Uncertainty Analysis and Pressurized Thermal Shock: An Opinion

N. Siu

White Paper Last Revised September 3, 1999

Introduction

To support current efforts regarding pressurized thermal shock (PTS) screening criteria in a manner consistent with NRC's current views on risk-informed decision making, probabilistic risk assessment (PRA) analysts need to: a) develop estimates of risk metrics such as core damage frequency (CDF) and large early release frequency (LERF), and b) characterize the uncertainties in these estimates. Typically, this characterization is in the form of a probability distribution (see Figure 1, where λ represents the frequency of interest and $\pi(\lambda)$ is the probability density function for that frequency). But what does this distribution mean? What uncertainties does it represent? Aren't CDF and LERF already measures of uncertainty? And how do we develop the CDF and LERF distributions for PTS?

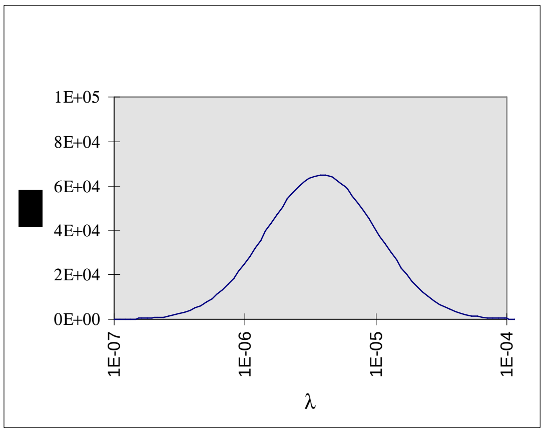


Figure 1 - Example Output of a PRA

This white paper answers these questions in two steps. First, it addresses the issues of uncertainty in a methodologically oriented discussion. This includes a definition of the two "types" of uncertainties currently distinguished in PRA, and a discussion of how they are treated. Second, based on this methodological discussion, it proposes an approach for addressing uncertainties in PTS; this approach integrates thermal hydraulic (T/H) and probabilistic fracture mechanics (PFM) analyses in a PRA framework. The proposed approach is then shown to be nearly identical with the current ("Method 2") PTS approach. Differences between the two approaches and their implications for PTS analysis are discussed.

It is recognized that, despite the agreement between the proposed approach and the current PTS approach, a number of details may need to be revised following input from domain experts; the intent of this paper is to provide an initial approach to the problem that is consistent with current PRA views on the treatment of uncertainty.

This paper also includes a list of references for further reading and three appendices covering probability concepts, aleatory and epistemic uncertainties, and parameter estimation.

Uncertainty Analysis Concepts

On the Meaning of "Frequency"

Although the analyses of CDF and LERF require the treatment of very different physical phenomena, they are, from a mathematical viewpoint, both frequencies of undesired events. This section discusses the notion of frequency *as it is typically used in PRA models*. It is shown that, in PRA, the frequency is a parameter in a probability distribution that quantifies random variability ("aleatory uncertainty") in an observable variable.

Let's start with some basic assertions that provide the foundation for subsequent discussion.

1. There are physical variables which are, in principle, observable. Examples include the time to failure of a particular component, the time at which an operator takes a particular action at a given point in an accident sequence, the average copper content in a particular subregion of a particular reactor vessel at a particular point in time.

2. We need to predict the values of a set of these variables as part of the PRA analysis.

3. Because of limitations in resources, lack of knowledge, or both, we choose to treat some of these variables as being the results of random processes. In other words, if we employ a thought experiment involving a number of repeatable trials, we envision observing a distribution of values (e.g., an empirical histogram) for the variable of interest. The "prediction," therefore, will be in terms of a probability distribution.

4. We also choose to treat the remaining variables as being deterministic. If we employ a thought experiment involving a number of repeatable trials, we envision observing a single value for the variable of interest (or, at least, a range of variability that is sufficiently small for the practical application). The prediction, therefore, will be in terms of a point value, at least in principle.

Note that because choice is involved, there is no fundamental principle as to when a variable should be modeled as being random or deterministic; the analyst needs to decide if the notion of repeatable trials makes sense for the problem being addressed. In PRAs, such things as pump failures and operator actions are modeled as being random; we treat pumps and operators as coming from populations of pumps and operators, and don't attempt to model individual pumps or individual operators. (One can argue that, even in the case of individual pumps and operators, the notion of random variability still makes sense due to such processes as environmental variation and renewal.) In the case of a reactor vessel, the choice may be less clear. A proposed approach is discussed later in this paper.

Note also that, in current PRAs, core damage events and large early release events are modeled as being the possible results of a set of interacting random processes, namely, those involving the initiating event that causes a plant transient, the response of mitigating systems to the transient, and the associated actions of human operators. The occurrences of core damage and large early release events are also, therefore, random processes.

For random events occurring over time, PRAs typically use a Poisson distribution to model event occurrence. This means that the probability of observing N core damage events in a time period T is given by:

$$P\{N \text{ events in time } T|\lambda\} = \frac{(\lambda \cdot T)^N}{N!} \cdot e^{-\lambda \cdot T} \qquad (1)$$

where λ, which is called a "frequency," is a simply a parameter characterizing the process. As λ increases, the likelihood of events also increases (see Figure 2). It can be shown that the average number of events occurring in time period T is equal to λT.

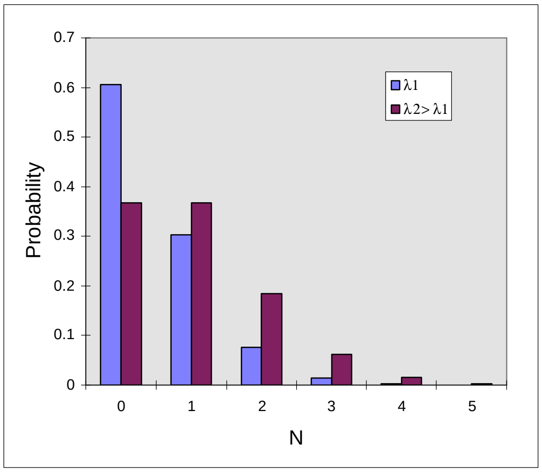


Figure 2 - Poisson Probability Distributions for Two Values of λ

It turns out that for a Poisson process, if T_1 is the time to the first event, then the distribution of T_1 is exponential, i.e.,

$$P\{T_1 < t|\lambda\} = 1 - e^{-\lambda \cdot t} \qquad (2)$$

As λ increases, the probability of observing the first event by a specified time also increases (see Figure 3). It can be shown that the average time to the first event is equal to $1/\lambda$. It can also be shown that

$$P\{T_1 < t|\lambda\} \approx \lambda \cdot t \quad \text{when } \lambda \cdot t < 0.1 \qquad (3)$$

As noted earlier, CDF and LERF are the frequencies of core damage events and large early release events, respectively. Thus, they are simply parameters of Poisson distributions. Knowing the values of CDF and LERF, we can make statements about the likelihood of observing a core damage event or a large early release event in, say, the next year. Of course, we don't know the values of CDF and LERF with a high degree of certainty. This issue is discussed in the following section.

Before concluding this discussion, it should be noted that the Poisson model, like all models, has some underlying assumptions. In particular, the Poisson model assumes that the process doesn't age, i.e., that λ does not change over time. In the case of CDF and LERF, this can be an unrealistic assumption. For example, if a severe accident really does occur, we can expect there to be significant changes in the industry (e.g., all plants might be shut down). Less dramatically, aging considerations might become important over time. For most PRA purposes, the Poisson model is adequate.

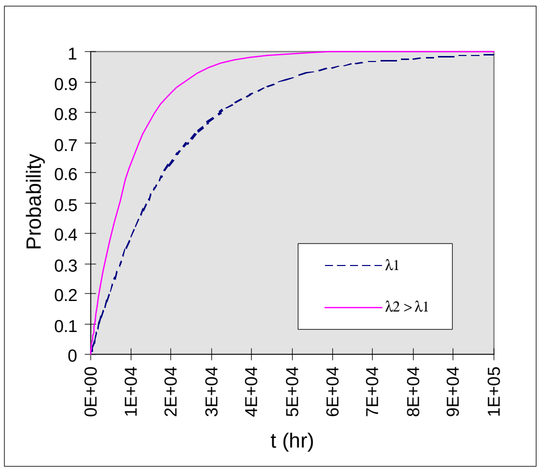


Figure 3 - Effect of Frequency on Time-to-Occurrence

Types of Uncertainties: Aleatory and Epistemic

The preceding discussion addresses uncertainties due to "inherent randomness". In earlier literature, they are often called "random uncertainties" or "stochastic uncertainties." Currently, following the terminology espoused by the ACRS, they are called "aleatory uncertainties."[1] Their principal characteristic is that they are (or are modeled as being) irreducible; they are defined by the form of the probability distribution (e.g., the Poisson distribution) and the value of the distribution parameters (e.g., λ).

Note that in the examples given earlier, the variability in the uncertain variable (e.g., N or T_1) is observable, at least in principle. In other words, repeated observations of the variable will result in an empirical distribution of values. This provides a way to think about aleatory uncertainties; if repeated trials of an idealized thought experiment (where the conditions are kept constant from trial to trial) will, assuming no measurement error, lead to a distribution of outcomes for the variable, this distribution is a measure of the aleatory uncertainties in the variable.

[1]According to Webster's, *aleatory* (adj.) comes from *alia* (a dice game); relevant definitions are: (1) depending on an uncertain event; (2) relating to good or bad luck.

Another type of uncertainty addressed in PRAs is “epistemic uncertainty,”[2] which has been called “state of knowledge uncertainty” in earlier papers because it is due to weaknesses in the current state of knowledge of the assessor. Uncertainties in a deterministic variable whose true value is unknown are epistemic. Repeated trials of a thought experiment involving the variable will, in principle, result in a single outcome, the true value of the variable.[3]

Unlike aleatory uncertainty, epistemic uncertainty is reducible with the collection of additional information. In PRAs, for example, it is typically assumed that the Poisson model is a good representation for the failure of equipment while running. Therefore, it is assumed that there is a particular failure rate for each component. Initially, we may not have much failure data for a component, and our (epistemic) uncertainties in the value of the failure rate will be large. After we collect a large enough sample of failure data, we can get a very good estimate of the failure rate, i.e., the epistemic uncertainties in the value of the failure rate will be small. The epistemic uncertainties are quantified using probability distributions (see Appendix A). Figure 4 shows how, in instance, the distributions are narrowed, i.e., the uncertainties are reduced, when additional information is collected. (N represents the number of observed failures and T represents the period of observation in hours.) The method for generating these distributions, given data, is discussed in the next section.

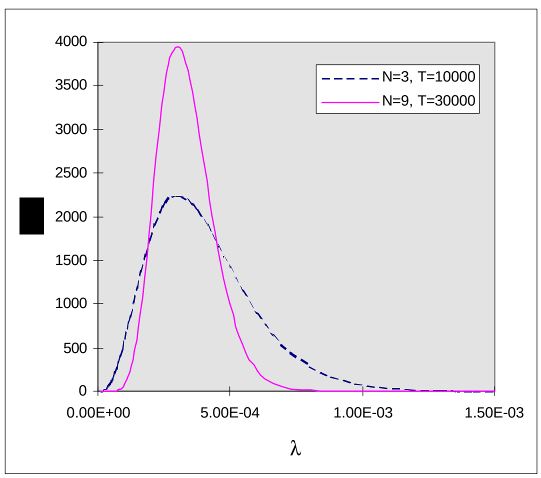


Figure 4 - Reduction in Epistemic Uncertainty with Increased Data

The answers to the first three questions posed at the beginning of this paper are therefore as follows. (1) The distribution in Figure 1 quantifies the analyst’s uncertainties in the value of the

---

[2]According to Webster’s, *epistemic* (adj.) comes from *epistemikos* (of knowledge, capable of knowledge); relevant definitions are: (1) of, having the character of, or relating to intellectually certain knowledge; (2) purely intellectual or cognitive; (3) subjective.

