Information management for estimating system reliability using imprecise probabilities and precise Bayesian updating

J.M. Aughenbaugh*

Applied Research Laboratories,
University of Texas at Austin,
Austin, TX 78713, USA
Email: jason@arlut.utexas.edu
*Corresponding author

J.W. Herrmann

Department of Mechanical Engineering and
Institute for Systems Research,
University of Maryland,
College Park, MD 20742, USA
Email: jwh2@umd.edu

Abstract: Engineering design decision-making often requires estimating system reliability based on component reliability data. Although this data may be scarce, designers frequently have the option to acquire more information by expending resources. Designers thus face the dual questions of deciding how to update their estimates and identifying the most useful way to collect additional information. This paper explores the management of information collection using two approaches: precise Bayesian updating and methods based on imprecise probabilities. Rather than dealing with abstract measures of total uncertainty, we explore the relationships between variance-based sensitivity analysis of the prior and estimates of the posterior mean and variance. By comparing different test plans for a simple parallel-series system with three components, we gain insight into the tradeoffs that occur in managing information collection. Our results show that to consider the range of possible test results is more useful than conducting a variance-based sensitivity analysis.

Keywords: reliability assessment; imprecise probabilities; information management.

Reference to this paper should be made as follows: Aughenbaugh, J.M. and Herrmann, J.W. (2009) 'Information management for estimating system reliability using imprecise probabilities and precise Bayesian updating', *Int. J. Reliability and Safety,* Vol. 3, Nos. 1/2/3, pp.35–56.

Biographical notes: Jason Matthew Aughenbaugh is currently a Research Associate for the Applied Research Laboratories at the University of Texas at Austin. He received his MSME and PhD from the George W. Woodruff School of Mechanical Engineering at the Georgia Institute of Technology, where he was a National Science Foundation Graduate Research Fellow from 2003 to 2006. He also holds a BSE in operations research and financial engineering from Princeton University. He is a member of INFORMS and ASME.

Jeffrey W. Herrmann is an Associate Professor at the University of Maryland, where he holds a joint appointment with the Department of Mechanical Engineering and the Institute for Systems Research. He is the Director of the Computer Integrated Manufacturing Laboratory and Associate Director for the University of Maryland Quality Enhancement Systems and Teams (QUEST) Honours Fellows Programme. He is a member of INFORMS, ASME, IIE, SME and ASEE. He earned his BS in Applied Mathematics from the Georgia Institute of Technology. As a National Science Foundation Graduate Research Fellow from 1990 to 1993, he received his PhD in Industrial and Systems Engineering from the University of Florida.

1 Introduction

Modelling uncertainty is an important activity in engineering applications. As engineers progress from deterministic approaches to non-deterministic approaches, the question of how to model the uncertainty in the non-deterministic approaches must be answered. Researchers have proposed various methods for modelling and propagating uncertainty, with developing and applying individual methods and others developing philosophical debates of the appropriateness of methods. Recently, there is a growing interest in practical comparisons of methods (Oberkampf et al., 2001; Nikolaidis et al., 2004; Soundappan et al., 2004; Aughenbaugh and Paredis, 2006; Hall, 2006; Kokkolaras et al., 2006; Aughenbaugh and Herrmann, 2008; Aughenbaugh and Herrmann, 2009).

Most of this work has focused on what we will call the *problem solution* stage of engineering decisions. In this stage, the engineer makes a decision about a product's design. For example, the engineer determines the dimensions of a component or chooses a particular architecture for the system. This stage follows and is distinct from the *problem formulation* phase, which includes tasks such as identifying design alternatives, eliciting stakeholder preferences and modelling the state of the world. In this phase, engineers also perform *information management*; they make decisions about what information to collect, how to collect it and how to process it. For example, the design of experiments falls into this stage. This paper focuses on modelling uncertainty to support information management decisions, with the specific application of a system reliability assessment, as discussed in Section 2.

Managing information collection is related to the concept of the value of information. This concept has been used in the context of engineering design for incorporating the cost of decision-making (Gupta, 1992), for model selection (Radhakrishnan and McAdams, 2005) and for catalogue design (Bradley and Agogino, 1994). Some recent work has considered this problem from a frequent updating perspective (Ling et al., 2006) and developed a method for managing multiple sources of information in engineering design using imprecise probabilities (Schlosser and Paredis, 2007) using the principles of *information economics* (Howard, 1966; Matheson, 1968; Marschak, 1974; Lawrence, 1999). At a basic level, these principles state that one should explicitly consider the expected net value of information.

Ideally, the value of information would be measured in terms of the value of the final product and the cost of the design process. However, such value and cost models are not always available, particularly early in the design process when the design is only very vaguely defined (Malak Jr. et al., 2009). It is thus important to have some statistical metrics for guiding information collection that are independent of the value context of the problem, while still adequately accounting for the information state and the known structure of the system being designed.

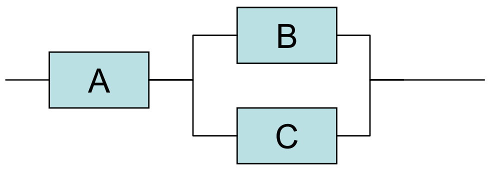
This paper presents and demonstrates approaches for evaluating information collection plans using metrics of uncertainty. In addition, we will compare the types of results that these different approaches give. These results can be used as the input to standard approaches for decision-making under uncertainty, including those for determining the economic value of information. The integration with such approaches we leave for future work. Therefore, we will focus on the required methods and demonstrating them with metrics that display the range of results.

This paper explores uncertainty metrics for information management by comparing different test plans for a simple parallel-series system with three components, conducting a variance-based sensitivity analysis of the prior, and estimating the posterior mean and variance. Section 2 presents the example problem. Section 3 reviews the precise and imprecise Bayesian statistical models. Section 4 discusses uncertainty metrics. Experimental results are presented and discussed in Section 5. Section 6 gives a general discussion and Section 7 concludes the paper.

2 Example problem description

We consider the case of a designer who is considering additional testing of key components in a system in order to improve estimates of the system reliability. From a reliability perspective, the example system can be modelled as the parallel-series system shown in Figure 1. We assume that the failures of each component are independent events. The designer has some vague prior information about the reliability of each component and hopes that additional testing will refine the estimate of system reliability.

Figure 1 Reliability block diagram for the system (see online version for colours)



It will be convenient to frame things in terms of failure probability instead of reliability. A component *i* with reliability R_i has a failure probability of $P_i = 1 - R_i$. Let θ be the failure probability of the system, which is the parameter of interest. Ideally, the designer would have enough data to make a precise assessment of P_A, P_B and P_C, such as '$P_A = 0.05$' and could calculate θ precisely. However, there are practical reasons why the designer cannot or is unwilling to make a precise assessment despite holding some initial beliefs about failure probabilities (Weber, 1987; Coolen, 2004; Groen and Mosleh, 2005; Malak Jr. et al., 2009).

The mathematical model for the reliability of the system shown in Figure 1 follows.

$$\theta = P_{sys} = P_A + P_B P_C - P_A P_B P_C \tag{1}$$

$$E[\theta] = E[P_A] + E[P_B]E[P_C] - E[P_A]E[P_B]E[P_C] \tag{2}$$

$$\begin{aligned} E[\theta^2] = E[P_A{}^2] + 2E[P_A]E[P_B]E[P_C] - 2E[P_A{}^2]E[P_B]E[P_C] + \\ + E[P_B{}^2]E[P_C{}^2] - 2E[P_A]E[P_B{}^2]E[P_C{}^2] + E[P_A{}^2]E[P_B{}^2]E[P_C{}^2] \end{aligned} \tag{3}$$

$$V[\theta] = E[\theta^2] - (E[\theta])^2 \tag{4}$$

We consider the case in which only component testing is feasible. Since testing requires resources, it is not reasonable to test every component a large number of times. In this work, we consider the number of tests as a surrogate for the cost of testing, which is reasonable if all tests require roughly the same amount of resources. The designer's information management problem is to decide which components to test in order to reduce the uncertainty of the system reliability estimate.

