

DRAFT SUBMISSION

DETC2007-35158

Updating uncertainty assessments: A comparison of statistical approaches

Jason M. Aughenbaugh and Jeffrey W. Herrmann

jason@arlut.utexas.edu

jwh2@umd.edu

ABSTRACT

The performance of a product that is being designed is affected by variations in material, manufacturing process, use, and environmental variables. As a consequence of uncertainties in these factors, some items may fail. Failure is taken very generally, but we assume that it is a random event that occurs at most once in the lifetime of an item. The designer wants the probability of failure to be less than a given threshold. This paper considers three approaches for modeling the uncertainty in whether or not the failure probability meets this threshold: a classical approach, a precise Bayesian approach, and a robust Bayesian (or imprecise probability) approach. In some scenarios, the designer may have some initial beliefs about the failure probability. The designer also has the opportunity to obtain more information about product performance (e.g. from either experiments with actual items or runs of a simulation program that provide the actual performance of interest or an acceptable surrogate for actual performance). This paper considers different approaches for forming and updating the designer's beliefs about the failure probability. The goal is to gain insight into the relative strengths and weaknesses of the approaches. Examples are presented for illustrating the conclusions.

INTRODUCTION

Engineers make decisions based on their beliefs. These beliefs depend upon the information that the engineer has gathered and can change based on new information (e.g. from either experiments with actual items or runs of a simulation program that provide the actual performance of interest or an acceptable surrogate for actual performance). Statistical reasoning includes a variety of approaches for updating beliefs based on new information. Most engineers have little training on these approaches, and some approaches may be completely unfamiliar, despite their usefulness and relevance. The main purpose of this article is to describe how one can apply different statistical approaches to update beliefs about the expected performance of a product and to discuss the relative strengths and weaknesses of each approach.

To answer this question, this paper discusses a specific design problem under a number of different scenarios. For each scenario, this paper shows how different statistical approaches can be used to update beliefs based on the results of new information (in this case, from the results of additional testing).

We assume that the specific designer has framed the most relevant design problem correctly, and that the designer may want to obtain and use more information to update his beliefs about the product's performance. In part the validity of this analysis will depend upon the reader's willingness to accept multiple theories of probability and statistics. Those who are highly committed to a specific philosophical position may find the arguments here unconvincing. This paper is intended for those who are open to trying any of these approaches but are unsure of their practical consequences.

From a practical perspective, it can be argued that the following properties are desirable for procedure of testing and analyzing reliability (or more generally probability) data. First, it can be valuable to incorporate existing information that may be relevant into the analysis; otherwise, existing information is essentially wasted. Second, the procedure should allow for new information to be incorporated into the estimate. Third, for experimental design and planning, the procedure should help the engineers to determine if they need more information, and if so, how much.

The analysis presented here implies that designers should carefully consider how to update their beliefs. It will help designers understand (though only in a qualitative sense) the value of the information they gather. Updating beliefs using a poor approach can lead to poor decisions.

This paper begins by stating the problem and then describes and compares the statistical approaches that will be considered. The paper introduces three scenarios that correspond to situations in which the designer has no information, substantial information, and partial information. Each statistical approach is considered for each scenario. A summary concludes the paper.

PROBLEM STATEMENT

The performance of a product that is being designed is unpredictable and often involves random behavior. Multiple items (instances of the product being designed) will be manufactured, sold, and used. Ideally, these items are identical. In practice, however, differences exist due to variations in materials and the manufacturing processes used to create them. Moreover, different items are used in different environments and in different ways.

We assume that the possibility of failures, though disagreeable to the engineer who is designing the product (i.e. "the designer"), is unavoidable for the above reasons. Some items will fail, while others will not fail. Here, "failure" is taken very generally, but we assume it is a one-time event in the lifetime of an item. In practice, the failure could be that the item's performance did not meet certain requirements during a test or that it catastrophically stops working at some point during its lifetime.

We assume that the designer wants the percentage of items that fail to be below a given threshold. In practice, this requirement would be based on customer expectations and the firm's tolerance of risk.

We assume that each item's failure is a random Bernoulli process. Let θ be the failure probability, which is the parameter of interest. Let θ_{crit} be the critical threshold such that the designer wants $\theta \leq \theta_{crit}$.

Ideally, the designer would have enough data to make a precise assessment of θ , such as " $\theta = 0.01$." If so, the designer must simply compare this assessment to the threshold, and the analysis is complete. However, there are many practical reasons why the designer cannot make such a precise assessment. Initially, the designer may have no relevant information (e.g. in the case of a completely novel design), or the designer may have data about the performance of one or more similar products from previous tests or other sources. In these cases, the designer may be unwilling to make a precise assessment despite holding some initial beliefs about θ .

We assume that the designer has the opportunity to obtain additional data from testing in order to update his beliefs. The designer can perform n tests in which the performance of the product—success or failure—is observed. We assume that test performance is the actual performance of interest or is an acceptable surrogate for actual performance. Let m be the number of failures observed (and consequently $n - m$ successes are observed). For convenience, we define a variable $x_i = 1$ if trial i is a failure, and $x_i = 0$ otherwise. Then $m = \sum_{i=1}^n x_i$. We assume each test is an independent Bernoulli trial, so m is a binomial random variable. The following function is the probability distribution for the number of failures in n trials, where j ranges from 0 to n .

$$P\{m = j\} = \binom{n}{j} \theta^j (1 - \theta)^{n-j}$$

This paper considers two interrelated questions. First, given the set of results $X = \{x_i\}_{i=1}^n$, how should the designer update his beliefs about the failure probability? Second, is the failure probability acceptable? Answering these requires understanding the relative strengths and weaknesses of the approaches, which may depend upon the scenario that occurs. This paper compares several approaches under different scenarios.

STATISTICAL APPROACHES FOR UPDATING PARAMETER ESTIMATES

This paper compares several approaches for analyzing data and updating beliefs about the parameter θ in different scenarios. The approaches can be classified using three categories:

1. Classical sampling theory,
2. Precise Bayesian, and
3. Robust Bayesian.

These approaches are described in the following subsections.

Classical sampling theory approach

Most engineers are familiar with classical, sampling theory approaches to statistical analyses, as these are generally emphasized in introductory texts such as [1]. Standard in these approaches is the adoption of a frequentist interpretation of probabilities. Under a frequentist interpretation, a probability represents the limit of the ratio of times that one outcome occurs compared to the total number of outcomes in an endless series of identical trials. For example, for a fair coin the probability of heads is 0.5 because, as the number of flips tends to infinity, the fraction of flips with the result heads approaches 0.5.

One can estimate an unknown probability from a random sample. Because the sample is random, the resulting conclusions are random and may be incorrect. Thus, we use confidence intervals to understand the probability of an incorrect conclusion. For example, if we repeatedly used a process to generate a 95% confidence interval, 95% of the resulting confidence intervals would contain the true value.

This approach focuses entirely on the observed data, that is the $\{x_i\}_{i=1}^n$. An unbiased point estimate of θ is the relative frequency of failures to trials in the sample. Specifically, one can estimate θ as $\hat{\theta} = \sum_{i=1}^n x_i / n = m / n$. A commonly used approximate one-sided $100(1 - \alpha)\%$ confidence interval that gives an upper bound on θ is given by the following [1]:

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

In this problem we use the upper bound because the designer would like the failure probability to be below the threshold. If this interval includes θ_{crit} , then one cannot reliably conclude that $\theta \leq \theta_{crit}$ based on the test results. We note that other confidence intervals for estimating this parameter have been proposed. For a full discussion of them, see Brown *et al.* [2].

Precise Bayesian approach

The Bayesian approach [3, 4] directly attacks a problem often encountered in science and engineering: *given existing knowledge and new knowledge, how can the two best be combined into a single estimate?* The term “Bayesian approach” comes from the use of Bayes’s Theorem for updating beliefs. A *prior* distribution for θ captures the designer’s current information about the failure probability. If we define Y to be additional information about θ (and let y be a particular set or realization of information), then the *posterior* distribution $\pi(\theta | Y = y)$ is the distribution *after* receiving information y . Bayes’s Theorem presents a way to update the prior using information from the experiment in order to form the posterior.

Bayes's Theorem is an extension of the rules of conditional probability, so its validity is generally accepted [4]. Some people use the term "Bayesian" to describe anything that uses Bayes's Theorem. Others use the term to refer to any procedure that uses a subjective interpretation of probabilities (more precisely known as subjective Bayesianism or subjectivism). Under a subjective interpretation, a probability is an expression of belief based on an individual's willingness to bet [5-8]. Bayes's Theorem provides an objective way to update *any* initial probability with objective data. If the prior is a subjective prior, then the posterior is necessarily subjective as well. However, one can also use *objective Bayesian analyses* in which the prior distributions are defined without subjective input. In these cases, the analysis can be viewed as entirely objective. See [9] for further discussion and references.

One of the requirements of Bayesian analysis is a starting prior distribution. Many of the strongest arguments both for and against Bayesian analysis involve the prior. On the one hand, if a designer has existing information about θ , it is important to consider this information in the new estimate of θ . This can be true even if the available information is only subjective in nature; for example, experts are often paid handsomely for their opinions.

On the other hand, if the designer has no specific information about θ , the introduction of a particular prior may distort the posterior estimate. The objective selection of a prior distribution in the absence of relevant prior information is a topic of extensive research. The approaches proposed include the use of *non-informative priors* [10, 11], maximum-entropy priors [12], and data-dependent empirical Bayes approaches [13].

