INL/CON-10-18285
PREPRINT

When the Details Matter – Sensitivities in PRA Calculations that Could Affect Risk-Informed Decision-Making

PSAM-10

Dana L. Kelly

June 2010

When the Details Matter – Sensitivities in PRA Calculations That Could Affect Risk-Informed Decision-Making

Dana L. Kelly[a1]
[a]Idaho National Laboratory, Idaho Falls, USA

Abstract: As the U.S. nuclear industry continues its efforts to increase its use of risk information in decision making, the detailed, quantitative results of probabilistic risk assessment (PRA) calculations are coming under increased scrutiny. Where once analysts and users were not overly concerned with figure of merit variations that were less than an order of magnitude, now factors of two or even less can spark heated debate regarding modeling approaches and assumptions. The philosophical and policy-related aspects of this situation seem to be well-recognized by the PRA community. On the other hand, the technical implications for PRA methods and modeling have not been as widely discussed. This paper illustrates the potential numerical effects of choices as to the details of models and methods for parameter estimation with four examples: 1) issues related to component boundary and failure mode definitions; 2) the selection of the time period data for parameter estimation; 3) the selection of alternative diffuse prior distributions, including the constrained non-informative prior distribution, in Bayesian parameter estimation; and 4) the impact of uncertainty in calculations for recovery of offsite power.

Keywords: PRA, risk-informed decision-making, Bayesian inference

1. INTRODUCTION

As the U.S. nuclear industry continues its efforts to increase its use of risk information in decision making, the detailed, quantitative results of probabilistic risk assessment (PRA) calculations are coming under increased scrutiny. Where once analysts and users were not overly concerned with figure of merit variations that were less than an order of magnitude, now factors of two or even less can spark heated debate regarding modelling approaches and assumptions. The philosophical and policy-related aspects of this situation are well-recognized by the PRA community. On the other hand, the technical implications for PRA methods and modelling have not been as widely discussed.

2. EXAMPLES

This paper illustrates the potential numerical effects of choices as to the details of models and methods for parameter estimation with four examples: 1) issues related to component boundary and failure mode definitions; 2) the selection of the time period data for parameter estimation; 3) the selection of alternative diffuse prior distributions, including the constrained non-informative prior distribution, in Bayesian parameter estimation; and 4) the impact of uncertainty in calculations for recovery of offsite power.

2.1 Issues Related to Failure Mode Assignment and Component Boundary Definition

In estimating parameter values, such as failure rate or failure-to-start probability for components in a PRA model, questions often arise as to the appropriate interpretation of a recorded event. Complex and sometimes ambiguous information must be processed in order to fit the event into the taxonomy of the PRA. The analyst must decide against which component an observed failure is to be counted, and for components with multiple failure modes, he must decide to which failure mode the event is assigned. As an example of how the failure mode assignment can have an impact on inputs to risk-

[1] Dana.Kelly@inl.gov

informed decisions, consider an emergency diesel generator (EDG), and assume there are two failure modes of interest: failure to start (FTS) and failure to run (FTR).[2] Assume that over a particular period of time, there have been 38 demands to start on the EDG, and that it has operated for 75 hours over this period. Also assume there have been two events in which the EDG has failed.[3]

We will use the usual binomial and Poisson aleatory models for FTS and FTR, with unknown parameters p and λ, respectively. For prior distributions on p and λ, we will use the industry-average priors from **(1)**. For p this is a beta(0.492, 97.9) distribution, and for λ it is a gamma(0.5, 625 hr) distribution. We will take the risk metric of interest to be the difference between the posterior and prior mean of the PRA basic event, multiplied by the event's Birnbaum importance.[4] For FTR, we will assume a 24-hr mission time for the EDG, so that $P(FTR) = 1 - \exp(-24 \times \lambda)$. We assume a Birnbaum importance of 3×10^{-5}/yr for both FTS and FTR.

