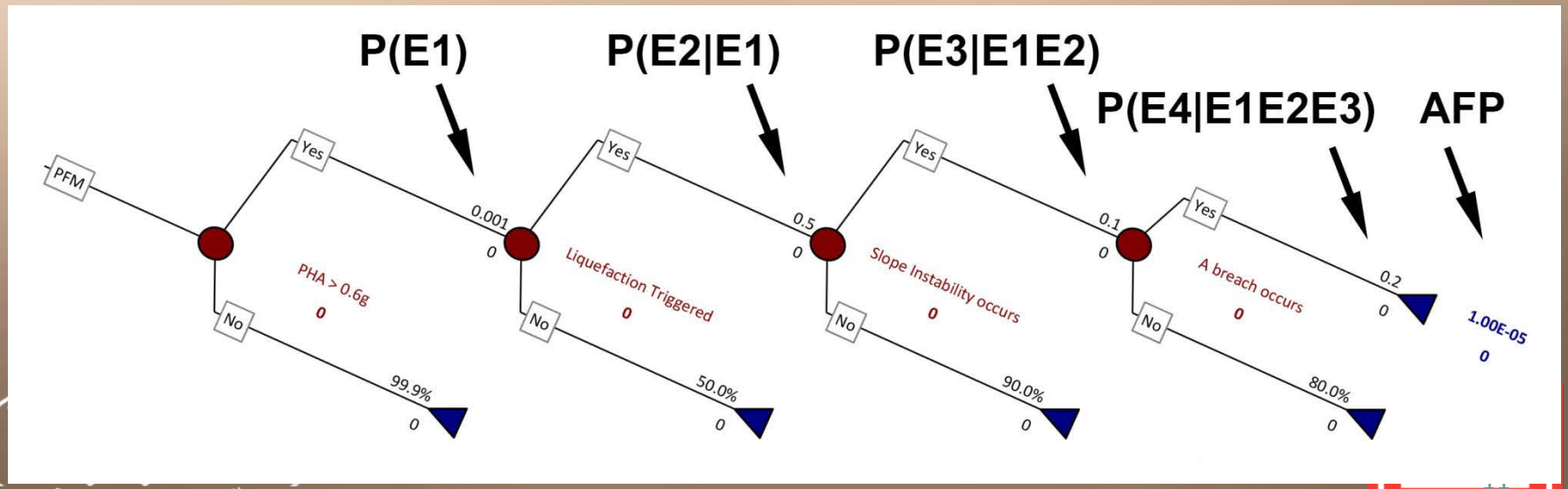


“Spreadsheet” solution

- Use the intersection probability formula
- For four events A, B, C, D, the formula is:
  - $P(ABCD) = P(A) * P(B|A) * P(C|AB) * P(D|ABC)$
- In our case,
  - $P(E1 \cap E2 \cap E3 \cap E4)$
    $= P(E1) * P(E2|E1) * P(E3|E1 \cap E2) * P(E4|E1 \cap E2 \cap E3)$
    $= 0.001 * 0.5 * 0.1 * 0.2$
    $= 0.00001$
- The AFP is equal to 0.00001

“Event tree ” solution

- Event would first be constructed using the event tree software
- Once probability estimates are placed in the correct slots, APF is automatically calculated



BUREAU OF RECLAMATION

Example 2

- Consider the following slightly modified PFM description:
  - In the year under consideration, an earthquake occurs at the site, triggering foundation liquefaction. Embankment slope instability occurs upstream, resulting in crest loss in excess of the available freeboard and a breach of the reservoir
- Because failure is no longer premised on the exceedance of a particular threshold, several different load ranges must now be considered

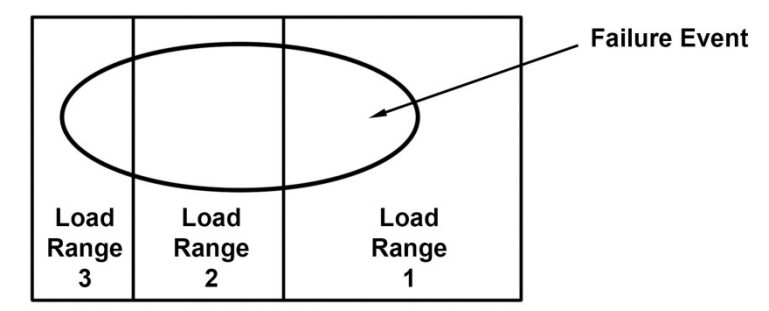
Example 2

- For the component events of the PFM, the following probabilities have been estimated:
  - $P(PHA < 0.3g) = 0.9$; $P(0.3g < PHA < 0.6g) = 0.099$; $P(PHA > 0.6g) = 0.001$
  - $P(E2|PHA < 0.3g) = 0$; $P(E2|0.3g < PHA < 0.6g) = 0.05$; $P(E2|PHA > 0.6g) = 0.5$
  - $P(E3|0.3g < PHA < 0.6g \cap E2) = 0.1$; $P(E3|\ PHA > 0.6g \cap E2) = 0.1$
  - $P(E4|0.3g < PHA < 0.6g \cap E2 \cap E3) = 0.2$; $P(E4|PHA > 0.6g \cap E2 \cap E3) = 0.2$
- What is the Annual Failure Probability (AFP)?

“Spreadsheet” solution

- Step 1. Use intersection probability formula
- For the PHA < 0.3g load branch, no failure
- For the 0.3g < PHA < 0.6g load branch, the “sub-AFP” is:
  - P(Load ∩ E2 ∩ E3 ∩ E4)
    = 0.099 * 0.05 * 0.1 * 0.2 = 0.00001
- For the PHA > 0.6g load branch:
  - P(Load ∩ E2 ∩ E3 ∩ E4)
    = 0.001 * 0.5 * 0.1 * 0.2 = 0.00001

