Reliability-Based Decision Making: A Comparison of Statistical Approaches

J. M. Aughenbaugh, *Applied Research Laboratories,*
University of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029, USA.
Email: jason@arlut.utexas.edu
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

This paper considers the problem of choosing between an existing component whose reliability is well established and a new component that has an unknown reliability. In some scenarios, the designer may have some initial beliefs about the new component's reliability. The designer may also have the opportunity to obtain more information and to update these beliefs. Then, based on these updated beliefs, the designer must make a decision between the two components. This paper examines the statistical approaches for updating reliability assessments and the decision policy that the designer uses. We consider four statistical approaches for modeling the uncertainty about the new component and updating assessments of its reliability: a classical approach, a precise Bayesian approach, a robust Bayesian approach, and an imprecise probability approach. The paper investigates the impact of different approaches on the decision between the components and compares them. In particular, given that the test results are random, the paper considers the likelihood of making a correct decision with each statistical approach under different scenarios of available information and true reliability. In this way, the emphasis is on practical comparisons of the policies rather than on philosophical arguments.

AMS Subject Classification: 62F10, 62F15, and 62F35

Keywords: Reliability assessment; Bayesian statistics; imprecise probabilities.

---

1 Introduction

Engineers make decisions based on their beliefs. These beliefs depend upon the information that the engineer has gathered, and they can change based on new information (e.g. from experts, experiments, and experience). Statistical reasoning includes a variety of approaches for updating beliefs based on new information. From a practical perspective, it can be argued that engineers would like a procedure for reliability assessment that can use existing information, incorporate new information, and help determine what additional information, if any, is needed. Previous work (Aughenbaugh and Herrmann, 2007) has considered the parameter estimation problem. However, it is ultimately the quality of the final decision that matters.

The goal of this article is to explore how different statistical methods for updating the designer's beliefs lead to correct, incorrect, or inconclusive decisions under different scenarios of available information and actual reliability. Naturally, this also involves comparing the use of different decision policies. We consider a specific design decision-making problem under a number of different information scenarios. The emphasis in the paper is on a practical, rather than a philosophical, comparison of the approaches in a defined context.

The paper begins by stating the problem and introducing three scenarios that correspond to situations in which the designer initially has different amounts of information. After briefly describing the statistical approaches that will be considered, the paper then discusses how they and the decision policy affect the accuracy of the decision in the different scenarios.

2 Reliability assessment problem

The performance of a product is unpredictable and often involves random behavior. Multiple items (i.e. instances of the product being designed) will be manufactured, sold, and used. Ideally, these items are identical. In practice, however, differences in materials, manufacturing, and usage exist, and consequently, the possibility of failures, though disagreeable to the designer, is unavoidable. Here, "failure" is taken very generally, but we assume it is a one-time event in the lifetime of an item in which at least one performance requirement is not met.

2.1 Problem statement

We consider the following example. A design team wishes to reduce the cost of a product by replacing one expensive component of the system with a new, lower cost component. Cost is a typical motivation, but other issues could motivate the replacement. Regulations require that the new component must be at least as reliable as the existing component. Consequently, the decision of whether to replace the component depends upon this criterion.

It will be convenient to frame things in terms of failure probability instead of reliability. The current component is assumed to have a failure probability of $\theta_{crit} = 0.05$. Let θ be the failure probability of the new component. In order to replace the old component with the new component, the designer requires that $\theta \leq \theta_{crit}$. We will

consider two cases for the new component's true failure probability: $\theta = 0.04 < \theta_{crit}$ (so the new component is preferred), and $\theta = 0.06 > \theta_{crit}$ (so the old component is preferred).

However, the designer does not know the true value of θ. He or she does have the opportunity to obtain additional data from testing in order to update his assessment of θ. The designer can perform n tests in which the performance—actual or an acceptable surrogate—of the new component is observed. Let $x_i = 1$ if trial i is a failure, and $x_i = 0$ otherwise. Then the number of failures observed is $m = \sum_{i=1}^{n} x_i$. We assume that each test is an independent Bernoulli trial, so m is a binomial random variable.

2.2 Information scenario descriptions

The feasibility and desirability of a given statistical approach for updating beliefs depend upon the scenario, including the amount of data currently available and the beliefs that are currently held. In this paper, we consider the following three scenarios.

Scenario 1: no prior information. In this scenario, the designer has concluded that no relevant information is available. For example, the new component may be completely new, or the environment may be so different that the designer believes that existing data have little value in predicting the performance of the new component. Essentially, the designer needs to construct an estimate of the reliability from scratch. All inferences about the component's reliability will be made using only the data received from the planned experiments.

Scenario 2: substantial prior information. In this scenario, there exists substantial information that the designer believes is relevant to the desired reliability analysis. The designer is considering testing as a way of verifying this information. For example, the new component may be a minor modification of the existing component. Alternatively, the new component may have been used in other similar settings, so its past performance is a good indication of its performance in this new setting.

