

Reliability Assessment of Legacy Safety-Critical Systems Upgraded with Off-the-Shelf Components

Peter Popov

Centre for Software Reliability, City University,
Northampton Square, London, UK
e-mail: ptp@csr.city.ac.uk

Abstract. Reliability assessment of upgraded legacy systems is an important problem in many safety-related industries. Some parts of the equipment used in the original design of such systems are either not available off-the-shelf (OTS) or have become extremely expensive as a result of being discontinued as mass production components. Maintaining a legacy system, therefore, demands using different OTS components. Trustworthy reliability assurance after an upgrade with a new OTS component is needed which combines the evidence about the reliability of the new OTS component with the knowledge about the old system accumulated to date. In these circumstances Bayesian approach to reliability assessment is invaluable. Earlier studies have used Bayesian inference under simplifying assumptions. Here we study the effect of these on the accuracy of predictions and discuss the problems, some of them open for future research, of using Bayesian inference for practical reliability assessment.

1. Introduction

The use of off-the-shelf (OTS) components with software becomes a practice increasingly widespread for both development of new systems and upgrading existing (i.e. legacy) systems as part of their maintenance. The main reason for the trend is the low cost of the OTS components compared with a bespoke development or older components being discontinued as mass production units.

In this paper we focus on reliability assessment of a legacy system upgraded with an OTS component which contains software. Two factors make the reliability assessment in this case significantly different from the assessment of a bespoke system. First, reliability data about the OTS component, if available at all, comes, as a rule, for an unknown, possibly different from the target, environment. The evidence of high reliability in different environment will give modest confidence in the reliability of the OTS component in the target environment. Second, acceptance testing of the upgraded system must be, as a rule, short. In some cases postponing the deployment of an upgraded system to undertake a long V&V procedure will simply prevent from gaining market advantage. In some other cases, e.g. upgrading a nuclear plant with smart sensors, it is simply prohibitively expensive or even impossible to run a long acceptance testing on the upgraded system before it is deployed. And yet in many

cases, e.g. in safety critical systems, there are very stringent requirements for demonstrably high reliability of systems in which OTS components are used. In these circumstances Bayesian approach to reliability assessment is very useful. It allows one to combine rigorously both, the *a priori* knowledge about the reliability of a system and its components, and the new (possibly very limited) evidence coming from observing the upgraded system in operation.

The simplest way to assess the reliability of a system is to observe its failure behaviour in (real or simulated) operation. If we treat the system as a black box, i.e. ignore the internal structure of the system, standard techniques of statistical inference can be applied to estimate its *probability of failure on demand (pfd)* on the basis of the amount of realistic testing performed and the number of failures observed. However, this ‘black-box’ approach to reliability assessment has severe limitations [1], [2]: if we want to demonstrate very small upper bounds on the *pfd*, the amount of testing required becomes very expensive and then infeasible. It is then natural to ask whether we can use the additional knowledge about the structure of the system to reduce this problem - to achieve better confidence for the same amount of testing. This is the problem which we address in this paper. We present a model of reliability assessment of a legacy system upgraded with a single OTS component and discuss the difficulties and limitations of its practical use.

In detail, section 2 presents the problem studied in the paper, in section 3 the main result is presented. In section 4 we discuss the implications of our results and the difficulties in applying the Bayesian approach to practical reliability assessment, some of them as open research questions. Finally, conclusions are presented in section 5.

2. The Problem

For simplicity we assume that the system under consideration is an *on-demand system*, i.e. it is called upon when certain predefined circumstances occur in the environment. A typical example of an on-demand system is a safety protection system intended to shut down a plant if the plant leaves its safety envelope.

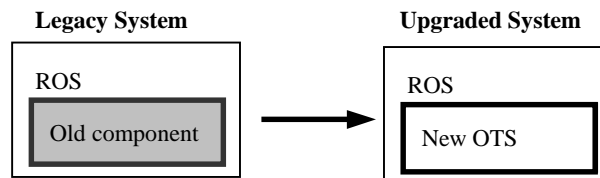


Fig. 1 Upgrading a legacy system with an OTS component.

