A Comparison of Information Management using Imprecise Probabilities and Precise Bayesian Updating of Reliability Estimates

J. M. Aughenbaugh[1] and J. W. Herrmann[2]

[1]Applied Research Laboratories
University of Texas at Austin
Austin, TX 78713-8029
jason@arlut.utexas.edu

[2]Department of Mechanical Engineering
and Institute for Systems Research University
of Maryland
College Park, MD 20742
jwh2@umd.edu

Abstract: Assessing the reliability of components and systems is an important problem in engineering design. Estimates of the reliability of a design can play a significant role in final design decisions. Data for making these estimates is often scarce during the design process. However, designers also frequently have the option to acquire more information by expending resources. Designers thus face the dual questions of how to update their estimates and whether it is valuable to collect additional information. Various statistical updating methods exist and can be used in reliability estimation, including precise Bayesian updating and methods based on imprecise probabilities. In this paper, we explore the management of information collection using these two approaches. These ideas combine elements from sensitivity analysis, value of information calculations, and uncertainty measures. Rather than dealing with abstract measures of total of uncertainty for a particular distribution or set of distributions, we explore the relationships between variance-based sensitivity analysis of the prior and posterior estimates of the mean and variance over all possible results of a particular test. The goal is to gain insight into the many tradeoffs that occur when comparing different information collection actions, especially when the exact outcome of the action is uncertain. These tradeoffs are explored using the example reliability modeling of a simple parallel-series system with three components.

Keywords: reliability assessment, imprecise probabilities, information management

1. Introduction

Modeling uncertainty is an increasingly important activity in engineering applications. As engineers progress from deterministic approaches to nondeterministic approaches, the question of how to model the uncertainty in the nondeterministic approaches must be answered. Researchers have proposed various methods for modeling and propagating uncertainty. A great deal of literature has been devoted to developing and applying individual methods, and others are devoted to philosophical debates of the appropriateness of different measures. More recently, there is a growing interest in practical comparisons of the 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, 2007).

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 modeling the state of the world. One step of this formulation phase is *information management*. In this step, the engineers 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. The focus of this paper is on how to model uncertainty in order to best support information management decisions. In particular, we consider the problem of system reliability assessment, as discussed in Section 2.

Managing information collection activities during engineering design is clearly 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 catalog design (Bradley and Agogino, 1994). Some recent work to improve engineering design processes has considered this problem from a frequentist 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). This work used 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*, which is the expected benefit of the information minus the cost of acquiring that information.

In this paper, we focus on a Bayesian updating problem and consider problem-independent measures of the value of the 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 et al., 2007). 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.

Section 2 presents the example problem. Section 3 reviews the precise and imprecise Bayesian statistical models. Section 4 presents the uncertainty metrics that we will use. 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 where a designer is considering additional testing of some key components in a system in order to get better estimates of the system reliability. From a reliability perspective, the example system can be modeled as a parallel-series system, as shown in Figure 1. We assume that the failures of each component are independent events. The designer has some prior information about the reliability of each component and thus can create an estimate of the system reliability. However, the engineer hopes that additional testing will refine the estimate.

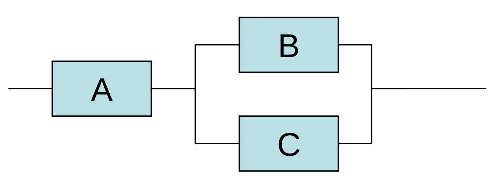

Figure 1. Reliability Block Diagram for the System.

It will be convenient to frame things in terms of failure probability instead of reliability. If component *i* has a reliability of R_i, then the corresponding failure probability of component *i* is

$P_i = 1 - R_i$. Let θ be the failure probability of the system, which is the parameter of interest. Ideally, in order to determine this, the designer would have enough data to make a precise assessment of P_A, P_B, and P_C, such as "$P_A = 0.05$." However, there are practical reasons why the designer cannot or is unwilling to make a precise assessment despite holding some initial beliefs about the failure probability (Malak et al., 2007).

We consider the case in which only component testing is feasible. No system-level tests are possible. The designer will use the results of additional testing to update his beliefs about the components' failure probabilities, which will yield an updated estimate of the system failure probability. In particular, the engineer is interested in knowing which component should be tested. 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 failure probabilities P_A, P_B, and P_C are modeled as independent random variables with a beta distribution. If $P_A \sim beta(\alpha_A, \beta_A)$, then

$$E\left[P_A\right] = \alpha_A / \left(\alpha_A + \beta_A\right);$$

$$V\left[P_A\right] = \frac{\alpha_A \beta_A}{\left(\alpha_A + \beta_A\right)^2 \left(\alpha_A + \beta_A + 1\right)};$$

$$E\left[P_A^2\right] = V\left[P_A\right] + \left(E\left[P_A\right]\right)^2 = \left(\frac{\alpha_A}{\alpha_A + \beta_A}\right)\left(\frac{\alpha_A + 1}{\alpha_A + \beta_A + 1}\right).$$

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

$$R_{sys} = R_A\left(1 - (1 - R_B)(1 - R_C)\right) \tag{1}$$

$$\theta = P_{sys} = P_A + P_B P_C - P_A P_B P_C \tag{2}$$

$$E[\theta] = E[P_A] + E[P_B]E[P_C] - E[P_A]E[P_B]E[P_C] \tag{3}$$

$$\begin{aligned} E[\theta^2] = {} & E[P_A{}^2] + 2E[P_A]E[P_B]E[P_C] - 2E[P_A{}^2]E[P_B]E[P_C] + \\ & + E[P_B{}^2]E[P_C{}^2] - 2E[P_A]E[P_B{}^2]E[P_C{}^2] + E[P_A{}^2]E[P_B{}^2]E[P_C{}^2] \end{aligned} \tag{4}$$

$$V[\theta] = E[\theta^2] - (E[\theta])^2 \tag{5}$$

3. Formalisms for modeling uncertainty

In this paper we will compare two different approaches for updating reliability assessments: the precise Bayesian and the imprecise beta model, which is useful for both the robust Bayesian approach and the imprecise probability approach. Some introductory material is provided here. For a more complete discussion, see the cited references and the discussion in Aughenbaugh and Herrmann (2007).

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 of the requirements 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 research and debate. The approaches proposed include the use of non-informative priors (Jeffreys, 1961; Zellner, 1977; Berger and Bernardo, 1992), maximum-entropy priors (Fougere, 1990), and data-dependent empirical Bayes approaches (Maritz and Lewin, 1989). Still, whether a single prior distribution, especially the uniform prior, can reflect all of the uncertainty is an open question to some observers. However, in many engineering problems, designers do have some prior information, such as data and experience from similar systems, and the Bayesian approach allows this information to be included in the analysis.

To support analytical solutions, the form of the prior is often restricted to conjugate distributions with respect to the measurement model, in which case the posterior distribution that results from the update has the same type as the prior. For the problem considered in this paper, in which the number of failures in a given number of tests is a binomial random variable, it is convenient to model the prior distribution of a component's failure probability as a beta distribution with parameters α and β. 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 simple addition and subtraction, an enormous improvement in efficiency 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 are mathematically similar, but differ in motivation.