[3]Note that measurement error arises from an aleatory process. However, if the measured variable is, in principle, deterministic, then the uncertainties in the variable are epistemic. The apparent contradiction can be resolved by clearly defining what uncertainties are being addressed in the PRA. This issue is further discussed in Appendix A.

parameter λ (which represents either CDF or LERF). Specifically, the integral of the curve (which is a probability density function) between any two limits, say λ_1 and λ_2, gives the probability that λ lies in the range (λ_1,λ_2). (2) These uncertainties are epistemic; they arise from the analyst's imperfect state of knowledge regarding the true value of λ. (3) CDF and LERF (which are typically computed in PRAs using conventional event tree/fault tree analysis) are frequencies (as defined earlier in this paper); they are parameters that quantify aleatory uncertainties in observable variables, e.g., the time to a core damage event. There are, of course, generally epistemic uncertainties in their values.

Figure 5 shows how these two types of uncertainty can be represented in the case of such variables as event occurrence times. (An analogous representation can be developed for variables representing the number of events in a given time period.) The heavy curves (solid and dashed) are the cumulative probability distributions quantifying the aleatory uncertainties in the event occurrence time. The light curve crossing these heavy curves is the probability density function quantifying the epistemic uncertainties in λ; it represents the same distribution as that illustrated in Figure 1. As shown by Equation (2), the aleatory distributions are conditioned on the value of λ; the four curves shown correspond to the 5^{th} percentile (λ05), median (λ50), mean (<λ>), and 95^{th} percentile (λ95) values of λ. Note that PRAs typically display results in the form of Figure 1 and not Figure 5; the aleatory uncertainties in the observable variable are assumed to be understood.

It should also be noted that fundamentally, as discussed by a number of authors (e.g., see Apostolakis, 1999) and noted in Appendix A, there is only one kind of uncertainty. Why does PRA distinguish between "aleatory" and "epistemic" uncertainties? The answer is due to the fact that PRA is used to support decision making; the distinction can be important for both interpreting the PRA output, and deciding what to do with this output. This is discussed in Appendix B.

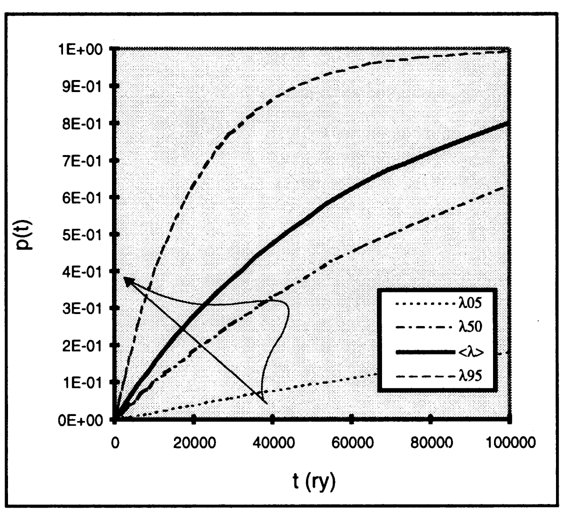


Figure 5 - Representation of Aleatory and Epistemic Uncertainties in Event Occurrence Time

Uncertainty Analysis in PRA

Current PRAs typically use two kinds of models to address aleatory uncertainties. The first, which is applied to events occurring over time (e.g., failures of already operating pumps), is the Poisson distribution already discussed. The second, which is applied to events occurring as the immediate consequence of a challenge (e.g., failures of standby pumps to start on demand), is the binomial distribution. This distribution quantifies the likelihood of outcomes resulting from a Bernoulli (or "coin flip") process. It is given by:

$$P\{R \text{ failures in N demands}|\phi\} = \frac{N!}{R!(N-R)!}\phi^R(1-\phi)^{N-R} \tag{4}$$

where ϕ is the probability of failure for a single demand. It can be seen that mathematically, ϕ plays the same role as λ; it is just a parameter characterizing a distribution. It can be shown that as the number of trials gets very large, the relative frequency of failures, R/N, approaches ϕ. Thus, ϕ can be interpreted as the fraction of times failures will occur in the long run.

Using the various λ's and ϕ's corresponding to the different components included in the PRA model, the CDFs and LERFs associated with various event sequences, as well as the overall CDF and LERF, can be computed. Symbolically,

$$\begin{aligned} CDF &= f_1(\underline{\lambda}, \underline{\phi}) \\ LERF &= f_2(\underline{\lambda}, \underline{\phi}) \end{aligned} \tag{5}$$

To quantify the epistemic uncertainties in CDF and LERF, the epistemic uncertainties in the λ's and ϕ's are propagated through f_1 and f_2. This is currently done on a routine basis using sampling schemes (e.g., direct Monte Carlo sampling).

The quantification of the uncertainties in the λ's and ϕ's involves the collection and interpretation of a variety of forms of evidence (e.g., model predictions, expert opinion, empirical data), and the application of an appropriate estimation procedure that uses this evidence. Formally, the estimation procedure involves the application of Bayes' Theorem. The general form of this theorem is:

$$\pi_1(\underline{\theta}|E) = \frac{L(E|\underline{\theta})\,\pi_0(\underline{\theta})}{\int_{\underline{\Theta}} L(E|\underline{\theta})\,\pi_0(\underline{\theta})d\underline{\theta}} \tag{6}$$

where $\underline{\theta}$ is the vector of parameters to be estimated; E is the evidence; $L(E|\underline{\theta})$ is the likelihood function, i.e., the probability of observing the evidence if is known; $\pi_0(\underline{\theta})$ is the prior distribution for $\underline{\theta}$, i.e., the probability distribution for $\underline{\theta}$ prior to observing the evidence; and the denominator on the right hand side of the equation is just a normalization constant.

While it may appear to be complicated, application of Equation (6) is straightforward in many practical cases. Consider the situation where we are estimating the failure rate (frequency) of a component, λ, and the evidence consists of an observation of R failures in a specified time interval T. The likelihood function is then the Poisson distribution as given by Equation (1); removing constants that appear in the numerator and denominator, Bayes' Theorem becomes:

$$\pi_1(\lambda|R,T) = \frac{\lambda^R e^{-\lambda\cdot T}\pi_0(\lambda)}{\int_0^\infty \lambda^R e^{-\lambda\cdot T}\pi_0(\lambda)d\lambda} \tag{7}$$

which has analytical solutions for some forms of the prior distribution, and which can be solved numerically using simple tools (e.g., spreadsheets or equation solving software) for arbitrary forms of the prior distribution. (The development of the prior distribution requires judgment, especially in the case where the data are sparse. Practical approaches are discussed in the paper by Siu and Kelly which is included in the list of references at the end of this paper. It is worth noting that for reasonable prior distributions, the precise shape of the distribution is unimportant when large amounts of data are available.)

It is important to observe that the likelihood function represents the aleatory model for the observable variable. In the above case, the observable variable (R), is assumed to be the result of a Poisson process; the Poisson distribution (which has the single parameter λ) is then appropriate for the likelihood function. To expand on this point, consider a slightly more complicated case where the observable variable, denoted by C, is assumed to be: a) random, and b) the result of a lognormal process, i.e., the aleatory uncertainties in C are quantified by a lognormal distribution. Assume an experiment is performed which results in N observations of C. Bayes' Theorem is then

$$\pi_1(\mu,\sigma|C_1,\ldots,C_N) = \frac{\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}\sigma C_i}\exp\left\{-\frac{1}{2}\left[\frac{\ell n C_i-\mu}{\sigma}\right]^2\right\}\pi_0(\mu,\sigma)}{\int_{-\infty}^{\infty}\int_0^{\infty}\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}\sigma C_i}\exp\left\{-\frac{1}{2}\left[\frac{\ell n C_i-\mu}{\sigma}\right]^2\right\}\pi_0(\mu,\sigma)d\sigma d\mu} \tag{8}$$

where μ and σ are the two parameters of the lognormal distribution and are related to the mean and variance of C. This equation can be solved using relatively simple software tools. An example is provided in Appendix C.

When the evidence is in more complicated forms (e.g., expert opinions), the use of Bayes' Theorem is not as straightforward. In such cases, current PRAs generally employ less formal procedures, e.g., subjective estimation of the probability distribution based on considerations of sample averages and ranges. Bayes' Theorem is an important tool for ensuring that the analyst updates his/her state of knowledge concerning the uncertain parameter in a manner consistent with the laws of probability, but it is just a tool.

Summary Points - Uncertainty Concepts

- Uncertainties in a variable are treated in PRAs as being aleatory when the variable is assumed to be the result of a random process, i.e., repeated trials of a thought experiment will lead to a distribution of values for the variable.
- Uncertainties in a variable are treated in PRAs as being epistemic when the variable is assumed to be deterministic, i.e., repeated trials of a thought experiment will lead to a single value for the variable.
- The distinction between aleatory and epistemic uncertainties is not always clear; drawing the line between the two is generally a modeling decision.
- PRAs generally address aleatory uncertainties in the behavior of model elements through the λ and ϕ parameters. The aleatory uncertainties in overall plant behavior are addressed using the CDF and LERF parameters; these are functions of the λ's and ϕ's.
- The epistemic uncertainties in the λ's and ϕ's are propagated through the PRA model to develop epistemic distributions for CDF and LERF.
- The formal approach for quantifying epistemic uncertainties in the λ's and ϕ's (or any other model parameter) involves the use of Bayes' Theorem. This is a straightforward process for many practical situations, and can be accomplished using spreadsheets or simple equation solving software.

Integrated PTS Analysis

To develop estimates of CDF and LERF associated with PTS, we know that thermal hydraulic (T/H) uncertainties and probabilistic fracture mechanics (PFM) uncertainties must be addressed in an integrated PRA framework. But how should this be done? Which uncertainties are aleatory? Which are epistemic? How should the results be presented? What does this mean in terms of the computational process used to generate the results?

This section proposes a particular approach for dealing with these questions. As indicated at the beginning of this paper, the intent is to provide an initial view and thereby stimulate constructive discussion. A final position cannot be developed without input from the PFM and T/H domain experts.

<u>The Problem</u>

Figure 6 shows a highly simplified view of the PTS problem with respect to the issue of CDF. (The discussion for LERF follows along very similar lines.) Using conventional PRA tools (e.g., event trees and fault trees), the scenarios resulting in PTS-related challenges to a particular reactor vessel (RV) at a particular plant can be identified and their frequencies (denoted in the figure by λ_i, i = 1,2,...,n) estimated. These frequencies characterize the aleatory uncertainties associated with the occurrence of the PTS challenge scenarios. Conventional PRA tools (e.g., Monte Carlo or Latin Hypercube sampling) can also be used to generate distributions quantifying the epistemic uncertainties in these frequencies.

Consider the ith PTS challenge scenario defined by the PRA. Using PFM models and judgment,[4] we can estimate ϕ_i, the conditional probability of vessel failure and core damage due to PTS, given the ith scenario. ***The parameter f_i is a measure of the aleatory uncertainty in the response of the vessel to the PTS challenge scenario***. It is perhaps best interpreted as the fraction of times PTS-induced core damage will be observed, given a large number of challenges of the type defined by scenario i. Care needs to be taken in defining which PFM uncertainties contribute to ϕ_i, and which contribute to the epistemic distribution for ϕ_i.