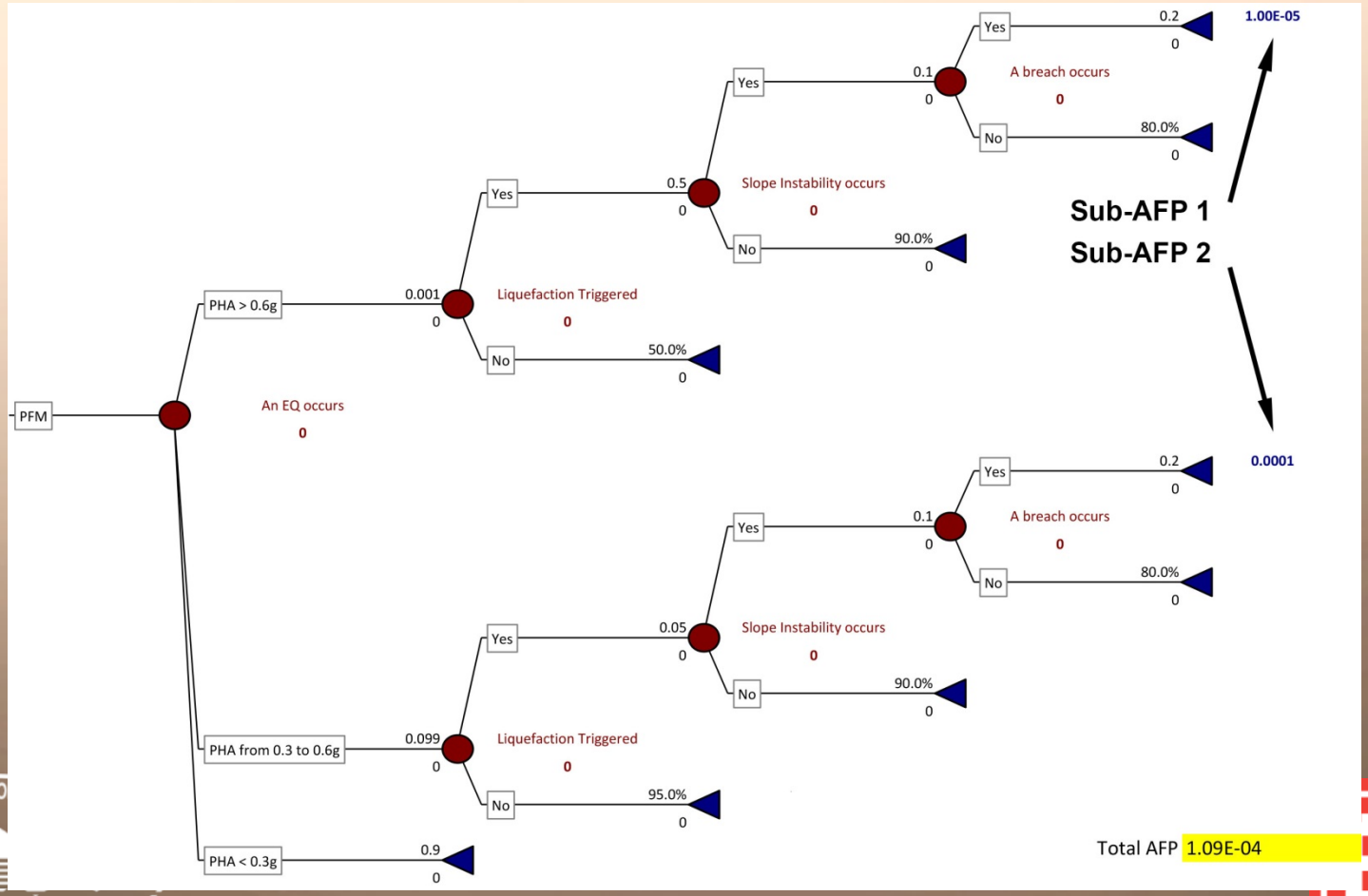
“Spreadsheet” solution



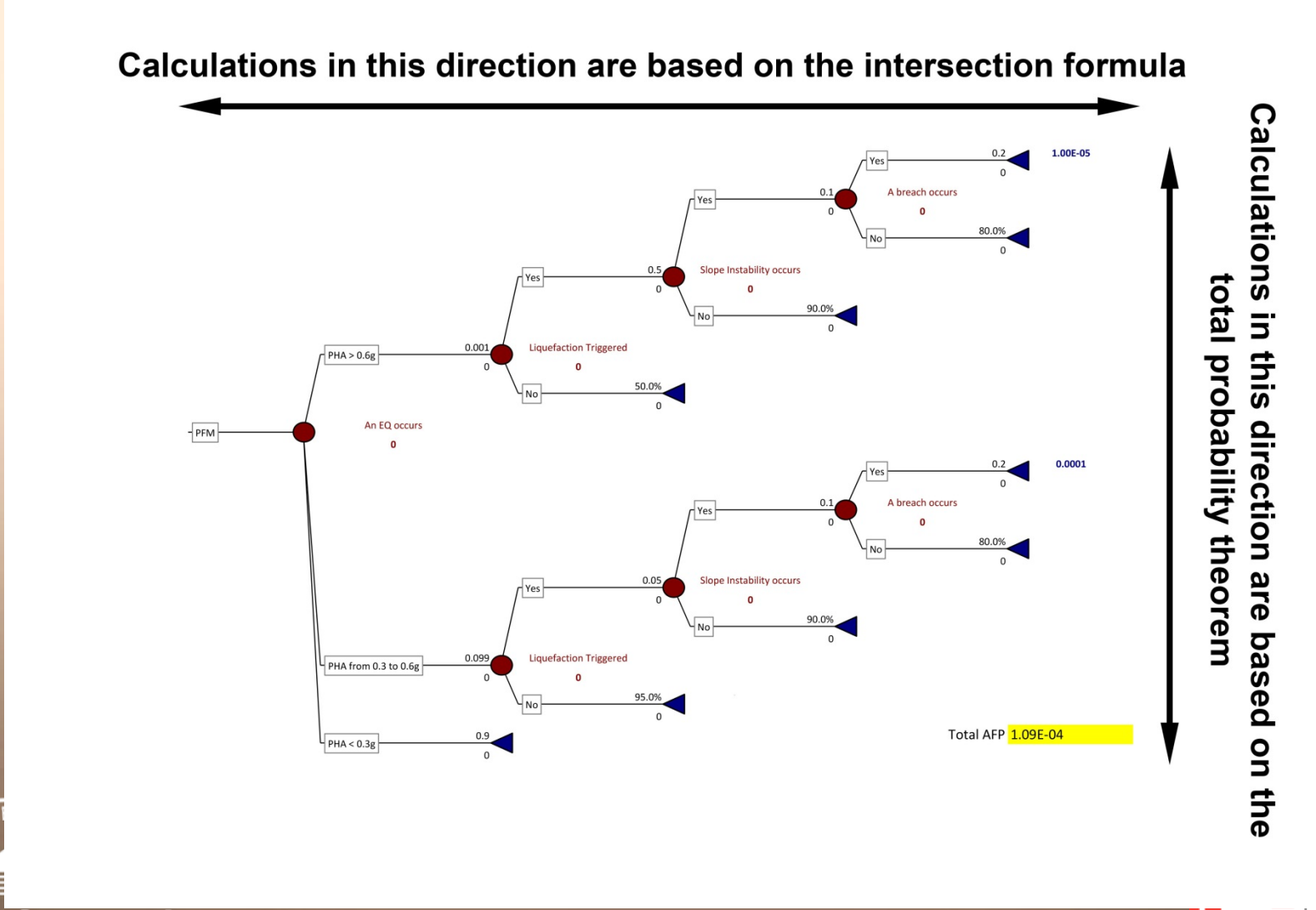
- Step 2. Use the total probability theorem
- AFP = {P(Fail|PHA < 0.3g)}*P(PHA < 0.3g)+ {P(Fail|0.3g < PHA < 0.6g)}*P(.3g < PHA < 0.6g) + {P(Fail|PHA>0.6g)} *P(PHA>0.6g) = 0 + 0.000099 + 0.00001 = 0.000109
- Note the expressions in the curly brackets are referred to as conditional failure probabilities

“Event tree ” solution

- Less messy than the spreadsheet solution



Summary of the basic concepts

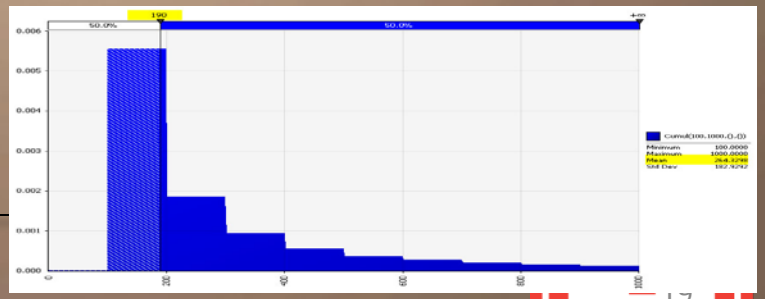


Partitioning the load range

- When the possibility of failure is tied to the exceedance of a simple threshold (such as in the case of a static PFM), it is relatively easy to assign a load probability. (→exceedance curve)
- For a seismic or hydrologic PFM, performance of the dam could vary depending on the magnitude of the loading. For example, there may be a higher conditional probability of failure at a 0.6g PHA than a 0.3g PHA.