We analyse the simplest possible case of system upgrade – the replacement of a *single* component with an OTS component which interacts with the *rest of a legacy system* (ROS), as illustrated in Fig. 1. In the rest of the paper we refer to ROS as sub-system A and to the new OTS component as sub-system B. The paper analyses a special case of a system in which *both sub-systems are used exactly once per demand*.

3. Reliability Assessment: Bayesian Approach

Bayesian approach to reliability assessment of an upgraded on-demand system is used. The *probability of failure on demand (pfd)* is the measure of interest.

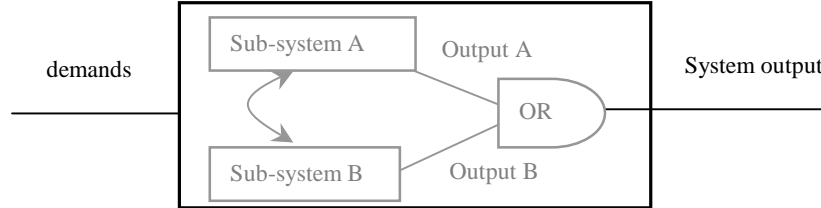


Fig. 2. Black-box model of a system. The internal structure of the system is unknown. The outputs of the sub-systems are not used in the inference. Only the system output is recorded on each demand and fed into the inference.

If the system is treated as a black box, i.e. we can only distinguish between *system* failures or successes (Fig. 2), the inference proceeds as follows. Denoting the system *pfd* as p , the posterior distribution of p after seeing r failures in n demands is:

$$f_p(x | r, n) \propto L(n, r | x) f_p(x), \quad (1)$$

where $f_p(\bullet)$ is the prior distribution of p , which represents the assessor's belief about p , before seeing the result of the test on n demands. $L(n, r | x)$ is the *likelihood* of observing r failures in n demands if the *pfd* were exactly x . This is given in this case (of independent demands) by the *binomial* distribution, $L(n, r | x) = \binom{n}{r} x^r (1-x)^{n-r}$.

The details of system behaviour which may be available but are ignored in the black-box model, such as the outcomes of the sub-systems which make up the system on a demand, are taken into account in the clear-box model. As a rule the predictions obtained with the clear-box and black-box models differ. We have shown elsewhere, in a specific context of a parallel system [3], that the black-box predictions can be over-optimistic or over-pessimistic and the sign of the error cannot be known in advance – it depends on the prior and the observations.

The Bayesian inference with a clear-box model is more complex than with the black-box model because a multivariate prior distribution and likelihood are used. The dimensions of the prior distribution depend on the number of sub-systems which make up the system and whether the system score (a success or a failure) is a deterministic or a non-deterministic function of the scores of the sub-systems involved¹. For instance, modelling a system with two sub-systems (e.g. Fig. 1) and a deterministic system score as a clear box requires a 3-variate prior distribution/likelihood. A clear–

¹ Examples of systems with deterministic system score are parallel systems [3] and serial systems (i.e. which fail if at least one of their sub-systems fails). An example of a system with a non-deterministic system score is the system in Fig. 1 if for the same sub-system scores (e.g. ROS fails, OTS component succeeds) it fails on some demands but succeeds on others. Non-deterministic system score must be explicitly modelled as a separate binary random variable.

box model of a system with the same number of sub-systems but a non-deterministic system score requires a 7-variate prior and likelihood. A clear box of a system with 3 sub-systems with a deterministic and a non-deterministic system score requires 7- or 15-variate distribution/likelihood, respectively, etc. Such an exponential explosion of complexity of the prior/likelihood with the increase of the number of the sub-systems poses two difficulties for using a clear-box Bayesian inference:

- Defining a multidimensional prior is difficult. A phenomenon is widely reported that humans are not very good at using probabilities [4]. Increasing dimensions of the prior distribution makes it very difficult for an assessor to justify a particular prior to match their informal belief about the reliability of the system and its sub-systems;
- Unless the prior and the likelihood form a conjugate family [5] the complexity of the Bayesian inference itself increases with the number of dimensions of the prior used in the inference because multiple integrals must be calculated.