3 Formalisms for modelling uncertainty

In this paper we will compare the precise Bayesian and the imprecise beta model approaches for updating reliability assessments. Some introductory material is provided here. For complete discussion, see the cited references and the discussion in Aughenbaugh and Herrmann (2008, 2009).

3.1 Precise Bayesian

The Bayesian approach (e.g. Box and Tiao, 1973; Berger, 1985) provides a way to combine existing knowledge and new knowledge into a single estimate by using Bayes's Theorem. One requirement of Bayesian analysis is a prior distribution that will be updated. The objective selection of a prior distribution in the absence of relevant prior information is a topic of extensive debate. However, in many engineering problems, designers have prior information such as data and experience from similar systems. The Bayesian approach allows this information to be included in the analysis.

The form of the prior is often restricted for computational reasons to *conjugate distributions*, in which case the posterior distribution has the same type (e.g. Gaussian) as the prior. For the binomial experiments considered in this example, it is convenient to model the unknown failure probabilities P_A, P_B and P_C with beta distribution. If the prior distribution is $Beta(\alpha_0, \beta_0)$ and one observes m failures out of n trials, then the posterior distribution is $Beta(\alpha_0 + m, \beta_0 + n - m)$. Consequently, the update involves only addition and subtraction, an enormous computational savings over the general case.

3.2 Robust Bayesian and imprecise probabilities approaches

Two alternatives to a precise Bayesian approach are the robust Bayesian and imprecise probabilities approaches. The two approaches can be mathematically similar, but they have different motivations.

The robust Bayesian approach addresses the problem of lack of confidence in the prior (Berger, 1984; Berger, 1985; Berger, 1993; Insua and Ruggeri, 2000). The core idea of the approach is to perform a 'what-if' analysis by considering different priors. The analyst considers several reasonable prior distributions and performs the update on each to get a set of posterior distributions. If there is no significant change in the decision solution across these posteriors, then it is *robust* to the selection of the prior.

This analysis is not possible with a single prior. For example, many Bayesians propose a uniform distribution when there is no prior information. This confounds two information scenarios: first, that nothing is known; second, that all failure probabilities between 0 and 1 are equally likely, which is actually substantial information.

In a large engineering project in which many individuals are cooperating concurrently, this is an important distinction. For example, suppose that one engineer (Designer A) has stated that two values for a system input are equally likely. Later in the design process, Designer B is working on a component that is very sensitive to this input. To Designer B, the following scenarios are very different: (1) Designer A narrowed the feasible inputs to these two values but had no further information, so he used a maximum entropy approach and assigned them equal probabilities; (2) Designer A performed a detailed analysis that concluded these two inputs occur with equal frequency of 0.5.

In the first scenario, it could be valuable for Designer B to collect additional information about the relative likelihoods of the two inputs before making expensive design modifications to accommodate both equally. In the second scenario, this analysis has already been completed. However, the mathematical representation–the precise probabilities–makes no distinction. This hinders information management decisions. The Robust Bayesian approach affords the design team the opportunity to make different, more appropriate decisions under the two scenarios.

Traditionally, the probability of an event is defined by a single number. However, various researchers find this definition too limiting when there exists incomplete or conflicting information. They have proposed theories of imprecise probabilities in which probabilities can be intervals or sets (Dempster, 1967; Shafer, 1976; Walley, 1991; Weichselberger, 2000) or alternatives to probability (Dubois and Prade, 1988). Imprecise probabilities have been considered in design decisions and set-based design (Aughenbaugh and Paredis, 2006; Malak Jr. et al., 2009) and in reliability analysis (Coolen, 1994; Coolen, 2004; Utkin, 2004a; Utkin, 2004b). For the problem of updating beliefs, imprecise probability theory essentially allows prior and posterior beliefs to be expressed as sets of density functions.

The admissibility of imprecision in beliefs is the primary difference in motivation between imprecise probabilities and robust Bayesian approaches. The imprecise probability view is that the analyst's beliefs can be imprecise whereas the robust Bayesian view is that there exists a single prior that captures the analyst's true beliefs perfectly, although it may be hard to identify this distribution in practice. Either motivation can lead to the consideration of sets of priors and posteriors.

For this problem, it is convenient to use the Imprecise Beta Model (IBM) (Walley, 1991; Walley et al., 1996). The IBM is a special case of imprecise probabilities. The motivation for its use is analogous to the precise Bayesian practice of using a conjugate distribution to simplify computations, even though some expressivity is lost. Improvements to the IBM have been suggested (Coolen, 1994), and non-parametric methods are also available (Coolen, 1998; Coolen-Schrijner and Coolen, 2007). Here, it is consistent to use the IBM to represent both imprecise probability distributions and a set of prior distributions and to apply Bayes's Theorem to update the IBM to determine the set of posterior distributions. However, not all methods for imprecise probabilities are compatible with robust Bayesian techniques.

The IBM re-parameterises the beta distribution so that the density of $beta(s,t)$ is as given in Equation (5) (Walley, 1991; Walley et al., 1996).

$$\pi_{s,t}(\theta) \propto \theta^{st-1}(1-\theta)^{s(1-t)-1} \tag{5}$$

To transform this to the standard $beta(\alpha,\beta)$ parameterisation, let $\alpha = s \cdot t$ and $\beta = s \cdot (1-t)$, which is equivalent to $s = \alpha + \beta$ and $t = \alpha/(\alpha+\beta)$. This parameterisation is convenient because t is the mean of the distribution. In addition, s describes how much information the prior represents. The model is updated as follows: if the prior parameters are s_0 and t_0 and m failures are observed in n trials, then the parameters of the posterior distribution are $s_n = s_0 + n$ and $t_n = (s_0 t_0 + m)/(s_0 + n)$, as derived from the relationship to the standard parameterisation using α and β.

Following Walley (1991), the parameters *t* and *s* can be imprecise. For example, the priors can be expressed as intervals $\left[\underline{t}_0, \overline{t}_0\right]$ and $\left[\underline{s}_0, \overline{s}_0\right]$ describing bounds on the mean and amount of prior information respectively. The priors are the set of beta distributions with $\alpha_0 = s_0 t_0$ and $\beta_0 = s_0\left(1-t_0\right)$ such that $\underline{t}_0 \le t_0 \le \overline{t}_0$ and $\underline{s}_0 \le s_0 \le \overline{s}_0$. After the test results are observed, each precise prior in the set is updated as described above. The posterior bounds on s are given by $\underline{s}_n = \underline{s}_0 + n$ and $\overline{s}_n = \overline{s}_0 + n$. The bounds on the mean estimate are given by:

$$\underline{t}_n = \min_{\underline{s}_0 \le s_0 \le \overline{s}_0} \{(s_0 \underline{t}_0 + m)/(s_0 + n)\} \tag{6}$$

$$\overline{t}_n = \max_{\underline{s}_0 \le s_0 \le \overline{s}_0} \left\{(s_0 \overline{t}_0 + m)/(s_0 + n)\right\} \tag{7}$$

It should be noted that the two intervals define a superset of the actual posterior region. The true posterior region is found by updating each distribution in the prior set, not by considering the bounds on the parameters.

4 Metrics of uncertainty

If the designer models the system performance (e.g. system failure probability) as a precise probability distribution, then the mean, variance and other statistics about that distribution are specific numbers for the prior distribution. If n identical tests are conducted, and each test result is a pass or fail, then there are $n+1$ possible test results. Characterising the quality of a test plan (which has uncertain outcomes) is a classic problem in decision-making and some decision-makers will want to know the complete distribution of outcomes, some will want the worst-case and others will want the 'average' value of a statistic.