In general, determining this posterior distribution can be computationally burdensome, as the output is a distribution over all possible values of θ . However, some code for performing these computations is available. See, for instance, the BUGS project at <http://www.mrc-bsu.cam.ac.uk/bugs/>. To support analytical solutions, the form of the prior is often restricted to conjugate distributions with respect to the measurement model. A prior distribution is considered *conjugate* if the posterior distribution is of 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 as a beta distribution (Equation 1); using this prior and a binomial likelihood function, the posterior will also be a beta distribution.

$$f(\theta; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \quad (2)$$

When one starts with a prior distribution of the form $Beta(\alpha_0, \beta_0)$ and observes m failures out of n trials, the posterior distribution is a beta distribution with updated parameters

$Beta(\alpha_0 + m, \beta_0 + n - m)$. Consequently, the update can be done analytically by simple addition and subtraction, an enormous improvement in efficiency.

The problem considered here is an estimation problem: the designer wants to know the true value of the probability of failure and to compare this to the critical value. One approach is to consider a point estimate of the parameter that is based on the posterior distribution. Common point estimates are the mean and median of the posterior and the maximum likelihood estimate (i.e., the value of θ that maximizes $\pi(\theta | Y = y)$).

Alternatively, one can express probabilities about the parameter. This is a key distinction from the classical approach. For example, the designer can calculate the probability that $\theta \leq \theta_{crit}$, as given in Equation (3).

$$P[\theta \leq \theta_{crit}] = \int_0^{\theta_{crit}} \pi(\theta | y) \cdot d\theta \quad (3)$$

If this probability is large enough, the designer can be comfortable that the failure probability is acceptably small.

Alternatively, one can use intervals to estimate the parameter. Using the posterior, any *credible interval* (or more generally, any *credible set*) can be found such that the probability that the true value is an element of the set is equal to a desired value. It is natural to calculate the $100(1 - \alpha)$ percentile of the posterior distribution, denoted $\theta_{1-\alpha}$ and form the $100(1 - \alpha)\%$ credible set $\{\theta | 0 \leq \theta \leq \theta_{(1-\alpha)}\}$. The decision criteria then becomes determining whether this interval contains θ_{crit} . If it does, then the probability that $\theta \leq \theta_{crit}$ is less than $100(1 - \alpha)\%$ and the design is unacceptable at that level.

Robust Bayesian approach

The third approach that we will consider uses an imprecise beta model for updating beliefs. This approach is compatible with the robust Bayesian approach and the theory of imprecise probabilities.

The core idea of robust Bayesian methods is to perform a “what-if” analysis by changing the prior [3, 14-16]. The analyst considers several reasonable prior distributions and performs the update on each to get a set of posterior distributions. If there is no significant change in the conclusion across this set of posteriors, then the conclusion is *robust* to the selection of the prior.

The theory of imprecise probabilities, formalized by Walley [17], has previously been considered in design decisions and set-based design [18, 19], but not in the context of updating parameter estimates in design settings. The theory begins with avoiding a sure loss (also known as a Dutch Book), the same fundamental notion of rationality as de Finetti [5, 20]. 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 for prior and posterior beliefs to be sets of density functions. Although the motivation is different from the robust Bayesian perspective, the computational approach is quite similar as it involves using Bayes's Theorem to find a set of posterior distributions from a set of prior distributions. For the problem considered in this paper, both approaches can be implemented using the imprecise beta model.

As explained previously, the beta distribution represents a convenient prior for the example problem in this paper under the precise Bayesian approach. For the robust updating approach, it is convenient to use the imprecise beta model, described by Walley [17, 21]. To introduce the imprecise beta model, one must first re-parameterize the beta, such that the density of $beta(s, t)$ is as given in Equation (4).

$$\pi_{s,t}(\theta) \propto \theta^{st-1}(1-\theta)^{s(1-t)-1} \quad (4)$$

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 the posterior parameters after n trials with m failures are given by $s_n = s_0 + n$ 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 and will be discussed with the different information scenarios. Following Walley [17], the learning parameters can be imprecise as well. That is, the priors are the set of beta distributions with $\alpha_0 = s_0 t_0$ and $\beta_0 = s_0 (1-t_0)$, with $\underline{t}_0 \leq t_0 \leq \bar{t}_0$ and $\underline{s}_0 \leq s_0 \leq \bar{s}_0$. The update equations can be expressed as follows (the min and max are taken over the range $\underline{s}_0 \leq s_0 \leq \bar{s}_0$):

$$\underline{t}_n = \min \{ (s_0 \underline{t}_0 + m) / (s_0 + n) \}$$

$$\bar{t}_n = \max \{ (s_0 \bar{t}_0 + m) / (s_0 + n) \}$$

$$\underline{s}_n = \underline{s}_0 + n$$

$$\bar{s}_n = \bar{s}_0 + n$$

The posteriors are the set of beta distributions with $\alpha_n = s_n t_n$ and $\beta_n = s_n (1-t_n)$, with $\underline{t}_n \leq t_n \leq \bar{t}_n$ and $\underline{s}_n \leq s_n \leq \bar{s}_n$.

The use of the imprecise beta model obviously affects the decision process in several ways, but most revealing is how it affects point estimation. Using a precise prior input, one gets a precise posterior output and a single point estimate. In the imprecise model, there are multiple posterior distributions, and consequently a range of point estimates. In particular, the interval $[\underline{t}_n, \bar{t}_n]$ is an interval estimate of θ . One approach for deciding whether or not $\theta \leq \theta_{crit}$ is to compare this interval to θ_{crit} . If $\bar{t}_n \leq \theta_{crit}$, then the designer knows that a decision that $\theta \leq \theta_{crit}$ which is based on the mean estimate of θ is robust across the range of prior assumptions. However, as with the precise case, a designer may wish to calculate the probability that $\theta \leq \theta_{crit}$ and compare this with some critical value. One can find upper and lower probabilities for this based on the set of posterior distributions. Although this varies across the set of posterior distributions, the minimal and maximal values occur at the extreme points of $[\underline{t}_n, \bar{t}_n] \times [\underline{s}_n, \bar{s}_n]$. Finally, one can determine a set of credible intervals from these posterior distributions.

DISCUSSION OF APPROACHES

The debate between Bayesians and non-Bayesians has been long, contentious, and without universal resolution [22-24]. However, there are certain properties of Bayesian analysis that are often attractive in practice. First, Bayesian analyses obey the *likelihood principle*, which basically states that all relevant information is contained in the likelihood function, and that two likelihoods contains the same information if they are proportional to each other (for example, see [3]). This implies that inferences should be made based on the data that was actually observed, not based on hypothetical data that may have been observed but was not. Lindley and Phillips [25] give the following example of the likelihood principle as it applies to Bayesian and sampling theory approaches, showing that sampling theory approaches depend on the stopping rule used in the experiment. Consider experiment A in which n trials are performed, and the number of failures m are counted. In experiment B, trials are performed until m failures are observed, which happens to take n trials in this example. Even though the same result of m failures out of n trials was observed in each experiment, a sampling theory approach can lead to different conclusions in a hypothesis test. This is because experiment A involves a binomial distribution, and experiment B involves a negative binomial distribution. However, the likelihood functions differ only by a scalar constant, and consequently the Bayesian inference will be the same in experiments A and B.

Second, a Bayesian analysis can take advantage of existing information. Thus, in some cases the Bayesian approach can incorporate more information than a classical approach. Of course, the impact of prior beliefs decreases as more data is obtained. Third, Bayesian approaches enable analysts to make direct probability statements about hypotheses and model parameters, statements that cannot be made in classical approaches.

However, the classical and Bayesian approaches are compatible under certain scenarios. Hogg and Tanis [1] point out that if the Bayesian prior is a constant non-informative one and the analyst uses the mode of the posterior distribution as the point estimate, then this is exactly the same as the maximum likelihood estimate in the classical approach. Other priors lead to a weighted maximum likelihood estimate.

Selecting the prior distribution is an important part of the Bayesian approach. For example, it was noted previously that priors are often stated in terms of conjugate distributions for computational convenience, but a particular conjugate distribution may not represent the prior information perfectly. When there is little prior information available, additional challenges occur. Using a single precise prior in this case seems to conceal the lack of information.

There are at least two arguments as to why a designer should not consider a single prior distribution—one practical, and one philosophical [26]. The first argument assumes that there exists a unique prior that best captures the prior information. However, the process of eliciting and assessing an individual's beliefs is resource intensive, so it will often be impractical to fully characterize them due to constraints such as bounded rationality, time, and computational ability [17, 27-29]. Consequently, only a partial—and therefore imprecise—characterization of subjective probabilities is normally available. This is view held by advocates of the robust Bayesian or Bayesian sensitivity analysis approaches [3, 14-16].

The second argument comes from the perspective that true beliefs need not be precise. For example, using the betting context of the subjectivists, *why should a person be willing to commit to either side of a gamble in the absence of any information about the gamble?* Precise Bayesian theory requires an individual to be able to state a fair price for the gamble paying p dollars if event A occurs and zero dollars otherwise, even when the individual is given no information about event A . 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 the imprecise probabilities view of uncertainty, in particular Walley and co-authors [17, 21, 26, 30].

The use of upper and lower probabilities has a number of advantages in this situation [21]. First, this approach yields posterior measures of uncertainty (the probability that a hypothesis is true). The conclusions are based on the results, not the stopping rule used in the experiments that generate the results (i.e., the analysis obeys the likelihood principle). The approach reflects the amount of information on which the conclusions are drawn; one can distinguish between probabilities based on a large of amount of data and those based on ignorance. The approach is able to model prior ignorance in a very intuitive way, and it includes common Bayesian models for prior ignorance as special cases. The models have good properties when considered from a classical frequentist perspective. Additional arguments for an imprecise model are given in the context of a multinomial problem in [30].

INFORMATION SCENARIO DESCRIPTIONS

The feasibility and desirability of a given approach for updating one's beliefs about an uncertain quantity depend upon the initial conditions (the amount of data currently available and the beliefs that are currently held).

