How Bayesian Reliability Analysis was Developed and Implemented for Production Decisions*

Bruce Baber[1], Allan T. Mense[2], C. Shane Reese[3]
[1]USAF, Eglin AFB, FL
[2]Raytheon Missile Systems, Tucson, AZ
[3]Brigham Young University, Provo, UT

Abstract
Often in the development of complex discrete functioning systems the systems level testing is very limited at the point significant decisions are made in the development process. One such point is typically the production decision to commit large resources to low rate initial production concurrent with the completion of developmental and operational testing. This condition introduces significant risk into the program and technical management of these type systems. Often these type systems are not designed from scratch, but utilize components and subsystems from previous programs that have extensive usage data in similar or identical environments. Most developments require extensive component and subsystem design verification testing and qualification testing across most of the environments that are expected to be encountered. A method is needed to utilize previous system development and production data and subsystem level test results combined with the systems level test data available when making reliability assessments. This paper explores Bayesian methodology to combine different types of data into a mathematically useful result for evaluating system reliability for these types of systems. The model presented is relatively simple, but allows combining expert opinion, previous system data, component and subsystem level testing with a limited amount of system level testing to develop a more comprehensive reliability case early in the system level test phase, but at point when significant program decisions must be made.

Key Words: Bayesian, reliability, production, discrete systems

1. Introduction

Often in the development of complex discrete functioning systems the systems level testing is very limited at the point significant decisions are made in the development process. One such point is typically the production decision to commit large resources to low rate initial production concurrent with the completion of developmental and operational testing. A method is needed to utilize previous system development and production data and subsystem level test

* Approved for public release; distribution is unlimited.

results combined with the systems level test data available when making reliability assessments.

It's not so important that we have an exact point estimate of our reliability, as it is that we have a method of measuring the confidence we have in meeting our decision criteria.

It is also important that our statistical data be correlated to identifying risk areas as we proceed with development and production. Being able to combine component, subsystem level, and system level data gives us this ability. The production decision reliability case study assesses confidence in meeting requirements along with a detailed description of the environmental and functional test exposure and failure modes identified and corrected. Failure modes are characterized by probability giving us a means of assessing the residual risk at the time of the production decision. The Bayesian approach of combining different test data allows us to assess the system across a wide array of environmental and functional exposure as well as evaluating increasing levels of complexity as the system is integrated at higher levels. This approach gives us both breadth and depth of evaluation. More complex versions of these type models have been developed by Reese and Mense to handle very complex systems with multiple mission phases and "different modalities distinguished by test fidelity and level of test (system vs. component)" [Reese]. However, we attempted to create a simpler model that would be more readily accepted and easily used by the practicing reliability engineer while providing the fidelity of result appropriate for the decision in question. <u>The significant question early in a program at the production decision point-is the system on track to meet its reliability requirements.</u> This is measured by comparing current reliability parameters to values on a planned growth curve. Typical statistical analysis usually "seeks objectivity by generally restricting the information" [Mense] used to system level testing in the actual usage environment which is "clearly relevant data" [Mense], but also very limited. The parameters such as a mean reliability are considered "fixed but unknown" [Mense] and estimated from a very small sample of data. The Bayesian approach considers these reliability parameters as "random, not fixed" [Mense] and uses previous system data and subsystem testing to develop a prior understanding of these reliability parameters and then modifies them using the system level likelihood data into a posterior distribution of the reliability parameters along with "credibility intervals" [Mense] for use in making inference statements about the maturity of the system relative to expectations. We need a method of evaluating a complex system at points in development when we have a limited amount of system level data requiring full functional exercise in the actual usage environment. An approach is needed to prevent underestimating or over estimating reliability early in the evaluation and test cycle. We want to protect against perfect assessment early on or low estimates if there are early failures. An effort should be taken to make use of all data available including previous data on similar systems and expert opinion as to component, subsystem and system reliabilities. Our desire is for reliability to converge to that given by actual test data as more and more data becomes available. We are seeking a method that correctly estimates actual confidence for the reliability relative to our decision criteria. Classical confidence intervals do not give an interval of interest. The posterior distribution used to make a production decision, will produce the prior distribution that will be used for later test phases.