Partitioning the load range

- To help develop estimates that are sensitive to the magnitude of the loading, the overall load range is split into bins, as in the case of the previous example
  - Load range 1: PHA < 0.3g; Load range 2: 0.3g < PHA < 0.6g Load range 3: PHA > 0.6g
- Continuous sampling of load probability is also possible, but beyond the scope of this presentation



Partitioning of the load range

- The appropriate number of load bins to use depends on the specific needs of the risk team, and on the information available
  - The use of not enough load bins could result in a critical load range being masked, or in broad uncertainty in the probability estimates
  - The use of too many load bins could result in an unmanageable number of risk estimates, and imply a degree of precision that is not really there

Partitioning of the load range

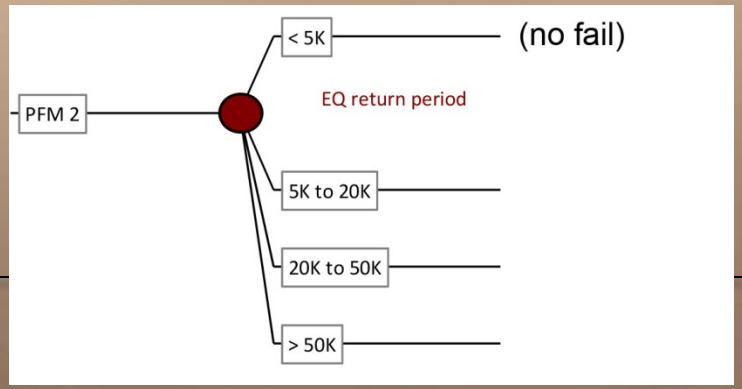
- In many cases, the type and number of available analysis results can provide the team with a sense of how many load bins to use
- For example, if finite element deformation results were available for 5K, 20K, and 50K ground motions, it would make sense to use four load bins (with the lowest not associated with a significant probability of failure)

Selecting load bin bounds

- Suppose that analysis results are available for the 5K (PHA = 0.3g), 20K (PHA = 0.6g), and 50K (PHA = 0.9g) ground motions.
- What would be the appropriate bin bounds to select given the available data?
  - Approach 1: Select bounds that correspond to the return periods associated with the results
  - Approach 2: Select bounds such that the return periods associated with the results are the "typical" return periods of the load bin

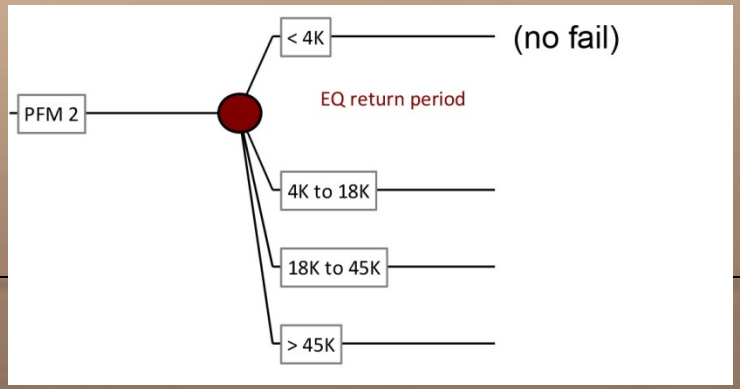
Approach 1

- Advantages
  - Simple and straightforward
- Disadvantages
  - Conditional probability estimates, in addition to taking into account other uncertainty sources, must also reflect the full range of responses spanned by the bounding sets of results



Approach 2

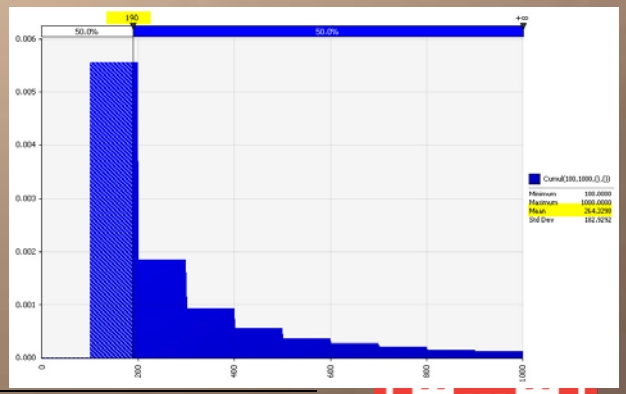
- Advantages
  - Conditional probability estimates can now be focused on a single “typical” set of analysis results, while still taking into account the uncertainty of those results
- Disadvantages
  - Requires a bit of thinking - how to select bounds?



Step 1: Define a rt. period distribution

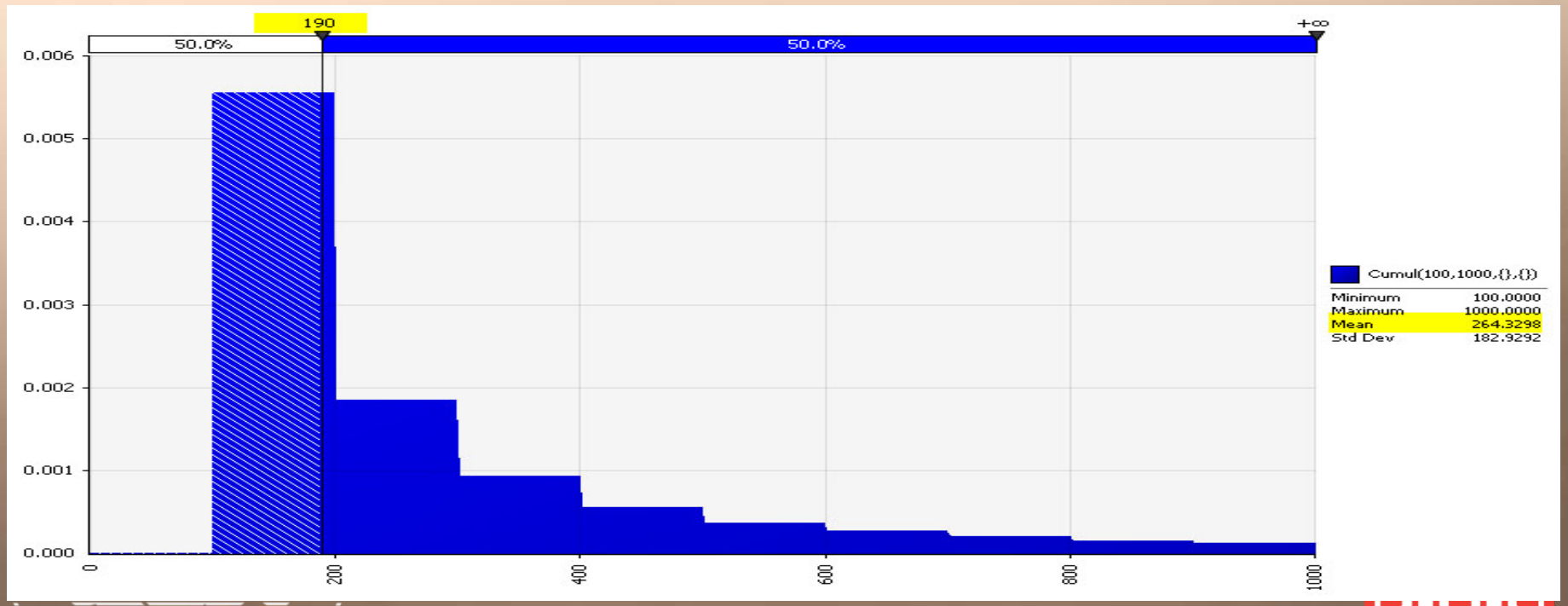
- The distribution of return periods for a given load branch (e.g. 100 to 1000-yr) can be approximated as a histogram with a finite number of bins (say, 9)
- This histogram can be generated by rewriting the intersection probability formula as P(A|B) = P(AB)/P(B), and applying it to each of the chosen bins
- For example, for the 100-200 yr bin, have P(100-200 yr eq | 100-1000 yr eq) =
  P(100-200 yr eq AND 100-1000 yr eq) ÷ P(100-1000 yr eq)
  = P(100-200 yr eq) ÷ P(100-1000 yr eq)
  = (1/100 – 1/200) ÷ (1/100 – 1/1000)
  = 0.56
- The histogram can then be programmed as an @risk distribution

Range	p
100 to 200 year	0.56
200 to 300 year	0.19
300 to 400 year	0.09
400 to 500 year	0.06
500 to 600 year	0.04
600 to 700 year	0.03
700 to 800 year	0.02
800 to 900 year	0.02
900 to 1000 year	0.01