Scenario 3: partial prior information. In this scenario, the designer believes that information relevant to the desired analysis is available, but the designer is considering testing in order to augment and verify this partial information. For example, perhaps the new component was used previously in another setting that did not stress the component as it will be stressed in the future, or the new component is a significant modification from existing designs so existing data is only partially relevant. Coolen (2004) discusses various scenarios in which only partial information may be available in more general reliability assessment problems.

3 Statistical approaches for updating reliability estimates

This paper considers four approaches for analyzing data and updating beliefs: classical sampling theory; precise Bayesian theory; robust Bayesian theory; and imprecise probability theory. The following subsections discuss these approaches.

3.1 Classical sampling theory approach

The classical, sampling theory approaches to statistical analyses, which are generally emphasized in introductory texts, focus entirely on the observed data—in this case the $\{x_i\}_{i=1}^{n}$. Because the prior information plays no role in the analysis, the classical approach is the same in all three information scenarios.

The ratio $\hat{\theta} = \sum_{i=1}^{n} x_i/n = m/n$, which is the relative frequency of failures to trials in the sample, is an unbiased point estimate of θ. Let z_α be the $100(1 - \alpha)$ percentile of the standard normal distribution. Then, a commonly used, approximate one-sided $100(1 - \alpha)\%$ confidence interval on θ is given by Eq. (3.1):

$$\left[0, \frac{m}{n} + z_\alpha \sqrt{\frac{m}{n} \left(1 - \frac{m}{n} \right) \frac{1}{n}} \right] \tag{3.1}$$

The actual coverage probability of this approximation fluctuates as n varies, an effect that is accentuated for small θ (Brown et al., 2001). We use the one-sided, upper bound for this problem because the designer would like the failure probability to be below the threshold, and only the upper limit of the interval matters for this decision. If this interval includes θ_{crit}, then one cannot reliably conclude that $\theta \leq \theta_{crit}$ based on the test results. Other confidence intervals for estimating this parameter have been proposed, as discussed in Brown et al. (2001).

3.2 Precise Bayesian approach

The Bayesian approach (cf. 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. This prior enables designers to include existing information into 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 (e.g., if the prior is Gaussian, so is the posterior). 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 as a beta distribution with parameters α and β. Specifically, when one starts with the prior distribution $Beta(\alpha_0, \beta_0)$ and observes m failures out of n trials, 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. Prior selection for the example is discussed in Section 3.6.

3.3 Robust Bayesian approach

The robust Bayesian approach, or Bayesian sensitivity analysis, addresses the problem of lack of confidence in the prior (Berger, 1985, 1993; Insua and Ruggeri, 2000; Berger, 1984). The core idea of the approach is to perform a "what-if" analysis: the analyst considers several reasonable prior distributions and performs the update on each to get a set of posterior distributions. This set of posteriors 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.

3.4 Imprecise probabilities approach

In traditional probability theory, the probability of an event is defined by a single number. However, various researchers have found this definition to be too limited when there exists incomplete or conflicting information. They have proposed theories of imprecise probabilities to generalize this to a concept in which probabilities can be intervals or sets, rather than precise numbers (Walley, 1991; Weichselberger, 2000; Dempster, 1967). Other researchers have proposed alternative uncertainty models (Shafer, 1976; Dubois and Prade, 1988).

In this paper, we consider the theory of imprecise probabilities as formalized by Walley (1991) and others. In this theory, uncertainty is still captured by probability, but rather than probabilities being single scalar numbers, a range of indeterminacy is allowed: in a subjective framework, these are 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, compared to the precise Bayesian requirement that exactly one distribution must be used to describe an individual's beliefs. Imprecise probabilities have previously been considered in reliability analysis (cf. Coolen, 2004; Utkin, 2004a,b; Coolen-Schrijner and Coolen, 2007), but a direct comparison to other methods that considers the quality of reliability-based design decisions, rather than philosophical or inference-based arguments, is not apparent in the literature.

For our example, it is convenient to use the imprecise beta model (IBM), described by Walley (Walley, 1991; Walley et al., 1996). The IBM is one special case of the more general 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. The IBM allows for simple updating when considering a Bernoulli process and prior information. Improvement to the IBM have been suggested (Coolen, 1994), and non-parametric methods are also available (Coolen, 1998; Coolen-Schrijner and Coolen, 2007).

For the IBM, the beta model is reparameterized so that the density of $beta(s,t)$ is as given in Eq. (3.2).

$$\pi_{s,t}(\theta) \propto \theta^{st-1}(1-\theta)^{s(1-t)-1} \tag{3.2}$$

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 and s reflects the quantity of information available.

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 Walley

(1991), the parameters can be imprecise. That is, the priors are the set of beta distributions $(\alpha_0, \beta_0) \in F_0 = \{(s_0 t_0, s_0 (1 - t_0)) : \underline{t}_0 \leq t_0 \leq \bar{t}_0, \underline{s}_0 \leq s_0 \leq \bar{s}_0\}$. The posteriors are the set of beta distributions with $\alpha_n = \alpha_0 + m$ and $\beta_n = \beta_0 + n - m$ $\forall (\alpha_0, \beta_0) \in F_0$.