2. Materials and Methods

This paper explores Bayesian methodology to combine different types of data into a mathematically useful result for evaluating system reliability for these types of systems. The model presented is relatively simple, but allows combining expert opinion, previous system data, component and subsystem level testing with a limited amount of system level testing to develop a more comprehensive reliability case early in the system level test phase, but at a point when significant program decisions must be made. The type of system under consideration is relatively complex electromechanical system with extensive imbedded software and several discrete functions and components. We break these systems down into several functional areas generally representative of these types of systems. The system is then further broken down into approximately 20 components for which we have expert opinions, previous system data, component and subsystem design verification and qualification testing. Finally, we utilize system level test and assigned failure modes for those tests in which we could identify a failed subsystem or component. Prior distributions were developed for components and subsystems. Likelihood functions were developed at the system level and for components and subsystems where they could be assigned. We are developing the prior distribution for later test phases by finding a posterior distribution based on subsystem priors and a variety of partial system level tests. The component priors have been created either from component test data or subject matter expert inputs and used to specify the values for nprior[i] and rprior[i] that are used in a beta distribution e.g.

$$\mathbf{f_{prior}(R)=Constant* R^{nprior*rprior} * (1\text{-}R)^{nprior*(1\text{-}rprior)}}$$

- rprior is the mode of the prior distribution and
- nprior is a weighting (or importance) factor .
- nprior=0 gives a uniform prior.

Computation was performed using Markov Chain Monte Carlo (MCMC). The updates were all performed using the Metropolis-Hastings (M-H) algorithm. Updates were done with beta proposal distributions centered at the previous value. The acceptance probability was adjusted to reflect the asymmetric proposal density. This Bayesian reliability model performs reliability analysis on the system whose functional components are assumed to all be in series from a reliability standpoint, i.e. if any component fails the system fails. The initial prior assessment is represented by a beta probability distribution with given parameters. The prior distribution is “multiplied” by the likelihood function using the data from development testing of the product. This results in a joint distribution with approximately 20 different reliabilities and from this one samples to obtain the required posterior distributions for the components as well as the system reliability distribution. Statistical inferences can be made from information conveyed by the posterior distribution.

1. Select an appropriate prior probability distribution
2. Obtain new evidence (data)
3. Choose a likelihood function, based on data type
4. Update the prior distribution with the new evidence to generate a posterior probability distribution.

2.1 Selecting the Prior

For the system under consideration "prior knowledge exists for each subsystem" therefore we used a beta prior distribution to account for this "existing knowledge". The beta distribution is in the form:

P_{ij} = estimate of reliability for the j^{th} component at the i^{th} iteration
N_m = accuracy weighting for the predicted reliability

This form of the beta distribution has been used previously by Mense and is called the Los Alamos formulation for a beta prior distribution. Where previous system test data existed to provide us with "confidence about the reliability of the item" we used it to establish a strong prior resulting in a "narrow and peaked distribution" that was not significantly affected by the actual test results. In most cases there was not a failure in the actual testing. For those cases where little previous system data or component or subsystem testing did not exist we used a weak prior, with a "weak distribution that is wide and relatively flat" to allow the posterior to be more influence by the actual system level test data [Mense].

2.2 Likelihood Function

The likelihood function was evaluated using system level test data. The priors were developed using previous system data and component and subsystem test data, no system level test data was used to form the priors. Therefore, independence was maintained. Since the system under consideration includes many discrete functions, a binomial likelihood function was chosen. The reliability of the component is R. The likelihood function indicates that n tests were performed on this component and of those n tests there were s successes and (n-s) failures. This results in the likelihood function having the form shown below:

$$L = \prod_i L(R_i \mid s_i, n_i) \propto \prod_i R_i^{s_i} (1 - R_i)^{n_i - s_i}$$

$$\ln(L) \propto \sum_i \left[s_i \ln(R_i) + (n_i - s_i) \ln(1 - R_i) \right]$$

2.3 Calculating the Posterior Distribution

Using the beta prior distributions and the binomial likelihood functions the posterior distribution was calculated using the following:

$$f_{\text{joint}}(R_1, R_2, ..., R_n) \propto \sum_i \left[\left(s_i + \mathrm{N}_{mi}\,\pi_i \right) \ln(R_i) + (\mathrm{N}_{Mi}(1-\pi_i) + (n_i - s_i)) \ln(1-R_i) \right]$$

For the complex multi-component system the posterior distribution is found numerically using MCMC sampling of the joint distribution and applying the M-H selection algorithm. From each subsystem's posterior distribution, we extracted the reliability for each subsystem and multiplying each of these reliabilities for each iteration of the MCMC we constructed the posterior distribution for the entire system. With the posterior distribution for the system we also determined the 80% "credibility interval" around the median reliability.

