Section 5: Uncertainty Analysis in Risk Assessment

- *Purpose*
  - *Students will see an overview of how Bayesian estimates are obtained in risk assessment*
- *Objectives*
  - *Through examples, students will learn about*
    - *Simulation of distributions with Monte Carlo sampling*
    - *Simulation of a “top event” probability by propagation of distributions through a logic model*
    - *Simple Monte Carlo sampling and Latin Hypercube sampling*

Risk

- *Three elements must always be considered*
  - *What things could happen?*
  - *What are their probabilities or frequencies?*
  - *What are their consequences?*
- *Must quantify answers, and assess uncertainty in the quantification*
- *In LOSP example*
  - *Events*
    - *Initiating event could occur*
    - *Then EDG power system could successfully operate or it could fail*
  - *Consequences*
    - *Electric power production, or something bad*
  - *Frequency of bad consequence is subject of this section*

Uncertainty Analysis in Risk Assessment

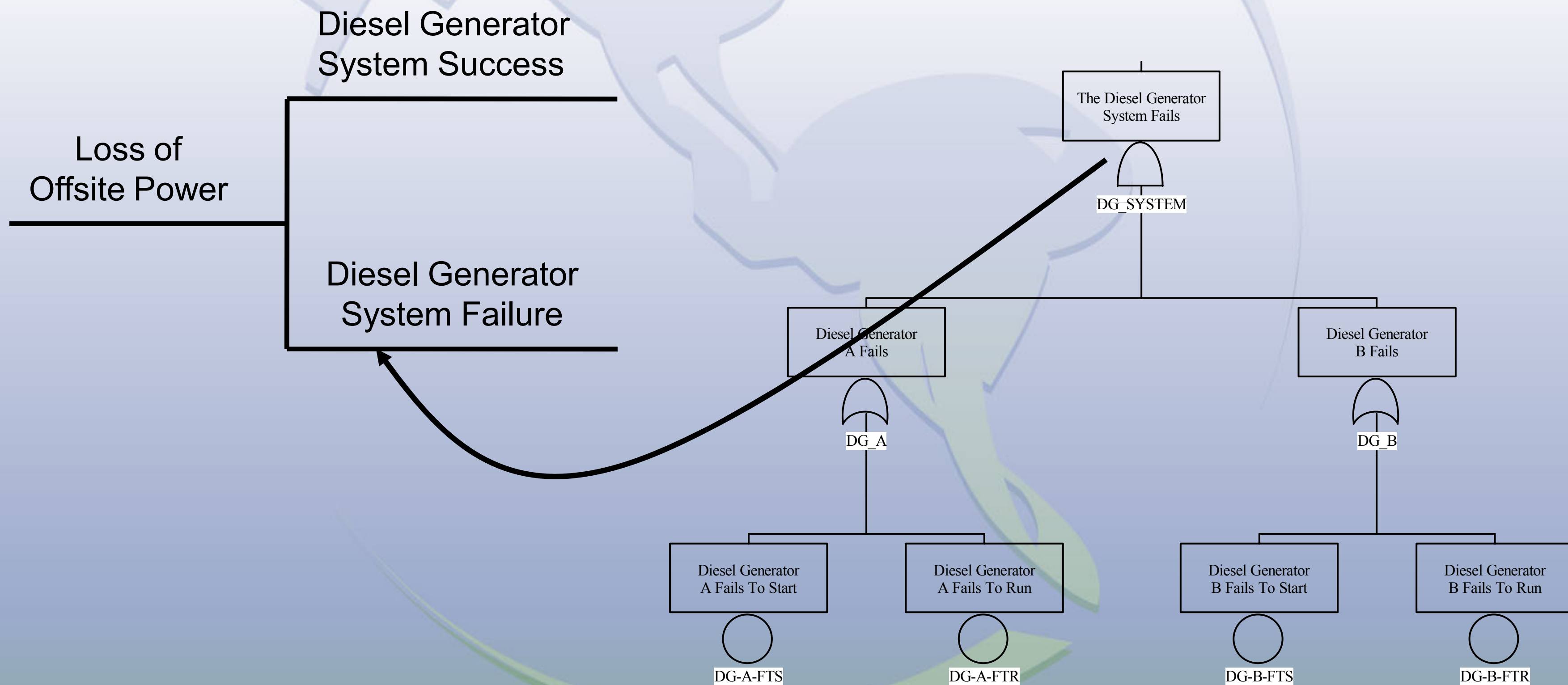
- ***Overall Approach***
- *In risk assessment, we estimate*
  - *Probability of "top event" (if looking at fault trees)*
  - *Frequency of "end state" (if looking at event trees)*
- *These estimates are based upon "minimal cut sets"*
  - *Minimal cut sets contain parameters such as*
    - *Failure rates*
    - *Probabilities of failure on demand*
  - *In the LOSP example, we develop $\lambda_{SystemFail}$ as a (fairly complicated) function of λ_{LOSP}, p_{FTS}, and λ_{FTR}*
  - *We approximate the Bayes distribution of the end-state frequency as follows*