3.5 Connection between robust Bayesian and imprecise probability approaches

There are at least two arguments as to why a designer should not consider a single prior distribution: one practical, and one philosophical. First, because eliciting and assessing an individual's beliefs is a resource intensive process, it will often be impractical to fully characterize them (Walley, 1991; Weber, 1987; Groen and Mosleh, 2005). Consequently, only a partial (imprecise) characterization will be available. This is the view held by advocates of the robust Bayesian approaches. The second argument claims that true beliefs need not be precise. A person may be unwilling to commit to either side of a gamble when he has no information about it. It is possible that one would be satisfied with some price, but this is not a condition for rationality. This view is held by advocates of imprecise probabilities (Walley, 1991, 1996).

Although the motivations differ between the robust Bayesian and the imprecise probabilities approaches, they have some similarities. Both use a set of distributions to analyze a problem. In our reliability assessment problem, 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.

3.6 Prior selection

Selecting priors is an important step. Lindley and Phillips (1976) discuss the process of choosing priors and give some relevant, practical examples. For our reliability assessment problem, we chose different sets of priors (shown in Tables 1 and 2) to illustrate the impact of different amounts of information of varying quality. Naturally, the priors selected affect the results of the comparison of methods. This is even part of the point; the selection of a single precise prior can affect the outcome of the decision problem, and thus one might want to consider a range of priors or at least the sensitivity of the decision to the prior.

With no prior information (Scenario 1), one natural choice for the precise prior is the uniform $[0, 1]$ distribution, which is a special case of the beta distribution ($\alpha_0 = 1$ and $\beta_0 = 1$). This prior is non-informative in that it maximizes the entropy. Others researchers may argue for the invariant prior ($\alpha_0 = 0.5$ and $\beta_0 = 0.5$), which minimizes the Fisher information and has the additional advantage that it is in variant under re-parameterization. Again, this lack of agreement is one motivation for considering sets of priors.

For the IBM, the appropriate starting point is a vacuous prior (so-called because it contains the least information). This prior has $\underline{t}_0 = 0$ and $\bar{t}_0 = 1$, meaning that the probability is somewhere between 0 and 1, the least specific statement that is possible. We must also choose an initial s_0. Because it reflects how much "importance" to assign

Table 1: Bayesian priors for $\theta = 0.06$ case

Approach	No prior info.	Partial prior info. (good)	Partial prior info. (bad)	Substantial prior info. (good)	Substantial prior info. (bad)
Precise Bayesian	$\alpha = 1$	$\alpha = 1.7$	$\alpha = 1.0$	$\alpha = 6$	$\alpha = 4$
	$\beta = 1$	$\beta = 23$	$\beta = 26$	$\beta = 94$	$\beta = 96$
Imprecise Beta Model	$\underline{t}_0 = 0$	$\underline{t}_0 = 0.05$	$\underline{t}_0 = 0.00$	$\underline{t}_0 = 0.055$	$\underline{t}_0 = 0.035$
	$\bar{t}_0 = 1$	$\bar{t}_0 = 0.10$	$\bar{t}_0 = 0.05$	$\bar{t}_0 = 0.065$	$\bar{t}_0 = 0.045$
	$\underline{s}_0 = 0$	$\underline{s}_0 = 20$	$\underline{s}_0 = 20$	$\underline{s}_0 = 80$	$\underline{s}_0 = 80$
	$\bar{s}_0 = 2$	$\bar{s}_0 = 30$	$\bar{s}_0 = 30$	$\bar{s}_0 = 120$	$\bar{s}_0 = 120$

Table 2: Bayesian priors for $\theta = 0.04$ case

Approach	No prior info.	Partial prior info. (good)	Partial prior info. (bad)	Substantial prior info. (good)	Substantial prior info. (bad)
Precise Bayesian	$\alpha = 1$	$\alpha = 1.1$	$\alpha = 1.8$	$\alpha = 4$	$\alpha = 6$
	$\beta = 1$	$\beta = 25$	$\beta = 22$	$\beta = 96$	$\beta = 94$
Imprecise Beta Model	$\underline{t}_0 = 0$	$\underline{t}_0 = 0.00$	$\underline{t}_0 = 0.05$	$\underline{t}_0 = 0.035$	$\underline{t}_0 = 0.055$
	$\bar{t}_0 = 1$	$\bar{t}_0 = 0.05$	$\bar{t}_0 = 0.10$	$\bar{t}_0 = 0.045$	$\bar{t}_0 = 0.065$
	$\underline{s}_0 = 0$	$\underline{s}_0 = 20$	$\underline{s}_0 = 20$	$\underline{s}_0 = 80$	$\underline{s}_0 = 80$
	$\bar{s}_0 = 2$	$\bar{s}_0 = 30$	$\bar{s}_0 = 30$	$\bar{s}_0 = 120$	$\bar{s}_0 = 120$

to the prior data and there is no prior data, one should select a small learning parameter. However, too small a parameter may cause the posterior to react too quickly to the data. As described by Walley (1991), $s_0 = 2$ has a number of good properties. One can also allow for a range of learning parameters, such as $\underline{s}_0 = 0$ and $\bar{s}_0 = 2$. Note that $t_0 = 0.5$ and $s_0 = 2$ corresponds to the uniform distribution and that $t_0 = 0.5$ and $s_0 = 1$ corresponds to the invariant prior distribution. Thus, this set of priors includes the single uniform prior considered in the precise Bayesian approach with no information and an alternative non-informative prior. This highlights the difference between a non-informative (precise) prior and a vacuous prior. The vacuous prior is a set of prior distributions, and this set contains multiple non-informative priors, including the uniform prior and the invariant prior; the vacuous prior is less specific (i.e. more general) than any one non-informative prior.

