Bayesian Reliability Demonstration*

F.P.A. Coolen and **P. Coolen-Schrijner**
Department of Mathematical Sciences, Durham University
Durham DH1 3LE, UK

Abstract

This paper presents several main aspects of Bayesian reliability demonstration, together with a concise discussion of key contributions to this topic.

1 Introduction

Many systems are required to be highly reliable, and prior to their practical use such reliability may need to be demonstrated through testing. This requires determination of the amount and type of testing that is needed in order to demonstrate a certain level of reliability. In many situations, in particular where high reliability is required, one would not accept faults occurring during such testing, as these are evidence against the system's ability to perform its task well and are likely to lead to rejection of the system. As such reliability demonstration testing clearly deals with uncertainty, and information to reduce this uncertainty, statistical methods can provide valuable insights and can guide such testing. We focus on Bayesian methods, which allow previous information to be taken into account in a subjective manner via the prior distribution chosen, and, as such, can lead to reduced test effort. However, when emphasizing the *demonstration* of reliability, it might not be considered appropriate to rely on prior information, be it based on historical data from similar systems or expert judgement. We will discuss this feature in detail, as it strongly influences the choice of prior distribution. But first, we provide a brief overview of some key contributions to the literature on Bayesian reliability demonstration, and we focus on the criterion used for 'reliability'.

*Contribution to *Wiley Encyclopedia of Statistics in Quality and Reliability*

In line with the increasing development and popularity of Bayesian statistical inference in the late 1960's and 1970's, many researchers in reliability developed Bayesian reliability demonstration (BRD) test philosophies and procedures during this period. An excellent overview of contributions to BRD during this period is Chapter 10 of Martz and Waller's well-known book 'Bayesian Reliability Analysis' [16]. A key topic is the choice of BRD criterion. Traditionally [16], criteria considered were closely linked to those in related statistical areas such as quality control, and mostly used producer's and consumer's risks expressed in terms of acceptance or rejection of a system after BRD testing, conditional on assumptions on characteristics of the system's failure distribution. Here, we use 'failure distribution' as a generic term for the probability distribution of the random variables of interest in BRD, which might e.g. include numbers of failures over defined periods of testing (as is typical in so-called 'attribute testing') or times between failures. The characteristics typically used are a parameter representing the proportion of failures, as a population characteristic in the standard statistical sense for Bernoulli random quantities (attribute testing) or Mean-Time-To-Failure (MTTF) or Mean-Time-Between-Failures (MTBF). It is important to notice that, although such characteristics have intuitive appeal, they are not directly observable and only meaningful in combination with additional statistical modelling assumptions (mostly exchangeability of the random quantities of interest). The Bayesian approach to reliability demonstration distinguishes itself from other statistical approaches to such problems via the use of prior distributions on these parameters, which can reflect prior knowledge - a crucial question is, of course, whether or not such prior knowledge should be included in BRD and, if perhaps not, what role the Bayesian prior then might have and how it can be chosen (we return to this important issue in Section 3). Once one accepts the more or less traditional framework for sample size determination via producer's and consumer's risks, a variety of probabilities can be considered for the detailed specification of the BRD criterion, and all of these lead to many different solutions, which are however closely related from philosophical perspectives. Such probabilities include posterior risks, average risks and rejection probabilities, and study of corresponding BRD approaches led to a significant number of papers in the literature until the late 1970s [16]. A typical contribution to this literature is the study of behaviour of some quantities used in BRD tests, by Higgins and Tsokos [11]. They show that the use of two different priors for MTBF, which cannot be easily distinguished between at the prior stage, can lead to substantial differences in the posterior risks in BRD. From this, we can conclude that choice of the prior distribution was known to be crucial and problematic in BRD, yet it has received relatively little attention in the BRD literature.

Since the early 1980s, BRD seemed to disappear as a main topic from the reliability literature, with further development mostly in industry, so less frequently reported in the academic literature, and often embedded in wider programmes for achieving quality and reliability. Examples of such progress on methods for reliability demonstration, including Bayesian approaches, in the automotive industry are presented by Lu and Rudy [13, 14]. Recently, Yadav, *et al.*, [22] raised interesting issues on reliability demonstration, emphasizing the need for a comprehensive approach which integrates reliability demonstration into the product development process. They emphasize that three dimensions of product design must be considered: physical structure, functional requirements, and time in service. They propose some statistical methodology for specific distributions and test plans, but although they mention the importance of including prior information, they do not formalize this nor do they use a Bayesian approach. This work, however, provides interesting arguments for renewed interest in BRD.