Uncertainty Analysis (cont.)

- *Randomly sample a value of each basic parameter*
  - *This sample comes from its posterior distribution*
- *Samples are used to calculate a desired*
  - *Top-event probability*
  - *End-state frequency*
- *This process is repeated*
  - *Use new sampled values of the basic parameters on each iteration*
  - *Obtain many calculated values of desired result*
  - *Resulting values are a random sample from the Bayesian distribution of the top-event prob. or end-state frequency*
    - *Together, they approximate the result distribution*

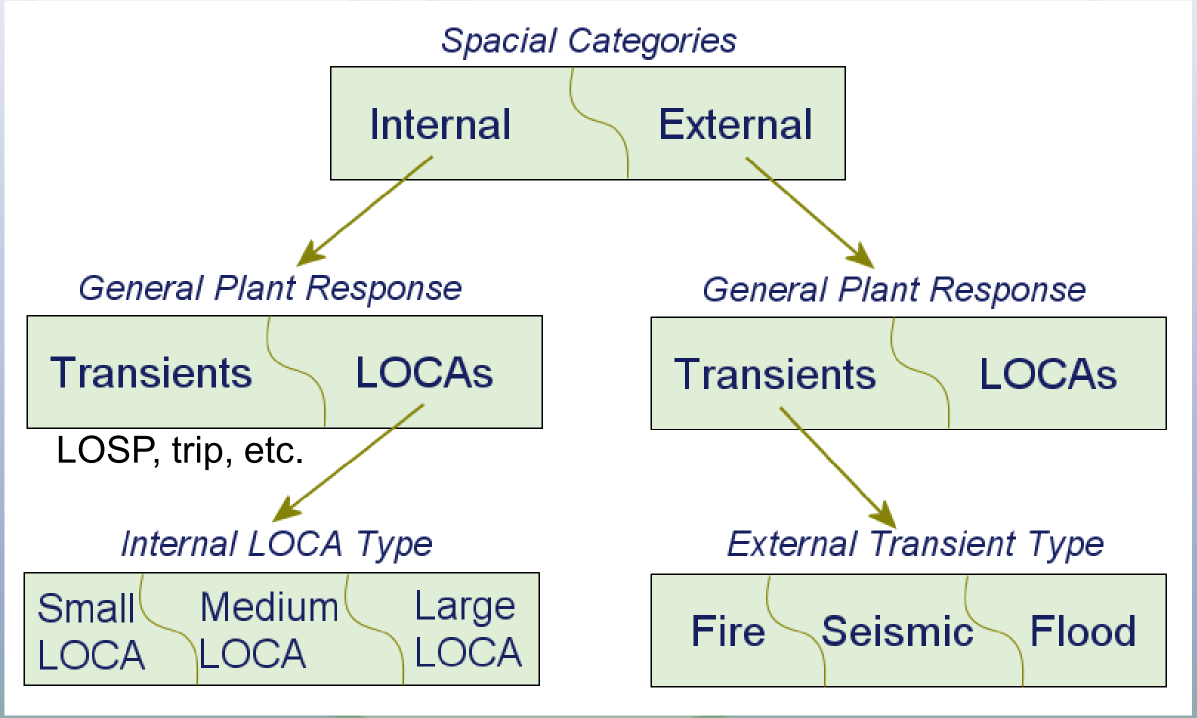
LOSP Example

- *Once again*



LOSP Example (cont.)