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 changing the prior. The analyst considers several reasonable prior distributions and performs the update on each to get a set of posterior distributions. After additional data is collected, each candidate prior is updated, resulting in a set of posterior distributions. This set of posterior distributions yields a range of point estimates and a set of credible intervals. If there is no significant change in the conclusion across this set of posteriors, then the conclusion is *robust* to the selection of the prior.

This analysis is not possible with a single prior. For example, if the designer is unsure about the failure probability, one precise Bayesian approach for dealing with this lack of confidence in the estimate is to increase the variance of the prior model, thus reflecting more uncertainty of some kind. Taken to the extreme, a complete lack of information generally leads to a uniform distribution. Unfortunately, the use of a uniform distribution confounds two cases: first, that nothing is known; second, that all failure probabilities between 0 and 1 are equally likely, which is actually substantial information.

In the context of a large engineering project in which there are many individuals, this is an important distinction. For example, one engineer's complete lack of knowledge about some aspect of the system may be offset by another engineer's expertise or by additional experimentation. However, if substantial analysis has already led to the conclusion that certain outcomes are equally likely, then it would be inefficient to expend additional resources examining those probabilities. A precise approach confounds these two scenarios, therefore adding confusion to information management decisions. The robust Bayesian approach allows one to consider the different scenarios independently rather than aggregating them together. This affords the design team the opportunity to make different, more appropriate decisions under the two scenarios.

The theory of imprecise probabilities, formalized by Walley (1991), has previously been considered in design decisions and set-based design (Aughenbaugh and Paredis, 2006; Rekuc et al., 2006) and in reliability analysis (Coolen, 1994; Coolen, 2004; Utkin, 2004a; Utkin, 2004b). The theory of imprecise probabilities uses the same fundamental notion of rationality as de Finetti's work (1974). However, the theory allows a range of indeterminacy—prices at which a decision-maker will not enter a gamble as either a buyer or a seller. These in turn correspond to ranges of probabilities. For the problem of updating beliefs, imprecise probability theory essentially allows prior and posterior beliefs to be expressed as sets of density functions.

The imprecise model captures two aspects of uncertainty: the imprecision in the prior beliefs (whether inherent or due to incomplete elicitation) and the probabilistic uncertainty in the parameter value. The distinction between these two types of uncertainty is not always obvious, but the distinction can be valuable in practice (Winkler, 1996).

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

For the robust updating approach, it is convenient to use the imprecise beta model and to re-parameterize the beta so that the density of $beta(s,t)$ is as given in Equation (6) (Walley, 1991; Walley et al., 1996).

$$\pi_{s,t}(\theta) \propto \theta^{st-1}(1-\theta)^{s(1-t)-1} \tag{6}$$

Compared to the standard parameterization of $beta(\alpha,\beta)$, this means that $\alpha = s \cdot t$ and $\beta = s \cdot (1-t)$ or equivalently that $s = \alpha + \beta$ and $t = \alpha/(\alpha+\beta)$. The convenience of this parameterization is that t is the mean of the distribution, which has an easily grasped meaning for both the prior assessment and the posterior analysis. The model is updated as follows: if the prior parameters are s_0 and t_0, then, after n trials with m failures, the posterior parameters are $s_n = s_0 + n$ and $t_n = (s_0 t_0 + m)/(s_0 + n)$. Since $s_n = s_0 + n$, s_0 can be interpreted to be a virtual sample size of the prior information; it captures how much weight to place on the prior compared to the observed data. Selecting this parameter therefore depends on the available information. Following Walley (1991), the parameters t and s can be imprecise and expressed as intervals $\left[\underline{t}_0, \overline{t}_0\right]$ and $\left[\underline{s}_0, \overline{s}_0\right]$. That is, 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 $\underline{s}_0 \le s_0 \le \overline{s}_0$. When the test results are collected, each prior in the set is updated as described above.

4. Metrics of uncertainty

When considering measures of uncertainty in context of planning additional tests to reduce uncertainty, it is important to keep the following points in mind.

First, if one is modeling 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. Thus, there are $n+1$ possible posterior distributions and $n+1$ possible means, variances, and other statistics. Characterizing the quality a test plan (which has uncertain outcomes) is a classic problem in decision-making. Decision-makers have different attitudes towards such situations. For example, 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.

Modeling the system performance as an imprecise probability distribution introduces an additional complexity: the mean, variance, and other statistics about that distribution are imprecise; that is, there is a set of means 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 imprecise posterior distributions, with $n+1$ possible sets of means, variances, and other statistics. Characterizing any specific result requires some way to describe the set, either by selecting a subset of the distributions or finding the range of values for that result. After that, one still has the problem of uncertain outcomes, as discussed above.

This paper presents and demonstrates approaches for evaluating information gathering 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 existing 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.

For the precise Bayesian approach, we will use a variance-based sensitivity analysis and the dispersion of the mean and variance of the posterior distributions of system failure probability. For the imprecise beta model, we will consider an imprecise variance-based sensitivity analysis (Hall, 2006), the imprecision in the mean before and after additional tests are conducted, and the range of the mean and variance of the prior and posterior distributions of system failure probability.

4.1. METRICS FOR PRECISE DISTRIBUTIONS

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 strategy considers the possible outcomes of a test plan.

4.1.1. Variance-based sensitivity analysis

One can avoid the problem of considering a large number of possible test results by ignoring them entirely and focusing on the current state of information. One such approach is variance-based sensitivity analysis, which calculates the total variance of the system performance and determines how each input variable contributes to this (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. The sensitivity index is the ratio of the variance of the conditional expectation to the total variance.

For test planning, a large sensitivity index indicates that reducing the variance of that variable can reduce the system performance variance. This suggests that, in order to reduce system performance variance, a test plan should focus on reducing that input variable's variance. A small sensitivity index for an input variable suggests that reducing that variable's variance should be a low priority for testing.

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:

$$
\begin{aligned}
SV_A &= \frac{1}{V(\theta)}\left(1 - E[P_B]E[P_C]\right)^2 V[P_A] \\
SV_B &= \frac{1}{V(\theta)}\left(E[P_C]\right)^2\left(1 - E[P_A]\right)^2 V[P_B] \\
SV_C &= \frac{1}{V(\theta)}\left(E[P_B]\right)^2\left(1 - E[P_A]\right)^2 V[P_C]
\end{aligned}
\tag{7}
$$

4.1.2. Observing Mean and Variance for different results

The variance of the probability distribution is a measure of uncertainty about the parameter that the probability distribution models. In general, a distribution with smaller variance means that there is less uncertainty about the parameter. 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 additional samples needed to satisfy an upper bound on the posterior variance for a random probability modeled with a beta distribution. Unfortunately this result is not directly applicable to reducing the variance of the system failure probability by testing only the components.

In the case considered here, a test plan conducts n_A tests of component A, n_B tests of component B, and n_C tests of component C. The test plan can be summarized as $T = \{n_A, n_B, n_C\}$. If there x_A failures of component A, x_B failures of component B, and x_C failures of component C, then the posterior distributions of the component failure probabilities are as follows: $P_A \sim beta(\alpha_A + x_A, \beta_A + n_A - x_A)$, $P_B \sim beta(\alpha_B + x_B, \beta_B + n_B - x_B)$, and $P_C \sim beta(\alpha_C + x_C, \beta_C + n_C - x_C)$. From these posterior distributions, one can 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 $(n_A + 1)\times(n_B + 1)\times(n_C + 1)$ possible test results.

