On Reliability Assessment When a Software-based System Is Replaced by a Thought-to-be-Better One

Bev Littlewood[a], Kizito Salako[a,*], Lorenzo Strigini[a], Xingyu Zhao[a,b]

[a]*The Centre for Software Reliability, School of Mathematics, Computer Science and Engineering,*
City, University of London, Northampton Square EC1V 0HB, United Kingdom
[b]*The Smart System Group, School of Engineering and Physical Sciences,*
Heriot-Watt University, Edinburgh, EH14 4AS, United Kingdom

Abstract

The failure history of pre-existing systems can inform a reliability assessment of a new system. Such assessments – consisting of arguments based on evidence from older systems – are attractive and have been used for quite some time for, typically, mechanical/hardware-only systems. But their application to software-based systems brings some challenges. In this paper, we present a conservative, Bayesian approach to software reliability assessment – one that combines reliability evidence from an old system with an assessor's confidence in a newer system being an improved replacement for the old one. We demonstrate, via different scenarios, what a thought-to-be-better replacement formally means in practice, and what it allows one to believe about actual reliability improvement. The results can be used directly in a reliability assessment, or to caution system stakeholders and industry regulators against using other models that give optimistic assessments. For instance, even if one is certain that some new software must be more reliable than an old product, using the reliability distribution for the old software as a prior distribution when assessing the new system gives optimistic, *not* conservative, predictions for the posterior reliability of the new system after seeing operational testing evidence.

Keywords: software reliability, safety-critical software, reliability assessment, similarity arguments, conservative Bayesian inference, software re-use, globally at least equivalent.

1. Introduction

Assessing the reliability of software-based systems can be challenging, particularly for systems with very stringent reliability requirements [1, 2]. For instance, such systems can require infeasible amounts of operational testing in order to demonstrate that they are sufficiently reliable. Faced with such difficulties, an assessor might turn to using their extensive experience with older "similar" systems, to support any operational testing evidence when assessing a new system. Informally, they may justify this as follows: "Our wealth of experience with good development processes – as evidenced by many (similar) reliable systems that have been built, with lots of operational exposure and very few failures – makes us confident that the new system is also very reliable."

Such arguments certainly have an informal appeal – by combining operational testing of the "new" with extensive reliability evidence from the "old", the new system could be justifiably claimed to be very reliable. Statistical approaches for combining reliability evidence have been employed. For instance, two-stage Bayesian models [3] have been advocated for the assessment of safety-critical systems (e.g. in the nuclear industry [4–6]). The basic idea, depicted in Fig. 1, is as follows. The old and new systems are considered similar, perhaps because of similar design, processes and techniques used in their fabrication/construction. However, their failure rates (i.e. Λ_i) are unknown. Conceptually, this uncertainty in failure rates is captured by some probability distribution over the possible values for the failure rates, where this distribution, itself, has an uncertain shape determined by an unknown hyper-parameter Q. So, the failure rates of the systems are assumed statistically independent and identically distributed (i.i.d.), with each of these rates distributed according to the same, hyper-parameter dependent, prior distribution. And the hyper-parameter is, in turn, determined by some hyper-distribution (i.e. the hyper-prior $P_Q(q)$).

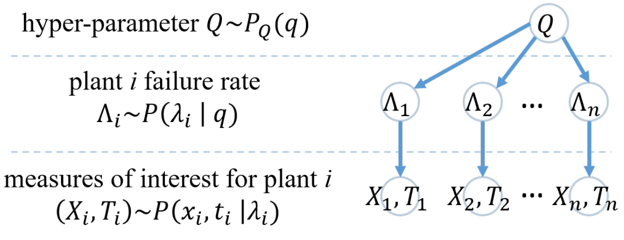
hyper-parameter $Q \sim P_Q(q)$
plant i failure rate
$\Lambda_i \sim P(\lambda_i \mid q)$
measures of interest for plant i
$(X_i, T_i) \sim P(x_i, t_i \mid \lambda_i)$
Q
Λ_1
Λ_2
$\cdots$
Λ_n
X_1, T_1
X_2, T_2
$\cdots$
X_n, T_n

Figure 1: A two-stage Bayesian hierarchical model used in [5].

Thus, Bayesian inference for the hyper-distribution, using failure data from different pre-existing systems, reveals how

[*] Corresponding author: Kizito Salako
Email addresses: k.o.salako@city.ac.uk (Kizito Salako), xingyu.zhao@hw.ac.uk (Xingyu Zhao)

good the commonly adopted development process for the systems is. And any new system should also possess a failure rate that is statistically independent, and distributed according to the same inferred failure rate distribution for the older systems (with the same associated hyper-parameter hyper-distribution). In this way, the failure track-record of pre-existing systems is (indirectly) integrated into the reliability assessment of a new system [7].

But this form of (Bayesian) reliability assessment has two main drawbacks. First, as an application of Bayesian inference, an assessor is required (compelled?) to specify a suitable probability distribution representing their prior beliefs about *every* possible hyper-parameter value. This is difficult enough when the assessment involves reliability evidence from only one system. When assessment involves evidence from two or more systems, the troubles only worsen. How can an assessor, realistically, characterise their beliefs about all of the possible relationships between the reliabilities of multiple systems? Secondly, experience gained from pre-existing systems must be entirely appropriate for making claims about a new system. This is the case for nominally identical systems operating in identical environments. However, it is often hard to adequately demonstrate such similarity in practice, especially for software-based systems. For example, manufacturers' evidence from the use of similar control or safety protection systems in different industries may differ in significant ways, thus undermining the assumption of sufficient similarity amongst these systems.

One common case where systems are, arguably, appropriately similar, is when a *single* bespoke system is replaced by a new, supposedly improved one, but the required functionality and operational environment remain unchanged (e.g. it is typical in the nuclear industry that a safety protection system – of a bespoke design for a given power plant – is replaced by a similarly bespoke system).

Intuitively, one might expect reliability evidence from a single precursor to only provide weak support for claims about a new system's reliability. However, an important consideration is an assessor's confidence that a new system is *not worse than the existing system* (NWTES). Informal NWTES notions are invoked by practitioners in different industries using different terminology. But, surprisingly, none of these notions appear to have been characterised formally, and their implications for assessment have not been studied. For example, under the U.S. Food and Drug Administration's (FDA) 510(k) "premarket notification" process for simplified approval of new medical devices, the new device must be demonstrated to be Substantially Equivalent (SE) to a device already on the market [8]. And the "*globally at least equivalent*" (GALE) requirement – e.g. in French law for railway safety [9, 10] – requires that system changes produce a safety level "at least equivalent" to that which existed before the change.

Usually, an assessor does not know if an NWTES claim does, in fact, hold – i.e. prior to testing a new system, an assessor has some confidence that this system is an improvement, based in part on their detailed knowledge of an older system. And such prior confidence, when suitably formalised in statistical terms, should influence the impact of operational testing evidence when assessing the new system. The question is "*how much*" prior confidence in NWTES is sufficient, to support reliability claims for a new system that is subjected to testing?

By addressing all of the foregoing, this paper makes a number of contributions to both (software) reliability theory and assessment practice. Upon formalising intuitive NWTES notions used by practitioners, we present a statistically principled approach to reliability assessment – one that results in conservative assessments for a new system (subjected to operational testing), while also incorporating additional reliability evidence from an older system. We show what various levels of confidence in NWTES, and failure-free testing evidence, allow one to claim for the reliability of a new system. And we determine how much confidence in NWTES, or testing evidence, is required to claim a certain level of reliability. Also, we give some indications about quantifying NWTES beliefs in practice.

While our assessment approach is Bayesian, it is applicable even when an assessor can (justifiably) specify only *some* beliefs about how the reliabilities of the old and new systems are related – this alleviates much of the burden of having to fully specify prior probability distributions. Moreover, the approach produces conservative assessments, as it is a novel extension of what we call *conservative Bayesian inference* (CBI) [11–17] – in this paper, for the first time, CBI is used to combine reliability evidence from multiple systems.

This work continues the authors' recent research into ways that practitioners' plausible intuitions about the assessment of critical software-based systems can be made rigorous in support of quantitative claims about reliability and safety. Our CBI methods help to check whether apparently "obviously plausible" claims can be trusted – revealing situations where such trust is inappropriate, and providing ways forward for these.

The outline of the paper is as follows. An overview of previous CBI applications is given next, in section 2, followed in section 3 by a description of the basic set-up for the CBI model developed in this paper. In section 4, four detailed scenarios with numerical examples show the implications of this model. Section 5 discusses practical considerations, modelling results, various contexts in which these results apply, and future work. Finally, the paper concludes with section 6.

2. A Review of CBI

The primary measure of reliability in this paper is the *probability of a system failing on a random demand* it receives from its environment (*pfd*). For sufficiently sophisticated software, an assessor is uncertain about the value of the *pfd*, and their uncertainty is formalised as a suitable prior probability distribution over all possible *pfd* values. In a Bayesian reliability assessment, the assessor updates their beliefs – i.e. this prior distribution – using evidence in the form of the observed failure (or lack thereof) of the software during operational testing.

The essential idea of CBI is that, instead of requiring an assessor specify a complete prior distribution, one considers the set of *all prior distributions* compatible with only *partial* prior knowledge specified by the assessor. This partial prior knowledge should be relatively easy for assessors to both state and justify with great confidence. And it will typically be much simpler than specifying an entire probability distribution. An instance, a confidence bound on system *pfd*. So, for instance, rather than convincing oneself[1] that the prior distribution of the software's *pfd* – prior to any operational testing – is precisely a Beta distribution with parameters $a = 1, b = 1000$, an assessor might only need much simpler prior knowledge like "the probability of $pfd < 10^{-3}$ is at least 80%". Building arguments to support the latter is a much easier task than the former. In this way, CBI allows for inference to proceed without having to fully specify a prior distribution for model parameters.

Then, depending on the specific posterior reliability prediction of interest (e.g. expected posterior *pfd*, the example we focus on in this paper), CBI determines a prior – out of *all* prior distributions satisfying the partial prior knowledge – whose posterior predicts the worst reliability.

The initial CBI idea published in [12] concerned the posterior expected *pfd* of software that passes n operational tests. Indeed, if one has a complete prior distribution of *pfd*, F say, then typical Bayesian inference gives[2]:

$$\mathbb{E}[pfd \,|\, \text{pass } n \text{ tests}] = \frac{\int_{[0,1]} x(1-x)^n \, \mathrm{d}F(x)}{\int_{[0,1]} (1-x)^n \, \mathrm{d}F(x)} \qquad (1)$$

[3]However, normally, one does not have a sound argument for adopting a specific F; instead, one has something much more limited, like a confidence bound on *pfd*:

$$P(pfd \leqslant y) = 1 - \alpha \qquad (2)$$

Such limited partial prior knowledge could be supported by evidence generated from the development process of some software, e.g. formal-technique-based program analysis, or prescriptive standard-based practice. For instance, one might claim $P(pfd \leqslant 10^{-4}) = 90\%$ on the basis of evidence that the software is strictly developed against IEC 61508 SIL-4 [18].[4] It was shown in [12] that, among all possible prior distributions F that satisfy (2), the two-point distribution shown in Fig. 2 gives the most conservative[5] posterior mean *pfd*, i.e. maximises (1). It has $1 - \alpha$ probability mass at $pfd = y$ and α probability mass at an optimality achieving point $pfd = z$ – where z, being a function of n, can be calculated numerically. Using this prior as our F in (1) results in the conservative upper-bound (3).

$$\mathbb{E}[pfd \mid \text{pass } n \text{ tests}] \leqslant \frac{y(1-y)^n(1-\alpha) + z(1-z)^n\alpha}{(1-y)^n(1-\alpha) + (1-z)^n\alpha} \qquad (3)$$

As illustrated, CBI starts from limited partial prior knowledge, resulting in an attainable upper-bound at the price of being conservative. To limit the conservatism, more partial prior knowledge can be elicited and used in CBI, e.g. a prior confidence in the perfection of the software is used in [12] to improve bound (3). And, the more partial prior knowledge an assessor incorporates into the assessment, the less conservative the CBI bound becomes. With a tension between the conservatism of CBI and one's burden in forming partial prior knowledge, an assessor must find a happy medium in practice.

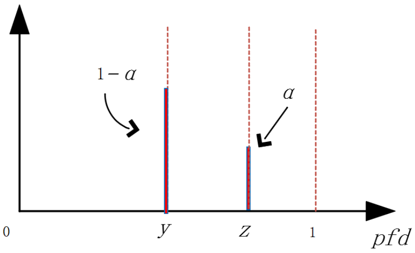

Figure 2: The most conservative prior distribution: one that maximises (1) while satisfying the partial prior knowledge (2).

CBI is applicable in many contexts and scenarios. For instance, CBI may be used with the following typical forms of partial prior knowledge (either solely or in combination):

- $\mathbb{E}[pfd] \leqslant m$: the prior mean *pfd* cannot be worse than a stated value;
- $P(pfd < p) = 1 - \alpha$: a prior confidence-bound on *pfd*;
- $P(pfd = 0) = \theta$: prior confidence in the perfection of the software;
- $\mathbb{E}[(1 - pfd)^n] \geqslant \gamma$: prior confidence in the reliability of passing n tests;
- $P((1 - pfd) \geqslant \frac{k}{n}) = 1 - \alpha$: a prior confidence-bound on the expected number $n(1-pfd)$ of successes for a system subjected to n i.i.d. tests.

And, CBI has been investigated for various objective functions, each with a "posterior" flavour:

- $\mathbb{E}[pfd \mid \text{pass } n \text{ tests}]$: the posterior expected *pfd* [12];
- $P(pfd \leqslant \epsilon \mid \text{pass } n \text{ tests})$: a posterior confidence bound on *pfd*. The ϵ normally represents a very small *pfd* of interest, such as that stipulated by some higher level system requirements [15];
- $P(pfd = 0 \mid \text{pass } n \text{ tests})$: the posterior probability of perfection [14];
- $\mathbb{E}[(1 - pfd)^t \mid \text{pass } n \text{ tests}]$: the posterior probability of the software passing the next t demands [13].

The CBI model in this paper is a novel extension of the CBI ideas in [12, 13] to a *multivariate* prior distribution case, with partial prior constraints on the relationship between the unknown *pfd*s of an old and a new system.

[1] Or industry regulators when submitting a safety case

[2] With standard assumptions, like the testing regime being a series of independent, identically distributed, Bernoulli trials, and the test inputs being statistically representative of operational use, e.t.c.

[3] These integrals are Lebesgue-Stieltjes integrals, defined with respect to the probability measure induced by the distribution function F. Such integrals are valid for all the distributions in this paper.

[4] This is just an example of a form of reasoning we have come across; we are not advocating it as a sound argument.

[5] We shall refer to such priors as being "*the most conservative*", but this is not meant to imply that they are unique. There may be other priors that give results that are equally conservative, but none give more conservative results.

3. The Basic Model

Our CBI model will characterise what one can expect the *pfd* of a new software-based system B to be, given beliefs (supported by testing evidence) about the *pfd* of an older software-based system A. Such a model would be useful if, say, one were considering the probability of a replacement, emergency-trip system failing to shut down a reactor when a sensor correctly detects a dangerous event, such as the temperature in a reactor reaching some threshold. Assume two systems are built to the same engineering requirements, where system A is old and has been in operation for some years, while system B is new. So:

1. pfd_A, pfd_B are the unknown *pfd*s of the pre-existing A system and the new B system;

2. There is a joint distribution F over all possible pairs of values for the A and B system *pfd*s – a distribution over the unit square $[0, 1] \times [0, 1]$. F captures an assessor's beliefs about the possible *pfd* values;

3. The marginal distributions for pfd_A and pfd_B are

$$F_A(x) = \int_{[1,0]\times[1,0]} \mathbf{1}_{u\in[0,x]}\mathbf{1}_{v\in[0,1]}\,\mathrm{d}F(u,v)\,,$$
$$F_B(y) = \int_{[1,0]\times[1,0]} \mathbf{1}_{u\in[0,1]}\mathbf{1}_{v\in[0,y]}\,\mathrm{d}F(u,v)\,,$$

 where $\mathbf{1}_{\mathrm{S}}$ is an indicator function – it equals 1 when predicate S is true, and 0 otherwise;

4. We formulate the NWTES belief as the requirement:

$$\int_{[0,1]\times[0,1]} \mathbf{1}_{v\leqslant u}\,\mathrm{d}F(u,v) \;=\; P(pfd_B \leqslant pfd_A) \;=\; c \qquad (4)$$

 That is, "I am $(c \times 100)\%$ confident that the new system's *pfd* is better than the older system's".

The NWTES formulation (4) is relatively simple and consistent with the notion of a diminished failure rate (e.g. due to fixing software bugs found in the software, without introducing new bugs) [19]. One could consider other forms of NWTES beliefs based on specific supporting evidence derived from a direct comparison of the two systems. For example, conditional on both systems' *pfd*s being better than 10^{-3}, some stated confidence that pfd_B is smaller than pfd_A. Or, another example, the marginal *pfd* distributions are stochastically ordered, i.e. the *Cumulative Distribution Function* (CDF) curve of pfd_A is believed to lie, everywhere, below that for pfd_B. In this paper we focus on the NWTES formulation (4), for its relative simplicity and the insight it brings concerning assessment challenges. We leave other NWTES forms for future work.