3. Results

Posterior distributions were developed for each component and subsystem as well as at the system level. The results are consistent with both a classically development point estimate for system level mission success testing and a demonstrated growth value developed using failure modes identified during system level testing. However, the credible interval was much narrower than the classical Clopper-Pearson estimate since the likelihood function includes tests that were each different in terms of which components were being tested. This information cannot be handled in a non-Bayesian Pass/Fail model. Also the classical Army Material Systems Analysis Activity (AMSAA) growth model does not exactly agree with the Bayesian approach as the posterior distribution takes into account all of the expert opinion, previous system data, and component and subsystem design verification and qualification testing. However, the combination of these three measurements provides an evaluation tool that has both community acceptance and achieves a greater confidence in results relative to the decision criteria. The component and subsystem posterior distributions when utilized in conjunction with the description of environmental and functional test exposure and the failure modes identified and corrected provides a subjective verification of the applicability of the model and data to the decision at hand. The same components and subsystems that degrade system level reliability are identified as being most susceptible to functional failure when exposed to environments. This provides intuitively satisfying results and a means to assess the risk in moving forward with development and production. The system level posterior distribution yielded a mean and median in the low 0.80s and the 80% credible interval from the low 0.70s to the mid to high 0.80s or about 0.15 (Fig. 1). This is well within the 80% confidence interval of a classical sample result of 10 successes out of 13 tests or 0.77 mean and a classical 80% confidence interval of [0.56 , 0.91]. The demonstrated growth value for the system level failure modes was in the low to mid 0.80s. This gives us high confidence that the cumulative probability of success is the mid 0.70s and the growth value is in the low to mid 0.80s. This allowed us to assess the system would meet the required reliability of approximately 0.90 in the middle of low rate production based on the planned reliability growth test.

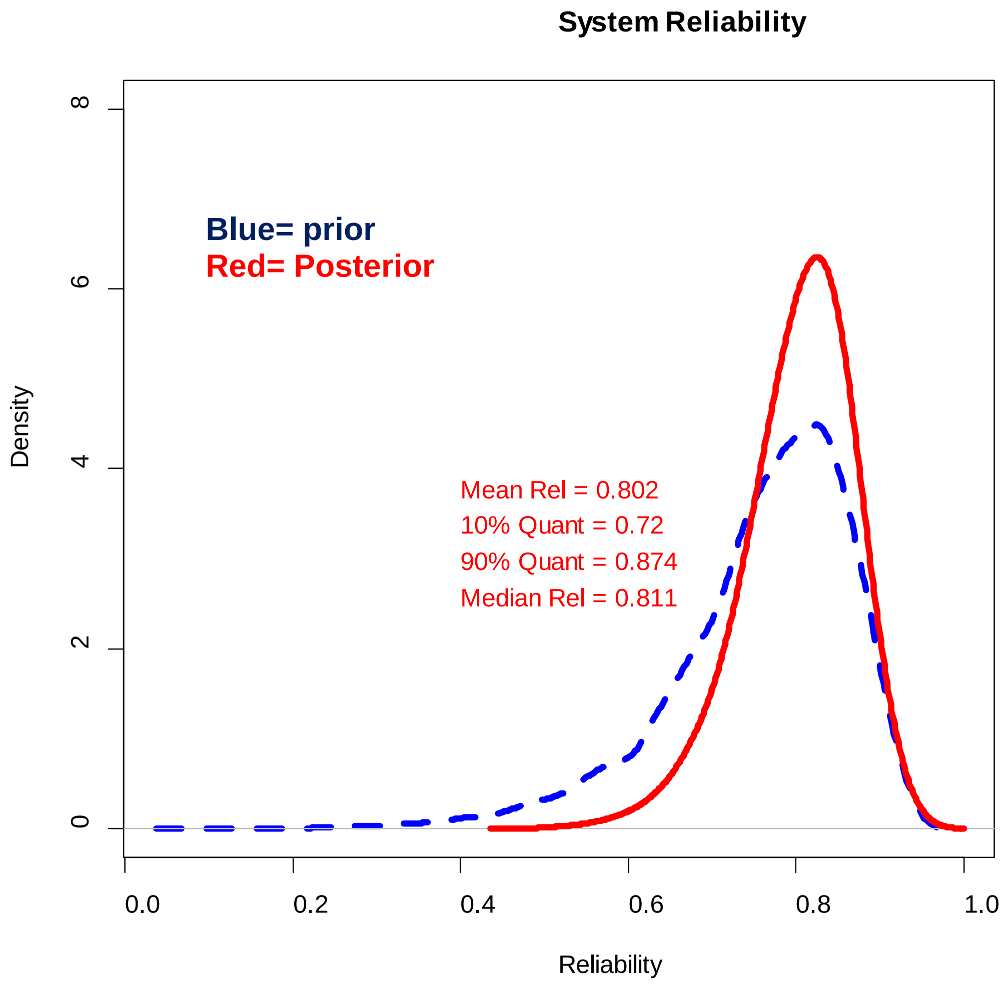

Figure 1: System level Bayesian prior and posterior distributions with mean, median, and 10% and 90% quantiles

Additionally, we're able to provide component and subsystem reliability distributions to substantiate the validity of the analysis and to correlate with identified failure modes for residual risk assessment. There were five components and subsystems that were of particular concern. These components and subsystems were give relatively low strength priors to account for the complexity and newness of the designs. System level failures on these components with weak priors yielded posterior distributions that were consistent with community perceptions of these type components expected reliability (Fig. 2). Components with priors that reflected lower reliability from previous systems will reflect that reliability until there is sufficient likelihood data to overcome the prior.