These two difficulties are a good incentive for one to try to simplify the multivariate prior distribution. One way of simplification is assuming various forms of independence between the variates of the distribution used which describe the assessor's knowledge (belief) about system reliability. Recently Kuball et al. [6] in *different context* used the assumption of independence between the failures of the subsystems, which is attractive. It allows one to avoid the difficulties in defining the *dependencies* that may exist between several failure processes, the most difficult part in defining a multivariate distribution. Once the failures of the subsystems are assumed independent, however, they will stay so despite what is observed in operation, even if overwhelming evidence is received that the sub-system failures are correlated (positively or negatively). This evidence of failure dependence is simply ignored; the only uncertainty affected by the inference is that associated with the *pdf* of the sub-systems. The multivariate Bayesian inference collapses to a set of univariate inferences, which are easily tractable. Kuball et al. assert that the predictions derived under the assumption of independence will be pessimistic at least in the case that *no failure* is observed. Even if this is true there is no guarantee that 'no failure' will be the only outcome to observe in operation, e.g. during acceptance testing after the upgrade. The justification that the 'no failure' case is the *only one of interest* for the assessor (since any other outcome would imply restarting the acceptance testing afresh) is known to have a problem. Littlewood and Wright have shown [7] that ignoring the previous observations (i.e. rounds of acceptance testing which ended with failures) can produce overoptimistic predictions. It is worth, therefore, studying the consequences of assuming independence between the failures of the sub-systems for a broader range of observations, not just for the 'no failure' case.

3.1 Experimental Setup

Now we formally describe the clear-box model of the system (Fig 1). The sub-systems A and B are assumed imperfect and their probabilities of failure - uncertain. The scores of the sub-systems, which can be observed on a randomly chosen demand, are summarised in Table 1.

Table 1. The combinations of sub-system scores which can be observed on a randomly chosen demand are shown in columns 1-2. The notations used for the probabilities of these combinations are shown in column 3. The number of times the score combinations are observed in N trials, r_0 , r_1 , r_2 and r_3 ($N = r_0 + r_1 + r_2 + r_3$) respectively, are shown in the last column

Sub-system scores			
Sub-system A	Sub-system B	Probability	Observed in N demands
0	0	p_{00}	r_0
0	1	p_{01}	r_1
1	0	p_{10}	r_2
1	1	p_{11}	r_3

Clearly, the probabilities of failure of sub-systems A and B, p_A and p_B , respectively, can be expressed as:

$$p_A = p_{10} + p_{11} \quad \text{and} \quad p_B = p_{01} + p_{11}.$$

p_{11} represents the probability of coincident failure of both sub-systems, A and B, on the same demand and hence the notation $p_{AB} \equiv p_{11}$ captures better the intuitive meaning of the event it is assigned to. The joint distribution $f_{p_A, p_B, p_{AB}}(\bullet, \bullet, \bullet)$ describes completely the a priori knowledge of an assessor about the reliability of the upgraded system. It can be shown that for a given observation (r_1 , r_2 , and r_3 in N demands) the posterior distribution can be calculated as:

$$f_{p_A, p_B, p_{AB}}(x, y, z | N, r_1, r_2, r_3) = \frac{f_{p_A, p_B, p_{AB}}(x, y, z) L(N, r_1, r_2, r_3 | p_A, p_B, p_{AB})}{\iint\limits_{p_A, p_B, p_{AB}} f_{p_A, p_B, p_{AB}}(x, y, z) L(N, r_1, r_2, r_3 | p_A, p_B, p_{AB}) dx dy dz} \quad (2)$$

$$L(N, r_1, r_2, r_3 | p_A, p_B, p_{AB}) = \frac{N!}{r_1! r_2! r_3! (N - r_1 - r_2 - r_3)!} (p_A - p_{AB})^{r_2} (p_B - p_{AB})^{r_1} p_{AB}^{r_3} (1 - p_A - p_B + p_{AB})^{N - r_1 - r_2 - r_3}$$

is the likelihood of the observation.