For the new B system, assume the assessor only has (4) as prior knowledge whilst, due to system A's age, the assessor can possess various forms of prior knowledge about system A. In section 4, scenarios with different forms of (partial) prior knowledge for system A are considered:

- Scenario 1: The assessor knows the old system's *pfd* with certainty, i.e. pfd_A is a known constant;

- Scenario 2: the assessor cautiously expresses a confidence bound, e.g."I am 99% sure the *pfd* of the old system A is less than 0.001";

- Scenario 3: the assessor, armed with convincing verification evidence for system A and evidence of its operating without failure, expresses beliefs about how likely $pfd_A = 0$ is (i.e. system A is "perfect"), along with a confidence bound on pfd_A if *not* zero [15, 20];

- Scenario 4: two cases where the assessor's evidence supports a complete, marginal prior distribution for the A-system *pfd* – with, and without, a probability of the old system being perfect.

Apart from scenario 1, for scenarios 2 through 4, marginal prior knowledge of the old system A is progressively modelled from very modest (i.e. only a confidence bound) to very detailed (i.e. an entire marginal distribution); and the added impact of failure-free evidence from operational testing is analysed. In section 4, we examine these scenarios in turn. Note that, in Bayesian terms, mathematical forms of prior knowledge constrain an assessor's candidate prior distributions, so we will use the phrase "prior constraints" instead of partial prior knowledge in discussing the models and their implications.

4. Implications of Various Forms of Prior Constraints On the Old System

4.1. A Known pfd for the Old System

To begin, consider the extreme (but simple) situation where the old A-system's *pfd* is known to be p_A. Together with the NWTES belief (4), these constrain the largest value for the expected *pfd* for the new B system:

$$\begin{aligned}&\mathbb{E}[pfd_B]\\ &= c\,\mathbb{E}[pfd_B\,|\,pfd_B \leqslant p_A] + (1-c)\,\mathbb{E}[pfd_B\,|\,pfd_B > p_A]\\ &\leqslant 1-(1-p_A)c \qquad (5)\end{aligned}$$

So, if one is certain that the new system is no worse than the old system (i.e. $c = 1$), the worst-case value for the expected B-system *pfd* is the known *pfd* value, p_A, from the old A-system. In other words, to make a conservative assessment, one must claim that the new system is just as reliable as the old one, even when the new system is known to be no worse. On the other hand, if the new system is known to be worse than the old one (i.e. $c = 0$), then the worst-case expected *pfd* value for the new system is 1 – i.e. expected to fail on every demand. This illustrates how extreme conservatism can result, even with an unrealistic amount of prior knowledge (i.e. being certain) about the old system A. Are the results less extreme if our assessor expresses *more* uncertainty about the A-system *pfd*?

4.2. A Confidence Bound on the Old System's pfd

Assume our assessor has a more modest form of belief about the old system's reliability, such as when only a confidence bound can be inferred from the history of the old A system:

$$P(pfd_A \leqslant p_A) = 1 - \alpha_A \qquad (6)$$

The assessor's confidence in the upper confidence bound p_A is $1-\alpha_A$ or, equivalently, their doubt about the bound is α_A. Together with the NWTES belief (4), any prior distribution F satisfying these two constraints must allocate probability masses to events accordingly, as illustrated in Fig. 3. There, the set of all possible pairs of *pfd* values – for the A and B systems – is the region defined by the unit square. Let M_i be the probability, according to F, that the pair of *pfd*s for the systems is some point in the region i. Then the two constraints, (4) and (6), may be restated as $M_4 + M_3 = c$ and $M_2 + M_3 = \alpha_A$, respectively. Of course, by definition, $M_1+M_2+M_3+M_4 = 1$. In some sense, region 4 contains the most desirable pairs of *pfd*s for the systems – those for which the old system is very good (i.e. $pfd_A \leqslant p_A$) and the new system is possibly even better (i.e. $pfd_B \leqslant pfd_A$). Contrastingly, region 2 contains the least desirable pairs.

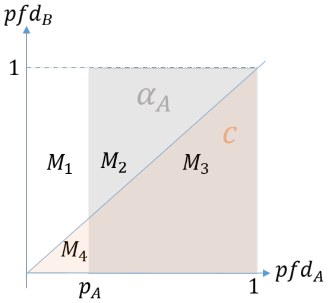

Figure 3: Any joint prior distribution F must satisfy $M_2 + M_3 = \alpha_A$ and $M_4 + M_3 = c$, where F assigns probability M_i to region i.

Fig. 3 does not uniquely define a prior distribution F, but rather a whole collection of distributions. So we may ask "what are the implications of our assessor's relatively weak beliefs for conservative assessment"? Our assessor's objective function – the expected value of pfd_B – can be written as:

$$\mathbb{E}[pfd_B] = \sum_{i=1}^{4} \mathbb{E}[pfd_B|\text{region } i]M_i = \sum_{i=1}^{4} p_i M_i \tag{7}$$

where $p_i := \mathbb{E}[pfd_B|\text{region } i]$. Being conservative would mean identifying a prior distribution that maximises $\mathbb{E}[pfd_B]$. Since $p_A \leqslant p_2 \leqslant 1$ and $0 \leqslant p_4 \leqslant p_A$, we must have

$$\begin{aligned}\mathbb{E}[pfd_B] &\leqslant 1\cdot M_1 + 1\cdot M_2 + 1\cdot M_3 + p_A\cdot M_4\\ &= 1 - M_4(1-p_A)\end{aligned} \tag{8}$$

This is a more general form of (5). Notice that the r.h.s. of (8) is a linearly decreasing function of M_4. Consequently, the maximum value of $\mathbb{E}[pfd_B]$, denoted S^*, occurs at the smallest values for M_4. This makes sense – conservatism dictates that as little confidence as possible be placed in the new system 1) being better than the old system, and 2) having a *pfd* better than p_A. In Appendix A, our assessor's beliefs force M_4 to be bounded below as $M_4 \geqslant \max(0, c-\alpha_A)$. So, there are two cases to consider, depending on how confident our assessor is in the new system (i.e. the value c) or how doubtful of the old system they are (i.e. the value α_A):

- if $M_4 = 0$ (the worst case if the assessor believes $c < \alpha_A$), the assessor is not excluding the new system having a *pfd* no better than p_A. And, without any contrary testing evidence, the worst-case expected *pfd* is $S^* = 1$. Clearly, this is too conservative;
- if, instead, $M_4 = c - \alpha_A$, the assessor can be fairly confident in both systems (i.e. $c \geqslant \alpha_A$). Despite this confidence, the worst-case expected *pfd* is still rather conservative (Table 1) at $S^* = 1-(c-\alpha_A)(1-p_A)$.

In fact, these two cases can be stated together as

$$S^* = 1-(c-\alpha_A)(1-p_A)\mathbf{1}_{c\geqslant\alpha_A} \tag{9}$$

where $\mathbf{1}_S$ is the indicator function[6] for the logical predicate S. Clearly, the worst-case (5) in subsection 4.1 is a special "best" case of (9) when the assessor has complete confidence in the old system *pfd* being better than p_A (so $\alpha_A = 0$). It seems our assessor is paying the price for more uncertainty about the old system with even more conservative bounds. Prior distributions attaining the worst-case (9) are depicted in Fig.s 4 and 5 as the asymptotic limits of joint distributions which satisfy the assessor's beliefs[7].

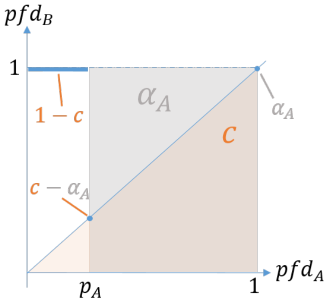

Figure 4: A joint distribution that attains the bound (9) when $c \geqslant \alpha_A$. A probability mass of $1-c$ is assigned uniformly to the $pfd_B = 1$ horizontal line-segment in region 1. The mass $c-\alpha_A$ is assigned to the point (p_A, p_A) in region 4, while the mass α_A is assigned to the point $(1,1)$ in region 3, with zero mass for region 2.

In both Table 1 and (9), as our assessor's doubt decreases, the worst-case S^* decreases. But, its value cannot be smaller than p_A. In fact, as one's beliefs tend to certainty (i.e. $\alpha_A \to 0$ and $c \to 1$), we find $S^* \to p_A$ from above. However, so far we have not been "Bayesian" – i.e. reliability evidence has not altered our beliefs. With such evidence – e.g. the new B system passing n tests – can the S^* bound improve beyond this p_A "floor"?

Rather than the expected B-system *pfd*, we now seek the posterior expected *pfd*, $\mathbb{E}[pfd_B \mid B \text{ passes } n \text{ tests}]$. However, we are not looking for the posterior value given a *specific* prior distribution, but for the worst-case posterior value, S^*, given a

[6] The function $\mathbf{1}_S$ has value 1 if S is true, and zero otherwise.

[7] These priors that attain the worst-case value for the posterior expected B-system *pfd* are not unique: for all CBI *most conservative priors* in this paper, there are other worst-case achieving joint priors that differ from these on (Lebesgue) null subsets of the unit square.

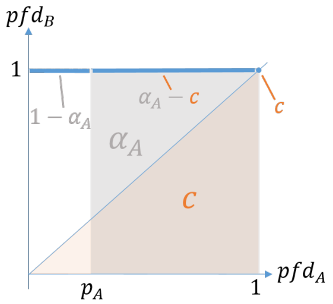

Figure 5: A joint distribution that attains the bound (9) when $c < \alpha_A$. A probability mass of $1 - \alpha_A$ is assigned uniformly to the $pfd_B = 1$ horizontal line-segment in region 1. The mass $\alpha_A - c$ is uniformly assigned to the $pfd_B = 1$ line-segment in region 2, while the mass c is assigned to $(1, 1)$ in region 3, with zero mass for region 4.

Table 1: Numerical examples for the worst-case bound S^* in (9) when $c \geqslant \alpha_A$

p_A	α_A	c	S^*
0.001	0.01	0.9	0.11089
0.001	0.01	1	0.01099
0.001	0	0.9	0.1009
0.001	0	1	0.001

range of prior distributions. Appendix B shows that S^* satisfies $\mathbb{E}[pfd_B \,|\, B \text{ passes } n \text{ tests}] \leqslant S^*$, where

$$S^* = 1 - \frac{(1-p_z)^{n+1}(1-M_4) + (1-p_A)^{n+1}M_4}{(1-p_z)^n(1-M_4) + (1-p_A)^n M_4} \tag{10}$$

and $M_1, \ldots, M_4$ satisfy (4) and (6), while p_z is the unique B-system *pfd* that satisfies both $p_A < p_z \leqslant 1$ and

$$p_z = 1 - \frac{n}{n+1}(1 - S^*) \tag{11}$$

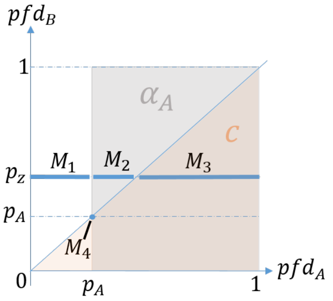

Figure 6: A prior distribution F^* that achieves the bound S^* in (10). Probability masses M_1, M_2 and M_3 are uniformly assigned, within their respective regions of the unit square, to the horizontal p_z-line (see (11)), where p_z must lie in the range $p_A \leqslant p_z \leqslant 1$. The mass M_4 is assigned to the closest point to the p_z-line within the region, (p_A, p_A). See Appendix B for details.

An immediate consequence of (11) is that the inclusion of failure-free evidence in the assessment has not eliminated the "floor" imposed by conservatism on the worst-case bound S^*. Indeed, (11) implies that as failure-free evidence increases (so $n \to \infty$), S^* will equal p_z at best. Our assessor remains unconvinced of the expected *pfd* being better than some value p_z, *no matter how much evidence they observe to the contrary*. Conservatism always allows for the (unlikely) possibility that a fairly unreliable system successfully executes a large number of test inputs.

On the other hand, the distribution F^* in Fig. 6 achieves the worst-case S^*, and illustrates how mounting failure-free evidence forces our conservative assessor to rule out the most unreliable *pfd* values for the B system. Unlike the beliefs in Fig.s 4 and 5, in the present case, the expected B-system *pfd* can be significantly better than 1.

So, being Bayesian has improved the S^* bound, but the undesirable lower bound on S^* (due to conservatism) remains. While failure-free evidence may be convincing enough to improve our expectations of how reliable the new B system is, it is not enough to overcome our skepticism about how reliable the old A system is, and therefore our skepticism about whether the new system *can* be any better. Perhaps a more explicit form for S^* might reveal how S^*'s value depends on both operational evidence and our assessor's skepticism. To obtain such a form, note that the largest value of S^* occurs at the smallest value of M_4 (see Appendix C). And, similar to (9), the smallest value for M_4 occurs either when $M_4 = c - \alpha_A$ (for $c \geqslant \alpha_A$) or when $M_4 = 0$ (for $\alpha_A > c$). That is,

$$S^* = 1 - \frac{(1-p_z)^{n+1}(1-c+\alpha_A) + (1-p_A)^{n+1}(c-\alpha_A)}{(1-p_z)^n(1-c+\alpha_A) + (1-p_A)^n(c-\alpha_A)}\mathbf{1}_{c\geqslant\alpha_A} \tag{12}$$

Unsurprisingly, the special case when $n = 0$, i.e. no failure-free evidence, reduces (12) to (9). And, like (9), we see a clear dependence of S^* (and, therefore, p_z) on α_A, c and p_A. For instance, it is curious that the model suggests the following. If it happened that our doubt in the old system's reliability is equal to our confidence in the new system being better than the old one (i.e. $c = \alpha_A$), then we could always expect the worst reliability ($S^* = 1$) for the new system, even if it is *much* better than the old one. So, for useful bounds S^*, one must simultaneously have a lot of confidence in the old system, and in the improvement the new system brings. In fact, due to S^* being a decreasing function of M_4, the smallest value of S^* achievable with an arbitrary amount of failure-free evidence occurs when the assessor has no doubts: i.e., as $c \to 1$ and $\alpha_A \to 0$, we have $p_z \to p_A$ and $S^* \to p_A$ from above, just as before. This lower bound on both S^* and p_z is further illustrated by the examples in Table 2 using (11) and (12).

Table 2 illustrates two lessons. The good news: in these scenarios – with strong confidence in a very reliable old system A and an even better new system B – failure-free evidence quickly yields an S^* (i.e. our conservative posterior claim for pfd_B) of the same order of magnitude as our prior confidence bound in pfd_A, that is p_A. In some practical situations, this will be sufficient to satisfy the reliability requirements. However, there is also some bad news: S^* cannot have a value better than p_A. CBI utilises all forms of doubt expressed. So, even if there is a slim chance that the new system is worse than the old, and that the old system has *pfd* worse than p_A, this chance is

Table 2: Numerical examples for the worst-case bound S^* in (12)

p_A	α_A	c	n	p_z	S^*
0.001	0.01	0.9	10	0.0956	0.00513
			1000	0.002	0.00104
			10000	0.0011	0.001
0.001	0.01	0.99	10	0.0925	0.00171
			1000	0.002	0.00101
			10000	0.0011	0.001
0.001	0.01	1	10	0.0921	0.00135
			1000	0.002	0.001
			10000	0.0011	0.001

an opportunity to be conservative. So, the only way for testing evidence to have a better impact on our bounds is if p_A is very small. And how small the S^* bound is shown to be in practice – particularly when assessing systems with ultra-high reliability requirements – must depend on how reasonable/feasible it is to make confidence statements using very small p_A values. It is quite convenient, then, that in a number of practical situations, some confidence in the perfection of the old A system (so $p_A = 0$) can be justified [13, 20–23]. The following section considers the benefit such claims about perfection can bring.

4.3. With Confidence in the Perfection of the Old System

So, in addition to (4) and (6), consider an assessor's belief in the possible perfection of the old A system,

$$P(pfd_A = 0) = \theta_A \tag{13}$$

for some θ_A. That is, there is an expressed belief in the perfection of the old system, *as well as* a belief in the old system being better than some p_A. In Fig. 7 we depict an allocation of probabilities M_i for any joint density function that satisfies the constraints (4), (6), and (13). Appendix D shows that

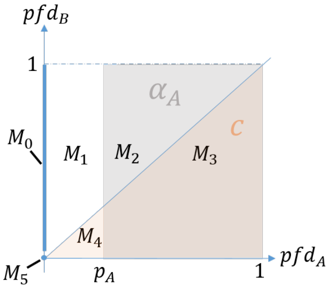

Figure 7: An allocation of probabilities for any prior distribution that satisfies (4), (6) and (13). That is, $M_5+M_4+M_3 = c$, $M_3+M_2 = \alpha_A$, and $M_5+M_0 = \theta_A$, where M_i is the probability associated with the depicted i-th subset of the unit square. Note, the mass M_5 is assigned to $(0,0)$, while M_0 is assigned to the segment $\{(0,y)\,|\,0 < y \leqslant 1\}$.