In Scenario 2 (substantial information), we consider priors that represent the substantial, relevant information. Larger values of α_0 and β_0 generally yield a prior distribution with less variance. We use both "good" priors that are good estimates of the true failure probability (the mean $\alpha_0 / (\alpha_0 + \beta_0)$ is close to the true value of θ) and "bad" priors that are not. The terms "good" and "bad" are relative in nature; the means of the "bad" priors are not as accurate as the means of the "good" priors. For the IBM, we choose $\bar{t}_0$ and $\underline{t}_0$ to be relatively close to each other (reflecting substantial knowledge and a well defined estimate of the mean) and $\bar{s}_0$ and $\underline{s}_0$ to be close to the sum $\alpha_0 + \beta_0$, which is 100 in this case. The priors for the IBM include the corresponding precise priors for each case.

In Scenario 3 (partial information), we use smaller parameter values to reflect the

smaller amount of information, as these values lead to a larger variance. For the IBM, the range for t_0 will be larger since there is more uncertainty in the estimates when there is less information, and the values of s_0 (like the values of α_0 and β_0) are smaller, indicating a smaller "pseudo" sample size for the prior. Again, the priors for the IBM include the corresponding precise priors for each case.

4 Decision accuracy within each scenario

To study the component selection decision, we must explicitly state the policies that the designer will use to make the decision. For convenience, let "accept" mean that the designer accepts the new component because its failure probability appears to be sufficiently small ($\theta < \theta_{\mathrm{crit}}$). Let "reject" mean that the designer rejects the new component because its failure probability appears to be too large. Let "no decision" mean that the designer does not decide on the acceptability of the new component immediately, instead preferring to continue collecting information or to apply an additional decision criteria (see Section 4.3). When the true failure probability is $\theta \leq \theta_{\mathrm{crit}}$, a correct decision accepts the new component, and a wrong decision rejects it. When the true failure probability is $\theta > \theta_{\mathrm{crit}}$, a correct decision rejects the new component, and a wrong decision accepts it.

The designer's decision depends upon the test results, which are random. For a given policy, we can determine which test results lead to which decision (accept or reject). The probability of a decision is the sum of the probabilities of the test results that lead to that decision. We will explore how the probability of a correct decision depends upon the quality and magnitude of the priors, the number of trials conducted, and the decision policy used.

4.1 Classical approach results

In the classical sampling approach, suppose that the designer will accept the new component if the upper bound of the 95% one-sided confidence interval (Eq. 3.1) on θ is less than or equal to θ_{crit}. Otherwise, he rejects it. This is analogous to a one-sided hypothesis test. Figure 1 shows the probability of a correct decision for each value of n in two cases: in the first, $\theta = 0.06 > \theta_{\mathrm{crit}}$; in the second, $\theta = 0.04 < \theta_{\mathrm{crit}}$. The probability of a wrong decision (not shown) equals one minus the probability of a correct decision.

There are several interesting behaviors apparent in Figure 1. First, the classical method performs better when $\theta > \theta_{\mathrm{crit}}$ than it does when $\theta < \theta_{\mathrm{crit}}$. This occurs because the upper bound of the confidence interval is sensitive to the observed test results. Only a limited number of results (those with few observed failures) yield an upper bound that is less than or equal to θ_{crit}, whereas there are many more results that lead to an upper bound that is greater than θ_{crit}. This, combined with the variability of the binomial distribution, make a correct decision (accepting the new component) unlikely when θ is slightly less than θ_{crit}. However, when θ is slightly greater than θ_{crit}, the large number of results that lead to a large upper bound makes the correct decision (rejecting the new component) highly likely.
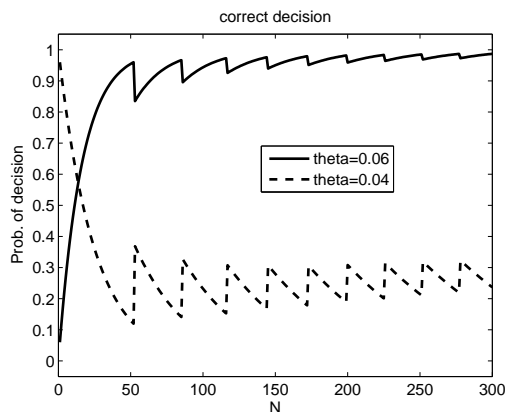
Figure 1: Probability of a correct decision for the classical approach.