Up to this point the inference will be the same no matter how the event 'system failure' is defined but calculating the marginal distribution of system *pdf*, P_S , is affected by how the event 'system failure' is defined. We proceed as follows:

1. A *serial system*: a failure of either of the sub-systems leads to a system failure. The posterior distribution, $f_{p_A, p_B, p_{AB}}(\bullet, \bullet, \bullet | N, r_1, r_2, r_3)$, must be transformed to a new distribution, $f_{p_A, p_B, p_S}(\bullet, \bullet, \bullet | N, r_1, r_2, r_3)$, where P_S is defined as: $P_S = P_A + P_B - P_{AB}$, from which the marginal distribution of P_S , $f_{p_S}(\bullet | N, r_1, r_2, r_3)$, will be calculated by integrating out the nuisance parameters P_A and P_B . If the system is treated as a black-box (Fig. 2) the system *pdf* can be inferred using formula (1) above. The marginal prior distribution of P_S , $f_{p_S}(\bullet)$, and a binomial likelihood of observing $r_1 + r_2 + r_3$ system failures in N trials will be used. If the failures of the sub-systems A and B are assumed independent then for any values of P_A and P_B the probability of joint failure, P_{AB} , of both sub-systems is $P_{AB} = P_A P_B$. Formally, the joint distribution can be expressed as:

$$f_{P_A, P_B, P_{AB}}^*(x, y, z) = \begin{cases} f_{P_A}(x)f_{P_B}(y)\delta(xy), & \text{if } z = xy \\ 0, & \text{if } z \neq xy \end{cases}$$

The failures of the two sub-systems remain independent in the posterior:

$$f_{P_A, P_B, P_{AB}}^*(x, y, z | N, r_1, r_2, r_3) = \begin{cases} f_{P_A}^*(x | N, r_1 + r_2)f_{P_B}^*(y | N, r_1 + r_3)\delta(xy), & \text{if } z = xy \\ 0, & \text{if } z \neq xy \end{cases}$$

$f_{P_A}^*(\bullet | N, r_1 + r_2)$ and $f_{P_B}^*(\bullet | N, r_2 + r_3)$ are the marginal posterior distributions of sub-systems A and B, respectively, inferred under independence. The inference for sub-system A proceeds according to (1) using the marginal prior of sub-system A, $f_{P_A}(\bullet)$, and binomial likelihood of observing $r_2 + r_3$ failures of sub-system A in N trials. Similarly, $f_{P_B}^*(\bullet | N, r_1 + r_3)$ is derived with prior $f_{P_B}(\bullet)$ and binomial likelihood of observing $r_1 + r_3$ failures of sub-system B in N trials. The posterior marginal distribution of system *pdf*, $f_{P_S}^*(\bullet | N, r_1, r_2, r_3)$, can be obtained from $f_{P_A, P_B, P_{AB}}^*(x, y, z | N, r_1, r_2, r_3)$ as described above: first the joint posterior is transformed to a form which contains P_S as a variate of the joint distribution and then P_A and P_B are integrated out.

2. The *system fails when sub-system A fails*. In this case the probability of system failure is merely the posterior *pdf* of sub-system A (ROS). The marginal distribution, $f_{P_A}(\bullet | N, r_1, r_2, r_3)$, can be calculated from $f_{P_A, P_B, P_{AB}}(\bullet, \bullet, \bullet | N, r_1, r_2, r_3)$ by integrating out P_B and P_{AB} . With black-box inference another marginal posterior can be obtained, $f_{P_A}^*(\bullet | N, r_2 + r_3)$, using (1) with the marginal prior of *pdf* of sub-system A, $f_{P_A}(\bullet)$, and binomial likelihood of observing $r_2 + r_3$ failures of sub-system A in N trials. Notice that the marginal distribution $f_{P_A}(\bullet | N, r_1, r_2, r_3)$ is different, as a rule, from the marginal distribution $f_{P_A}^*(\bullet | N, r_2 + r_3)$, obtained with the black-box inference.

3.2. Numerical Example

Two numerical examples are presented below which illustrate the effect of various simplifying assumptions used in the inference on the accuracy of the predictions.

The prior, $f_{P_A, P_B, P_{AB}}(\bullet, \bullet, \bullet)$, was constructed under the assumption that $f_{P_A}(\bullet)$ and $f_{P_B}(\bullet)$ are both Beta distributions, $B(\bullet, a, b)$, in the interval $[0, 0.01]$ and are independent of each other, i.e. $f_{P_A, P_B}(\bullet, \bullet) = f_{P_A}(\bullet)f_{P_B}(\bullet)$. The parameters a and b for the two distributions were chosen as follows: $a_A = 2$, $b_A = 2$ for sub-system A and $a_B = 3$, $b_B = 3$ for sub-system B.