$\mathbb{E}[pfd_B \,|\, B \text{ passes } n \text{ tests}] \leqslant S^*_{pp}$, where

$$S^*_{pp} = S^*_{ppLHS}\mathbf{1}_{p_z > p_A} + S^*_{ppRHS}\mathbf{1}_{p_z \leqslant p_A} \tag{14}$$

for which

$$S^*_{ppLHS} := 1 - \frac{(1-p_z)^{n+1}(1-M_4-M_5) + (1-p_A)^{n+1}M_4 + M_5}{(1-p_z)^{n}(1-M_4-M_5) + (1-p_A)^{n}M_4 + M_5}\,,$$

$$S^*_{ppRHS} := 1 - \frac{(1-p_z)^{n+1}(1-M_2-M_5) + (1-p_A)^{n+1}M_2 + M_5}{(1-p_z)^{n}(1-M_2-M_5) + (1-p_A)^{n}M_2 + M_5}\,,$$

the masses $M_0, \ldots, M_5$ are given, they satisfy the constraints, and p_z is the unique *pfd* value that satisfies

$$p_z = 1 - \frac{n}{n+1}(1 - S^*_{pp}) \tag{15}$$

The forms of the bounds S^*_{ppRHS} and S^*_{ppLHS} are consistent with the story so far, with p_A playing an explicit role in controlling the size of the bounds. However, unlike previous S^* forms, the bounds now contain the probability M_5 of both systems being perfect. The conservative prior distributions that, together, result in (14) and (15) are shown in Fig. 8. The value of S^*_{pp} depends on whether $p_z > p_A$ or $p_z \leqslant p_A$, and which of these holds depends on the value of p_z (itself, a function of the M_is). Therefore, to determine how large S^*_{pp} can get, one determines those values of M_i that achieve this.

Our bound S^*_{ppRHS} (and, therefore, S^*_{pp}) can now fall below p_A, but only when *both systems can be perfect*! That is, we need $M_5 > 0$. In practice, this will be a requirement of an assessor who expresses some confidence in NWTES and the possible perfection of the old A system. Otherwise, even if the A-system could be perfect, failure-free evidence will not convince them of the expected reliability of the new system beyond some value p_A. But with $M_5 > 0$, Appendix E and Appendix F show that when $\alpha_A \leqslant 1 - \theta_A \leqslant c$, then the probability masses (consistent with this ordering of the constraint parameters) that give the largest S^*_{ppRHS} value are $\{M_2 = 0,\ M_4 = 1 - \theta_A - \alpha_A,\ M_5 = c - 1 + \theta_A\}$. And the bounds in (14) become

$$S^*_{ppLHS} = 1 - \frac{(1-p_z)^{n+1}(1-c+\alpha_A) + (1-p_A)^{n+1}(1-\theta_A-\alpha_A) + c - 1 + \theta_A}{(1-p_z)^{n}(1-c+\alpha_A) + (1-p_A)^{n}(1-\theta_A-\alpha_A) + c - 1 + \theta_A}\,,$$

$$S^*_{ppRHS} = 1 - \frac{(1-p_z)^{n+1}(2-c-\theta_A) + c - 1 + \theta_A}{(1-p_z)^{n}(2-c-\theta_A) + c - 1 + \theta_A}\,. \tag{16}$$

In particular, (16) implies that $S^*_{pp} \xrightarrow{n\to\infty} 0$. So now, our assessor is forced to accept that the old system might be perfect. And therefore, via NWTES, that failure-free evidence suggests the new system might be perfect too! So the more failure-free runs that *are* observed, the more likely this perfect pair becomes, and the smaller our conditional worst-case expected *pfd* is. Table 3 illustrates this when $p_z \leqslant p_A$ occurs, with p_z approaching 0 as n grows.

4.4. Fully Specified Prior Knowledge for the Old System

The combination of a claim about the perfection of both systems (even if made implicitly as in the previous section), and NWTES, allows failure-free evidence to "save" our conservative bounds from extreme conservatism. We now see why the first scenario in section 4.1 had shortcomings – the assessor's confidence in the value of the old system's *pfd* may have been

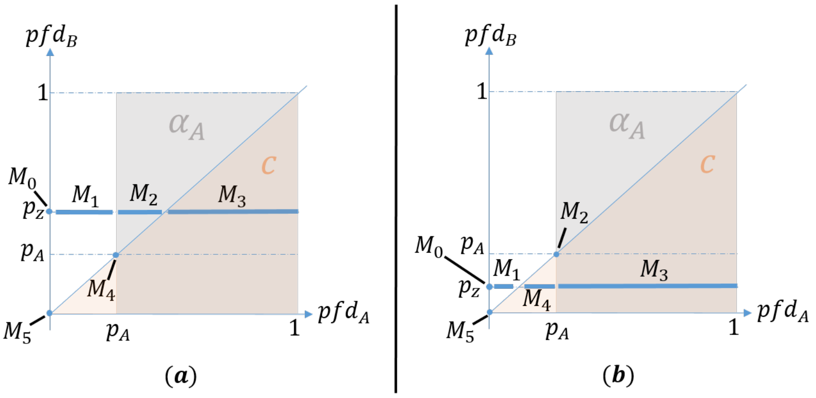

Figure 8: Two joint prior distributions that, together, achieve the bound S^*_{pp} in (14). When $p_z > p_A$, the distribution **(a)** achieves S^*_{ppLHS}, while **(b)** achieves S^*_{ppRHS} for $p_z \leqslant p_A$. Within each depicted subset of the unit square, probability masses are assigned to be as close as possible to, or uniformly assigned on, the horizontal p_z-line (see (15) for p_z). When necessary, to satisfy constraints, masses are assigned to points, such as M_0 assigned to the point $(0, p_z)$ or M_5 to $(0, 0)$. See Appendix D for details.

Table 3: Numerical examples for the worst-case bound S^*_{pp} in (14)

θ_A	p_A	α_A	c	n	p_z	S^*_{pp}	Prior
0.5	0.001	0.01	0.9	10	0.0952	0.0047	LHS
				1000	0.0014	0.00036	LHS
				10000	0.00014	0.000037	RHS
0.95	0.001	0.01	0.9	10	0.0947	0.0042	LHS
				1000	0.00106	0.000061	LHS
				10000	0.00011	0.0000061	RHS
0.5	0.001	0.01	0.99	10	0.09201	0.00121	LHS
				1000	0.0013	0.00028	LHS
				10000	0.00013	0.000029	RHS
0.95	0.001	0.01	0.99	10	0.09159	0.00075	LHS
				1000	0.001	0.000023	LHS
				10000	0.0001	0.0000023	RHS
0.5	0.001	0.01	1	10	0.0917	0.00085	LHS
				1000	0.00127	0.00027	LHS
				10000	0.00013	0.000028	RHS
0.95	0.001	0.01	1	10	0.0913	0.00039	LHS
				1000	0.001	0.000019	LHS
				10000	0.0001	0.0000019	RHS

extreme (i.e. a certainty in *pfd* p_A), but the *pfd* value itself was not (i.e. $p_A \neq 0$). Their bounds would have been better, had they spread out their beliefs a bit more to allow the possibility of the old system being much better than could be confidently claimed. So, suppose the assessor has enough information to specify a complete marginal prior distribution, i.e. some density function f_A for a continuous distribution over $[0, 1]$. This distribution and the confidence bound (4), together, constrain the worst-case, conditional expected *pfd* for the new *B*-system, having observed n failure-free runs. Appendix G shows that $\mathbb{E}[pfd_B \,|\, B \text{ passes } n \text{ tests}] \leqslant S^*_{cmplt}$, where

$$1 - S^*_{cmplt} = \frac{(1-p_z)^{n+1}(1 - c + \int_{p_z}^{1} f_A(x)\,\mathrm{d}x) + \int_{p_{1-c}}^{p_z} (1-x)^{n+1} f_A(x)\,\mathrm{d}x}{(1-p_z)^{n}(1 - c + \int_{p_z}^{1} f_A(x)\,\mathrm{d}x) + \int_{p_{1-c}}^{p_z} (1-x)^{n} f_A(x)\,\mathrm{d}x} \qquad (17)$$

for the unique *pfd* value p_{1-c} that satisfies $\int_0^{p_{1-c}} f_A(x)\,\mathrm{d}x = 1 - c$, and p_z is the unique *pfd* value that satisfies

$$p_z = 1 - \frac{n}{n+1}(1 - S^*_{cmplt}) \qquad (18)$$

Table 4: Numerical examples for the worst-case bound S^*_{cmplt} in (17)

$f_{beta}(x; a, b)$	c	p_{1-c}	n	p_z	S^*_{cmplt}
$a = 1$ $b = 1000$ $\mathbb{E}[X] = 9.99e{-4}$ $\sigma^2 = 9.96e{-7}$	0.9	$1.05e{-4}$	0	1	0.201
			1000	$1.67e{-3}$	$6.73e{-4}$
			10000	$3.6e{-4}$	$2.61e{-4}$
			100000	$1.45e{-4}$	$1.35e{-4}$
	0.99	$1.01e{-5}$	0	1	$1.11e{-1}$
			1000	$1.5e{-3}$	$5.38e{-4}$
			10000	$2.6e{-4}$	$1.6e{-4}$
			100000	$4.92e{-5}$	$3.92e{-5}$
$a = 0.9$ $b = 900$ $\mathbb{E}[X] = 9.99e{-4}$ $\sigma^2 = 1.11e{-6}$	0.9	$8.58e{-5}$	0	1	$2.0e{-1}$
			1000	$1.64e{-3}$	$6.45e{-4}$
			10000	$3.37e{-4}$	$2.37e{-4}$
			100000	$1.24e{-4}$	$1.14e{-4}$
	1	0	0	1	$9.99e{-4}$
			1000	$1.498e{-3}$	$4.997e{-4}$
			10000	$2.38e{-4}$	$1.38e{-4}$
			100000	$3.63e{-5}$	$2.63e{-5}$

The joint distribution F^* that achieves this worst-case posterior value is schematically depicted[8] in Fig. 9.

To illustrate (17) and (18), Table 4 contains numerical examples using a beta-density $f_{beta}(x; a, b)$ for the marginal $f_A(x)$. Some observations from Table 4 are:

- if one doubts an NWTES assumption, so that $c \neq 1$, then as the number of failure-free test runs increases, the value of S^*_{cmplt} tends to its smallest value p_{1-c}, but no smaller – a result of no probability mass lying below $pfd_B = p_{1-c}$ in Fig. 9. This agrees with all of our earlier results (e.g. $\lim\limits_{n\to\infty} S^* = p_A$ in (12));

[8] F^* is not absolutely continuous with respect to the Lebesgue measure over the unit square – e.g. F^* assigns non-zero probability to line segments. So, F^* is not characterised solely by a joint density function.

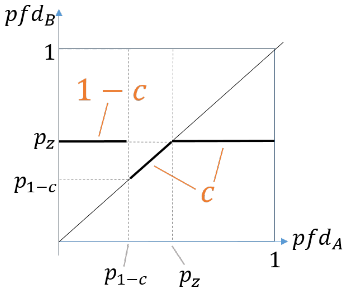

Figure 9: This schematically depicts a conservative joint prior distribution F^* that achieves the worst-case posterior expected *pfd* for the B system (17), given f_A – a completely specified, continuous marginal density for pfd_A. The bold line-segments have the indicated probability masses, c and $1-c$, assigned to them by F^*. F^* is the limit of a weakly convergent sequence of conservative joint prior distributions. See Appendix G for details.

- when the new system cannot be worse than the old one ($c = 1$), S^*_{cmplt} is arbitrarily small for large n;
- when $c \neq 1$ and $n = 0$, S^*_{cmplt} is much larger than the mean pfd_A, $\mathbb{E}[pfd_A]$. Intuitively, without any failure-free evidence from the new system, although the assessor is almost certain (e.g. $c = 0.99$) that the new system has a smaller *pfd* than the old system, CBI utilizes whatever little doubt the assessor has, to specify a joint prior distribution with a larger mean *pfd* for the new system, $\mathbb{E}[pfd_B]$. However, with a large enough c, such overly pessimistic results are easily overcome with a modest number n of observed failure-free runs, e.g. $n = 1000$;
- when $c = 1$ and $n = 0$, we have $S^*_{cmplt} = \mathbb{E}[pfd_A]$. So, even when one is certain the new system is better than the old, conservatism forces one to assume the expected *pfd* of the new system is the same as that of the old system;
- when $n \neq 0$ and $c = 1$, by the definition of S^*_{cmplt} we have (see Fig. 10):

$$S^*_{cmplt} \geqslant \mathbb{E}[\,pfd_B \,|\, B \text{ passes } n \text{ tests}] \overset{\text{in particular}}{=} \frac{\int_0^1 x(1-x)^n f_A(x)\,\mathrm{d}x}{\int_0^1 (1-x)^n f_A(x)\,\mathrm{d}x} \tag{19}$$

  In essence, (19) says, even when one is certain the B system is better than the old A system (i.e. $c = 1$), S^*_{cmplt} will still not be better than it would be if the Bayesian inference for the B-system *pfd* distribution utilised the continuous marginal distribution of pfd_A as the prior distribution of pfd_B. However, on the other hand, this unpleasant result also provides an important warning against naively using the old system's *pfd* distribution – in a bid to be conservative – as the prior distribution for the new one. Doing this might not be conservative at all, and may in fact be too optimistic. This, even when one is certain the new system is better than the old one.

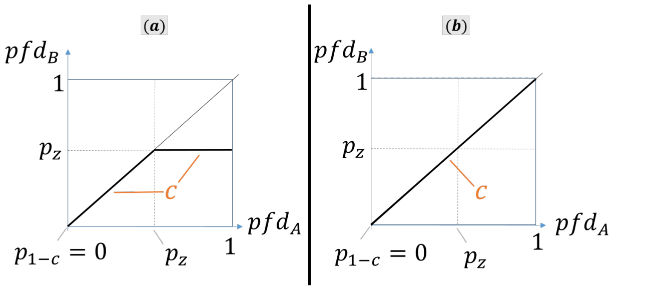

Figure 10: With continuous marginal density f_A for pfd_A, the conservative joint prior distribution in **(a)** achieves the worst-case S^*_{cmplt} when $c = 1$ (this is a special case of F^* in Fig. 9). Any other similarly constrained joint prior distribution – such as that in **(b)** with a marginal B-system *pfd* density equal to f_A – must give a posterior expected pfd_B smaller than S^*_{cmplt} (i.e. inequality (19) holds, with its r.h.s. expression given by using density f_A in (1)). The B-system *pfd* distributions in **(a)** and **(b)** are also depicted in Fig. 12.

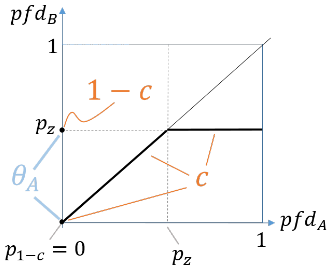

Figure 11: A conservative joint prior distribution that achieves the worst-case posterior bound S^*_{cmplt}, and has a fully specified marginal distribution for pfd_A, F_A – with θ_A prior confidence in the A-system being perfect. In this example, it is assumed that $\theta_A > 1-c$.