In recent years, the current authors have presented, in part jointly with Rahrouh, a different perspective on BRD, focussing solely on attribute testing [5, 6, 7, 8, 19]. This work has mainly been motivated by the observation that methods presented in the literature seemed not to take explicitly into account characteristics of the process in which the system is to be used after BRD testing, nor practical constraints on such testing. For example, they distinguished between the cases of catastrophic and non-catastrophic system failure in the process after testing, and they also studied optimal decisions for both BRD and choice of redundancy level in a system, from overall cost perspective. They also considered the role and choice of the prior distribution in BRD in detail, which is crucial yet far from trivial and had not yet received much attention. We briefly summarize main results in Sections 2 and 3, illustrated by some examples in Section 4, and discuss some further aspects in Section 5.

2 Zero-failure reliability demonstration

Martz and Waller [15] considered the before mentioned risk-based BRD from the perspective of only tests that reveal zero failures. This has two advantages. First, in practical demonstration of, particularly, high reliability, occurrence of one or more failures goes against the very nature of reliability demonstration, and is likely to lead to improvement actions of the system, where possible, followed by renewed reliability demonstration. This cycle brings with it very important challenges, also called 'test-fix-test', which are beyond the scope of this paper. A second advantage of restricting attention to inference only valid when testing reveals zero failures, is the relative mathematical simplicity of the resulting Bayesian posterior and corresponding predictive distributions, enabling these to be embedded in wider decision-theoretic frameworks involving aspects of the process in which the system is required after testing, and inclusion of costs and time considerations and constraints. This and the following two sections provide some details of such a predictive approach to zero-failure BRD, with attention restricted to attribute testing [5, 6, 7, 8, 19].

We consider systems which have to deal satisfactorily with tasks of k different types, where we assume complete independence between tasks of different type (both in the success probability of the system to deal with such different tasks, and in the statistical sense). In a standard Bayesian setting [1, 9, 16], we use a Binomial model for the number of failures in n_i tests of tasks of type i, for $i = 1, \ldots, k$, with parameter θ_i representing the unknown probability of a failure when the system has to deal with a task of type i. We assume a Beta(α_i, β_i) prior distribution for θ_i, and we further assume that zero failures will be observed in the n_i tests, denoted as data $(n_i, 0)$, so these inferences are only valid for zero-failure BRD. The corresponding posterior predictive probability of zero failures in m_i tasks of type i, in the process after such testing, is

$$P(0|m_i, (n_i, 0)) = \prod_{j=1}^{m_i} \frac{j + \beta_i + n_i - 1}{j + \alpha_i + \beta_i + n_i - 1}. \quad (1)$$

For $n_i > 0$, (1) holds for $\alpha_i > 0$ and $\beta_i \geq 0$. Because no failures are observed we cannot use $\alpha_i = 0$, as the corresponding posterior distribution would be improper [1]. The hyperparameters α_i and β_i can be interpreted in terms of results of an imaginary earlier test of $\alpha_i + \beta_i$ tasks of type i, in which the system failed to deal with α_i such tasks, but performed β_i such tasks successfully. For BRD, this implies that inferences are quite insensitive to choice of β_i, where an increase in β_i means that the minimal required n_i is reduced by the same number. However, such inferences are highly sensitive to the choice of α_i, as effectively the reliability demonstration requires n_i tests without failures to counter the prior information of α_i 'imaginary test failures'. We discuss the choice of the prior distribution in more detail in Section 3, where we will advocate the use of Beta prior distributions with $\alpha_i = 1$ and $\beta_i = 0$. It should be noted that this chosen value for β_i leads to an improper prior distribution, so formally not really a probability distribution as it does not integrate to one over the parameter space. No statistical inferences can be based on improper distributions, but in this setting we assume explicitly that, for posterior inference for demonstrating reliability, the data contain at least a positive number of observed successes, which ensures that the corresponding posterior distribution is proper.