If the sub-systems are assumed to fail independently the parameters above are a sufficient definition of the prior distribution.

If the sub-systems are not assumed to fail independently we specify the conditional distributions, $f_{P_{AB}|P_B, P_A}(\bullet | P_A, P_B)$, for every pair of values of P_A and P_B , as Beta

distributions, $B(\bullet, a, b)$ in the range $[0, \min(P_A, P_B)]$ with parameters $a_{AB} = 5$, $b_{AB} = 5$ which complete the definition of the trivariate distribution, $f_{P_A, P_B, P_{AB}}(\bullet, \bullet, \bullet)$.

We do not make any claims that the priors used in the examples should be used in practical assessment. They serve illustrative purposes only and yet, have been chosen from a reasonable range. Each of the sub-systems, for instance, has an average *pdf* of $5 \cdot 10^{-3}$, which is a value from a typical range for many applications.

Two sets of observations were used for the calculations with the same number of trials, $N = 4000$:

- Observation 1 (The sub-systems never failed together): $r_3 = 0$, $r_1 = r_2 = 20$;
- Observation 2 (Sub-systems always fail together): $r_3 = 20$, $r_2 = r_3 = 0$.

The number of failures of the sub-systems has been chosen so that they are indistinguishable under the assumption of failure independence – in both cases each of the sub-systems failed 20 times². The observations, however, provide evidence of different correlation between the failures of the sub-systems: in the first observation - of strong positive - while in the second observation - of strong negative correlation.

The inference results under various assumptions for both observations are summarised in Table 2 and 3, respectively, which show the percentiles of the marginal prior/posterior distributions of system *pdf*:

Table 2. Observation 1: Strong negative correlation between the failures of the sub-systems ($N = 4000$, $r_3 = 0$, $r_1 = r_2 = 20$). The upper part of the table shows the posteriors if the upgraded system were a ‘serial’ system while the lower part of the table shows the posteriors if the system failed only when sub-system A failed.

	50 %	75%	90%	95%	99%
Serial system					
prior system <i>pdf</i> , $f_{p_S}(\bullet)$	0.0079	0.0096	0.0114	0.0124	0.0144
‘proper’ posterior <i>pdf</i> , $f_{p_S}(\bullet N, r_1, r_2, r_3)$	0.01	0.0118	0.012	0.0126	0.0137
Posterior <i>pdf</i> with independence, $f_{p_S}^*(\bullet N, r_1, r_2, r_3)$	0.0103	0.0112	0.0122	0.0128	0.01393
Black-box posterior with independence	0.01	0.011	0.012	0.0125	0.0136
Black-box posterior without independence	0.0095	0.01035	0.0113	0.0118	0.0128
Failure of sub-system A (ROS) only leads to a system failure					
Prior system <i>pdf</i> , $f_{p_A}(\bullet)$	0.0049	0.0066	0.0080	0.0086	0.0093
‘proper’ posterior <i>pdf</i> , $f_{p_A}(\bullet N, r_1, r_2, r_3)$	0.0051	0.0059	0.0066	0.0071	0.0079
Posterior system <i>pdf</i> with independence, $f_{p_A}^*(\bullet N, r_1 + r_2)$	0.005	0.0058	0.0065	0.0069	0.0078

² equal to the expected number of failures of each of the sub-system in 4000 demands as defined by the prior.

The results in Table 2 reveal that the black-box inference produces optimistic posteriors: there is stochastic ordering between the posteriors obtained with the clear-box model, no matter whether independence of failures is assumed or not. Comparing the clear-box predictions with and without failure independence reveals another stochastic ordering: the predictions with independence are conservative. This is in line with the result by Kuball et al. The differences between the various posteriors are minimal. The tendency remains the same (the same ordering between the posteriors was observed) for a wide range of observations with negative correlation between the failures of the sub-systems. Finally, for a non-serial system the independence produces more optimistic predictions than without independence (the last two rows of the table). In other words, the independence is not guaranteed to produce conservative predictions – the ordering depends on how the event ‘system failure’ is defined.