A second note is the nature of the curves in Figure 1. They are not smooth, as one initially might expect. Instead, they have a saw-tooth aspect. This leads to non-intuitive results. For example, in the case with $\theta = 0.04$, the probability of a correct decision is larger when $n = 54$ than when $n = 278$. How can more information lead to worse performance? It is due primarily to the discrete nature of the binomial distribution, and it is seen for the other approaches, too. Brown et al. (2001) discuss this further.

4.2 Precise Bayesian approach results

In the precise Bayesian approach, the designer will conclude that $\theta \leq \theta_{\text{crit}}$ and accept the new component if the posterior probability that $\theta \leq \theta_{\text{crit}}$ is at least 0.8. Otherwise, he will conclude that $\theta \geq \theta_{\text{crit}}$ and reject the new component. The selection 0.8 is somewhat arbitrary; some analysts may consider 0.5 sufficient, while others would require a higher value such as 0.95. This issue is addressed further in Section 5. The plots in Figure 2 show the probability of a correct decision for each value of n in two cases: in the first, $\theta = 0.06 > \theta_{\text{crit}}$; in the second, $\theta = 0.04 < \theta_{\text{crit}}$. The different plots in each graph correspond to the different priors from Tables 1 and 2.

The upper plot shows the probability of making a correct decision when $\theta = 0.06$. This probability is highest when the designer starts with a good prior that is based on substantial information. Starting with a good prior that is based on partial information and starting with non-informative prior yield similar results. For small n, the probability of a correct decision is lowest when the designer starts with a bad prior that is based on partial information. For larger n, the probability of a correct decision is lowest when the designer starts with a bad prior that is based on substantial information. The partial, bad prior is more easily influenced and therefore requires less information to push the probabilities over the decision threshold of 0.80. As n increases, this responsiveness causes its posterior to move away from the bad prior faster than in the substantial prior information case.

The lower plot shows the probability of making a correct decision when $\theta = 0.04$.
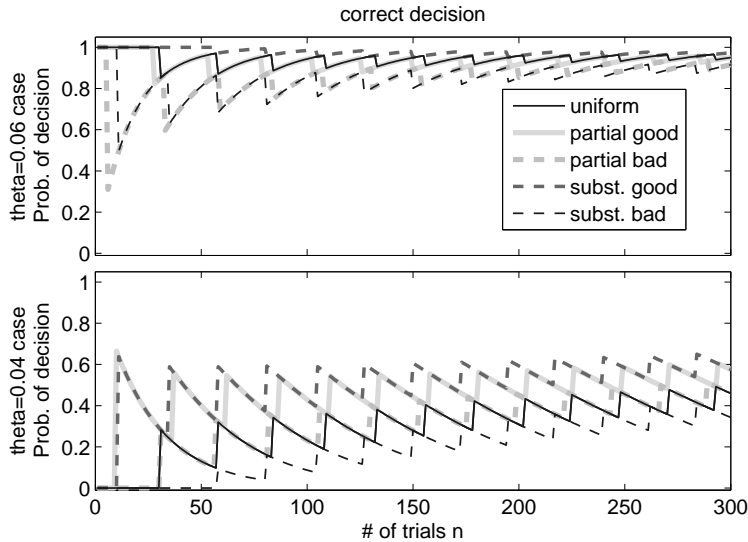
Figure 2: Probability of a correct decision for the precise Bayesian approach.

In general, the results are as expected: having substantial, good prior information leads to the best performance, followed by having partial good information; having substantial, bad prior information leads to the worst performance. Interestingly, the partial, bad prior scenario performs slightly better than the no information, uniform $[0, 1]$ prior. The mean of the partial, bad prior is inaccurate (0.076 is greater than the true value of 0.04), but it is more accurate than the mean (0.50) of uniform [0, 1] prior. Essentially, the partial information, while not accurate, has helped to eliminate very large probabilities of failure from consideration. Thus, one can see that the "noninformative" uniform prior actually provides a substantial amount of "bad" information.