We briefly state some results from [7], on required test numbers n_i, such that zero failures in testing ensures a probability p of no failures in the process, either for deterministic numbers of tasks or for a stochastic process considered over a fixed period of time. This setting is, for example, suitable for catastrophic failures, i.e. a failure leads to break-down of the whole system, in the sense that the clear target is to demonstrate reliability via an inferred (very) high probability of zero failures in the process after testing, when the system is actually in service. In such cases (e.g. alarm systems for chemical plants), it is often difficult to bring cost considerations into the argument, either due to inherent lack of knowledge of consequences of rare events or due to ethical issues, so requirements directly on the predictive probability of zero failures are attractive. Of course, such cost considerations are likely to influence choice of an appropriate value of p.

For the deterministic case, where the system must perform a known number m_i of tasks of type i in the process, the n_i must be such that

$$P(0|m_i, (n_i, 0), i = 1, \ldots, k) = \prod_{i=1}^{k} \prod_{j=1}^{m_i} \frac{j + \beta_i + n_i - 1}{j + \alpha_i + \beta_i + n_i - 1} \geq p. \tag{2}$$

If we aim at satisfying this requirement with a minimal total number of tests, there is no closed-form solution for n_i. For the special case with $\alpha_i = \alpha$, $s_i = \beta_i + n_i = s$ and $m_i = m$, for all i, an approximation for the minimum required value of s to meet this reliability requirement, with p close to 1, is

$$\frac{p^{1/(k\alpha)}}{1 - p^{1/(k\alpha)}} \times m. \tag{3}$$

For the case with $\alpha = 1$ and $\beta_i = 0$, for all $i = 1, \ldots, k$, which we advocate for use in reliability demonstration (see Section 3), (3) gives the exact value of the optimal (real-valued) n_i's [7] (the integer-valued solution is found by checking this probability value for the neighbouring integer-valued k-vectors). The fact that, to minimise the total number of zero-failure tests required in this situation, one would indeed wish to use equal values of s_i, is proven in [7].

For p close to 1, the expression (3) is close to

$$\frac{p}{1 - p} \times k\alpha m, \tag{4}$$

which can be proven via the same argument as used in [7] to show that (3) is approximately linear in k. The expression (4) makes clear that the required number of zero-failure tests is very sensitive to the choice of α. Clearly, the required value of n_i is less sensitive to small changes in β_i, as the sum of n_i and β_i must be a constant. The influence of the α_i and β_i on the required number of zero-failure tests is similar in more general situations with unequal α_i, s_i and m_i.

Of course, in many real-world processes the numbers of tasks m_i are not deterministic, which is for example typically the case in alarm systems. An important and interesting result is the following: let M_i be the random number of tasks of type i that the system has to perform in the process after testing. Then, for all possible probability distributions for M_i with the same expected value $E(M_i)$, the most BRD testing is actually required for the deterministic case with $m_i = E(M_i)$ (this is a consequence of Jensen's inequality). Hence, as long as one can specify an expected value for the number of tasks the system will be required to perform (or of course an upper bound to stay on the safe side), then the amount of testing required by assuming this number to be deterministic leads to conservative test numbers. It also turned out from many examples studied, although formal proof was not achieved, that if M_i actually has a Poisson distribution, which is often suitable to model accidental events, then the optimal numbers of BRD tests of tasks of type i are very close to the corresponding optimal test numbers in the deterministic case, which in practice means that over-testing would be prevented in such cases if you based your calculations upon the deterministic case instead of the Poisson case [7]. Example 1 in Section 4 illustrates this approach.

A different predictive BRD approach was presented in [8], which is suited for non-catastrophic system failure during the process after testing. This implies that the system will be able to continue operating in case such a failure occurs, and also that cost considerations become more important, so both costs of testing and of failure are taken into account. That approach aims at minimisation of overall expected costs for testing and the process after testing, and takes constraints on testing (both budget and time) into account. Optimal BRD plans are derived by solving relatively straightforward constrained optimisation problems, with several analytical results making clear the influence of constraints on testing. The reader is referred to [8] for more details, we wish to emphasize that it is again the predictive nature of this approach, with reliability criteria explicitly expressed in terms of the system's performance in the process after testing, and using only observable random quantities (mostly the number of failures per type of task), which makes this an attractive solution to BRD.