If both EDG failures are assigned to the FTS failure mode, then the posterior mean of p will be 0.018, and the risk metric of interest will be $(0.018 - 0.005) \times (3 \times 10^{-5}/\text{yr}) = 4.0 \times 10^{-7}/\text{yr}$. If both failures are assigned to the FTR mode the posterior mean of λ will be 0.004/hr and the risk metric will be $[(0.004 - 0.0008)/\text{hr} \times 24\ \text{hr}] \times (3 \times 10^{-5}/\text{yr}) = 1.9 \times 10^{-6}/\text{yr}$. If a risk metric value of 10^{-6}/yr constitutes a decision threshold, then the failure mode assignment in this example could influence a risk-informed decision using this metric.

Ref. **(2)** reviewed U.S. industry data contained in the EPIX database **(3)**, and found that the most common problem with the data was failure mode assignment. For EDGs, there were 104 FTS events over the review period, and **(2)** agreed with this assignment for only 64 of these events. It determined that 17 of the events should have been classified as fail to load/run, 4 should have been FTR, and 19 should have been classified as test or maintenance unavailability. Thus, the possibility that failure mode assignment could affect the risk significance of an event appears to be a real one.

A similar issue can exist with respect to the definition and treatment of component boundaries. Consider the EDG fuel oil transfer pump (FOTP), which transfers fuel oil from the fuel oil storage tank (large capacity) to the smaller capacity EDG day tank to support EDG operation over the 24-hr mission time. Assume that failure of the transfer pump will cause the EDG to fail to run.[5] We will assume that the prior distribution for the FOTP pump failing to start is a beta(0.498, 498) distribution, taken from **(1)**.[6] Assume that during a 24-hr run of the EDG, there are 6 demands on the FOTP to start, and that there have been 50 demands on the FOTP during the time period over which data have been collected.

We will assume that failure of the FOTP causes EDG FTR, so we will use the EDG Birnbaum value of 3×10^{-5}/yr from above. The contribution to EDG FTR from FOTP failure will be the probability of a failure in 6 demands over the 24-hr EDG mission time, or $1 - (1 - p_{FOTP})^6$. If the FOTP were considered to be within the EDG component boundary, then 2 failures of the FOTP would be treated as 2 EDG FTR failures, and the risk metric would be the value of 1.9×10^{-6}/yr calculated above. However, if the FOTP were modeled separately, the posterior mean of p_{FOTP} would be 0.0045, and the risk metric would be 6.3×10^{-7}/yr. Again, if a risk metric value of 10^{-6}/yr constituted a threshold, as it does in the U.S. Nuclear Regulatory Commission's Mitigating System Performance Index (MSPI), the component boundary classification could affect the risk significance.

---

[2] In a typical PRA today there are three failure modes: failure to start, failure to load, and failure to run. This additional complication is ignored for clarity of presentation, but it does not affect the point to be made.
[3] These data are representative values for a three-year period, with monthly EDG surveillance tests.
[4] This is an estimate of the change in an overall risk metric, such as core damage frequency.
[5] This is the case at a minority of U.S. plants; most U.S. plants have multiple fuel oil transfer pumps, such that a single pump failure will not fail the EDG.
[6] For simplicity, we ignore failure to run of the FOTP.

2.2. Selection of Time Period for Parameter Estimation

Ref. (4) introduced a so-called baseline approach to selecting the time period over which to estimate an initiating event frequency. As (4) states on p. 44, "The goal was to choose a baseline period…with the most constant performance." Operationally, this was done via statistical hypothesis testing, where the null hypothesis was a constant initiating event frequency, and the alternative hypothesis was a loglinear time trend: $\log \lambda = a + bt$. The parameter b in the loglinear model determines whether the trend is an increasing ($b > 0$) or decreasing one ($b < 0$). If $b = 0$, the loglinear model simplifies to a model with constant frequency (no trend). Restating the hypotheses in terms of the parameters of the loglinear model, we have:

H_o: $b = 0$
H_1: $b \neq 0$

Ref. (4) selected the baseline period as follows (cf. p. 44). The data period considered was 1988-2002. If there was one event or less during that period, the entire period was used. If there were only two events, and they occurred during the first three years, then these years were discarded and the baseline period started in the first year with no events. Except for these cases, *the baseline period was the one in which the p-value against a trend (i.e., against H_1) was maximized* [emphasis added].[7]

Note that this approach, although it may be intuitively appealing, can discard data based on statistically *insignificant* trends[8], and this can impact frequency estimates in a way that tends to be nonconservative when the underlying frequency giving rise to the data is relatively small, meaning initiating events are relatively infrequent. Put another way, when there is no underlying trend in the frequency (i.e., H_o is true), the baselining approach of (4) will err and discard data a significant portion of the time. How often? It is well known that the p-value against a simple null hypothesis, as we have here, has an asymptotic uniform distribution over the interval (0, 1). The baselining approach will discard the first year's data if the p-value with the first year removed is higher than the p-value considering the entire time period. Asymptotically, the probability that this happens, given that H_o is true (the frequency is constant) is 0.5. So, about half the time, at least one year of data is discarded, even though there is no time trend present. This is easily borne out by simulation.

As a concrete illustration, consider the following example data, simulated from a Poisson distribution with constant frequency of 0.025/yr. The exposure times are shutdown-reactor-years, and are taken from data from the NRC's public website for a particular U. S. plant.

---

[7] Recall that the frequentist p-value is the probability of observing data at least as extreme as those that were observed, given that H_o is true. Thus, smaller p-values correspond to more evidence for the presence of a loglinear trend.

[8] A widely used criterion for statistical significance, which has been adopted in past NRC work, is p-value < 0.05.

Table 1: Simulated event counts from Poisson distribution with constant frequency (0.025/yr)

Year	Number of Events	Exposure Time (yr)
1987	1	28.11
1988	1	27.85
1989	2	30.62
1990	0	27.23
1991	0	24.73
1992	0	24.37
1993	1	24.01
1994	0	21.2
1995	1	18.28
1996	0	21.97
1997	0	26.47
1998	0	18.6
1999	0	12.27
2000	0	10.11
2001	0	9.04
2002	0	8.12
2003	0	10.39
2004	0	8.07
2005	0	9.09
2006	0	8.659
2007	0	4.821

As described above, the baselining approach to determining the time period to use in estimating the Poisson frequency fits a loglinear trend model to the entire dataset, and then removes one year at a time, starting in 1987, until the p-value against the loglinear trend is maximized.[9] Note that the starting p-value does not need to be < 0.05, the usual criterion for identifying a statistically significant trend. Let us apply this approach to the datasets above. The R open-source statistical package was used to do this (5). The results are shown in Table 2. The p-value against the loglinear trend is maximized with the data from 1987-1989 removed.

Table 2: Baselining method for data in Table 1, discarding data for 1987-1989

Starting Year	Slope (*b*)	Std. Error	p-value
1987	-0.217	0.1235	0.0788
1988	-0.2428	0.1475	0.0998
1989	-0.2651	0.1770	0.134
1990	-0.1247	0.1882	**0.508**
1991	-0.2028	0.2314	0.381

With 4 of 6 failures discarded (along with the associated exposure times), the posterior mean frequency obtained by updating the Jeffreys prior drops from 0.02/yr to 0.009/yr, about a factor of 2. Recall that the data in Table 1 are simulated counts from a Poisson distribution with constant frequency of 0.025/ yr.

[9] "Remove" implies removing both the event counts and exposure time. Removing event counts is the critical item due to the (relatively) small number of events compared to the total exposure time. However, shortening the exposure time will increase the uncertainty in the resulting estimate.

