SANDIA REPORT
SAND 2002-3953
Unlimited Release
Printed March 2003

An Application of Bayesian Methods For Combining Data From Different Test Modalities

David G. Robinson and I. Ann Chang

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550
Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of
Energy under Contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865)576-8401
Facsimile: (865)576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800)553-6847
Facsimile: (703)605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND 2002-3953
Unlimited Release
Printed March 2003

An Application of Bayesian Methods for Combining Data from Different Test Modalities

David G. Robinson and I. Ann Chang
Risk and Reliability Department
P.O. Box 5800, MS 0748
Sandia National Laboratories
Albuquerque, NM 87185-0748

drobin@sandia.gov
ichang@sandia.gov

Abstract

This report documents the research into the application of hierarchical Bayesian methods for characterizing the population failure rate (i.e. probability of defect) of an electronic component based on test data from a number of different test modalities. Classical statistical methods, those based on a frequency approach permit the combination of point estimates but stumble when characterizing the resulting confidence limits. Classical Bayesian methods permit the logical combination of test data, but are not fully efficient in incorporating all available information. In particular, classical Bayesian methods assume that the articles under test are not related in any manner even though the articles may be identical. Alternatively, hierarchical Bayesian methods permit the relationship between test articles to be explicitly included in the analysis. Data from four different test modalities are considered in the analysis. Comparisons are made between the current analysis approach (using traditional statistical methods), classical Bayesian methods and a hierarchical Bayesian approach.

Contents

Figures

Tables

An Application of Bayesian Methods for Combining Data from Different Test Modalities

Abstract

This report documents the research into the application of hierarchical Bayesian methods for characterizing the population failure rate (i.e. probability of defect) of an electronic component based on test data from a number of different test modalities. Classical statistical methods, those based on a frequency approach, permit the combination of point estimates but stumble when characterizing the resulting confidence limits. Classical Bayesian methods permit the logical combination of test data, but are not fully efficient in incorporating all available information. In particular, classical Bayesian methods assume that the articles under test are not related in any manner even though the articles may be identical. Alternatively, hierarchical Bayesian methods permit the relationship between test articles to be explicitly included in the analysis. Data from four different test modalities are considered in the analysis. Comparisons are made between the current analysis approach (using traditional statistical methods), classical Bayesian methods and a hierarchical Bayesian approach.

Background

System Description

For security reasons, the actual component to be characterized can only be described in abstract terms. The component consists of two identical subunits in hot standby redundancy. That is, both subunits are powered up and operational through the entire mission. A subunit consists of an electronic package utilizing monolithic microwave circuit technology, a number of cables, and an antenna. Failure of a subunit is assumed to occur when it does not provide a correct signal (go/no-go) at the selected range given the proper inputs. Component failure occurs when neither of the subunits provides a correct signal.

Test Modalities

Testing is accomplished at both the subunit as well as component level. In addition, there are situations were individual devices within the subunit are tested. While the techniques discussed in this report are applicable when including device test data, this data is not included in the analysis that follows. There are four major types of testing where data is available:

1. Tool Made Sample (TMS) – testing on the initial design production
2. E&D
    a. Environmental Testing (E-test) – subunit testing involving a series of non-destructive environmental tests, for example, temperature cycling performed on samples to assure stockpile reliability and end-use performance. Non-failing units are returned to the stockpile.

b. Destructive Testing (D-test) – subunit testing involving destructive environmental testing, e.g., vibration and shock combined with temperature cycling.

3. REST Lab Testing (Stockpile Lab Testing) – system level testing
4. REST Flight Testing (Stockpile Flight Testing) – system level testing; component and subunits may/may not be completely energized.

Of these tests, the most comprehensive and rigorous test regime is that accomplished for E&D. This testing puts each subunit through a wide range of possible input signals under a variety of simulated stress conditions. On the other extreme are the system tests accomplished as part of the Stockpile Flight Testing program. In this case, while the test is the most realistic, not all functions of the subunit will be interrogated during a particular flight test.

Table 1 provides a summary of the number of tests and the number of associated <u>subunit</u> failures available for the four different test modalities.

Mode	Tests	Failures
TMS	20	2
E & D	124	0
REST Lab (SLT)	36	0
REST Flight (SFT)	2	0

Table 1. Summary of Test Results

Traditional Confidence Interval Assessment

Suppose that *n* independent tests are conducted with the probability of failure *p*. The observed number of failures, *r*, occurring in these *n* tests has a binomial distribution:

$$f(X = r; p) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \tag{1}$$

with cumulative distribution:

$$F(x;p) = \sum_{x=r}^{n} \frac{n!}{r!(n-r)!} p^x (1-p)^{n-x} \tag{2}$$

Let γ = desired confidence with p_L and p_U being the associated lower and upper confidence limits. The confidence limits can be found by solving the following equations for p_L and p_U:

$$\sum_{x=0}^{r} \binom{n}{x} p_U^x (1-p)_U^{n-x} = (1-\gamma)/2 \tag{3a}$$

$$\sum_{x=r}^{n} \binom{n}{x} p_L^x (1-p)_L^{n-x} = (1-\gamma)/2 \qquad (3b)$$

Since x is a discrete random variable, an exact solution may or may not be available. Thus if *n* samples are tested an infinite number of times and compute confidence limits each time, then if $\gamma = 0.90$, these limits will contain the true value of *p* 90 times out of 100.

For the particular series of tests under consideration, it is desired to assess the upper 10 and lower 90 percent confidence limits, i.e. $\gamma = 0.80$. Table 2 provides a summary of the evaluation of equations 3a and 3b for the data in Table 1. The median values are determined from evaluation of either of the equations at $\gamma = 0.50$.