In [19] a similar approach, yet again more suitable for catastrophic failures, is considered in which reliability of a system, to be demonstrated, is controlled both by tests and decisions on in-built system redundancy. This also considers minimisation of total expected costs, taking into account the costs of additional redundancy which of course enables fewer zero-failure tests in order to demonstrate a required level of reliability, the latter typically included as a constraint for the optimisation problem. We refer to [19] for further details, Example 2 in Section 4 shows the potential of this approach.

The important considerations of the role and choice of the prior distribution is similar for all the different predictive BRD approaches mentioned here, and is commented on in the next section.

3 The prior distribution

A main concern in Bayesian statistics is the choice of the prior distribution. In many applications, if sufficient data are observed then the inferences based on the posterior distribution are pretty robust with regard to the choice of the prior distribution. In the zero-failure BRD setting, however, the posterior distribution remains highly sensitive to some changes in the prior distribution. Hence, it is not straightforward to use so-called 'non-informative' priors [1] for such zero-failure testing [6, 7].

Design of experiments within the Bayesian framework is also strongly influenced by the prior distribution used [1]. It is usually advocated that the main role of the prior distribution is to take subjective information into account. From this perspective, the main aim of design of experiments, in Bayesian statistics, is for the person whose beliefs are modelled, to learn in some optimal manner from the data resulting from the experiment [3]. In case of testing for high reliability, one would often have quite strong prior beliefs about the quality of the system, in the sense that one would already expect the system to be highly reliable, such that one would be somehow surprised if there were still one or more failures observed during testing. If one takes such beliefs into account via a prior distribution, then this will lead to relatively few zero-failure tests being required in order to meet a reliability criterion. Using one's prior beliefs as such, the intention of the tests would be to learn, in a way that is dependent on the subjective input. In the extreme case, if one is already so confident about the system's reliability that one judges testing unnecessary, one can choose a corresponding prior distribution such that indeed no further testing is required. Of course, this may not convince others. A similar difference between personal learning and convincing others, affects general Bayesian design of experiments and the role of randomization [2].

We propose that the prior distribution used for Bayesian reliability *demonstration*, plays a different, non-subjective, role, as reliability demonstration is not normally intended to convince only the person whose beliefs are modelled, of the reliability of a system, but to satisfy externally imposed rules or to convince others of the system's reliability. From this perspective, the prior distribution can be regarded as reflecting neutral, or even pessimistic, prior information, which the test data must counter in order to *demonstrate* reliability.

From this perspective of reliability demonstration, we advocate the use of the Beta prior for θ_i, with hyperparameters $\alpha_i = 1$ and $\beta_i = 0$. The choice of $\beta_i = 0$ is natural, given its interpretation and its effect on the required number of zero-failure tests [7]. An appropriate value of α_i is more open to discussion, and this choice has great impact on the number of zero-failure tests required. With the interpretation as 'having seen α_i tests failing' before testing, our test criterion demands sufficiently many zero-failure tests, such that, on the basis of the combined information represented by this prior distribution and the test results, the reliability requirement is satisfied. The choice $\alpha_i = 1$ seems reasonable from this perspective, as it asks for sufficient zero-failure tests to outweigh one suggested earlier failure. We consider $\alpha_i = 1$ a fairly conservative choice, in the sense that many zero-failure tests are required to demonstrate high reliability. Hartigan [10] suggested the same prior distribution for similar conservative inferences. A further argument in favour of this choice of α_i and β_i results from analysis of this same problem from a different foundational perspective [5].

4 Examples

Example 1.
This example illustrates the BRD approach from [7], with failures assumed to be catastrophic. It considers the optimal test numbers for a system that has to perform $k = 4$ types of tasks, using prior distributions with $\alpha_i = 1$ and $\beta_i = 0$ for all $i = 1, \ldots, 4$. We aim at minimisation of the total number of tests, assuming they reveal no failures, such that the resulting predictive probability (2) of zero failures in the process is at least p. For the deterministic case, let $\underline{m} = (1, 2, 4, 9)$. The corresponding optimal test numbers are given in Table 1.

p	n_1	n_2	n_3	n_4
0.90	70	98	139	207
0.95	144	204	287	429
0.99	737	1042	1474	2209
0.995	1479	2091	2956	4433
0.999	9457	10553	13943	21505

Table 1: Optimal test numbers.