It is worth mentioning that, when a marginal prior distribution of old system A's *pfd* (*with* some prior confidence in A being perfect) is fully specified, CBI gives arbitrarily small worst-case posteriors from mounting failure-free evidence. For example, assume there is a θ_A probability mass at the origin of a marginal prior distribution F_A for pfd_A, and in the range of $x \in (0, 1]$, F_A takes the form of a scaled Beta distribution – with parameters a, b and density $(1-\theta_A) f_{beta}(x; a, b)$. That is, for $0 \leqslant x \leqslant 1$,

$$P(pfd_A \leqslant x) = F_A(x) := \theta_A + (1-\theta_A)\int_0^x f_{beta}(u; a, b)\,\mathrm{d}u \tag{20}$$

Given such prior knowledge of the old A-system, there are two cases to consider:

- When $\theta_A \leqslant 1-c$, the worst-case joint prior is still the one in Fig. 9, exemplified in table 4;
- When $\theta_A > 1-c$, similar to how the joint prior for Fig. 9 is derived, one obtains a conservative joint prior distribution by first ensuring that as much of the NWTES mass, c, as is possible, lies beneath the diagonal and on the far right in the unit square (while satisfying constraints on the mass. See Appendix H for a detailed argument). Upon doing this, we must have $p_{1-c} = 0$, because the inequality $\theta_A > 1-c$ implies there must still be some probability mass at the origin no matter how much mass one relocates from the origin elsewhere within the unit square. This generalises the simpler case in section 4.3,

which also had a probability of the A-system being perfect (and a mass-moving argument for that case is given in Appendix E). Consequently, by following near identical arguments to those given in Appendix G, one deduces the conservative joint prior distribution depicted in Fig. 11. And given the F_A and NWTES constraints, this distribution attains the largest value for the posterior expected B-system *pfd*, S^*_{cmplt},

$$1 - S^*_{cmplt} = \frac{Nu_1 + Nu_2}{De_1 + De_2} \tag{21}$$

where

$$\begin{aligned}
Nu_1 &= (1-p_z)^{n+1}\left(1 - c + (1-\theta_A)\int_{p_z}^{1} f_{beta}(x;a,b)\,\mathrm{d}x\right) \\
Nu_2 &= \theta_A - 1 + c + (1-\theta_A)\int_0^{p_z}(1-x)^{n+1} f_{beta}(x;a,b)\,\mathrm{d}x \\
De_1 &= (1-p_z)^{n}\left(1 - c + (1-\theta_A)\int_{p_z}^{1} f_{beta}(x;a,b)\,\mathrm{d}x\right) \\
De_2 &= \theta_A - 1 + c + (1-\theta_A)\int_0^{p_z}(1-x)^{n} f_{beta}(x;a,b)\,\mathrm{d}x
\end{aligned}$$

with an associated, unique *pfd* value p_z that satisfies

$$p_z = 1 - \frac{n}{n+1}(1 - S^*_{cmplt}) \tag{22}$$

Table 5 gives numerical examples in which we observe 2-3 orders of magnitude improvement, compared with the results in Table 4, thanks to both the prior confidence in the perfection of the old A-system and strong confidence in an NWTES assumption.

Table 5: Numerical examples for the worst-case bound S^*_{cmplt} in (21) resulting from a joint prior distribution with the complete marginal prior distribution F_A given in (20)

θ_A	a	b	c	n	p_z	S^*_{cmplt}
0.5	1	1000	0.9	10*e*3	1.25*e*–3	2.52*e*–4
				10*e*4	1.28*e*–4	2.79*e*–5
				10*e*5	1.36*e*–5	3.76*e*–6
0.95	1	1000	0.9	10*e*3	1.05*e*–3	5.58*e*–5
				10*e*4	1.06*e*–4	6.04*e*–6
				10*e*5	1.06*e*–5	6.10*e*–7

5. Discussion

5.1. Conservatism in Reliability Assessment

Conservatism in reliability assessment is a fine balancing act. On the one hand, in being conservative, one seeks to "err on the side of caution" by coming up with reliability estimates that indicate worse reliability than the actual unknown value of the system reliability. On the other hand, one also wants the weight of evidence to drive assessments, and alter them in principled ways. Tip the scales too far to the left and our estimates are doomed to be too conservative to be meaningful. Too far to the right, and we risk being unpleasantly surprised by failure events that, otherwise, would *not* have been surprising (as these events would have been judged, conservatively, to be unacceptably likely). Finding that useful middle ground is a judgement call an assessor makes, typically (necessarily?) on a case-by-case basis, by a combination of formal and informal reasoning. Reasoning that, when formalised in probabilistic terms, requires that one's uncertainty about uncertainties be adequately expressed[9] – a nesting of uncertainties. Quite often, such reasoning is difficult, more so when it involves probabilities of rare failure events. And these difficulties are only compounded when one tries to use failure evidence from the assessment of one system in the assessment of another. For while Bayesian inference provides a principled approach to evidence-based reliability assessment, we *are* sympathetic to the plight of an assessor faced with the challenge of coming up with a suitably rich prior probability distribution that captures their prior beliefs about the system having *any* stated plausible reliability level. So, to the assessor seeking that conservative, yet useful, middle ground, we say "be Bayesian, *but* be conservative in how you go about being Bayesian".

5.2. CBI Applied to "Not Worse than the Existing System" Arguments

This is where CBI enters into the picture. Since the publication of [12], a number of CBI applications have been studied, inferring the reliabilities of software-based systems [13–15]. The key novelty of CBI is, instead of assuming a complete prior distribution, only partial prior knowledge of the distribution is required. This knowledge defines a constrained set of prior distributions, each compatible with the specified partial prior knowledge, from which one chooses a "most conservative" prior distribution – i.e. a prior distribution that produces the "most conservative" value for a posterior estimate of interest. Precisely *which* prior is "most conservative" will depend on *which* posterior estimate is of interest. For example, the constrained prior distribution that minimises the probability of perfection [14] is different from the constrained prior that maximises the expected *pfd* [12]. With CBI, one avoids much of the difficulty in eliciting a complete prior distribution when applying Bayesian inference in practice.

Perhaps the very act of *not* having to articulate an entire prior probability distribution over the possible *pfd* values is, itself, a "conservative" act. In articulating only partial knowledge about a suitable prior, one is not compelled to possibly claim more than one can reasonably justify. And yet, such minimalist beliefs can be used to identify a suitable prior – one guaranteed to lead to conservative posterior estimates.

We have extended CBI from previous applications to the present context – that of assessing a new system, where the assessment is based, in part, on an assessment of an older system. Here, a multivariate prior distribution is sought for the inference. Clearly, in most cases, fully specifying a joint-distribution

[9] e.g. in the form of a distribution of system *pfd*.

is infeasible; the assessor would normally only have sparse information of the new system, relatively rich knowledge of the old one, and evidence prior to operational testing to suggest that the new system is probably better than the old one. Hence, to begin with, they may assume nothing, marginally, about the statistical properties of the new system's *pfd*, but have some partial knowledge of the old system's *pfd* distribution and some confidence in an NWTES claim (see (4)). Altogether, these constrain the unknown joint prior in an application of our CBI extension.

With these prior constraints, CBI gives conservative posterior reliability estimates; in this paper, the estimate of interest is a posterior expected *pfd* for the new system. But recall our earlier conservatism "balancing act": when are these estimates *not* too conservative, and to what extent can evidence temper healthy skepticism? Indeed, the analyses and examples in section 4 show that NWTES-based models can produce results that are too conservative, either suggesting the new system will fail on every input it executes on, or suggesting that the new system can only be as reliable as the old one, but no better. And all of this despite failure-free evidence to the contrary. For instance, the worst-case reliabilities in Table 1 and Table 2 are each worse than some "floor" – a lower non-zero limit, *despite having 100% confidence* in NWTES and *infinitely many* observed failure-free runs of the new system!

But these shortcomings come as no surprise. There are 2 influences on claims about the new B-system's reliability that, together, provide the means for extreme conservatism to occur. These influences are "*confidence in* pfd_A" and "*confidence in* NWTES". These are formalised as probabilities of "*less than or equal to*" events. For example, an assessor is $(1 - \alpha_A)\%$ confident that $pfd_A \leqslant p_A$, or is $c\%$ confident that $pfd_B \leqslant pfd_A$. So, to be conservative, CBI exploits the "equals to" possibilities in both of these. That is, to be conservative, the new system can only be "as good as" the old system, but no better, and the old system can only be "as good as" having the *pfd* value p_A, but no better. These assertions may still be held, even when faced with vast amounts of failure-free evidence, *ceteris paribus*. Because, for any finite number of failure-free tests, there is a non-zero probability that the pfd_B value is consistent with these assertions and, nevertheless, the new system survives that many tests. In this sense, these 2 influences on B-system claims are considered "weak"; allowing an assessor to be unreasonably conservative, and yet be formally consistent in their conservatism. Failure-free evidence from the B system can only, at best, convince one that the B system is "as good as" the best *pfd* value believed for the old system – the value p_A. Yes, this is clearly an unhelpfully bullish way to be cautious.

The solution is clear: if "worst-case" reasoning requires we be resolute in our conservative beliefs – that the new system can only be as good as the old system, but no better, and that the old system can, at best, have *pfd* p_A – then we should express beliefs about the old system being, possibly, very reliable indeed. We should articulate how (un)certain we are that the old system has a *pfd* value much better than p_A. This might require collecting extra evidence or analysing more in-depth the available evidence for system A. In fact, in a number of practical contexts, one can "go all the way" and articulate a probability that the old system is altogether fault-free/perfect. And, due to the NWTES assumption, this implies a non-zero probability that the new system is fault-free too. In Table 3, the worst-case expected *pfd*s improve in response to testing evidence, with no limits on the reliability that can be suggested by these estimates. And Table 4 shows such improvement does not need perfection beliefs – our worst-case estimates also improve, with increasing confidence in an NWTES assumption, when a complete marginal *pfd* distribution of the old system is used *without* some confidence in the perfection of the old system. Of course, things only get better when a complete marginal distribution *with* a probability mass for perfection is specified (see Table 5).

The redeeming role of "probability of perfection", and approximations of this, is in line with the findings of previous CBI applications [11–13, 21, 22]. The smaller the *pfd* p_A in our CBI model, the lower our worst-case expected *pfd* can become with increasing failure-free evidence. "Perfection" or "fault-freeness" can be seen as the special case when $p_A = 0$. How best to learn about the probability that a pre-existing system *is* fault-free remains an open question, and an active research area. Work in [14] explores statistical inference using evidence of a good operational history, and [11] uses statistical evidence garnered from other products – but still in the same product line – to infer the probability of perfection.

Of course, depending on how confident an assessor is in NWTES, the amount of failure-free evidence required for a certain reliability claim can vary quite a bit. Consider that a "classical" estimate for how many failure-free runs n are needed, to claim a new system *pfd* "pfd_{claim}" with a confidence level of 99%, is given by [24]:

$$P(\text{surviving } n \text{ tests}) = (1 - pfd_{claim})^n = 1 - 0.99 \qquad (23)$$

We can compare the n suggested[10] by (23) with the number of runs stipulated by our CBI:

- Using the fifth entry from the bottom in Table 4, with an NWTES confidence of 90%, to claim the new system *pfd* is $1.14e{-}4$ requires 10^5 runs. This is significantly more than the 40394 runs suggested by (23);
- In the last entry in Table 4, with 99% or more confidence in NWTES, to claim a $2.63e{-}5$ *pfd* requires 10^5 runs. This is just over half of the number of runs, $1.75e5$, obtained from (23).

So, less confidence in NWTES can result in significantly more convincing needed from successful test executions.

We do not want to give the impression that conservative estimates will always be preferred to more optimistic ones. The practical dictates of a given situation might mean our CBI estimates are simply too conservative to be used directly in building arguments in safety cases. So, an assessor might have to turn to other more optimistic models – ones that also incorporate the history of an old system's development and its operation. Such models could be based on informal or less rigorous reasoning,

[10] Strictly speaking, n is the smallest integer larger than the solution to (23).

and may contain within them implicit assumptions. But, recall that useful "middle ground" an assessor needs to find. By performing *what-if* calculations using the CBI/NWTES model, our assessor is armed with an estimate of how optimistic an alternative model to CBI actually is. Such perspective can give pause, and provide useful caution against an over-reliance on optimistic results. For instance, they could determine the required model-input values for the NWTES model that give matching results between it and a more optimistic alternative model. Would such an alternative (implicitly?) require very high confidence in NWTES? And/or ideal knowledge of the old system (e.g. a high probability of perfection)? Perhaps such knowledge is clearly unsupported in a given context, in which case this should cast doubt on optimistic reliability claims made with the alternative. The works reported in [25, 26] are in this spirit, where caution is given against misleading, optimistic models being used in system-safety claims, arguments and regulations.

Even without contrasting against alternative models, CBI reveals a counter-intuitive case that serves as an admonition to assessors/regulators. This is in the last few entries of Table 4, where the confidence in NWTES is $c = 1$ and a completely specified *pfd* distribution of the old system is given. An assessor might be tempted to reason as follows: "Since I am 100% certain the new system is better, I will be conservative by using the *pfd* distribution for the old one as the prior *pfd* distribution for the new system, in a Bayesian assessment with testing evidence from the new system." And doing this might seem to be both conservative and convenient – the old system distribution is known, trusted, and using it avoids grappling with multidimensional priors. However, (19) shows this to be optimistic.

But why *is* there this lack of conservatism resulting from using the old system's marginal *pfd* distribution as a prior? Well, initially, there isn't. Note that complete confidence in NWTES means that, without seeing any failure-free evidence, it *is* conservative to assume that the new system's *pfd* distribution *is* the same as the old system's. So, the expected *pfd* for the new system *is* the same as that of the old system. However, once we begin to observe failure-free evidence from the new system, at some point, we are forced to accept that the new system cannot be that bad. But (being conservative) we nevertheless maintain that if the old system *is* very reliable, the new system is no better. So, the prior distribution for the new system is identical to that of the old one at first. Then evidence begins to alter it, but conservatively – beliefs about the new system not being very reliable are slowly changed by the gathering failure-free evidence, but beliefs about the new system being very reliable are still, conservatively, identical to those for the old system (see Fig. 12). Even ardent skeptics can change their minds, but only with a lot of convincing. Possibly much more convincing than that required to change the old system's distribution when not being conservative. The result is that our conservative posterior expected *pfd* for the B system does improve with evidence, *but (much) more slowly than it would when not being conservative.*

5.3. Obtaining Confidence in an NWTES Assumption

Our objective in this paper is to present details of our new CBI/NWTES model. To aid readers' understanding we have presented various numerical examples, but we must emphasise that these have been chosen arbitrarily to represent scenarios in which we envisage this kind of reasoning being used. We do not claim that the actual numbers used in these examples are realistic for actual scenarios. Despite informal NWTES reasoning being used in practice, we are not aware of cases where a belief, c, in NWTES has been expressed quantitatively. Nevertheless, we believe that this may be feasible in some real cases. In what follows we discuss, informally, some situations in which NWTES beliefs would be natural, and where confidence, c, in NWTES would be less than 1.

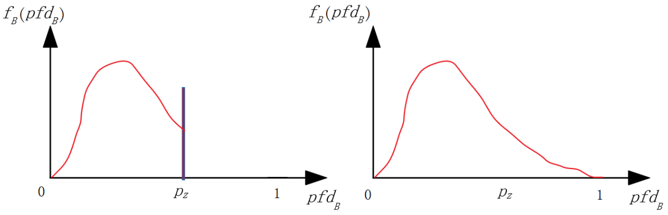

Figure 12: The worst-case achieving prior distribution for pfd_B in the NWTES model (l.h.s) and the optimistic model of using the A-system's distribution as the B-system's prior (r.h.s). See related joint distributions in Fig. 10.

- In a bug-fixing scenario, new software is essentially a modification of older, faulty software. One's confidence in perfect debugging leads to the NWTES belief of (4). Empirical evidence could be used to quantify such a belief, e.g. since the programmer intends to decrease the failure rate by their debugging actions [19], and empirically we know they are more likely to succeed rather than make things worse (implying $c \geqslant 0.5$ in (4)), one may choose $c = 0.5$ which, by a simple monotonicity analysis of our CBI results, gives conservative estimates.

- It may seem obvious that replacing a component of a software-based system with a more reliable one produces an upgraded new system that is at least as reliable as the old one. But there are exceptions e.g. if the new component introduces a *system design fault* [27, pp. 43–46], that is, the specification used for building the component is a misrepresentation of the behaviour actually required from it by the system [28]. The new component could be shown to be more reliable than the old one *according to its explicit specification*, and yet make the system less reliable. So, one's confidence in the absence of system design faults forms a belief in NWTES.

- Information about system architecture can provide clues to inform one's confidence in an NWTES assumption. For example, the new system could be obtained by augmenting the old system with additional runtime checks, safety checks, or monitoring channels [18, 29] so that some situations that would cause the old system to fail are tolerated by the new (augmented) system. Then, there will be strong confidence of the new system being more reliable than the old one (with respect to the kinds of failures mitigated), although c might not be 1 because of the small risk of having introduced system design faults.

- Advanced program analysis techniques can be used, e.g. "probabilistic symbolic execution" [30] can check a massive number of execution paths, symbolically, against a formal specification to estimate bounds on reliability. If one assumes such formal analysis is correct and can check all execution traces, and the results for the old and new software happen to show that the *pfd* of the new one is better, then one might claim 100% confidence in NWTES. However, due to inevitable uncertainties in any formal method for sufficiently complex software [31], instead of being certain, one may quantify one's confidence in the formal analysis, on the basis e.g., of statistics of the effectiveness of the verification tools [32].

Admittedly, quantifying NWTES beliefs can be an involved task, and it requires further study of rigorous probabilistic approaches to it. However, even when numbers are hard to come by, it might still be practical, and easier, to argue *qualitatively* for certainty in an NWTES assumption. Our NWTES model is suited for such situations (*cf.* the numerical examples when $c = 1$), and allows one to check the sensitivity of the reliability claim to the value of c, to avoid making wildly optimistic reliability claims on the basis of informal reasoning.

In passing, we note that the CBI worst-case achieving joint prior distributions *all* have the property that they produce posterior confidence in NWTES that *is no worse than prior confidence in* NWTES. And, the posterior confidence increases with increasing failure-free evidence[11]. Indeed, the "good news" of seeing the new system behave so well increases one's confidence that it is, at least, as reliable as the old one.