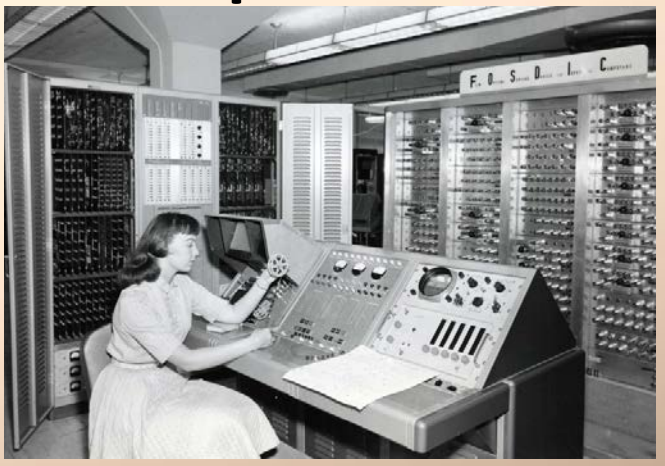
Step 2. Look for a “typical” rt. period

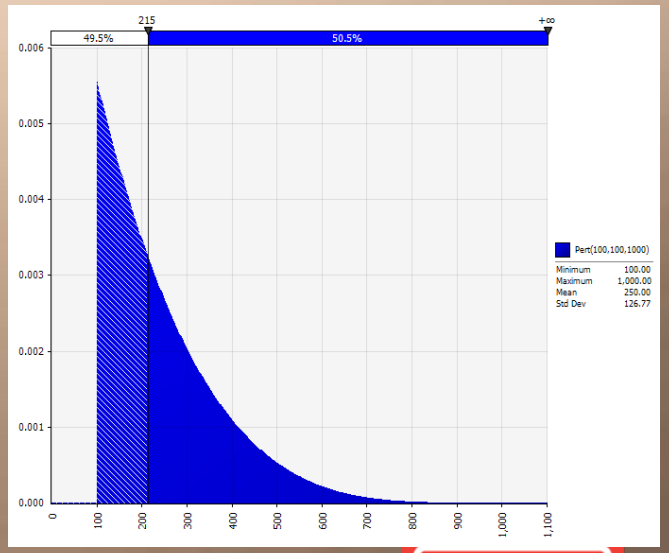
- We need to determine which point estimator to focus on for selecting the “typical” return period of each load branch
- For the 100-1000 year distribution, μ = 264 and x_{50} = 190
- Mode undefined for the histogram but presumably = 100



Step 3: Select a “typical” rt. period

- Selecting the mode would make sense if we were using Approach 1, since the mode is equal to the lower bound, and for Approach 1, the upper and lower bounds are the ones associated with the available results
- Either the mean μ or the median x_{50} EQ of the return period distribution should be used to represent the “typical EQ”
- The mean is often selected, since it is more familiar to many estimators than x_{50}
- Note that if we had simply assumed a PERT distribution, would have gotten nearly the same results. E.g. for the 100-1000 yr branch, μ = 250 and x_{50} = 217, using PERT(100,100,1000)





Background calculations

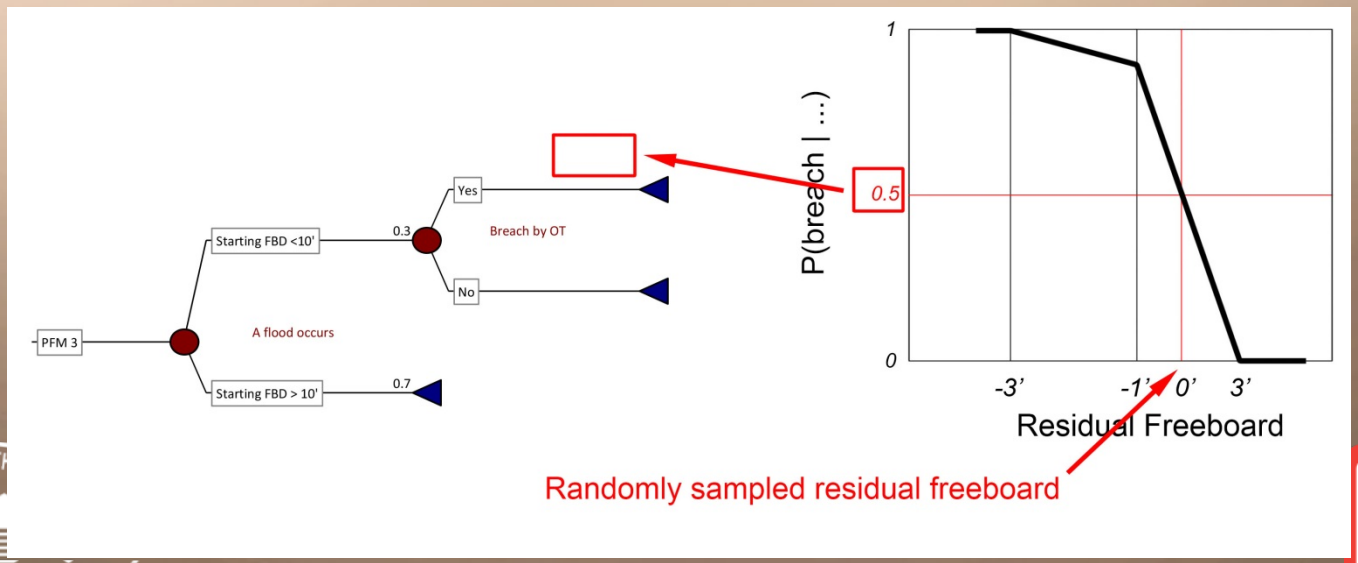
- In many cases, the conditional probability estimates are plugged directly into the event tree.
- In some cases, instead of being directly estimated, these probability estimates are calculated from other quantities that have been directly estimated

Background calculations

- In these cases, the event tree must be linked to background calculations, often located in an adjacent spreadsheet tab
- Occurs in the context of seismic and hydrologic PFMs
- In evaluating the conditional probability of breach by overtopping, a risk team often develops a system response (“fragility”) curve

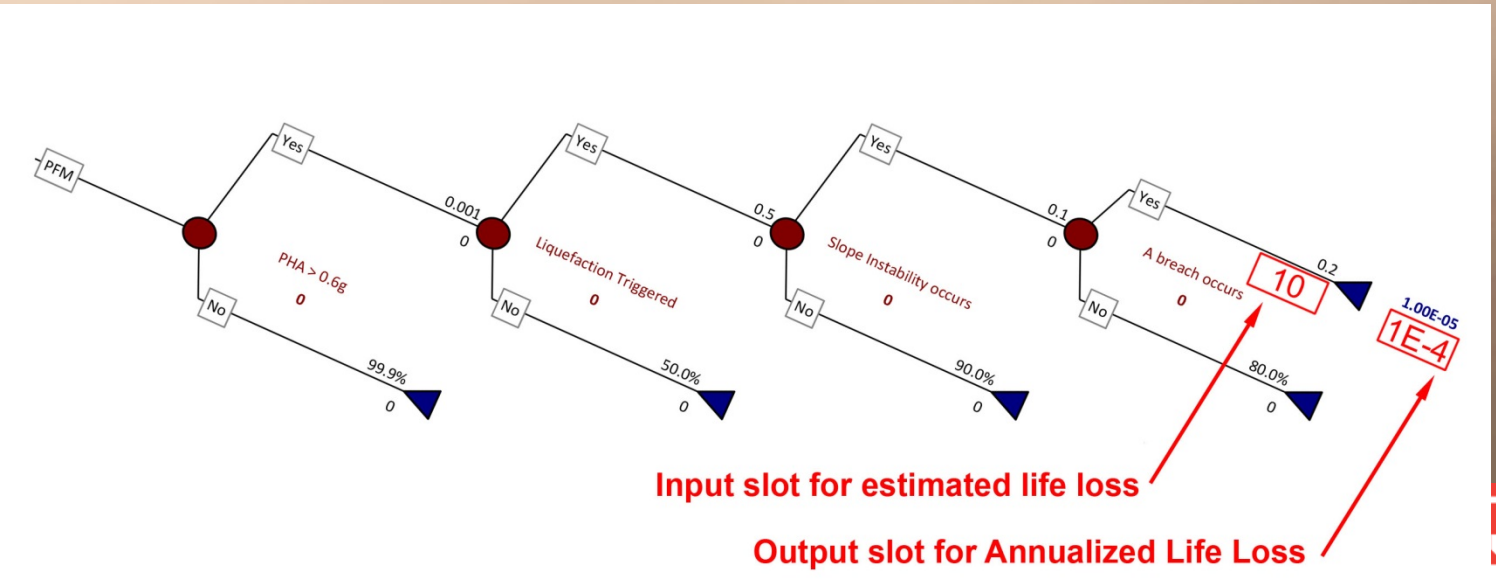
System response curve

- Allows the conditional probability of breach to be evaluated over a continuous range of residual freeboards
- Residual freeboard can be randomly sampled