Inherent within the assumption of the different test modalities are independent is that there is no similarity or correlation between the articles being tested. Therefore, from a theoretical view, the test results can not be grouped together to make an overall 'population estimate'.

Mode	Tests	Failures	10% LCL	Median	90% UCL
TMS	20	2	0.05650	.1315	.245
E & D	124	0	0.001	0.006	0.0185
REST Lab (SLT)	36	0	0.003	0.0195	0.062
REST Flight (SFT)	2	0	0.0515	0.293	0.684

Table 2. Summary of Traditional Confidence Interval Analysis

Classical Bayesian Analysis

In the traditional technique (also known as the Frequentist approach) applied above, the results implicitly depend on the fixed, but unknown value of *p*. In a Bayesian analysis, *p* is treated as simply an unknown parameter that can be better estimated through the accumulation of information. As a result, the conditional dependence of observing *r* failures given the variable *p* is stated explicitly:

$$f(X=r \mid p) = \frac{n!}{r!(n-r)!}(1-p)^{n-r}(p)^s \qquad (4)$$

Let $g(p)$ be a probability distribution function characterizing the <u>initial</u> understanding that the variable *p* takes on a particular value. Bayes theorem states the information gained from actually observing the random process (the failure data) can be combined with the initial understanding to gain a better understanding of the variable *p*:

$$g(p \mid r) = \frac{g(p)f(X=r \mid p)}{\int_0^1 g(p)f(X=r \mid p)dp} = \frac{g(p)f(r \mid p)}{f(r)} \qquad (5)$$

Selection of a prior

The most controversial aspect of Bayesian analysis is the selection of the prior $g(p)$ since a poorly characterized prior can dramatically impact the distribution $g(p \mid r)$. One approach to addressing this problem is through the use of a prior distribution that contains little information, i.e. a *non-informative* prior.

In the reliability field, the typical non-informative prior is the uniform distribution:

$$g(p) = \begin{cases} 1, & 0 < p < 1 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

It will be convenient to recall that the uniform probability density function is a special case of the beta density function:

$$g(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} \quad 0 \leq p \leq 1 \tag{7}$$

where: $\Gamma(n) = \int_0^\infty x^{n-1}\exp(-x)dx$, and a, b are parameters of the distribution. In the case of the uniform distribution the parameters take on the values: a=1, b=1. The beta form for the prior is particularly appealing since the posterior is therefore also a beta distribution function:

$$g(p \mid s) = \frac{\Gamma(a+b+n)}{\Gamma(a+r)\Gamma(b+n-r)} p^{(a+r)-1}(1-p)^{(b+n-r)-1} \quad 0 \leq p \leq 1 \tag{8}$$

This traditional non-informative prior assumed on p in the form of a uniform beta distribution function was used as the first approach.

It is however rather unrealistic in that given the system being analyzed, both engineering judgment and historical records indicate that the underlying failure rate is very low. Therefore as an alternative, a prior distribution that captures this information will be investigated. In particular, p will be assumed to be a Weibull distributed random variable, with scale parameter λ and shape parameter β (e.g. Figure 1):

$$g(p \mid \lambda, \beta) = \lambda\beta p^{\beta-1}\exp[-\lambda p^{\beta}], \quad 0 \leq p \leq 1. \tag{9}$$

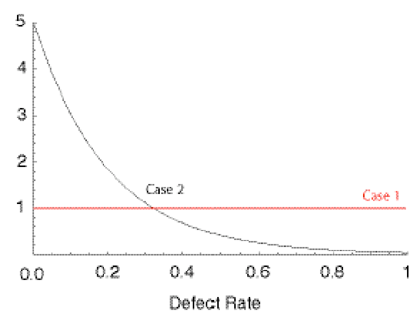


Figure 1. Weibull Prior

The Weibull distribution is chosen because it is often used to describe product life and it models either increasing or decreasing failure rates easily (see Figure 1). It permits a more realistic description of the historical information.

Results

For ease of analysis, Markov Chain Monte Carlo (MCMC) will be used to assess the confidence intervals for each of the prior distributions assumed above. The analysis for this study was accomplished using the WinBugs software package (Spiegelhalter, et al. 1996). All runs in this report were carried out with 1000 burn-in iterations and a single long-run of 10000 iterations.

Case A

In a classical Bayesian analysis, as with the traditional Frequentist approach, each of the tests is assumed to be independent from the other. Figure 2 depicts the probability density function of the probability of observing a particular defect rate for each of the tests. Table 3 presents a summary of the resulting confidence limit analysis.

It should be noted that since p is considered a random variable, the confidence intervals represent that probability that the true value of p is above or below a particular value; this is quite different than the definition of confidence intervals from a Frequentist perspective. From a theoretical viewpoint, the comparison between these two perspectives on confidence intervals is a bit like mixing apples and oranges, but for the purposes of this discussion the differences will be overlooked.

Finally, note that since the tests were assumed to be independent (as was the case in the Frequentist analysis), no information is exchanged between tests, and therefore no estimate of the population characteristics is available. This is similar to a ***fixed effects*** model common in statistical analyses.