5.4. Future Work, including a Mathematical Dual with Applications to Software re-use

As indicated in our CBI review (section 2), there is a great deal of flexibility in both the beliefs that may be expressed, and the worst-case posterior estimates that may be sought, using CBI. We will explore many of these in future work, all within the "two-system" context presented in this paper. In general, we expect different combinations of beliefs and posterior estimates to "pick out" different worst-case priors, and show how being conservative can (significantly) change from situation to situation. For instance, if our CBI extension is to be used in the assessment framework proposed in [11, 15], then the posterior estimate of interest for the new B system becomes the probability of perfection given n failure-free executions by the system.

One might envisage multi-criterion optimisation, where multiple posterior estimates are simultaneously optimised for the new B system. For instance, how would one conservatively expect the new B system to have both a low *pfd* and a high probability of surviving t future executions?

So far, our work has focused on using knowledge about a single, older system; this can be extended to multiple older, similar systems. Or multiple past versions of the same software, each bringing valuable reliability information to bear on the new system's assessment. In fact, even older systems with multiversion architectures may be included in the assessment, likewise CBI could be used when upgrading one of the diverse software-channels in such a fault-tolerant architecture.

At present, CBI lacks explicit *feedback* mechanisms. But we know, just as older software can inform a reliability assessment of newer software, the reverse is true – evidence about (un)reliability in a newer system might cause an assessor to re-think conclusions made for an older system. Or additional (un)reliability evidence from the old system can arise, if the old system hasn't actually been replaced but is still in operation elsewhere. Moreover, beliefs expressed for these systems – such as the NWTES belief – should be amenable to change and updated over time. Especially if both systems are run side-by-side, and assessment is continuous and ongoing.

We will also explore other types of NWTES assumption. One possible alternative is that of *stochastically ordered pfd distributions*, analogous to how ordered distributions of failure rates were used in [19] to model the effects of imperfect bug-fixing. Indeed, if the new software is produced and verified using a new and improved development process – one which is likely to result in software with fewer faults across all fault types – then the cumulative distribution of the new system's *pfd* may be characterised as lying, everywhere, above that for the old one. Will this NWTES alternative overcome the shortcomings of NWTES as used in this paper? Or, do alternative forms of NWTES simply have their pros and cons, with no clear preference amongst them? In which case, one might consider using an *ensemble* of CBI models employing these NWTES alternatives, and developing techniques to evaluate how "good" the *pfd* estimates/forecasts coming from these models are. Such techniques, using the *principles of prequential statistics*, have been successfully developed in other contexts [33].

Our notion in this paper of two versions of a software-based system operating in a single environment is reminiscent of the probabilistic models for multiversion software in fault-tolerant systems, of some years ago. In [34] for example, several theorems are proved about the efficacy of such multiversion software architectures in delivering high reliability. These theorems exhibit a mathematical duality between the situation where many versions operate in a single environment, and the situation where

[11] *Proof*: for non-negative integers n, k, we wish to show that

$$\begin{aligned} & P(pfd_B \leqslant pfd_A \mid B \text{ passes } n+k \text{ tests}) \\ \geqslant\ & P(pfd_B \leqslant pfd_A \mid B \text{ passes } n \text{ tests})\,, \end{aligned}$$

which holds *iff* $\frac{\mathbb{E}[(1-Y)^{n+k}\mathbf{1}_{Y\leqslant X}]}{\mathbb{E}[(1-Y)^{n+k}]} \geqslant \frac{\mathbb{E}[(1-Y)^{n}\mathbf{1}_{Y\leqslant X}]}{\mathbb{E}[(1-Y)^{n}]}$ holds for random probabilities X and Y. Simplifying this by using the identity $\mathbb{E}[(1-Y)^r] = \mathbb{E}[(1-Y)^r\mathbf{1}_{Y\leqslant X}] + \mathbb{E}[(1-Y)^r\mathbf{1}_{Y>X}]$ with $r = n, n+k$, one obtains

$$\frac{\mathbb{E}[(1-Y)^{n+k}\mathbf{1}_{Y\leqslant X}]}{\mathbb{E}[(1-Y)^{n}\mathbf{1}_{Y\leqslant X}]} \geqslant \frac{\mathbb{E}[(1-Y)^{n+k}\mathbf{1}_{Y>X}]}{\mathbb{E}[(1-Y)^{n}\mathbf{1}_{Y>X}]}$$

which is the identity

$$\begin{aligned} & \mathbb{E}[(1-Y)^k \mid Y \leqslant X \,\&\, B \text{ passes } n \text{ tests}] \geqslant \\ & \mathbb{E}[(1-Y)^k \mid Y > X \,\&\, B \text{ passes } n \text{ tests}]\,, \end{aligned}$$

which is self-evidently true, since the function $(1-x)^k$ is a monotonically decreasing function on $[0, 1]$. ■

a single version operates in many environments. These are formally "*dual*", in the sense that for every theorem in the first model there is an exactly corresponding theorem in the second model – as if, for every scientific paper about the first model, there is another paper about the dual model that "writes itself".

Such a duality exists here too. Our CBI model concerns a single operational environment, and two programs/systems, one of which is believed to be no worse than the first. The dual situation would concern a single program/system, and two operational environments (one environment is believed to be no more "stressful" than the other). This situation does arise, we think, in some cases of software re-use: when a pre-existing system is re-used in a new environment that is believed to be no more stressful than the old one. We call this NWTEE (*No Worse Than Existing Environment*), to correspond to NWTES in the dual model in the current paper. A famous example of re-use concerned the inertial platform of the Ariane IV launch vehicle, which was re-used some years ago in the new Ariane V. It turned out that, for this inertial platform, the belief that Ariane V was NWTEE compared with Ariane IV was misplaced [28].

We believe that the mathematical results in the current paper apply directly to this dual situation of software re-use. Because of the ubiquity and importance of software re-use, and the widespread use of *off-the-shelf* components, these ideas seem worthy of further investigation.

In section 5.2 we showed why using the old system's *pfd* distribution as a prior for the new system's *pfd* is not being conservative. This warning, viewed through the dual NWTEE "lens", also shows why assuming that a changing environment has no significant effect on one's beliefs about a program's *pfd* can result in very optimistic posterior *pfd* estimates for the program's *pfd* in its new environment.

6. Conclusions

It is easy to see why using the historical development and operation of a pre-existing system, in the reliability assessment of a new system, is attractive – if little is known of the new system, this brings to bear the wealth of experience/evidence gained from a similar, pre-existing system.

But this comes with challenges, especially when trying to do this in a statistically principled way. For instance, Bayesian methods necessitate that an assessor adequately express their beliefs about the reliabilities of the old and new system, and how these reliabilities might be related. Not an easy thing to do, since such beliefs express uncertainty about unknown reliability measures, like the probability of a system failing on a random demand it receives from its environment (*pfd*).

In this regard, we present a novel extension of *conservative Bayesian inference* methods, here applied to this "two-system" assessment problem. Successfully used in "single-system" assessment work, CBI provides worst-case reliability estimates based on minimalist sets of beliefs expressed. As in previous CBI applications, we help assessors reason conservatively about the implications of their prior beliefs, using only partial prior knowledge of an old system's *pfd* distribution rather than fully specified priors. In particular, we study the implications of expressing a belief that the new system is "*not worse than the existing system*" (NWTES), and identify those beliefs that, when expressed together with NWTES, result in useful worst-case posterior estimates of the new system's reliability. We highlight the important role of failure-free testing evidence from the new system, and the need to express beliefs about the old system possibly being very reliable – or fault-free even – in order to gain useful conservative reliability estimates.

This work continues the authors' recent research [11–17], which has looked at ways that practitioners' plausible intuitions about the assessment of critical software-based systems can be formalised, and made more rigorous in support of quantitative claims about reliability and safety. Our CBI methods provide checks on whether apparently "obviously plausible" claims about system dependability can indeed be trusted. Our approach reveals circumstances in which such trust is inappropriate, and we provide ways forward for these cases. In summary, this paper makes the following contributions:

1. we formalise intuitive NWTES notions used by assessors in practice, and demonstrate the consequences of these for conservative reliability assessment;
2. we illustrate the conservative assessment of a software-based system subjected to operational testing, using a statistically principled incorporation of reliability evidence from a pre-existing, similar system;
3. we demonstrate how, with a justifiable (minimal) set of expressed prior beliefs, an assessor can use a conservative Bayesian approach to reliability assessment;
4. we extend CBI beyond previous "single system" applications;
5. we show how confidence in NWTES and claims about an older system's reliability severely limit, or strengthen, the impact of operational testing evidence on conservative posterior estimates of reliability;
6. we outline how these results may also be applied to the conservative reliability assessment of a system subject to a changing operational environment;
7. the work re-affirms the important role of claims about software perfection/fault-freeness to conservative assessment, found in previous CBI applictaions;
8. we demonstrate how CBI may be used to identify dangerously optimistic assessments. This includes seemingly reasonable, but ultimately naive, attempts at conservative reliability assessment.

Finally, whilst we have used the language of software faults in this paper, the results here, and in previous CBI work we have cited, may apply more widely to general design faults – for example, faults in the design of complex hardware/software systems. Our notions of fault-freeness, and thus "perfection", for example, seem equally applicable in these wider cases.

Acknowledgement

We are grateful to the anonymous reviewers of the paper, especially one whose detailed technical comments were helpful in improving the presentation and proofs. This work was supported in part via project AQUAS by the ECSEL Joint Undertaking, grant no. 737475.

References

[1] R. W. Butler and G. B. Finelli, "The infeasibility of quantifying the reliability of life-critical real-time software," *IEEE Trans. Softw. Eng.*, vol. 19, no. 1, pp. 3–12, Jan. 1993. [Online]. Available: https://doi.org/10.1109/32.210303

[2] B. Littlewood and L. Strigini, "Validation of ultra-high dependability for software-based systems," *Communications of the ACM*, vol. 36, pp. 69–80, 1993.

[3] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*. CRC press Boca Raton, FL, 2014, vol. 2.

[4] K. Pörn, "The two-stage Bayesian method used for the T-Book application," *Reliability Engineering & System Safety*, vol. 51, no. 2, pp. 169 – 179, 1996.

[5] C. Bunea, T. Charitos, R. M. Cooke, and G. Becker, "Two-stage Bayesian modelsapplication to ZEDB project," *Reliability Engineering & System Safety*, vol. 90, no. 2, pp. 123 – 130, 2005.

[6] C. Atwood, J. LaChance, H. Martz, D. Anderson, M. Englehardt, D. Whitehead, and T. Wheeler, "Handbook of parameter estimation for probabilistic risk assessment," U.S. Nuclear Regulatory Commission, Washington, DC, Report NUREG/CR-6823, 2003.

[7] B. Littlewood and D. Wright, "A Bayesian model that combines disparate evidence for the quantitative assessment of system dependability," in *Proceedings of the 14th International Conference on Computer Safety*. Belgirate, Italy: Springer London, 1995, pp. 173–188.

[8] US Food and Drug Administration, "The 510(k) program: Evaluating substantial equivalence in premarket notifications [510(k)] guidance for industry and food and drug administration staff," 2014.

[9] "Arrêté du 19 mars 2012 fixant les objectifs, les méthodes, les indicateurs de sécurité et la réglementation technique de sécurité et d'interopérabilité applicables sur le réseau ferré national, NOR:TRAT1208556A." [Online]. Available: https://www.legifrance.gouv.fr/eli/arrete/2012/3/19/TRAT1208556A/jo/texte

[10] European Committee for Electrotechnical Standardization, "EN 50126: railway applications – the specification and demonstration of reliability, availability, maintainability and safety (rams)," 2017.

[11] X. Zhao, B. Littlewood, A. Povyakalo, L. Strigini, and D. Wright, "Conservative claims for the probability of perfection of a software-based system using operational experience of previous similar systems," *Reliability Engineering & System Safety*, vol. 175, pp. 265 – 282, 2018.

[12] P. Bishop, R. Bloomfield, B. Littlewood, A. Povyakalo, and D. Wright, "Toward a formalism for conservative claims about the dependability of software-based systems," *IEEE Transactions on Software Engineering*, vol. 37, no. 5, pp. 708–717, 2011.

[13] L. Strigini and A. A. Povyakalo, "Software fault-freeness and reliability predictions," in *International Conference on Computer Safety, Reliability and Security*, vol. 8153. Springer, 2013, pp. 106–117.

[14] X. Zhao, B. Littlewood, A. Povyakalo, and D. Wright, "Conservative claims about the probability of perfection of software-based systems," in *26th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2015, pp. 130–140.

[15] X. Zhao, B. Littlewood, A. Povyakalo, L. Strigini, and D. Wright, "Modeling the probability of failure on demand (pfd) of a 1-out-of-2 system in which one channel is 'quasi-perfect'," *Reliability Engineering & System Safety*, vol. 158, pp. 230–245, 2017.

[16] X. Zhao, V. Robu, D. Flynn, F. Dinmohammadi, M. Fisher, and M. Webster, "Probabilistic model checking of robots deployed in extreme environments," in *The 33rd AAAI Conference on Artificial Intelligence (In Press)*, Honolulu, Hawaii, USA, 2019.

[17] X. Zhao, V. Robu, D. Flynn, K. Salako, and L. Strigini, "Assessing the safety and reliability of autonomous vehicles from road testing," in *the 30th Int. Symp. on Software Reliability Engineering (ISSRE)*. Berlin, Germany: IEEE, 2019, in press.

[18] (IEC) International Electrotechnical Commission, "(IEC) 61508: Functional safety of electrical/ electronic/programmable electronic safety related systems," 2010.

[19] B. Littlewood and J. L. Verrall, "A Bayesian reliability growth model for computer software," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 22, no. 3, pp. 332–346, 1973. [Online]. Available: http://www.jstor.org/stable/2346781

[20] A. Bertolino and L. Strigini, "Assessing the risk due to software faults: estimates of failure rate vs evidence of perfection," *Software Testing, Verification and Reliability*, vol. 8, no. 3, pp. 155–166, 1998.

[21] B. Littlewood and J. Rushby, "Reasoning about the reliability of diverse two-channel systems in which one channel is 'possibly perfect'," *IEEE Transactions on Software Engineering*, vol. 38, no. 5, pp. 1178–1194, 2012.

[22] J. Rushby, B. Littlewood, and L. Strigini, "Evaluating the assessment of software fault-freeness," in *Workshop Planning the Unplanned Experiment: Assessing the Efficacy of Standards for Safety Critical Software At the European Dependable Computing Conference*, Newcastle upon Tyne, May 2014.

[23] X. Zhao, "On the probability of perfection of software-based systems," PhD Thesis, City, University of London, 2016.

[24] B. Littlewood and D. Wright, "Some conservative stopping rules for the operational testing of safety critical software," *IEEE Transactions on Software Engineering*, vol. 23, no. 11, pp. 673–683, 1997.

[25] ——, "The use of multilegged arguments to increase confidence in safety claims for software-based systems: A study based on a bbn analysis of an idealized example," *IEEE Transactions on Software Engineering*, vol. 33(5), pp. 347 – 365, 2007.

[26] B. Littlewood and A. Povyakalo, "Conservative bounds for the pfd of a 1-out-of-2 software-based system based on an assessor's subjective probability of "not worse than independence"," *IEEE Transactions on Software Engineering*, vol. 39, no. 12, pp. 1641–1653, 2013.

[27] P. A. Lee and T. Anderson, *Fault Tolerance: Principles and Practice*, 2nd ed., J. C. Laprie, A. Avizienis, and H. Kopetz, Eds. Berlin, Heidelberg: Springer-Verlag, 1990.

[28] Ariane 501 Inquiry Board, "Ariane 5: Flight 501 failure," 1996. [Online]. Available: https://esamultimedia.esa.int/docs/esa-x-1819eng.pdf

[29] L. Strigini, "Fault tolerance against design faults," in *Dependable Computing Systems: Paradigms, Performance Issues, and Applications*, H. Diab and A. Zomaya, Eds. John Wiley & Sons, 2005, pp. 213 – 241. [Online]. Available: http://openaccess.city.ac.uk/278/

[30] A. Filieri, C. S. Psreanu, and W. Visser, "Reliability analysis in symbolic pathfinder," in *35th International Conference on Software Engineering*. IEEE, 2013, pp. 622–631.

[31] P. Fonseca, K. Zhang, X. Wang, and A. Krishnamurthy, "An empirical study on the correctness of formally verified distributed systems," in *12th European Conference on Computer Systems*. ACM, 2017, pp. 328–343.

[32] D. Beyer, "Software verification with validation of results," in *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*. Springer, 2017, pp. 331–349.

[33] P. Dawid, "Present position and potential developments: some personal views. statistical theory: the prequential approach (with discussion)," *Journal of the Royal Statistical Society (Series A)*, vol. 147, pp. 278 – 292, 1984.