In some applications, the designer may have relevant information about a product before he plans specific testing. For example, if a new product is a derivative of a previous product, certain aspects of its performance may be expected to be similar. In this case, data from experiments with the predecessor product give a rough idea of the performance of the new product, and testing is pursued to confirm and refine these assumptions. In other applications, the new product may be so novel or employed in such a different environment that the designer believes that existing data have little value in predicting the performance of the new product. In this paper, three scenarios are considered:

1. No relevant prior information is available.
2. Partially relevant information or a small amount of relevant information is available.
3. Substantial, highly relevant data histories are available.

The placement of an analysis task into one of these categories is a subjective judgment of the designer, and his assessment certainly could be incorrect; in this case, it may be valuable for subsequent testing and analysis to indicate this error.

Scenario 1: no prior information

In this scenario, it is assumed that the designer has concluded that there is no available information that is relevant to the desired reliability analysis. Essentially, the designer needs to construct an estimate of the reliability from scratch. All inferences about the product's reliability will be made using only the data samples received from the planned experiments, specifically $X = \{x_i\}_{i=1}^n$.

Classical sampling approach

From one perspective, a classical approach is perfectly suited to a scenario in which no prior information exists. A classical approach deals precisely with analyzing the probability of the data given a particular truth, and therefore focuses entirely on the observed data, that is, $X = \{x_i\}_{i=1}^n$. Prior information plays no role in the analysis. Consequently, the classical approach is identical in all information scenarios.

Precise Bayesian approach

A key element of the precise Bayesian approach is the choice of a prior distribution for θ . In the case of no prior information, the use of a *non-informative* prior is generally advocated [10-12, 31-33].

Non-informative priors are designed such that they favor no value of θ over any others. In this example, a natural choice is the uniform $[0, 1]$ distribution. Conveniently, the uniform $[0, 1]$ distribution is a special case of the beta distribution, expressed by letting $\alpha_0 = 1$ and $\beta_0 = 1$. As discussed earlier, the beta distribution is a conjugate prior for a Binomial process, and, after observing n trials with m failures, the posterior distribution for θ will be a beta distribution with parameters $\alpha_n = 1 + m$ and $\beta_n = 1 + n - m$.

Robust Bayesian approach

The robust Bayesian approach begins with a set of priors. When there is a complete lack of prior information, then the appropriate starting point is a vacuous prior (so-called because it contains the least information). In the imprecise beta model, this means setting $\underline{t}_0 = 0$ and $\bar{t}_0 = 1$. Because we have no information, we place no constraints upon θ .

We must also choose a “learning parameter” s . Because it reflects how much “importance” to assign to the prior data and there is no prior data, one could select a small learning parameter. However, too small a parameter may cause the posterior to react too quickly to the data. In particular, $s_0 = 2$ has a number of good properties, as described by Walley [17]. As described early, one can also allow for a range of learning parameters. Note that $t_0 = 0.5$ and $s_0 = 2$ corresponds to the beta distribution with $\alpha_0 = 1$ and $\beta_0 = 1$. Thus, this set of priors includes the single uniform prior considered in the precise Bayesian approach with no information.

Comparing behavior of estimates

It is clear from the above discussion that the choice of approach will affect the conclusions that are drawn from the test results. The classical approach will give a confidence interval on the failure probability. The precise Bayesian approach will give a single distribution for θ . The imprecise beta model yields a set of distributions for θ .

To illustrate the differences, consider the following example: the designer wants θ to be less than $\theta_{crit} = 0.05$. The confidence interval is a 95% one-sided confidence interval. The Bayesian prior is $\alpha_0 = 1$ and $\beta_0 = 1$. The imprecise beta prior is the vacuous prior $[0, 1]$ with a learning parameter of $\underline{s}_0 = \bar{s}_0 = 2$.

If the designer conducts $n = 100$ trials, Table 1 describes the conclusions for some of the possible outcomes. As the number of observed failures increases, the confidence interval on θ increases, and the posterior beta distributions shift, which increases the 95% credible intervals for both the precise Bayesian approach and the imprecise beta model. In this example, the 95% credible intervals for the precise

Bayesian approach are significantly larger than the confidence intervals. The smallest credible intervals from the imprecise beta model are nearly the same size as the confidence intervals.

Table 1. Conclusions based on number of observed failures in 100 trials

Outcome	Classical approach	Precise Bayesian	Imprecise beta
m	95% confidence interval	95% credible interval	Range of 95% credible intervals
0	[0, 0.000]	[0, 0.029]	[0, 0] to [0, 0.046]
1	[0, 0.026]	[0, 0.046]	[0, 0.029] to [0, 0.061]
2	[0, 0.043]	[0, 0.061]	[0, 0.046] to [0, 0.075]
3	[0, 0.058]	[0, 0.075]	[0, 0.061] to [0, 0.088]
4	[0, 0.072]	[0, 0.088]	[0, 0.075] to [0, 0.101]
5	[0, 0.086]	[0, 0.101]	[0, 0.088] to [0, 0.114]
6	[0, 0.099]	[0, 0.114]	[0, 0.101] to [0, 0.126]
7	[0, 0.112]	[0, 0.126]	[0, 0.114] to [0, 0.138]
8	[0, 0.125]	[0, 0.138]	[0, 0.126] to [0, 0.150]
9	[0, 0.137]	[0, 0.150]	[0, 0.138] to [0, 0.162]
10	[0, 0.149]	[0, 0.162]	[0, 0.150] to [0, 0.174]

Alternatively, consider how the conclusions would change based on the number of trials, for a fixed relative frequency of observed failures to trials. For instance, if $m/n = 0.04$, then the conclusions change based on the number of trials as shown in Tables 2 and 3. The confidence intervals decrease as the number of trials increases because the large number of trials reduces the sample variance. Similarly, the credible intervals decrease because the variance of the posterior distribution decreases. This also increases the probability of $\theta \leq \theta_{crit}$. The probability from the precise Bayesian approach is in the middle of the range of upper and lower probabilities that result from the imprecise beta model. Note also that this range significantly decreases as the number of trials increases. The additional information reduces the imprecision in the probability.

Table 2. Conclusions based on number of trials (relative number of observed failures is 4%).

Number of trials	Classical approach	Precise Bayesian	Imprecise beta
n	95% confidence interval	95% credible interval	Range of 95% credible intervals
25	[0, 0.104]	[0, 0.170]	[0, 0.109] to [0, 0.223]
50	[0, 0.086]	[0, 0.118]	[0, 0.090] to [0, 0.145]
100	[0, 0.072]	[0, 0.088]	[0, 0.075] to [0, 0.101]
200	[0, 0.063]	[0, 0.071]	[0, 0.064] to [0, 0.077]
500	[0, 0.054]	[0, 0.057]	[0, 0.055] to [0, 0.060]
1000	[0, 0.050]	[0, 0.052]	[0, 0.051] to [0, 0.053]
2000	[0, 0.047]	[0, 0.048]	[0, 0.047] to [0, 0.048]

Table 3. Conclusions based on number of trials (relative number of observed failures is 4%).

Number of trials	Precise Bayesian	Imprecise beta
n	$P\{\theta \leq \theta_{crit}\}$	$[P\{\theta \leq \theta_{crit}\}, \bar{P}\{\theta \leq \theta_{crit}\}]$
25	0.376	[0.139, 0.736]
50	0.473	[0.251, 0.731]
100	0.573	[0.393, 0.749]
200	0.679	[0.552, 0.791]
500	0.824	[0.762, 0.875]
1000	0.920	[0.895, 0.941]
2000	0.980	[0.974, 0.985]

Scenario 2: substantial prior information

In this scenario, there exists substantial information that the designer believes is relevant to the desired reliability analysis. This information may be based on previous tests, the performance of similar products, or other sources. The designer is considering testing as a way of verifying this information.

As before, the designer can perform n tests in which the performance of each item—success or failure—is observed. (We assume that test performance is the actual performance of interest or is an acceptable surrogate for actual performance.) Let m be the number of failures observed (and consequently $n-m$ successes are observed).

Classical approach

The classical approach does not include any prior information. As mentioned before, it focuses entirely on the observed data. Thus, the approach does not change for this scenario.

Precise Bayesian approach

The key to the performance of this approach is the choice of a prior distribution for θ . The designer should choose a prior that represents his substantial relevant information. It will be convenient once again to use a beta distribution. Larger values of α_0 and β_0 yield a prior distribution with less variance.

After observing n trials with m failures, the posterior distribution for θ will be a beta distribution with parameters $\alpha_n = \alpha_0 + m$ and $\beta_n = \beta_0 + n - m$. From this posterior distribution the designer can determine the probability that $\theta \leq \theta_{crit}$.

To illustrate this, consider two different priors for the following scenario: the designer wants θ to be less than $\theta_{crit} = 0.05$. Suppose that $\theta = 0.04$. The first (good) prior is $\alpha_0 = 4$ and $\beta_0 = 96$. The second (bad) prior is $\alpha_0 = 6$ and $\beta_0 = 94$ (it is a bad prior because it gives a mean-point estimate for θ

that is too high). Suppose the number of observed failures $m = 0.04n$. Then Figure 1 shows how the probability that $\theta \leq \theta_{crit}$ increases as more trials are conducted. Additional cases are considered subsequently in the paper, and the priors are given in Tables 4 and 5.