4.2. METRICS OF UNCERTAINTY FOR IMPRECISE DISTRIBUTIONS

As we did with the precise 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 strategy considers the possible outcomes of a test plan.

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 (see (Klir and Smith, 2001) for an overview). 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) presented an approach to extend variance-based sensitivity analysis to imprecise probability distributions. The generalization 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 the following:

$$\begin{aligned}\underline{SV}_i &= \min_{p\in F}\left(SV_{i,p}\right)\\ \overline{SV}_i &= \max_{p\in F}\left(SV_{i,p}\right)\end{aligned} \tag{8}$$

The difficulty in calculating these is the need to optimize these indices 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:

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

This yields N_e+2 values for this parameter. We repeat for the other five parameters. We then take all of the combinations, which yields $\left(N_e+2\right)^6$ joint prior distributions. We choose $N_e=3$, which was determined to be adequate for this problem. More complex systems will require a more complex parameter sampling scheme.

4.2.2. Dispersion of mean and variance

In Section 4.1.2, the dispersion of the mean and variance were considered for a precise prior. Given an imprecise prior, a specific test result will yield an imprecise posterior distribution, which has a range of means and a range of variances, as discussed above. The dispersion of the mean and variance (over the possible test results) is no longer a sequence of points, as in the precise case; it is instead a sequence of sets of mean-variance pairs.

In the case considered here, given imprecise priors for the failure probabilities of the three components, we can compare different test plans (e.g. test just Component A or test just Component B) and determine how they affect the dispersion of the mean and variance.

As before, let F be the entire set of prior joint distributions for the component failure probabilities, and consider a test plan that conducts n_A tests of component A, n_B tests of component B, and n_C tests of component C. If there x_A failures of component A, x_B failures of component B, and x_C failures of component C, then this result yields a set $F'\left(x_A, x_B, x_C, n_A, n_B, n_C\right)$ of posterior distributions. There is a different F' for every test result. Each posterior distribution $p' \in F'\left(x_A, x_B, x_C, n_A, n_B, n_C\right)$ is determined by updating a prior distribution $p \in F$ as described in Section 4.1.2. From the posterior distribution, one can calculate the mean and variance of the system failure probability as discussed in Section 2. We will select distributions from *F* using the procedure 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 imprecise beta model, this measure is simply $\overline{t}-\underline{t}$. We can measure this range for the prior distribution and for each imprecise posterior distribution that results from a specific test result. Each result has a particular posterior imprecision (of the mean) associated with it. Ideally, the analyst would like this posterior imprecision to be as small as possible over all results, so we consider the maximum imprecision that results across all results.

In the case considered here, given the range of means for the failure probabilities of the three components, it is easy to see that the minimal failure probability of the system is determined by the components' minimal failure probabilities. Likewise, the maximal failure probability of the system is determined by the components' maximal failure probabilities. Therefore, the prior imprecision in the system failure probability can be calculated as follows:

$$\begin{aligned}\Delta_0(\theta) &= \max_{p\in F} E\left[\theta \middle| p\right] - \min_{p\in F} E\left[\theta \middle| p\right] \\ &= \left(\overline{t}_{0A}+\overline{t}_{0B}\overline{t}_{0C}-\overline{t}_{0A}\overline{t}_{0B}\overline{t}_{0C}\right)-\left(\underline{t}_{0A}+\underline{t}_{0B}\underline{t}_{0C}-\underline{t}_{0A}\underline{t}_{0B}\underline{t}_{0C}\right)\end{aligned} \tag{9}$$

Each possible result of a test plan that conducts a total of *n* tests will yield a set F' of posterior distributions. The posterior imprecision in the system failure probability (given this result) can be determined as follows:

$$\begin{aligned}\Delta_{n,F'}(\theta) &= \max_{p'\in F'} E\left[\theta \middle| p'\right] - \min_{p'\in F'} E\left[\theta \middle| p'\right] \\ &= \left(\overline{t}_{nA}+\overline{t}_{nB}\overline{t}_{nC}-\overline{t}_{nA}\overline{t}_{nB}\overline{t}_{nC}\right)-\left(\underline{t}_{nA}+\underline{t}_{nB}\underline{t}_{nC}-\underline{t}_{nA}\underline{t}_{nB}\underline{t}_{nC}\right)\end{aligned} \tag{10}$$

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:

$$\Delta_{\max}(\theta) = \max_{F'} \left\{ \Delta_{n,F'}(\theta) \right\} \tag{11}$$

One can also consider the average mean over the results for a given prior. For each prior and possible test result, that prior is used to determine the probability of that test result and the posterior mean given that result. Here, let $\mu\left(x,n,p_0(\cdot)\right)$ be the posterior mean, which depends upon the prior $p_0(\cdot)$ and the test result. These posterior means and prior probabilities of each result are then used to calculate an average mean for that prior: $\tilde{\mu}_{p_0} = E_{p(x)}\left[\mu(x,n,p_0)\right]$. Across a set of priors $p_0 \in F$, one can find the minimum and maximum of $\tilde{\mu}_{p_0}$.

4.2.4. Imprecision in the variance

Just as the mean of the posterior depends on both the priors and the experimental results, so does the variance. The variance is a traditional measure of uncertainty in precise formulations of probability. Even in an imprecise approach, the analyst would like the variance to be as small as possible. However, the variance is no longer a precise measure, but rather an interval for each possible result. Strictly speaking, if the analyst requires a posterior variance below some threshold, then he must consider the maximum variance across all combinations of prior distributions and possible experimental results.

As we did with the mean, the analyst could calculate the expected posterior variance across all results given the prior. When the prior is precise, this yields a single number. When the prior is imprecise, this also results in an interval. The motivation for such as approach is that although the experimental results may not match the prior mean estimate, the analyst believes (by definition) that the priors are a reasonably accurate model of the results. For example, assume the prior mean is the range $[0.05, 0.10]$. Then if one performs 20 experiments, it is highly unlikely that one will observe 20 failures. Thus, this result is discounted by its improbability, unlike the maximum variance approach that would consider this extreme case on an equal footing with all others.

One can also consider measuring the imprecision using the range of the variance across the set of probability distributions. The imprecision in the variance 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 prior imprecision in the variance, the posterior imprecision in the variance given a particular result, and the maximum imprecision over all results are shown in Equations (12)–(14) respectively.

$$\Delta_0(V) = \max_{p \in F} V\left[\theta \middle| p\right] - \min_{p \in F} V\left[\theta \middle| p\right] \tag{12}$$

$$\Delta_{n,F'}(V) = \max_{p' \in F'} V\left[\theta \middle| p'\right] - \min_{p' \in F'} V\left[\theta \middle| p'\right] \tag{13}$$

$$\Delta_{\max}(V) = \max_{F'} \left\{\Delta_{n,F'}(V)\right\} \tag{14}$$

To estimate these measures, we will select distributions for each F' by updating the distributions from F that are generated using the procedure described in Section 4.2.1.