This illustrates that very high reliability can only be *demonstrated* by very many zero-failure tests. For this setting, if tasks are assumed to arrive according to Poisson processes, with their expected numbers equal to the m_i used above, then the numbers of zero-failure tests required are indeed nearly identical to those in Table 1. If we increase the m_i, then the required test numbers increase by about the same factor, for example optimal testing for $\underline{m} = (10, 20, 40, 90)$, and $p = 0.90$, is achieved by taking $\underline{n} = (699, 985, 1387, 2067)$.

Example 2.

In this example we briefly illustrate the optimal zero-failure test numbers and numbers of components, for systems with redundancy [19]. We consider a 2-out-of-y system, so for the system to function a minimum of 2 out of y components must function, and we assume components to function independently. We only consider the deterministic case, with the numbers of tasks of $k = 3$ independent types required in the process after testing equal to 1, 3 and 6, respectively. The optimal BRD zero-failure test numbers and optimal number of components y are presented in Table 2 for several cases. The cost per component installed is Q, C is the cost incurred if the system does not deal successfully with all 10 tasks in the process after testing, and $\underline{c}$ represents the costs of one test per type of task. Again we use Beta prior distributions with $\alpha_i = 1$ and $\beta_i = 0$ for all types of tasks, and the optimisation criterion chosen is minimal expected overall costs, so for testing and the system's functioning during the process considered after testing, with the constraint that the posterior predictive probability of zero failures in those 10 tasks after testing should be at least $p = 0.95$. The cases considered are:

(1) $\underline{c} = (1, 1, 1)$, $C = 10,000$
(2) $\underline{c} = (20, 50, 50)$, $C = 10,000$
(3) $\underline{c} = (20, 50, 50)$, $C = 1,000,000$

	$\underline{n}$, y		
Case(s)	$Q = 300$	$Q = 1,000$	$Q = 3,000$
(1)	$(47, 68, 86)$, 3	$(47, 68, 86)$, 3	$(201, 347, 489)$, 3
(2)	$(8, 9, 10)$, 5	$(12, 12, 15)$, 4	$(26, 28, 35)$, 3
(3)	$(14, 14, 16)$, 8	$(20, 21, 23)$, 6	$(41, 43, 51)$, 4

Table 2: Optimal numbers of zero-failure tests ($\underline{n}$) and components (y).

Obviously, higher built-in redundancy (larger y) requires fewer zero-failure tests. Cases (1) and (2) in Table 2 illustrate the fact that increasing testing cost c_i per test of type i, reduces the optimal number of zero-failure tests n_i, and may increase the optimal y. For Case (2) with $Q = 1,000$, Table 2 shows that the optimal solution is to install $y = 4$ components with $\underline{n} = (12, 12, 15)$ zero-failure tests. If we increase the cost per component to $Q = 3,000$ in the same setting, the optimal solution for Case (2) is

$y = 3$ components with $\underline{n} = (26, 28, 35)$ tests. If we consider a 2-out-of-2 system instead (not presented in Table 2), the optimal zero-failure test numbers, in this case, are $(296, 322, 454)$. This illustrates that the option of building in redundancy can greatly reduce the test requirements and the corresponding expected costs. Cases (2) and (3) illustrate that increasing process failure cost C requires an increased number of components and/or an increased number of zero-failure tests to minimise the total costs and to demonstrate the required reliability level.

5 Summary and related topics

Bayesian reliability demonstration is a topic of great interest, both for applications as from the perspective of foundations of statistics, the latter due to wide variety of possible reliability criteria and the non-standard role prior distributions play when test data are required to *demonstrate* reliability. The predictive approach to BRD, recently presented by the current authors and discussed above, is promising in its flexibility with regard to reliability criteria used and the possible inclusion of constraints in the optimisation problems considered to take practical considerations into account [18]. As this approach has so far only been developed for attribute testing, there are important challenges for a similar way for BRD involving (continuous) time random quantities. This is particularly challenging in zero-failure situations, where observations would only consist of observed periods without failures, so all observations would be right-censored (see [4] for a nonparametric predictive approach to a related problem in probabilistic safety assessment). There are many topics in statistics, quality control and reliability, which consider problems that are closely related to BRD, for example software reliability verification and testing [12, 20, 21] and acceptance sampling, as well as topics such as fault removal and re-testing, and accelerated life testing [17], and linking these to BRD for large-scale systems and developing engineering projects is a major challenge, so BRD promises to offer exciting opportunities for research and applications over the next decades.