Mode	Tests	Failures	Analysis Method	10% LCL	Median	90% UCL
TMS	20	2	Frequentist	0.0565	0.1315	0.245
			Classical Bayes	0.0543	0.1227	0.238
E & D	124	0	Frequentist	0.001	0.006	0.0185
			Classical Bayes	0.001	0.0058	0.01895
REST Lab (SLT)	36	0	Frequentist	0.003	0.0195	0.062
			Classical Bayes	0.0033	0.01898	0.0663
REST Flight (SFT)	2	0	Frequentist	0.0515	0.293	0.684
			Classical Bayes	0.0354	0.202	0.523

Table 3. Summary of Classical Bayes Confidence Interval Analysis: Uniform Prior

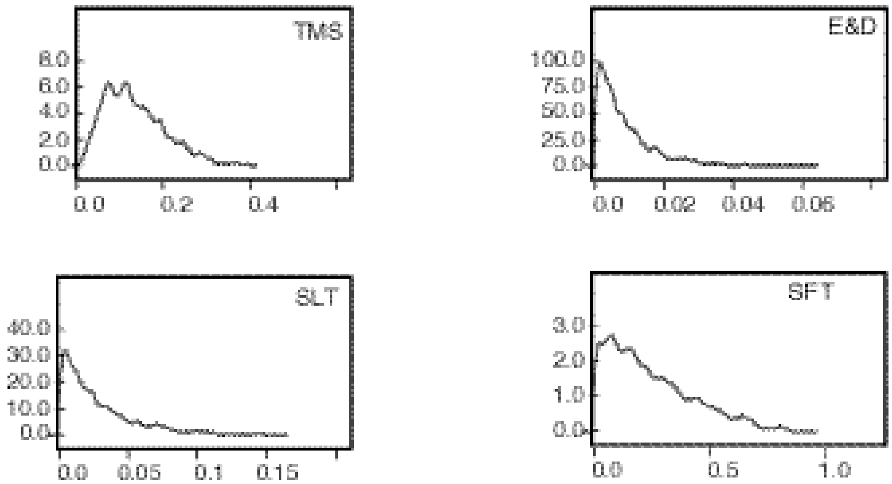


Figure 2. Posterior Density Functions of Failure Rates: Classical Bayes Analysis with Uniform Prior

Case B

As mentioned previously, the uniform prior, while a popular choice that avoids controversy, does not reflect the available knowledge of the system being analyzed. The particular system is a very reliable system with strict controls over both the manufacturing and testing process. A more logical representation of the available information would be a prior that emphasized the low likelihood of finding a defective subunit but didn't overpower the test data. A prior that fits this description is a Weibull density function with a shape parameter of 1.0 and a scale parameter of 5.0 truncated with a maximum value of 1 (depicted in Figure 1). As seen in Figure 1, these numbers suggest failure rates more realistically – most likely lower than those of the uniform distribution. The results under the assumption of this prior are provided in Table 4 and the associated probability density functions of the probability of observing a particular

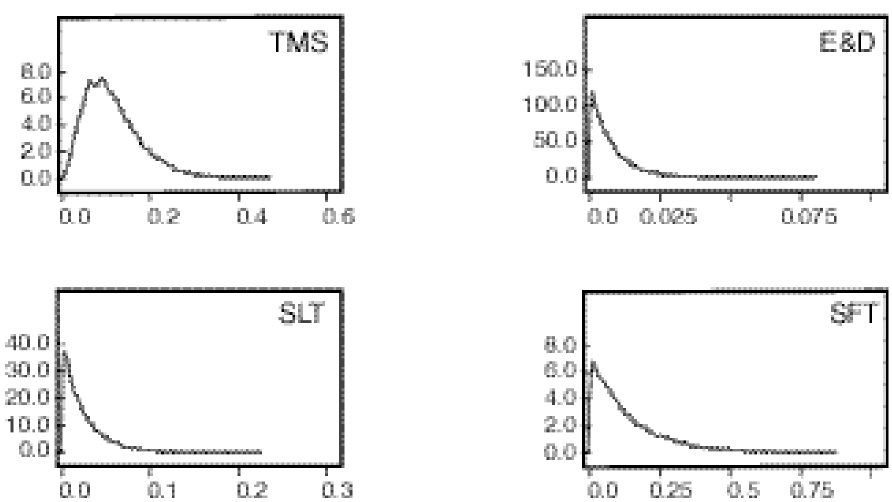


Figure 3. Posterior Density Functions of Defect Rates: Classical Bayes Analysis with Weibull Prior

defect rate for each of the tests are presented in Figure 3.

Mode	Tests	Failures	Analysis Method	10% LCL	Median	90% UCL
TMS	20	2	Frequentist	0.0565	0.1315	0.245
			Classical Bayes	0.0449	0.1037	0.199
E & D	124	0	Frequentist	0.001	0.006	0.0185
			Classical Bayes	0.0007	0.0052	0.0174
REST Lab (SLT)	36	0	Frequentist	0.003	0.0195	0.062
			Classical Bayes	0.0025	0.01657	0.0547
REST Flight (SFT)	2	0	Frequentist	0.0515	0.293	0.684
			Classical Bayes	0.0140	0.09217	0.2939

Table 4. Summary of Classical Bayes Confidence Interval Analysis: Weibull Prior

It is clear from both Table 4 and Figure 3 that the assumption that the underlying subunit defect rate can be characterized by the suggested Weibull density function has relatively minor impact on the results; when the available test data clearly dominates the final median and confidence interval estimates (the E&D test data). As would be expected, with the very limited number of stockpile flight tests (<u>and</u> no observed failures) the estimate is strongly influenced by the optimistic Weibull prior.

Hierarchical Bayesian Analysis

As mentioned earlier, the major suspicion of Bayesian methods lies in the construction of the prior density function to characterize the initial knowledge. Hierarchical Bayesian (HB) methods provide some relief from this concern in that they reduce the sensitivity of the results to the specifics of the prior density function. Classical Bayesian methods require the specification of unique parameters for the density function(s) associated with the prior information and it is the uniqueness of the parameters that lead to a sensitivity in the results. Alternatively, hierarchical methods permit these parameters to be random variables, which in turn relax the specification of the prior and results in the data having an even stronger influence on the posterior characterization of the subunit. For a thorough introduction to hierarchical Bayesian methods see the review by Robinson (2001). Conceptually what is happening is that the parameters of the prior are assumed to take on a range of values with a specified probability of a particular value being observed as characterized by the probability density function.