- *Note that LOSP is just one initiating event…this analysis process is carried out for all results from all initiating events*



PRA Minimal Cut Sets

- *In every minimal cut set there are "basic events"*
- *Every basic event stored in PRA database has some uncertainty about the value used for the event*
  - *The **propagation** of this uncertainty through cut sets must be performed in order to understand the overall uncertainty*

Minimal Cut Sets in LOSP Example

- $\lambda_{SystemFail} = \lambda_{LOSP} \times Pr[EDG\ system\ fails]$
- Pr[EDG system fails]

$$= Pr[(FTS_A \text{ and } FTS_B) \text{ or } (FTS_A \text{ and } FTR_B) \text{ or } (FTS_B \text{ and } FTR_A) \text{ or } (FTR_A \text{ and } FTR_B)]$$

$$\approx Pr(FTS_A \text{ and } FTS_B) + Pr(FTS_A \text{ and } FTR_B) + Pr(FTS_B \text{ and } FTR_A) + Pr(FTR_A \text{ and } FTR_B)$$

(rare event approximation)

$$= Pr(FTS_A) \times Pr(FTS_B) + Pr(FTS_A) \times Pr(FTR_B) + Pr(FTS_B) \times Pr(FTR_A) + Pr(FTR_A) \times Pr(FTR_B)$$

assume EDGs A and B fail independently

Minimal Cut Sets in LOSP Example (cont.)

- Generic forms for basic event probabilities
  - $Pr(FTS) = p_{FTS}$
  - $Pr(FTR) = 1 - e^{-\lambda_{FTR} t_{mission}} \approx \lambda_{FTR} t_{mission}$
- $Pr(FTS_A) \times Pr(FTS_B) = ?$
  - p_{FTS}^2 ? (one estimated parameter)
  - $p_{FTS\text{-}A} \times p_{FTS\text{-}B}$?(two estimated parameters)

How Many Distinct Parameters in Example?

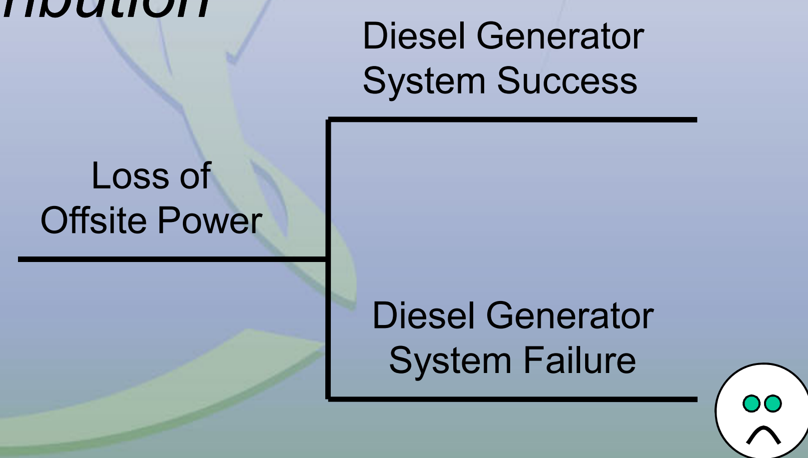
- If we distinguish between $p_{FTS\text{-}A}$ and $p_{FTS\text{-}B}$
  - Assume that the two p's differ to important degree
  - Use only data from EDG i to estimate $p_{FTS\text{-}i}$
    - Have relatively more uncertainty in each estimate
    - Same prior for each $p_{FTS\text{-}i}$?
- If we model only a single p_{FTS}
  - Assume that the two p's are nearly equal
  - Use data from both EDGs, and generic prior, to estimate the one p
    - Have relatively less uncertainty in the one estimate
    - Use generic prior

How Many Distinct Parameters in Example? (cont.)