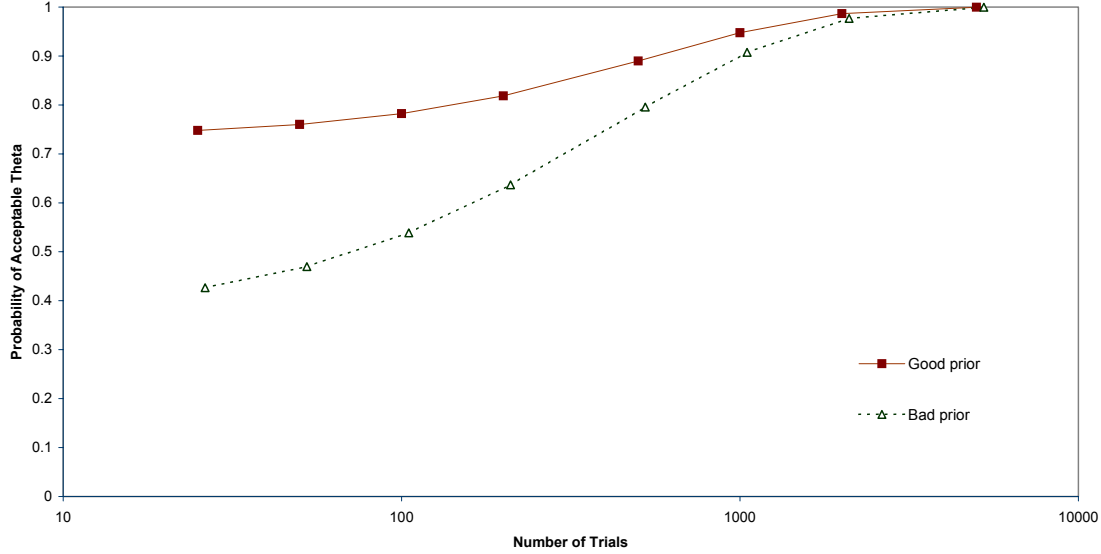


Figure 1. The probability that $\theta \leq \theta_{crit}$ as the number of trials increases for a good prior and a bad prior using the precise Bayesian approach.

Robust Bayesian approach

The robust Bayesian approach begins with a choice of priors. Given substantial prior information, the designer can choose \bar{t}_0 and \underline{t}_0 to be upper and lower bounds on θ .

The designer must also choose the learning parameter. If the designer is unsure of how much “importance” to assign to the prior data, then the designer could specify upper and lower learning parameters \bar{s}_0 and \underline{s}_0 . In the case of substantial information, we would expect \bar{t}_0 and \underline{t}_0 to be close to each other and \bar{s}_0 and \underline{s}_0 to be large. The set of posterior distributions can be determined as discussed earlier.

It is clear from the above discussion that the choice of approach will affect the conclusions that are drawn from the test results. The classical approach will give a confidence interval on the failure probability. The precise Bayesian approach will give a single distribution for θ . The imprecise beta model yields a set of distributions for θ .

An important consideration is whether the prior information is correct. Naturally, the quality of the information affects the conclusions. To illustrate the differences, consider again the previous example: the designer wants θ to be less than $\theta_{crit} = 0.05$. Given that the true $\theta = 0.04$, we will consider two different sets of prior information: one that is correct and a second that overestimates θ .

The first set has the following ranges of parameters: $[\underline{t}_0, \bar{t}_0] = [0.035, 0.055]$ and $[\underline{s}_0, \bar{s}_0] = [80, 120]$. The second set has the following ranges of parameters: $[\underline{t}_0, \bar{t}_0] = [0.055, 0.065]$ and $[\underline{s}_0, \bar{s}_0] = [80, 120]$.

Each set of posterior distributions (the one based on good priors and the one based on incorrect priors) gives a range of values for the probability that θ is acceptable. Figure 2 shows these ranges for each set as the number of trials increases. In all cases, the number of observed failures is the expected value; that is, $m = 0.04 n$. As the number of trials increases, the posteriors based on the good set of priors converge, and their variance decreases, significantly increasing the probability that θ is acceptable. The posteriors based on the bad set of priors move more slowly, which keeps the probability that θ is acceptable low. Interestingly, if we consider the results from Table 3, we see that the posteriors based on the vacuous priors lead to higher probabilities that θ is acceptable than the posteriors based on the bad set of priors. In one case, the new information needs to overcome a lack of knowledge; in the other, the new information needs to overcome a large amount of incorrect knowledge. This also illustrates that the precise Bayesian approach does not capture the amount of information on which the estimate is based.

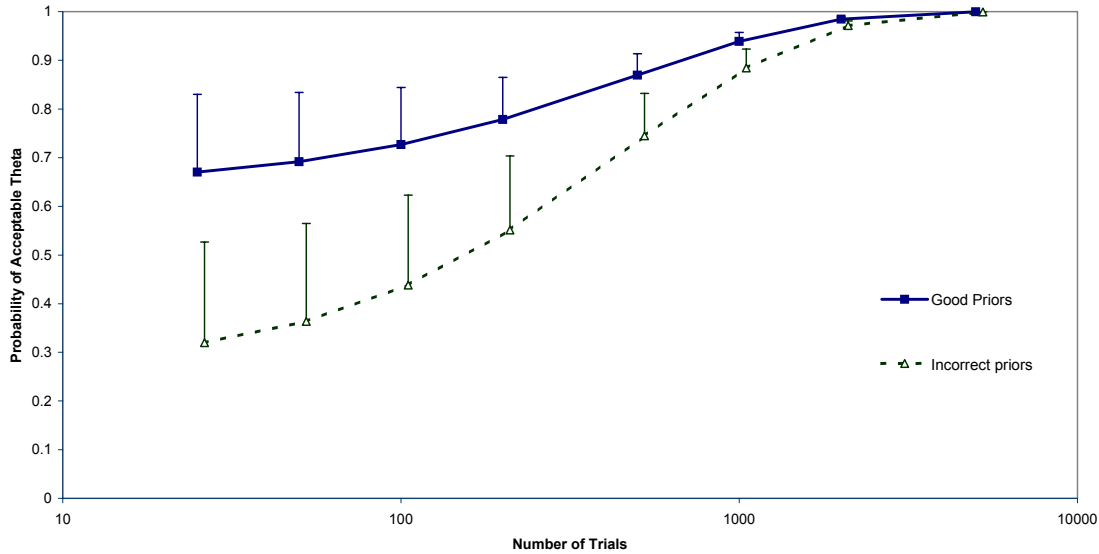


Figure 2. Range of probability of acceptable θ for different sets of priors in the robust Bayesian approach.

Scenario 3: Partial prior information

In this scenario, there exists some information that the designer believes is relevant to the desired reliability analysis. This partial information may be based on some previous tests, the performance of several diverse products, or other sources. The designer is considering testing in order to augment this partial information.

Classical approach

The classical approach does not include any prior information. As mentioned before, it focuses entirely on the observed data. Thus, the approach does not change for this scenario.

Precise Bayesian approach

The key to the performance of this approach is the choice of a prior distribution for θ . The designer should choose a prior that represents his partial information. It will be convenient once again to use a beta distribution. It is natural to choose the parameters α_0 and β_0 such that the mean $\alpha_0 / (\alpha_0 + \beta_0)$ is close to where the designer believes that the true value of θ may be. In this scenario, the designer should choose small absolute values to reflect the inadequate amount of information. Choosing $\alpha_0 \geq 1$ ensures that the density function has a bell-shaped curve. Lindley and Phillips [25] discuss this process and give some relevant (though non-engineering) examples.

For illustrating these approaches, we choose the priors shown in Table 4 for the case when $\theta = 0.06$ and the priors shown in Table 5 for the case when $\theta = 0.04$. We chose these priors so that they have a mean either near the true θ (for the good prior case) or farther away from the true θ (for the bad case). However, compared to the priors chosen for the significant information scenario, these priors have smaller parameter values and larger variance. After observing m failures in n trials, the posterior distribution for θ will be a beta distribution with parameters $\alpha_n = \alpha_0 + m$ and $\beta_n = \beta_0 + n - m$. Because of the smaller initial parameter values, the posteriors react more quickly to new data, thus de-emphasizing the importance of the prior.

Robust Bayesian approach

As discussed in Scenario 2, this approach starts with a set of prior distributions for θ and creates a set of posterior distributions for θ . The chosen priors are summarized in Table 4 for the case when $\theta = 0.06$ and in Table 5 for the case when $\theta = 0.04$. Since these priors reflect less information than in Scenario 2, the range of probabilities is wider, and the values of s_0 are smaller, indicating a smaller “pseudo” sample size for the prior. In general, the range for t_0 will be larger since there is more uncertainty in the estimates when there is less information.

Summary of scenarios

The priors selected for the precise Bayesian and the robust Bayesian approaches are summarized in Table 4 for the case when $\theta = 0.06$ and in Table 5 for the case when $\theta = 0.04$. The classical approach does not use any prior information, so it is unnecessary to describe it. It is important to note that although some priors are labeled “good” or “bad”, not all good (or all bad) priors are created equal. The term

“good” merely refers to a prior with an estimated mean for θ that is relatively closer to the truth than for the “bad” prior. It is impossible to define an equivalent “good” for a substantial information case and a partial information case, because the quality of the prior is a combination of the accuracy and precision of the estimate, where accuracy refers to how on target an estimate is, and precision describes the variance (and in the robust Bayesian approach, also the width of the intervals for s_0 and t_0).

Table 4. Bayesian priors for $\theta = 0.06$ case

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

Table 5. Bayesian priors for $\theta = 0.04$ case

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

COMPARISONS ACROSS SCENARIOS

In this section we will compare the approaches in a variety of ways. Both the precise Bayesian approach and the imprecise beta model depend upon the set of priors chosen. We will consider both the quality of the prior distribution (is it close to θ or not) and the magnitude of the prior (does it represent no information, partial information, or substantial information). We will consider how the conclusions depend upon the quality and magnitude of the priors and the number of trials conducted.

After considering the case in which the number of observed failures equals the expected number, we will consider the randomness of the outcomes, which affects the designer’s decision to accept the design. Specifically, given policies for deciding when θ is acceptably small, how does the probability that the designer will make a correct decision depend upon the quality and magnitude of the priors and the number of trials conducted? First, we will look at decisions made based on the results of the classical approach. Then we will consider the decisions made based on the results of the precise Bayesian approach. After that, we will examine the decisions made based on the results of the imprecise beta model. Finally, we will compare the decisions made for all of the approaches in the no information scenario.