The final very important point to be made is that it will be assumed that the items being tested have similar underlying defect rates. The similarity of those defect rates is captured through the use of common underlying characteristics for the distribution of the parameters for the prior. The easiest way to view the differences between HB and classical Bayes is via a directed graph.

Graphical Representation of Hierarchical Models

With the concern regarding selection of the initial prior necessary in classical Bayesian methods, hierarchical models have become increasingly popular and have a potential application for solving very complex problems. Hierarchical methods permit the uncertainty associated with selection of the parameters for the prior to be explicitly recognized. The result is that the final estimates are less sensitive to the particular prior chosen.

The structure of hierarchical models lends itself easily to a graphical depiction of the relationships between various model constructs using directed graphs. These graphical cartoons are useful for organizing information and also for constructing the posterior distribution functions discussed above.

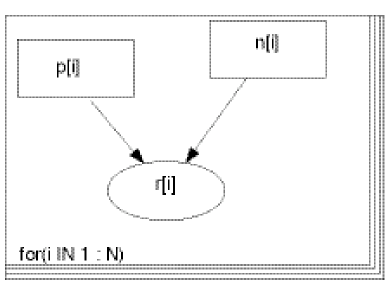
Directed graphs are essentially a set of nodes connected with a set of directed edges or arrows which depict the informational dependencies between the nodes. Those nodes that feed information to subsequent nodes are considered *parent* nodes. Each node is considered independent of all other nodes except parent nodes and those nodes for which that node is a parent.

There are three types of nodes:

1. Constant nodes have no parents and represent fixed quantities in the analysis. These nodes are represented by rectangles.
2. Stochastic nodes may have parents or children and represent, typically, unobserved random variables. They are commonly represented by circles/ellipses on the graph.
3. Functional or deterministic nodes represent functions of other variables in the graph.

In addition there are constructs referred to as plates, which indicate iteration over an index variable, typically to represent data input.

Figure 4 depicts the directed graphs associated with the classical Bayesian analysis performed in the previous section as well as the graph associated with the Frequentist approach discussed earlier. In both cases the number of defects, *r*, observed is assumed to be a random variable (enclosing ellipse) and the number of test samples is assumed known (enclosing rectangle). However, note that in one case the defect rate is assumed to be known (enclosing rectangle) and in the other it is assumed to be a random variable. In



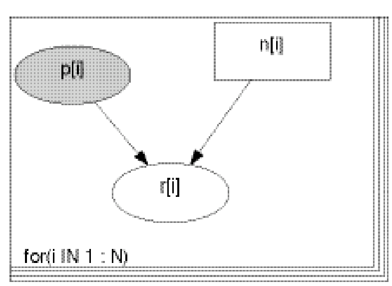


Figure 4. Directed Graphs for Frequentist and Classical Bayesian Approaches

the second situation, the number of defects in a sample, r, is assumed to be a binomially distributed random variable with parameters p, the defect probability, and n, the number of samples. The probability of finding a defect is also assumed to be a random variable as indicated by the surrounding ellipses with a beta distribution characterized with parameters α and β. In both cases, there are data available from $i=1,\ldots, N$ different tests as indicated by notation on the plates.

Contrast this with the structure depicted in Figure 5. In this case, the number of failures is still a binomially distributed random variable with parameters p and n. The defect rate, p, is again assumed to be a random variable; however, now the parameters α and β are assumed to be unknown and are further characterized with specific probability distributions.

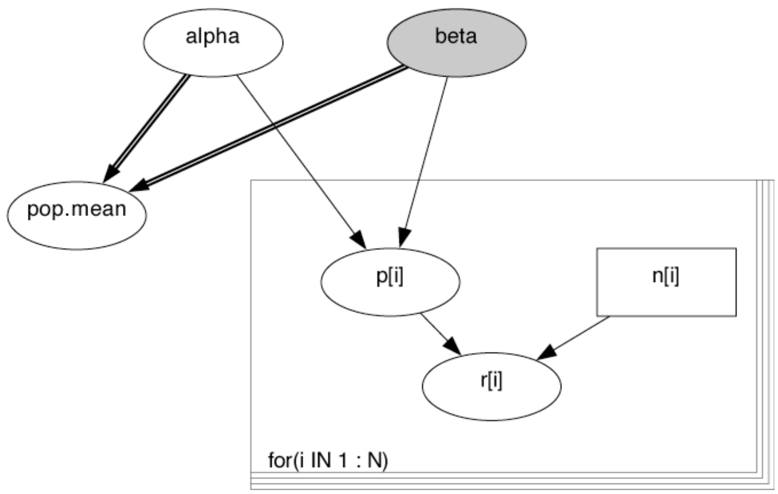


Figure 5. Directed Graph of Hierarchical Bayes Model Using Beta Distribution as Prior for Defect Rate (A)

Case A

Figure 5 depicts the directed graph for the case where the defect rate p is a beta-distributed random variable with parameters α and β, which are in turn, random variables (Weibull and uniform respectively).

$$g(\alpha \mid \lambda,\nu) = \lambda\nu p^{\nu-1} \exp[-\lambda\alpha^{\nu}], \quad 0 \le \lambda,\nu \le \infty \tag{11}$$

$$g(\beta) = \begin{cases} 1, & 0 < \beta < 10 \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

It is important to recognize that the underlying defect rates for the subunits in all tests are assumed to be similar as depicted by the parameters α and β. This permits the results available from each test to be merged into an estimate of the overall population defect rate (*pop.mean*). The choice of the beta distribution as the prior is for computational simplicity, where the posterior distribution follows the same parametric form as the prior distribution, and for practical advantage, where the posterior mean is interpretable as the experimental/current data plus the additional prior data. As with the classical Bayesian analysis, the HB analysis for this study was accomplished using the WinBugs software package.