- *If assign independent Bayes distributions to $p_{FTS\text{-}A}$ and $p_{FTS\text{-}B}$*
  - $E(p_{FTS\text{-}A} \times p_{FTS\text{-}B}) = E(p_{FTS\text{-}A}) \times E(p_{FTS\text{-}B})$
- *If assign Bayes distribution to p_{FTS}*
  - $E(p_{FTS}^2) > E(p_{FTS}) \times E(p_{FTS})$
- *So if the two parameters are really the same*
  - *Modeling them with independent distributions yields too small a mean.*
- *In SAPHIRE, to force $p_{FTS\text{-}A}$ and $p_{FTS\text{-}B}$ to equal each other, i.e. to equal p_{FTS}*
  - *Assign them to a single* ***correlation class.***

End-State Frequency in LOSP Example

- Assume single p_{FTS}, single λ_{FTR}
- $\lambda_{SystemFail} \approx \lambda_{LOSP} \times [p_{FTS}^2 + 2p_{FTS}\lambda_{FTR}t_{mission} + (\lambda_{FTR}t_{mission})^2]$
- Approximate the Bayes distribution of $\lambda_{SystemFail}$ by a (large) random sample from the distribution



Uncertainty Analysis for NMSS Application

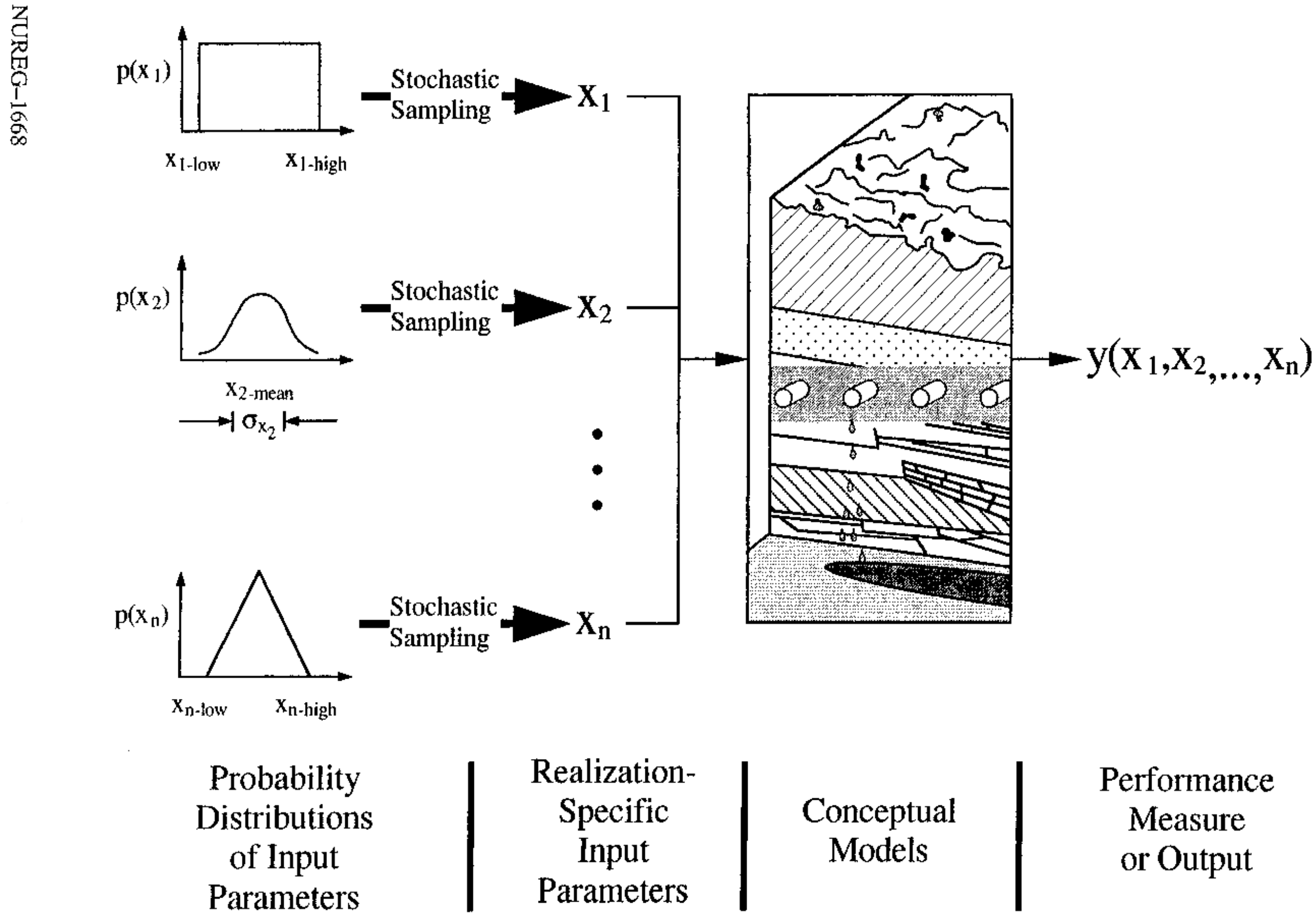