Table 3. Observation 2: Strong positive correlation between the failures of the two sub-systems ($N = 4000$, $r_3 = 20$, $r_1 = r_2 = 0$). The same arrangement of the results is given as in Table 2

	50 %	75%	90%	95%	99%
Serial system					
Prior system pdf , $f_{p_S}(\bullet)$	0.0079	0.0096	0.0114	0.0124	0.0144
‘proper’ posterior pdf , $f_{p_S}(\bullet N, r_1, r_2, r_3)$	0.0051	0.0058	0.0065	0.0069	0.0076
Posterior pdf with independence, $f_{p_S}^*(\bullet N, r_1, r_2, r_3)$	0.0103	0.0113	0.0123	0.0128	0.0139
Black-box posterior with independence	0.0055	0.0063	0.0071	0.0075	0.0084
Black-box posterior without independence	0.0059	0.0066	0.0073	0.0078	0.0088
Failure of sub-system A (ROS) only leads to a system failure					
Prior system pdf , $f_{p_A}(\bullet)$	0.0049	0.0066	0.0080	0.0086	0.0093
‘proper’ posterior pdf , $f_{p_A}(\bullet N, r_1, r_2, r_3)$	0.00495	0.0056	0.0063	0.0066	0.0074
Posterior system pdf with independence, $f_{p_A}^*(\bullet N, r_1 + r_2)$	0.005	0.0058	0.0065	0.0069	0.0078

The results from the black-box inference in Table 3 reveal a pattern different from Table 2. Black-box predictions here are more pessimistic than the ‘proper’ posteriors, i.e. with clear-box without assuming independence. This is true for both the serial system and the system which only fails when sub-system A fails. The fact that the sign of the error of the black-box predictions changes (from over-estimation in Table 2 to underestimation in Table 3) is not surprising, it is in line with our result for parallel systems, [3]. If we compare the clear-box predictions – with and without independence – the same stochastic ordering is observed no matter how the event ‘system failure’ is defined. If the system failure is equivalent to a failure of sub-system A, the predictions with independence (i.e. if the effect of sub-system B on sub-system

A is neglected) are more pessimistic than the predictions without independence. In other words the ordering for this type of system is the opposite to what we saw in Table 2. For serial systems, the predictions shown in Table 3 obtained with independence are, as in Table 2 - more pessimistic than without independence. The pessimism, however, in this case is much more significant than it was in Table 2.

The values predicted under independence are almost twice the values without independence: the conservatism for a serial system may become significant. With the second observation (Table 3) the assumption of statistical independence of the failures of the sub-systems is clearly unrealistic! If the independence were true the expected number of joint failures is 0.1 failure in 4000 trials, while 20 were actually observed!

4. Discussion

The results presented here are hardly surprising! They simply confirm that simplifications of models or model parameterisation may lead to errors. If the errors caused by the simplifications were negligible or at least consistently conservative, reporting on them would not have made much sense. What seems worrying, however, and therefore we believe worth pointing out, is that the errors are neither guaranteed to be always negligible nor consistently conservative. Simplifying the model and using black-box inference may lead to over- or under-estimation of system reliability. We reported elsewhere on this phenomenon with respect to a parallel system. Here we present similar results for alternative system architectures. One seems justified in concluding that using a more detailed clear-box model always pays off by making the predictions more accurate. In some cases, the accuracy may imply less effort on demonstrating having reached a reliability target, i.e. makes the V&V less expensive. In some other cases, it prevents from over-confidence in system reliability although at the expense of longer acceptance testing. In the particular context of this study – reliability assessment of an upgraded system – there is another angle of why clear-box must be preferred to black-box. We would like to reuse the evidence available to date in the assessment of the upgraded system. This evidence, if available at all, is given for the sub-systems: for sub-system A and, possibly, for sub-system B but not for the upgraded system as a whole. Using the evidence available seems only possible by first modelling the system explicitly as a clear box and plugging-in the pieces of evidence into the definition of the joint prior. From the multivariate prior, the marginal prior distribution of system *pdf* can be derived and used in a marginal Bayesian inference. It does not seem very sensible, however, to use the marginal inference after carrying out the hard work of identifying the plausible multivariate prior. The gain will be minimal, only in terms of making the inference itself easier, too little of a gain at the expense of inaccurate predictions.