References

[1] Bernardo, J.M. and Smith, A.F.M. (1994). *Bayesian Theory.* Wiley, Chichester.

[2] Berry, S.M. and Kadane, J.B. (1997). Optimal Bayesian randomization. *Journal of the Royal Statistical Society, Series B*, **59**, 813-819.

[3] Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: a review. *Statistical Science*, **10**, 273-304.

[4] Coolen, F.P.A. (2006). On probabilistic safety assessment in case of zero failures. *Proceedings of the Institute of Mechanical Engineers, Part O: Journal of Risk and Reliability*, to appear.

[5] Coolen, F.P.A. and Coolen-Schrijner, P. (2005). Nonparametric predictive reliability demonstration for failure-free periods. *IMA Journal of Management Mathematics*, **16**, 1-11.

[6] Coolen, F.P.A. and Coolen-Schrijner, P. (2006). On zero-failure testing for Bayesian high reliability demonstration. *Proceedings of the Institute of Mechanical Engineers, Part O: Journal of Risk and Reliability*, **220**, 35-44.

[7] Coolen, F.P.A., Coolen-Schrijner, P., and Rahrouh, M. (2005). Bayesian reliability demonstration for failure-free periods. *Reliability Engineering and System Safety*, **88**, 81-91.

[8] Coolen, F.P.A., Coolen-Schrijner, P. and Rahrouh, M. (2006). Bayesian reliability demonstration with multiple independent tasks. *IMA Journal of Management Mathematics*, **17**, 131-142.

[9] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2003). *Bayesian Data Analysis* (2nd edition). Chapman & Hall, London.

[10] Hartigan, J.A. (1983). *Bayes Theory*. Springer, New York.

[11] Higgins, J.J. and Tsokos, C.P. (1976). On the behavior of some quantities used in Bayesian reliability demonstration tests. *IEEE Transactions on Reliability*, **25**, 261-264.

[12] Kuo, L. (1999). Software reliability. *Encyclopedia of Statistical Sciences*, S. Kotz (ed.), update vol. 3, 671-680.

[13] Lu, M.W. and Rudy, R.J. (2001). Reliability demonstration test for a finite population. *Quality and Reliability Engineering International*, **17**, 33-38.

[14] Lu, M.W. and Rudy, R.J. (2001). Laboratory reliability demonstration test considerations. *IEEE Transactions on Reliability*, **50**, 12-16.

[15] Martz, H.F. and Waller, R.A. (1979). A Bayesian zero-failure (BAZE) reliability demonstration testing procedure. *Journal of Quality Technology*, **11**, 128-138.

[16] Martz, H.F. and Waller, R.A. (1982). *Bayesian Reliability Analysis*. Wiley, New York.

[17] Meeker, W.Q. and Escobar, L.A. (1998). *Statistical Methods for Reliability Data*. Wiley, New York.

[18] O'Connor, P.D.T. (2001). *Test Engineering: a concise guide to cost-effective design, development, and manufacture*. Wiley, Chichester.

[19] Rahrouh, M., Coolen, F.P.A. and Coolen-Schrijner, P. (2006). Bayesian reliability demonstration for systems with redundancy. *Proceedings of the Institute of Mechanical Engineers, Part O: Journal of Risk and Reliability*, to appear.

[20] Singpurwalla, N.D. and Wilson, S.P. (1999). *Statistical Methods in Software Engineering*. Springer, New York.

[21] Wooff, D.A., Goldstein, M. and Coolen, F.P.A. (2002). Bayesian graphical models for software testing. *IEEE Transactions on Software Engineering*, **28**, 510-525.

[22] Yadav, O.P., Singh, N. and Goel, P.S. (2006). Reliability demonstration test planning: a three dimensional consideration. *Reliability Engineering and System Safety*, **91**, 882-893.