Figure 4-1. A diagram illustrating the use of the Monte Carlo method in performance assessment.

Propagation of Uncertainty

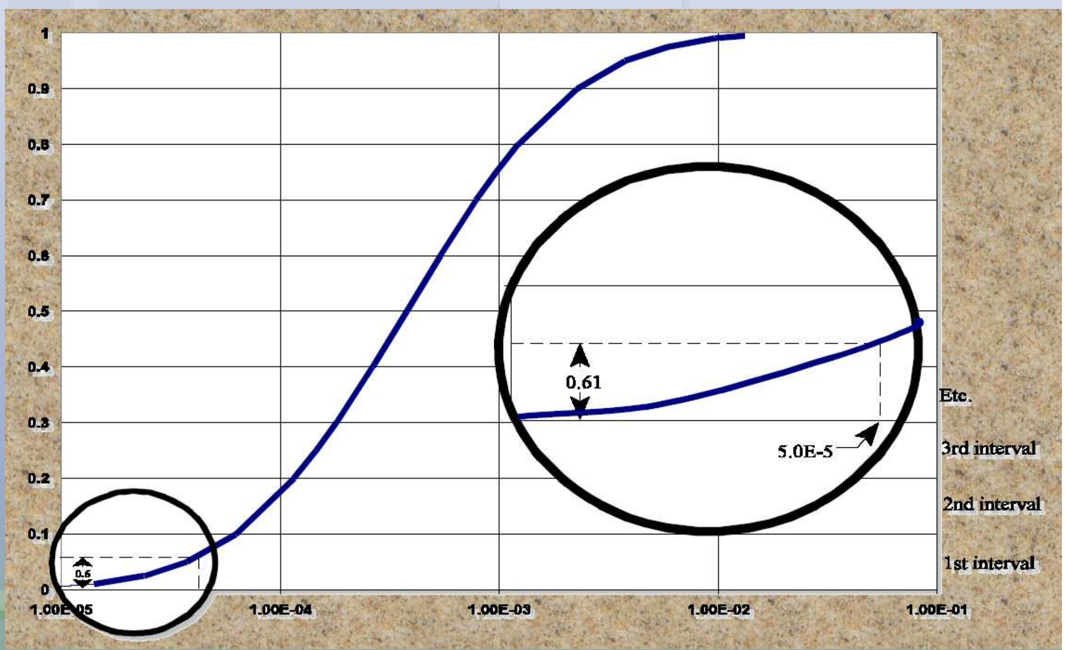
- *Prior to performing the uncertainty analysis, the PRA requires answers to three queries:*
  1. *The type of sampling*
     - *Simple random sampling (SRS)*
     - *Latin hypercube sampling (LHS)*
  2. *The number of iterations (i.e., samples)*
     - *For example, if we specify 2,000 samples and there are 10 unique basic events, we generate 20,000 random numbers*
  3. *The random number generator seed value*

Two Kinds of Sampling

- ***Simple Random Sampling (SRS)***
- *In simple random sampling, each parameter is sampled at random*
  - *The sampled values are entered into the logic model(s)*
  - *The frequency/probability of the top event is calculated*
  - *This process is repeated many times (up to the number of samples specified)*

Two Kinds of Sampling (cont.)

