A Bayesian Solution to Incompleteness in Probabilistic Risk Assessment

Chris Everett[a], Homayoon Dezfuli[b]
[a] ISL, New York, NY, USA
[b] NASA, Washington, DC, USA

Abstract: The issue of incompleteness is a persistent challenge in probabilistic risk assessment (PRA), where the probabilities of accident consequences such as loss of crew (LOC) are systematically underestimated. The source of the problem is that quantification of the underlying logic model implicitly assumes model completeness when in fact the model represents only the known accident causes, which can be just a small fraction of the total set of causes, especially for new systems. This paper presents a Bayesian approach to logic model quantification, in which the quantification of every event in the model is treated inferentially, based not just on the logical decomposition of the event but also on belief as to the completeness of that decomposition. The result is an assessment methodology that accounts for both known and unknown accident causes, and whose quantitative results can legitimately be said to represent belief about the actual risk of the system. The benefits of the approach are manifold, including improved risk acceptance decision-making, improved risk prioritization, and explicit quantification in risk terms of the value of testing (at any level of integration) and the value of operational successes.

Keywords: PRA, Bayesian Inference, Model Completeness, Total Risk, Risk Acceptance.

1. INTRODUCTION

The issue of incompleteness is a persistent challenge in probabilistic risk assessment (PRA), where the probabilities of occurrence of the end states of concern are systematically underestimated due to the presence of unknown failure causes that are not represented in the failure logic. This paper explores the implications of incorporating unknown failure causes into the failure logic, such that every event in the logic model can occur either as a result of the explicitly modeled antecedents, or as a result of some unknown antecedent. Quantification of such a logic model cannot be wholly done using the traditional "bottom up" approach of standard PRA, since the unknown failure causes have no logic structure and their probabilities of occurrence are therefore not anchored to the probabilities of occurrence of any set of basic events. Instead, quantification is treated as a Bayesian inference problem, where the quantified probability of an event's explicitly modeled antecedents is interpreted as evidence for the probability of the event itself, through a likelihood model that accounts for belief concerning the completeness of the antecedents in addressing the event's causes.

Under this modeling framework it can be seen that traditional PRA represents, epistemically, the special case of complete knowledge of the causes of system failure. However, complete knowledge is seldom, if ever, the case, resulting in a disconnect between PRA results and reality. A more honest accounting of the state of knowledge concerning the failure of the system not only results in credible top event probabilities, but also provides vectors for incorporating information pertaining to the presence of unknown failure causes, such as the failure history of similar systems, failure-free operation of the current system, the technology readiness levels (TRLs) of various subsystems, and the completeness of failure cause identification throughout the logic model.

A failure model that is, by definition, complete, and which can incorporate diverse information, also provides a more powerful basis for risk management than a traditional PRA, for example by identifying areas in the system where unknown failure causes are disproportionately believed to reside, so that resources can be directed towards their discovery and reduction.

2. INCOMPLETENESS OF PRA

As traditionally performed, PRA is a "synthetic" analysis technique, in that it produces risk estimates by explicitly constructing an accident scenario set and aggregating the risk contributions from each scenario to obtain an estimate of risk at the system level. As such, it relies on the ability of the analyst to identify (or bound) all scenarios that can befall the system, so that the system-level risk result represents the complete system risk. One principal technique used in PRA for developing an accident scenario set is to develop a master logic diagram (MLD) of the system, from which initiating events of accidents can be identified. These initiating events are the starting points for the development of accident scenarios, using event sequence diagrams (ESDs), event trees (ETs), and fault trees (FTs) to probabilistically characterize the possible ways that the initiating event might propagate through the system and lead to accident consequences. To quantify the probability of occurrence of each scenario, the events that the scenario entail are decomposed to a level where data exists. To the extent that uncertainties remain in the event probabilities, the probabilities are characterized as epistemically uncertain values rather than as point values, and these uncertainties, which originate at the "bottom" level of scenario decomposition, propagate through the analysis to produce uncertainty in the system-level risk result. The process is illustrated in Figure 1.