Case B

Figure 6 depicts the hierarchical model of the analysis where the number of defects *r* is a binomial random variable, with parameters *n* and *p*. The number of subunit tests *n* is considered a constant for each sample, but *p* is again considered a random variable. A logit transformation $b_i = \log(p_i/(1 - p_i))$ is taken. Both theoretical and practical considerations suggest that when the response *r* is an indicator variable (success or failure), the response function can be modeled assuming b_i is a normally distributed random variable with mean μ and precision τ. As indicated by the ovals surrounding μ and τ, they are considered random variables with unique statistical characteristics.

The key point to observe is that the subunits evaluated in each of the *N* test programs

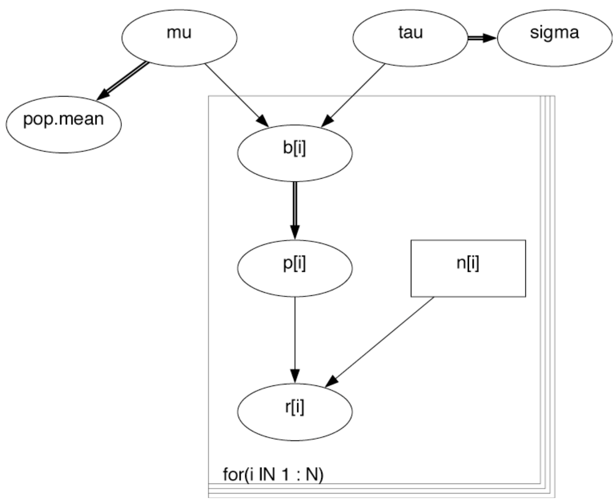


Figure 6. Directed Graph of Hierarchical Bayes Model Using Logit Transformed Normal Distribution Prior for Defect Rate

share common characteristics, μ and τ. Although the true values of μ and τ remain unknown, it is assumed that the subunits being tested are similar enough that their true defect rates come from the same family of distributions.

Hierarchical Analyses and Results

Case A

As depicted in Figure 5, the probability of observing a defect is considered a beta distributed random variable with parameters □ and □. Since these parameters are not known, it is assumed that they are in turn random variables. Based on engineering judgement, the following hyperparameter values were chosen. The uncertainty in parameter □ is assumed to be characterized by a Weibull density function with shape parameter □=1, and scale parameter, □=10. The parameter beta is also a random variable and is assumed to be uniformly distributed random variable over the interval (0,10]. The results obtained from simulating these two distributions are depicted in Figure 7.

In an abstract sense, by assuming that the parameters of the prior distribution are not known exactly, the prior distribution can be thought of as being sampled from a family of distributions. For Case 1, a random set of 30 prior distributions is depicted in Figure 8. As can be seen a wide variety of outcomes are possible, each representing a possible description of the prior information on the true defect rate.

