Efficient Sampling Strategies to Estimate Extremely Low Probabilities

Cédric J. Sallaberry, Robert E. Kurth, Frederick W. Brust, Elizabeth A. Kurth-Twombly

Engineering Mechanics Corporation of Columbus (Emc^2)
3518 Riverside Drive - Suite 202
Columbus, OH 43221-1735
Phone: (614) 459-3200
Fax: (614) 459-6800

Context

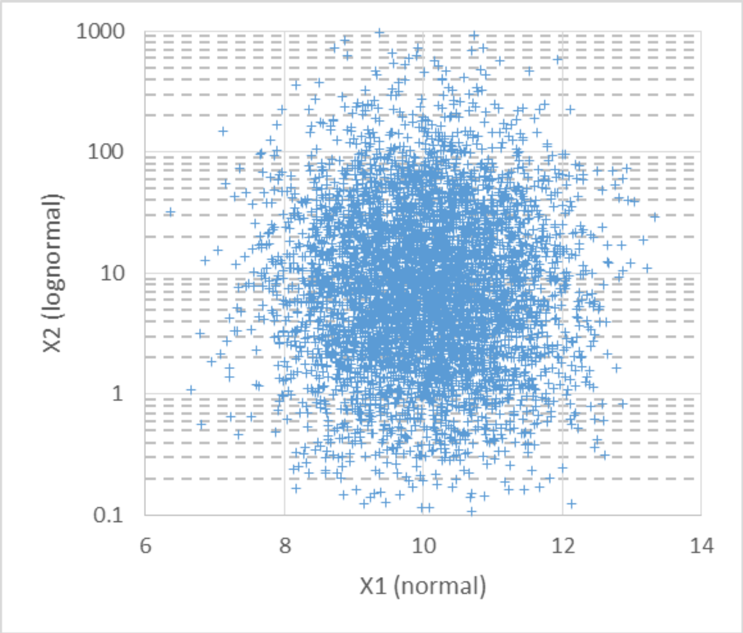
- **The result of active degradation mechanisms, primarily Primary Water Stress Corrosion Cracking (PWSCC), presented challenges to the Leak Before Break (LBB) analysis**
- **The US Nuclear Regulatory Commission (USNRC) and the Electric Power Research Institute (EPRI) develop jointly a probabilistic fracture mechanics code to assess LBB: xLPR (extremely Low Probability of Rupture)**
- **The extremely low occurrence of events (10^{-6} range) requires optimization of the sampling based methods in order to generate adverse events when using reasonable sample size.**
- **Approach derived:**
  - Sampling strategy with denser covering of either each input or the multidimensional space
  - Sensitivity Analysis to identify uncertain parameters driving the output of interest (topic of another presentation PSAM-102)
  - Use of Importance Sampling

Sampling strategies

- **Simple Random Sampling (SRS)**
  - Corresponds to the original Monte Carlo approach
  - Used as reference
- **Latin Hypercube Sampling (LHS)**
  - Stratification of each input into equiprobable segments.
  - Sample each segment once
  - Assemble randomly (or using correlation control) between inputs
- **Discrete Probability Distribution (DPD)**
  - Discretize each distribution according to a number of strata
  - Use the conditional expected value for each strata (not the mid-point)
  - Sample randomly the resulting discrete distribution

Example problem for comparison

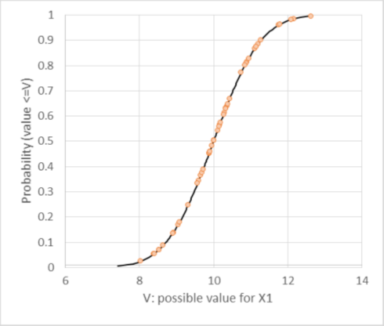
- **Sampling of two variables $X_1 \sim N(\mu = 10; \sigma = 1)$ and $X_2 \sim LN(\mu = 2; \sigma = 1.5)$**
- **Sample of size 50. DPD uses 10 strata**

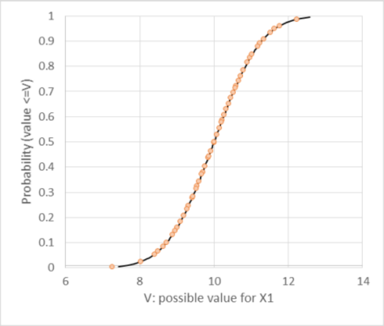


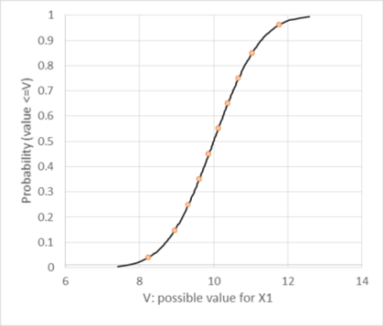
example of sampled values for a sample of size 50