5. Results

As mentioned above, in this paper we are primarily concerned with determining which component should receive more tests. To illustrate the approaches to evaluating test plans, we will consider the example of Section 2 using both precise priors and imprecise priors about the failure probabilities of the three components. We consider two scenarios for each approach.

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. The high prior mean for Component C is chosen for illustrative purposes; it is unlikely that any real system would have a component with such a high mean estimate of probability of failure. For the distribution of the system failure probability, the prior mean equals 0.2201, and the prior variance equals 0.0203.

Table 1. Priors for Scenario 1

Component	A	B	C
Beta parameters	$t_0 = 0.15$ $s_0 = 10$	$t_0 = 0.15$ $s_0 = 2$	$t_0 = 0.55$ $s_0 = 10$

5.1.1. Scenario 1: Variance-based Sensitivity Analysis

The variance-based sensitivity analysis gives the following values:

$$SV_A = 0.4814$$
$$SV_B = 0.4583$$
$$SV_C = 0.0181$$

These values indicate that the output variance is similarly dependent on the variance of the failure probabilities for Components A and B. It is highly insensitive to component C. This suggests that testing be focused on Components A and B, but it does not suggest the appropriate allocation between them.

5.1.2. Scenario 1: Observing Mean and Variance for different results

Figure 1 shows the dispersion of the mean and variance for seven test plans: (1) 12 tests of Component A, (2) 12 tests of Component B, (3) 12 tests of Component C, (4) 4 tests of each component, (5) 6 tests of Components A and B, (6) 6 tests of Components A and C, and (7) 6 tests of Components B and C. The multiple points for each test plan correspond to the set of possible outcomes of the test.

Figure 1 reveals that the test plan makes a significant difference in the mean and variance of the possible posterior distributions. Because $E\left[P_C\right]$ is near 0.5 and the $s_0 = 10$, test plan 3 can change $E\left[P_C\right]$ very little for any test result. When Component C fails, Component B becomes serially connected to Component A and its influence on the system failure is greatly increased.

Test plans 1 and 2, on the other hand, can change the mean a great deal, from a low near 0.15 to a max near 0.65, and can substantially reduce $V(\theta)$. Test plans 4, 5, 6, and 7 likewise have a large range of possible posterior means. Test plans 4, 6, and 7 have generally larger posterior variance than test plan 5, which tests only the two components with the largest sensitivity indices. In this scenario, it seems that testing the components with the largest sensitivity indices is a

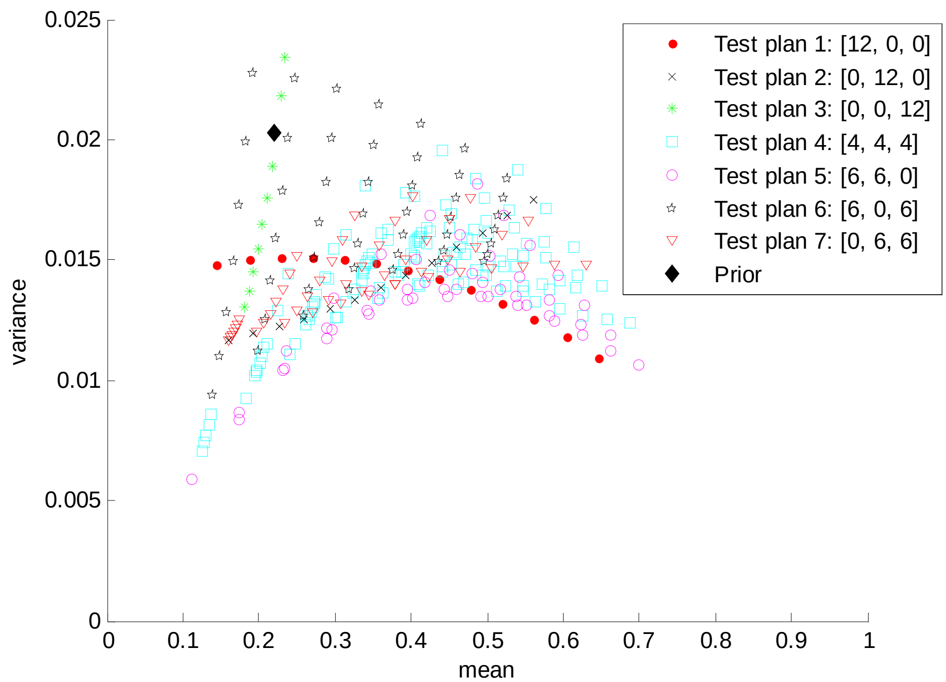

Figure 1. Dispersion of the mean and variance of the system failure probability for different test plans for Scenario 1.

worthwhile plan because this testing can reduce the variance of the corresponding component failure probability distributions, which has a large impact on the variance of the system failure probability distribution.

5.1.3. Posterior variance

Table 2 shows the minimum and maximum variance of $V(\theta)$, the posterior system failure probability distribution, for each of the seven test plans (taken over the possible results for each plan). Test plans 1 and 2 test the components with the largest sensitivity indices and reduce variance significantly. Test plan 5 can significantly reduce variance, as it has a significantly lower minimum variance. Its maximum variance is moderate, as poor test results for both Components A and B would increase the $V(P_A)$ and $V(P_B)$, increasing $V(\theta)$. Test plan 4 has similar performance. Test plan 3 has the largest minimum and maximum posterior variance. Reducing the variance of Component C's failure probability distribution cannot reduce $V(\theta)$ much, but poor test results could increase $E[P_C]$ greatly, which in this case increases $V(\theta)$ by making it more sensitive to the highly uncertain performance of Component B as in Equation (7).

Table 2. Posterior variance for scenario 1

Test Plan	Posterior Variance Across Test Results	
$\#:\{n_A, n_B, n_C\}$	Min	Max
$1:\{12,0,0\}$	0.0110	0.0151
$2:\{0,12,0\}$	0.0117	0.0175
$3:\{0,0,12\}$	0.0131	0.0291
$4:\{4,4,4\}$	0.0071	0.0195
$5:\{6,6,0\}$	0.0059	0.0181
$6:\{6,0,6\}$	0.0094	0.0228
$7:\{0,6,6\}$	0.0117	0.0177

5.2. SCENARIO 2

In the second scenario, as in the first, the priors for the failure probability distributions are precise beta distributions. The parameters are shown in Table 3. The difference from Scenario 1 is that Component C now has the same distribution as Component A. For the distribution of the system failure probability, the mean equals 0.1691, and the variance equals 0.0116.

Table 3. Priors for Scenario 2

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

5.2.1. Scenario 2: Variance-based Sensitivity Analysis

The variance-based sensitivity analysis gives the following values:

$$SV_A = 0.8982$$
$$SV_B = 0.0560$$
$$SV_C = 0.0153$$

The important position of Component A yields a large sensitivity index. Compared to Scenario 1, $E[P_C]$ is now much smaller, which reduces SV_B, as suggested by Equation (7). These values suggest that reducing $V(P_A)$ by testing Component A should reduce $V(\theta)$ significantly.

5.2.2. Scenario 2: Observing Mean and Variance for different results

Figure 2 shows the dispersion of the mean and variance of the posterior mean and variance of the system failure probability distribution for seven different test plans (the same test plans used in Scenario 1). 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 $V(\theta)$. In this scenario, the mean-variance dispersion confirms the suggestion made by the variance-based sensitivity analysis: testing Component A appears to be the best strategy.