Before discussing a proposed treatment of aleatory and epistemic uncertainties in PFM which is based on the discussions provided earlier in this paper, we first need to address the question of why there should be a ϕ_i term at all. In other words, is the behavior of the reactor vessel deterministic, given the ith PTS challenge scenario?

---

[4]Judgment comes in when we are deciding what PFM endpoint is equivalent to core damage. Some possible endpoints are, in order of decreasing conservatism and increasing PFM uncertainty: RV crack initiation, RV through-wall crack, and catastrophic RV failure (i.e., failure of the RV beyond the capacity of available makeup). The general discussion in this paper is intended to cover all of these endpoints; the specific examples employed focus on crack initiation.

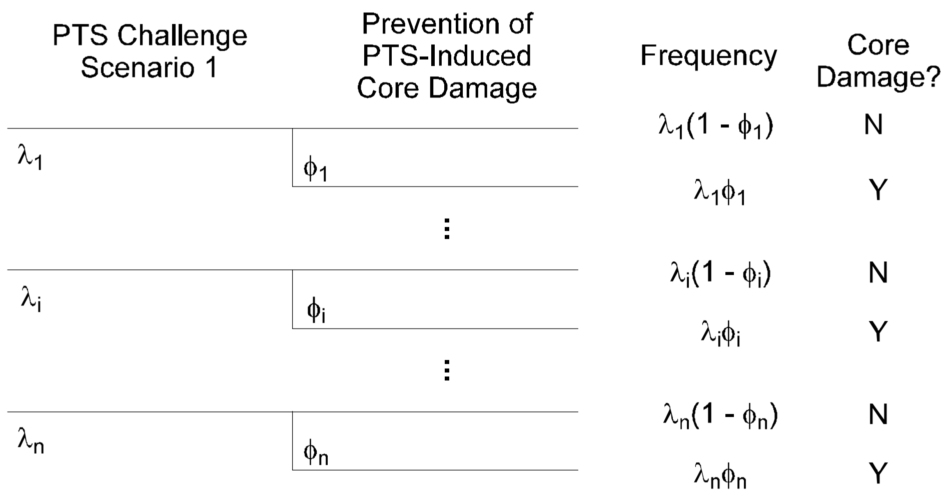


Figure 6 - Simplified PRA Representation of PTS Problem

I believe that variability in the response of the reactor vessel should be expected. This variability certainly arises because of the manner in which the PRA defines the PTS challenge scenarios. It may also arise due to modeling simplifications in the PFM analysis, even for such relatively well defined problems as crack initiation.

Consider first the issue of scenario definition. The PTS challenge scenarios identified by conventional PRAs are defined in terms of initiating events (e.g., steam line breaks) and successes or failures of mitigating equipment and actions (e.g., isolation of main feedwater on demand). Two important modeling approximations in this characterization are: a) all equipment and operator behaviors are treated as being binary (either successful or failed), and b) the timing of events is important only to the extent that it affects the definition of “success” or “failure.” The T/H response of the plant to the initiating event is clearly affected by these issues.

For example, a PRA might treat two states of a pressurizer PORV block valve: the block valve closes (on demand), and the block valve fails to close. If the block valve only closes midway or takes too long to close, the PRA might (depending on the precise success criteria employed) treat these as being equivalent to a situation where the valve gate doesn’t move at all. However, these different situations could lead to different temperature and pressure transients, and, therefore, different reactor vessel responses.

As another example, each initiating event treated in the PRA actually represents a set of potential accident initiators. For instance, the PRA groups steam line breaks of different sizes and locations. Again, these differences could lead to different temperature and pressure transients and different reactor vessel responses.

In general, it can be seen that each PRA-defined scenario actually represents a bundle of possible T/H scenarios. Even if reactor vessel behavior were a deterministic function of the T/H scenario, an experiment involving multiple occurrences of a particular PRA-defined PTS challenge scenario would be expected to lead to multiple outcomes due to variations in the T/H scenarios included in the PRA scenario.

Next consider the behavior of the reactor vessel. It is for the PFM analysts to decide if there can be any *significant* variations in the response of a specified reactor vessel to a well-defined T/H scenario. However, if the current PFM approach[5] includes models for material behavior that do not explicitly account for all potentially important factors (see the scatter data for K_{Ic}), then vessel behavior could vary, even if all PFM model input parameters (including those defining the T/H scenario) are fixed.

Based upon the preceding arguments, it appears that the concept of aleatory uncertainties in the behavior of the reactor vessel when subjected to a PTS-challenge scenario (as defined by the PRA) is valid. The term ϕ_i is therefore relevant and needs to be estimated.

<u>Analysis Interfaces</u>

Before discussing a proposal concerning how ϕ_i is to be estimated, a short discussion on the interfaces between the PRA, T/H, and PFM analyses is useful. This will provide a context for the discussion on estimation.

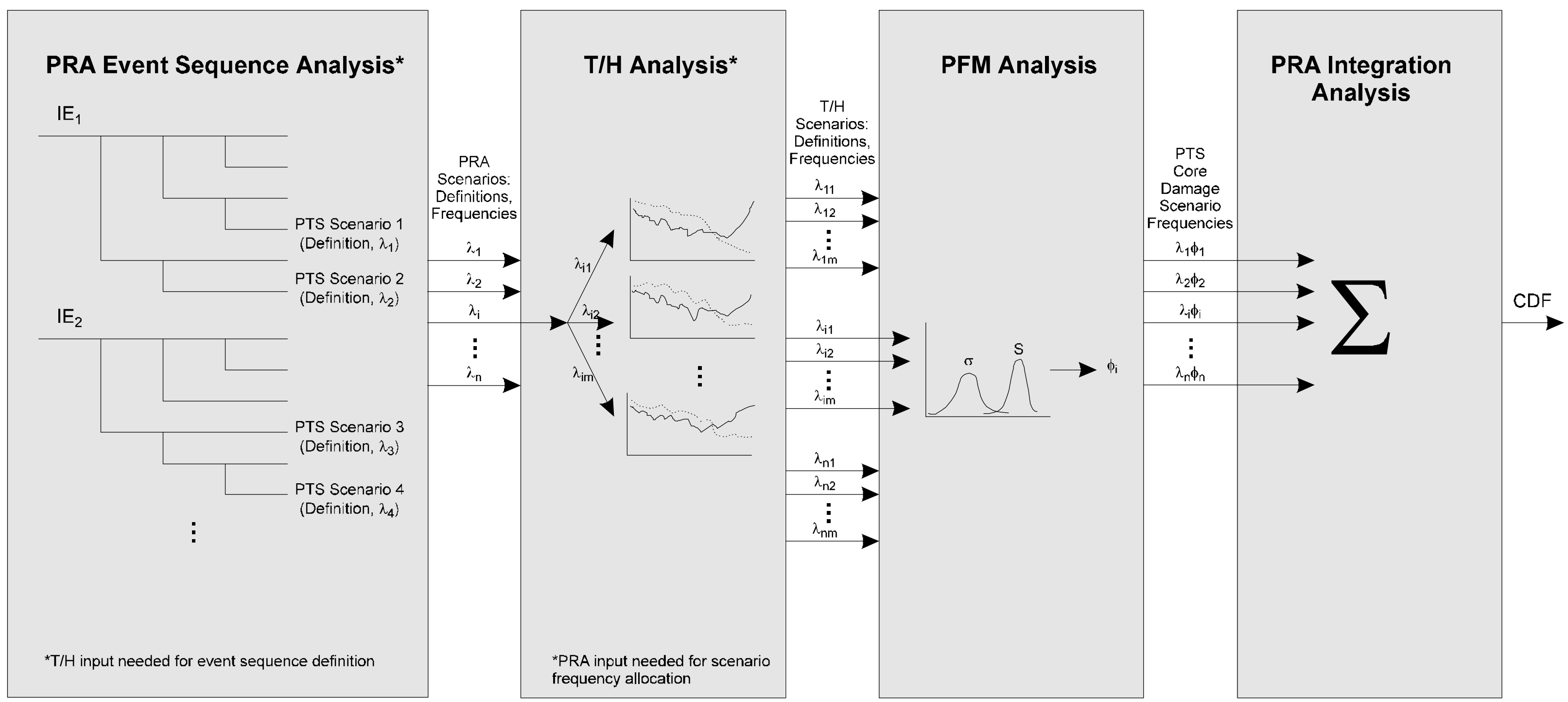
Figure 7 outlines a conceptual approach for defining the interfaces. In this approach, a PRA analysis (with some input from T/H analyses, e.g., regarding system success criteria) defines the PTS challenge scenarios in terms of initiating events (IEs) and associated equipment/operator successes and failures, and then estimates the frequencies (λ_i) of these scenarios.[6] These PRA scenario definitions and frequencies are provided to a T/H analysis. For each PRA scenario, a set of representative T/H scenarios is defined (with some additional input from the PRA analysis, e.g., regarding the likelihood of various failure times). Each representative T/H scenario, which is chosen to represent a bundle of similar T/H scenarios, is assigned an appropriate fraction of the PRA scenario frequency, and is analyzed using an appropriate T/H model. (Note that the effect of aleatory uncertainties in key T/H parameters, if any, should be factored into the T/H scenario frequencies; the effect of epistemic uncertainties in key parameters should be addressed through the epistemic uncertainties in both the scenario frequencies and the T/H output for each T/H scenario.) The results of each T/H scenario analysis, together with an estimate of the scenario frequency, are then provided to a PFM analysis. The PFM analysis then generates an estimate of ϕ_i.[7] The ϕ_i are then combined with the λ_i in an integrated assessment of CDF (shown) and LERF.

---

[5]All references to the "current PFM approach" refer to the proposed Method 2 presented at the joint NRC-industry meeting on PTS held on April 20, 1999 and discussed in subsequent NRC meetings.

[6]The estimation process is assumed to include the quantification of epistemic uncertainties.

[7]A decision needs to be made whether some reactor vessel endstate is going to be used to represent core damage, or if additional analysis between, say, through-wall crack propagation, and core damage is to be performed.



Note: the quantification of epistemic uncertainties in all parameters is not shown explicitly, but is assumed.

Figure 7 - Conceptual Interfaces Between PRA, T/H, and PFM Analyses for PTS

This approach appears to be nearly identical to that discussed at the April 30, 1999 and June 9, 1999 NRC meetings on PFM/PRA integration. Two minor differences are as follows. First, the proposed approach requires a slightly different aggregation of results (on a PRA scenario basis, rather than on an overall basis). Second, it requires that PRA scenario frequencies be explicitly allocated to the constituent T/H scenarios in a manner consistent with the PRA model.

Proposed Approach for Estimating ϕ_i

The PFM variables and parameters considered as being uncertain in the current approach to PTS are listed in Table 1. (This table is based on discussion at the June 9, 1999 NRC meeting on PFM/PRA.) My understanding is that the uncertainties in the variables and parameters listed as being "inside FAVOR," as well as the uncertainties in the T/H scenarios (each T/H scenario is effectively assigned a probability), are currently being addressed via Monte Carlo simulation in two ways (see Figure 8). First, most of the Table 1 variables and parameters (e.g., copper content, fluence, flaw size) are sampled to characterize a particular reactor vessel. Second, the possible T/H scenarios are sampled to estimate what fraction of these scenarios will lead to the failure of the given vessel. As shown in Figure 8, the first (reactor vessel-related) round of sampling effectively treats the sampled variables as being deterministic; the associated uncertainties are therefore epistemic. The second (T/H-related) round of sampling effectively treats the sampled variables as being random; the associated uncertainties are therefore aleatory. (Note that in Figure 8, the "ϕ" and "P_{FM}" terms correspond to the "λ" and "ϕ" terms, respectively, of this paper.)