Modelling the system reliability with an imprecise probability distribution introduces an additional complexity: now the mean, variance and other statistics about that distribution are imprecise. If *n* identical tests are conducted, and each test result is a pass or fail, then there are $n+1$ possible imprecise posterior distributions, with the corresponding $n+1$ sets of means, $n+1$ sets of variances and so on. Characterising any specific result requires some way to describe the set.

For both precise and imprecise priors, we will consider two different strategies. The first is a variance-based sensitivity analysis of the prior distribution, which allows one to ignore the possible test results. The second considers the possible outcomes of a test plan.

4.1 Metrics for precise distributions

We will begin by considering the case of precise prior distributions.

4.1.1 Variance-based sensitivity analysis

One can avoid the problem of considering a large number of possible test results by focusing on the current state of information. In variance-based sensitivity analysis, one determines how each input variable contributes to the total system variance (Sobol, 1993; Chan et al., 2000). The sensitivity of the system performance to an input variable X_i is described by the sensitivity index SV_i, defined as the ratio of the variance of the conditional expectation for that input to the total variance. A large sensitivity index indicates that reducing the variance of that variable can reduce the system variance relatively more than other variables and it suggests that a test plan should focus on reducing that input variable's variance.

In the case considered here, the failure probabilities of the three components (P_A, P_B and P_C) are the input random variables and the failure probability of the system is the system performance (or output random variable). In particular, we can calculate the sensitivity indices for our example system as follows:

$$\begin{aligned} SV_A &= \left(\left(1-E\left[P_B\right]E\left[P_C\right]\right)^2 V\left[P_A\right]\right)/V\left(\theta\right) \\ SV_B &= \left(\left(E\left[P_C\right]\right)^2\left(1-E\left[P_A\right]\right)^2 V\left[P_B\right]\right)/V\left(\theta\right) \\ SV_C &= \left(\left(E\left[P_B\right]\right)^2\left(1-E\left[P_A\right]\right)^2 V\left[P_C\right]\right)/V\left(\theta\right) \end{aligned} \tag{8}$$

4.1.2 Observing mean and variance for different results

The variance of the probability distribution is one measure of uncertainty. In general, a distribution with smaller variance means that there is less uncertainty about the parameter it models. For the problem of test planning, we may hope to conduct tests that will yield a posterior distribution with a variance that is smaller than some threshold. Pham-Gia and Turkann (1992) derived lower bounds on the number of samples needed to satisfy an upper bound on the posterior variance for a beta distribution. Unfortunately, this result is not directly applicable to this system-level example problem.

A test plan conducts n_i tests of component i and is summarised as $T = \{n_A, n_B, n_C\}$. If there is x_i failures of component i, then the posterior distributions of the component failure probabilities are $P_i \sim beta(\alpha_i + x_i, \beta_i + n_i - x_i)$. These can be used to calculate the mean and variance of the system failure probability as discussed in Section 2. Of course, this must be repeated for each of the $\left(n_A+1\right)\times\left(n_B+1\right)\times\left(n_C+1\right)$ possible test results.

4.2 Metrics of uncertainty for imprecise distributions

One of the motivations for using imprecise probabilities instead of precise probabilities is that they allow the total uncertainty to be captured more adequately by separating imprecision and probability. If the variance generally captures the variability, the natural question follows: *how can imprecision be measured*? Or more generally, how can we measure the total uncertainty? This issue has been pursued by various authors, as summarised by Klir and Smith (2001). In short, the search for a single, useful measure of total uncertainty has been largely unsuccessful. We begin our examination of the problem by considering the extension of precise measures to the imprecise case.

4.2.1 Imprecise variance-based sensitivity analysis

Hall (2006) extends variance-based sensitivity analysis to imprecise probability distributions. The generalisation from the precise case is to consider the minimum and maximum sensitivity indices across the set of input distributions. Let F be the set of input distributions (jointly across all inputs). Let $SV_{i,p}$ be the sensitivity to input i given the input distribution p. Then, the bounds are given by:

$$\underline{SV}_i = \min_{p \in F}\left(SV_{i,p}\right), \qquad \overline{SV}_i = \max_{p \in F}\left(SV_{i,p}\right) \tag{9}$$

The difficulty in calculating these is the need to optimise over the set F. In the case considered here, each input distribution p is a joint distribution over the component failure probabilities. Each marginal distribution comes from the imprecise prior distribution for that component. We will use a numerical approach that selects distributions from the set F in the following way. First, we select a parameter N_e that determines the number of intermediate values for each parameter. For parameter s_A, we calculate the following set of values:

$$\left\{\underline{s}_{0A}, \underline{s}_{0A} + \frac{1\left(\overline{s}_{0A} - \underline{s}_{0A}\right)}{N_e + 1}, \underline{s}_{0A} + \frac{2\left(\overline{s}_{0A} - \underline{s}_{0A}\right)}{N_e + 1}, \ldots \underline{s}_{0A} + \frac{N_e\left(\overline{s}_{0A} - \underline{s}_{0A}\right)}{N_e + 1}, \overline{s}_{0A}\right\} \tag{10}$$

This yields $N_e + 2$ values for this parameter. We repeat for the other five parameters. Then, we take all combinations, which yield $\left(N_e + 2\right)^6$ joint prior distributions. Due to monotonicity in some aspects of this problem, much of our analysis can be done using just the extreme points. However, for complete results in later analysis, $N_e = 3$ was determined to be adequate for this problem. Increases in N_e or the number of components in the system will require a more complex parameter sampling scheme or the use of alternative methods due to the explosion in total number of samples required. However, alternative methods have their own difficulties. For example, probability bounds analysis (Ferson and Tucker, 2006) reduces the computational cost but loses some accuracy due to repeated variables in the problem. The application of imprecise methods to complex problems remains an area of research.

4.2.2 Dispersion of mean and variance

Given an imprecise prior, a specific test result will yield an imprecise posterior distribution. The dispersion of the mean and variance (over the possible test results) is no longer a sequence of points, as in the precise case of Section 4.1.2; it is instead a sequence of sets of mean-variance pairs. Given imprecise priors for the failure probabilities of the three components, we can compare different test plans (e.g. test only Component A or test only Component B) and determine how they affect the dispersion of the mean and variance for all outcomes.

As before, let F be the entire set of prior joint distributions for the component failure probabilities, and consider a test $T = \{n_A, n_B, n_C\}$. If there is x_i failures of component i, then this one result yields a set $F'(x_A, x_B, x_C, n_A, n_B, n_C)$ of posterior distributions. A different result yields a different F'. Each posterior $p' \in F'(x_A, x_B, x_C, n_A, n_B, n_C)$ is determined by updating a prior distribution $p \in F$ as described in Section 4.1.2. From the posterior, one can calculate the mean and variance of the system failure probability as discussed in Section 2. We will select distributions from F as described in Section 4.2.1.