[34] B. Littlewood and D. R. Miller, "Conceptual modeling of coincident failures in multiversion software," *IEEE Trans. Softw. Eng.*, vol. 15, no. 12, pp. 1596–1614, 1989. [Online]. Available: https://doi.org/10.1109/32.58771

[35] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis: a Hitchhiker's Guide*, 3rd ed. Berlin; London: Springer, 2006.

[36] M. Spivak, *Calculus On Manifolds: A Modern Approach To Classical Theorems Of Advanced Calculus*, ser. Mathematics monograph series. Avalon Publishing, 1971.

[37] R. L. Schilling, *Measures, Integrals and Martingales*. Cambridge University Press, 2005.

[38] V. Bogachev, *Weak Convergence of Measures*, ser. Mathematical Surveys and Monographs. American Mathematical Society, 2018. [Online].

Available: https://books.google.co.uk/books?id=g8bKuQEACAAJ
[39] R. M. Dudley, *Real Analysis and Probability*, 2nd ed., ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002.

Appendix A.

From Fig. 3, we know $M_4 = c - M_3 = c - (\alpha_A - M_2) = M_2 + c - \alpha_A$. And, we know $0 \leqslant M_2 \leqslant \min\{\alpha_A, 1 - c\}$. So, the range of M_4 is $\max\{0, c - \alpha_A\} \leqslant M_4 \leqslant \min\{c, 1 - \alpha_A\}$.

Appendix B.

Problem 1. *Consider the set $\mathcal{D}$ of all probability distributions over the unit square, each distribution representing a potential joint prior distribution of pfds for the A and B systems. Constrain the members of $\mathcal{D}$ as follows. Partition the unit square into four convex regions with known non-zero probabilities $M_i := P(\text{region } i)$ for $i = 1, \ldots, 4$, as shown in Fig. 3, such that $M_2 + M_3 = \alpha_A$ and $M_4 + M_3 = c$ are required to hold for known α_A, c. We seek to*[12]

$$\begin{aligned} &\underset{F \in \mathcal{D}}{\textit{maximise}} && \mathbb{E}[\, pfd_B \,|\, B \textit{ passes } n \textit{ tests}] \\ &\textit{subject to} && P(\textit{region } i) = \int_{\textit{region } i} \mathrm{d}F = M_i, \; i = 1, \ldots, 4 \end{aligned}$$

Solution: A prior $F^* \in \mathcal{D}$ maximises the objective function (see Fig. 6) and has an associated *pfd* value p_z satisfying (11). With F^*, the posterior expected *B*-system *pfd* (after observing *n* failure-free tests) is the upper bound S^* in (10).

Proof. The proof will progress in three stages:

1. First, the feasible set for the optimisation can be restricted from $\mathcal{D}$ to a smaller subset, $\mathcal{D}^* \subset \mathcal{D}$, of joint prior distributions that have discrete marginal distributions for the *B* system. Over $\mathcal{D}^*$, the optimisation becomes a constrained minimisation of $\Phi(w_1, w_2, w_3, w_4)$ – a rational function;

2. Secondly, with respect to each non-empty subset $\mathcal{W}$ of the variables $w_1, \ldots, w_4$, the function Φ is continuously differentiable and has a global minimum at the unique stationary point with respect to the variables in $\mathcal{W}$;

3. Lastly, from stages 1 and 2 we deduce a prior F^* that, in terms of a unique *pfd* p_z, maximises the objective function at the value S^*.

Let us proceed:

stage 1) Let *Y* be the unknown *pfd* for the *B* system. For any joint distribution in $\mathcal{D}$, the objective function can be written in terms of a quotient of expectations. By first expanding these expectations as conditional expectations – each conditional on one of the 4 regions of the unit square – and, then, bounding those conditional expectations in the numerator of the quotient,

$$\begin{aligned} \mathbb{E}[\, pfd_B \,|\, B \text{ passes } n \text{ tests}] &= 1 - \frac{\mathbb{E}[(1-Y)^{n+1}]}{\mathbb{E}[(1-Y)^n]} \\ &= 1 - \frac{\sum_i \mathbb{E}[(1-Y)^{n+1} \,|\, i\,] M_i}{\sum_i \mathbb{E}[(1-Y)^n \,|\, i\,] M_i} \\ &\leqslant 1 - \Phi(w_1, \ldots, w_4) \qquad \text{(B.1)} \end{aligned}$$

where $\Phi(w_1, \ldots, w_4) := \left(\frac{w_1^{\frac{n+1}{n}} M_1 + \ldots + w_4^{\frac{n+1}{n}} M_4}{w_1 M_1 + \ldots + w_4 M_4} \right)$, in which $w_i := \mathbb{E}\left[(1-Y)^n \,|\, i\,\right]$ is the probability of the *B* system surviving *n* tests, when the *B*-system *pfd* is some value from a point in the *i*-th region. The inequality in (B.1) follows from the relationship

$$\mathbb{E}[(1-Y)^{n+1} \,|\, i\,] \geqslant \left(\mathbb{E}[(1-Y)^n \,|\, i\,]\right)^{\frac{n+1}{n}} \qquad \text{(B.2)}$$

that holds for each region *i*. This is an application of either *Jensen's inequality* or *the monotonicity of L^p norms*[13].

Note that w_i, as a conditional expectation of a continuous random variable, satisfies[14] $w_i = (1 - y_i)^n$ for some point (x, y_i) in the region *i*, with y_i being unique. This allows us to define, for any given $F \in \mathcal{D}$, a related joint distribution that is also in $\mathcal{D}$ but with a discrete marginal distribution for the *pfd* of the B system[15], and the y_is are the possible *pfd* values for *Y*. Moreover, this related distribution, when used as a joint prior, gives a value for the objective function that is at least as bad as that resulting from *F*. Consequently, our optimisation task may now proceed by considering only those distributions in $\mathcal{D}$ that are of this kind – this is a subset of $\mathcal{D}$ we denote by $\mathcal{D}^*$. So, we will minimise Φ in (B.1) over $\mathcal{D}^*$.

stage 2) Now consider a non-empty subset $\mathcal{W}$ of the variables $w_1, \ldots, w_4$ and fix the values of those w_i that are not in $\mathcal{W}$. With respect to the w_i in $\mathcal{W}$, Φ is a rational function of continuously differentiable functions of these w_i. Therefore, Φ is continuously differentiable and its partial derivatives determine how Φ changes with respect to each $w_i \in \mathcal{W}$ at an arbitrary feasible "*point*". With respect to each variable $w_i \in \mathcal{W}$, the partial derivative of Φ is

$$\frac{\partial \Phi}{\partial w_i} = \frac{M_i \left(\frac{n+1}{n} w_i^{\frac{1}{n}} - \Phi \right)}{w_1 M_1 + \ldots + w_4 M_4}, \qquad \text{(B.3)}$$

[12] The integral in the constraints is with respect to a Lebesgue-Stieltjes measure defined over Borel sets of the unit square – each $F \in \mathcal{D}$ induces a Lebesgue-Steiltjes measure over the unit square.

[13] For each prior distribution $F \in \mathcal{D}$ and the Borel sigma-algebra $\mathcal{B}$ on the unit square, consider the probability space $([0,1]\times[0,1], \mathcal{B}, F)$. Then, for any random variable $X\colon [0,1]\times[0,1] \to \mathbb{R}$ and $1 \leqslant p < \infty$, one may compute the conditional expectation $\mathbb{E}[\,|X|^p \,|\, i\,]$, conditional on region *i* of the unit square.

Let us denote the *L^p-norm of X*, $\mathbb{E}[\,|X|^p \,|\, i\,]^{\frac{1}{p}}$, by $\|X\|_p$ for short. The *monotonocity of L^p norms* is the guarantee that, for $1 \leqslant r < q < \infty$, we have $\|X\|_q \geqslant \|X\|_r$ (see [35], page 463). In particular, for $r = n$, $q = n + 1$ and $X := (1 - Y)$, one has (as claimed in (B.2))

$$\|(1-Y)\|_{n+1} \geqslant \|(1-Y)\|_n \,.$$

[14] "Satisfies", because the *intermediate value theorem* applied over each convex region means the bounded, continuous function $(1-y)^n$ takes on all values between its maximum and minimum on each region, and the properties of expectations then imply that this bounded function must attain its expected value.

[15] This marginal distribution is $P(Y = y) = \sum_{i=1}^{4} M_i \mathbf{1}_{y=y_i}$

so that the sign of $\frac{\partial \Phi}{\partial w_i}$ is completely determined by the sign of $(\frac{n+1}{n} w_i^{\frac{1}{n}} - \Phi)$. That is, the sign of the numerator in (B.3) is a rule for how to change w_i in order to minimise Φ, and this rule is in terms of an explicit relationship between a feasible "*point*" $(w_1, \ldots, w_4)$ and the value of the objective function at that point, $\Phi(w_1, \ldots, w_4)$. At any given "*point*", there are three possible directions for changing $w_i \in \mathcal{W}$, resulting from the sign of $\frac{\partial \Phi}{\partial w_i}$:

1. Φ decreases with decreasing w_i *if, and only if,*
$$w_i^{\frac{1}{n}} > \frac{n}{n+1}\Phi(w_1, \ldots, w_4) \ ;$$
2. Φ decreases with increasing w_i *if, and only if,*
$$w_i^{\frac{1}{n}} < \frac{n}{n+1}\Phi(w_1, \ldots, w_4) \ ;$$
3. Φ is stationary *if, and only if,*
$$w_i^{\frac{1}{n}} = \frac{n}{n+1}\Phi(w_1, \ldots, w_4) \ .$$

We restate these possibilities in terms of B-system *pfd*s as follows. Since, for each i, we have $w_i = (1 - y_i)^n$ for some unique *pfd* y_i, the possibilities become:

1. Φ decreases with increasing *pfd* y_i *if, and only if,*
$$y_i < 1 - \frac{n}{n+1}\tilde{\Phi}(y_1, \ldots, y_4) \ ;$$
2. Φ decreases with decreasing *pfd* y_i *if, and only if,*
$$y_i > 1 - \frac{n}{n+1}\tilde{\Phi}(y_1, \ldots, y_4) \ ;$$
3. Φ is stationary *if, and only if,*
$$y_i = 1 - \frac{n}{n+1}\tilde{\Phi}(y_1, \ldots, y_4) \ ;$$

where, for notational convenience,
$$\tilde{\Phi}(y_1, \ldots, y_4) := \Phi\left((1 - y_1)^n, \ldots, (1 - y_4)^n\right) \qquad \text{(B.4)}$$

The existence of stationary points of Φ (with respect to the $\mathcal{W}$ variables) follows from the existence of zeroes of the rational functions $\frac{\partial \Phi}{\partial w_i}$. Rational functions have a finite number of zeroes and, consequently, Φ has at most a finite number of isolated stationary points. The convexity of Φ with respect to $\mathcal{W}$ follows from Φ having a unique stationary point at which it attains a global minimum; this fact can be deduced from the properties of Φ's Hessian matrix of second-order partial derivatives. At a stationary point, the Hessian for Φ is necessarily a diagonal matrix $\mathbf{diag}(\frac{(n+1)w_i^{1/n} M_i}{n^2 w_i \sum_j w_j M_j})$, where $w_i \in \mathcal{W}$. Clearly, this matrix is positive-definite for all reasonable w_i (and associated *pfd*s y_i)[16], since all the diagonal entries are positive and the off-diagonal entries are zero. So Φ must be a local minimum at each isolated stationary point. But isolated stationary points are not possible since, if we assume that there is more than one isolated local minimum, then any path between these local minima must pass through a stationary point at which the Hessian is not positive-definite; a contradiction. Hence, there are no isolated stationary points, and Φ has a global minimum at the unique stationary point with respect to the $\mathcal{W}$ variables. This shows Φ is convex with respect to $\mathcal{W}$. And thus, to minimise Φ, one picks an arbitrary feasible "*point*" and changes the $w_i \in \mathcal{W}$ in the direction of the global Φ minimum; a direction indicated by the signs of the partial derivatives $\frac{\partial \Phi}{\partial w_i}$ given in (B.3).

In particular, the minimum value of Φ is obtained at a feasible "*point*" $((1 - \widehat{y_1})^n, \ldots, (1 - \widehat{y_4})^n)$ such that: 1) each $\widehat{y_i}$ satisfies one of the partial derivative relationships, 2) moving away from this "*point*" increases Φ and, 3) at a unique *pfd* value p_z, achieved by some $\widehat{y_i}$,
$$p_z = 1 - \frac{n}{n+1}(1 - S^*) \qquad \text{(B.5)}$$
where S^* is the maximum value of the objective function.

stage 3) Since S^* must satisfy $0 \leqslant S^* \leqslant 1$, (B.5) implies that $0 \leqslant p_z \leqslant 1$. So, there is a line segment in the unit square such that Φ is at a minimum when each $\widehat{y_i}$ either equals p_z, or is as close to p_z as constraints in regions $1, \ldots, 4$ will allow.

Clearly (see Fig. B.13), both $\widehat{y_1}$ and $\widehat{y_3}$ can assume the value p_z. However, depending on the value of p_z and p_A, only one of $\widehat{y_2}$ or $\widehat{y_4}$ can reach the value p_z, while the other must take the value p_A since the point (p_A, p_A) will be the closest reachable point to the p_z-line. These two possibilities, shown in Fig. B.13, are two possible forms for a "most conservative" joint prior distribution F^*. And the corresponding two S^*s are
$$S^*_{LHS} = 1 - \frac{(1 - p_z)^{n+1}(M_1 + M_2 + M_3) + (1 - p_A)^{n+1} M_4}{(1 - p_z)^n (M_1 + M_2 + M_3) + (1 - p_A)^n M_4}\mathbf{1}_{p_z \geqslant p_A}$$
$$S^*_{RHS} = 1 - \frac{(1 - p_z)^{n+1}(M_1 + M_4 + M_3) + (1 - p_A)^{n+1} M_2}{(1 - p_z)^n (M_1 + M_4 + M_3) + (1 - p_A)^n M_2}\mathbf{1}_{p_z < p_A} \qquad \text{(B.6)}$$

However, the following argument shows that the "RHS" case is impossible. Using $S^* = S^*_{RHS}$ and $p_z = 1 - \frac{n}{n+1}\left(1 - S^*_{RHS}\right)$ (by (B.5)), we have:
$$\frac{n+1}{n} = \frac{(1 - p_z)^{n+1}(1 - M_2) + (1 - p_A)^{n+1} M_2}{(1 - p_z)^{n+1}(1 - M_2) + (1 - p_A)^n (1 - p_z) M_2} \qquad \text{(B.7)}$$

Since $\frac{n+1}{n} > 1$, the r.h.s. of (B.7) is also bigger than 1. This inequality implies that $p_z > p_A$ – a contradiction, since the condition $(p_z < p_A)$ must hold for $S^* = S^*_{RHS}$. Consequently, the only valid form for F^* is that in Fig. B.13 (a). ■

Some Remarks: This proof readily extends to any finite partition of the unit square into convex sets, and any finite number of pre-existing systems.

Appendix C.

Lemma 1. *Let the M_i, p_A, and p_z, be defined/constrained as in Appendix B. Then, p_z is a decreasing function of M_4.*

[16] "Unreasonable" is a *pfd* of 1 – i.e. the software always fails.

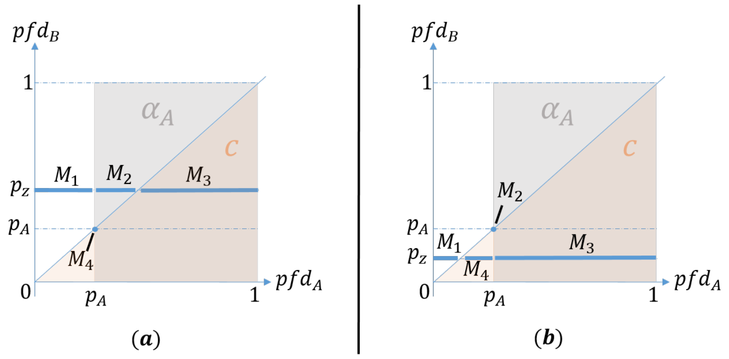

Figure B.13: Two potential forms for a "most conservative" joint prior F^*. When $p_z \geqslant p_A$, our choice of F^* takes the form **(a)** and gives a posterior value S^*_{LHS}. Otherwise, for $p_z < p_A$, we have the form **(b)** instead, with posterior value S^*_{RHS}. The probability mass in region i, i.e. M_i, concentrates uniformly on the portion of the p_z-line within the region, or at the point closest to the p_z-line within the region. Case **(b)** can be shown to be impossible, so our F^* must take the form **(a)**.

Proof. Appendix B shows $p_z \geqslant p_A$. Define (for $n \geqslant 1$)

$$Nu := (1-p_z)^{n+1}(1-M_4) + (1-p_A)^{n+1}M_4$$
$$De := (1-p_z)^{n}(1-M_4) + (1-p_A)^{n}M_4$$

so that, using these definitions, (10) and (11) imply

$$\frac{n+1}{n}(1-p_z) = \frac{Nu}{De} \tag{C.1}$$

Note, by Appendix B, the identity (C.1) holds for any set of M_i masses and their associated p_z value, where the M_is satisfy the "α_A" and "c" sum constraints.