4.3 Imprecise beta model results

With the IBM, the designer will accept the new component if (over the set of posteriors) the minimum probability that $\theta \leq \theta_{\text{crit}}$ is at least 0.80. The designer will reject the new component if (over the set of posteriors) the maximum probability that $\theta \leq \theta_{\text{crit}}$ is less than 0.80. Otherwise, he or she makes no decision on the acceptability of the new component, essentially concluding that the information is still insufficient information to make this decision. In this case, the decision maker may choose to delay the decision (about component acceptability) and collect additional information with the goal of reducing the indeterminacy in the existing information, or he or she may choose to apply an additional decision criterion, such as a mean, an upper bound, or something more formal (cf. Rekuc et al., 2006). In a sense, this is still a decision, but it is not a direct decision in the original context. It involves either extending the problem to allow for additional information collection as an option or refining the decision criteria to allow new methods.
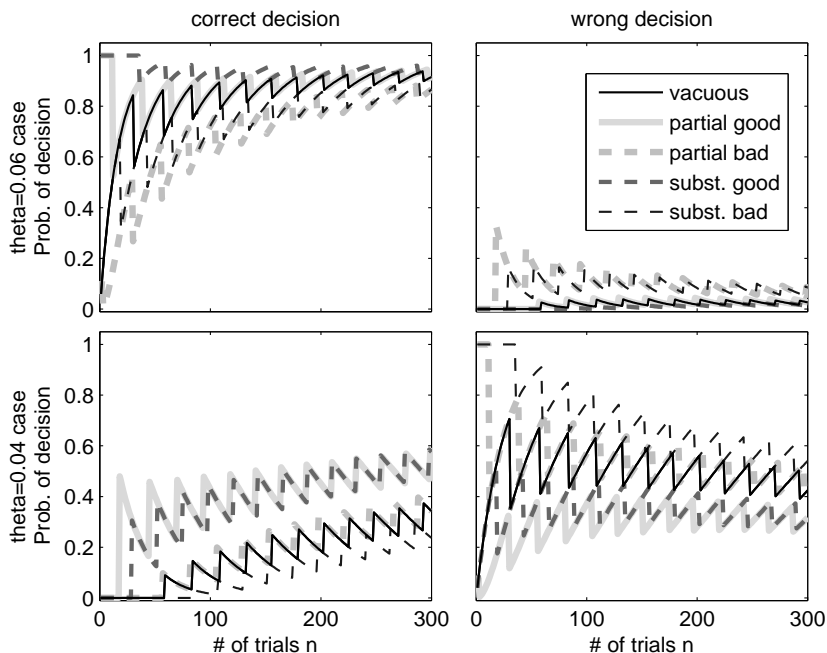
Figure 3: Decision probabilities for the imprecise beta model.

Figure 3 shows the decision probabilities for each value of n in two cases: in the first, $\theta = 0.06 > \theta_{\text{crit}}$; in the second, $\theta = 0.04 < \theta_{\text{crit}}$. The probability of no decision is not shown due to space constraints, but it can be deduced from the other two

4.3.1   Results for the $\theta = 0.04$ scenario

We begin by considering the bottom row of charts in Figure 3, which are for the $\theta = 0.04$ scenario. The probability of making a correct decision (i.e. accepting the new design) is lowest when the designer starts with a bad prior that is based on substantial information. The probability of making a correct decision is higher when starting with a bad prior that is based on partial information because there is less incorrect prior information to overcome. The probability of making a correct decision is highest when the designer starts with a good prior that is based on substantial information or a good prior that is based on partial information.

The partial, bad prior and the vacuous prior lead to similar performance; starting with a small amount of bad information and starting with no information are similar. With the partial, bad prior, one needs enough additional information to counter that prior's inaccuracies. In the no information scenario, one requires enough additional information to lower the upper bound of the point estimate from 1.0 to something less than 0.05.

The performance of the partial, good prior actually exceeds that of the substantial,

good prior for small n, with the performance shifting to favor the substantial, good prior as n approaches 400. This is related to the quality and responsiveness of the priors. With the priors listed in Table 1, the initial bounds on the probability that $\theta \leq \theta_{\text{crit}}$ are $[0.60, 1.00]$ for the partial, good prior case and $[0.64, 0.83]$ for the substantial, good prior case, so both lead to no decision initially. If $n = 20$ trials are performed, the probability of observing $m = 0$ failures equals 0.44 when $\theta = 0.04$. In this case, for the partial, good prior, the posterior bounds on the probability that $\theta \leq \theta_{\text{crit}}$ are $[0.82, 1.00]$, so the designer accepts the new component. For the substantial, good prior, the posterior bounds are $[0.76, 0.93]$, so no conclusion can be reached. As expected, the partial prior is more sensitive to the data. Although this is advantageous in the $\theta = 0.04$ case, it is detrimental in the $\theta = 0.06$ case, as discussed next.

4.3.2 Results for the $\theta = 0.06$ scenario

The probability of making a correct decision when $\theta = 0.06$ is higher in general than it is when $\theta = 0.04$. The probability is highest when the designer starts with a good prior that is based on substantial information. Starting with a good prior that is based on partial information and starting with a vacuous prior perform similarly to each other.

It is interesting that having partial, bad prior information is actually worse than having substantial, bad prior information over the region of n shown. To some extent, this is an artifact of the particular priors that were chosen. For example, the partial, bad prior gives a point estimate of $[0.00, 0.05]$ for θ, whereas the substantial, bad prior gives the interval of $[0.035, 0.045]$. Here, the substantial information estimate is more narrowly focused on a value that is closer to the truth than the broad estimate based on partial information. The partial information scenarios includes the possibility that θ is near zero, which tends to lead to incorrect conclusions in this scenario. For example, the probability that $\theta \leq \theta_{\text{crit}}$ is the interval $[0.18, 0.62]$ for the partial, bad information scenario, whereas the substantial, bad prior gives the interval $[0.27, 0.48]$. The upper bound for the partial information case (0.62), is much closer to the threshold (0.80) for making a bad decision than the upper bound for the substantial information case (0.48).