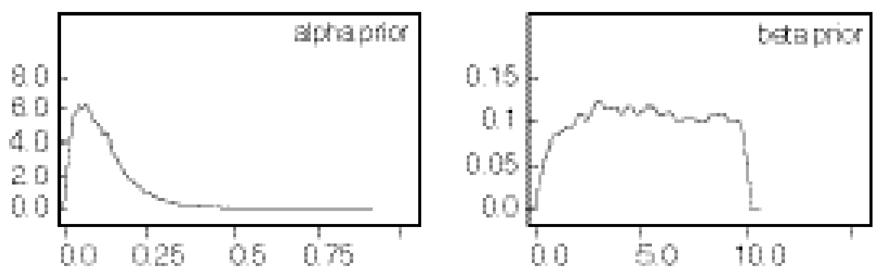


Figure 7. Parameter Prior PDF – Hierarchical Bayes: Beta Prior

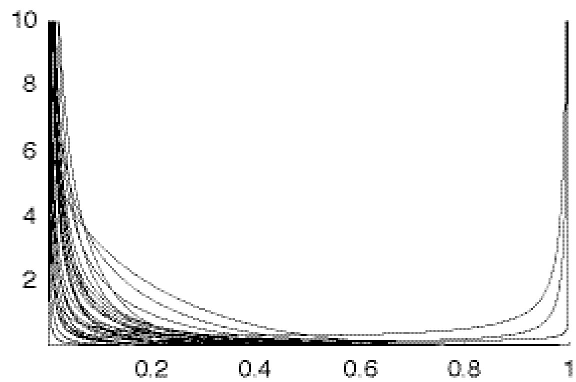


Figure 8. Sample of Possible Beta Priors

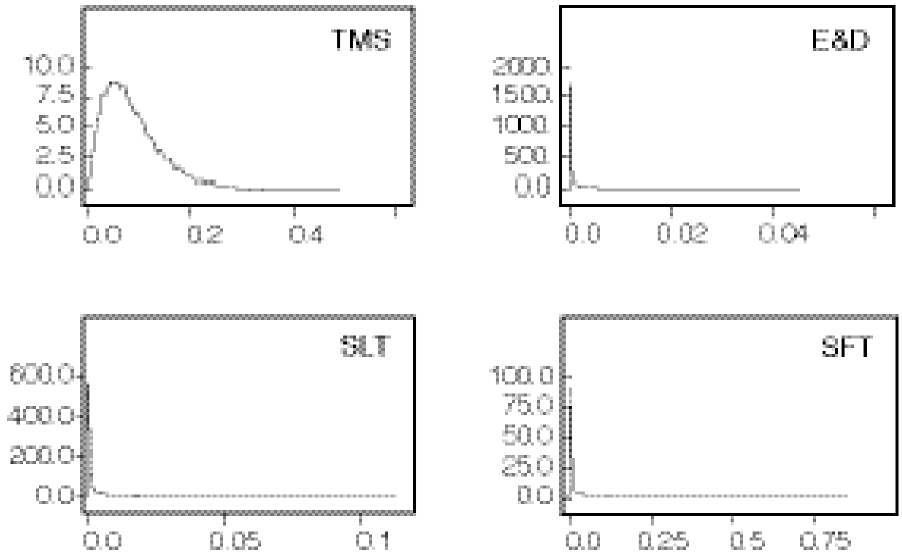


Figure 9. Posterior Density Functions of Defect Rates - Hierarchical Bayes: Beta Prior

Figure 9 depicts the probability density functions of the defect rates associated with each of the four tests. Since the subunits being tested are assumed to be related, an estimate of the population density function is available (Figure 10). Table 5 summarizes the results of the analysis and contrasts the results with those from the previous traditional Frequentist analysis; included are comparisons between the traditional confidence intervals and the probability intervals that result from the Bayesian analysis. As we can see in Table 5, the 90% credible sets are significantly reduced when using the Bayesian approach.

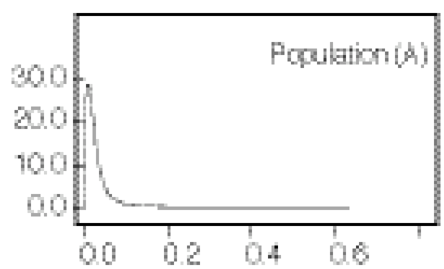


Figure 10. Population Posterior Density Function - Hierarchical Bayes: Beta Prior

Mode	Tests	Failures	Analysis Method	10% LCL	Median	90% UCL
TMS	20	2	Frequentist	0.0565	0.1315	0.245
			Hierarchical (A)	0.0250	0.0731	0.1609
E & D	124	0	Frequentist	0.001	0.006	0.0185
			Hierarchical (A)	5. E-19	1.0 E-6	0.00027
REST Lab (SLT)	36	0	Frequentist	0.003	0.0195	0.062
			Hierarchical (A)	2. E-18	1.2 E-5	0.0547
REST Flight (SFT)	2	0	Frequentist	0.0515	0.293	0.684
			Hierarchical (A)	6. E-18	6.6 E-5	0.2939
Population			Hierarchical (A)	0.0060	0.0209	0.0663

Table 5. Summary of Hierarchical Bayes Confidence Interval Analysis: Beta Prior

Case B

The second case is depicted in Figure 6. In this situation the characterization of a non-informative prior that is common in the statistical literature is assumed. As before, the hierarchical model assumes that the number of defects *r* is a binomial random variable, with parameters *n* and *p*. The number of tests *n* is considered a constant for each sample, but *p* is considered a random variable. However, in this case, the defect rate is the result of the transformation: $p_i = \exp[b_i] / \{1 □ \exp[b_i]\}$ where b_i is a normally distributed random variable with mean □ and precision □. Since information regarding the density function for b_i is limited □ and □ are assumed to be random variables. However, it is known that the equipment under each test is similar, so it is assumed that the b_i share common □ and □. Following the suggestions in the statistical literature when

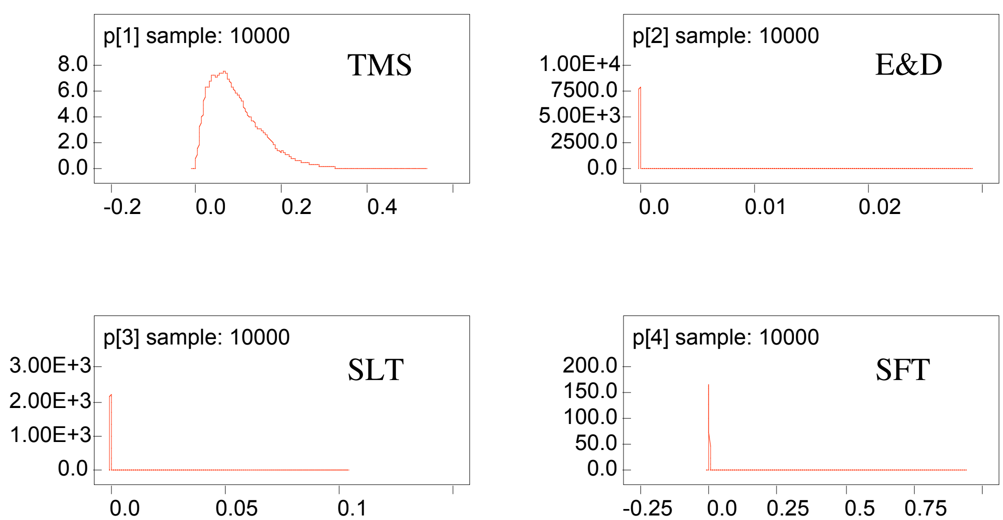


Figure 11. Probability Density Function of Defect Rates – Hierarchical Bayes: Logit Transformation