2.3. Choice of Diffuse Prior Distribution

A salient feature of Bayesian inference is its ability to incorporate information from a variety of sources into the inference model, via the prior distribution. However, there is an often-voiced criticism of subjectivity in the choice of prior, which analysts sometimes attempt to avoid by choosing prior distributions which are in some sense minimally informative, that is, they are diffuse over the region where the likelihood function is non-negligible, but incorporate some information about the parameters being estimated, such as a mean value. The most widely used minimally informative prior in the U.S., the so-called constrained noninformative (CNI) prior, was put forth in (6).

The CNI prior is formulated as a conjugate distribution for the most commonly encountered aleatory models in PRA, making it mathematically convenient; but it has a relatively light tail, as do conjugate prior generally, as noted in (7) for example, and the posterior distribution can be somewhat under-influenced by updates with sparse data, an issue discussed by (8) in the context of the NRC's Mitigating System Performance Index. Some recent work at Idaho National Laboratory described in (9) (this conference) presents three alternatives to the CNI prior, which produce posterior distributions that are influenced more by sparse data than is the CNI prior, especially for highly reliable components (small failure probability or failure rate). The example presented in (9) is failure to start of a motor-driven pump, assumed to have a mean failure probability of 0.001. We will assume there are 50 demands on this pump, and that failures to start are described by a binomial distribution with parameters p and 50. From (6), the beta distribution that approximates the CNI prior has parameters $\alpha = .498$ and $\beta = 498$. Therefore, the posterior distribution for p is beta($\alpha + x$, $\beta + 50 - x$), where x is the number of failures observed in 50 demands.

One of the alternatives explored by (9) is the logistic-normal distribution, which can be viewed as a lognormal distribution constrained to lie in the interval [0, 1]. Historically, the lognormal distribution has been used widely in PRA to represent parameter uncertainty, but the fact that it is not conjugate to common aleatory models in PRA has been a hindrance to its use as a prior distribution in Bayesian inference. The logistic-normal distribution is also nonconjugate to the binomial likelihood, so Bayesian inference must be done numerically, a simple matter with today's tools. The R package was used to find the posterior means (5).

Ref. (9) chose the parameters of the logistic-normal distribution so that it would have the same mean and 95th percentile as the CNI prior. This choice preserves the mean value, which is often taken as a point estimate of an overall industry average value. The 95th percentile is a commonly used upper percentile value in PRA. Thus, both distributions will have 5% area above the 95th percentile, but the distribution of this 5% will be different. Figure 1 shows the posterior means for p, for 1 to 4 failures in 50 demands. As this figure shows, the difference caused by the choice of diffuse prior can be significant. In applications such as MSPI, which estimate an increase in p above a baseline value, the difference could lead to a change in risk classification.

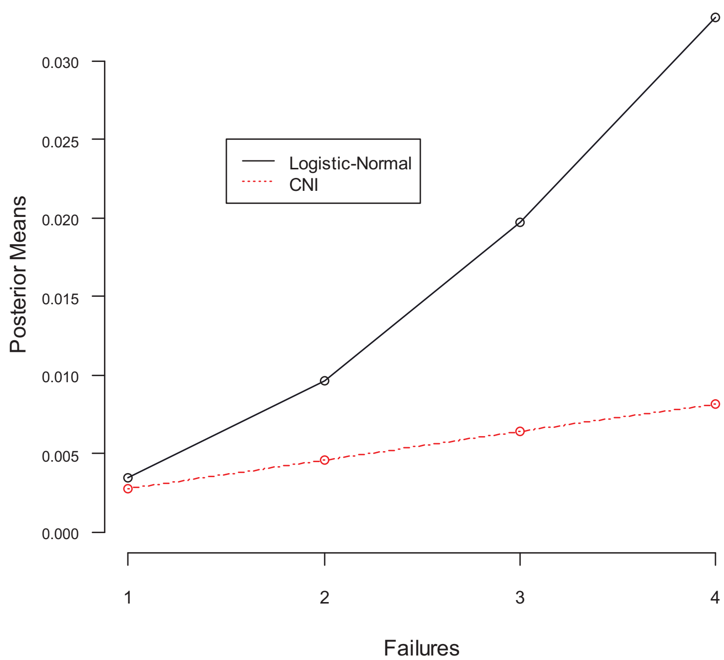

Figure 1 Posterior means for *p* for 1-4 failures in 50 demands, comparing CNI and logistic-normal prior, both with same mean and 95th percentile