Since the partial information prior is also more sensitive to the data, more of the possible random results of an experiment will lead to accepting the new component, which is the wrong decision. A more substantial, yet also incorrect, prior will be more robust to the variation in additional samples. These results reflect the complicated trade-offs between the quality and quantity of available information. While the distinctions are sometimes difficult to discern even in the imprecise models, these details are lost completely in a precise analysis. Note that after 200 trials, the probability of a wrong decision is equal for both bad priors. This is because the acquired data is moving the estimate based on the partial, bad prior information through the "no decision" region, as described in the following.

4.3.3 The option of indecision

A major difference between these results and those for the precise Bayesian approach is the option of making no decision when the available information is incomplete or otherwise imprecise. The probability of reaching no conclusion depends greatly upon

the quality and amount of information available. When the prior is a vacuous prior or a bad prior, then there is a larger probability of reaching no conclusion. In the $\theta = 0.06$ scenario, the probability of not reaching a conclusion is highest for the partial, bad prior and the substantial, bad prior. This is desirable; when the prior and data conflict, something is amiss and more data are needed to resolve the conflict.

When there is substantial, good prior information, there is a much smaller probability of reaching no conclusion because on average the prior and data agree with each other. In particular, for the substantial, good prior case, there is initially no chance of reaching no conclusion because the prior always would lead to a correct decision. Even as more data is collected, the probability of getting test results that cancel out the influence of the substantial, good prior is very small, so the probability of making a correct decision remains high.

As the number of samples increases, the probability of no decision will eventually tend to zero, regardless of the prior. However, the probability of no decision may increase initially, a situation that reflects conflict between the prior and the data. The ability of the IBM to reflect both the overall quality and consistency of the combined prior and data evidence appears to be an advantage over the other methods.

5   Comparing decision accuracy across policies in the no prior information scenario

All of the results about the decision probabilities depend upon the policies used to make decisions. Therefore, it is also important to consider two different sets of policies and compare their performance in a specific scenario (we will use Scenario 1, the no information scenario).

The first set uses "generous" policies. For the classical approach, an 80% confidence interval based on Eq. 3.1 is used. For the precise Bayesian approach, the designer accepts the new component if the probability that $\theta \leq \theta_{\mathrm{crit}}$ is at least 0.80. When using the IBM, the minimum probability must be at least 0.80 to accept, and the maximum probability must be less than 0.80 to reject. The results of these policies are shown in Figure 4. The second set uses "strict" policies. The classical approach uses a 95% confidence interval based on Eq. 3.1, and the Bayesian approaches uses a probability of 0.95 instead of the 0.80 of the "generous" policies. The results of these policies are shown in Figure 4.

In the $\theta = 0.04$ scenario, the primary difference between the approaches is most obvious; this difference is their relative caution in concluding that the new component is acceptable. The classical approach is the least conservative approach, while the IBM is the most conservative. Hence, when the new component is acceptable, the classical approach has the highest probability of a correct decision, while the IBM has the lowest probability of a correct decision. The bias towards the "safe" decision (reject the new component) is also apparent in the $\theta = 0.06$ scenario, because all methods perform well when the correct decision is the "safe" decision. This trade-off shares some aspects of the trade-off between type I and type II errors in traditional hypothesis testing, and it confirms some of the interactions discussed in Section 4.

The conservativeness of the approaches is more pronounced when using the stricter
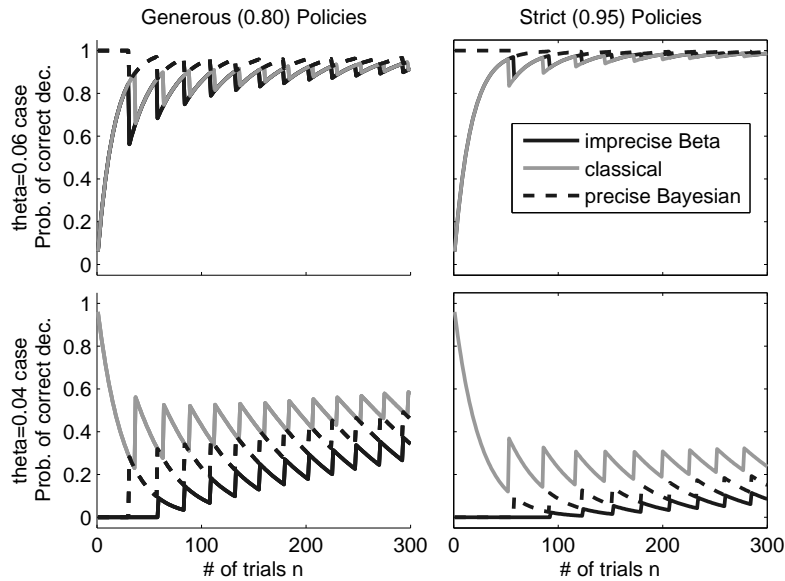
13

Figure 4: Comparing probabilities of correct decisions across policies.

decision policies. This also reveals a tradeoff between decision policy and statistical approach. For a small number of trials, the classical, strict approach yields a higher probability of a correct decision than the IBM, generous approach yields. However, as the number of trials increases, the IBM, generous approach gains the advantage.