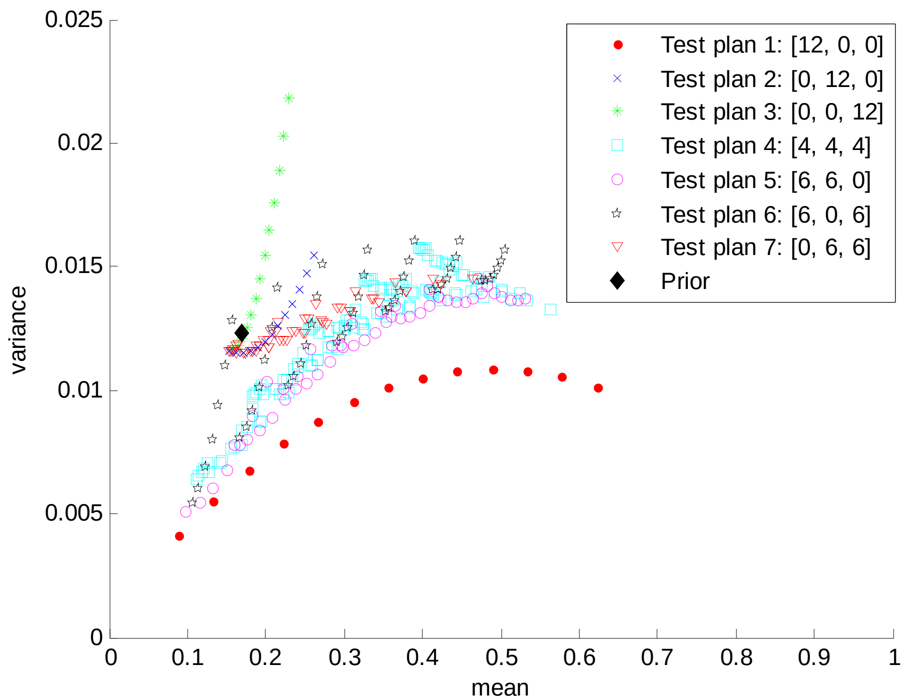

Figure 2. Dispersion of the mean and variance of the system failure probability for different test plans for Scenario 2.

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 $V(P_C)$. Moreover, the fact that $SV_B > SV_C$ suggests that this impact will be larger on $V(\theta)$ than on $E[\theta]$, which we see by comparing test plans 2 and 3 and then comparing test plans 5 and 6. Testing Component B reduces $V(\theta)$ more than testing Component C.

If we look at the possible results of test plan 2, which tests Component B, we see an interesting "hook" pattern that occurs because the test results with a small number of failures tend to confirm the prior, which reduces $V(P_B)$. However, results with more failures increase both $E[P_B]$ and $V(P_B)$, which increase $E[\theta]$ and $V(\theta)$.

5.2.3. Posterior variance

Table 4 shows the minimum and maximum variance of $V(\theta)$, the posterior system failure probability distribution, for each of the seven test plans (taken over the possible results for each plan). Test plans 1, 4, 5, and 6 all have low minimum $V(\theta)$ because all test Component A and can lower $V(P_A)$, which make a large impact, as we know because Component A has the largest sensitivity index. Test plan 1 also has a low maximum $V(\theta)$, which makes this test plan particularly desirable. As in Scenario 1, test plan 3 has the largest minimum and maximum posterior variance. Reducing the variance of Component C's failure probability distribution cannot reduce $V(\theta)$, but poor test results could increase $E[P_C]$ greatly, which in this case increases $V(\theta)$ because the large variance $V(P_B)$ becomes more important, as in Equation (7).

Table 4. Posterior variance for scenario 2

Test Plan #: $\{n_A, n_B, n_C\}$	Posterior Variance Across Test Results	
	Min	Max
1: {12, 0, 0}	0.0042	0.0109
2: {0, 12, 0}	0.0115	0.0155
3: {0, 0, 12}	0.0116	0.0218
4: {4, 4, 4}	0.0064	0.0158
5: {6, 6, 0}	0.0051	0.0145
6: {6, 0, 6}	0.0054	0.0160
7: {0, 6, 6}	0.0115	0.0145

5.3. SCENARIO 3

In the third scenario, prior distributions are imprecise (we use the imprecise beta model parameterized by *t* and *s*). Table 5 lists the parameters for each component's failure probability distribution. We assume that less is known about Component B than the other components, as indicated by the small values for s and the large range for t. The estimated probability of failure of Component C is assumed to be quite large; while these values may not make sense in a real system, they are illustrative of interesting information management behavior. Note that the precise priors given for the first scenario (Table 1) are included in these sets. For selecting priors for the numerical results, as discussed in Section 4.2.1, we use $N_e = 3$.

Table 5. Imprecise priors for Scenario 3

Component	A	B	C
Imprecise beta parameters	$\underline{t}_0 = 0.15$ $\overline{t}_0 = 0.20$ $\underline{s}_0 = 10$ $\overline{s}_0 = 12$	$\underline{t}_0 = 0.15$ $\overline{t}_0 = 0.55$ $\underline{s}_0 = 2$ $\overline{s}_0 = 5$	$\underline{t}_0 = 0.55$ $\overline{t}_0 = 0.60$ $\underline{s}_0 = 10$ $\overline{s}_0 = 12$

The mean of the system failure probability distribution ranges from 0.2201 to 0.4640, which is an imprecision of 0.2439. The variance ranges from 0.0136 to 0.0332, which has an imprecision of 0.0196.

5.3.1. Scenario 3: Variance-based sensitivity analysis

The imprecise variance-based sensitivity analysis yields the results shown in Table 2. These results suggest that it is similarly important to test A and B (using the upper bounds), and it is much less important to test component C. The maximum for component C is about the same as the minimum for Component B, so the values strongly suggest that testing B is more valuable than testing C.

Table 6. Imprecise variance-based sensitivity analysis Scenario 3

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

5.3.2. Scenario 3: Dispersion of mean and variance

We will consider the same seven test plans used in Scenarios 1 and 2. Based on the sensitivity indices, 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 3. In general, a figure showing all of these points gets very difficult to display and view due to overlap. Consequently, we generally will display just the convex hull of each result of each test plan, as shown in Figure 4, which are clearer when viewed in color. While these sets of points are not always convex, this approximation is reasonable for the qualitative analysis performed with them.

All three test plans (1, 2, and 5) can significantly change $E[\theta]$. The impact of test plan 2 (0, 12, 0) is mitigated by the system structure, in which Component B is parallel to Component C. The maximum $V(\theta)$ of test plan 1 (12, 0, 0) is much greater than 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 5 shows the convex hulls for the results of test plans 3 (0, 0, 12), 6 (6, 0, 6), and 7 (0, 6, 6). Test plan 6 leads to results that have a wide range of means and variances. Test plan 3 also has results with large variance, though not as large a range as test plan 6. The results for test plan 7 are similar to those of test plan 5 (shown in Figure 4), but the variance is not as small. As suggested by the sensitivity indices, testing A has more impact than testing C.