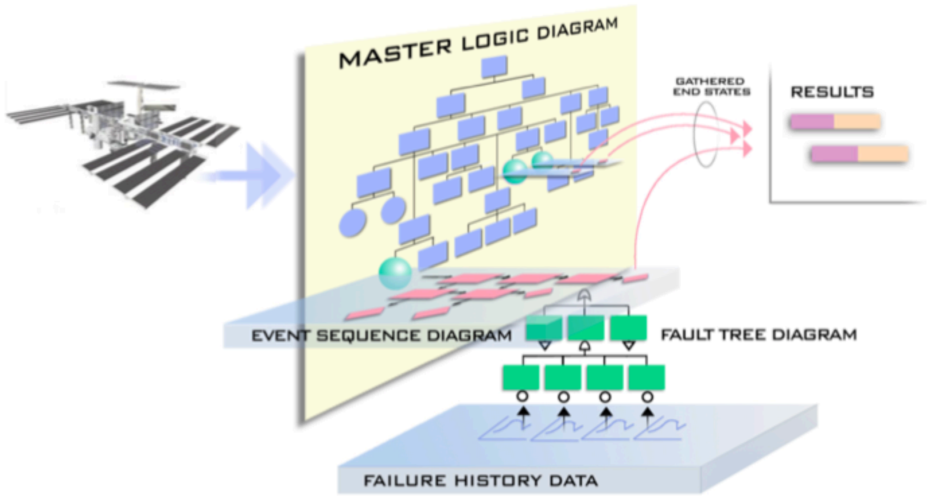

Figure 1. PRA Model Architecture

However, although PRA methods have a history of providing insight into the relative risk significance of potential accident scenarios that might occur in a system, and into the relative safety performance of different systems, it has long been recognized that there are challenges inherent in using synthetic methods such as PRA to quantify a system's actual risk, due to the inherent incompleteness of the scenario sets identified by these methods. The unaccounted-for scenarios typically involve organizational issues and/or complex intra-system interactions that may have little to do with the intentionally engineered functional relationships of the system. Such underappreciated interactions (along with other factors) were operative in both the Challenger and Columbia accidents. In the Challenger disaster, O-ring blow-by impinged on the external tank, leading to tank rupture and subsequent loss of crew. In the Columbia accident, insulating foam from the external tank impacted the wing leading edge reinforced carbon-carbon (RCC), puncturing it and allowing an entryway for hot plasma upon reentry into the Earth's atmosphere. PRA incompleteness is a particular challenge early in the operational life of a system when real-world data is sparse and insight into system risk is most heavily dependent on analysis. This is illustrated in Figure 2, reproduced from Volume 2 of the NASA System Safety Handbook [1], which shows that early in the Shuttle program the Shuttle PRA identified only a small fraction of the risk that was ultimately revealed (red curve vs. green curve).*

* The results shown in Figure 2 leverage the work of Hamlin, et al., as described in [2].

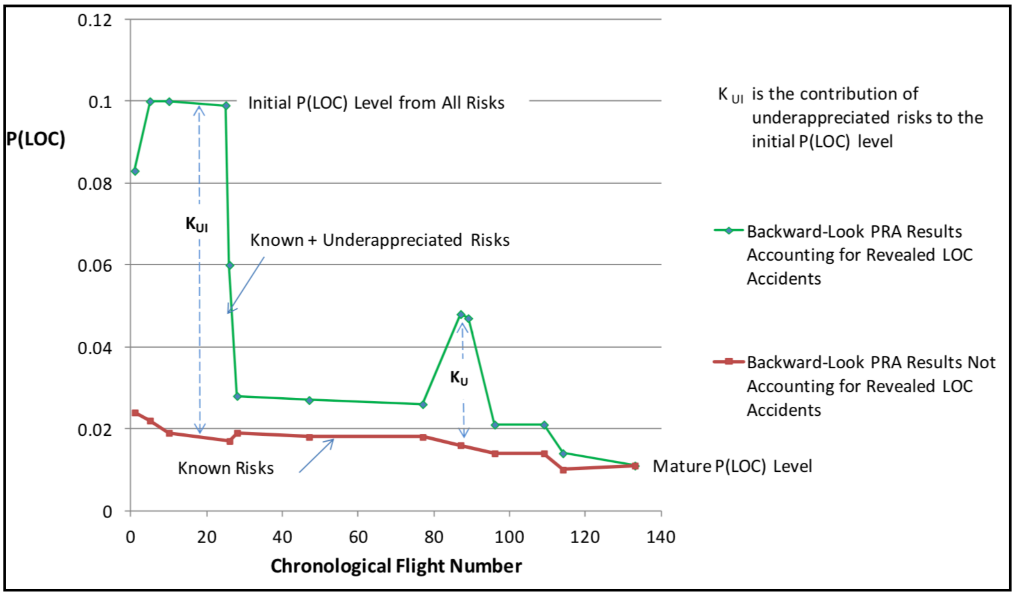

Figure 2. Comparison Between Known Risk and Total Risk for the Space Shuttle

3. TOWARDS A METHOD FOR INCORPORATING UNKNOWN FAILURE CAUSES INTO PRA

3.1. Inclusion of Unknown Failure Causes as Undeveloped Events

This paper explores the idea of incorporating unknown failure causes into PRA through the use of undeveloped events that act as stand-ins for the risk contributions of unidentified accident scenarios. The situation is that of Figure 3, which shows a system fault tree that has been augmented at two levels of system decomposition (System A and Subsystem B) with the inclusion of OR gates that accommodate unknown failure causes of the respective systems. OR gates are appropriate because the unknown failure causes represent unidentified accident scenarios that are in addition to those that are explicitly modeled. This can be seen in Equation 1 by solving the logic model to produce the cut sets that produce system failure. Since each cut set represents a distinct accident scenario, the event, “Unknown Failure Causes of X,” represents an additional clause in the disjunction of otherwise “known” cut sets.

