

System Initiating Event Frequency Estimation using Uncertain Data

Kurt G. Vedros*

NuScale Power, LLC, Corvallis, Oregon, United States

Abstract: Presented is an application of Bayesian inference methods for quantifying a prior distribution used for a system initiating event with no directly applicable system information available. New safety systems in development require the use of uncertain data and estimations where the applicability must be determined by the analyst. Presented is an approach for utilizing available, germane data that both captures the appropriateness of the contributing data and quantifies and maintains the uncertainty of the data.

Keywords: Probabilistic Risk Assessment, Parameter Estimation, Initiating Events.

1. INTRODUCTION

When a new safety system to be included in a Probabilistic Risk Assessment (PRA) is developed there is rarely a direct example with operational experience from which to draw an initiating event frequency. Additionally, there may be little to no test data on the system from which to draw for updating the estimation. Using the Bayesian inference software OpenBUGS[†] [1], the system initiating event can be modeled through its component parts with their associated failure modes to develop one distribution for use as a system initiating event prior. Similar system data can also be used for another prior distribution modeling the initiating event. I show how to incorporate all data available and how to apply proportional weights to the priors to find a weighted posterior distribution. I then show how to fit this developed distribution to a probability distribution that can be used as an initiating event frequency for PRA.

2. METHODOLOGY FOR UTILIZATION OF MULTIPLE DATA SOURCES

Sometimes there are multiple data sources available for the same initiating event. In other cases, the initiating event is unique to the plant design and it becomes necessary to compile applicable data associated with contributors to the initiating event to develop the complete system initiator. In both situations, a methodology is required to incorporate all applicable data to develop an initiating event frequency distribution for use in the PRA.

Event frequencies are reported in data sources such as NUREG/CR-5750 [2] and the NUREG/CR-6928 update [3] as a mean value with an uncertainty distribution parameter. When adding events for system initiator estimation, it is important not to lose the uncertainty information by just adding the means. Likewise, when utilizing multiple data sources for the same initiator it is important to not lose the unique uncertainty surrounding each source of information by only utilizing their means then applying some generic uncertainty to the result.

There is not a standard distribution to use when quantifying uncertainty. Several distributions can be utilized for parameterization for demand based failures or for rate based failures. Therefore, it is important to utilize a method that can mathematically manipulate different distributions. One method for this is using Bayesian inference and the software package OpenBUGS. OpenBUGS utilizes Markov Chain Monte Carlo (MCMC) to converge to a Bayesian inference solution and is practical for

* Kurt Vedros email address: kvedros@nuscalepower.com

† OpenBUGS is free for download at <http://www.openbugs.net>

use with any distribution. Prior to high speed processors and MCMC programs such as OpenBUGS, the multiple integration of the denominator in a Bayesian inference calculation made the method impractical due to the amount of time required to solve.

Instructions for running OpenBUGS, including multiple ways of checking for convergence and sample scripts that utilize these methods are found in NASA/SP-2009-569 [4], where it is referred to as its non-open source, still available predecessor, WinBUGS. The following paragraphs explain the process for adding and combining initiating events in OpenBUGS. If one does not have knowledge of OpenBUGS script, the model is defined between the “{ }” and anything behind a “#” is a comment to help in understanding what that line of code does. Note that the examples were run with 1,000 non-counted samples for convergence, followed by 200,000 samples for results.

2.1. Using OpenBUGS to Add Initiating Events

When an initiating event considered for inclusion in the plant model is a combination of failure events that are quantified separately in documentation it becomes necessary to add the events. An example of this would be where a system initiator includes a piping failure with multiple failure modes such as weld failure per weld, per year and pipe rupture per foot, per year. When adding event frequencies, it is also necessary to add the distributions. In OpenBUGS it is as easy to add dissimilar distributions as it is to add like distributions.