Now, by the *implicit function theorem*, p_z is a continuously differentiable function of M_4 for $De > 0$. So, differentiating (C.1) w.r.t. M_4 gives

$$\frac{n+1}{n}\left(-\frac{\partial p_z}{\partial M_4}\right) = \frac{1}{De}\left(\frac{\partial Nu}{\partial M_4} - \frac{Nu}{De}\frac{\partial De}{\partial M_4}\right) \tag{C.2}$$

Since the M_i are constrained to ensure $De > 0$, (C.2) shows that the sign of the derivative $\frac{\partial p_z}{\partial M_4}$ is the "negative" of the sign of $\frac{\partial Nu}{\partial M_4} - \frac{Nu}{De}\frac{\partial De}{\partial M_4}$. To determine the sign of the r.h.s. of (C.2), observe that $\frac{\partial Nu}{\partial M_4}$ and $\frac{\partial De}{\partial M_4}$ evaluate as

$$\frac{\partial Nu}{\partial M_4} = (n+1)\left(-\frac{\partial p_z}{\partial M_4}\right)(1-M_4)(1-p_z)^n + (1-p_A)^{n+1} - (1-p_z)^{n+1}$$
$$\frac{\partial De}{\partial M_4} = n\left(-\frac{\partial p_z}{\partial M_4}\right)(1-M_4)(1-p_z)^{n-1} + (1-p_A)^n - (1-p_z)^n$$

With these partial derivatives of Nu, De, and the relationship (C.1), we may expand the expression within the brackets on the r.h.s. of (C.2) to obtain

$$\begin{aligned}\frac{\partial Nu}{\partial M_4} - \frac{Nu}{De}\frac{\partial De}{\partial M_4} &= -(1-p_z)^n\left((1-p_z) - \frac{n+1}{n}(1-p_z)\right)\\ &\quad + (1-p_A)^n\left((1-p_A) - \frac{n+1}{n}(1-p_z)\right)\\ &= -g(p_z) + g(p_A)\end{aligned} \tag{C.3}$$

where $g : [0,1] \to [-1,1]$ is an auxilliary function defined as

$$g(x) = (1-x)^n\left((1-x) - \frac{n+1}{n}(1-p_z)\right) \tag{C.4}$$

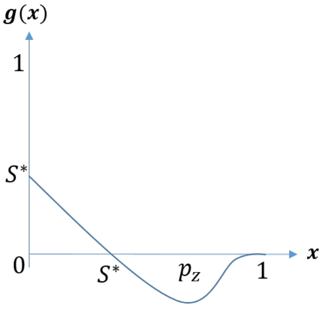

Figure C.14: The g function.

for fixed p_z.

The properties of g determine the sign we seek. In fact, g monotonically decreases to a stationary point over the range $0 \leqslant x \leqslant p_z$ and then monotonically increases over the range $p_z \leqslant x \leqslant 1$ (see Fig. C.14). This is because g is continuously differentiable over $(0,1)$, and its derivative

$$g'(x) = (n+1)(1-x)^{n-1}(x-p_z) \tag{C.5}$$

implies g's stated monotonic behaviours and unique minimum at p_z. Moreover, $g(S^*) = 0$, $g(1) = 0$, and $g(0) = 1 - \frac{n+1}{n}(1 - p_z) = S^* \geqslant 0$, where this inequality must hold in Appendix B for all worst-case priors[17]. Consequently,

$$\frac{\partial Nu}{\partial M_4} - \frac{Nu}{De}\frac{\partial De}{\partial M_4} = g(p_A) - g(p_z) \geqslant 0$$

which implies, from (C.2), that $\frac{\partial p_z}{\partial M_4} \leqslant 0$. That is, p_z is a decreasing function of M_4. ■

A Remark: The proof can be viewed as an argument that moves probability mass from M_4 (and consequently, M_2) to M_1

[17] Note that, since the objective function in the optimisation problem of Appendix B is an expected probability (so it must lie between 0 and 1), any worst-case prior distribution F^* (for consistent M_i) must give a worst-case value for the objective function, S^*, that satisfies $0 \leqslant 1 - S^* \leqslant 1$. And therefore, $0 \leqslant \frac{Nu}{De} \leqslant 1$ and $0 \leqslant \frac{n+1}{n}(1-p_z) \leqslant 1$ must also hold, since $1 - S^* = \frac{Nu}{De} = \frac{n+1}{n}(1-p_z)$ by (10) and (C.1).

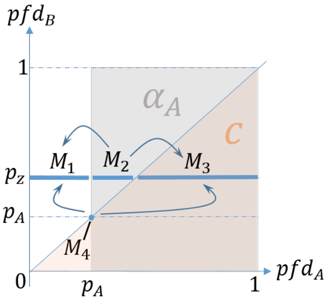

Figure C.15: For the worst-case prior from Appendix B, the constrained movement of probability mass – from M_4 (and consequently, M_2) to M_1 and M_3 – increases the mass lying on the p_z-line, while reducing the mass that does not. Hence, p_z should increase since, in this worst-case assignment of probability masses w.r.t p_z, the B-system has only become more likely to be less reliable.

and M_3 in a constrained manner (see Fig. C.15). Given that the closer the probability masses over the unit square are to lying on the horizontal line segment defined by p_z, the greater the value of p_z becomes[18], one can see that p_z *must* be a decreasing function of M_4. Because, reducing M_4 results in *all of the probability mass over the unit square either being as close, or closer, to the p_z-line*, than the masses were before the M_4 reduction. That is, reducing M_4 must increase M_1 (because $M_1 + M_4 = 1 - \alpha_A$), must increase M_3 (because $M_3 + M_4 = c$), and must reduce M_2 (because $M_2 + M_3 = \alpha_A$). Consequently, more of the probability masses from the various regions are closer to lying on the p_z-line segment.

Appendix D.

Problem 2. *Consider the set $\mathcal{D}$ of all probability distributions over the unit square, each distribution representing a potential joint prior distribution of pfds for the A and B systems. Constrain the members of $\mathcal{D}$ as follows. Partition the unit square into 6 sets: the origin and five convex regions, including the line segment $\{(0, y) : 0 \leqslant y \leqslant 1\}$. Let these sets have known non-zero probabilities $M_i := P(set\ i)$ for $i = 0, \ldots, 5$, as shown in Fig. 7. We require $M_3 + M_2 = \alpha_A$, $M_5 + M_0 = \theta_A$ and $M_5 + M_4 + M_3 = c$ to hold for known α_A, c, θ_A. We seek to*

$$\begin{aligned} &\underset{F \in \mathcal{D}}{\mathit{maximise}} && \mathbb{E}[\, pfd_B \,|\, B\ \mathit{passes}\ n\ \mathit{tests}] \\ &\mathit{subject\ to} && P(\mathit{set}\ i) = \int_{\mathit{set}\ i} \mathrm{d}F = M_i,\ i = 0, \ldots, 5 \end{aligned}$$

Solution: There is a prior $F^* \in \mathcal{D}$ that maximises the objective function. It is illustrated in Fig. 8, where the value of p_z satisfies (15). Upon using this prior F^*, the posterior expected B-system *pfd* (after observing n failure-free tests) achieves the upper bound S^*_{pp} in (14).

Proof. The proof follows an almost identical development to that given in Appendix B. However, now, the feasible priors F assign probability mass to the origin and the 1-dimensional convex set that is $\{(0, y) \,|\, 0 < y \leqslant 1\}$ (see Fig. 7). So, the objective function is bounded as follows:

$$\begin{aligned} \mathbb{E}[\, pfd_B \,|\, B \text{ passes } n \text{ tests}] &= 1 - \frac{\mathbb{E}[(1-Y)^{n+1}]}{\mathbb{E}[(1-Y)^n]} \\ &= 1 - \frac{\sum_{i=0}^{4} \mathbb{E}[(1-Y)^{n+1} \,|\, i\,] M_i + M_5}{\sum_{i=0}^{4} \mathbb{E}[(1-Y)^{n} \,|\, i\,] M_i + M_5} \\ &\leqslant 1 - \Phi\,(w_1, \ldots, w_4) \end{aligned}$$

where $\Phi\,(w_1, \ldots, w_4) := \left(\frac{\sum_{i=0}^{4} w_i^{\frac{n+1}{n}} \;\, M_i + M_5}{\sum_{i=0}^{4} w_i \, M_i + M_5} \right)$ and $w_i := \mathbb{E}\left[(1-Y)^n \;|\, i\,\right]$ is the probability of the B system surviving n tests, when the B-system *pfd* is some value from a point in the i-th set. The rest of the proof is identical[19] to that given in Appendix B.

Notice, from the definition of the w_i and the maximum vertical ranges in the various convex sets, we must have

$$\begin{aligned} 0 < y_0 \leqslant 1,\ 0 < y_1 \leqslant 1,\ 0 \leqslant y_3 \leqslant 1, \\ 0 \leqslant y_4 \leqslant p_A < y_2 \leqslant 1 \end{aligned} \tag{D.1}$$

so y_0, y_1 and y_3 can all reach the value p_z of (15) within their range, but only one of y_2 or y_4 can reach p_z in any given case. The two possible cases, shown in Fig. 8, have corresponding bounds (14) on the objective function. ■

Appendix E.

Lemma 2. *Let the M_i, p_A, and p_z, be defined and constrained as in Appendix D. Then, p_z is a decreasing function of M_5, for either fixed M_4 (when $p_z \geqslant p_A$) or fixed M_2 (when $p_z < p_A$).*

Proof. The proof is similar to that of Appendix C, but there are now two cases to consider. We show $\frac{\partial p_z}{\partial M_5} \leqslant 0$ in the first of these two cases – that of $p_z \geqslant p_A$ with a fixed M_4. The proof for the second case, that of a fixed M_2 and $p_z < p_A$, is essentially identical but with M_4 replaced by M_2.

To begin, define (for $n \geqslant 1$)

$$\begin{aligned} Nu &:= (1-p_z)^{n+1}(1 - M_4 - M_5) + (1-p_A)^{n+1} M_4 + M_5 \\ De &:= (1-p_z)^{n}(1 - M_4 - M_5) + (1-p_A)^{n} M_4 + M_5 \end{aligned}$$

so that, by using (14) and (15) together, we have

$$\frac{n+1}{n}(1-p_z) = \frac{Nu}{De} \tag{E.1}$$

As in Appendix C, the identity (E.1) holds for any set of M_i masses and their associated $p_z (\geqslant p_A)$ value, where the M_i satisfy the "α_A", "c" and "θ_A" sum constraints. And, similar to

[18] p_z should increase since, if it doesn't, this implies that the B-system has only become more likely to be less reliable, but without an increase in S^*_{LHS} – the worst-case posterior expected *pfd* upon observing no failures in n tests.

[19] Note, the "w_5" coefficients of M_5 are fixed at 1. Consequently, the function Φ is constant w.r.t. "w_5", so the M_5 terms in Φ add no further complications in using the arguments of Appendix B here.

Appendix C, the *implicit function theorem* implies p_z is continuously differentiable w.r.t. M_5 for $De > 0$. So, differentiating (E.1) w.r.t. M_5 gives

$$\frac{n+1}{n}\left(-\frac{\partial p_z}{\partial M_5}\right) = \frac{1}{De}\left(\frac{\partial Nu}{\partial M_5} - \frac{Nu}{De}\frac{\partial De}{\partial M_5}\right) \tag{E.2}$$

Since the M_i are feasible and constrained to ensure $De > 0$, (E.2) shows that the sign of $\frac{\partial p_z}{\partial M_5}$ is the "negative" of the sign of $\frac{\partial Nu}{\partial M_5} - \frac{Nu}{De}\frac{\partial De}{\partial M_5}$. To determine the sign of the r.h.s. of (E.2), observe that $\frac{\partial Nu}{\partial M_5}$ and $\frac{\partial De}{\partial M_5}$ evaluate as

$$\begin{aligned}\frac{\partial Nu}{\partial M_5} &= (n+1)\left(-\frac{\partial p_z}{\partial M_5}\right)(1-M_4-M_5)(1-p_z)^n + 1 - (1-p_z)^{n+1}\\ \frac{\partial De}{\partial M_5} &= n\left(-\frac{\partial p_z}{\partial M_5}\right)(1-M_4-M_5)(1-p_z)^{n-1} + 1 - (1-p_z)^n\end{aligned}$$

With these partial derivatives of Nu, De, and the relationship (E.1), we can rewrite the expression within the brackets on the r.h.s. of (E.2) to obtain

$$\frac{\partial Nu}{\partial M_5} - \frac{Nu}{De}\frac{\partial De}{\partial M_5} = -\,g(p_z) + g(0)$$

where g is the auxilliary function defined in (C.4). From the properties of g, $g(0) - g(p_z) \geqslant 0$ (see Fig. C.14, with S^* replaced by S^*_{pp}). So, by (E.2), $\frac{\partial p_z}{\partial M_5} \leqslant 0$; i.e. p_z decreases as M_5 increases.

The proof for the second case, where $p_z < p_A$ with fixed M_2, follows an identical argument. Define (for $n \geqslant 1$)

$$\begin{aligned}Nu &:= (1-p_z)^{n+1}(1-M_2-M_5) + (1-p_A)^{n+1}M_2 + M_5\\ De &:= (1-p_z)^n(1-M_2-M_5) + (1-p_A)^n M_2 + M_5\end{aligned}$$

And, upon using these definitions with (14) and (15), the proof proceeds like before to show that $\frac{\partial p_z}{\partial M_5} \leqslant 0$. ■

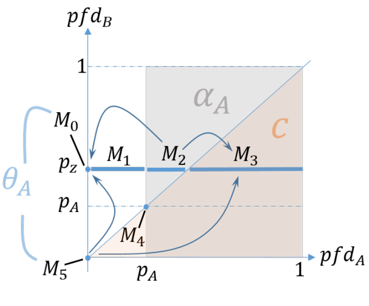

Figure E.16: For a worst-case prior from Appendix D, the constrained movement of probability mass – from M_5 (and consequently, M_2) to M_0 and M_3 – increases the mass lying on the p_z-line, while reducing the mass that does not. Hence, p_z should increase, since the B-system is now only more likely to be less reliable with the same p_z. Shown here is the worst-case prior when $p_z \geqslant p_A$.

Some Remarks:

1. The proof, e.g. when $p_z \geqslant p_A$ and M_4 is fixed, can be viewed as an argument that moves probability mass from M_5 (and consequently, M_2) to M_0 and M_3 in a constrained manner (see Fig. E.16), analogous to Appendix C. Like there, here, reducing M_5 increases the probability mass lying on the p_z-line while reducing the mass not touching the line. That is, reducing M_5 must increase M_0 (because $M_5 + M_0 = \theta_A$), must increase M_3 (because $M_3 + M_4 + M_5 = c$), and must reduce M_2 (because $M_2 + M_3 = \alpha_A$). And, since M_1 is fixed (because M_4 is, and $M_1 + M_4 = 1 - \theta_A - \alpha_A$), more of the probability masses from the various regions are closer to lying on the p_z-line segment. So, p_z should increase, because the B-system is now only more likely to be less reliable with the same p_z, so that the worst-case posterior expectation of it failing should increase. And, by (15), so too should p_z;

2. This "mass moving" viewpoint generalises (Appendix G), giving the sign of $\frac{\partial p_z}{\partial M_i}$ for any M_i. In fact, the proof above mirrors that in Appendix C as follows. The vertical "line" on the left edge of the unit square in Fig. E.16 – i.e. the union of region 0 (with mass M_0) and point "5" (with mass M_5) – is analogous to the vertical "strip" in Fig. C.15 comprised of region "2" (with mass M_2) and region "3" (with mass M_3). Referring to both of these as "strips", each strip has a fixed total mass (θ_A and $1 - \alpha_A$, respectively). And, mass from the lower half of each strip is either moved to the upper half of the strip (regions "0" and "1", respectively) or to the lower half of another strip lying to the right (region "3" in both cases). So, both $\frac{\partial p_z}{\partial M_4} \leqslant 0$ and $\frac{\partial p_z}{\partial M_5} \leqslant 0$ hold for nearly identical reasons.

Appendix F.

For fixed, constrained M_i, Appendix B and Appendix D give the forms of the worst-case posterior expected B-system *pfd*, and the discrete priors that ensure the attainment of these. But some M_i give worse (i.e. larger) posterior expected *pfd*s than others, even when these M_i are subject to the same constraints. In this appendix, we illustrate how to find those M_i values that give the worst value for S^*_{pp}. By considering the ordering $\alpha_A \leqslant 1 - \theta_A \leqslant c$ amongst the constraint parameters α_A, c and θ_A, we give the form for the value of the largest S^*_{pp} and its related M_i, all in terms of these parameters.