Results

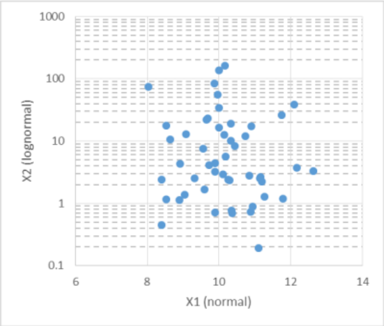
SRS | **LHS** | **DPD**

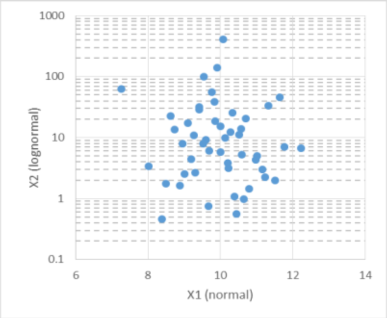


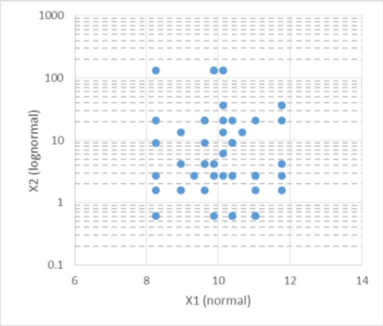




Dense coverage of each input with LHS. Helps if extreme values are needed.



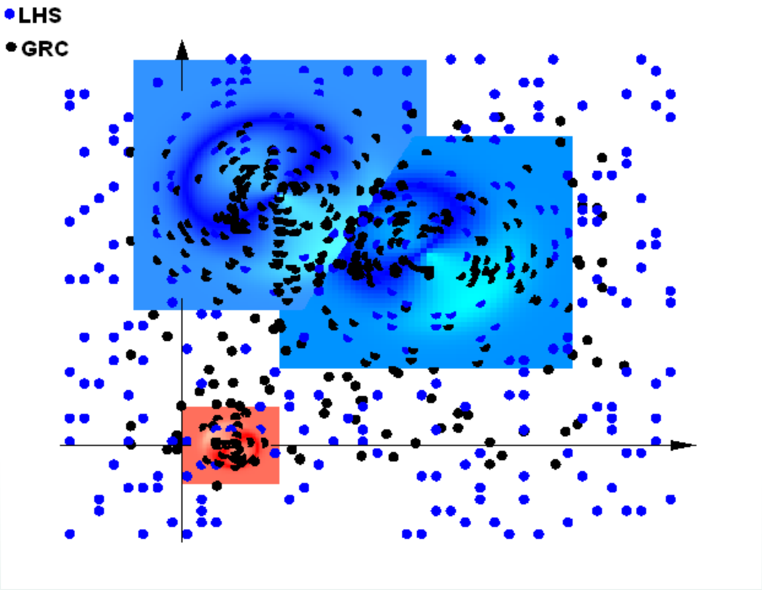




Better multi-dimensional coverage for DPD. Helps in conjoint influence.

Importance Sampling

- Variance reduction technique
- Purpose is to oversample the region of interest and under-sample the regions with no events.



- Weighting of each realization to reflect the area it represents.

Limitations

- **Importance sampling requires knowledge of:**
  - Which input parameters need to have importance sampling
  - Which region needs to be over sampled
- **Too many parameters with importance sampling brings dilution (not efficient importance sampling) or too much focus (weights are all too small to represent the problem correctly)**
- **Missing the region or choosing the wrong inputs leads to results potentially worse than with regular sampling**