The results with the simplified clear-box inference seem particularly worth articulating. Our two examples indicated conservative predictions obtained for a serial system under the independence assumption. One may think that this is universally true for serial systems as Kuball et al. asserted. Unfortunately, this is not the case! We have found examples of priors/observations when the conservatism does not hold. An

example of such a set of prior/observation is the following: $f_{P_A}(\bullet)$ and $f_{P_B}(\bullet)$ assumed independent Beta distributions in the range $[0,1]$ with parameters $a_A = 2, b_A = 20$, for $a_A = 2, b_A = 2$. The conditional distribution, $f_{P_{AB}|P_B, P_A}(\bullet | P_A, P_B)$, for a pair of values P_A and P_B assumed to be a Beta distributions in the range $[0, \min(P_A, P_B)]$ with parameters $a_{AB} = 5, b_{AB} = 5$, observations: $N = 40, r_1 = 0, r_2 = 12, r_3 = 12$. In this case the posterior system *pdf* under the assumption of independence is *more optimistic* than the posterior without independence. The point with this counterexample is that there exist cases in which the assumption of independence *may lead to over-optimism*. Since the exact conditions are unknown under which the predictions become over-optimistic assuming independence between the failures of the sub-systems may be dangerous: it may lead to unacceptable errors such as overconfidence in achieved system reliability.

A second problem with the independence assumption exists that is that even when the predictions under this assumption are conservative, the level of conservatism may be significant which is expensive. This tendency seems to escalate with the increase of the number of sub-systems. In the worse case it seems that the level of conservatism in the predicted system reliability is proportional to the number of sub-systems used in the model. For a system of 10 sub-systems, for example, the worst case underestimation of system reliability can reach an order of magnitude. The implications are that by using the predictions based on the independence assumption the assessor may insist on *unnecessary long acceptance testing until unnecessary conservative targets are met*. We call them unnecessary because the conservatism is merely due to the error caused by the independence assumption.

Using the independence assumption in Bayesian inference is in a sense ironic because it is against the Bayesian spirit to let data ‘speak for itself’. Even if the observations provide an overwhelming evidence of dependence between the failures of the sub-systems, the strong assumption of independence precludes from taking this knowledge into account. In the posteriors the failures of the sub-systems will continue to be modelled as independent processes.

Having pointed out problems with the simplified solutions is not a solution of the problem of reliability assessment of a system made up of sub-systems. The full inference requires a full multivariate prior to be specified which for a system with more than 3 components seems extremely difficult unless a convenient parametric distribution, e.g. a Dirichlet distribution [5], is used, which in turn, is known to be problematic as reported in [3]. In summary, with the current state of knowledge *it does not seem reasonable to go into detailed structural reliability modelling* because of the intrinsic difficulties in specifying the prior without unjustifiable assumptions.

Our example of a system upgraded with an OTS component is ideally suited for a ‘proper’ clear-box Bayesian inference because only a few sub-systems are used in the model. It is applicable if one can justify that the upgrade is ‘perfect’, i.e. there are no (design) faults in integrating the new OTS component with the ROS. If this is the case the following assumptions seem ‘plausible’ in defining the joint prior after the upgrade:

- the marginal distribution of *pdf* of sub-system A (ROS) is available from the observations of the old system before the upgrade.

- the marginal distribution of *pdf* of the OTS component will be, generally, unknown for the new environment (in interaction with sub-system A and the system environment). We should take a "conservative" view here and assume that the new OTS component is *no better* in the new environment than it is reported to have been in other environments. It may be even worth assuming it less reliable than the component it replaces, unless we have very strong evidence to believe otherwise. The strong evidence can only come from the new OTS component being used extensively in an environment *similar* to the environment created for the new OTS component by the system under consideration. The new OTS component may have a very good reliability record in various environments. This, however, cannot be used 'automatically' as *strong evidence* about its reliability in the upgraded system.
- the *pdf* of sub-systems A and B are *independently distributed* (as we assumed in the examples) unless there is evidence to support assuming otherwise. In the latter case, the supporting evidence will have to be used in defining the dependence between the *pdf* of the two sub-systems.
- specifying the *pdf* of joint failure of sub-system A and sub-system B we can use 'indifference' within the range of possible values, but sensitivity analysis is worth applying to detect if 'indifference' leads to gaining high confidence in high system reliability too quickly.