$$\text{System X Fails} = (\text{Cut Set X})_1 \vee \ldots \vee (\text{Cut Set X})_n \vee (\text{Unknown Failure Causes of X}) \qquad (1)$$

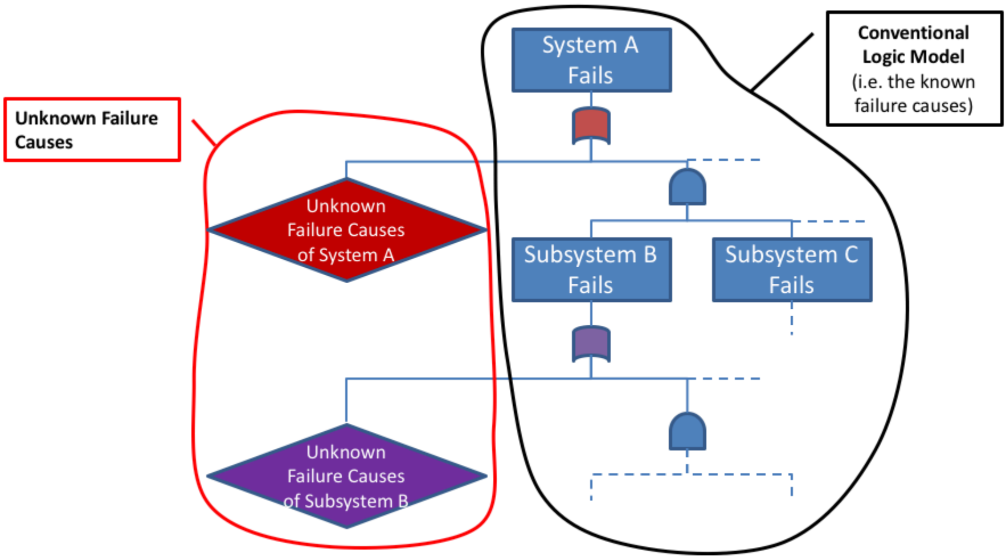
Unknown Failure Causes
Conventional Logic Model (i.e. the known failure causes)
System A Fails
Unknown Failure Causes of System A
Subsystem B Fails
Subsystem C Fails
Unknown Failure Causes of Subsystem B

Figure 3. Incorporation of Unknown Failure Causes into System Fault Tree

In order to quantify the probability of system failure due to unknown causes, the method of this paper depends on the existence of belief about the probability of failure of the system that is not merely the result of synthetic analysis, but which reflects an overall judgment about the probability of failure of the system from any cause, known or unknown, based on familiarity with systems of the type in question. The existence of this system-level belief is not a substantive constraint, since to lack such belief is to have no grounds for even suspecting the possible presence of unknown failure causes. Indeed, given the evolutionary, rather than revolutionary, nature of most engineered systems, there is typically a general consensus as to what constitutes a reasonable belief about system failure probability, even for new systems. For example, given that launch vehicle failure probabilities have historically bottomed out at around 1 in 200 per launch, it would be reasonable to believe that a new launch vehicle design is unlikely to have a failure probability below, say, 1 in 500. Thus, it could be strongly argued that the lack of a mechanism to incorporate multi-level belief about system failure probability is a significant analytical deficiency that squanders important information that might otherwise be brought to bear.

Because the event *System A Fails* is the disjunction of known and unknown causes, the probability that System A fails is:

$$\begin{aligned}\text{Pr(System A Fails)} &= \text{Pr(Known Failure)} + \text{Pr(Unknown Failure)}\\ &\quad - \text{Pr(Known Failure)} \cdot \text{Pr(Unknown Failure | Known Failure)} \qquad (2)\end{aligned}$$

At this point the assumption is made that the known and unknown causes of System A failure are independent of each other. This is justified on the grounds that the known causes of failure will have been investigated to a point where their implications are reasonably well understood, so it is likely that unknown causes of failure are unrelated to them. Thus, equation 2 simplifies to:

$$\begin{aligned}P_T\text{(System A Fails)} &= P_K\text{(System A Fails)} + P_U\text{(System A Fails)}\\ &\quad - P_K\text{(System A Fails)} \cdot P_U\text{(System A Fails)} \qquad (3)\end{aligned}$$

where

P_T is the total probability due to any cause
P_K is the probability due to known causes
P_U is the probability due to unknown causes

Quantification of P_U is illustrated in Figure 4, where, given belief about the probability of failure of System A and the results of a PRA of System A, the probability of failure of System A due to unknown causes is:

$$P_U = (P_T - P_K) / (1 - P_K) \qquad (4)$$

where

$$P_K \leq P_T$$

Figure 5 plots P_U over the range of allowable values of P_K and P_T.