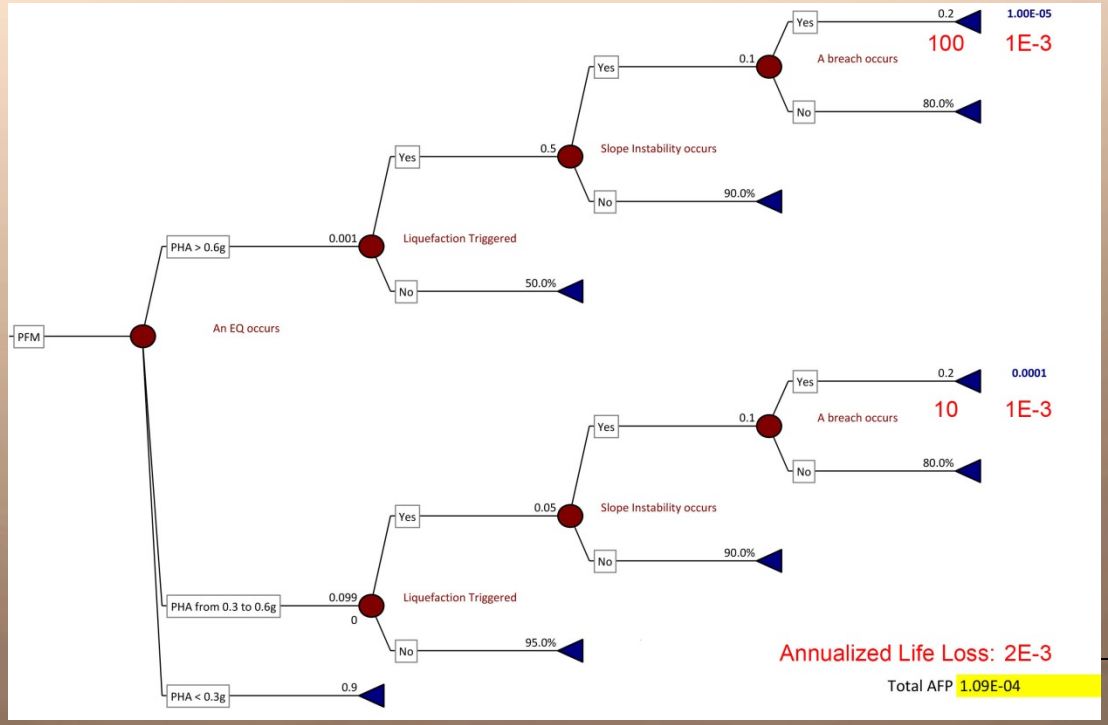
Annualized Life Loss

- An event tree built using event tree software often has “slots” for consequence estimates
- The number shown to the right of the slot is referred to as the Annualized Life Loss



Annualized Life Loss

- In a multiple-branch event tree, the total Annualized Life Loss would be calculated by summing across branches



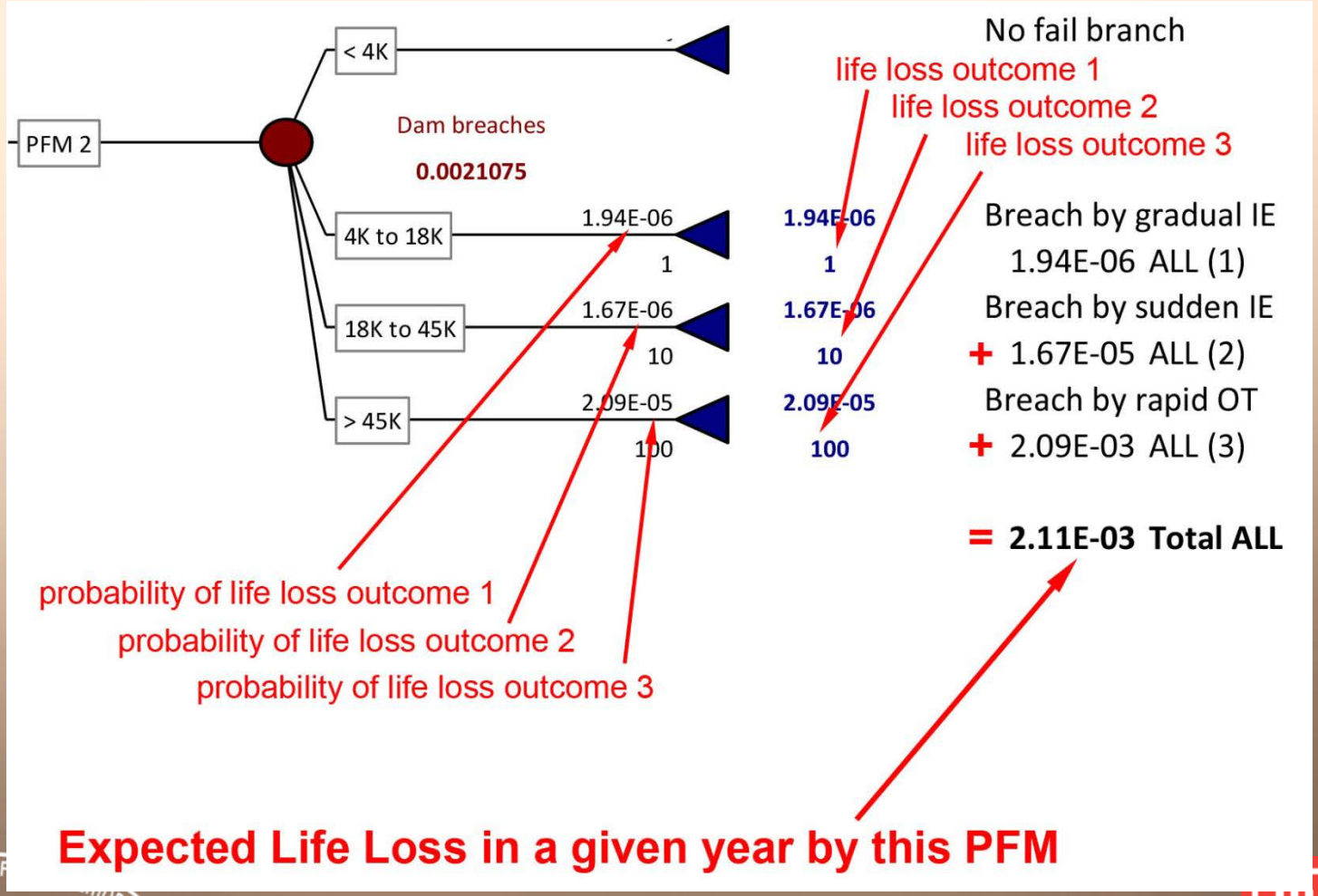
Annualized Life Loss

- The theoretical basis for these operations is different than that of the AFP calculations
- Annualized Life Loss is interpreted as the expected value of Life Loss, itself interpreted as a random variable (RV) whose “outcomes” are the life loss estimates developed for different breach mechanisms