Table 1 - Uncertain Variables and Parameters in PFM

Inside FAVOR[a]	**Outside FAVOR**[a]
copper content	weld residual stresses
nickel content	cladding thickness
neutron fluence	stress-free temperature
flaw size	flaw size distributions[b]
flaw location	flaw density[b]
RT_{NDT} margin	T/H pressure-temperature curve[b]
reactor vessel temperature	
reactor vessel stress	
K_I	
K_{Ic} scatter	

[a]Based on current version of FAVOR
[b]Might be able to move inside FAVOR without modifying loading/stress intensity libraries

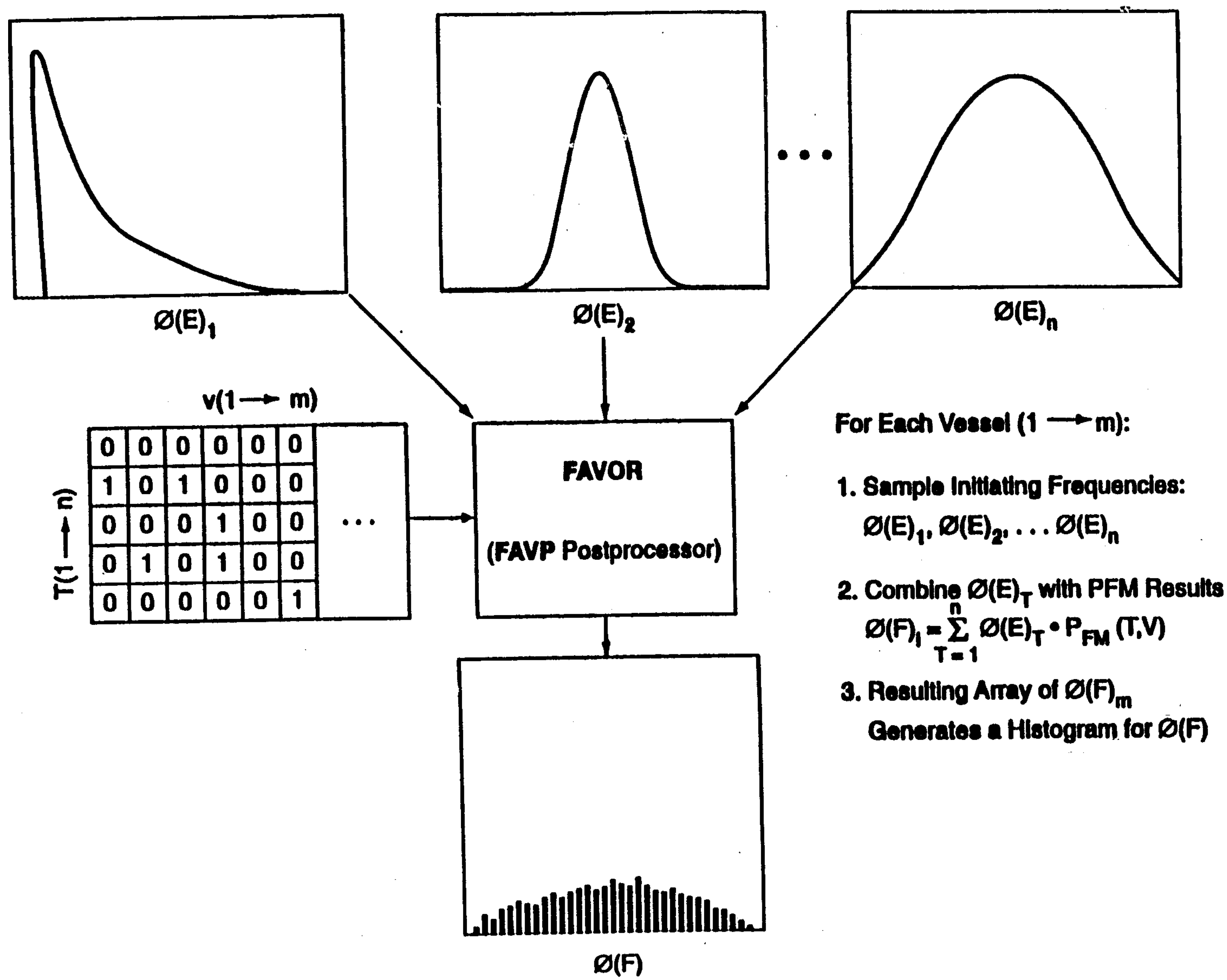


Figure 8 - Method 2 PTS PFM Analysis

This section of the paper re-examines the variables and parameters listed in Table 1 in light of the philosophical discussion provided in the first section of the paper. It then provides recommendations as to whether their uncertainties should be categorized as being aleatory, epistemic, or both.[8] It concludes with a discussion of the implications of any changes in categorization on FAVOR.

Modeling Observations and Assumptions Concerning the Reactor Vessel

As mentioned early in this paper, the distinction between aleatory and epistemic uncertainties is, to some degree, a matter of modeling. The discussion therefore starts with some modeling observations and assumptions that will be used to provide a basis for the discussion on uncertainty-based categorization.

First, it should be recognized that, under the current PTS program, analyses will be performed for a set of specified plants and reactor vessels. Thus, the although the results will be used in developing a generic screening criterion, the analyses themselves are not generic.

Second, looking at a specific reactor vessel, the vessel's material properties are essentially deterministic. In other words, the concept of "the true value" for such variables as the copper content at a specified point[9] is meaningful, whether or not there are problems with our current ability to reliably measure those variables. Other reactor vessel spatially dependent physical characteristics that can be viewed as being deterministic on a pointwise basis are the weld residual stresses, the vessel cladding thickness, and flaws in the vessel. Regarding the latter, it appears that the flaws in the reactor vessel are those created during manufacturing, i.e., non-catastrophic operational transients cannot initiate or propagate flaws with any significant likelihood. If this observation is incorrect, then random variations in the timing and magnitude of such transients would then lead to random variations in flaw density, size, and location.

Regarding external influences on the reactor vessel prior to the PTS challenge, it seems reasonable to assume that the spatially dependent neutron fluence can be treated as being deterministic. (There are random fluctuations in neutron flux, but time averaging will tend to smooth out these fluctuations.) Regarding external influences during the challenge, it seems that reactor coolant temperature and pressure can also be treated as being deterministic, i.e., that the impact of random fluctuations will be small (due to vessel thermal and mechanical inertia).

Third, many of the reactor vessel properties and external influences will vary with location (r,θ,z). This means, for example, that a sampling of the copper content over a specified vessel subregion will result in an empirical distribution of values for that property. (This distribution can be fairly broad and can be multimodal.) It should be emphasized that the existence of a sampling distribution reflects aleatory uncertainty in the sampling process. It does not necessarily mean that the pointwise values are themselves random.

---

[8]Random variables whose distributions are uncertain have both aleatory and epistemic uncertainties.

[9]The "value of a continuously distributed variable at a point in the reactor vessel" is understood to mean the average value in a suitably small subvolume about that point.

Proposed Categorization of Uncertain Variables and Parameters

The following proposals concerning the categorization of the variables and parameters listed in Table 1 are based upon the preceding observations and assumptions

- Copper Content: Epistemic

  In the current PFM approach, which is done on a subregion basis, the copper content is sampled once per flaw. This is done because the concern is not with the average copper content in the entire subregion (whose characteristic dimensions can range from several centimeters to even a few meters), but rather with the copper content local to the flaw (and at the time of the PTS challenge). The sampling is done using a distribution derived from empirical data. As noted earlier, the procedure essentially treats the uncertainties in copper content as being epistemic in nature.

  Both the flaw location and the local copper content are, in principle, deterministic. (They are essentially determined when the vessel is manufactured.) Thus, it seems reasonable (i.e., consistent with the principles described in the first part of this paper) to treat the uncertainty in the copper content as being epistemic. Sampling based distributions can be used to quantify epistemic uncertainties,[10] but they should not be used as aleatory distributions. Note that the current assumption that the uncertainty distribution for copper content is Gaussian may need to be revisited; the investigation can be done in a straightforward manner using standard statistical tools.

- Nickel Content: Epistemic

  See the discussion for copper content.

- Neutron fluence: Epistemic

  In the current PFM approach, the neutron fluence is sampled once per flaw (to support the calculation of the extent of embrittlement near the flaw). The sampling is done using a distribution derived from expert judgment concerning the accuracy of neutronics calculations. The procedure essentially treats the uncertainties in fluence as being epistemic in nature.

  As argued earlier, although there are random fluctuations in the neutron flux (and therefore fluence), the time averaging used to calculate the fluence should tend to reduce the impact of these fluctuations. It therefore appears reasonable to treat the uncertainty in the fluence as being epistemic in nature. Expert judgment, which could involve a more detailed treatment which explicitly addresses the key sources of uncertainty, can be used to quantify the uncertainty.

[10] In cases where the assessor chooses to use the sampling distribution directly as a representation of his/her state of knowledge, they are numerically identical.

! Flaw size: Epistemic

In the current PFM approach, uncertainties in the crack geometry are effectively treated being treated as being epistemic in nature. Since non-catastrophic operational transients apparently have little effect on flaw initiation or growth, it appears that the geometry of a given flaw should be deterministic. (It is essentially determined when the vessel is manufactured.) Therefore, it appears reasonable to treat the uncertainties in flaw size as being epistemic. As is the case with copper and nickel content, sampling based distributions can be used to quantify the epistemic uncertainties in flaw size, but they should not be used as aleatory distributions.

! Flaw location: Epistemic

See the preceding discussions on copper content and flaw size.

! RT_{NDT} margin: Epistemic

In the current PFM approach, this term is used to account for uncertainties in both the initial, unirradiated value of RT_{NDT}, i.e., RT_{NDT0}, and uncertainties in the correlation used to predict the neutron radiation-induced shift in RT_{NDT}, i.e., ΔRT_{NDT}. As with most of the other variables and parameters discussed, the uncertainties are treated as being epistemic in nature.

This treatment appears to be reasonable. The parameter RT_{NDT0} is derived experimentally under a specified protocol. For the purposes of the PTS analysis, it appears that it can be considered as a material property. This means that the uncertainties in RT_{NDT0} can be treated as being epistemic. For similar reasons, the parameter ΔRT_{NDT} can also be considered as a material property, and its uncertainties can be treated as being epistemic in nature.

Note that the comparison of correlation results for ΔRT_{NDT} with experimental data will lead to a sampling distribution for error in the correlation (due to the effect of factors not included in the correlation). This sampling distribution can be used to develop the epistemic distribution for ΔRT_{NDT}, but it should not be taken to mean that ΔRT_{NDT} at a given point (the location of the flaw) is itself aleatory.

Also note that the correlation for ΔRT_{NDT} requires values of copper content, nickel content, and fluence, all of which are uncertain. Estimation of the uncertainties in ΔRT_{NDT} due solely to modeling needs to be done recognizing these uncertainties. Bayesian methods have been developed to address this problem.

! Reactor vessel temperature: Deterministic

In the current PFM approach, the spatial distribution of temperature inside the reactor vessel is computed deterministically based on the temperature-time curves provided by the T/H analysis. (Presumably, the heat transfer coefficients and material thermal properties, e.g., thermal diffusivities, are assumed to be constant.) Uncertainties in the

T/H input will lead to uncertainties in the vessel temperature, but there are no other sources of uncertainty considered.

Unless the effect of uncertainties in the heat transfer coefficients and the material thermal properties are believed to be important, there is no need to perform any additional sampling.

! Reactor vessel stress: Deterministic

In the current PFM approach, the spatial distribution of stress inside the reactor vessel is also computed deterministically (based on a number of factors, including the time-dependent temperature profile, the vessel geometry, and the weld residual stresses.) Unless there are any significant uncertainties in these calculations, there is no need to perform any additional sampling.