It is clear from Equation 4 and Figure 5 that P_T must be greater than or equal to P_K, since the accident scenarios that contribute to P_K also contribute to P_T. But because the values given to P_K and P_T come from different sources (synthetic analysis vs. prior belief), there is the possibility that this constraint will be violated, indicating conflicting assumptions within the analysis. This issue is problematic when dealing with point values of P_K and P_T, but becomes tractable when epistemic uncertainty concerning P_K and P_T is incorporated into the analysis.

3.2. Accounting for Epistemic Uncertainty

When epistemic uncertainty is considered, P_K and P_T can no longer be characterized as point values, but must be treated as probability density functions, i.e., $f_K(P_K)$ and $f_T(P_T)$. Likewise, P_U is no longer a point value but must also be treated as a probability density function, i.e, $f_U(P_U)$. Moreover, P_K and P_U may be correlated, so they must in general be treated as a joint density function, $f_{KU}(P_K, P_U)$, rather than as separate functions. The analytically modeled density function, $f_K(P_K)$, is now seen as a *marginal* density function of $f_{KU}(P_K, P_U)$, i.e.:

$$f_K(P_k) = \int_0^1 f_{KU}(P_K, P_U) \cdot dP_U \tag{5}$$

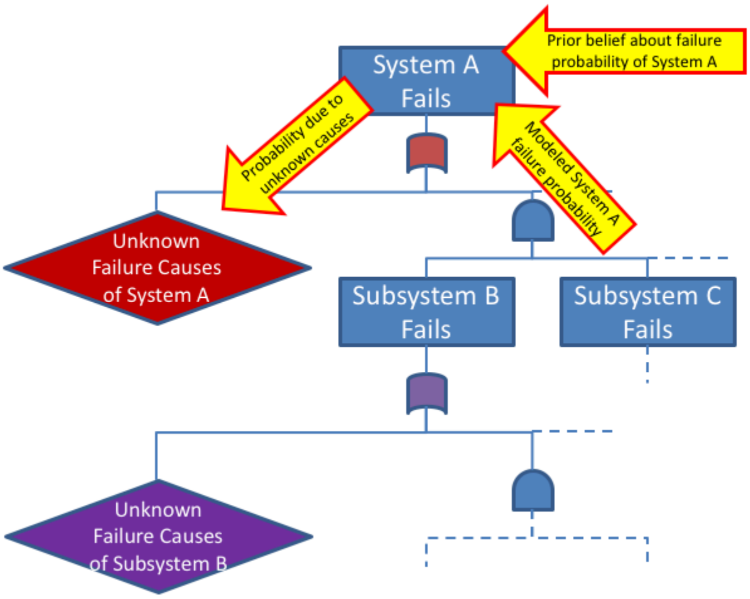
System A Fails
Prior belief about failure probability of System A
Probability due to unknown causes
Modeled System A failure probability
Unknown Failure Causes of System A
Subsystem B Fails
Subsystem C Fails
Unknown Failure Causes of Subsystem B

Figure 4. Quantification of Unknown Failure Causes

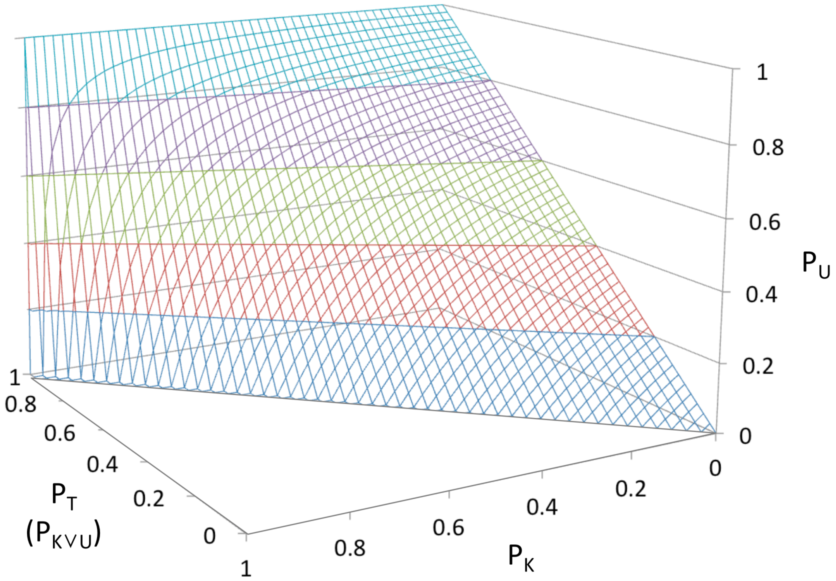

Figure 5. P_U vs. P_K and P_T

The density function for the total probability of failure, $f_T(P_T)$, is now definable as an integral of $f_{KU}(P_K, P_U)$. Equation 6 describes the situation, in which $f_{KU}(P_K, P_U)$ contributes to $f_T(P_T)$ at every combination of P_K and P_U where $P_T = P_K + P_U - P_K \cdot P_U$. For other combinations of P_K and P_U, a delta function ensures that no contribution is made.