Probability of an acceptable design

The probability that $\theta \leq \theta_{crit}$ is a key outcome of the precise Bayesian approach. This probability depends upon the quality and magnitude of the prior as well as the number of trials conducted and the number of failures observed. For simplicity, we first assume that the number of observed failures equals the expected number; that is, $m = \theta n$, with $\theta = 0.04$. Figure 3 shows the change in this probability for five priors: the non-informative prior (from Scenario 1), a correct prior based on substantial information (from Scenario 2), an incorrect prior based on substantial information (from Scenario 2), a correct prior based on partial information (from Scenario 3), and an incorrect prior based on partial information (from Scenario 3). All of the priors are listed in Table 5.

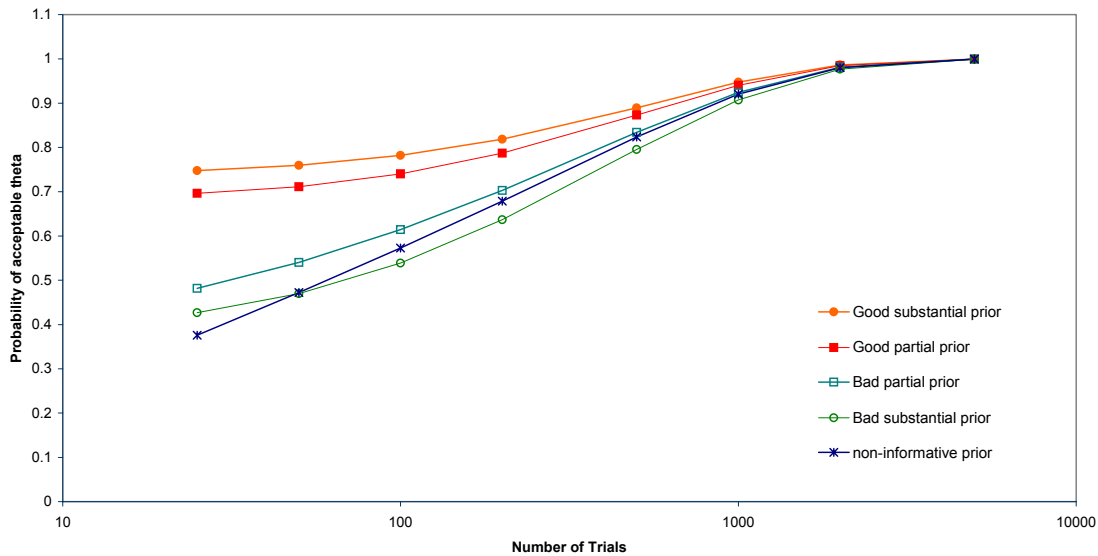


Figure 3. For the precise Bayesian approach, the posterior probability of an acceptable θ depends upon the quantity and quality of the prior information.

The plot reveals that, as expected, more good information leads to a better assessment. However, it also reveals that having a small amount of bad prior information is useful (compared to the non-informative prior), but a lot of bad information is detrimental. A “bad” estimate that is near the truth but has little weight essentially provides a good starting point for fast updates based on the observed data. A bad estimate that is given a lot of weight will not allow the posterior adapt quickly enough to the observed data.

Point Estimates for θ

The imprecise beta model can be used to determine upper and lower bounds on θ . These bounds depend upon the quality and magnitude of the prior as well as the number of trials conducted (assuming that the number of observed failures equals the expected number; that is, $m = \theta n$). When $\theta = 0.04$, Figure 4 shows the change in this probability for five priors: the non-informative prior (from Scenario 1), a correct prior based on substantial information (from Scenario 2), an incorrect prior based on substantial information (from Scenario 2), a correct prior based on partial information (from Scenario 3), and an incorrect prior based on partial information (from Scenario 3). (To improve the readability of the chart somewhat, the upper bounds are shown using the error bars, and the abscissa for each point for an informative prior is shifted to the right or left of the point for the non-informative prior.) All of the priors are located in Table 5.

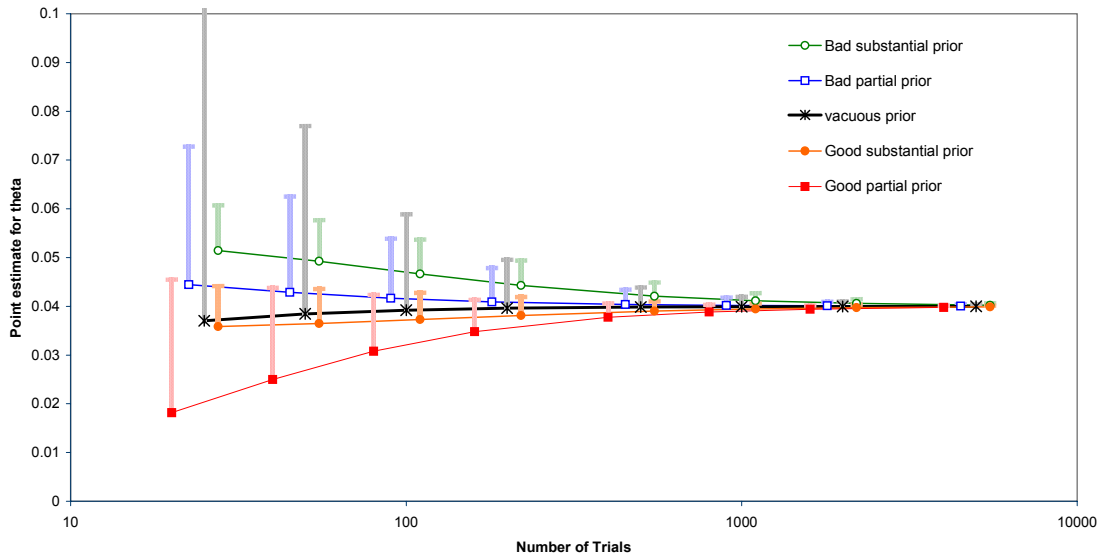


Figure 4. For the robust Bayesian approach, the range of point estimates for θ depends upon the quantity and quality of the prior information.

In all cases, the lower and upper bounds converge as the number of trials increases. Additionally, it should be noted that initially the bad substantial prior result is a subset of the bad partial prior result, which in turn is a subset of the vacuous prior result. As n increases, the vacuous prior and partial prior results adapt faster to the observed data and converge more quickly to the truth than the substantial prior case does.

Accepting the Design

The above discussion is based upon the expected results, in which the number of observed failures is the expected number for any number of trials (that is, $m = \theta n$). However, the observed results are random. Thus, if the designer uses a fixed policy for judging that θ is acceptably small and therefore accepting the design, the designer's decision is random. Therefore, for a given policy, we can look at the range of possible outcomes and calculate the probability of the resulting decision. For convenience, let "accept" mean that the designer decides that the design is acceptable ($\theta \leq \theta_{\text{crit}}$). Let "reject" mean that the designer decides that the design is not acceptable ($\theta > \theta_{\text{crit}}$). Let "no decision" mean that the designer does not decide.

Classical Approach Results

In the classical sampling approach, suppose that the designer will accept the design if the upper bound of the 95% one-sided confidence interval on θ is less than or equal to θ_{crit} . Otherwise, he rejects the design. Figure 5 shows the decision probabilities for each value of n in two cases: in the first, $\theta = 0.06$ (and is thus larger than θ_{crit}); in the second, $\theta = 0.04$ (and is thus smaller than θ_{crit}).

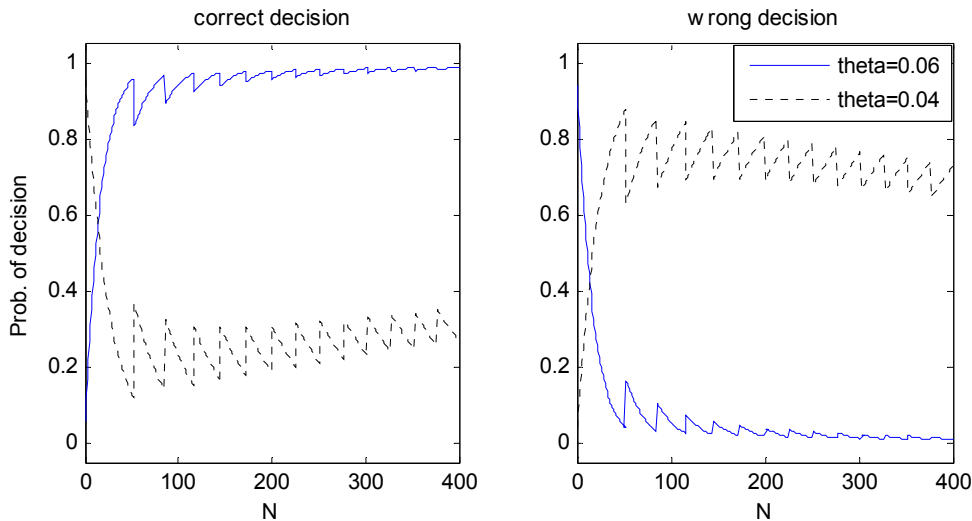


Figure 5. Decision probabilities for the classical approach. A "correct decision" rejects the design when $\theta = 0.06$ and accepts the design when $\theta = 0.04$. Otherwise, the decision is wrong.

There are several interesting behaviors apparent in Figure 5. First, note that the classical method performs better when $\theta > \theta_{\text{crit}}$. This occurs because the classical approach, using a 95% one-sided confidence interval, is (in a sense) biased towards concluding $\theta > \theta_{\text{crit}}$. In a hypothesis testing framework, this is equivalent to formulating $\theta > \theta_{\text{crit}}$ as the null hypothesis and then requiring a substantial amount of information to reject it in favor of the alternative hypothesis that $\theta \leq \theta_{\text{crit}}$.