Annualized Life Loss

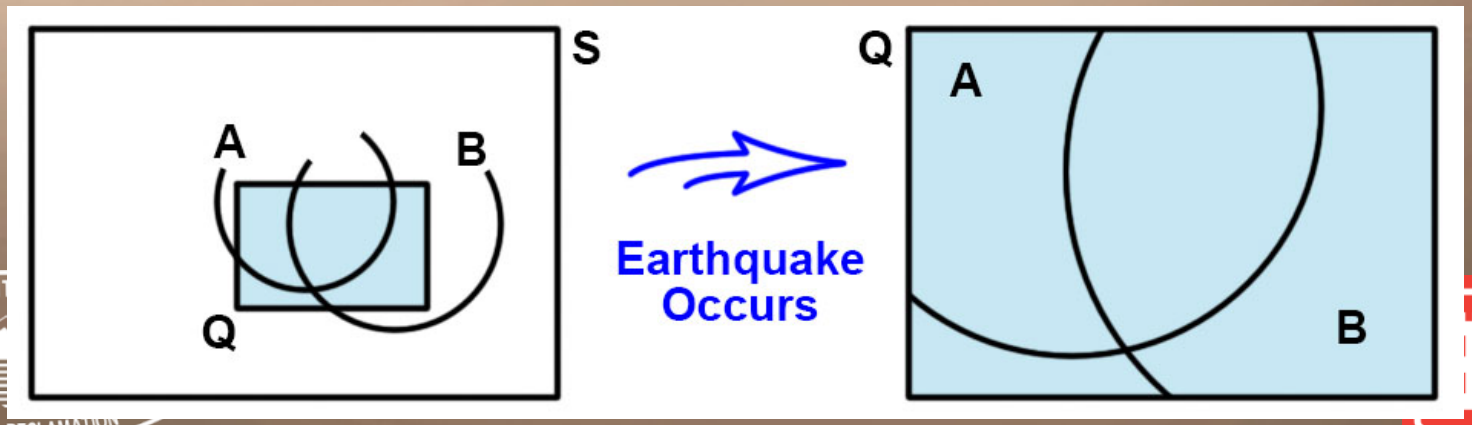
- The formula governing this operation is the expectation formula for a discrete RV:
  - Let L be a random variable representing the life loss associated with the failure of a dam or levee
  - For an event tree with n branches leading to breach, E[L] = $\sum_1^n l_i * P(l_i)$, where l is the life loss estimated for the ith branch and $P(\ l_i\ )$ is the sub-AFP calculated for the ith branch
- Annualized Life Loss = E[L]

Annualized Life Loss: summary



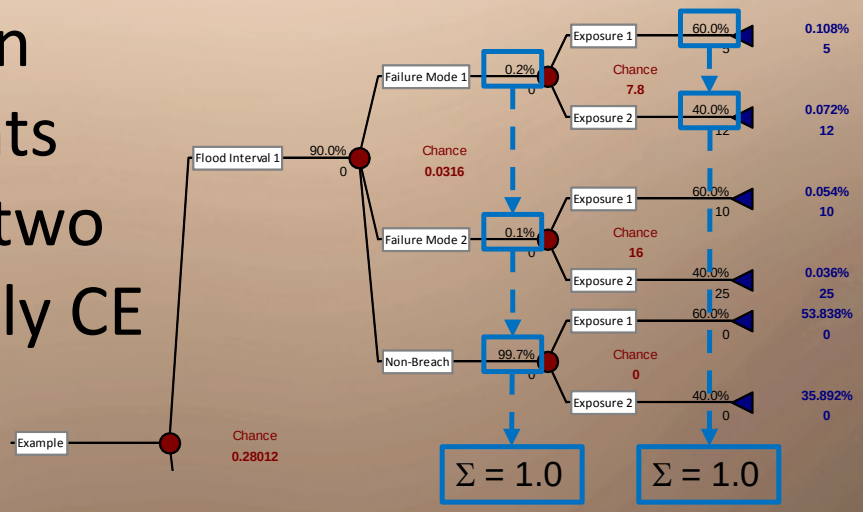
Common Cause Adjustment

- This topic is covered in detail in the *Combining and Portraying Risks* presentation
- The purpose of these slides is to illustrate how the use of event tree techniques could help identify the need to perform a CCA
- Note: could apply at various levels of event tree



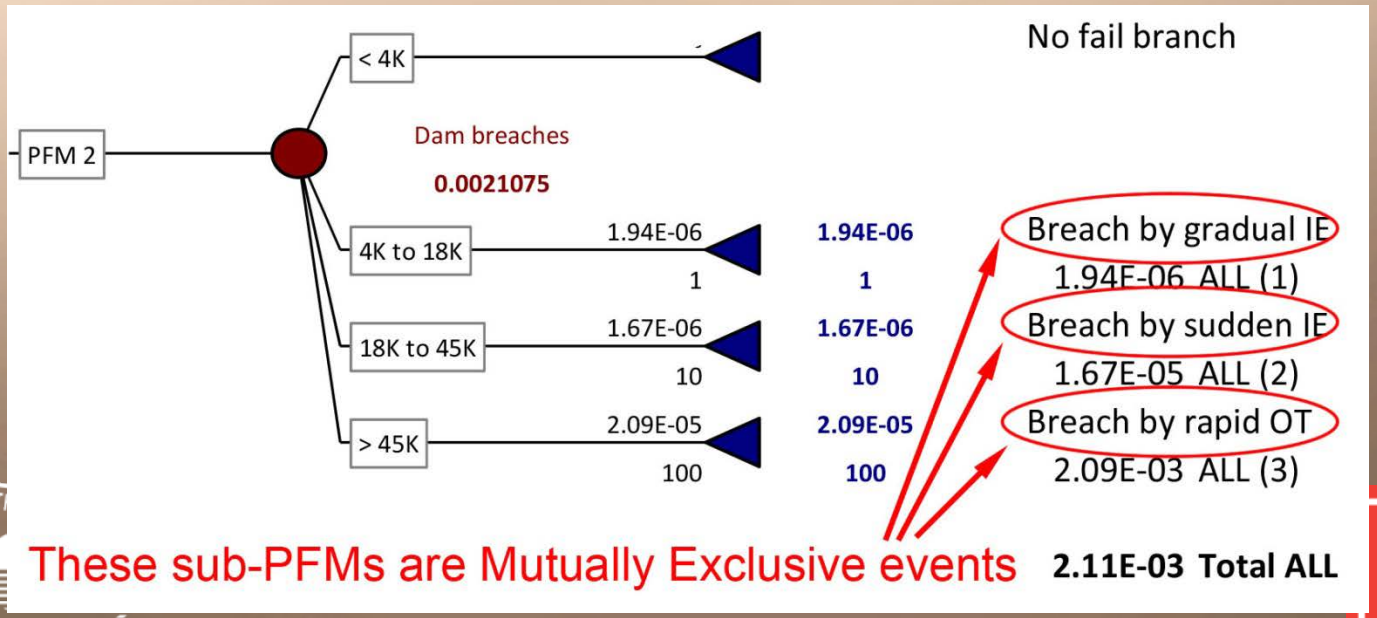
Background

- A split in the event tree represents a partition of the total event space (or a reconditioned sample space) into mutually exclusive and collectively exhaustive sub-events.
  - At a binary split in an event tree, the events represented by the two branches are not only CE and ME, but also Complementary.



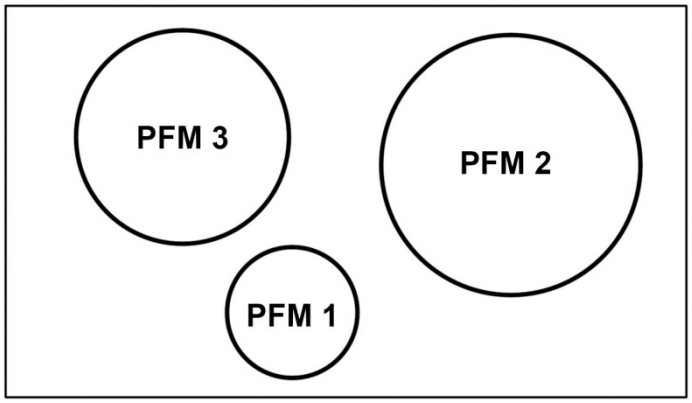
Background

- Since they are the result of the partitioning of event space into ME events, the intersection events associated with end-branches are also ME (but not CE, unless non-breach considered)



Background

- What this means is that if all the PFMs for a dam or levee were portrayed in a single properly constructed event tree, there would be no need to correct for the intersection area