$$f_T(P_T) = \iint_{0,0}^{1,1} f_{KU}(P_K, P_U) \cdot \delta(P_T - (P_K + P_U - P_K \cdot P_U)) \cdot dP_K \, dP_U \tag{6}$$

Equation 6 can be simplified to:

$$f_T(P_T) = \int_0^1 f_{KU}(P_K, (P_T - P_K) / (1 - P_K)) \cdot dP_K \tag{7}$$

The task, then, is to deconvolve Equation 7 to solve for $f_{KU}(P_K, P_U)$, from which the marginal density, $f_U(P_U)$, can be projected. This is a non-trivial problem that is beyond the scope of this paper. In general, deconvolution is done through simulation. Closed form derivation is only possible under very severe assumptions (for example, assuming exponential inputs and independence) [3]. However, in the special case of total correlation between P_K and P_U, the math simplifies considerably. The question of correlation is a challenging one, and difficult to address absent the specifics of a system in question. We cannot know what events contribute to P_U, since they are, by definition, unknown, along with other specifics of the unidentified accident scenarios. Under our assumption that they are novel and do not share events with the known scenarios (see Equation 3), an argument can be made that P_U is independent from P_K just as *A Fails due to Unknown Causes* is independent from *A Fails due to Known Causes*. Conversely, assuming that systemic issues such as industrial hygiene or safety management program quality are the principal sources of uncertainty concerning the values of P_U and P_K, then an argument can be made that P_U is highly correlated with P_K. It is under the latter assumption that the following illustration is valid.

3.2.1. Illustration: P_U Totally Correlated with P_K

In the case where P_U is totally correlated with P_K, $P_U|P_K = F_U^{-1}(F_K(P_K))$. Since P_T increases both with increasing P_K and P_U, the immediate implication of this is that F_K, F_U, and F_T always occur at corresponding values of C from their respective distributions, and:

$$P_U = F_U^{-1}(C) = [F_T^{-1}(C) - F_K^{-1}(C)] / [1 - F_K^{-1}(C)] \tag{8}$$

In this case, the inequality $P_K \le P_T$ results in the constraint $F_K^{-1}(C) \le F_T^{-1}(C)$ at all percentiles C, or equivalently, $F_K(P_K) \ge F_T(P_T)$ for all values of P, as illustrated in Figure 6.

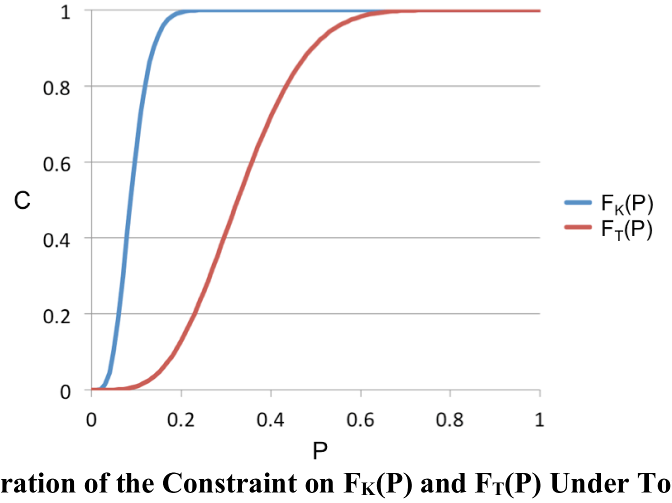

Figure 6. Illustration of the Constraint on $F_K(P)$ and $F_T(P)$ Under Total Correlation

A Monte-Carlo analysis was developed using Microsoft Excel to generate $f_U(P_U)$ and $F_U(P_U)$. The analysis takes, as input, parameters defining beta distributions for P_K and P_T, either in terms of values for the mean and standard deviations or in terms of alpha and beta. It outputs a distribution for P_U, along with a beta distribution fit that matches the mean and standard deviation of the numerical result. Figure 7 and Figure 8 show the results for:

$f_K(P_K) = \text{beta}(3.5, 32)$	(mean = 0.1, SD = 0.05),
$f_T(P_T) = \text{beta}(6, 14)$	(mean = 0.3, SD = 0.1).