A second point worth noting is the nature of the curves in Figure 5. They are not smooth, as one might expect, but are instead almost saw-tooth in nature. 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? The answer is due to the discrete nature of the binomial and the approximation made in Equation (1). The actual coverage probability of the approximate 95% interval fluctuates as n varies, an effect that is accentuated for small θ [2]. Similar patterns are seen for the other approaches and metrics, too, again due to the discrete nature of the underlying problem.

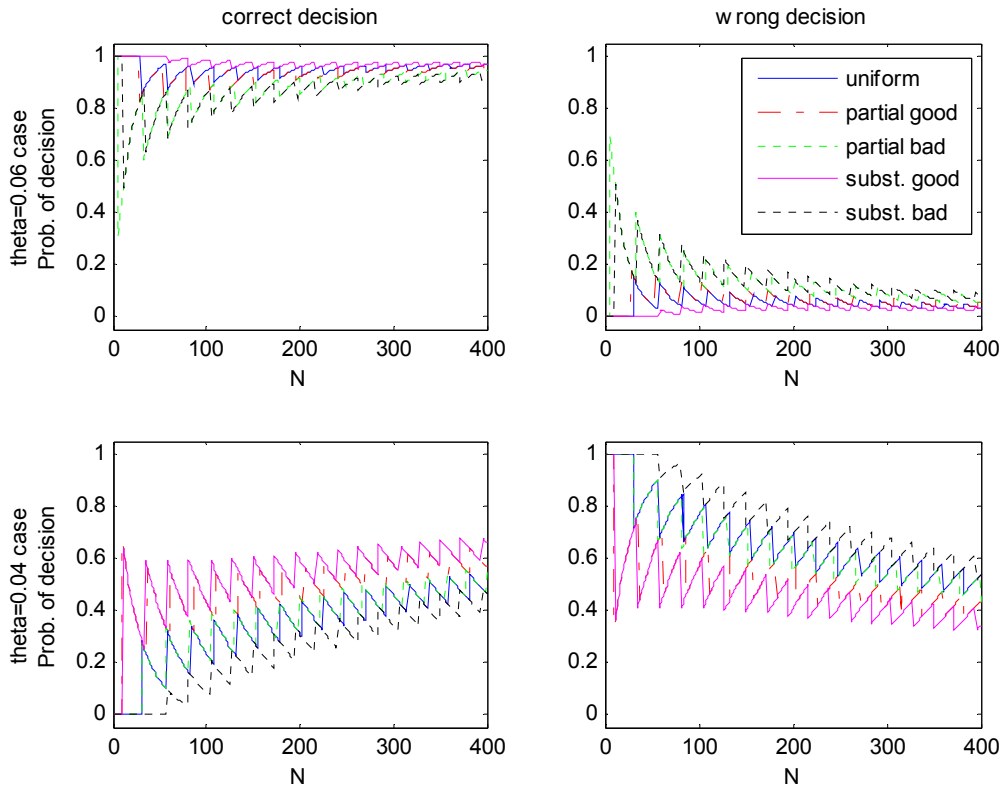


Figure 6. Decision probabilities for the precise Bayesian approach. A “correct decision” rejects the design when $\theta = 0.06$ and accepts the design when $\theta = 0.04$. Otherwise, the decision is wrong.

Precise Bayesian Results

If the designer uses the precise Bayesian approach, the designer will accept the design (because he concludes that $\theta \leq \theta_{\text{crit}}$) if the posterior probability that $\theta \leq \theta_{\text{crit}}$ is at least 0.8. Otherwise, he will reject the design (concluding that $\theta \geq \theta_{\text{crit}}$). Figure 6 shows the decision probabilities for each value of n in two cases: in the first, $\theta = 0.06$ (and is thus larger than θ_{crit}); in the second, $\theta = 0.04$ (and is thus smaller than θ_{crit}).

We begin by examining the upper left plot, which 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 or a bad prior. The partial, bad prior performs worse than the substantial, bad prior for very small n because it has a higher variance and is more easily influenced by data, and therefore requires less information to push the probabilities over the decision threshold on 0.80. In the long run, this higher responsiveness to the data leads to a faster adjustment away from the bad prior.

The lower left plot shows the probability of making a correct decision when $\theta = 0.04$. In general, the results are as expected, since 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 prior. As defined, the partial, bad prior was “bad” in that the mean of the prior was 0.076, which is much greater than the true 0.040. Despite the error in the mean, this prior is still in the vicinity of the truth, especially compared to the mean of the uniform, which is at 0.500.

Robust Bayesian Results

In the robust Bayesian (imprecise beta) approach, the designer will accept the design if (over the set of posteriors) the minimum probability that $\theta \leq \theta_{\text{crit}}$ is at least 0.8. The designer will reject the design if (over the set of posteriors) the maximum probability that $\theta \leq \theta_{\text{crit}}$ is less than 0.8. Otherwise, he makes no decision, essentially concluding that there is still not sufficient information to warrant a firm decision. Figure 7 shows the decision probabilities for each value of n in two cases: in the first, $\theta = 0.06$ (and is thus larger than θ_{crit}); in the second, $\theta = 0.04$ (and is thus smaller than θ_{crit}).

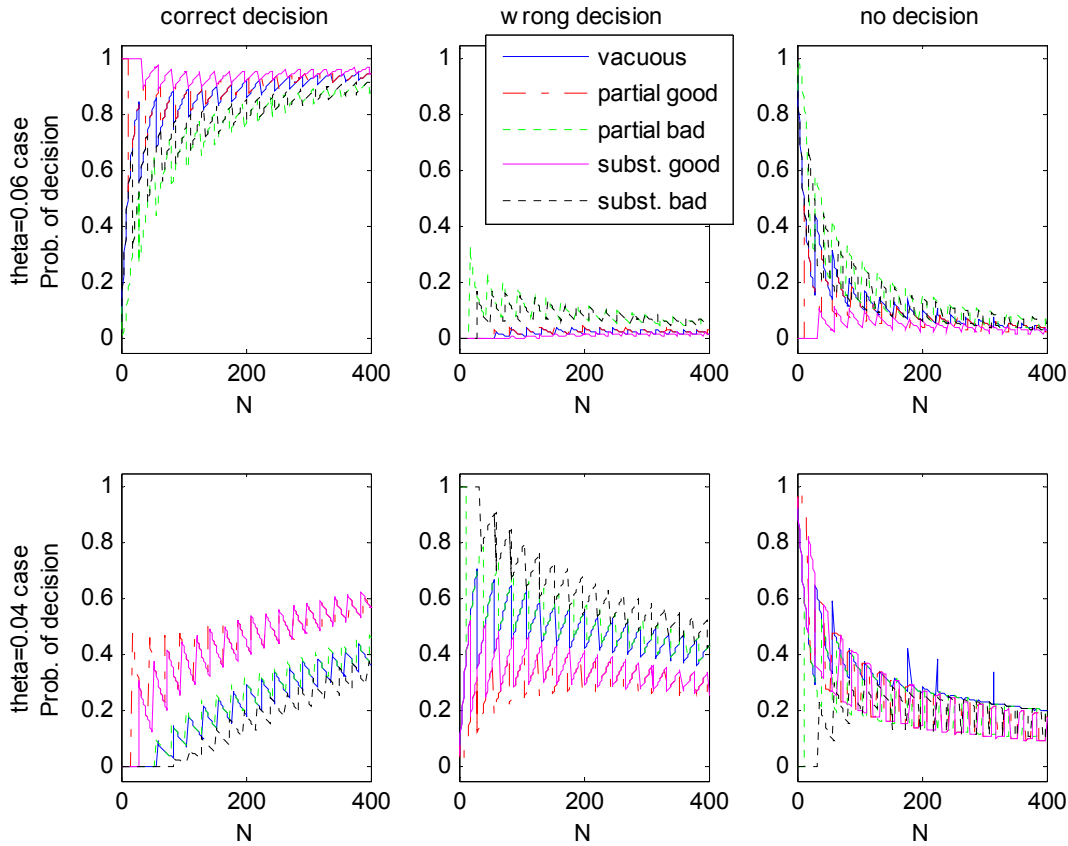


Figure 7. Decision probabilities for the robust Bayesian approach. A “correct decision” rejects the design when $\theta = 0.06$ and accepts the design when $\theta = 0.04$. Otherwise, the decision is wrong.

We begin by considering the bottom row of charts, which are for the $\theta = 0.04$ scenario. The probability of making a correct decision when $\theta = 0.04$ 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 probability of a correct decision is lowest when the designer starts with a bad prior that is based on substantial information. The probability is higher when the starting with a bad prior that is based on partial information, because there is less information to overcome.

The partial, bad prior and the vacuous prior lead to similar performance. Starting with no information and starting with a small amount of bad information are similar, in that it requires a lot of additional information in order to conclude that the design is acceptable. With the partial, bad prior, one needs enough additional information to counter the small amount of bad prior information that, while bad, is still in the vicinity of the truth. 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 0.04.

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 5, 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 actually lead to no decision.

Suppose $n = 20$ trials are performed and $m = 0$ failures are observed. This outcome has a probability of 0.44 when $\theta = 0.04$. For the partial, good prior, the posterior bounds on the probability that $\theta \leq \theta_{\text{crit}}$ are [0.82, 1.00], so the conclusion is that the design is acceptable. For the substantial, good prior, the posterior bounds are [0.76, 0.93] and no conclusion can be reached. Thus, it is evident that as expected, the partial prior is more sensitive to small sample sizes. Although this is advantageous in the $\theta = 0.04$ case, it is detrimental in the $\theta = 0.06$ case, as discussed next.

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. The probability of a correct decision is lowest when the designer starts with a bad prior that is based on substantial information or a bad prior that is based on partial information.