! K_I: Deterministic

This variable is currently computed deterministically as a function of other variables. Unless it is postulated that the computation process itself introduces additional uncertainties, there is no need to perform any additional sampling.

! K_{Ic} scatter: Aleatory and Epistemic

In the current PFM approach, the scatter in K_{Ic} is sampled once per time step for each flaw. (The sampling distribution is based on a comparison of K_{Ic} predictions with experimental data.) Based on when the sampling is done (K_{Ic} is a function of local temperature, which is a function of the thermal hydraulic transient), it appears that the uncertainties in K_{Ic} are being treated as being aleatory in nature.

At first glance, it appears that K_{Ic}, which is computed as a function of T - RT_{NDT}, is a temperature-dependent material property and should therefore be deterministic (at a given point). However, consider the crack initiation model which uses K_{Ic}. This model predicts crack initiation whenever K_I, which is a computed function of a number of factors (e.g., crack geometry and applied stress), exceeds K_{Ic}. Applying this model to experimental results, it would not be surprising for the model would be correct for some trials and incorrect for others. (The graph showing variability in K_{Ic} for fixed values of T - RT_{NDT} may be an indication of this aleatory uncertainty. Note that models, by definition, are simplified representations of the real world, and generally don't address all factors that can potentially affect the results.) Thus, although the uncertainties in K_{Ic} are epistemic, there are aleatory uncertainties in the results of the model which uses K_{Ic}.

Note that in a mathematically analogous problem involving aging-related failures of piping, Apostolakis (1999) argues that model uncertainty should be treated as being epistemic in a PRA. It is currently planned that a small task group reinvestigate the treatment of the scatter in K_{Ic}. The task group will need to determine if the current PFM distribution for K_{Ic} appropriately addresses the model uncertainty and how epistemic uncertainties in the model should be addressed.

! Weld residual stresses: Epistemic

In the current PFM analysis, these are treated as being deterministic. (They affect the finite element stress calculations, and therefore cannot be easily incorporated into the current computational scheme used by FAVOR to address uncertainties.)

Since weld residual stresses are essentially determined at the time of vessel manufacture, the uncertainties in these stresses are epistemic in nature. Given the difficulty of addressing these uncertainties within FAVOR, a scheme for doing this outside of FAVOR is outlined later in this section.

! Cladding thickness: Epistemic

In the current PFM analysis, this is treated as being deterministic. (It affects the finite element stress calculations, and therefore cannot be easily incorporated into the current computational scheme used by FAVOR to address uncertainties.)

Since the vessel dimensions (including the cladding thickness) are essentially determined at the time of vessel manufacture, the uncertainties in this thickness (for a given subregion) are epistemic in nature. Given the difficulty of addressing these uncertainties within FAVOR, a scheme for doing this outside of FAVOR is outlined later in this section.

! Stress-free temperature: Epistemic

In the current PFM analysis, this is treated as being a deterministic parameter. Presuming that, for a given reactor vessel, there is a temperature at which the stress between the cladding and the vessel base material is zero, it appears that this treatment is reasonable. The uncertainties in the parameter are, therefore, epistemic.

! Flaw size distributions: Epistemic

In the current PFM analysis, uncertainties in the flaw size distribution (e.g., regarding its shape and parameter values) are not treated. Since, as noted earlier, the uncertainties in the flaw characteristics are epistemic in nature, the uncertainties in the distribution of characteristics is also epistemic. From a computational point of view, the proposed treatment of flaw characteristics accounts for uncertainties in the flaw size distribution; no additional treatment is needed.

! Flaw density: Epistemic

Following the discussion of other flaw characteristics, the flaw density is determined at the time of vessel manufacture and the uncertainties in this density are epistemic.

! T/H pressure-temperature curve: Aleatory and Epistemic

In the current PFM analysis, T/H uncertainties are used directly in the computation of the ϕ_i; this procedure treats the T/H uncertainties as being aleatory.

The proposed treatment of T/H uncertainties has been discussed earlier in this paper. It recognizes that there is an aleatory component (quantified by the frequency of the parent PRA scenarios and the fraction of this frequency associated with the bundle of T/H scenarios modeled through the use of a single representative T/H scenario) and an epistemic component (quantified by distributions for the T/H scenario frequencies and the conditional T/H model output).

Table 2 summarizes the results of the preceding discussions on the categorization of uncertain PFM variables and parameters. In general, the conceptual treatment of uncertainties in the variables and parameters used by the current PFM approach appears to be consistent with the principles described in the first part of this paper (although a PRA-based description would describe the process somewhat differently[11]). The impact of changes in categorization are discussed in the following section.

Implications for FAVOR

Table 2 shows that, from the standpoint of PFM uncertainty analysis, four classes of variables/parameters have been identified.

1. Variables/parameters which do not need to be explicitly included in sampling schemes used to perform the uncertainty analysis. These are generally deterministic functions of other uncertain variables/parameters. Uncertainties in these will be automatically dealt with as part of the uncertainty analysis process.

2. Variables/parameters which have both aleatory and epistemic uncertainties. The epistemic uncertainties can be addressed within FAVOR.

3. Variables/parameters which have epistemic uncertainties. The epistemic uncertainties can be addressed within FAVOR.

4. Variables/parameters which have epistemic uncertainties. The epistemic uncertainties cannot be addressed within FAVOR (at least without considerable restructuring of the code).

The discussion in the previous section and Table 2 show that the current PFM categorization of variables and parameters is generally reasonable. Furthermore, Figure 8 shows that the computational approach used by FAVOR appropriately distinguishes between aleatory and epistemic uncertainties. Thus, the following points, which address recommended changes in the PFM uncertainty analysis, do not appear to require significant changes in the FAVOR code.

[11]For example, as noted earlier in this paper, the term "stochastic" is typically used in the PRA literature to refer to random or aleatory issues. My understanding is that the process of "stochastically generating vessels" actually addresses epistemic uncertainties. I recommend that future descriptions of the PFM analysis use the terminology of this white paper.

Table 2 - Recommendations for Categorization of Uncertain Variables and Parameters in PFM

Variable/Parameter	**Recommended Uncertainty Category**[a]
copper content	epistemic
nickel content	epistemic
neutron fluence	epistemic
flaw size	epistemic
flaw location	epistemic
RT_{NDT} margin	epistemic
reactor vessel temperature	deterministic[b]
reactor vessel stress	deterministic[b]
K_I	deterministic[b]
K_{Ic} scatter	aleatory and epistemic
weld residual stresses	epistemic
cladding thickness	epistemic
stress-free temperature	epistemic
flaw size distributions	epistemic[c]
flaw density	epistemic
T/H pressure-temperature curve	aleatory and epistemic

[a]Underline indicates a change from the current PFM approach.
[b]Variable is a deterministic function of other, uncertain variables; no additional treatment of uncertainty is required.
[c]Uncertainties in flaw size distribution should be addressed as part of the uncertainty analysis for flaw size.

! Category 2 Variables and Parameters: K_{Ic} scatter and T/H temperature/pressure

In general, the parameters of aleatory distributions are uncertain. If these uncertainties are significant (methods for quantifying these uncertainties were discussed in the first section of this paper), they need to be addressed in the sampling process. This can be done in a very straightforward manner within the FAVOR code.

Assume, for example, that the distribution of K_{Ic} is lognormal with uncertain parameters μ and σ. At the time FAVOR is sampling the reactor vessel parameters (e.g., copper content, which have epistemic uncertainties), it should also sample a value for μ and a value for σ. Then, when FAVOR is actually sampling for K_{Ic}, it should use the sampled values of μ and σ in defining the lognormal distribution for K_{Ic}.

! <u>Category 4 Variables and Parameters: weld residual stresses, cladding thickness, and stress-free temperature</u>

Although the epistemic uncertainties in these variables and parameters are fundamentally of the same nature as the epistemic uncertainties in other variables and parameters, it appears for computational efficiency reasons that they should be addressed outside of the FAVOR code. It appears that this can be done relatively simply using Latin Hypercube Sampling (LHS) techniques; LHS is used to define sets of inputs (with appropriate probability weights) that are then provided to FAVOR.

Summary Points - Integrated PTS Analysis

C The proposed approach for integrating PRA, T/H, and PFM analyses described in this paper (see Figure 7) is nearly identical to that discussed at the April 30, 1999 and June 9, 1999 NRC meetings on PFM/PRA integration. Two minor differences are: 1) the proposed approach requires the aggregation of results on a PRA scenario basis, rather than on an overall basis; and 2) the approach requires that PRA scenario frequencies be explicitly allocated to the constituent T/H scenarios in a manner consistent with the PRA model.

C Although it doesn't use the same terminology, the uncertainty analysis framework employed by the current PFM approach correctly distinguishes between epistemic and aleatory uncertainties.

C The current PFM categorization of uncertain PFM variables and parameters (in terms of whether the uncertainties are epistemic, aleatory, or both) appears to be generally reasonable. A few changes in categorization are recommended (see Table 2). Some of these changes can be addressed within the current FAVOR code; others will need to be addressed outside of the code.

C The quantification of aleatory uncertainties in K_{Ic} and of the epistemic uncertainties in this distribution needs to be looked at further.

C The current quantification of uncertainties for many of the PFM variables and parameters can be updated using relatively simple tools.

Recommended for Further Reading

Apostolakis, G., "Probability and risk assessment: the subjectivistic viewpoint and some suggestions," *Nuclear Safety*, *9*, 305-315(1978). [*Provides a seminal discussion on the interpretation of probability appropriate to PRA*.]

Apostolakis, G., "The concept of probability in safety assessments of technological systems," *Science*, 250, 1359-1364(1990). [*An update of the Nuclear Safety paper*.]

Apostolakis, G., "A commentary on model uncertainty," in *Model Uncertainty: Its Characterization and Quantification*, A. Mosleh, N. Siu, C. Smidts, and C. Lui, eds., Center for Reliability Engineering, University of Maryland, College Park, MD, 1995, pp. 13-22. [*Provides first reference to the terminology "aleatory" and "epistemic" uncertainties in a PRA context*.]

Apostolakis, G., "The distinction between aleatory and epistemic uncertainties is important: an example from the inclusion of aging effects into PSA," *Proceedings of Probabilistic Safety Assessment International Topical Meeting (PSA '99)*, Washington, DC, 1999. [*Provides a detailed discussion of aleatory and epistemic uncertainties in the context of a PSA aging analysis. The analysis includes a phenomenological model for piping failure, and is mathematically similar to the PTS problem addressed in this paper*.]

Iman, R.L, and M.J. Shortencarier, "A FORTRAN Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models," NUREG/CR-3624, 1984. [*Provides a brief introduction to Latin Hypercube Sampling as well as the program*.]

Kaplan, S. and B.J. Garrick, "On the quantitative definition of risk," *Risk Analysis*, 1, 11-37(1981). [*Provides a pioneering discussion of the purpose of risk assessment and the need to address uncertainties*.]

Parry, G.W. and P.W. Winter, "Characterization and evaluation of uncertainty in probabilistic risk analysis, *Nuclear Safety*, *22*, 28-42(1981). [*An early discussion of various sources of uncertainty relevant to PRA, including model uncertainty*.]

Helton, J.C. and Burmaster, D.E., Guest Editors, "Treatment of Aleatory and Epistemic Uncertainty," Special Issue of *Reliability Engineering and System Safety*, 54(1996). [*Includes contributions from a number of authors on the topic*.]

Siu, N. and D.L. Kelly, "Bayesian Probability and Statistics in PRA," *Reliability Engineering and System Safety*, *62*, 89-116, 1998. [*Provides a tutorial level discussion on Bayesian estimation*.]

