Sampling Size Issue in PRA Uncertainty Analysis

Chunrui Deng

2018.9

Introduction

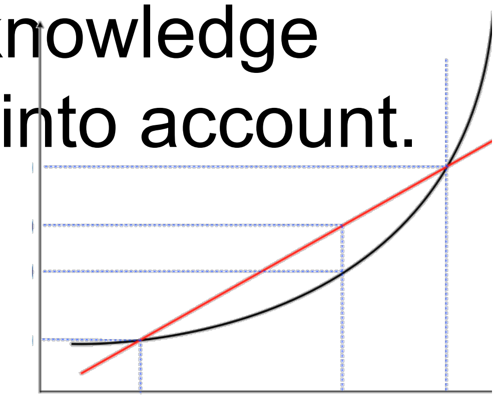
- Uncertainty is inherent
  - epistemic, aleatory
  - parameter, model, completeness
- Sampling is a natural choice
- In risk-informed application, “mean” value is often a preferred statistics, as it usually corresponds to a high quantile.
- Quantile is another concern

Introduction

- The issue
  - If a formal uncertainty propagation is necessary
  - Sample size needed for mean value estimate
  - Sample size needed for quantile estimate
- This paper
  - Survey the mathematics behind
  - Give illustrative examples

Point estimate vs. Mean

- An often oberserved phenomenon
  - The point estimate with mean parameter inputs is approximately the same as the sampling mean, when state-of-knowledge correlation (SOKC) is not taken into account.
- $E[f(\lambda)] = f(E\lambda)$?
  - Jensen inequality
    - for a convex function, $g(px + (1-p)y) \leq pg(x) + (1-p)g(y)$, $g(EX) \leq Eg(X)$, the equality relationship holds if and only if X is constant or g is a linear function.



Point estimate vs. Mean

- Without SOKC
  - The probability of a basic event can be approximated to a linear function of input parameters.
  - The basic events in a cutset are independent
- With SOKC
  - The basic events in a cutset may be correlated
  - $E[X^2] = E[X]^2 + Var(X)$
  - A formal uncertainty propagation could be necessary

Point estimate vs. Mean

- 2-redundant components, with or w/o SOKC

	Failure rate λ with a gamma distribution with parameters α and β			Mean failure probability of 1 component	Point estimate of the failure probability of 1 component

- The difference is significant.
- When possible, sampling is recommended.

Sample size for mean estimate

- Central limiting theorem
  - for i.i.d. random variables with limited v $\sqrt{n}(\bar{X}$ — asympotically follows standard normal ($\lim G_n(x) = \int^{x} \frac{1}{}$
- Slutsky theorem

$$\frac{\sqrt{n}(\bar{X}_n-\mu)}{} \rightarrow$$

- thus, for large samples, we can use sample variance to substitue population variance.

Sample size for mean estimate

- We can construct a 1-α confidence interval for mean value

$$\left\{\mu : \bar{x} - z_{\alpha/2}\frac{s}{\sqrt{\ }} \le \mu \le$$

- The length of the interval is

$$2z_{\alpha/2}$$

- We can define the relative deviation as a measure of the accuracy of the estimate

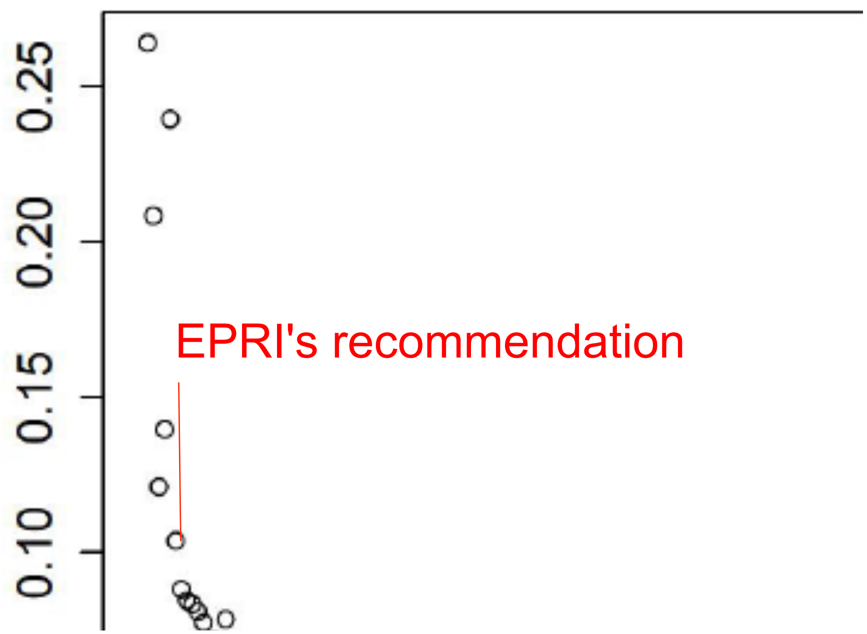
$$2z_{\alpha/2}\frac{s}{\sqrt{\ }}$$

- Thus, the size needed is

$$n > \frac{4z^{2}}{}$$

Sample size for mean estimate

- Example:
  – Lognormal with mean 1E-5, EF=10
  – 90% confidence level, required relative deviation 2%，necessary sample size 170,000



Sample size for mean estimate

- Example: $n > \frac{4z}{\;}$ depends on σ/μ

	Mean μ
Single	$E(1 - e^{-\lambda t})$
	1.39E-4
	1.94E-8
Double redundancy (with SOKC)	$E(1 - e^{-\lambda t})^2$
	5.81E-8
Double redundancy (CCF)	Note 2
	9.28E-6
Triple redundancy	$E^3(1 - e^{-\lambda t})$
	2.70E-12

Sample size for mean estimate

- CCF does not increase the needed sample size
- SOKC increases the sample size moderately
- redundancy increases the sample size greatly (no CCF)

Sample size for mean estimate

- 2 cutsets
- sample size is determined by $\frac{\sigma_1^2 + \ldots}{\ldots}$
- It can be concluded that the sample size is mainly determined by the dominant cutset.
- For new reactor with high redundancy, the dominant cutsets are usually CCF cutsets, it can be judged the needed sample size is not different than that of the old designs.

Sample size for quantile estimate

- Asympotical normal distribution is also valid, but the order statistics and approximation theory has to be used.
- A 1-α confidence interval

$$\frac{k_{1n}}{n} = p - z_{\alpha/2}\sqrt{p(\ldots}$$

- The interval length

$$C(X) = \frac{2z_{\alpha/2}\ldots}{\ldots}$$

is dependent on the desity function.

Sample size for quantile estimate

- Similarly, define the relative deviation, we can obtain the sample size needed

$$n \geq \frac{4z^{2}_{\alpha/2} \cdots}{\cdots}$$

- It depends on the local property, the sample size may be larger than that for mean estimate.

Lognormal distribution parameters	Sample size for mean estimate
mean = 1E − 5, EF = 10	170,000

Conclusions

- SOKC can significantly impact the mean, a formal uncertainty propagation is recommended when available
- The required sample size for mean value estimate may be tens of thousands, more samples are needed for quantile estimation
- The dominant cutsets also dominate the sample size

Issues still in doubt

- For quantile estimation, the probability density function estimate is difficult.
- In risk-informed applications, when significances are used, if we should use their mean values, then how many samples are necessary?

Thanks for your attention!