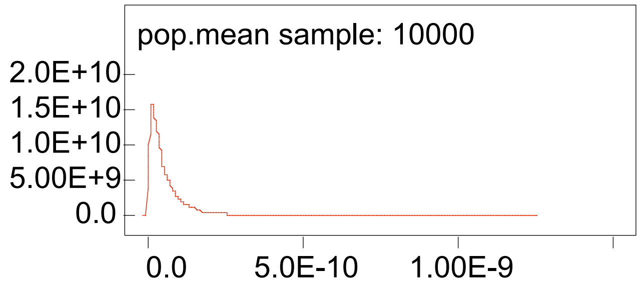


Figure 12. Population Posterior Density Function - Logit Transformation

characterizing a non-informative prior in conjunction with the logit transformation, The random variable □is assumed to be a gamma distributed random variable with a shape parameter 0.001 and a scale parameter of 0.001. The □ is assumed to be a normally distributed random variable with a precision of 0.001 and a mean of –24. The value of □24 for □ is informative however and equivalent to the mean of logit(p_i) in Case A above which was determined based on engineering judgement. Note that *precision* is defined as 1/variance. [Leonard (1972), Congdon, p17-18, (2001)].

Figure 11 depicts the probability density functions for the defect rate for each of the different test modalities and Figure 12 depicts an estimate of the defect rate for the entire population based on all four test modalities. As we can see in Table 6, the 90% credible sets are significantly reduced when using the Bayesian approach.

Mode	Tests	Failures	Analysis Method	10% LCL	Median	90% UCL
TMS	20	2	Frequentist	0.0565	0.1315	0.245
			Hierarchical (B)	0.0255	0.0816	0.182
E & D	124	0	Frequentist	0.001	0.006	0.0185
			Hierarchical (B)	0.0	2.0E-13	6.0E-6
REST Lab (SLT)	36	0	Frequentist	0.003	0.0195	0.062
			Hierarchical (B)	0.0	3.2E-13	2.0E-5
REST Flight (SFT)	2	0	Frequentist	0.0515	0.293	0.684
			Hierarchical (B)	0.0	7.1E-13	1.1E-4
Population			Hierarchical (B)	1.1E-11	4.0E-11	1.4E-10

Table 6. Summary of Hierarchical Bayes Confidence Interval Analysis: Logit Prior

Summary

The objective of this report was to explore the various approaches currently available to characterize the reliability of systems based on data available from a variety of test modalities. The three approaches investigated were a traditional Frequentist approach, a classical Bayesian approach with non-informative and informative priors and a hierarchical Bayesian approach using two analysis structures. Table 7 and Figure 13 summarize the results investigated in this report. *Note that all results are accurate to the third decimal place only*.

The most notable difference in the three approaches is that the hierarchical Bayesian approach permits the integration of test data from a variety of test modalities. While it is relatively common practice to merge data in all three analysis techniques, only with the HB technique is the practice with a sound theoretical basis.

With the objective of the report said, a few issues related to the analyses in the report were not addressed and should be evaluated in the immediate future. First of all, all analysis results indicate that TMS data may be different from the other 3 testing data. The differences may be due to a burn-in period when the TMS data was taken or due to the specifics of the testing procedures. The poolability of the data sources should be carefully evaluated. Secondly, the time period when the data was collected was over a course of several years. The possible degradation trend was not examined. Finally, case B in the hierarchical Bayesian analysis gave a tighter confidence interval than case A; however the behaviors of these prior distributions are quite complicated and their application in these types of problems is suspect. Further research into this type of prior warrants further investigation before it is used in critical situations.

Mode	Tests	Failures	Analysis Method	10% LCL	Median	90% UCL
TMS	20	2	Frequentist	0.0565	0.1315	0.245
			Classical Bayes (A)	0.0543	0.1227	0.238
			Classical Bayes (B)	0.0449	0.1037	0.199
			Hierarchical (A)	0.0250	0.0731	0.1609
			Hierarchical (B)	0.0255	0.0816	0.182
E & D	124	0	Frequentist	0.001	0.006	0.0185
			Classical Bayes (A)	0.001	0.0058	0.01895
			Classical Bayes (B)	0.0007	0.0052	0.0174
			Hierarchical (A)	5. E-19	1.0 E-6	0.00027
			Hierarchical (B)	0.0	2.0E-13	6.0E-6
REST Lab (SLT)	36	0	Frequentist	0.003	0.0195	0.062
			Classical Bayes (A)	0.0033	0.01898	0.0663
			Classical Bayes (B)	0.0025	0.01657	0.0547
			Hierarchical (A)	2. E-18	1.2 E-5	0.0547
			Hierarchical (B)	0.0	3.2E-13	2.0E-5
REST Flight(SFT)	2	0	Frequentist	0.0515	0.293	0.684
			Classical Bayes (A)	0.0354	0.202	0.523
			Classical Bayes (B)	0.0140	0.09217	0.2939
			Hierarchical (A)	6. E-18	6.6 E-5	0.2939
			Hierarchical (B)	0.0	7.1E-13	1.1E-4
Population			Hierarchical (A)	0.0060	0.0209	0.0663
			Hierarchical (B)	1.1E-11	4.0E-11	1.4E-10

Table 7. Summary of Hierarchical Bayes Confidence Interval Analysis

As is expected, as the sample size increases all methods provide increasingly similar results. When the number of test samples is very small, e.g. n=2, the Bayesian methods are clearly influenced by the use of information beyond that available from the immediate test regime. However, as the sample size increases even slightly, e.g. n=20, the test data begins to dominate whatever prior assumptions were made about the underlying defect rate. This is clearly evident when comparing the results from using the two very different priors in the classical Bayesian analysis. With that said, there is something unique about the hierarchical Bayesian analysis.

The results from the HB method clearly give 'optimistic' results when compared to either of the other methods. This is a direct result of assuming that the subunits in the various tests are similar. The key here is that if you believe that the E &D testing truly provides an accurate representation of the underlying defect rate, then the information from these 124 tests can be used to increase the confidence in the results from the two flight tests.