4.2.3 Imprecision in the mean

A fundamental measure of imprecision is the range of the mean value across the set of probability distributions. For the IBM, this is simply $\overline{t} - \underline{t}$. Each possible result of a test plan that conducts a total of n tests will yield a set F' of posterior distributions, and each set has a particular imprecision of the mean associated with it, given by:

$$\Delta_{n,F'}(\theta) = \max_{p' \in F'} E\left[\theta \middle| p'\right] - \min_{p' \in F'} E\left[\theta \middle| p'\right] \tag{11}$$

In this example, monotonicity in the mean parameter allows simplification to:

$$\Delta_{n,F'}(\theta) = \left(\overline{t}_{nA} + \overline{t}_{nB}\overline{t}_{nC} - \overline{t}_{nA}\overline{t}_{nB}\overline{t}_{nC}\right) - \left(\underline{t}_{nA} + \underline{t}_{nB}\underline{t}_{nC} - \underline{t}_{nA}\underline{t}_{nB}\underline{t}_{nC}\right) \tag{12}$$

The maximum posterior imprecision over all possible F' (that is, over all possible results for this test plan, $0 \le x_A \le n_A$, $0 \le x_B \le n_B$, $0 \le x_C \le n_C$) can be denoted as follows:

$$\Delta_{\max}(\theta) = \max_{F'}\left\{\Delta_{n,F'}(\theta)\right\} \tag{13}$$

Given a prior, one can also calculate the expected posterior mean. Let the posterior mean be $\mu(x, n, p_0(\cdot))$, which depends upon the prior $p_0(\cdot)$, the test setup (n) and the test result (x). One can calculate an average posterior mean for that prior and setup across possible results as $\tilde{\mu}_{p_0} = E_{p(x)}\left[\mu(x, n, p_0)\right]$, and then find the minimum and maximum of $\tilde{\mu}_{p_0}$ across all priors $p_0 \in F$.

4.2.4 Imprecision in the variance

The designer would like the variance to be as small as possible, but the variance for each result is now an interval. One can calculate the minimum and maximum expected variance across all the priors. One can also consider measuring the imprecision using the range of the variance across the set of posterior distributions, which reflects how well the variance is known. Ideally, an analyst would pick a test design that will result in a posterior variance that tends to be low and well known. The posterior imprecision in the variance given a particular result and the maximum imprecision over all results are given in Equations (14) and (15). We approximate each F' using the procedure described in Section 4.2.1.

$$\Delta_{n,F'}(V) = \max_{p' \in F'} V\left[\theta \middle| p'\right] - \min_{p' \in F'} V\left[\theta \middle| p'\right] \tag{14}$$

$$\Delta_{\max}(V) = \max_{F'} \left\{\Delta_{n,F'}(V)\right\} \tag{15}$$

5 Results

In this paper, we are using component test plans as an example of information collection. A test plan is a particular scheme for evaluating component reliability. For example, one test plan is to sample 6 units of Component A, 4 units of Component B and 2 units of Component C. For brevity in subsequent text, such a test plan will be referred to as $[6,\ 4,\ 2]$. For the example system of Section 2, we now consider specific scenarios for demonstrating the uncertainty metrics described in Section 4.

5.1 Scenario 1

In the first scenario, the priors for the failure probability distributions are precise beta distributions. The parameters are shown in Table 1.

Table 1 Priors for Scenario 1

Component	*A*	*B*	*C*
Precise Beta parameters	$t_0 = 0.15$ $s_0 = 10$	$t_0 = 0.15$ $s_0 = 2$	$t_0 = 0.15$ $s_0 = 10$

5.1.1 Scenario 1: variance-based sensitivity analysis

The variance-based sensitivity analysis gives the following values: $SV_A = 0.8982$. $SV_B = 0.0560$ and $SV_C = 0.0153$. These values suggest that testing Component A, which should reduce $V(P_A)$, will have the most impact on $V(\theta)$.

5.1.2 Scenario 1: observing mean and variance for different results

Figure 2 shows the dispersion of the posterior mean and variance of the system failure probability distribution for seven different test plans. Although the prior distributions for the failure probabilities for Components A and C are the same, testing Component A (which is essential for system operation and has a much greater sensitivity index) makes a bigger change in $E[\theta]$ and $V(\theta)$. Test plans 1, 4, 5 and 6 all have low minimum $V(\theta)$, while test plan 1 also has a low maximum $V(\theta)$. Components B and C have the same position in the system, but Component B has a smaller *s* parameter and a higher variance. Therefore, testing Component B makes a bigger change in $E[P_B]$ and $V(P_B)$ than the same number of tests of Component C would make in $E[P_C]$ and $V(P_C)$. Test plan 3 actually has the largest minimum and maximum posterior variances. These results support ranking established by the variance-based sensitivity analysis.

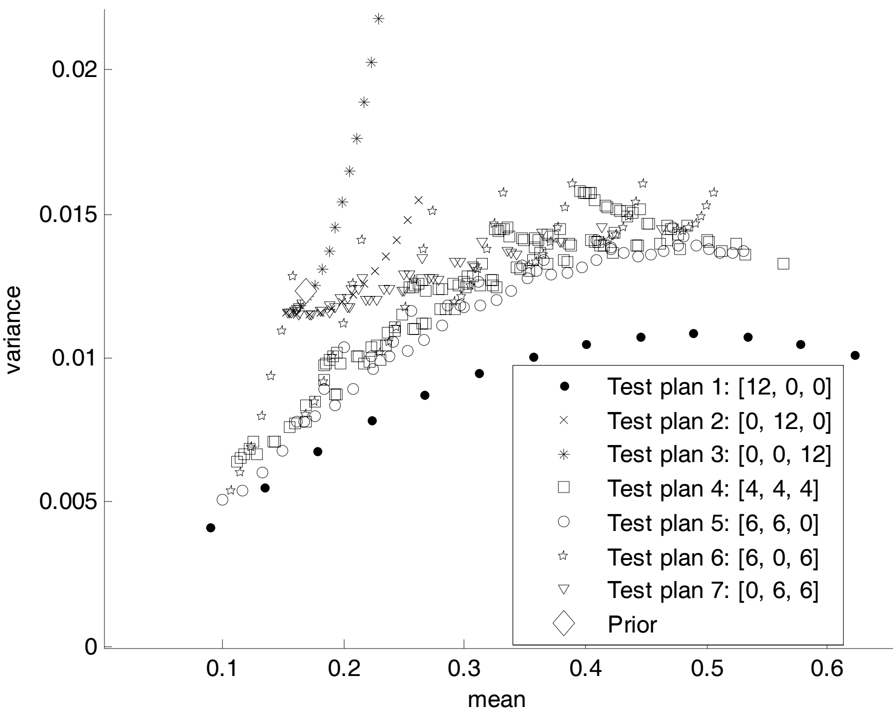
5.2 Scenario 2

In the second scenario, the prior distributions are imprecise (we use the imprecise beta model). The imprecise prior distributions are given in Table 2. Note that the precise priors given for the first scenario are included in these sets.

Table 2 Imprecise priors for Scenario 2

Component	*A*	*B*	*C*
IBM parameters	$\underline{t}_0 = 0.15$, $\overline{t}_0 = 0.20$ $\underline{s}_0 = 10$, $\overline{s}_0 = 12$	$\underline{t}_0 = 0.15$, $\overline{t}_0 = 0.55$ $\underline{s}_0 = 2$, $\overline{s}_0 = 5$	$\underline{t}_0 = 0.15$, $\overline{t}_0 = 0.20$ $\underline{s}_0 = 10$, $\overline{s}_0 = 12$

Figure 2 Dispersion of the mean and variance for different test plans for Scenario 1



5.2.1 Scenario 2: variance-based sensitivity analysis

The imprecise variance-based sensitivity analysis yields the results shown in Table 3.

Table 3 Imprecise variance-based sensitivity analysis Scenario 2

Component i	*A*	*B*	*C*
$\min\{SV_{ij}\}$	0.5438	0.0210	0.0095
$\max\{SV_{ij}\}$	0.9590	0.1819	0.2515

These results suggest that testing Component A and reducing its variance will have the most impact on reducing $V(\theta)$. Reducing the variance of Components B and C will have much less impact.

5.2.2 Scenario 2: dispersion of mean and variance

We will consider the same seven test plans used in the previous scenario. In general, a figure showing all of these points (every test result for a large sampling of priors) gets very difficult to display and view due to overlap. Consequently, we have chosen to display only the approximate regions that these points span, as shown in Figures 3 and 4. While these sets of points generally do not have such smooth boundaries, this approximation is reasonable for the qualitative analysis performed with them. For additional discussion, see Figure 6 in Section 5.3.2. Based on the sensitivity indices, it appears that test plan 1 (12, 0, 0) has the most potential to reduce $V(\theta)$. Because testing Component B can reduce the large imprecision in $E[P_B]$, we expect that test plans that include Component B will reduce the imprecision in $E[\theta]$.