It is interesting to note 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 “bad” being defined ambiguously; in truth, not all bad information is equally bad, similar to the example for the good priors in the $\theta = 0.04$ scenario. For example, the partial, bad prior gives the probability that $\theta \leq \theta_{\text{crit}}$ as the interval [0.18, 0.62], 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 results will lead to accepting the design, which is the wrong decision. A more substantial, yet also incorrect, prior will be more robust to the variation in small samples.

A major difference between these results and those for the precise Bayesian approach is the option of making no decision. 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. Recall that the decision policy is to accept the design only when the minimum probability that $\theta \leq \theta_{\text{crit}}$ is at least 0.80. For the partial, bad prior scenario, the set of prior distributions actually leads to no conclusion, and hence the initially high probability of no conclusion. As evidence is gathered that $\theta > \theta_{\text{crit}}$, eventually the prior belief is countered by the evidence, and both the probability of reaching no conclusion and the probability of

reaching the wrong decision (i.e. accepted the design when $\theta > \theta_{\text{crit}}$) both decrease (and naturally the probability of the correct decision increases).

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, and with a small data set, the probability of getting “bad” data is very small.

In the $\theta = 0.04$ scenario, the probability of not reaching a conclusion is highest for the vacuous prior. This reflects that when there is very little information, acceptance of the design requires that this information be strongly in favor of concluding that the design is acceptable, and this has a very low probability of occurring. It may appear that there is an inconsistency compared to the $\theta = 0.06$ scenario in that the “bad” priors now have the lowest probabilities of reaching no decision. Additionally, the substantial, bad prior actually has a generally increasing probability of no decision for n between 0 and 200. This results from the complicated tradeoffs between prior information, observed data, and the “bias” in the decision rule.

Recall that a design is accepted only if the minimum probability that $\theta \leq \theta_{\text{crit}}$ is at least 0.8. Thus, when there is substantial prior information that the decision is not acceptable, the designer will tend to reject the design. As data is collected that starts to suggest that the design may be acceptable, the designer moves from concluding it is unacceptable to reaching no conclusion. As the data increasingly suggests that $\theta \leq \theta_{\text{crit}}$, the prior becomes less relevant, and eventually the designer will conclude that the design is acceptable.

More generally, the probability of reaching no conclusion reflects the overall weight of evidence available. 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 robust Bayesian approach to reflect both the overall quality and consistency of the combined prior and data evidence appears to be an advantage over the other methods, although perhaps only qualitatively.

Results across methods for 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). A 95% confidence interval is used with the classical approach as a benchmark for both sets. The first set uses “generous” policies for the Bayesian approaches. The designer accepts the design if the probability that

$\theta \leq \theta_{\text{crit}}$ is at least 0.80. (In the robust Bayesian, the minimum probability must be at least 0.8 to accept, and the maximum probability must be less than 0.8 to reject.) The results of these policies are shown in Figure 8. The second set uses “strict” policies that accept the design if the probability that $\theta \leq \theta_{\text{crit}}$ is at least 0.95. (In the robust Bayesian, the minimum probability must be at least 0.95 to accept, and the maximum probability must be less than 0.95 to reject.) The results of these policies are shown in Figure 9. Both Figure 8 and Figure 9 show how the decision probabilities change for two possible truth scenarios: in the first, $\theta = 0.06$ (and is thus larger than θ_{crit}); in the second, $\theta = 0.04$ (and is thus smaller than θ_{crit}).

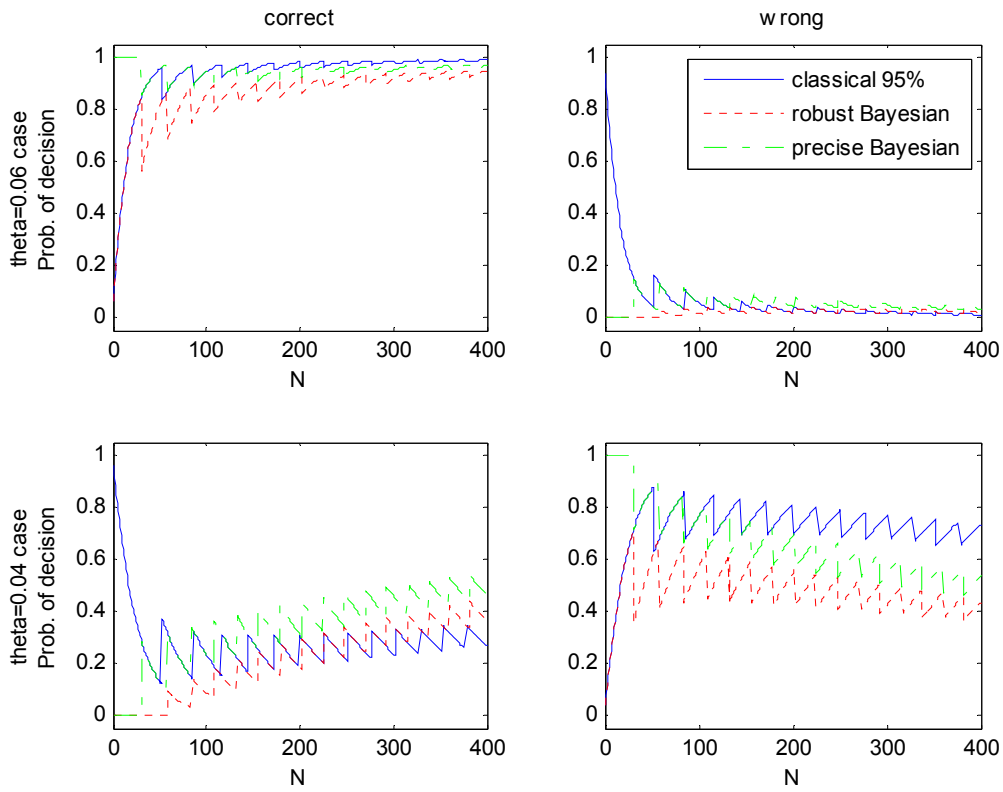


Figure 8. Decision probabilities for all three approaches in the no information scenario using the “generous” policies (0.80 probability test). A “correct decision” rejects the design when $\theta = 0.06$ and accepts the design when $\theta = 0.04$. Otherwise, the decision is wrong.

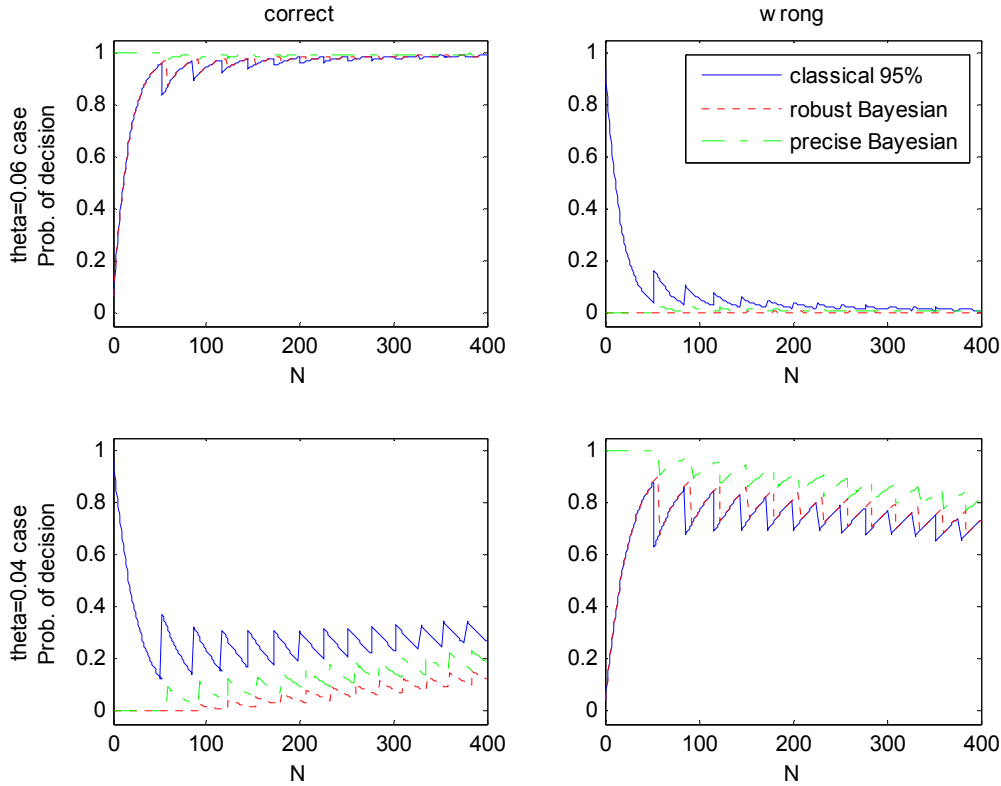


Figure 9. Decision probabilities for all three approaches in the no information scenario using the “strict” policies (0.95 probability level). A “correct decision” rejects the design when $\theta = 0.06$ and accepts the design when $\theta = 0.04$. Otherwise, the decision is wrong.

Under either set of policies, because it has the “no decision” option, the robust Bayesian approach leads to a lower probability of correct decision and a lower probability of a wrong decision. When $\theta = 0.06$, the generous policies lead to a lower probability of a correct decision (compared to the strict policies) because they are more likely to accept the design (which is wrong in this case). When $\theta = 0.04$, the generous policies lead to a higher probability of a correct decision (compared to the strict policies), because accepting the design is the correct decision.

The results for the $\theta = 0.06$ scenario in Figure 9 all indicate that performance is very good. Using the strict decision rule, the decision is biased towards concluding that the design is not acceptable. In the classical approach, the 95% confidence level denotes that, when using the approach many times, the true θ does not fall in the interval at most 5% of the time. However, if the confidence interval upper bound is greater than θ_{crit} and the true θ is less than θ_{crit} , then θ is in the confidence interval but the decision will be incorrect (it rejects an acceptable design).