Siu, N., D. Karydas, and J. Temple, "Bayesian Assessment of Modeling Uncertainty: Application to Fire Risk Assessment," *Analysis and Management of Uncertainty: Theory and Application*, B.M. Ayyub, M.M. Gupta, and L.N. Kanal, eds., North-Holland, 1992, pp. 351-361. [*Provides a Bayesian methodology for quantifying model uncertainty when model input parameters are also uncertain*.]

U.S. Nuclear Regulatory Commission, “An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis,” Regulatory Guide 1.174, July 1998.March 25, 1999. [*Discusses how uncertainties are to be treated in one particular risk-informed approach to regulatory decision making*.]

Winkler, R.L., “Model uncertainty: probabilities for models?”, in *Model Uncertainty: Its Characterization and Quantification*, A. Mosleh, N. Siu, C. Smidts, and C. Lui, eds., Center for Reliability Engineering, University of Maryland, College Park, MD, 1995, pp. 109-118. [*Discusses a Bayesian approach for addressing model uncertainty*.]

Appendix A - Probability Definitions and Concepts

Probability

Probability is a subjective (internal) measure of likelihood.[12] Thus, P{A} is the quantity that measures the assessor's degree of certainty (or uncertainty) as to the truth of proposition A.[13] P{A|B}, the conditional probability of A, given B, measures the assessor's belief that proposition A is true, given (assuming) that proposition B is true. Some important observations are as follows:

1. Although there is no "true" or "correct" probability for a given proposition, useful probabilistic assessments are not arbitrary; they must adhere with the rules established by the calculus of probabilities. It turns out that this requirement forces convergence of subjective and frequentist probabilities when there is a large amount of data.

2. For a probability to be meaningful, the proposition must be carefully defined. Lack of clarity can lead to misunderstandings and misuses of probabilistic analysis results.

3. All probabilities are conditional; they are all based on the assessor's current state of knowledge concerning the proposition in question. As that state of knowledge changes, the (conditional) probability of the proposition changes as well.

4. The definition of probability does not distinguish between "aleatory" and "epistemic" uncertainties. Uncertainties of both types contribute to the overall probability. However, they contribute in different manners, as illustrated by an example at the end of this appendix.

Probability Distributions

Let X be a continuous variable (e.g., the copper content at a specific point in the reactor vessel) whose precise value is unknown. Some generic propositions of interest are:

$$\{X \le x\}$$
$$\{X > x\}$$
$$\{x \le X < x + \Delta x\}$$

where x is a given value. The probabilities of these propositions being true clearly can change as functions of x. Because of their usefulness, these functions have been given specific names:

---

[12] Although there are other definitions of probability, e.g., the "frequentist" definition which takes the probability to be the limiting ratio of successes to trials in an infinite series of repeatable, identical experiments, the subjectivist definition is appropriate for use in PRA, as it is an integral part of current theories on decision making under uncertainty.

[13] A proposition is a statement that is either true or false.

Cumulative Distribution Function (*CDF*): $F(x) \equiv P\{X \leq x\}$

Complementary Cumulative Distribution Function (*CCDF*): $\overline{F}(x) \equiv P\{X > x\}$

Probability Density Function (*pdf*): $f(x) \equiv \lim_{\Delta x \to 0} \frac{P\{x \leq X < x + \Delta x\}}{\Delta x}$

Some useful relationships following from these definitions and the axioms of probability are given in Table A.1.

It is important to observe that all of the above distribution functions are probabilities which quantify the assessor's subjective beliefs as to whether the true value of X lies in a specified range. Thus, for example (see Figure A.1), a highly peaked pdf indicates that the assessor is, correctly or incorrectly, very confident in his knowledge about X; a more shallow pdf indicates a lower level of confidence.

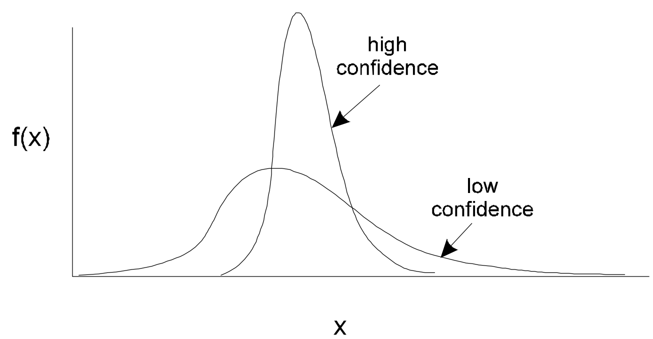


Figure A.1 - Probability Density Functions (pdfs) and Confidence

It should also be noted that neither the definitions of distributions nor the relationships in Table A.1 are dependent on the particular form of the distribution. This means that, in principle, probability distributions do not have to members of any particular parametric family, e.g., normal (Gaussian), lognormal, gamma, Weibull, or exponential. However, for mathematical and computational convenience, it is often useful to approximate the assessor's distribution using a particular parametric form. Specific forms and their characteristics (e.g., mean value, variance, key percentiles) can be found in numerous handbooks and textbooks.

The above discussion focuses on a single uncertain variable. Similar propositions and associated distribution functions can be developed for multiple uncertain variables, albeit with more complexity. In dealing with multiple variables, care needs to be taken that dependencies between the variables are accounted for because, in general,

$$P\{a \leq X < b, c \leq Y < d\} \neq P\{a \leq X < b\} \cdot P\{c \leq Y < d\}$$

Table A.1 - Some Useful Relationships Between Distribution Functions

$$\overline{F}(x) = 1 - F(x)$$

$$F(x) = \int_{-\infty}^{x} f(x')dx'$$

$$P\{a \le X < b\} = F(b) - F(a) = \int_{a}^{b} f(x')dx'$$

$$f(x) = \frac{dF(x)}{dx}$$

On The Meaning of Probability Distributions: Examples

In probabilistic risk assessments (PRAs) and other probabilistic analyses, probability distributions are routinely used to represent the uncertainties in key variables and parameters. However, the meaning of each distribution, which is directly related to the specific proposition addressed by the distribution, is not always clearly specified. This can lead to misunderstandings or even misuses of the distributions and, therefore, of the analysis results.

Example 1: Reactor Vessel Copper Content

Define the variable C as the copper content (in weight percent) at the location of a specific flaw in a particular subregion of the vessel.[14] From an engineering analysis perspective, it is reasonable to assume that there is a fixed, “true value” of C, whether or not there are problems with our current ability to reliably measure C. The proposition of interest, therefore, is that the true value of C lies in a specific range of values, e.g., $(c, c+\Delta c)$.

For the sake of this simple example, assume that, following some data analysis (see Appendix C for example calculations), the assessor determines that his state of knowledge regarding C is adequately represented by a lognormal distribution function with a mean value of 0.20 and a standard deviation of 0.05. The pdf is shown in Figure A.2.

---

[14] C is clearly a function of position; its explicit dependence on (r,θ,z) is not shown for notational convenience.

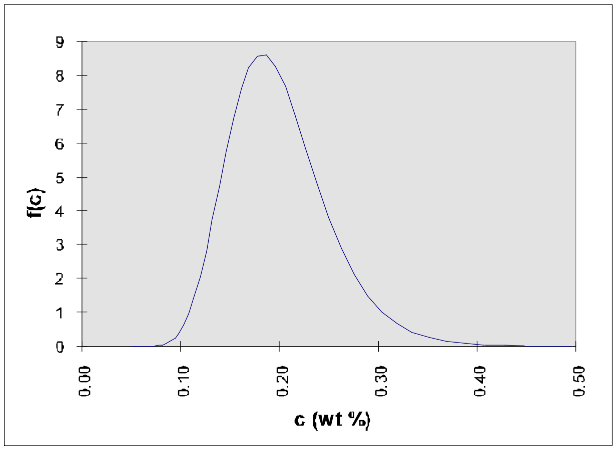


Figure A.2 - Copper Content Probability Density Function (Example)

Using the properties of the lognormal distribution function, it can be shown that some of the key percentiles of this distribution are as follows.

$$C_{05} = 0.129, \text{ i.e., } P\{C \le 0.129\} = 0.05$$
$$C_{50} = 0.194, \text{ i.e., } P\{C \le 0.194\} = 0.50$$
$$C_{95} = 0.291, \text{ i.e., } P\{C \le 0.291\} = 0.95$$

It can be seen that the assessor is very confident (with 95% probability) that the true (but unknown) value of C is less than 0.291. It also can be seen, using the third relationship in Table A.1, that the assessor is very confident (with 90% probability) that the true value of C lies between 0.129 and 0.291.

Note that the uncertainties modeled by this distribution of C are purely epistemic and should be treated as such. If the uncertainties are treated in an analysis as being aleatory,[15] this would imply that C could vary randomly over time (e.g., from pressurized thermal shock event to event), which contradicts the basic modeling assumption that there is a fixed, true value of C.

Example 2: On Measurement Errors and Epistemic Uncertainties

Consider a situation where the copper content of a particular sample is measured in a series of tests. It can be expected that random variations in the measurement process will lead to random variability, i.e., aleatory uncertainty, in the measurement outcomes, and that this variability can be represented by a distribution. Does this mean that the copper content is an aleatory variable?

[15] Appendix B provides additional discussion on the treatment of epistemic and aleatory uncertainties in a probabilistic analysis.

The answer depends on what is meant by “the copper content,” i.e., what is the underlying proposition.

If the proposition is that the value of the next measurement of copper content falls in some range, e.g., $(c, c+\Delta c)$, the uncertainty in the truth of this proposition is indeed aleatory. (This follows directly from the description of the situation.) The observed distribution provides a good indication of what the next measurement might be, as long as key factors (e.g., the test procedure, the sample itself) are not changed.

On the other hand, if the proposition is that the copper content at some specified (r,θ,z) in a given reactor vessel falls in some range, then the model of the previous example still holds: the uncertainties in this copper content are epistemic. The distribution of measured values for the sample is evidence which affect the assessor’s distribution for $C(r,\theta,z)$, but it is not the assessor’s distribution. Even if, as a practical matter, the assessor decides to make his distribution for $C(r,\theta,z)$ numerically identical to the distribution of measured values, his distribution must be treated in subsequent analyses as being epistemic (rather than aleatory), or else the analysis will be inconsistent.

Appendix B - Aleatory and Epistemic Uncertainties

Is It Important To Make The Distinction?

In order to make most effective use of the results of any analysis, it is important that the user understand the fundamental modeling assumptions underlying the analysis. In particular, in the case of a probabilistic risk assessment (PRA), it is important to understand how the analysis deals with uncertainties that arise because of issues not explicitly modeled and those that arise because of imperfect knowledge concerning the issues that are explicitly modeled. This understanding will affect how the user perceives and uses the analysis results in subsequent decision making activities.

Consider a situation where a reactor pressure vessel (RPV) could be subjected to a pressurized thermal shock (PTS) event. Assume the PTS event arrival is governed by a Poisson process and has characteristic frequency λ. We are uncertain as to whether the RPV will fail because of a PTS event; the associated conditional probability of failure, given a PTS event, is ϕ. Depending on the interpretation of ϕ, the analysis user could have very different pictures of the situation.

Two extreme interpretations are as follows (see Figures B.1 and B.2).

1) The uncertainty quantified by ϕ arises only because of issues not explicitly modeled (e.g., causal factors underlying differences in the timing of component actuations and failures, which, in turn, lead to different thermal hydraulic subscenarios) and is entirely aleatory. Under this treatment, if we hypothesize a very large number of PTS events, we would expect to see RPV failure for a fraction ϕ of these events.