Figure 3 Dispersion for Scenario 2, test plans 1, 2, 5 and 6

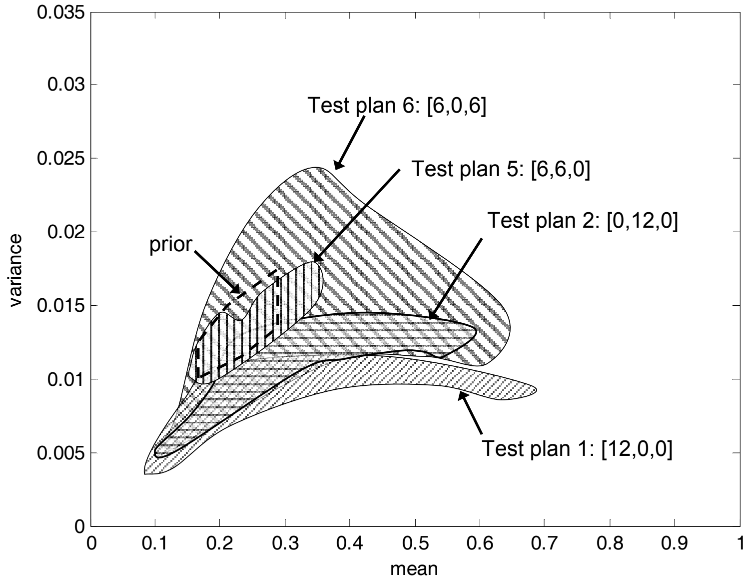
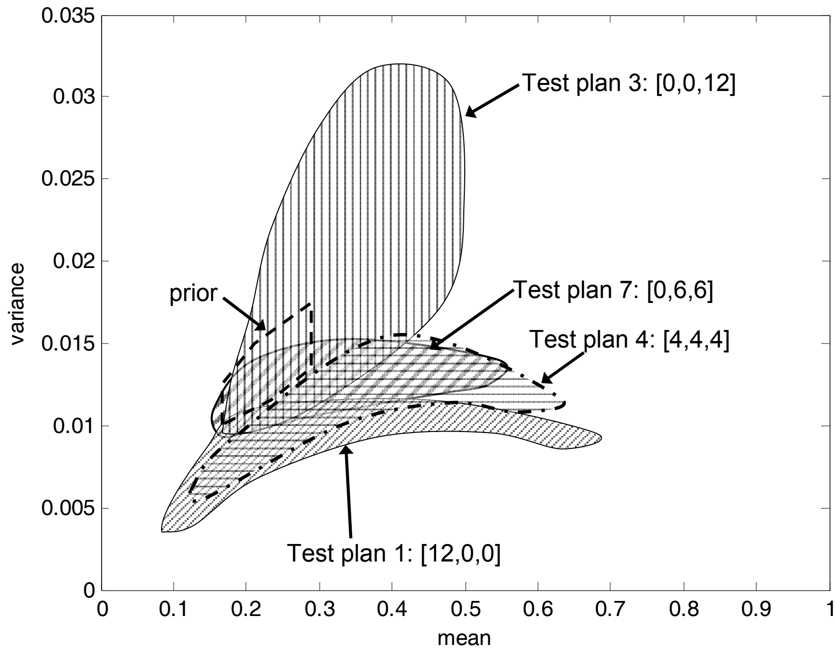


Figure 4 Dispersion for Scenario 2, test plans 1, 3, 4 and 7



Test plan 1 has the greatest range of $E[\theta]$, which reflects the critical location of Component A in the system. Moreover, this plan reduces $V(\theta)$ significantly, as the sensitivity index suggested. Test plan 5 has a slightly smaller range of $E[\theta]$ and does not reduce $V(\theta)$ as much, though it does more than test plan 2. In the results of test plan 2, we see that, for a given prior, when the test results confirm the prior, testing Component B reduces $V(P_B)$. However, poor test results (i.e. ones with many observed failures) increase both $E[P_B]$ and $V(P_B)$, which increase $E[\theta]$ and $V(\theta)$.

Testing Component C (test plan 3) is not helpful. Test results that confirm the prior tend to shrink the range of $E[\theta]$ compared to the prior. However, poor test results increase both $E[P_C]$ and $V(P_C)$, which increase $E[\theta]$ and $V(\theta)$. Test plan 6 can also give high-variance results because it does not reduce $V(P_B)$, which is relatively large, and poor results can increase both $V(P_A)$ and $V(P_C)$. Test plan 7 can reduce both $V(P_B)$ and $V(P_C)$, but, as the sensitivity indices suggest, this cannot reduce $V(\theta)$ as much as reducing $V(P_A)$. Test plan 4 can reduce all three component-level variances, but the limited number of test results means that the $V(P_A)$ is not reduced as much as it is in test plan 1, which limits the reduction of $V(\theta)$.

5.2.3 Scenario 2: imprecision in the mean

Table 4 describes the imprecision of $E[\theta]$ that can result from the various test plans. In each row, the first column is the test plan. The second column ('Minimum minimum') is the minimum possible mean over all possible distributions and test results. The third column ('Maximum maximum') is the maximum possible mean over all possible distributions and test results. The fourth column ('Minimum average') is the minimum average mean (see Section 4.2.3). The fifth column ('Maximum average') is the maximum-average mean over all the priors. The sixth and seventh columns are different. Here, the imprecision in $E[\theta]$ is calculated for each possible test result using Equation (11), and the minimum and maximum are taken over the possible test results.

Table 4 Posterior mean analysis for scenario 2

Test Design	$E[\theta]$				*Imprecision in* $E[\theta]$	
#: $\{n_A, n_B, n_C\}$	*Minimum minimum*	*Maximum maximum*	*Minimum average*	*Maximum average*	*Minimum*	*Maximum*
Prior	0.1691	0.2880	n.a.	n.a.	0.1189	
1: {12,0,0}	0.0891	0.6764	0.1643	0.2934	0.0918	0.1099
2: {0,12,0}	0.1527	0.3497	0.1671	0.2916	0.0732	0.1041
3: {0,0,12}	0.1587	0.4800	0.1682	0.2912	0.0853	0.2567
4: {4,4,4}	0.1120	0.6367	0.1656	0.2918	0.0709	0.1817
5: {6,6,0}	0.0988	0.5888	0.1647	0.2955	0.0685	0.1100
6: {6,0,6}	0.1065	0.6375	0.1644	0.2926	0.0809	0.2190
7: {0,6,6}	0.1530	0.5550	0.1667	0.2936	0.0737	0.1790

Test plan 1 yields the most extreme values of minimum-minimum and maximum-maximum because no failures (or all failures) significantly affects $E[P_A]$, which has a large impact on $E[\theta]$ due Component A's position in the system. Most of the test plans have the same minimum-average and maximum-average, which are close to the minimum and maximum prior $E[\theta]$. This is not surprising since extreme test results (and large changes from a prior to its posterior) such as observing all failures are unlikely when the number of tests is large enough.

The test plans that include Component B reduce the large imprecision in $E[P_B]$, which reduces the imprecision in $E[\theta]$. Test plans 3 and 6, which do not include Component B, not only fail to reduce the large imprecision in $E[P_B]$ but also add imprecision when a large number of failures for Component C add imprecision to $E[P_C]$. Similarly, test plan 4 can add some imprecision. Testing just Component A does not significantly reduce the imprecision, suggesting that the sensitivity index is not a good predictor of which tests will do well on this measure. The greatest potential reduction in imprecision can occur when both A and B are tested equally in test plan 5.