For the Bayesian approaches, the strict policy requires the probability that $\theta \leq \theta_{crit}$ to be greater than 0.95 in order to accept the design. Even if the true θ is less than θ_{crit} , there is a significant

probability that, due to the randomness of the outcomes, the posterior probability that $\theta \leq \theta_{\text{crit}}$ will be less than 0.95, especially when the number of trials is small. This policy has a stricter acceptance criterion than the 0.80 policy of Figure 8, so the results are even more biased towards concluding that $\theta > \theta_{\text{crit}}$. When this is the true conclusion (e.g. when $\theta = 0.06$), then the performance of the method is very good. When this is the false conclusion (e.g. when $\theta = 0.04$), this method performs badly.

In all cases, the robust Bayesian method yields a lower probability of making a wrong decision compared to the precise Bayesian method. It also yields a lower probability of making a wrong decision than the classical approach, except for the $\theta = 0.04$ scenario using the strict decision policy (Figure 9), for which the results are similar. This avoidance of a wrong decision comes at the expense of not being able to make a correct decision either for small n ; instead of reaching a concrete decision based on little or conflicting information, the robust Bayesian approach acknowledges the inability to reach a conclusion.

Because the probability of a correct decision with $\theta = 0.06$ is generally high with either set of policies, the designer may seek to increase the probability of a correct decision with $\theta = 0.04$, which would lead to recommending the generous policies.

Finally, as part of the classical approach, one could use a different confidence interval for the decision policy. Increasing the confidence interval (e.g., using a 99% confidence interval) will make accepting the design more difficult. Decreasing the confidence interval (e.g., using a 80% confidence interval) will make accepting the design easier.

SUMMARY

A variety of statistical approaches are available for updating beliefs about the expected performance of a product based on test results. This paper has compared the classical approach, the precise Bayesian approach, and the robust Bayesian approach. This last approach uses the imprecise beta model, which makes it compatible with the imprecise probabilities approach.

The classical approach cannot take into account any information that might be available before testing. This feature has the advantage of being insensitive to the quality of the information. If one ignores the information, it doesn't matter if it is wrong.

The Bayesian approaches, on the other hand, can use high-quality prior information to help the designer make a correct decision more quickly. However, they are greatly affected by the quality and amount of existing information. In general, it appears that using a prior based on partial information is less "risky": if it is "bad," new information corrects it quickly and good decisions can be made. If it is good, it leads to decisions nearly as good as those obtained using a good prior based on significant

information. Interestingly, if it is not too “bad,” using a prior based on partial, incorrect information leads to better results than using the non-informative prior because new information corrects it quickly.

Using a prior based on significant information can lead to poor results if it is a “bad” prior. Overcoming the incorrect information will take a great amount of new information. Of course, if it is “good” then good decisions can be made quickly. Using partial information may be a good compromise, therefore. A prior (or set of priors) based on such information provides an effective head start compared to the non-informative prior (or vacuous priors) with minimal investment in information gathering, but, if information is incorrect, new information can overcome the priors more easily than it would if the prior was based on significant information.

The robust Bayesian approach allows the designer to use ranges to express his beliefs about the value of the parameter (via a range of means) and the importance of the prior data (via a range of learning parameters). Such caution about the prior information reduces the probability of making a bad decision. However, the probability of making a good decision also decreases because outcomes for which the available information is not conclusive explicitly result in no decision being made. Thus, the robust Bayesian method reflects both the quality and quantity of information available, including indirectly any conflict between the prior information and the new data.

Given a failure probability threshold that should not be exceeded, accepting a design when it is truly acceptable is a difficult problem because the threshold 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. The probability of incorrectly accepting a design whose failure probability is too high remains very low.

These conclusions have been illustrated using a specific example. More generally, a designer can use the methods presented here to get insight into the tradeoffs that exist in a specific domain.

ACKNOWLEDGEMENTS

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

REFERENCES

- [1] Hogg, R. V., and Tanis, E. A., 2001, *Probability and Statistical Inference*, Sixth Ed., Prentice Hall, Upper Saddle River, NJ.
- [2] Brown, L. D., Cai, T. T., and DasGupta, A., 2001, "Interval Estimation for a Binomial Proportion," *Statistical Science*, 16.2, pp. 101-133.
- [3] Berger, J. O., 1985, *Statistical Decision Theory and Bayesian Analysis*, 2nd Ed., Springer.

- [4] Box, G. E. P., and Tiao, G. C., 1973, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, Massachusetts.
- [5] de Finetti, B., 1974, *Theory of Probability Volume I: A Critical Introductory Treatment*, Wiley, New York.
- [6] Lindley, D. V., 1982, "Subjectivist View of Decision Making," *European Journal of Operational Research*, 9.3, pp. 213.
- [7] Winkler, R. L., 1996, "Uncertainty in Probabilistic Risk Assessment," *Reliability Engineering & System Safety*, 54.2-3, pp. 127-132.
- [8] Savage, L. J., 1972, *The Foundation of Statistics*, Dover Publications, New York.
- [9] Berger, J. O., 2000, "Bayesian Analysis: A Look at Today and Thoughts of Tomorrow," *Journal of the American Statistical Association*, 95.452, pp. 1269-1276.
- [10] Jeffreys, H., 1961, *Theory of Probability*, 3rd Ed., Oxford University Press, London.
- [11] Zellner, A., 1977, "Maximal Data Information Prior Distributions," *New Methods in the Applications of Bayesian Methods*, Aykac, A. and Brumat, C. eds., North-Holland, Amsterdam.
- [12] Fougere, P., eds, 1990, *Maximum Entropy and Bayesian Methods*, Klower Academic Publishers, Dordrecht.
- [13] Maritz, J. S., and Lewin, T., 1989, *Empirical Bayes Methods*, 2nd Ed., Chapman and Hall, London.
- [14] Berger, J. O., 1993, "An Overview of Robust Bayesian Analysis," Purdue University, West Lafayette, IN.
- [15] Insua, D. R., and Ruggeri, F., 2000, *Robust Bayesian Analysis*, Springer, New York.
- [16] Berger, J. O., 1984, "The Robust Bayesian Viewpoint," *Robustness of Bayesian Analysis*, Kadane, J. B. eds., North-Holland, New York, pp. 63-124.
- [17] Walley, P., 1991, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, New York.
- [18] Aughenbaugh, J. M., and Paredis, C. J. J., 2006, "The Value of Using Imprecise Probabilities in Engineering Design," *Journal of Mechanical Design*, 128.4, pp. 969-979.
- [19] Rekuc, S. J., Aughenbaugh, J. M., Bruns, M., and Paredis, C. J. J., 2006, "Eliminating Design Alternatives Based on Imprecise Information," *Society of Automotive Engineering World Congress 2006-01-0272*.
- [20] de Finetti, B., 1980, "Foresight. Its Logical Laws, Its Subjective Sources (Translated)," *Studies in Subjective Probability*, Kyburg, H. E. and Smokler, H. E. eds., E. Krieger Publishing Company.
- [21] Walley, P., Gurrin, L., and Burton, P., 1996, "Analysis of Clinical Data Using Imprecise Prior Probabilities," *The Statistician*, 45.4, pp. 457-485.
- [22] Rothenberg, T. J., 1974, "The Bayesian Approach and Alternatives in Econometrics-II," *Studies in Bayesian Econometrics and Statistics: In Honor of Leonard J. Savage*, Fienberg, S. E. and Zellner, A. eds., North-Holland Publishing Company and American Elsevier Publishing Company, New York, pp. 55-67.
- [23] Zellner, A., 1974, "The Bayesian Approach and Alternatives in Econometrics-I," *Studies in Bayesian Econometrics and Statistics: In Honor of Leonard J. Savage*, Fienberg, S. E. and Zellner, A. eds., North-Holland Publishing Company and American Elsevier Publishing Company, New York, pp. 39-54.
- [24] Clarotti, C. A., 1993, "Decision Making: Drawbacks of the Classical Statistics Approach," *Reliability and Decision Making*, Barlow, R. E., Clarotti, C. A., and Spizzichino, F. eds., Chapman & Hall, New York, pp. 1-12.
- [25] Lindley, D. V., and Phillips, L. D., 1976, "Inference for a Bernoulli Process (a Bayesian View)," *American Statistician*, 30.112-119.
- [26] Pericchi, L. R., and Walley, P., 1991, "Robust Bayesian Credible Intervals and Prior Ignorance," *International Statistical Review / Revue Internationale de Statistique*, 59.1, pp. 1-23.
- [27] Groen, F. J., and Mosleh, A., 2005, "Foundations of Probabilistic Inference with Uncertain Evidence," *International Journal of Approximate Reasoning*, 39.1, pp. 49-83.

- [28] Weber, M., 1987, "Decision Making with Incomplete Information," *European Journal of Operational Research*, 28.1, pp. 44.
- [29] Savage, L. J., 1971, "Elicitation of Personal Probabilities and Expectations," *Journal of the American Statistical Association*, 66, pp. 783-801.
- [30] Walley, P., 1996, "Inferences from Multinomial Data: Learning About a Bag of Marbles," *Journal of the Royal Statistical Society. Series B (Methodological)*, 58.1, pp. 3-57.
- [31] Berger, J. O., and Bernardo, J. M., 1992, "On the Development of the Reference Prior Method," *Bayesian Statistics 4*, Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. eds., Oxford University Press, Oxford.
- [32] Bernardo, J. M., 1979, "Reference Posterior Distributions for Bayesian Inference (with Discussion)," *Journal of the Royal Statistical Society*, 41, pp. 113-147.
- [33] Jaynes, E. T., 1983, *Papers on Probability, Statistics and Statistical Physics*, Rosenkrantz, R. eds., Reidel, Dordrecht.