2) The uncertainty quantified by ϕ arises only because of imperfect knowledge regarding modeled processes (e.g., sparsity and relevance of data for the copper content at a specific point in the RPV) and is entirely epistemic. Under this treatment, the RPV will either fail or it won't, regardless of the number of challenges. Thus, for N hypothesized PTS events, one of two hypotheses will be true: i) there will be N RPV failures, or ii) there will be N RPV successes. The likelihood that the first hypothesis is true is ϕ; the likelihood that the second hypothesis is true is 1 - ϕ.

Under the first interpretation, the expected number of PTS-induced RPV failures in a fixed time interval T is given by:

$$E[\#\ \text{RPV failures in } (0,T) | \text{interpretation } 1] = \lambda\phi T$$

The probability of N such events is given by:

$$P\{N\ \text{RPV failures in } (0,T) | \text{intepretation } 1\} = \frac{(\lambda\phi T)^N}{N!} e^{-\lambda\phi T}$$

Under the second interpretation, the expected number of events and the probability of N such events are given by:

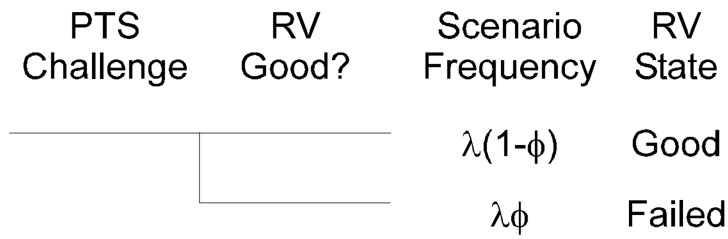


Figure B.1 - Risk Model for Aleatory Interpretation of RPV Conditional Failure Probability

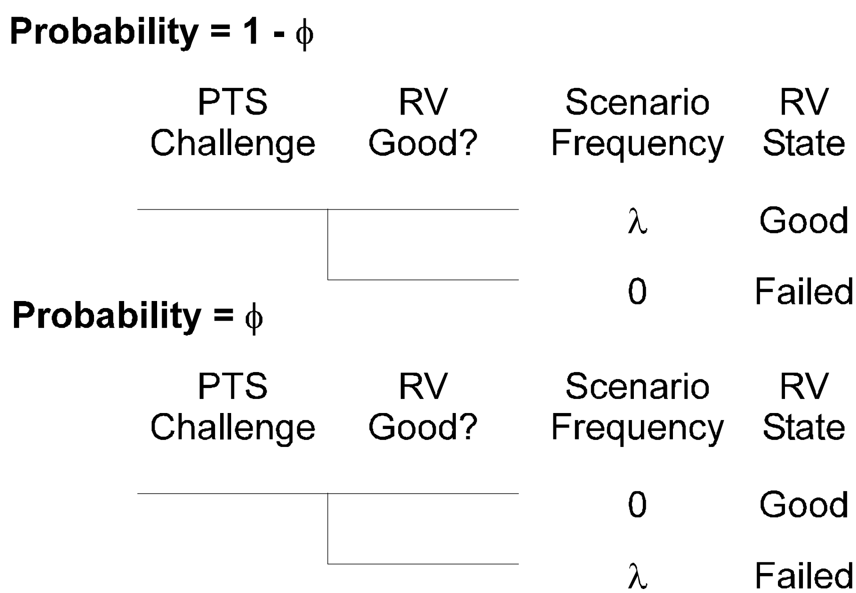


Figure B.2 - Risk Model for Epistemic Interpretation of RPV Conditional Failure Probability

$$E[\#\ \text{RPV failures in } (0,T) | \text{interpretation 2}] = \lambda\phi T$$

$$P\{N \text{ RPV failures in } (0,T) | \text{intepretation 2}\} = \phi \cdot \frac{(\lambda T)^N}{N!} e^{-\lambda T}$$

It can be seen that if the user only cares about the expected number of events in a fixed time interval T, both interpretations will lead to the same value: $\lambda\phi T$. However, if the user has a non-linear consequence function for PTS-induced RPV failure (e.g., if one event is barely tolerable but two events spell utter doom), the differences in interpretation can make a difference.

Reinforcing the points raised above, Apostolakis (1999) points out that the distinction between aleatory and epistemic uncertainties can make a difference at the detailed technical analysis level. In particular, he questions the concept of a failure rate for components when the failure mechanisms are essentially deterministic (albeit, with uncertain governing parameters). The problem involves the passive failure of an aging pipe under steady-state load conditions, and corresponds mathematically to the situation shown in Figure B.2.

In general, it might be expected that there are aleatory and epistemic contributions to the RPV conditional failure probability. Operational issues in dealing with such situations are discussed in the following section.

Treating Aleatory and Epistemic Uncertainties

For situations where there are aleatory and epistemic contributions to uncertainty, these contributions need to be separated for the reasons discussed above. In our example of the PTS-induced RPV failure, this separation is shown in Figure B.3. The aleatory contribution (ϕ') is dealt with in the event tree (i.e., as a “conditional split fraction”). The epistemic uncertainty in ϕ' is treated when epistemic uncertainties are propagated through the event tree model. Neglecting the epistemic uncertainties in λ for simplicity, the expected number of failures and the probability of N events are given by:

$$E[\#\ \text{RPV failures in } (0,T)] = \lambda T \cdot \int_0^1 \phi' \pi(\phi') d\phi' = \lambda T \cdot E[\phi']$$

$$P\{N \text{ RPV failures in } (0,T)\} = \int_0^1 \frac{(\lambda\phi' T)^N}{N!} e^{-\lambda\phi' T} \pi(\phi') d\phi'$$

where $\pi(\phi')$ is the epistemic pdf for ϕ'. Note that ϕ, the overall conditional probability of PTS-induced RPV failure given a PTS event, is given by:

$$\phi = \int_0^1 \phi' \pi(\phi') d\phi' = E[\phi']$$

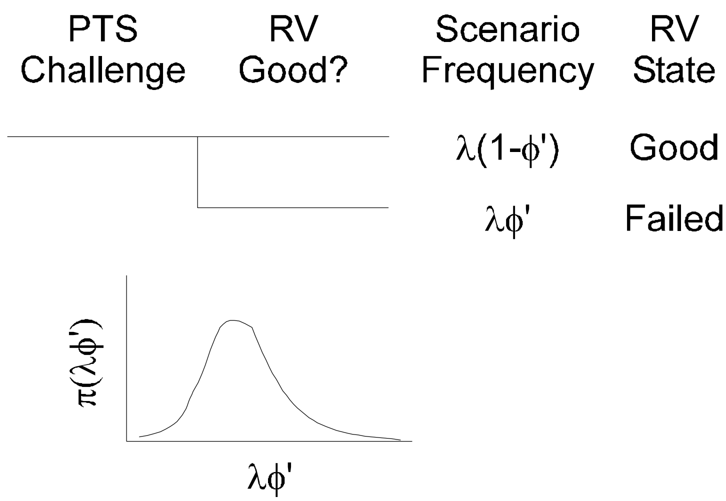


Figure B.3 - Risk Model for General Interpretation of RPV Conditional Failure Probability

It is important to note that this quantity is only used in mean value computations (e.g., when computing the expected number of events, as above, or when computing the mean PTS-induced core damage frequency). When uncertainty analyses are performed, ϕ‘ is the appropriate quantity to use.

A Computational Note

In situations where Monte Carlo sampling is used to address both aleatory uncertainties and epistemic uncertainties, it is still important to treat these two contributions separately. In the example of the PTS-induced RPV failure probability, an appropriate approach is illustrated in Figure B.4. Here, an inner sampling loop is used to estimate ϕ‘, which is conditioned on a number of deterministic (but unknown) parameters, represented by the vector $\underline{\omega}$. (Recall that ϕ‘ quantifies the aleatory uncertainties in RPV failure.) The epistemic uncertainties in the deterministic parameters, represented by the joint distribution $\pi(\underline{\omega})$ are addressed via an outer sampling loop. (This is the so-called “propagation of uncertainties” phase of the PRA.) Failure to properly perform this sampling (e.g., by addressing epistemic uncertainties in the inner loop) will lead to confusion in the interpretation of results.

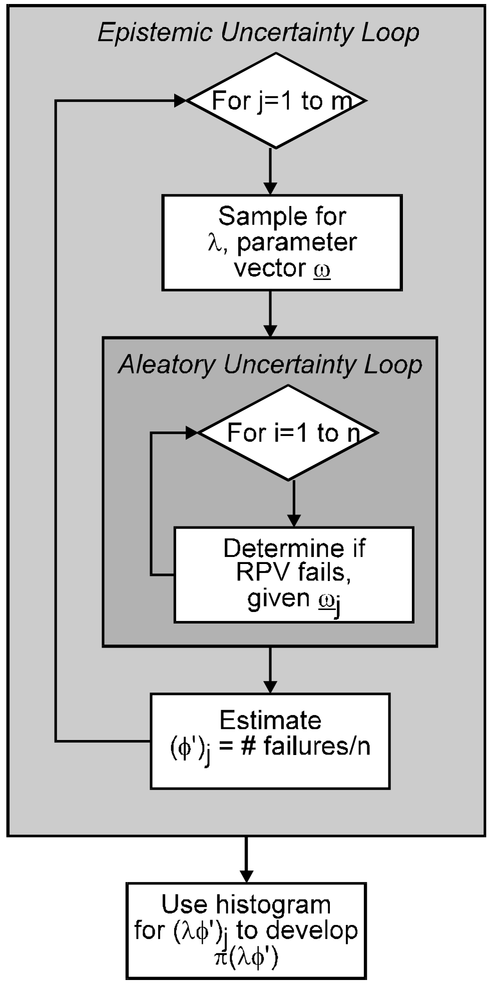


Figure B.4 - Schematic of Sampling Scheme for Addressing Aleatory and Epistemic Uncertainties

Appendix C - Example Application of Bayes' Theorem for A Lognormal Variable

Problem

Consider a situation where C, a random variable, is believed to be lognormally distributed. In other words, the likelihood that C takes on a value in any specified range, e.g., $(c,c+\Delta c)$, is governed by a probability density function (pdf) of the form

$$f(c|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma c}\exp\left[-\frac{1}{2}\left(\frac{\ell n c-\mu}{\sigma}\right)^2\right]$$

where μ and σ are parameters of the distribution. Note that the mean value and variance of C can be determined from μ and σ, if they are known:

$$E[C] = e^{\mu+0.5\sigma^2}$$
$$Var[C] = (E[C])^2\left(e^{\sigma^2}-1\right)$$

In general, μ and σ are not known and must be estimated based on available data.