5.2.4 Scenario 2: imprecision in the variance

Table 5 describes the imprecision of $V[\theta]$ that can result from the various test plans. The table structure and types of results shown are similar to those of Table 4. In these results, test plan 1 is notable for its low values on almost all of the measures (the only exception being the maximum imprecision). This plan can substantially reduce $V(P_A)$, which reduces $V[\theta]$, as the sensitivity indices indicate. As we saw in Scenario 1, poor test results for Component C can greatly increase $V[\theta]$, and we see that here in the maximum-maximum variance for test plan 3. Unlike the results for the mean, here we see that testing Component A, as test plans 1, 5 and 6 do, can reduce the minimum-average and maximum-average (compared to the prior) because they substantially reduce $V(P_A)$, which reduces $V[\theta]$, as the sensitivity indices indicate. The other test plans have less impact, as suggested by the sensitivizty indices. All of the test plans reduce the imprecision in $V[\theta]$ except test plans 3 and 6 (which can greatly increase $V[\theta]$).

Table 5 Posterior variance analysis for scenario 2

Test Design #: $\{n_A, n_B, n_C\}$	$E[\theta]$				*Imprecision in* $E[\theta]$	
	Minimum minimum	*Maximum maximum*	*Minimum average*	*Maximum average*	*Minimum*	*Maximum*
Prior	0.0100	0.0173	n.a.	n.a.	0.0073	
1: {12,0,0}	0.0034	0.0117	0.0053	0.0110	0.0015	0.0068
2: {0,12,0}	0.0097	0.0181	0.0098	0.0155	0.0045	0.0061
3: {0,0,12}	0.0098	0.0325	0.0100	0.0160	0.0048	0.0189
4: {4,4,4}	0.0059	0.0162	0.0074	0.0117	0.0020	0.0054
5: {6,6,0}	0.0048	0.0146	0.0067	0.0116	0.0019	0.0062
6: {6,0,6}	0.0049	0.0243	0.0068	0.0119	0.0024	0.0165
7: {0,6,6}	0.0097	0.0155	0.0098	0.0145	0.0028	0.0050

5.3 Additional scenarios

For the sake of comparison, we now consider two slightly different scenarios (Scenarios 3 and 4) that increase the failure probability of Component C, thus changing the relationship between the components, as shown in Table 6. Note that the precise case is contained in the imprecise case. We will consider the same seven test plans used in Scenarios 1 and 2.

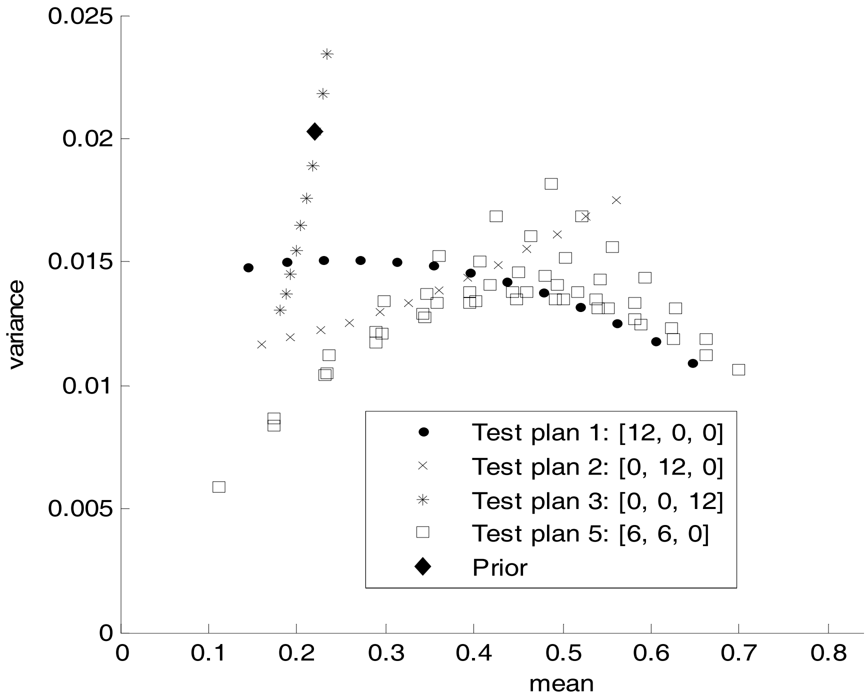
5.3.1 Scenario 3

For Scenario 3, the variance-based sensitivity analysis gives $SV_A = 0.4814$, $SV_B = 0.4583$, $SV_C = 0.0181$. Although these results indicate that testing Components A and B will be the most useful, they do not suggest the appropriate allocation between them. By comparison, Figure 5 reveals more about the tradeoffs involved. In agreement with the sensitivity indices, test plan 3 can change the mean very little for any test result, whereas test plans 1, 2 and 5 can change the mean a great deal. However, it also shows the tradeoffs between A and B, suggesting that testing both equally could be better or worse than testing either alone, depending on the actual results and the particular mean-variance tradeoffs preferences of the problem.

Table 6 Priors for Scenarios 3 and 4

Component	*A*	*B*	*C*
Scenario 3 (Precise)	$t_0 = 0.15$ $s_0 = 10$	$t_0 = 0.15$ $s_0 = 2$	$t_0 = 0.55$ $s_0 = 10$
Scenario 4 (IBM)	$\underline{t}_0 = 0.15,\ \overline{t}_0 = 0.20$ $\underline{s}_0 = 10,\ \overline{s}_0 = 12$	$\underline{t}_0 = 0.15,\ \overline{t}_0 = 0.55$ $\underline{s}_0 = 2,\ \overline{s}_0 = 5$	$\underline{t}_0 = 0.55,\ \overline{t}_0 = 0.60$ $\underline{s}_0 = 10,\ \overline{s}_0 = 12$

Figure 5 Dispersion of the mean and variance for different test plans for Scenario 3



5.3.2 Scenario 4

For Scenario 4, the imprecise variance-based sensitivity analysis yields the results shown in Table 7. These results also suggest that it is important to test A and B (using the upper bounds), while testing Component C is less important. Note that the maximum for

Component C is about the same as the minimum for Component B, but there is additional overlap between A and C. Therefore, it appears that test plans 1 (12, 0, 0), 2 (0, 12, 0), and 5 (6, 6, 0) should have the most potential to reduce $V(\theta)$.

We begin by examining the dispersion of the mean and variance estimates across all possible experimental results for these three test plans, as shown in Figure 6. First, we have shown the actual cloud of points for test plan 1 in order to reveal some additional details that were lost in the previous analysis. Here the non-smooth nature of the boundaries is apparent. The cloud of points actually consists of the union of many clusters – one cluster for each possible test result. Unlike the precise case, where each result led to a point in Figure 5, now each result leads to a set of points that reflect the imprecision in the prior.

All three test plans (1, 2 and 5) can significantly change $E[\theta]$. The impact of test plan 2 (0, 12, 0) is reduced by the system structure, in which Component B is parallel to Component C. The maximum and range for $V(\theta)$ of test plan 1 (12, 0, 0) are much greater than those for the other two test plans. The significant imprecision in the priors, especially when combined, leads to large imprecision for any test result, especially in test plan 1 (12, 0, 0). Because the *s* parameters for Component B are smaller than those for Component A, testing Component B reduces $V(P_B)$ more than testing Component A reduces $V(P_A)$. Of course, testing both components (as in test plan 5) can reduce both component variances, which is quite effective at reducing $V(\theta)$ while still being responsive to the mean.