Figure 6 includes the convex hulls for the results of test plan 4 (4, 4, 4), as well as the promising plans of 2, 5, and 7. Plan 4 yields results that are quite close to those of test plan 5. Because it tests all three components and can change all three mean values, the max $E[\theta]$ is larger in the results of test plan 4. The extreme results of test plan 5 were limited by no change in $E[P_C]$.

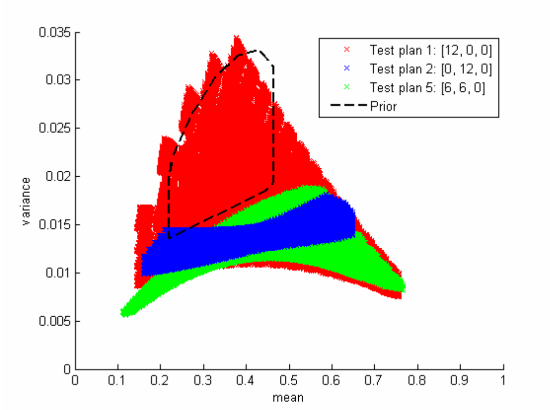

Figure 3. Sample results for test plans 1, 2, and 5 for Scenario 3.

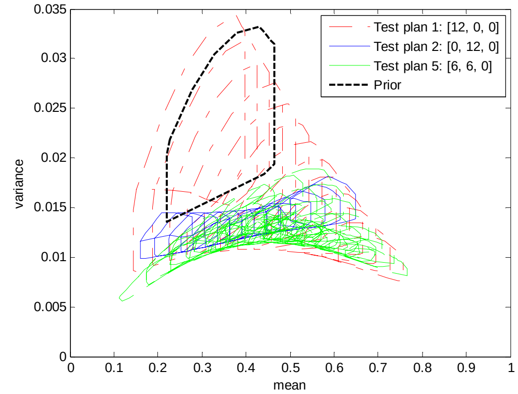

Figure 4. Convex hull for each result of test plans 1, 2, and 5 for Scenario 3.

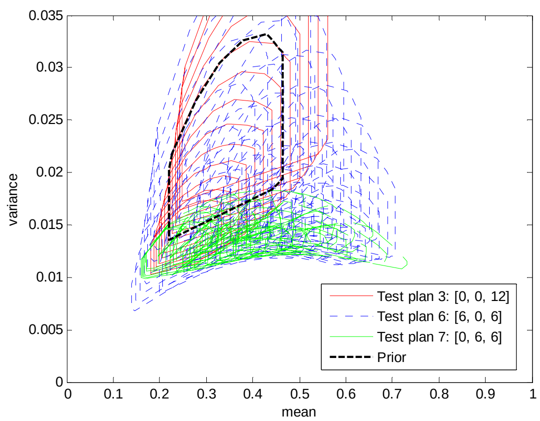

Figure 5. Convex hull for each result of test plans 3, 6, and 7 for Scenario 3.

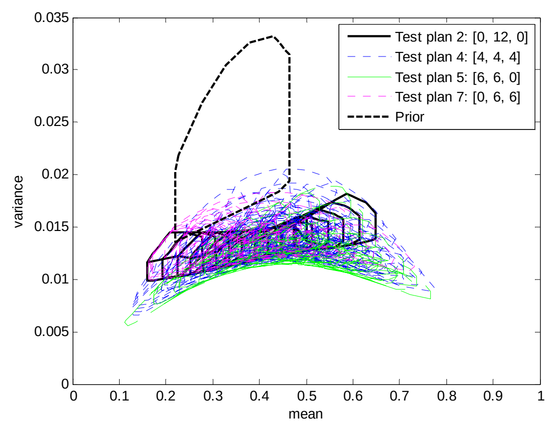

Figure 6. Convex hull for each result of test plans 2, 4, 5, and 7 for Scenario 3.

5.3.3. Scenario 3: Imprecision in the mean

Table 7 describes the imprecision of the $E[\theta]$ 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 (10), and the minimum and maximum are taken over the possible test results.

In these results, test plan 4 (4, 4, 4) stands out for its low minimum minimum, low minimum average, and low maximum average. This occurs because this test plan is more likely to have zero failures (than other test plans that run more tests of a component) and it includes the possibility of zero failures of all three components. Either result would significantly reduce the components' means and thus $E[\theta]$. Such a result would also leave little imprecision in $E[\theta]$, as indicated in its very low minimum imprecision. 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 to not test Component C based on the sensitivity indices is contradicted.

Table 7. Posterior mean analysis for scenario 3

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	0.2201	0.4640	n.a.	n.a.	0.2439	
1: {12,0,0}	0.1451	0.7564	0.2143	0.4680	0.1463	0.2519
2: {0,12,0}	0.1600	0.6491	0.2113	0.4766	0.1085	0.1486
3: {0,0,12}	0.1819	0.5600	0.2192	0.4678	0.1501	0.3112
4: {4,4,4}	0.1119	0.6367	0.1644	0.2916	0.0709	0.1817
5: {6,6,0}	0.1124	0.7662	0.2116	0.4778	0.1172	0.1952
6: {6,0,6}	0.1405	0.7063	0.2153	0.4690	0.1474	0.3019
7: {0,6,6}	0.1610	0.7325	0.2151	0.4773	0.1126	0.2174

5.3.4. Scenario 3: Imprecision in the variance

Table 8 describes the imprecision of $V[\theta]$ that can result from the various test plans. The table structure and results shown are similar to those of Table 7. In these results, test plans 4 and 5 (which test both Components A and B) are notable for their low values on all of the measures. Both plans reduce the variance associated with these components' failure probability distributions, which can significantly reduce $V[\theta]$, as the sensitivity indices indicate.

As mentioned above, 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 $V(P_A)$. Consequently, when comparing plans that test Component B to those that test Component A, we see that test plan 2 reduces $V[\theta]$ and the imprecision in $V[\theta]$ more than test plan 1, and test plan 7 reduces these measures more than test plan 6. The exceptions are the minimum-minimum and minimum average because test plans 1 and 6 include the possibility of

Table 8. Posterior variance analysis for scenario 3

Test Design #: $\{n_A, n_B, n_C\}$	$V[\theta]$				Imprecision in $V[\theta]$	
	Minimum minimum	Maximum maximum	Minimum average	Maximum average	Minimum	Maximum
Prior	0.0136	0.0332	n.a.	n.a.	0.0196	
1: {12,0,0}	0.0075	0.0344	0.0094	0.0304	0.0046	0.0259
2: {0,12,0}	0.0099	0.0181	0.0103	0.0153	0.0035	0.0051
3: {0,0,12}	0.0103	0.0465	0.0134	0.0310	0.0070	0.0293
4: {4,4,4}	0.0059	0.0162	0.0075	0.0118	0.0020	0.0054
5: {6,6,0}	0.0056	0.0189	0.0083	0.0150	0.0022	0.0063
6: {6,0,6}	0.0068	0.0458	0.0107	0.0295	0.0041	0.0309
7: {0,6,6}	0.0100	0.0183	0.0109	0.0183	0.0026	0.0060

dramatically reducing $V(P_A)$ and $V[\theta]$ if no failures are observed.