The result is a distribution for P_U that fits to:

$$f_U(P_U) = \text{beta}(7.9, 27) \qquad \text{(mean = 0.23, SD = 0.07).}$$

The Monte-Carlo simulation utilized 10,000 trials and 500 histogram bins.

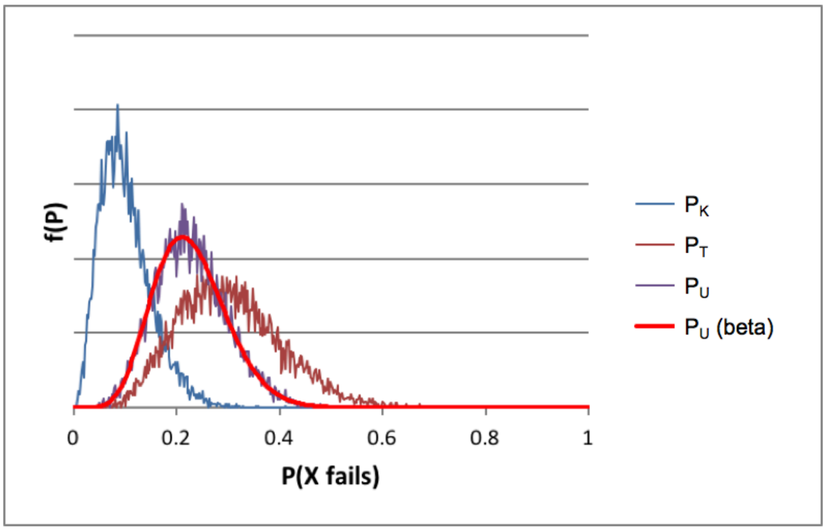

Figure 7. P_K, P_T, and P_U for Correlated P_K, P_U (pdfs)

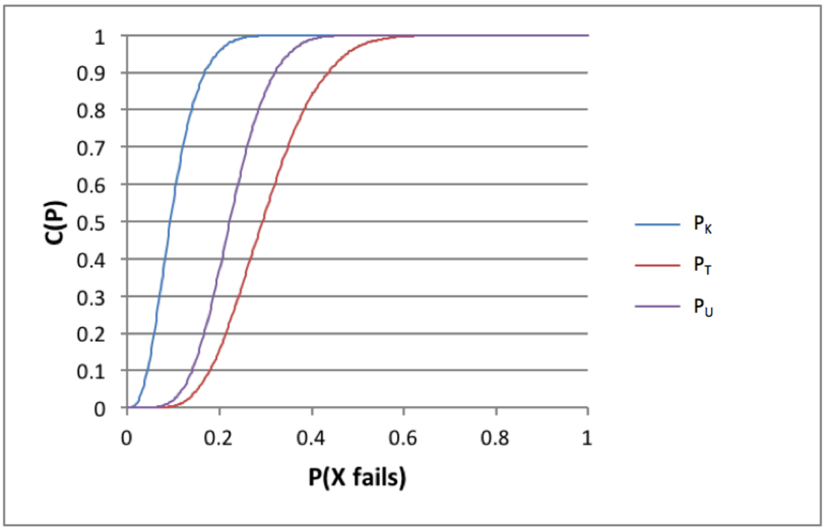

Figure 8. P_K, P_T, and P_U for Correlated P_K, P_U (CDFs)

3.3. Determining $f_T(P_T)$

So far it has been assumed that a probability density function, $f_T(P_T)$, is available for use in Equation 7. Indeed, the premise of this paper is that prior belief about P_T exists and should be incorporated into the PRA along with belief about P_K, as constituted by the results of synthetic analysis. However, there is a difference between $f_T(P_T)$ prior to analysis and $f_T(P_T)$ after analysis, since belief about P_T is affected by knowledge of $f_K(P_K)$. Thus, a Bayesian treatment of P_T is needed in order to get from prior belief about P_T, i.e., $f_T(P_T)$, which is what is elicited from the relevant subject matter experts, to posterior belief about P_T, i.e., $f_T(P_T|f(P_K))$, which is what must be used in Equation 7 in order to be consistent with the presence of $f_K(P_K)$.

Disregarding normalization, Bayes' theorem, applied to the current problem, is:

$$f_T(P_T|f_K(P_K)) \propto f_T(P_T) \cdot L(f_K(P_K)|P_T) \tag{9}$$

where

$f_T(P_T)$ is the prior probability density function of P_T
$L(f_K(P_K)|P_T)$ is the likelihood of $f_K(P_K)$ given P_T

In order to develop a likelihood function for this problem, an *analysis completeness factor* C_A is introduced, that represents the fraction of P_T identified by the synthetic analysis:

$$C_A = P_K/P_T \tag{10}$$

Moreover, since the synthetic analysis is as much an art as a science, and its effectiveness is both hard to quantify and hard to reproduce, C_A is uncertain and therefore properly characterized as a probability density function $f_C(C_A)$.