- ***Latin Hypercube Sampling (LHS)***
- *In Latin hypercube sampling, each parameter is sampled in a stratified way, to guarantee that **each** portion of the range of the distribution is represented*
- *An example with 10 stratifications is shown*
  - *Within each portion, we randomly sample*



LHS (cont.)

- *For example, let us denote one parameter by p*
  - *Bayesian distribution of p is known, the posterior distribution of p based on prior information and relevant data*
  - *If 10 samples were to be constructed*
    1. *p would be sampled randomly from interval ($p_{0.0}$, $p_{0.10}$), giving a value that we denote as p_1*
    2. *Again, sample randomly from interval ($p_{0.10}$, $p_{0.20}$), giving a value that we denote at p_2*
    3. *Repeat process until we have p_{10} [from interval ($p_{0.90}$, $p_{1.0}$)]*
  - *This is **stratified sampling**, in which the sampled points are forced to cover entire range of the distribution*

LHS (cont.)

- *After all parameters in the model have been sampled in this stratified way, they are randomly matched to each other*
  - *In example with λ_{LOSP}, p_{FTS}, and λ_{FTR}, one of the sampled values of each parameter would be chosen*
  - *However, they would be chosen so that the largest value of one parameter is **not** necessarily matched with largest or smallest values of other parameters*
  - *Instead, the choice of each pairing is random*
  - *For the chosen values, the top-event is calculated*
  - *Then another set of sampled parameter values is chosen, using values that have not been chosen yet*
- *In this way, a number (10 in this example) of values are calculated for the end-state frequency*

Differences Between Sampling Types

- *While there are computational differences between the two techniques (SRS and LHS):*
  - *One should not be too concerned about which technique is selected for a particular analysis*
  - *Instead, one should be concerned about **convergence** of the numeric calculation*
  - *Convergence may be checked by noting change (or lack thereof) of uncertainty results as the number of samples is varied*
- *The samples from either method **converge** to the Bayes distribution of the end-state frequency or top-event probability*

The Seed Value

- *A seed value tells software where, in sequence of possible random numbers, to* ***start selecting*** *random numbers*
  - *The random number generator gives a sequence of "random" integers (which are typically converted to real numbers)*
  - *A seed of "51" may tell us to start at the i'th random integer*
  - *A seed of "1,236" may tell us to start at the j'th random integer*
  - *etc.*
- *Again, checking for* ***convergence*** *should make seed selection irrelevant*
  - *But, to reproduce analysis results, one must use the same seed and same number of samples*

Accuracy of Sampling

- *Accuracy of a simple random sample is roughly **proportional** to square root of sample size*
  - *For example, if $\lambda_{SystemFail}$ is sampled from its distribution n times*
    - *Mean of the distribution is estimated by **average** of n sampled values (the sample mean), and this average has standard deviation proportional to $1/\sqrt{n}$*
    - *Estimate of this quantity is the standard error*
      - *A confidence interval equals the sample mean $\pm$ a multiple of the standard error*
- *LHS is more **complicated** than simple random sampling*
  - *But requires **fewer samples** for comparable accuracy*
  - *Therefore, it is justified if each calculation of top-event is expensive or time-consuming*

Uncertainty Analysis Results

- *Every result from the PRA is uncertain*
  - *Individual basic events (FTS, FTR, etc.)*
  - *Initiating event frequency*
  - *System failure probability*
  - *Overall results such as core damage frequency*