This report has demonstrated a mathematical foundation for an approach to bring together data from a variety of sources and suggested an approach to combine this data in a logical fashion. It also demonstrates that the confidence of the results is increased due to this use of pooled data. In Figure 13, the confidence sets are greater/narrower for the population mean where 4 test data were pooled together. The application of a hierarchical Bayesian approach will result in an increased confidence in the reliability assessment of the nuclear weapons stockpile.

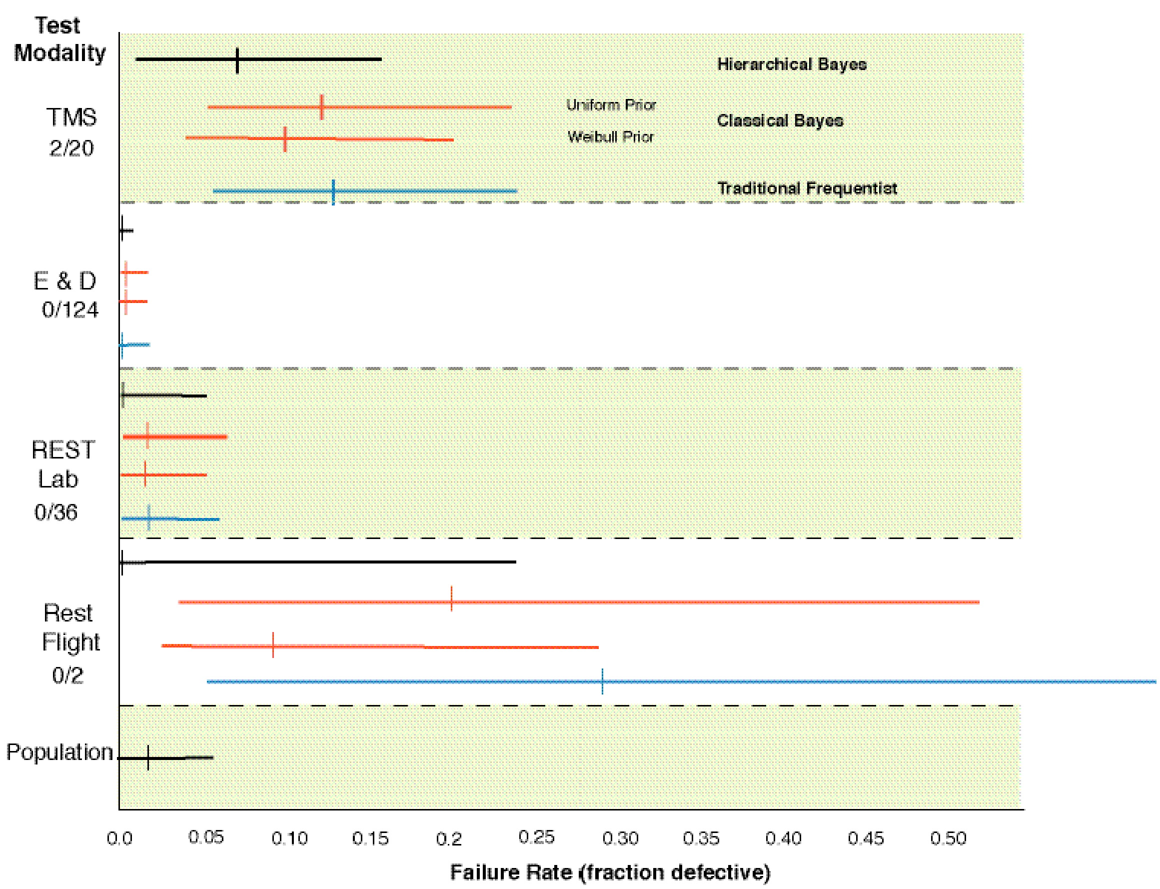


Figure 13. Summary of Analysis Results

References

Models with Bayesian Updating, Applied Mathematics and Computation, Vol. 73, pp
Azaiez, M., V. Bier (1995), *Perfect Aggregation for a Class of General Reliability*
281-302. (GB)

Leonard, T., (1972), *Bayesian Methods for Binomial Data*, Biometrika, Vol. 59, No. 3, pp581-589.

Congdon, P. (2001), Bayesian Statistical Modeling, Wiley and Sons, New York, NY.

Spiegelhalter, D. J., A. Thomas, N. Best, W. Gilks, (1996), BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.5, (version ii), MRC Biostatisics Unit, UK.

Robinson, D. (2001), *A Hierarchical Bayes Approach to System Reliability Analysis*, SAND 2001-3513, Sandia National Laboratories, Albuquerque, NM, USA.

Distribution

1	MS 1407	R. J. Salzbrenner	(1805)
1	MS0479	J. O. Harrison	(2113)
1	MS 0639	R. W. Lizut	(2950)
1	MS 9033	A. J. West	(2253)
1	MS 9005	R. Miller	(2820)
1	MS 0747	G. D. Wyss	(6410)
5	MS 0748	D. G. Robinson	(6413)
1	MS 0829	J. M. Sjulin	(12323)
1	MS 0405	T. R. Jones	(12333)
1	MS 0830	K. V. Diegert	(12335)
1	MS 0830	E. W. Collins	(12335)
1	MS 1176	L. A. Swiler	(15312)

1	MS 9018	(8945-1)	Central Technical Files
1	MS0161	(11500)	Patent and Licensing Office
2	MS 0899	(9616)	Technical Library
1	MS 0612	(9612)	Review & Approval Desk for DOE/OSTI

TOTAL COPIES = 20