Use of Sensitivity Analysis as support

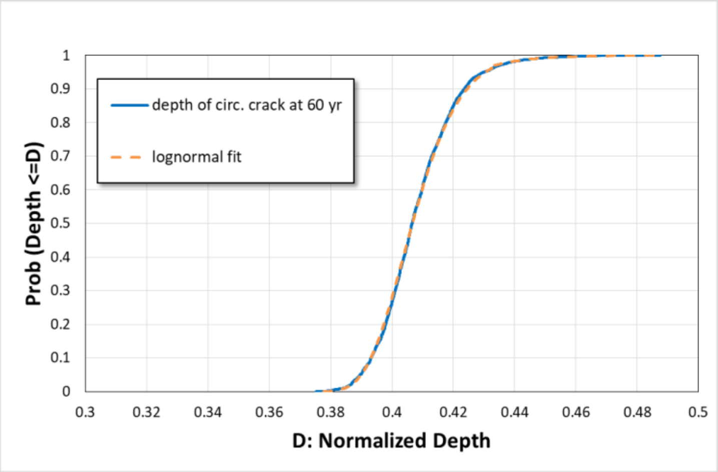
- Sensitivity Analysis rank the input variables in term of importance to the uncertainty in the output of interest
- Scatterplots can help identify the region of most importance
- Using Sensitivity Analysis on a first set of sample results *providing that one has at least some events occurring* helps set up a more appropriate importance sampling strategy
- If no events occur, surrogate outputs can be used (for instance, if no cracks rupture, select through wall cracks or deepest/largest cracks)

Computational model used

- xLPR V2.0 soon to be released
- Deterministic Models, using more than 500 inputs (most can be uncertain in the probabilistic version)
- Large number of submodels
  - Crack Initiation (coalescence and transition)
  - Crack Growth (SCC and fatigue) – growth at deepest and surface points
  - Crack Stability (J-tearing theory)
  - ISI
  - COD
  - Leak Rate and detection capability
  - Weld Residual Stresses – play a key role for SCC
  - K-solutions (weight function approach)
  - WRS mitigation
- One deterministic calculation takes ~4 seconds.

Example Study

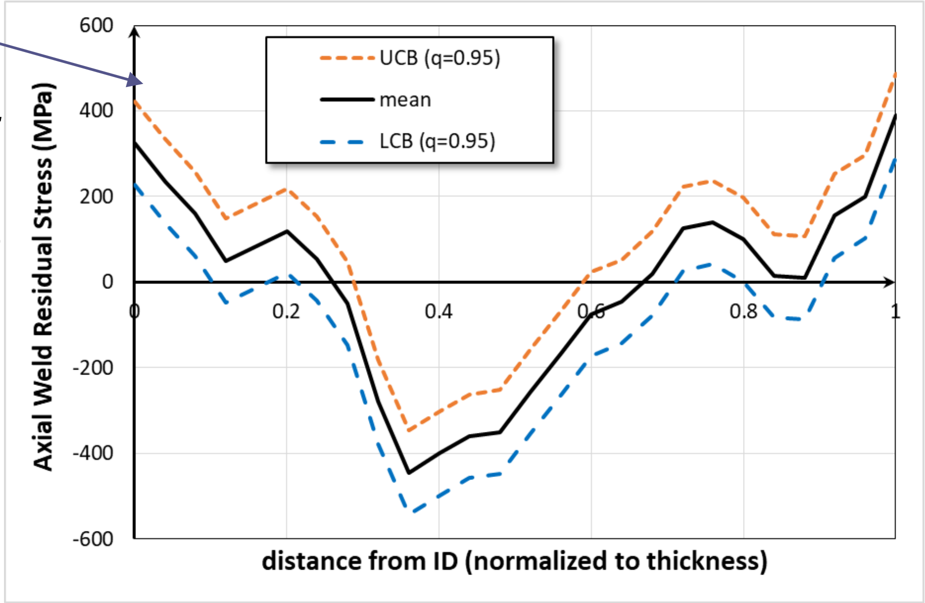
- Double V-Groove weld with bottleneck around 40% to 50% through circumference
- No TWC or rupture for 2,500 and 10,000 realizations
- Fitting distribution on depth and extrapolating gives probability of rupture in $[10^{-6}, 10^{-4}]$ range with most likely around 10^{-5}

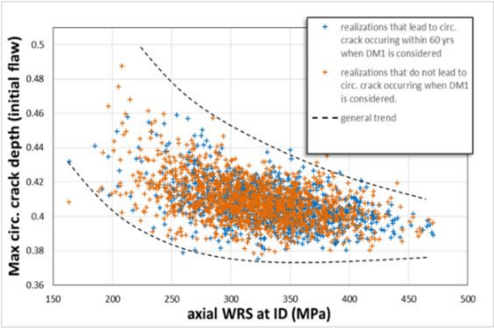


a	$P(X > a)$
0.5	1.28×10^{-4}
0.51	5.88×10^{-5}
0.52	2.73×10^{-5}
0.53	1.29×10^{-5}
0.54	6.14×10^{-6}
0.55	2.97×10^{-6}