If justifying a 'perfect' upgrade is problematic at least a 7-variate prior distribution must be used to allow for system failures in operation to be accommodated in the inference which are neither failures of ROS nor of the new OTS component. In this case the marginal distributions of the *pdfs* of the two sub-systems, A and B, are the only 'obvious' constraints which can be used in defining the prior. These, however, are likely to be insufficient to set the parameters of the 7-variate joint prior distribution and additional assumptions are needed which may be difficult to justify.

5. Conclusions

We have studied the effect of the model chosen and of a simplifying assumption in parameterising a clear-box model on the accuracy of Bayesian reliability assessment of a system upgraded with a single OTS component. We have shown:

- that simplifications may lead to overestimation or underestimation of system reliability and the sign of the predictions is not known in advance. The simplified inference, therefore, is not worthy recommending for predicting the reliability of safety-critical systems.
- that even when the simplified predictions are conservative, e.g. the predictions for a serial system under the assumption of independence of failures of the sub-systems, they may be too conservative. In the worst case the conservatism is proportional to the number of sub-systems used in the model. This leads to unjustifiably conservative reliability targets achieving which is expensive.
- that detailed clear-box modelling of a system is intrinsically difficult because: i) the full inference without simplifications requires specifying a multivariate prior

which is difficult with more than 3 variates, ii) the simplified inferences (black-box or clear-box with simplifications) have problems with the accuracy of the predictions.

- how the available knowledge about the reliability of the sub-systems before the upgrade can be reused in constructing a multivariate prior when the the upgrade is free of design faults.

The following problems have been identified and are open for further research:

- clear-box inference with the simplifying assumption that the sub-systems fail independently has been shown to lead to over- or underestimation of system reliability. Identifying the conditions under which the simplified clear-box inference produces conservative results remains an open research problem.
- Further studies are needed into multivariate distributions which can be used as prior distributions in a (non-simplified) clear-box Bayesian inference.

Acknowledgement

This work was partially supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under the 'Diversity with Off-the-shelf components (DOTS)' project and the 'Interdisciplinary Research Collaboration in Dependability of Computer-Based Systems (DIRC)'.

References

1. Littlewood, B. and L. Strigini, *Validation of Ultra-High Dependability for Software-based Systems*. Communications of the ACM, 1993. **36**(11): p. 69-80.
2. Butler, R.W. and G.B. Finelli. *The Infeasibility of Experimental Quantification of Life-Critical Software Reliability*. in *ACM SIGSOFT '91 Conference on Software for Critical Systems*, in *ACM SIGSOFT Software Eng. Notes*, Vol. 16 (5). 1991. New Orleans, Louisiana.
3. Littlewood, B., P. Popov, and L. Strigini. *Assessment of the Reliability of Fault-Tolerant Software: a Bayesian Approach*. in *19th International Conference on Computer Safety, Reliability and Security, SAFECOMP'2000*. 2000. Rotterdam, the Netherlands: Springer.
4. Strigini, L., *Engineering judgement in reliability and safety and its limits: what can we learn from research in psychology?* 1994.
<http://www.csr.city.ac.uk/people/lorenzo.strigini/ls.papers/ExpJudgeReport/>
5. Johnson, N.L. and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley Series in Probability and Mathematical Statistics, ed. R.A. Bradley, Hunter, J. S., Kendall, D. G., Watson, G. S. Vol. 4. 1972: John Weley and Sons, INc. 333.
6. Kubal, S., May, J., Hughes, G. *Structural Software Reliability Estimation*. in *SAFECOMP '99, 18th International Conference on Computer Safety, Reliability and Security*. 1999. Toulouse, France: Springer.
7. Littlewood, B. and D. Wright, *Some conservative stopping rules for the operational testing of safety-critical software*. IEEE Transactions on Software Engineering, 1997. **23**(11): p. 673-683.