In BUGS language distributions are designated by the letter “d” followed by an abbreviation of the distribution used. The parameters are then given in the parenthesis immediately following. For instance a Gamma distribution is denoted by $x \sim \text{dgamma}(r, \mu)$ and read aloud as: “x is distributed as a gamma distribution with parameters r and mu”. Parameters r and mu are defined in the OpenBUGS documentation accessible within the “help” option of the program. It is recommended to review these parameterizations because OpenBUGS may not use the same parameters you may be familiar with, sometimes the parameters are manipulated differently or inverse.

A length of pipe is used as an example system. Assume that two events (pipe weld failure and pipe rupture) are independent and either event can cause the system initiating event. Adding frequency distributions within the OpenBUGS language is accomplished simply by listing the distributions then denoting another expression to add the distributions. For instance, diverse distributions (Gamma and log-normal) for two components contributing to a system frequency of a failed pipe system could be added as follows:

Script 1: Adding Distributions in OpenBUGS

```
Model {
  A ~ dgamma(r, mu)           # Failure Mode (FM) A pipe failure per length per hour
  Lambda.A <- A * length * h.y # FM A rate per year (length of pipe, hours per year)
  Lambda.B ~ dlnorm(mu.ln,tau) # FM B weld failures
  m <- welds * p.weld/h.y     # mean weld failures/yr
  mu.ln <- log(m) - pow(log(EF)/1.645,2)/2 # lognorm parms given mean & EF
  tau <- pow(log(EF)/1.645,-2)
  Lambda.Sum <- lambda.A + lambda.B # rate of system leak is sum of two FM rates
}

Data
list(r = 1.0, mu = 1.0E+10, length = 100, h.y = 8760, welds = 20, p.weld = 8.0E-4, EF = 10)
```

Note that the data used are not from any reference and are used solely for the purpose of illustrating the methodology.

Table 1: Results of addition of initiating event distributions

Parameter	Mean	5 th Percentile	Median	95 th Percentile
Lambda.A	8.8E-5	4.6E-6	6.1E-5	2.6E-4
Lambda.B	1.8E-6	6.7E-8	6.8E-7	6.9E-6
Lambda.Sum	8.9E-5	5.9E-6	6.3E-5	2.6E-4

It can be seen from the results that means can be added, but the uncertainties must be added through the distributions. Frequencies are number of occurrences over a year and while the frequency is small in this example, it is not out of the question to find a frequency above 1.0.

2.2. Combining Posteriors of Analogous Systems

When more than one data source of an initiating event frequency is appropriate for use, a methodology used to consider all inputs is the weighted average of the posterior distributions. For instance, if the system initiating event is similar in most regards to historical data from published sources, the data should be considered as input to the estimation. Another valid data point would be to piece together sub-components of the system, for instance: piping, welds, heat exchangers, etc for a unique system initiating event estimation. The analyst would then use both estimates and give a weight to them to determine the posterior when the distributions are combined and inferred.

Consider a hypothetical “System A” which is made up of piping. There is a similar, but not identical, system found in the NRC operational database NUREG/CR-5750. System A can also be modeled by adding the frequencies of its component’s initiators as above or a completely different failure mechanism can be considered. For instance, consider the following data sources for “System A”:

- Datum #1: NUREG/5750, similar but not identical system
- Datum #2: The added frequency distributions of two component failure modes that make up the system
- Datum #3: An published study covering a unique failure mode applicable to the entire system that would encompass the failures in Datum #2

All three data sources are valid and meaningful to the estimation of the system initiating event frequency. The methodology used to incorporate their appropriate influence on the final estimation is to give each datum a weight based on its importance to the system where all weights used in the analysis sum to 1.0. This is accomplished by utilizing the categorical distribution in OpenBUGS. The categorical distribution requires that its components add up to a value of 1.0. Use the number of categorical components (Lambdas in the example) equal to the number of data inputs with each index number equivalent to the index in the script. Then place a proportional value on the components equal to the confidence the analyst has in the applicability of the data to the design of the system. The result is an informed posterior distribution which keeps input from all of the data, but applies proportional emphasis on each element. Of course, if all input data are equally important, then the elements of the categorical distribution are listed as equal.