2.4 Impact of Uncertainties in Offsite Power Recovery Calculations

Uncertainty, both aleatory and epistemic, can have a significant impact on estimated probabilities of recovering from loss of offsite power (LOSP) within a specified time window, and such probabilities are an input to risk-informed decisions as to the significance of inspection findings in the U.S. Nuclear Regulatory Commission's Reactor Oversight Process.

Ref. (10) (this conference) presents an example in which three alternative aleatory models are used for a sample of recovery times, and epistemic parameter uncertainties are propagated through the models in a fully Bayesian analysis using Markov chain Monte Carlo (MCMC) sampling, as described in (11). The probability of not recovering offsite power within a representative time period (8 hr) is estimated, including epistemic uncertainty, for each of the aleatory models. A comparison is also made to a typical point estimate calculation that uses the maximum likelihood parameter estimates for each aleatory model, without considering epistemic uncertainty. Table 3 summarizes the results of these calculations. The three aleatory models for recovery time give significantly varying mean nonrecovery probabilities at 8 hours when parameter uncertainty is properly accounted for in a fully Bayesian analysis. Note that using the maximum likelihood parameter estimates gives point estimate nonrecovery probabilities at 8 hours that are significantly less than the posterior means from the fully Bayesian analysis.

Table 3: Comparison of point estimate nonrecovery probabilities at 8 hours with posterior mean values from fully Bayesian analysis

Model	Point estimate nonrecovery probability at 8 hours using MLEs	Posterior mean nonrecovery probability at 8 hours	Ratio of posterior mean to point estimate
Exponential	6.7E-4	2.8E-3	4.2
Weibull	2.8E-3	0.012	4.3
Lognormal	8.8E-3	0.023	2.6

3. CONCLUSIONS

This paper has presented four examples to illustrate how the quantitative results of PRA can be relatively sensitive to details associated with model and data selection. In the past, before these inputs came to be used as inputs to risk-informed decisions, PRA analysts often tolerated considerable variation in the results, with variation less than about an order of magnitude typically not being an issue for concern. Today, however, variations of a factor of two or even less can determine on which side of a numerical threshold a PRA result lies. The potential for such sensitivity needs to be taken into consideration by the analyst (12), whose responsibility it is to communicate such sensitivities to the decision-maker, who can then take them into account in a risk-informed process.

With regard to the baselining approach for selecting a time period for parameter estimation, it is not the author's impression that this has led to widespread underestimation of initiating event frequencies. However, it certainly has the potential to underestimate frequencies, as illustrated by the example herein. Thus, there is also the concern that this approach could be adopted for PRA parameters beyond initiating event frequencies, such as equipment failure rates, where it has the same potential to underestimate parameter values, particularly for highly reliable equipment with sparse failure data.

Setting aside theoretical arguments, it is the author's opinion that a practical reason for the widespread adoption of the CNI prior distribution in the U.S. is that, because the CNI prior is a conjugate distribution to the binomial and Poisson aleatory models, it is a mathematically convenient choice that simplifies the Bayesian inference calculations. However, because it is conjugate, and thus has a relatively light tail, it can overly influence the posterior distribution (particularly summary measures such as the posterior mean) in updates with sparse data. Other diffuse priors have heavier tails and thus give more robust results, that is, the posterior distribution is influenced more by the observed data. These other priors are typically nonconjugate, and so Bayesian inference must be done numerically. However, modern software tools, some of which are freely available and open-source, make Bayesian inference with nonconjugate priors straightforward, and so mathematical convenience no longer needs to be a principal attribute influencing the choice of prior distribution. Perhaps there is a need to make the PRA community more aware as to the availability and use of such tools.