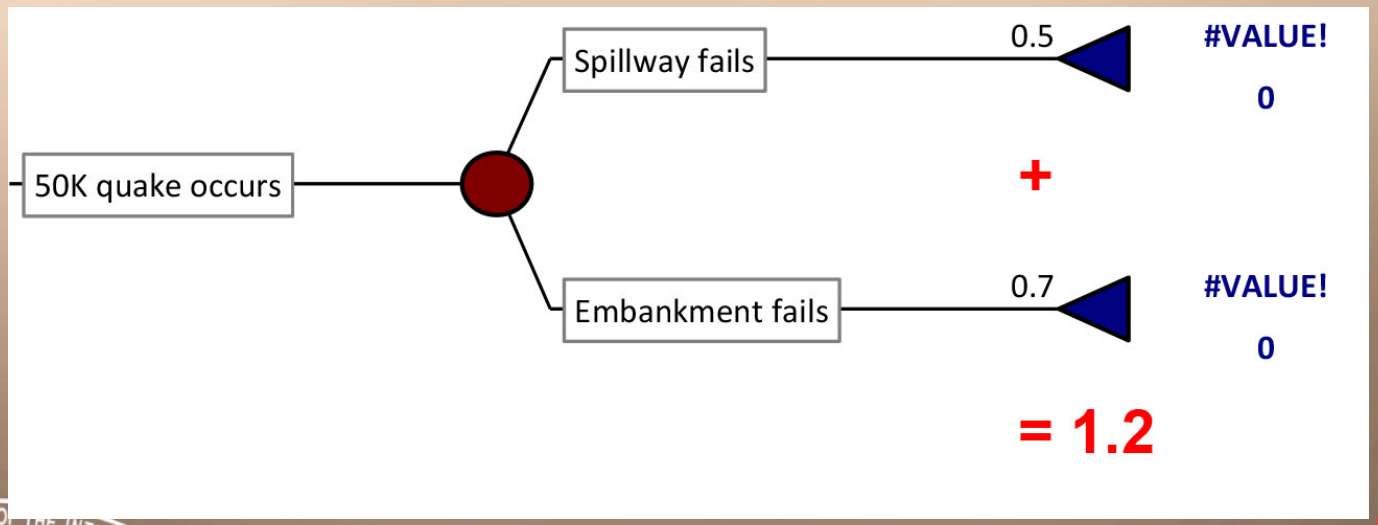


Common Cause Adjustment

- Unfortunately, the risks of different PFMs are often estimated at different times, or by entirely different teams
- Consider for example a seismic spillway PFM and a seismic embankment liquefaction PFM
- For the controlling > 50K EQ, spillway team has estimated a conditional failure probability of 0.5, whereas the embankment team has estimated a 0.7 conditional failure probability

Common Cause Adjustment

- When these results are portrayed in a single event tree, one of the axioms of probability appears to be violated (P[S] = 1 must still hold in a reconditioned sample space)

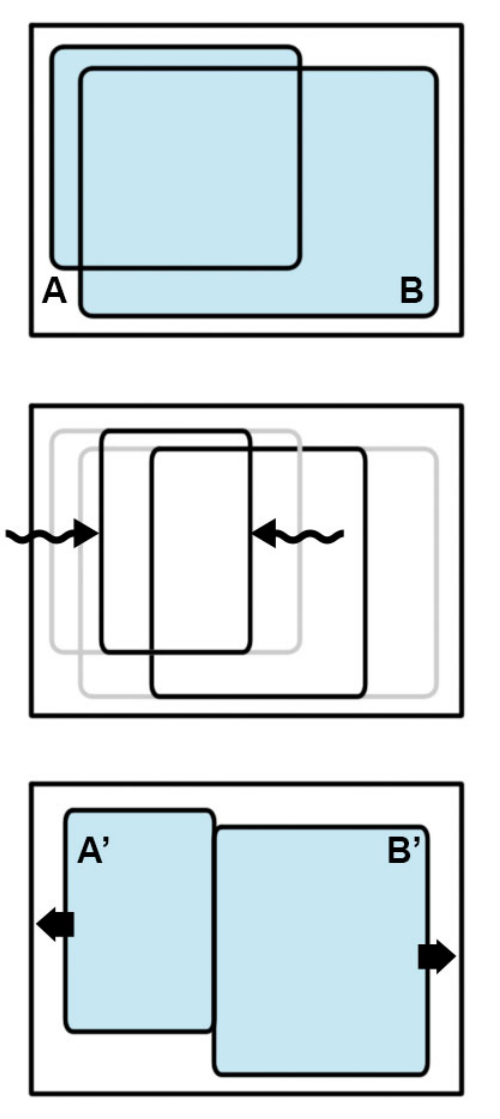


Common Cause Adjustment

- The problem is not that the risk estimates are "wrong", only that the events that they have been estimated for intersect.
- A finite intersection would be expected for two statistically independent events such as spillway failure and embankment breach
- Intersection area could be large in a space conditioned by occurrence of an EQ or flood
- Solution: Common Cause Adjustment (SI events)

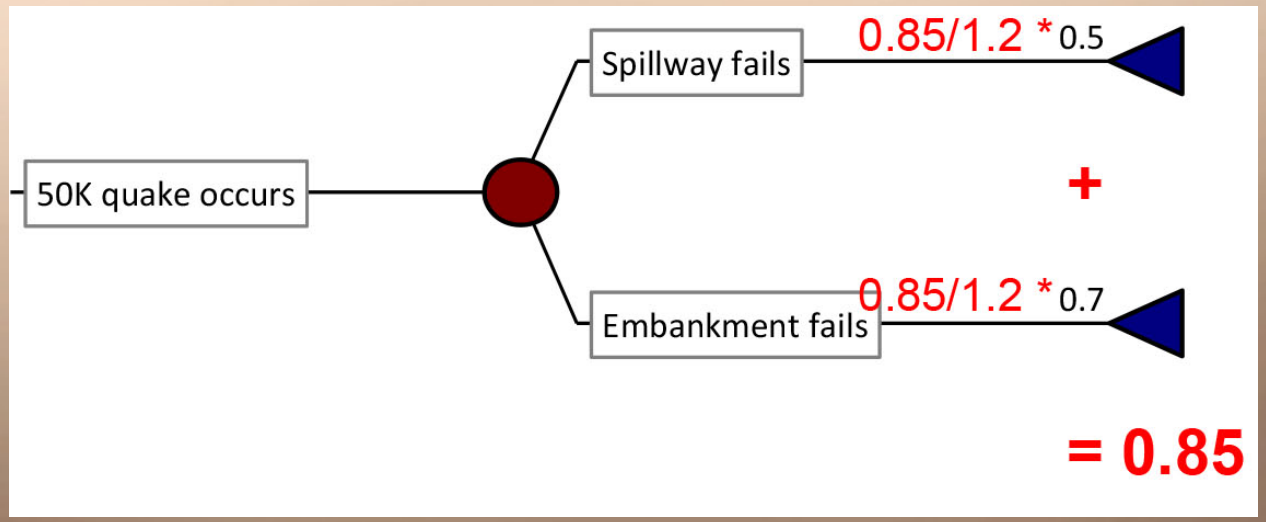
Common Cause Adjustment

- Purpose is to correct the <u>total</u> AFP
- For consistency (and bookkeeping), the intersection area is reallocated to individual PFMs, allowing their AFPs to sum directly to the total AFP
- For the purposes of calculating the total AFP, the adjusted PFMs can thus be treated as ME. Note this does <u>not</u> make them ME. The PFMs remain SI, and the original PFM risk estimates are still correct (events just intersect)



Common Cause Adjustment

- Adjusted conditional failure probabilities for the example (for details, see *Combining and Portraying Risks* presentation)



Monte Carlo Simulation

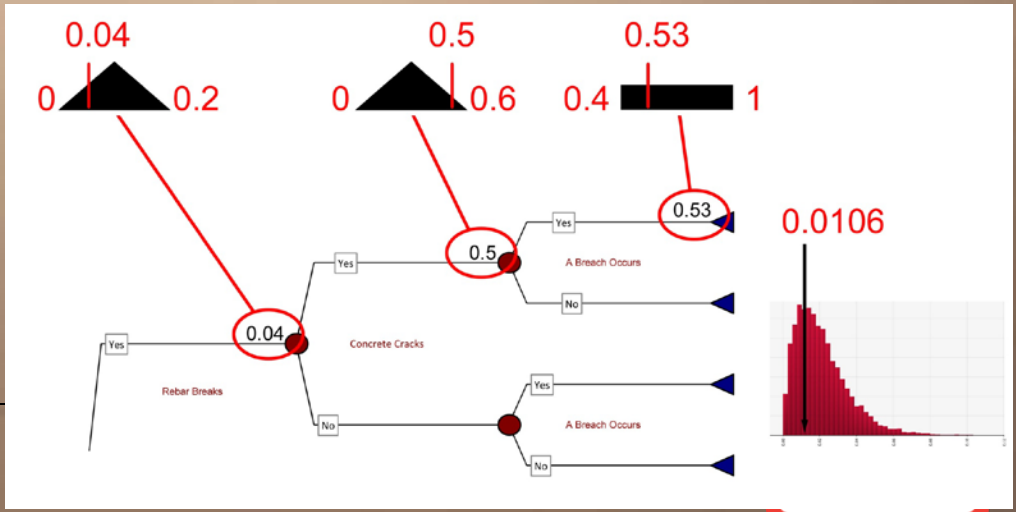
- Most of the examples and event trees in this presentation involve point estimates of conditional probability
- In practice, conditional probabilities are nearly always estimated as ranges or distributions
- The purpose of this is to capture uncertainty
- The most practical way to reflect the input estimate uncertainty in the output AFP is to perform a Monte Carlo simulation