Figure 6 Sample results for test plans 1, 2 and 5 for Scenario 4

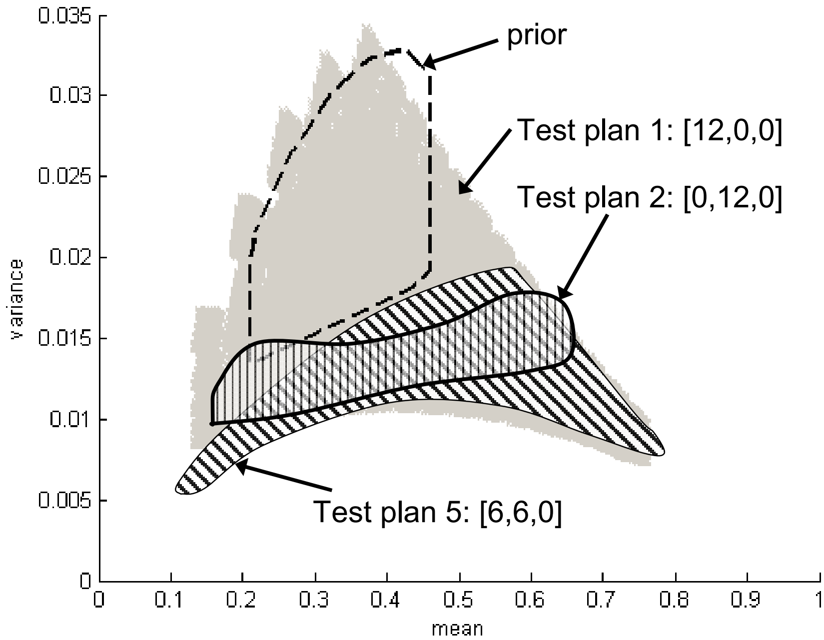


Table 7 Imprecise variance-based sensitivity analysis Scenario 4

Component i	*A*	*B*	*C*
$\min\{SV_{ij}\}$	0.1363	0.2406	0.0116
$\max\{SV_{ij}\}$	0.7204	0.6960	0.2512

Some additional results are shown in Table 8. Test plans 2, 4, 5 and 7 have a maximum imprecision that is less than the imprecision in the prior. All of these plans test Component B and reduce the large imprecision in $E[P_B]$, which reduces the imprecision in $E[\theta]$. Note that testing Component A (as in test plans 1 or 6) does not reduce the imprecision significantly, suggesting that the sensitivity indices are not good predictors of which tests will do well on this measure. Similarly, the suggestion does not test Component C based on the sensitivity indices is contradicted by small imprecision for test plan 4.

Table 8 Posterior analysis for Scenario 4

Test Design #: $\{n_A, n_B, n_C\}$	*Imprecision in* $E[\theta]$		$V[\theta]$	
	Minimum	*Maximum*	*Minimum average*	*Maximum average*
Prior	0.2439		n.a.	n.a.
1: {12,0,0}	0.1463	0.2519	0.0094	0.0304
2: {0,12,0}	0.1085	0.1486	0.0103	0.0153
3: {0,0,12}	0.1501	0.3112	0.0134	0.0310
4: {4,4,4}	0.0709	0.1817	0.0075	0.0118
5: {6,6,0}	0.1172	0.1952	0.0083	0.0150
6: {6,0,6}	0.1474	0.3019	0.0107	0.0295
7: {0,6,6}	0.1126	0.2174	0.0109	0.0183

The bounds on average variance are also interesting. Because testing component B significantly reduces imprecision, plans that test it lead to lower maximum-average variances. However, test plan 4, which also tests component C, has the lowest maximum-average variance. This provides significantly more insight than the sensitivity values gave.

6 Discussion

The above results, though for specific scenarios and a specific system design, demonstrate some principles that we believe are generally applicable to the problem of managing information collection activities.

First, examining the dispersion of the mean and variance is a useful way to determine the possible outcomes of a test plan. This analysis, which can be done before any tests are performed, can identify those plans that are most likely to reduce system-level variance and have a large impact on system-level mean.

Next, the variance-based sensitivity analysis is not a substitute for looking at the dispersion of the mean and variance, especially in the imprecise scenarios. It does give some prediction into which components should be tested. Because it is computationally less expensive to calculate the sensitivity indices than the potential posteriors across all results, this is important. In particular, testing a component with a high sensitivity index can reduce system-level variance substantially if the number of tests is large enough relative to the *s* parameter (a small number of tests would not change the component-level variance enough if the *s* parameter is large). However, testing a component with a small sensitivity index may greatly increase system-level variance; only examining the dispersion of the mean and variance can reveal that.

Moreover, the sensitivity indices do not give adequate insight into the impact of test plans that test multiple components. In Scenario 2, the sensitivity indices clearly suggested that testing Components B and C were much less important than testing A. However, as shown in Table 5, testing both A and B (test plan 5) reduced the maximum-maximum posterior variance more than testing both A and C (test plan 6) or testing all three components equally (test plan 4). Test plan 5 also reduces the maximum imprecision in the variance (compared to test plan 6), which means that its worst case result leads to the most information about the variance than any other test's worst case. This is ideal because the variance not only has the smallest maximum but also will be known accurately, whatever the actual result. It should be noted that one could also consider joint sensitivity indices, an analysis that was not performed in this study and should be considered in future work.

A variance-based sensitivity index does not give much insight into how testing that component will affect the imprecision of the system-level mean. The adjustment from the precise sensitivity indices to the imprecise ones is necessary when using imprecise probabilities, but it does not sufficiently capture all important aspects of the imprecision. For example, in Scenario 2, the sensitivity indices clearly suggest that testing Component A is most important. However, test plan 5, which tests both A and B, has the smallest minimum imprecision of the mean.

Testing a component with large imprecision in its mean failure probability is useful because it reduces the component-level imprecision, which reduces the system-level imprecision. However, if the component-level imprecision is low, testing that component may increase the imprecision of the system-level mean and variance if the results contradict the prior information. Again, the dispersion plot will show this potential.

The minimum and maximum-average measures (for system-level mean and variance) are computationally expensive and not very useful. In Scenario 2, they change very little from the values for the prior. Additionally, the minimum-minimum and maximum-maximum metrics yield similar rankings to those from the minimum-average and maximum-average respectively.

In this example, many posterior statistics were analytically computable, as shown in Equations (1–4) and Equation (8). In general, the posterior system distribution would need to be calculated using a double-loop Monte Carlo simulation, or a more advanced method (for a summary, see Bruns and Paredis, 2006). This greatly increases the computational costs over this example. However, having an estimate of the posterior distribution allows one to use other uncertainty metrics, such as the entropy, the Aggregate Uncertainty (Klir and Smith, 2001), or imprecise posterior breadth measures (Ferson and Tucker, 2006). Consideration of these metrics is left for future work.

7 Summary

This paper has presented different uncertainty metrics for evaluation information collection activities and has applied these strategies for evaluating the possible results of component test plans intended to improve system reliability estimates. Our approach is designed for collecting information that will be used to update beliefs that are described using prior distributions for the unknown parameters. The approach here does not apply when maximum likelihood estimators or other estimation techniques will be used. Aughenbaugh and Herrmann (2008, 2009) compare different statistical approaches for assessing reliability.

The strategies presented here provide useful information for managing information collection in general and planning reliability tests in particular. In this paper we have not considered specific approaches for making decisions in the presence of uncertainty or estimating the economic value of the information, since these depend on the problem context and the preferences of the decision-maker.

Instead, we considered the variance and imprecision of the posterior distributions more directly. In some cases, this will be sufficient to make a decision. Future work will need to consider how to integrate the approaches presented here with approaches in information economics, decision analysis and optimisation to help one select the best test plan.

Acknowledgements

This work was sponsored in part by Applied Research Laboratories at the University of Texas at Austin IR&D grant 07–09.

References

Aughenbaugh, J.M. and Herrmann, J.W. (2008) 'A comparison of statistical approaches for assessing reliability', *International Journal of Reliability and Safety*, Vol. 2, No. 4, pp.265–285.