The strategy used in this paper for developing the likelihood function, $L(f_K(P_K)|P_T)$, is to first develop the likelihood function $L(P_K|P_T, C_A)$, from which $L(f_K(P_K)|P_T, C_A)$ can then be constructed by treating $f_K(P_K)$ as the result of a large number n of individual samples P_{Ki}, each drawn from $f_K(P_K)$.

The likelihood function must account for analysis completeness C_A, and in the case where the analysis is, in fact, complete, it should reproduce $f_K(P_K)$. In other words, the likelihood function should have the property that when $f_C(C_A) = \delta(1 - C_A)$, $L(P_K|P_T, C_A) = f_K(P_K)$.

The following likelihood function satisfies this constraint:

$$L(P_K|P_T, C_A) = f_K(P_K - (E[P_K] - C_A \cdot P_T)) \tag{11}$$

Equation 11 is just $f_K(P_K)$ translated so that it is anchored to the true probability of occurrence of a known failure, $C_A \cdot P_T$, rather than the mean analyzed value, $E[P_K]$. So, for example, in the case where $C_A = E[P_K]/P_T$, the translation is zero and $L(P_K|P_T, E[P_K]/P_T) = f_K(P_K)$. In the case where both parameter uncertainty and analysis incompleteness are eliminated, $C_A = 1$ and $f_K(P_K) = \delta(P_K - P_T)$, and the likelihood function reduces to $L(P_K|P_T, 1) = \delta(P_K - P_T)$, as one would expect.

Now, if $f_K(P_K)$ is seen as consisting of a large number n of individual samples P_{Ki}, each drawn from $f_K(P_K)$, then:

$$L(f_K(P_K)|P_T, C_A) = L(P_{K1} \wedge P_{K2} \wedge \ldots \wedge P_{Kn}|P_T, C_A) = \prod_{i=1}^{n} [L(P_{Ki}|P_T, C_A)] \tag{12}$$

So far, the derivation of the likelihood has been conditioned on a single given value for C_A. However, C_A is uncertain, hence equation 12 must be integrated over all possible values of C_A to produce the final likelihood $L(f_K(P_K)|P_T)$:[†]

$$L(f_K(P_K)|P_T) = \int_0^\infty \{L(f_K(P_K)|P_T, C_A) \cdot f_C(C_A)\} \cdot dC_A \tag{13}$$

Substituting equations 11 and 12 into equation 13 yields:

$$L(f_K(P_K)|P_T) = \int_0^\infty \{\prod_{i=1}^{n} [f_K(P_{Ki} - (E[P_K] - C_A \cdot P_T))] \cdot f_C(C_A)\} \cdot dC_A \tag{14}$$

[†] The upper limit of integration is ∞ rather than 1 to allow for the possibility that P_K overestimates P_T.

3.3.1. Illustration: Bayesian Treatment of P_T

Reconciliation of $f_T(P_T)$ with $f_K(P_K)$ via Bayesian updating was done using the probability density functions from section 3.2.1. A beta distribution for completeness factor C_A was used, with the following values:

$$f_C(C_A) = \text{beta}(6.0, 14) \qquad \text{(mean = 0.3, SD = 0.1)}.$$

Figure 9 shows $f_C(C_A)$.

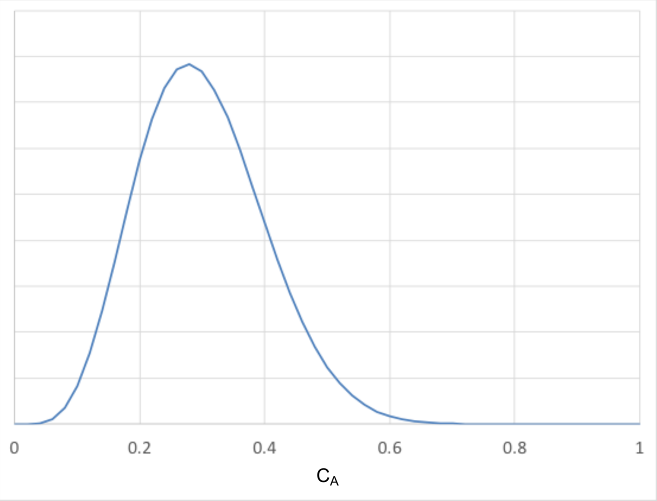

Figure 9. Analysis Completeness, C_A

A likelihood function was developed consistent with Equation 13, with n = 100. The results of updating the density function $f_T(P_T)$ from section 3.2.1 with $f_K(P_K)$ from that same section is shown in Figure 10.