These results are more consistent with the sensitivity indices. Testing just Component C leads to the worst performance (according to most metrics). However, test 4, in which all three components are tested, performs very well, even though it includes testing C. This is because testing A and B change the actual sensitivities. This is related to the difference between batch testing and sequential (i.e. one-at-a-time) testing.

5.4. SCENARIO 4

For the fourth scenario, consider the imprecise prior distributions given in Table 9. The difference from Scenario 3 is only in component C: we now assume the probability of failure is believed to be much lower and more realistic. Note that the precise priors given for the first scenario are included in these sets. For selecting priors for the numerical results, as discussed in Section 4.2.1, we use $N_e = 3$.

Table 9. Imprecise priors for Scenario 4

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

The mean of the system failure probability distribution ranges from 0.1691to 0.2880, which is an imprecision of 0.1189. The variance ranges from 0.0100 to 0.0173, which is an imprecision of 0.0073.

5.4.1. *Scenario 4: Variance-based sensitivity analysis*

The imprecise variance-based sensitivity analysis yields the results shown in Table 10. SV_A and SV_C have remained roughly the same. Compared to Scenario 3, SV_B has dropped due to the drop in $E[P_C]$. These results suggest that testing Component A and reducing its variance will have the most impact on reducing $V(\theta)$.

Table 10. Imprecise variance-based sensitivity analysis Scenario 4

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

5.4.2. *Scenario 4: Dispersion of mean and variance*

We will consider the same seven test plans used in the previous scenarios. Based on the sensitivity indices, it appears that test plan 1 (12, 0, 0) should have 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 7–Figure 9 show the convex hull of each result of each test plan.

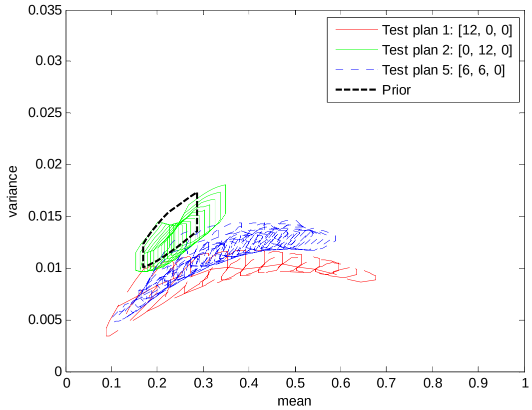

Figure 7. Convex hull of each result for Scenario 4, test plans 1, 2, and 5.

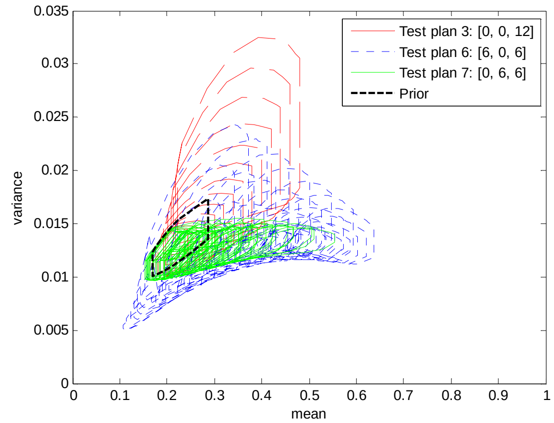

Figure 8. Convex hull of each result for Scenario 4, test plans 3, 6, and 7.

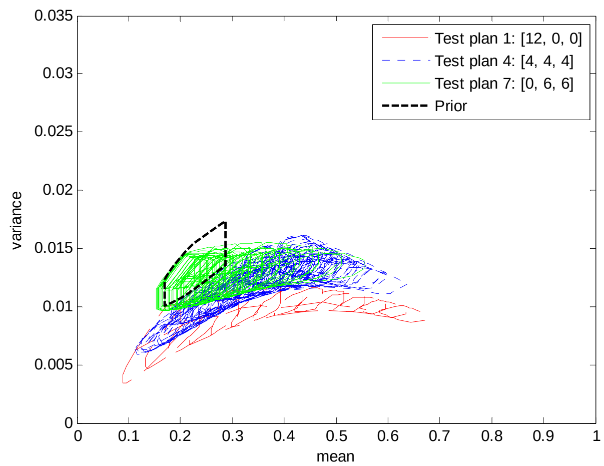

Figure 9. Convex hull of each result for Scenario 4, Tests 1, 4 and 7

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 again the behavior noted in Scenario 2 (the "hook" in Figure 2), but now multiplied for a number of priors. The entire convex hull follows this trajectory. For a given prior, when the test results confirm the prior, testing Component B reduces $V(P_B)$. However, poor test results 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 sensivity 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 $V(\theta)$.

5.4.3. Scenario 4: Imprecision in the mean

Table 11 describes the imprecision of $E[\theta]$ that can result from the various test plans. The table structure and the types of results shown are identical to those of Table 7. 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.

As in Scenario 3, 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 don't 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, though not to the same degree, test plan 4 can add imprecision. As noted in the results of Scenario 3, testing just Component A (as in test plans 1) 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 A and B are tested equally in test plan 5.

Table 11. Posterior mean analysis for scenario 4

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	0.1691	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

5.4.4. Scenario 4: Imprecision in the variance

Table 12 describes the imprecision of $V[\theta]$ that can result from the various test plans. The table structure and types of results shown are identical to those of Table 8. 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 2, poor test results for Component C can greatly increase $V[\theta]$, and we see that here in the maximum maximum for test plan 3.

Unlike the results for the mean in Table 11, 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 $V\left(P_A\right)$, which reduces $V\left[\theta\right]$, as the sensitivity indices indicate. The other test plans have less impact because the sensitivity indices of the other components are smaller. All of the test plans except test plans 3 and 6 (which can greatly increase $V\left[\theta\right]$) reduce the imprecision in $V\left[\theta\right]$.

Table 12. Posterior variance analysis for scenario 4

Test Design #: $\{n_A, n_B, n_C\}$	$V\left[\theta\right]$				Imprecision in $V\left[\theta\right]$	
	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

6. Discussion

The above results, though for specific scenarios and a specific system design, demonstrate some principles that we believe are generally applicable to problems of this type.

First, examining the dispersion of the mean and variance is a useful way to determine the possible outcomes of a test plan. Comparing different dispersions can identify which plans 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 won't 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.

However, the sensitivity indices do not give adequate insight into joint testing—that is, testing multiple components. In Scenario 3, the sensitivity indices clearly suggested that testing Component C was much less important than testing A or B. However, the smallest maximum-maximum and second smallest minimum-minimum posterior variances actually occur with test plan 4, which tests all three components equally (see Table 8). This test plan also yields the smallest maximum imprecision in the variance, 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 in that not only does the variance have the smallest maximum, but it 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 sensitivity index does not give much insight into how testing that component will affect the imprecision of the system-level mean. While expected, since they deal with the variance and not the mean, the results confirm this result. 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 4, the sensitivity indices clearly suggest that testing Component A is most important, and from a variance perspective, it is. However, the best reduction in the imprecision of the mean actually occurs in test plan 5, when both A and B are tested. Similarly, in Scenario 3, the best reduction in imprecision in the mean goes from either testing just A or testing all three equally (Table 7), although the sensitivity indices clearly suggested that testing C was unimportant, and were relatively inconclusive between A and B.