Assume that there are N data points for C: $\{c_1,c_2,...,c_N\}$. If N is large, μ and σ can be estimated using a number of different methods (e.g., the method of maximum likelihood, Bayes' Theorem). If N is small, Bayes' Theorem provides an appropriate tool. In the case of this example, Bayes' Theorem states that the joint distribution for μ and σ is given by:

$$\pi_1(\mu,\sigma|c_1,\ldots,c_N) = \frac{\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}\sigma c_i}\exp\left\{-\frac{1}{2}\left[\frac{\ell n c_i-\mu}{\sigma}\right]^2\right\}\pi_0(\mu,\sigma)}{\int_{-\infty}^{\infty}\int_{0}^{\infty}\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}\sigma c_i}\exp\left\{-\frac{1}{2}\left[\frac{\ell n c_i-\mu}{\sigma}\right]^2\right\}\pi_0(\mu,\sigma)d\sigma d\mu}$$

where $\pi_0(\mu,\sigma)$ is the joint probability distribution for μ and σ prior to the collection of the data set $\{c_1,c_2,...,c_N\}$. The predictive pdf for C, i.e., the pdf to be used for predictive purposes, is the average lognormal distribution function, where the posterior distribution for μ and σ is used as the weighting function.

$$f(c|c_1,\ldots,c_N) = \int_{-\infty}^{\infty}\int_{0}^{\infty} f(c|\mu,\sigma)\,\pi_1(\mu,\sigma|c_1,\ldots,c_N)d\sigma d\mu$$

Example Application

Consider the following data set:

Table C.1 - Sample Data Set

i	C_i (dimensionless)
1	0.20
2	0.13
3	0.44
4	0.18
5	0.19
Sample Mean	0.228
Sample Variance	0.0118

Using a non-informative prior distribution (in this case, a distribution proportional to $1/\sigma$)[16], Bayes' Theorem and the predictive distribution for C can be readily evaluated using commercial spreadsheet or equation solving software. An example solution using Mathcad 6.0 is attached. The mean, variance, 5th, 50th, and 95th percentiles of the predictive distribution are as follows:

$E[C] = 0.23$
$Var[C] = 0.018$
$C_{05} = 0.076$
$C_{50} = 0.21$
$C_{95} = 0.57$

Computation Notes

1. The Mathcad worksheet has been written for clarity of presentation and not computational efficiency. For example, the integration symbols used in the worksheet invoke the Mathcad-supplied automatic integrator. For the problem of interest, pre-computing the posterior distribution at a specified set of points and using a single-pass trapezoidal integration scheme will lead to results of comparable accuracy with significantly less computation time. (On a 133 MHz Pentium PC, the provided worksheet takes about 5 minutes to solve. With the indicated modifications, only 30 seconds are required for the same problem.)

---

[16]See for example G.E.P. Box and G.C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA, 1973.

2. The maximum likelihood estimates (MLEs) for μ and σ can be found using the sample moments shown in Table C.1 and the relationships between E[C], Var[C], μ, and σ specified earlier. Figure C.1 compares the pdf based on these estimates with the pdf developed using Bayes' Theorem. It can be seen that the MLE-based pdf is narrower; this is because the MLE-based pdf does not account for the uncertainties in μ and σ due to the limited sample size.

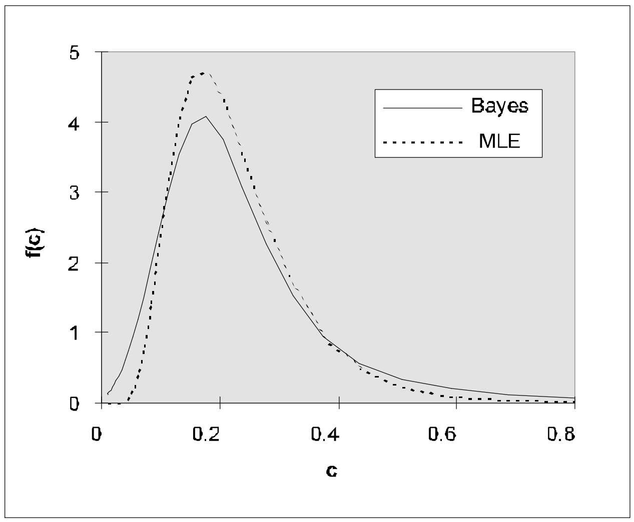


Figure C.1 - Comparison of Bayesian and MLE pdfs

Example Application of Bayes' Theorem for A Lognormal Variable
Mathcad 6.0 Worksheet, Last Revised September 3, 1999

Data

$$C := \begin{bmatrix} 0.20 \\ 0.13 \\ 0.44 \\ 0.18 \\ 0.19 \end{bmatrix} \qquad N := \mathrm{length}(C) \qquad \mathrm{mean}(C) = 0.228$$

$$i := 0 .. N - 1 \qquad \mathrm{var}(C) = 0.0118$$

(Mathcad uses "0" as the first index of a vector/matrix)

Functions

$$\pi0(\mu,\sigma) := \frac{1}{\sigma}$$ (Prior distribution)

$$L(c,\mu,\sigma) := \left(\frac{1}{\sqrt{2\cdot\pi}\cdot c\cdot\sigma}\right)\cdot\exp\left[-0.5\cdot\left(\frac{\ln(c)-\mu}{\sigma}\right)^2\right]$$ (Likelihood function - 1 data point)

$$LN(c,\mu,\sigma) := \prod_{i=0}^{N-1} L\left(c_i,\mu,\sigma\right)$$ (Likelihood function - N data points)

Initial Plot (Unnormalized Posterior Distribution)

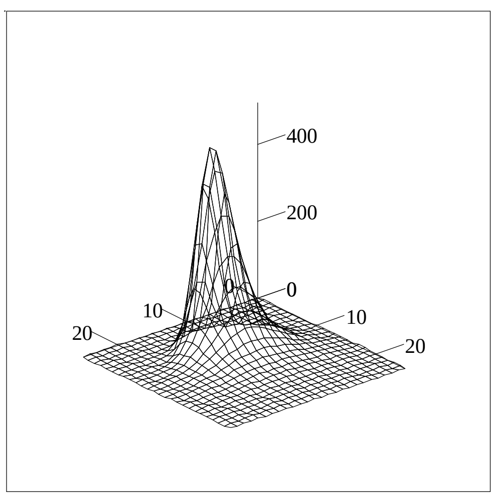
$m := 25 \qquad j := 0 .. m \qquad k := 0 .. m$ (Linear grid for μ and σ)

$\mu min := -2.5 \qquad \mu max := -.5 \qquad \sigma min := .01 \qquad \sigma max := 1.5$

$$\mu_j := \mu min + \frac{j}{m}\cdot(\mu max - \mu min) \qquad \sigma_k := \sigma min + \frac{k}{m}\cdot(\sigma max - \sigma min)$$

$$\pi1u_{j,k} := LN\left(C,\mu_j,\sigma_k\right)\cdot\pi0\left(\mu_j,\sigma_k\right)$$ (Unnormalized posterior distribution)

(A plot of the unnormalized posterior is useful for defining appropriate integration bounds.)



π1u

Bayes' Theorem Integration Constant

$$k := \int_{\sigma min}^{\sigma max}\int_{\mu min}^{\mu max} LN(C,\mu,\sigma)\cdot\pi0(\mu,\sigma)\,d\mu\,d\sigma$$

$k = 85.986$

Normalized Posterior Distribution Function

$$\pi1(\mu,\sigma) := \frac{1}{k}\cdot LN(C,\mu,\sigma)\cdot\pi0(\mu,\sigma)$$

Posterior Distribution Moments

(The moments of μ and σ are not needed to develop the predictive distribution but they can be useful. Note that the calculations would be more efficient if a precalculated posterior and a single-pass numerical integration scheme were used; the Mathcad automatic integrator is used for clarity of presentation.)

$$E\mu := \int_{\sigma min}^{\sigma max} \int_{\mu min}^{\mu max} \mu \cdot \pi 1(\mu, \sigma)\, d\mu\, d\sigma$$

$$E\mu 2 := \int_{\sigma min}^{\sigma max} \int_{\mu min}^{\mu max} \mu^2 \cdot \pi 1(\mu, \sigma)\, d\mu\, d\sigma \qquad Var\mu := E\mu 2 - E\mu^2$$

$$E\sigma := \int_{\sigma min}^{\sigma max} \int_{\mu min}^{\mu max} \sigma \cdot \pi 1(\mu, \sigma)\, d\mu\, d\sigma$$

$$E\sigma 2 := \int_{\sigma min}^{\sigma max} \int_{\mu min}^{\mu max} \sigma^2 \cdot \pi 1(\mu, \sigma)\, d\mu\, d\sigma \qquad Var\sigma := E\sigma 2 - E\sigma^2$$

$$E\mu\sigma := \int_{\sigma min}^{\sigma max} \int_{\mu min}^{\mu max} (\mu \cdot \sigma) \cdot \pi 1(\mu, \sigma)\, d\mu\, d\sigma \qquad Cov\mu\sigma := E\mu\sigma - E\mu \cdot E\sigma$$

$$Corr\mu\sigma := \frac{Cov\mu\sigma}{\sqrt{Var\mu \cdot Var\sigma}}$$

$E\mu = -1.568$ (Mean value of μ)

$Var\mu = 0.063$ (Variance of μ)

$E\sigma = 0.541$ (Mean value of σ)

$Var\sigma = 0.047$ (Variance of σ)

$Cov\mu\sigma = 7.618 \cdot 10^{-4}$ (Covariance of μ and σ)

$Corr\mu\sigma = 0.014$ (Correlation coefficient: μ and σ)

Predictive Distribution for C

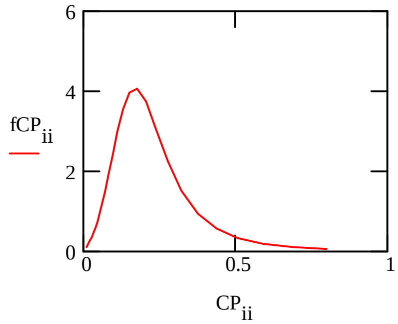
$npoint := 30 \qquad ii := 0 .. npoint - 1$

$CPmin := 0.01 \qquad CPmax := 0.80$

(The pdf is calculated at specified points on a grid. A logarithmic grid is used here.)

$$CP_{ii} := CPmin \cdot \left(\frac{CPmax}{CPmin}\right)^{\frac{ii}{npoint - 1}}$$

$$fCP_{ii} := \int_{\sigma min}^{\sigma max} \int_{\mu min}^{\mu max} \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma \cdot CP_{ii}}} \cdot \exp\left[-.5 \cdot \left(\frac{\ln(CP_{ii}) - \mu}{\sigma}\right)^2\right] \cdot \pi 1(\mu, \sigma)\, d\mu\, d\sigma$$



(Plot of predictive pdf)

Cumulative Distribution and Moments

(Trapezoidal integration is used for ease and efficiency; CDF and moments can also be found using built-in integration functions.)

$jj := 1 .. npoint - 1$

$FCP_0 := 0$

$FCP_{jj} := FCP_{jj-1} + 0.5 \cdot \left(fCP_{jj} + fCP_{jj-1}\right) \cdot \left(CP_{jj} - CP_{jj-1}\right)$

$$ECP := \sum_{jj=1}^{npoint-1} 0.5 \cdot \left(CP_{jj} \cdot fCP_{jj} + CP_{jj-1} \cdot fCP_{jj-1}\right) \cdot \left(CP_{jj} - CP_{jj-1}\right)$$

$$ECP2 := \sum_{jj=1}^{npoint-1} 0.5 \cdot \left[\left(CP_{jj}\right)^2 \cdot fCP_{jj} + \left(CP_{jj-1}\right)^2 \cdot fCP_{jj-1}\right] \cdot \left(CP_{jj} - CP_{jj-1}\right)$$

$VarCP := ECP2 - ECP^2$

$C0 := \min(CP)$

(Use linear interpolation to find percentiles of C)

Given
$\mathrm{linterp}(CP, FCP, C0) = 0.05$
$C05 := \mathrm{Find}(C0)$

Given
$\mathrm{linterp}(CP, FCP, C0) = 0.50$
$C50 := \mathrm{Find}(C0)$

Given
$\mathrm{linterp}(CP, FCP, C0) = 0.95$
$C95 := \mathrm{Find}(C0)$

$ECP = 0.229$ (Mean value of C)
$VarCP = 0.0177$ (Variance of C)
$C05 = 0.076$ (5th percentile of C)
$C50 = 0.209$ (50th percentile of C)
$C95 = 0.571$ (95th percentile of C)

Output Results To File case1.prn

$M_{ii} := \left(CP_{ii} \quad fCP_{ii} \quad FCP_{ii}\right)$

$\mathrm{WRITEPRN}(case1) := M$