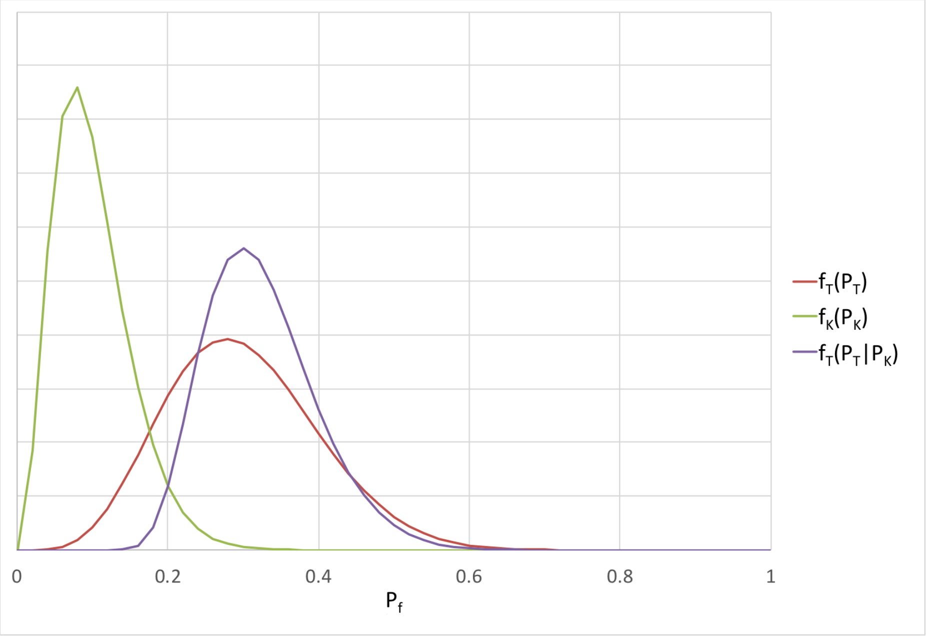

Figure 10. Updating of $f_T(P_T)$ with $f_K(P_K)$

It is clear that in this example, belief about $P_T|P_K$ is substantially different from belief about P_K, and that a naïve use of $f_K(P_K)$ as a proxy for $f_T(P_T|P_K)$ could potentially lead to poor risk-related decision-making and failure to meet system risk expectations.

4. CONCLUSION

This paper points a way towards the explicit incorporation of unknown failure causes into PRA. It does not tackle all the issues, nor all the math, associated with doing so, but it does demonstrate a theoretically well-founded approach. The benefits of incorporating unknown failure causes into the analysis are manifold. First, it results in a "complete" risk model, in that it captures the full scope of belief concerning system failure probability, at any level of logical decomposition where such belief exists. Such a model is appropriate for risk acceptance decision-making in a way that "synthetic-only" PRA is not. In particular, PRA top event probabilities such as the probability of loss of crew, P(LOC), are routinely communicated to decision-makers as uncertainty distributions that reflect only the uncertainties in the basic event probabilities. History is clear that this uncertainty is seldom realistic. All it does is create the false impression that uncertainty has been adequately addressed in the analysis, leading to overconfidence in the use of what history has also shown to be systematically and often unreasonably optimistic risk results. Second, it provides a means of allocating uncertainty throughout the logic model, informing risk management decisions such as margin determination by indicating what parts of the system may be more likely than others to be harboring vulnerabilities. With this additional information it may turn out that the subsystems that dominate total system risk are not the same subsystems that a synthetic-only analysis would identify as the dominant contributors. Finally, it results in a model with the capability to incorporate diverse information such as successful operating experience, which reduces the likelihood that high-probability unknown failure causes are lurking in the system but has little to no effect on the results from a standard PRA.

It might be argued that this approach places an undue data burden on the analyst, since it requires prior distributions on all the event probabilities in the logic model, as well as distributions for the completeness of the logical decomposition of all but the basic events. However, when standard PRA is seen as a special case of this more general method, it becomes clear that the response of standard PRA to this legitimate data burden is to make the most optimistic assumption possible, namely that the logical decomposition is complete and that knowledge of failure causes is total. This is certainly unfounded. A better default approach would be to begin with non-informative priors and update them with whatever diverse evidence can be brought to bear, such as the histories of similar systems, expert opinions, TRL analyses, etc.

Acknowledgements

The authors would like to thank the NASA Office of Safety and Mission Assurance (OSMA) for their sponsorship of the work underlying this study.

References

[1] NASA/SP-2014-612, *NASA System Safety Handbook Vol. 2*, NASA, November 2014.
[2] T. Hamlin, et al., "Shuttle Risk Progression: Use of the Shuttle PRA to Show Reliability Growth," AIAA SPACE Conference & Exposition. 2011.
[3] Correspondence between C. Everett and M. Modarres, University of Maryland, November 29, 2017.