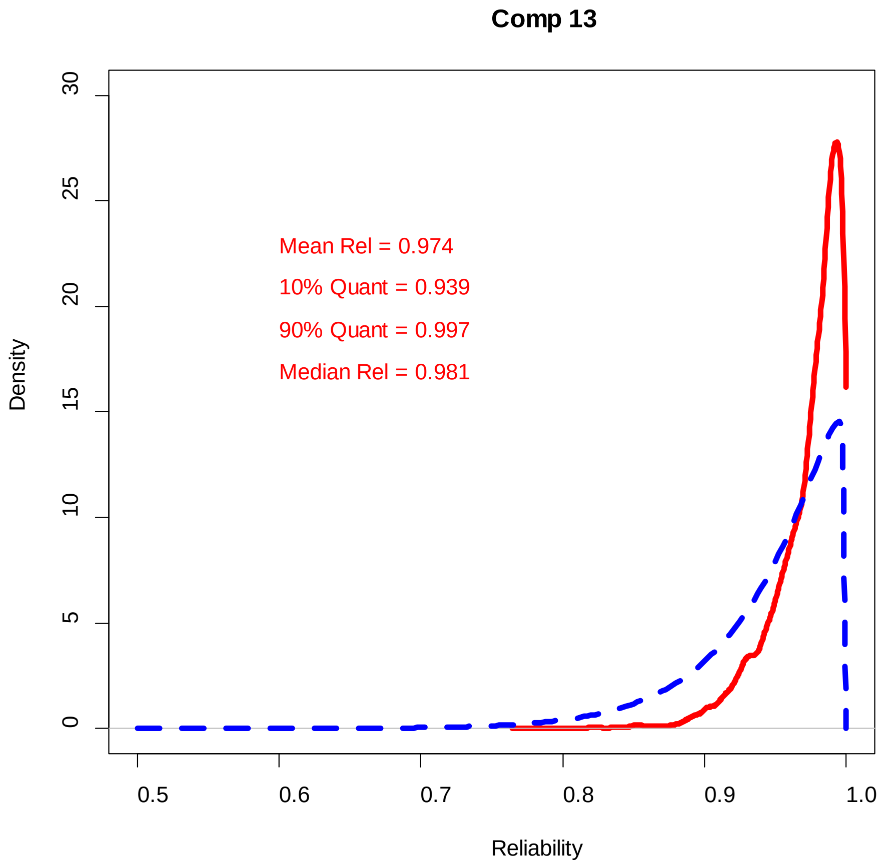

Figure 2: Component with weak prior and sufficient likelihood data to overcome. Results are consistent with community perceptions for these type components.

These results and simply the visibility into the component and subsystem level enhance the credibility of the model. This approach would appear to be superior to simply relying on the presence or absence of a failure mode of a certain component or subsystem to accurately reflect the expected contribution during longer term testing.

4. Discussion

The Bayesian model presented here builds on earlier work by simplifying their application to a specific, but significant use, and estimating the system reliability early in the program at the point of making a production decision. The model is useful, when used in conjunction with classical results and reliability growth analysis, for characterizing system level discrete reliability at this early stage of the test program. The model can be easily adapted for use on similar systems by simply renaming the functional areas and subsystems and components. The flexibility of utilizing prior information from expert opinion, previous system test and production, and subsystem and component design verification and qualification testing makes the model adaptable to almost any program situation. The use of attribute data, success or failure, in the likelihood function allows the use of most system level test data with relatively small interpolations. Posterior statistical results provided at the system level provide the ability to make system level reliability evaluations with confidence against program requirements. Additionally, posterior results provided at the subsystem and component level provide the ability to statistically evaluate risk areas. These can be especially useful when used in conjunction with a qualitative assessment of test results in terms of subsystem and component exposure to environments and functional requirements with failure modes and corrective actions identified. This Bayesian reliability model is simple enough to promote widespread use and opens up several possibilities for enhancement and adaptation.

References

Reese, C Shane, Hamada, Michael, and Robinson, David. 2005. “Assessing System Reliability by Combining Multilevel Data from Different Test Modalities.
Mense, Allan. “Bayesian Reliability Point Paper”.