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 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 not very useful. In Scenario 4, they change very little from the values for the prior. In Scenario 3, they can change significantly, but the dispersion plot will show this as well. Additionally, the minimum-minimum and maximum-maximum metrics yield similar rankings to those from the minimum-average and maximum-average respectively. Theoretically, the average metrics give a more accurate insight into the actual posterior means and variances that would result from test plans, but as far as choosing a test plan, it is only the ranking that matters. Additionally, the average values are computationally more expensive to compute.

In this example, many posterior statistics were analytically computable, as shown in Equations (1)-(5) and Equation (7). In general, the posterior system distribution would need to be calculating 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 and compared different strategies for measuring the uncertainty of precise and imprecise distributions for use in making test planning decisions. 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 optimization 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 J. W. Herrmann. Updating Uncertainty Assessments: A Comparison of Statistical Approaches. *2007 ASME International Design Engineering Technical Conferences*. Las Vegas, NV, 2007.

Aughenbaugh, J. M. and C. J. J. Paredis. The Value of Using Imprecise Probabilities in Engineering Design. *Journal of Mechanical Design* **128**(4): 969-979, 2006.

Berger, J. O. The Robust Bayesian Viewpoint. *Robustness of Bayesian Analysis*. J. B. Kadane, ed. New York, North-Holland**:** 63-124, 1984.

---. *Statistical Decision Theory and Bayesian Analysis* (2nd edn.), Springer, 1985.

---. *An Overview of Robust Bayesian Analysis*. Report num. Technical Report #93-53c. West Lafayette, IN, Purdue University, 1993.

Berger, J. O. and J. M. Bernardo. On the Development of the Reference Prior Method. *Bayesian Statistics 4*. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds. Oxford, Oxford University Press, 1992.

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

Bradley, S. R. and A. M. Agogino. An Intelligent Real Time Design Methodology for Component Selection: An Approach to Managing Uncertainty. *Journal of Mechanical Design* **116**(4): 980-988, 1994.

Bruns, M. and C. J. J. Paredis. Numerical Methods for Propagating Imprecise Uncertainty. *2006 ASME International Design Engineering Technical Conferences and Computers in Information Engineering Conference*, Philadelphia, PA, DETC2006-99237, 2006.

Chan, K., S. Tarantola, A. Saltelli and I. M. Sobol. Variance-Based Methods. *Sensitivity Analysis*. A. Saltelli, K. Chan and E. M. Scott, eds. New York, Wiley, 2000.

Coolen, F. P. A. On Bernoulli Experiments with Imprecise Prior Probabilities. *The Statistician* **43**(1): 155-167, 1994.

---. On the Use of Imprecise Probabilities in Reliability. *Quality and Reliability in Engineering International* **20**: 193-202, 2004.

de Finetti, B. *Theory of Probability Volume 1: A Critical Introductory Treatment*. New York, Wiley, 1974.

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

Fougere, P., ed. *Maximum Entropy and Bayesian Methods*. Dordrecht, Klower Academic Publishers, 1990.

Gupta, M. M. Intelligence, Uncertainty and Information. *Analysis and Management of Uncertainty: Theory and Applications*. B. M. Ayyub, M. M. Gupta and L. N. Kanal, eds. New York, North-Holland**:** 3-11, 1992.

Hall, J. W. Uncertainty-Based Sensitivity Indices for Imprecise Probability Distributions. *Reliability Engineering & System Safety* **91**(10-11): 1443-1451, 2006.

Howard, R. A. Information Value Theory. *IEEE Transactions on Systems Science and Cybernetics* **SEC-2**(1): 22, 1966.

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

Jeffreys, H. *Theory of Probability* (3rd edn.). London, Oxford University Press, 1961.

Klir, G. J. and R. M. Smith. On Measuring Uncertainty and Uncertainty-Based Information: Recent Developments. *Annals of Mathematics and Artificial Intelligence* **32**(1-4): 5-33, 2001.

Kokkolaras, M., Z. P. Mourelatos and P. Y. Papalambros. Impact of Uncertainty Quantification on Design: An Engire Optimisation Case Study. *International Journal of Reliability and Safety* **1**(1/2): 225-237, 2006.

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

Ling, J. M., J. M. Aughenbaugh and C. J. J. Paredis. Managing the Collection of Information under Uncertainty Using Information Economics. *Journal of Mechanical Design* **128**(4): 980-990, 2006.

Malak, R. J., Jr., J. M. Aughenbaugh and C. J. J. Paredis. Multi-Attribute Utility Analysis in Set-Based Conceptual Design. *Journal of Computer Aided Design* **(in press)**, 2007.

Maritz, J. S. and T. Lewin. *Empirical Bayes Methods* (2nd edn.). London, Chapman and Hall, 1989.

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

Matheson, J. E. The Economic Value of Analysis and Computation. *IEEE Transactions on Systems Science and Cybernetics* **SSC-4**(3): 325-332, 1968.

Nikolaidis, E., S. Chen, H. Cudney, R. T. Haftka and R. Rosca. Comparison of Probability and Possibility for Design against Catastrophic Failure under Uncertainty. *Journal of Mechanical Design* **126**(3): 386-394, 2004.

Oberkampf, W. L., J. C. Helton, C. A. Joslyn, S. F. Wojtkiewicz and S. Ferson. *Challenge Problems: Uncertainty in System Response Given Uncertain Parameters*, 2001.

Pham-Gia, T. and N. Turkkan. Sample Size Determination in Bayesian Analysis. *The Statistician* **41**(4): 389-397, 1992.

Radhakrishnan, R. and D. A. McAdams. A Methodology for Model Selection in Engineering Design. *Journal of Mechanical Design* **127**(May): 378-387, 2005.

Rekuc, S. J., J. M. Aughenbaugh, M. Bruns and C. J. J. Paredis. Eliminating Design Alternatives Based on Imprecise Information. *Society of Automotive Engineering World Congress* Detroit, MI, 2006-01-0272, 2006.

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

Sobol, I. M. Sensitivity Analysis for Non-Linear Mathematical Models. *Mathematical modeling and computational experiment* **1**(1): 407-414, 1993.

Soundappan, P., E. Nikolaidis, R. T. Haftka, R. Grandhi and R. Canfield. Comparison of Evidence Theory and Bayesian Theory for Uncertainty Modeling. *Reliability Engineering & System Safety* **85**(1-3): 295-311, 2004.

Utkin, L. V. Interval Reliability of Typical Systems with Partially Known Probabilities. *European Journal of Operational Research* **153**(3 SPEC ISS): 790-802, 2004a.

---. Reliability Models of M-out-of-N Systems under Incomplete Information. *Computers and Operations Research* **31**(10): 1681-1702, 2004b.

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

Walley, P., L. Gurrin and P. Burton. Analysis of Clinical Data Using Imprecise Prior Probabilities. *The Statistician* **45**(4): 457-485, 1996.

Winkler, R. L. Uncertainty in Probabilistic Risk Assessment. *Reliability Engineering & System Safety* **54**(2-3): 127-132, 1996.

Zellner, A. Maximal Data Information Prior Distributions. *New Methods in the Applications of Bayesian Methods*. A. Aykac and C. Brumat, eds. Amsterdam, North-Holland, 1977.