Figure 4 reveals that the IBM yields a lower probability of making a correct decision, but the IBM also yields a very low probability of making wrong decision (not shown), equaling or exceeding the performance of the other approaches. Instead of reaching a concrete decision based on little or conflicting information (as the precise Bayesian approach might do), the IBM acknowledges the inability to reach a conclusion. For these values of θ, if avoiding an incorrect decision is the primary goal, then using the IBM is preferable. However, if making a correct decision with as little information as possible is the goal, then either the strict or generous classical approach performs better, with the correct choice depending on the underlying truth, which is not known *a priori*.

6 Summary

A variety of statistical approaches are available for updating assessments of the reliability based on test results. This paper has compared the classical sampling theory approach, the precise Bayesian approach, and the imprecise beta model, which is a subset of imprecise probabilities and is equivalent to the robust Bayesian approach when considering conjugate beta distributions for a Bernoulli testing problem. We have focused on the quality of the decisions reached as more information is collected. Of

course, the decision quality also depends upon the priors chosen, as the results show. We use those results to illustrate the strengths and weaknesses of each approach.

The example considered in this paper was very amenable to the use of conjugate prior distributions for the precise Bayesian and robust Bayesian (imprecise probability) approaches. The computational expense of these methods increases greatly when conjugate distributions cannot be used, especially for the robust Bayesian approaches. This adds another layer of trade-offs that was not considered in this study.

Given an existing component with a failure probability that should not be exceeded, accepting the new component when its failure probability is truly lower is a challenge because the failure probability is close to zero and there is a significant probability of test results that lead to a wrong decision. This indicates that the designer should use generous policies for making the acceptance decision. However, such policies increase the probability of wrongly accepting the new product. The use of the IBM allows the designer to avoid making incorrect decisions more often than the other approaches, but it can also significantly delay correct decisions. The IBM better reflects the trade-offs between the quantity and quality of available information, a distinction that can be important when communicating about uncertainty and managing information.

These conclusions have been illustrated using a specific example decision. More generally, a designer can use the results presented here to get insight into the trade-offs that exist in a specific domain. Each method has particular advantages and disadvantages, and the appropriate choice depends not only on the unknown underlying truth and available information but also on the designer's preferences for the quality and speed of decisions.

Acknowledgments

References

Aughenbaugh J.M., Herrmann J.W., 2007. Updating uncertainty assessments: A comparison of statistical approaches. In *2007 ASME International Design Engineering Technical Conferences*. Las Vegas, NV.

Berger J.O., 1984. The robust bayesian viewpoint. In J.B. Kadane (ed.), *Robustness of Bayesian analysis*, 63–124. North-Holland, New York.

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

Berger J.O., 1993. An overview of robust bayesian analysis. Technical Report 93-53c, Purdue University.

Brown L.D., Cai T.T., DasGupta A., 2001. Interval estimation for a binomial proportion. *Statistical Science*, 16(2), 101–133.

Coolen F.P.A., 1994. On bernoulli experiments with imprecise prior probabilities. *The Statistician*, 43(1), 155–167.

Coolen F.P.A., 1998. Low structure imprecise predictive inference for bayes' problem. *Statistics and Probability Letters*, 36, 349–357.

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

Coolen-Schrijner P., Coolen F.P.A., 2007. Non-parametric predictive comparison of success-failure data in reliability. *Proc. IMechE, Part 0: J. Risk and Reliability*, 221, 319–327.

Dempster A.P., 1967. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38, 325–339.

Dubois D., Prade H.M., 1988. *Possibility theory: an approach to computerized processing of uncertainty*. Plenum Press, New York.

Groen F.J., Mosleh A., 2005. Foundations of probabilistic inference with uncertain evidence. *International Journal of Approximate Reasoning*, 39(1), 49–83.

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

Lindley D.V., Phillips L.D., 1976. Inference for a bernoulli process (a bayesian view). *American Statistician*, 30, 112–119.

Rekuc S.J., Aughenbaugh J.M., Bruns M., Paredis C.J.J., 2006. Eliminating design alternatives based on imprecise information. In *Society of Automotive Engineering World Congress*. Detroit, MI.

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

Utkin L.V., 2004a. Interval reliability of typical systems with partially known probabilities. *European Journal of Operational Research*, 153(3 SPEC ISS), 790–802.

Utkin L.V., 2004b. Reliability models of m-out-of-n systems under incomplete information. *Computers and Operations Research*, 31(10), 1681–1702.

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

Walley P., 1996. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 3–57.

Walley P., Gurrin L., Burton P., 1996. Analysis of clinical data using imprecise prior probabilities. *The Statistician*, 45(4), 457–485.

Weber M., 1987. Decision making with incomplete information. *European Journal of Operational Research*, 28(1), 44–57.

Weichselberger K., 2000. The theory of interval probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24, 149–170.