Aughenbaugh, J.M. and Herrmann, J.W. (2009) Reliability-based decision-making: a comparison of statistical approaches', *Journal of Statistical Theory and Practice*, Vol. 3, No. 1, pp.289–304.

Aughenbaugh, J.M. and Paredis, C.J.J. (2006) 'The value of using imprecise probabilities in engineering design', *Journal of Mechanical Design*, Vol. 128, No. 4, pp.969–979.

Berger, J.O. (1984) 'The robust Bayesian viewpoint', in Kadane, J.B. (Ed.): *Robustness of Bayesian Analysis*, North-Holland, New York, pp.63–124.

Berger, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer.

Berger, J.O. (1993) *An Overview of Robust Bayesian Analysis*, Technical Report #93–53c, Purdue University, IN, West Lafayette.

Box, G.E.P. and Tiao, G.C. (1973) *Bayesian Inference in Statistical Analysis*, Reading, Addison-Wesley, Massachusetts.

Bradley, S.R. and Agogino, A.M. (1994) 'An intelligent real time design methodology for component selection: an approach to managing uncertainty', *Journal of Mechanical Design*, Vol. 116, No. 4, pp.980–988.

Bruns, M. and Paredis, C.J.J. (2006) 'Numerical methods for propagating imprecise uncertainty', *2006 ASME International Design Engineering Technical Conferences and Computers in Information Engineering Conference*, Philadelphia, PA.

Chan, K., S. Tarantola, A. Saltelli and Sobol, I.M. (2000) 'Variance-based methods', in Saltelli, A. Chan, K. and Scott, E.M. (Eds): *Sensitivity Analysis*, Wiley, New York.

Coolen-Schrijner, P. and Coolen, F.P.A. (2007) 'Non-parametric predictive comparison of success-failure data in reliability', *Proceeding IMechE, Part 0: Journal of Risk and Reliability*, Vol. 221, pp.319–327.

Coolen, F.P.A. (1994) 'On Bernoulli experiments with imprecise prior probabilities', *The Statistician*, Vol. 43, No. 1, pp.155–167.

Coolen, F.P.A. (1998) 'Low structure imprecise predictive inference for Bayes's problem', *Statistics and Probability Letters*, Vol. 36, pp.349–357.

Coolen, F.P.A. (2004) 'On the use of imprecise probabilities in reliability', *Quality and Reliability in Engineering International*, Vol. 20, pp.193–202.

Dempster, A.P.(1967) 'Upper and lower probabilities induced by a multi-valued mapping', *The Annals of Statistics*, Vol. 28, pp.325–339.

Dubois, D. and Prade, H. (1988) *Possibility Theory: An Approach to Computerised Processing of Uncertainty*', Plenum Press, New York.

Ferson, S. and Tucker, W.T. (2006) *Sensitivity in Risk Analyses with Uncertain Numbers*, Sandia National Laboratories, Albuquerque, NM.

Groen, F.J. and Mosleh, A. (2005) 'Foundations of probabilistic inference with uncertain evidence', *International Journal of Approximate Reasoning*, Vol. 39, No. 1, pp.49–83.

Gupta, M.M. (1992) 'Intelligence, uncertainty and information', in Ayyub, B.M., Gupta, M.M. and Kanal, L.N. (Eds): *Analysis and Management of Uncertainty: Theory and Applications*, North-Holland, New York, pp.3–11.

Hall, J.W. (2006) 'Uncertainty-based sensitivity indices for imprecise probability distributions', *Reliability Engineering and System Safety*, Vol. 91, Nos. 10–11, pp.1443–1451.

Howard, R.A. (1966) 'Information value theory', *IEEE Transactions on Systems Science and Cybernetics*, Sec-2, No. 1, p.22.

Insua, D.R. and Ruggeri, F. (2000) *Robust Bayesian Analysis*, Springer, New York.

Klir, G.J. and Smith, R.M. (2001) 'On measuring uncertainty and uncertainty-based information: recent developments', *Annals of Mathematics and Artificial Intelligence*, Vol. 32, Nos. 1–4, pp.5–33.

Kokkolaras, M., Mourelatos, Z.P. and Papalambros, P.Y. (2006) 'Impact of uncertainty quantification on design: an engine optimisation case study', *International Journal of Reliability and Safety*, Vol. 1, Nos. 1–2, pp.225–237.

Lawrence, D.B. (1999) *The Economic Value of Information*, Springer-Verlag, New York.

Ling, J.M., Aughenbaugh, J.M. and Paredis, C.J.J. (2006) 'Managing the collection of information under uncertainty using information economics', *Journal of Mechanical Design*, Vol. 128, No. 4, pp.980–990.

Malak Jr., R.J., Aughenbaugh, J.M. and Paredis, C.J.J. (2009, March) 'Multi-attribute utility analysis in set-based conceptual design', *Computer-Aided Design*, Vol. 41, No. 3, pp.214–227.

Marschak, J. (1974) *Economic Information, Decision, and Prediction: Selected Essays*, D. Reidel Publishing Company, Boston.

Matheson, J.E. (1968) 'The economic value of analysis and computation', *IEEE Transactions on Systems Science and Cybernetics*, SSC-4, No. 3, pp.325–332.

Nikolaidis, E., Chen, S., Cudney, H., Haftka, R.T. and Rosca, R. (2004) 'Comparison of probability and possibility for design against catastrophic failure under uncertainty', *Journal of Mechanical Design*, Vol. 126, No. 3, pp.386–394.

Oberkampf, W.L., Helton, J.C., Joslyn, C.A., Wojtkiewicz, S.F. and Ferson, S. (2004) 'Challenge problems: uncertainty in system response given uncertain parameters', *Reliability Engineering and System Safety*, Vol. 85, Nos. 1–3, pp.11–19.

Pham-Gia, T. and Turkkan, N. (1992) 'Sample size determination in Bayesian analysis', *The Statistician*, Vol. 41, No. 4, pp.389–397.

Radhakrishnan, R. and McAdams, D.A. (2005, May) 'A methodology for model selection in engineering design', *Journal of Mechanical Design*, Vol. 127, pp.378–387.

Schlosser, J. and Paredis, C.J.J. (2007) *Managing Multiple Sources of Epistemic Uncertainty in Engineering Decision Making*, SAE World Congress, Detroit, MI.

Shafer, G. (1976) *A Mathematical Theory of Evidence*, Princeton University Press, NJ, Princeton.

Sobol, I.M. (1993) 'Sensitivity analysis for non-linear mathematical models', *Mathematical Modelling and Computational Experiment*, Vol. 1, No. 1, pp.407–414.

Soundappan, P., Nikolaidis, E., Haftka, R.T., Grandhi, R. and Canfield, R. (2004) 'Comparison of evidence theory and Bayesian theory for uncertainty modelling', *Reliability Engineering and System Safety*, Vol. 85, Nos. 1–3, pp.295–311.

Utkin, L.V. (2004a) 'Interval reliability of typical systems with partially known probabilities', *European Journal of Operational Research*, Vol. 153, No. 3, SPEC ISS, pp.790–802.

Utkin, L.V. (2004b) 'Reliability models of M-out-of-N systems under incomplete information', *Computers and Operations Research*, Vol. 31, No.10, pp.1681–1702.

Walley, P. (1991) *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, New York.

Walley, P., Gurrin, L. and Burton, P. (1996) 'Analysis of clinical data using imprecise prior probabilities', *The Statistician*, Vol. 45, No. 4, pp.457–485.

Weber, M. (1987) 'Decision making with incomplete information', *European Journal of Operational Research*, Vol. 28, No. 1, p.44.

Weichselberger, K. (2000) 'The theory of interval-probability as a unifying concept for uncertainty', *International Journal of Approximate Reasoning*, Vol. 24, Nos. 2–3, pp.149–170.