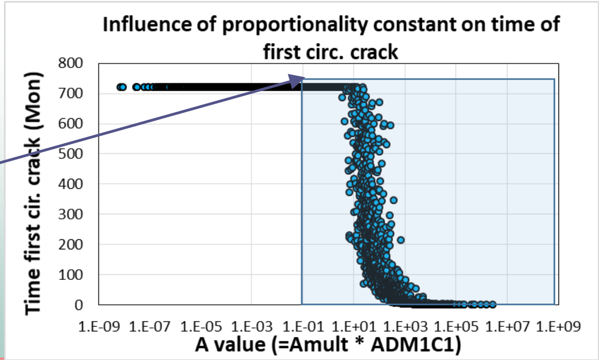
Example study : selection of importance sampling

Due to axisymmetric conditions, high values at ID leads to crack stopping 40% through the thickness



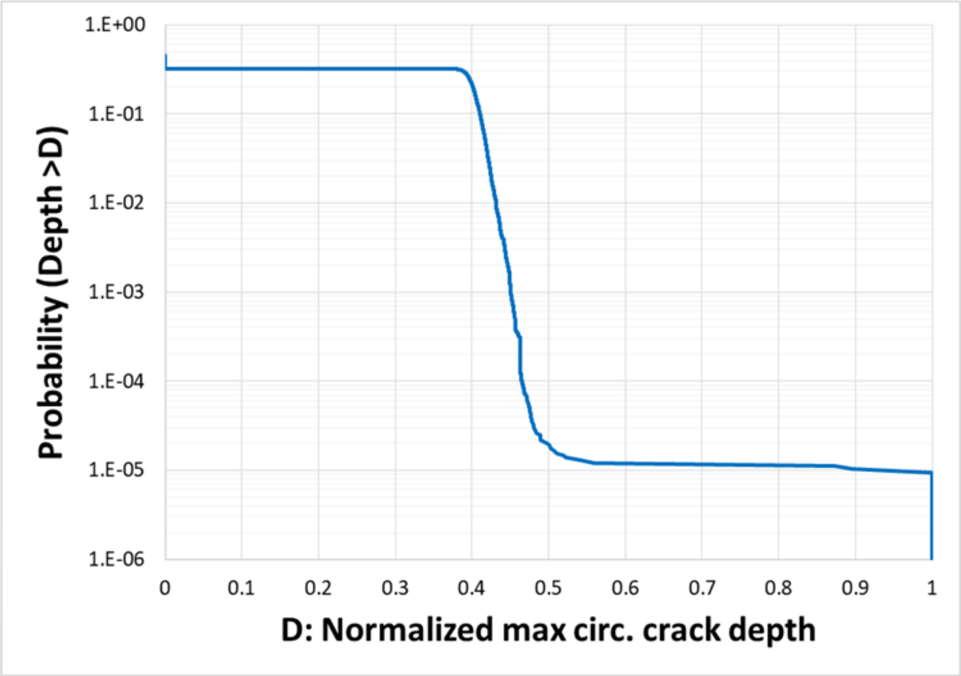


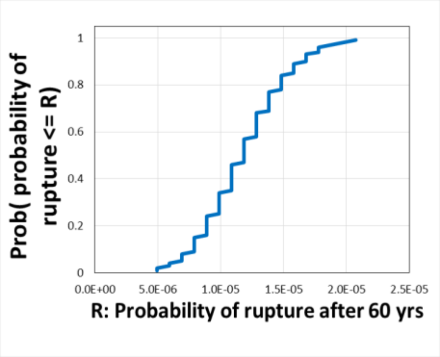
Low value of proportionality constant leads to no crack occurring

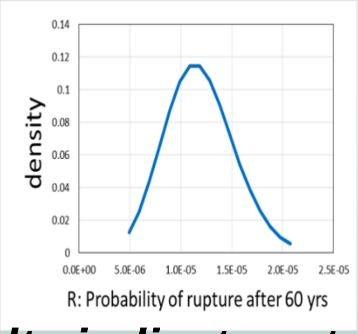


Results

- **Importance sampling on A parameter (high values) and WRS (low values) gives probability around 10^{-5}**







Bootstrapping of results indicates stability

Conclusion

- Extremely low probabilities require good understanding of the problem and a robust approach to estimate results with confidence using a reasonable sample size
- LHS and DPD have proven successful as a first step for the approach to identify the input drivers
- The use of Sensitivity Analysis helps in the selection of those drivers
- Importance sampling can then be applied with a greater chance of success