These tools also enable sensitivity calculations with alternative aleatory models, and accommodate fully Bayesian treatment of the epistemic uncertainty associated with the parameters in these models. Without these tools, the analyst's choice is confined to what may be overly simple aleatory models (e.g., exponential model for LOSP recovery time), and Bayesian inference with conjugate priors. If more complex aleatory models, such as a lognormal distribution are used, the analyst has to resort to frequentist inference with either no treatment of parameter uncertainties, or an approximate one at best. Modern Bayesian tools enable the analysis of quite complex aleatory models, without having to rely upon conjugate prior distributions. These tools also support Bayesian model-checking, as illustrated in (10) and (11).

It is the impression of the author that not very many PRA analysts are aware of the kinds of sensitivities described herein, or are trained in methods for addressing them. If this is the case, then it is relatively straightforward to make improvements via education and training. However, an unexplored possibility is that some numerical decision criteria are low enough that “noise” in the PRA outputs limits their usefulness. This may become an even more serious problem in the future, as enhanced safety features in new reactor designs strain the ability of current PRA models and data to provide accurate estimates of increasingly small risks.

Acknowledgements

The author would like to thank Nathan Siu of the U.S. Nuclear Regulatory Commission for his helpful comments on this paper. However, the thoughts in the paper are the author’s alone and do not represent the position of any U.S. government agency.

This manuscript has been authored by Battelle Energy Alliance, LLC under Contract No. DE-AC07-05ID14517 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References

1. Nuclear Energy Institute. *Regulatory Assessment Performance Indicator Guideline.* Washington, DC : Nuclear Energy Institute, 2007. NEI-99-02, Rev. 5.
2. Wierman, Thomas E., et al. *Characterization of EPIX Data Quality.* Idaho National Laboratory. 2009. Internal report. Published as INL/EXT report?.
3. Institute for Nuclear Power Operations. *Equipment Performance and Information Exchange System (EPIX): Reporting Requirements.* 2007. INPO 98-001, Rev. 5.
4. Eide, Steven L., et al. *Industry-Average Performance for Components and Initiating Events at U. S. Commercial Nuclear Power Plants.* U. S. Nuclear Regulatory Commission, 2007. NUREG/CR-6928.
5. R Development Core Team. *R: A language and environment for statistical computing.* Vienna, Austria : R Foundation for Statistical Computing, 2009.
6. Atwood, Corwin L. "Constrained Noninformative Priors in Risk Assessment." 1996, *Reliability Engineering and System Safety*, Vol. 53, pp. 37-46.
7. Berger, James O. *Statistical Decision Theory and Bayesian Analysis, Second Edition.* Springer, 1985.
8. Dube, D. A., et al. *Independent Verification of the Mitigating Systems Performance Index (MSPI) Results for the Pilot Plants.* U. S. Nuclear Regulatory Commission. Washington, D.C., 2005. NUREG. NUREG-1816.
9. Kelly, Dana L., Youngblood, Robert W. and Vedros, Kurt. "Minimally Informative Prior Distributions for PSA." Seattle, 2010. International Probabilistic Safety Assessment and Management.
10. Kelly, Dana L. "Impact of Uncertainty on Calculations for Recovery from Loss of Offsite Power." Seattle, 2010. International Probabilistic Safety Assessment and Management 10.
11. Kelly, Dana L. and Smith, Curtis L. "Bayesian Inference in Probabilistic Risk Assessment - The Current State of the Art." 2009, *Reliability Engineering and System Safety*, Vol. 94, pp. 628-643.
12. Drouin, M., et al. *Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decision Making.* Washington, DC : U.S. Nuclear Regulatory Commission, 2009. NUREG-1855.