The way this looks in an OpenBUGS script is shown in Script 2:

Script 2: System Initiating Event Analysis with Weighted Posterior Averaging

```

Model {
  Lambda[1] ~ dgamma(r[1], mu[1])      # Datum 1
  A ~ dgamma(r[2], mu[2])             # Datum 2 gamma dist. frequency/ryr
  Lambda.A <- A * length * h.y
  Lambda.B ~ dlnorm(mu.ln,tau)        # Datum 2 lognorm dist frequency/ryr
  m <- welds * p.weld/h.y             # mean weld failures/yr
  mu.ln <- log(m) - pow(log(EF)/1.645,2)/2 # lognorm parm given mean & EF
  tau <- pow(log(EF)/1.645,-2)
  Lambda[2] <- Lambda.A + Lambda.B    # Datum #2
  Lambda[3] ~ dgamma(r[3], mu[3])     # Datum #3
  Lambda.avg <- Lambda[S]
  S ~ dcat(p[])                       # p=array of weights given to the data
}

Data
list(r=c(1.6, 1.0, 2.3), mu=c(365000, 1.0E+10, 170000), welds = 20, p.weld = 8.0E-4, p=c(0.3,
0.4, 0.3), EF=10, h.y = 8760, length = 100)

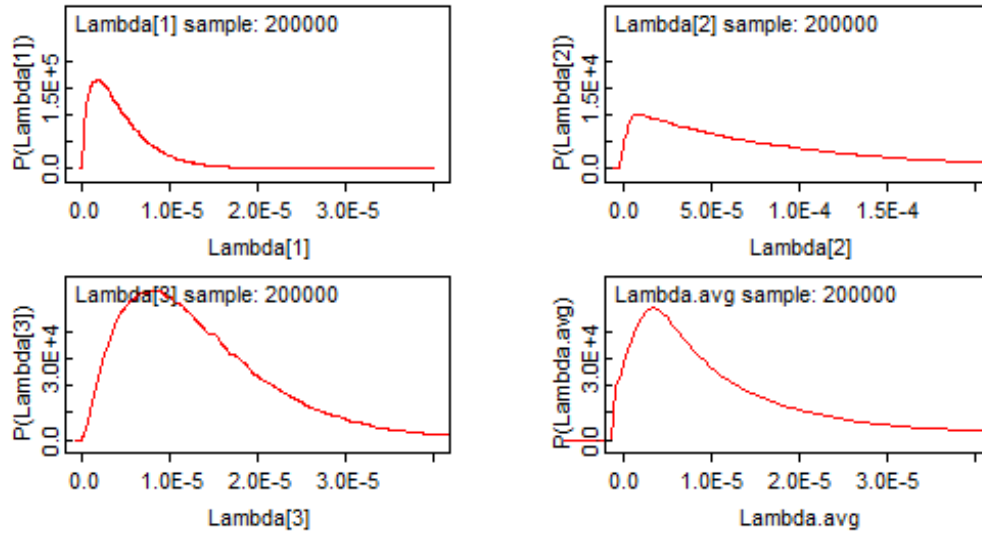
```

The input distributions are prior likelihoods and their output distribution is a posterior which is averaged with the categorical distribution using the weights given to Datum 1, 2, and 3 as 0.3, 0.4, and 0.3 respectively. Weights are subjective and left to engineering discretion. Documentation should support the weighting decisions. The result, once fit to a proper distribution, will be used for initial estimation of system performance. This estimation should be updated once actual performance of the system is gathered by using it as an informed prior in a Bayesian update. Note the resulting statistics and density graphs and how each of the prior's features can be seen influencing the posterior Lambda.avg.