Monte Carlo Simulation

- Monte Carlo simulation involves the random sampling of input distributions (such as conditional probability estimates, life loss estimates) in accordance with their shapes
- For each trial, an output value of AFP (or Annualized Life Loss) is recorded
- Many event tree software packages (such as Precision Tree) are set up to perform MC

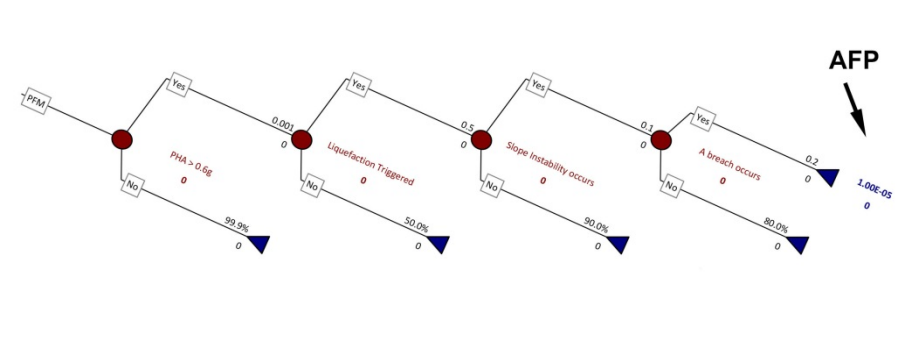
Understanding MC

- The AFP is a product function of the *n* probability distributions that form the basic inputs of the event tree
- The MC simulation process allows an output distribution to be built through repeated random sampling of inputs
- If we could easily develop a closed form solution for the PDF of the AFP, there would be no reason to perform MC
- Since we can't, the MC output distribution pro-vides a good model of what the analytical AFP distribution looks like



Product Distributions

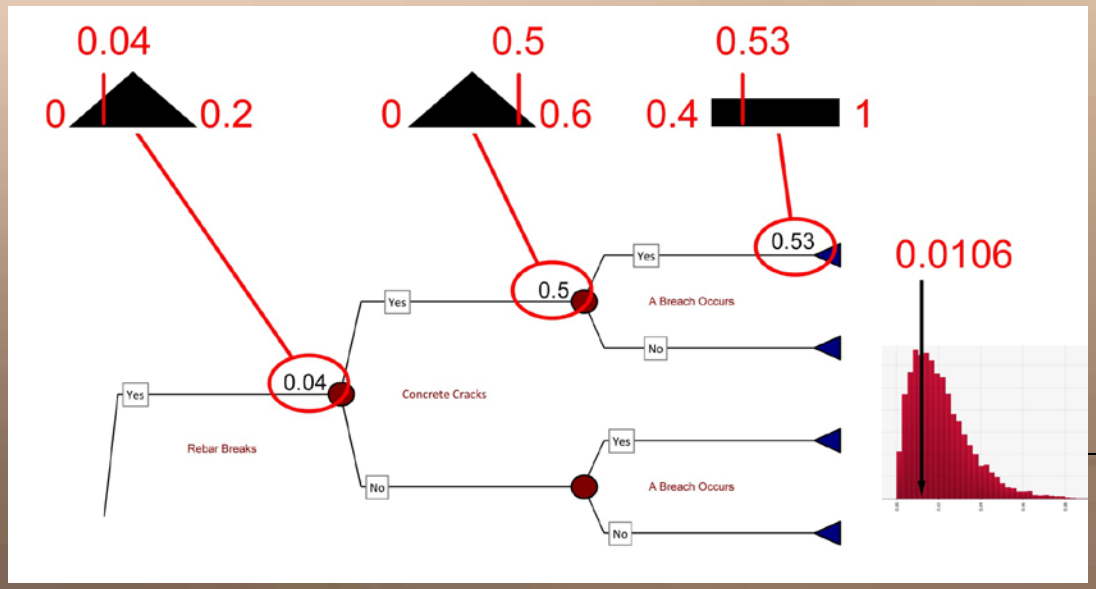
- For simple event trees, the mean AFP can be estimated by multiplying out the means of the input distributions. This is typically equal to the number shown in the AFP output cell with MC not running

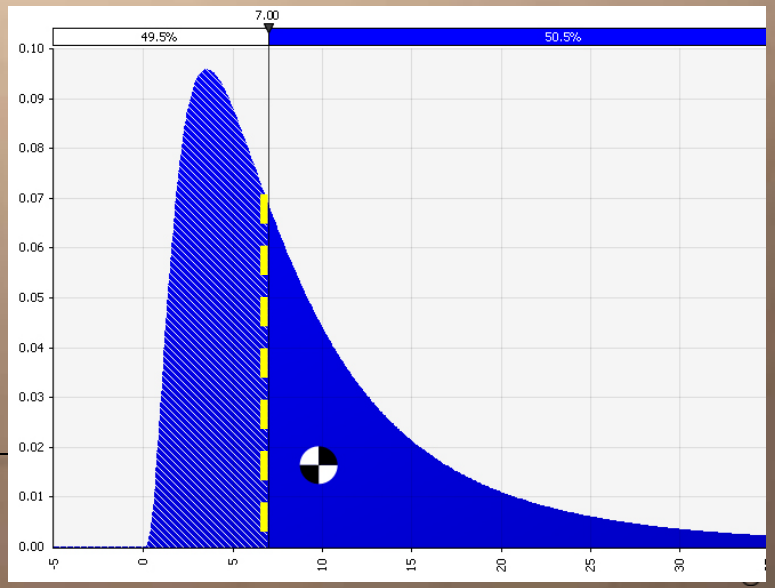
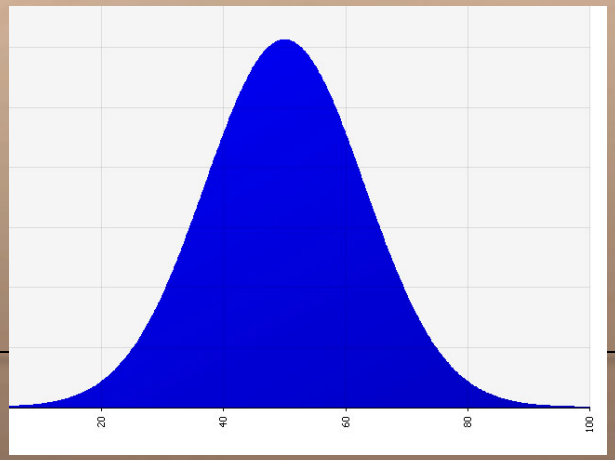


- For complex event trees involving background calculations and system response curves, the number shown in the AFP output cell with Monte Carlo not running is typically meaningless
- When probability distributions are used as inputs the mean AFP should <u>always</u> taken as the mean of the MC output distribution

Central Limit Theorem

- The Central Limit Theorem states that a product distribution of *n* random variables will tend toward the Lognormal as *n* → large # (also predicts that summed distributions approach Gaussian)
- Holds true regardless of what the input distributions look like, and the effect can be observed even when *n* is relatively small
- For a lognormal distribution, have $\mu > x_{50}$, so even if the input distributions are symmetric, the mean of the output distribution will typically not be equal to the product of the input means





Questions