Table 2: Results of initiating event analysis using posterior averaging

Parameter	Mean	5 th Percentile	Median	95 th Percentile
Lambda[1]	4.4E-6	5.6E-7	3.5E-6	1.1E-5
Lambda[2]	8.9E-5	5.9E-6	6.3E-5	2.6E-4
Lambda[3]	1.4E-5	2.8E-6	1.2E-5	3.1E-5
Lambda.avg	4.1E-5	1.2E-6	1.3E-5	1.8E-4

Figure 1: Component and posterior probability density functions



2.2. Fitting Averaged Posteriors as an Initiating Event Frequency

Whether the initiating event is calculated through the process of adding dissimilar distributions or averaging distribution posteriors, the resulting distribution is improper and not belonging to any distribution family. Visually, this can be seen by reviewing the probability density function graph for the initiating event calculated in the previous section (Lambda.avg in Figure 1). The resulting distribution is not smooth and is “pulled” in the direction of its components. In extreme cases, these posterior averaged distributions can be multi-modal and have misshapen tails. Fortunately, the posterior’s parameters generated by OpenBUGS can be used to fit a proper distribution. The log-normal distribution is a good choice for this. The log-normal distribution is typically represented by a mean and an error factor. The log-normal error factor (EF) is a measure of deviation and can be determined in a few ways, all of which are a measure of the spread of the distribution akin to variance [5] [6]:

The error factor of a log-normal distribution is defined in one of three ways:

$$\text{Log-normal Error Factor: } \sqrt{\frac{95\text{th Percentile}}{5\text{th Percentile}}} \approx \frac{95\text{th Percentile}}{\text{Median}} \approx \frac{\text{Median}}{5\text{th Percentile}} \quad (1)$$

While all three formulas are accepted for determining the EF, the shape of the distribution has an effect on the EF calculation. The three formulas for determining the EF do not provide identical results unless it is close to a normal distribution. If the distribution has a long trailing tail, the ratio of the 5th to the median will be the largest EF. If the distribution has a long preceding tail, the ratio of the 95th to the median will be largest EF. The square root of the ratio of the 95th and 5th percentiles will give a result in the middle of the other two methods and this is the method chosen for the example.

Taking the results of the example initiating event posterior above for the 5th and 95th percentile, the log-normal EF is:

$$EF_{\text{Lambda.avg}} = \sqrt{\frac{1.8E-4}{1.2E-6}} = 12.2 \quad (2)$$

Therefore, the log-normal fit parameters are the mean posterior from the analysis (Lambda.avg) of 4.1E-5 and an EF calculated of 12.2.

Insert the following lines of code into Script 2 to verify the resulting parameter fit for the log-normal distribution:

Script 3: Script Additions to Compare Log-Normal Fit

```
# insert these lines under the S~dcat(p[]) line:
LogNormFit ~ dlnorm(mu.Infit, tau.Infit)
mu.Infit <- log(mean) - pow(log(EFln)/1.645,2) / 2
tau.Infit <- pow(log(EFln)/1.645,-2)

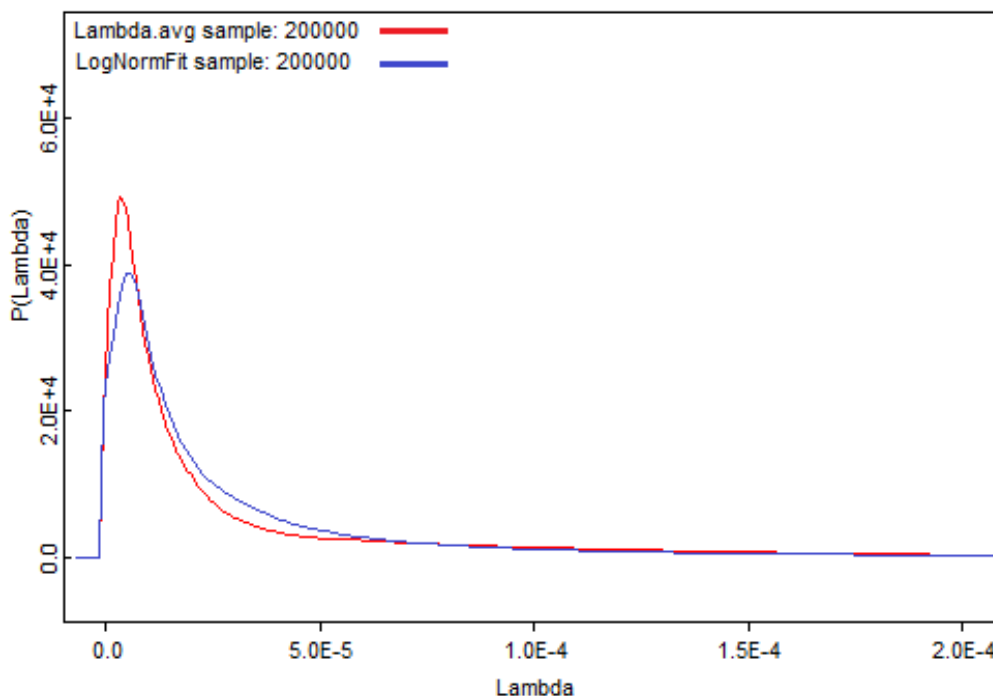
# and add to the data list:
mean = 4.1E-5, EFln = 12.2
```

A comparison of the improper posterior distribution of the analysis and the proper log-normal fit distribution is shown in Table 3 and Figure 2.

Table 3: Distribution Statistics for Resultant Distribution and Log-normal Fit

Parameter	Mean	5 th Percentile	Median	95 th Percentile
Lambda.avg	4.1E-5	1.2E-6	1.3E-5	1.8E-4
LogNormFit	4.1E-5	1.1E-6	1.3E-5	1.6E-4

Figure 2: Density comparison of log-normal fit to initiating event analysis posterior distribution



The fitted log-normal function from the results of the analysis (mean = 4.1E-5, Error Factor = 12.2) can now be used as a PRA initiating event frequency input.

3. CONCLUSION

A practical use of existing methodology using Bayesian inference software to develop a system initiating event was presented. This method illustrates a process of parameter estimation that utilizes the entire scope of the information available, preserving uncertainty and providing a mechanism to attach proper influence to each element of information. Although the exercises in the use of OpenBUGS and fitting a Log-Normal distribution from other distribution parameters are not in and of themselves new processes, the entire scheme illustrated brings together a useful technique for initiating event frequency estimation using uncertain data.

Going forward, this estimate could be used as an informed prior to update the system initiating event with operational performance in a similar manner.

References

- [1] D. Lunn, D. Spiegelhalter, A. Thomas, N. Best. “*The BUGS project: Evolution, critique, and future directions*”, *Statistics in Medicine*, volume 28, pp. 3049-3067, (2009).
- [2] U.S. Nuclear Regulatory Commission. “*Rates of Initiating Events at U.S. Nuclear Power Plants: 1987-1995*”, NUREG/CR-5750, (1999).
- [3] U.S. Nuclear Regulatory Commission. “*Initiating Event Data Sheets, 2010 Update*”, Update of NUREG/CR-6928, (2012).
- [4] H. Dezfuli, et al. “*Bayesian Inference for NASA Probabilistic Risk and Reliability Analysis*”, NASA/SP-2009-569, (2009).
- [5] D. Kelly, C. Smith. “*Bayesian Inference for Probabilistic Risk Assessment: A Practitioner’s Guidebook*”, Springer-Verlag, (2011), London.
- [6] M. Rausand, A. Høyland. “*System Reliability Theory: Models and Statistical Methods*”, John Wiley & Sons, 2nd edition, (2004), Hoboken.