Lemmas 1 and 2 of Appendix C and Appendix E – and the constrained "probability-mass moving" operations that underlie them – are particular examples of a more general lemma that shows how to move probability mass between regions in the unit square to increase p_z. And, thereby, increase S^*_{pp} too, since $1 - p_z = \frac{n}{n+1}(1 - S^*_{pp})$ by (15). The general lemma proved in Appendix H, *a fortiori*, gives stronger justification for what we present here. However, for now, using the arguments and intuition from Appendix C and Appendix E, we will construct "worst-case achieving" M_i, starting from *any initial feasible* M_i.

In what follows we require $1 - \theta_A \geqslant \alpha_A$ for the joint prior distributions to be consistent, since $P(pfd_A > p_A) = \alpha_A$ must be smaller than $P(pfd_A > 0) = 1 - \theta_A$.

So, consider any feasible M_is constrained as in Appendix D, with an associated p_z value satisfying $p_z \geqslant p_A$ (so the S^*_{pp} for these M_i is given by (14) and (15) together). Also, suppose

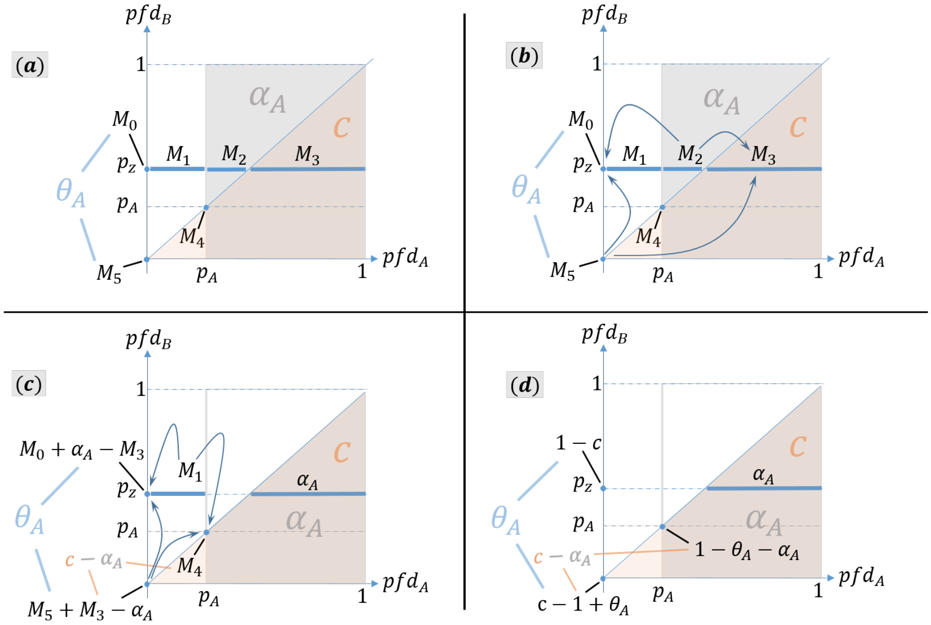

Figure F.17: Given α_A, θ_A and c, that satsify $\alpha_A \leqslant 1-\theta_A \leqslant c$, depicted here is a sequence of constrained probability mass movements to determine those values of $M_0, \ldots, M_5$ that give the largest value for posterior, expected B-system *pfd* S^*_{pp}. Begin with **(a)** an arbitrary collection of consistent probability masses $M_0, \ldots, M_5$, and their related worst-case prior distribution. Next, **(b)** transfer probability mass from M_5 and M_2 to M_0 and M_3 in a constrained manner. Then, once $M_3 = \alpha_A$, **(c)** transfer mass from M_5 and M_1 to M_0 and M_4 in a constrained manner, until $M_4 = 1-\theta_A-\alpha_A$, resulting in the most conservative prior distribution **(d)**.

that $\alpha_A \leqslant 1-\theta_A \leqslant c$. This requirement forces $M_5 \geqslant c-1+\theta_A$ since, otherwise, the parameters do not define consistent probabilities[20]. Now, if $M_5 > c-1+\theta_A$ then, by the reasoning of lemma 2, we may increase p_z by moving probability mass from M_5 to either M_4 (if $M_4 < 1-\theta_A-\alpha_A$) or M_3 (if $M_3 < \alpha_A$); at least one of these mass movements must be possible since, otherwise, the parameters do not define consistent probabilities[21]. We move all of the mass we can from M_5 until $M_5 = c-1+\theta_A$, at which point we must have $M_3 = \alpha_A$ and $M_4 = 1-\theta_A-\alpha_A$, otherwise, the parameters do not define consistent probabilities[22]. Fig. F.17 shows a sequence of mass moving operations from M_5, resulting in the worst-case S^*_{pp}, and the masses M_i that achieve it, as:

$$S^*_{pp_{LHS}} = 1 - \frac{(1-p_z)^{n+1}(1-c+\alpha_A) + (1-p_A)^{n+1}(1-\theta_A-\alpha_A) + c - 1 + \theta_A}{(1-p_z)^{n}(1-c+\alpha_A) + (1-p_A)^{n}(1-\theta_A-\alpha_A) + c - 1 + \theta_A} \qquad \text{(F.1)}$$

where this form of S^*_{pp} is given by using (14) (when $p_z \geqslant p_A$) with probability masses

$$M_0 = 1-c\,, \quad M_1 = 0\,, \quad M_2 = 0\,,$$
$$M_3 = \alpha_A\,, \quad M_4 = 1-\theta_A-\alpha_A\,, \quad M_5 = c-1+\theta_A \qquad \text{(F.2)}$$

Similar mass-moving arguments apply, if the initial M_is implied $p_z < p_A$ instead. The resulting worst-case M_is would still be (F.2). And, if these masses still implied $p_z < p_A$, then S^*_{pp} would now take the alternative form

$$S^*_{pp_{RHS}} = 1 - \frac{(1-p_z)^{n+1}(2-c-\theta_A) + c - 1 + \theta_A}{(1-p_z)^{n}(2-c-\theta_A) + c - 1 + \theta_A} \qquad \text{(F.3)}$$

obtained from (14) for the case when $p_z < p_A$. If (F.3) holds then it implies that $S^*_{pp_{RHS}} \xrightarrow{n\to\infty} 0$.[23] That is, the more failure-free runs are observed, the smaller the worst-case, conditional expected B-system *pfd* becomes. Table 3 illustrates this with examples where $p_z \leqslant p_A$ and p_z approaches 0 for ever larger n.

Appendix G.

Problem 3. *Consider the set $\mathcal{D}$ of all probability distributions over the unit square – each distribution is a joint prior distribution of pfds for the A and B systems. Constrain members of $\mathcal{D}$ as follows. $F \in \mathcal{D}$ has the continuous, marginal distribution for the A-system, denoted f_A, and F satisfies $P(pfd_B \leqslant pfd_A) = \int_{y\leqslant x} \mathrm{d}F = c$ for given c. We seek to*

$$\underset{F\in\mathcal{D}}{\mathit{maximise}} \quad \mathbb{E}[\, pfd_B \,|\, B \text{ passes } n \text{ tests}]$$

Solution: A prior $F^* \in \mathcal{D}$ maximises the objective function. Upon using F^*, the posterior expected B-system *pfd* (after observing n failure-free tests) achieves the upper bound S^*_{cmplt} in (17) with an associated, unique, p_z-value satisfying (18).

[20] *Proof*: If $M_5 < c-1+\theta_A$ is possible, then $c = M_3+M_4+M_5 < M_3+M_4+c-1+\theta_A$, so that $1-\theta_A < M_3+M_4$. But this, in turn, implies $1-\theta_A < M_3+M_4 \leqslant M_1+M_2+M_3+M_4 = 1-\theta_A$. So, $1-\theta_A < 1-\theta_A$; a contradiction. ■

[21] *Proof*: Suppose $M_5 > c-1+\theta_A$. And, by the constraints on the M_i, both $M_3 \leqslant \alpha_A$ and $M_4 \leqslant 1-\theta_A-\alpha_A$ hold. So, in particular, if both $M_3 = \alpha_A$ and $M_4 = 1-\theta_A-\alpha_A$ hold, then $c = M_3+M_4+M_5 = 1-\theta_A+M_5 > 1-\theta_A+c-1+\theta_A = c$. That is, $c > c$; a contradiction. ■

[22] *Proof*: Suppose $M_5 = c-1+\theta_A$. If either $M_4 < 1-\theta_A-\alpha_A$ or $M_3 < \alpha_A$, then $1-\theta_A = \alpha_A+1-\theta_A-\alpha_A > M_3+M_4 = c-M_5 = 1-\theta_A$. That is, $1-\theta_A > 1-\theta_A$; a contradiction. ■

[23] This limiting behaviour may also occur if (F.1) holds: note that (F.1) "becomes" (F.3) when $p_z = p_A$

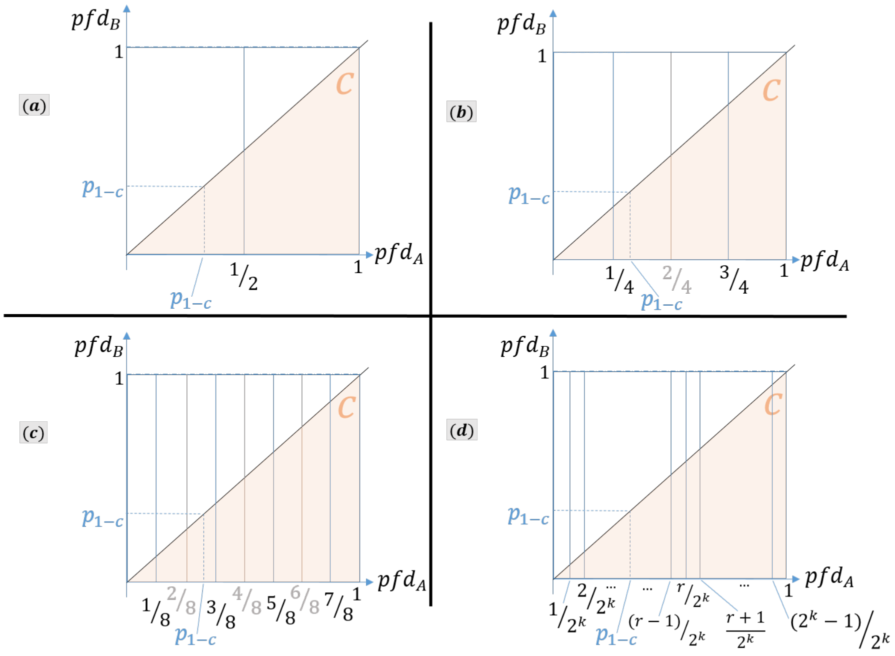

Figure F.18: A sequence of partitions of $[0, 1]$ consisting of dyadic rationals. Given the continuous marginal density for the A-system *pfd*, f_A, and the NWTES constraint parameter c, the value p_{1-c} is defined as the unique *pfd* value that satisfies $1 - c = \int\limits_0^{p_{1-c}} f_A(x)\mathrm{d}x$.

Proof. By considering a sequence of partition refinements over the interval $[0, 1]$, with their mesh sizes tending to zero[24], the proof involves a number of convergence arguments to deduce S^*_{cmplt} and p_z, deduce the forms (17) and (18), and deduce a joint prior F^* that achieves these. The stages for these arguments are:

1. First, for each partition k, the unit square is partitioned into equal vertical "strips", each strip having an associated probability mass defined by f_A. For any given collection M of probability masses over these "strips" – where the masses M are consistent with the f_A and NWTES constraints – there is a worst-case posterior expected *pfd* $S^*_{k,M}$ with a unique, associated *pfd* value $p_{z,k,M}$;
2. For each partition k, there is a worst-case achieving prior F^*_k with associated S^*_k ($\geqslant S^*_{k,M}$) and $p_{z,k}$ ($\geqslant p_{z,k,M}$) for all consistent collections of probability masses M;
3. Finally, since f_A and $(1 - x)^n$ are continuous functions over the compact set $[0, 1]$, and since the endpoints of the intervals in the dyadic partitions form a dense subset of $[0, 1]$, we prove the following:
   (a) the existence of p_z and $p_{z,k} \overset{k\to\infty}{\longrightarrow} p_z$;
   (b) the existence of S^*_{cmplt} and $S^*_k \overset{k\to\infty}{\longrightarrow} S^*_{cmplt}$;
   (c) the aforementioned limits prove (17) and (18);
   (d) with a guessed, feasible pre-(probability)measure F^* defined on the *semi-algebra* of half-open rectangles on the unit square (denoted $\mathcal{R}$), we show that $F^*_k(R) \overset{k\to\infty}{\longrightarrow} F^*(R)$ for $R \in \mathcal{R}$. That is, the limit of the sequence of conservative priors F^*_k agrees with the pre-measure F^* on $\mathcal{R}$. Therefore, such agreement in the limit also holds – between the limit of the conservative priors and the unique extension of F^* – on the *Borel sigma-algebra* generated by $\mathcal{R}$, $\sigma(\mathcal{R})$. Existence and uniqueness[25] of F^*'s extension is guaranteed by *Carathéodory's extension theorem*, and shows the *weak convergence* of the prior probability measures F^*_k to the unique prior probability measure in $\mathcal{D}$ that is the extension of F^*. This prior satisfies the constraints on members of $\mathcal{D}$.

Let us proceed:

stage 1) First, some preliminaries. Consider a sequence of *dyadic partitions* of the horizontal axis of pfd_A-values in the unit square. The k-th dyadic partition divides the axis into 2^k intervals of equal length $1/2^k$ (see Fig. F.18). This induces a partition of the unit square into 2^k vertical "strips"[26]. For any

[24] The mesh size, of a partition of $[0, 1]$ into sub-intervals, is the length of the largest of these sub-intervals.

[25] A clarification on uniqueness. F^*'s extension to $\sigma(\mathcal{R})$ is unique – any other prior distribution in $\mathcal{D}$ that agrees with F^* on $\mathcal{R}$ must also agree with F^* on $\sigma(\mathcal{R})$. However, F^* is not unique as a "worst-case achieving" prior distribution – other priors that differ from F^* on (Lebesgue) null sets also achieve S^*_{cmplt}.

[26] Our use of dyadic partitions is w.l.o.g. – *any* sequence of partition refinements over $[0, 1]$ with mesh sizes tending to zero will do. However, dyadic

feasible prior in $\mathcal{D}$, the masses associated with these strips are constrained in the following two ways:

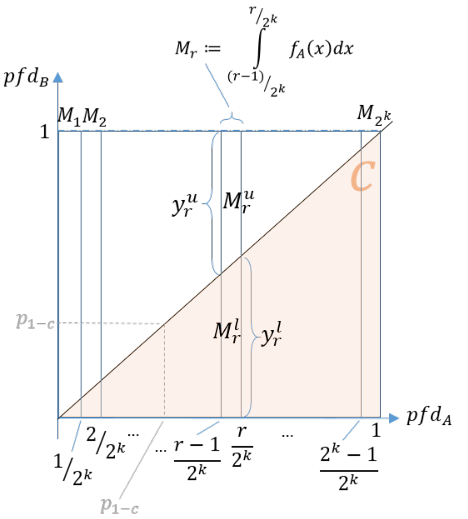

Figure G.19: The k-th dyadic partition of the pfd_A-axis into 2^k intervals of length $\frac{1}{2^k}$. A joint prior distribution will associate probabilities M_r^u and M_r^l with the upper and lower parts, respectively, within the vertical rectangular strip associated with the rth interval of the partition. The total probability for the r-th strip is $M_r := \int_{(r-1)/2^k}^{r/2^k} f_A(x)\mathrm{d}x$; that is, $M_r = M_r^u + M_r^l$. Also shown are the ranges of values for y_r^u and y_r^l in the r-th strip (see (G.3)).

1. the continuous, marginal A-system *pfd* distribution, f_A, implies that interval r must have an associated probability mass (see Fig. G.19)

$$M_r = \int_{(r-1)/2^k}^{r/2^k} f_A(x)\mathrm{d}x \tag{G.1}$$

   Given each M_r, a feasible $F \in \mathcal{D}$ allocates mass M_r^u to the region of the strip above the primary diagonal of the unit square, and mass M_r^l below the diagonal, and these satisfy $M_r = M_r^l + M_r^u$. Of course, $\sum_{i=1}^{2^k} M_i = 1$ holds;

2. the NWTES assumption implies that a total probability mass c must be associated with the region below the primary diagonal of the unit square. Note that there exists a unique *pfd* value p_{1-c} satisfying

$$1 - c = \int_0^{p_{1-c}} f_A(x)\mathrm{d}x \tag{G.2}$$

Now p_{1-c} lies in one of the 2^k intervals of the k-th partition, denoted interval k_{1-c}. But *which* interval it lies in, i.e. what the value of k_{1-c} is, depends on k. This is because the $(k+1)$-th partition is a *refinement* of the k-th partition – it has intervals with endpoints that consist of all of the endpoints of the intervals in the k-th partition, as well as the midpoints between them. A changing k_{1-c} is illustrated in Fig. G.21, where $k_{1-c} = r$ for the k-th partition, and $k_{1-c} = 2r - 1$ for the $(k+1)$-th partition.

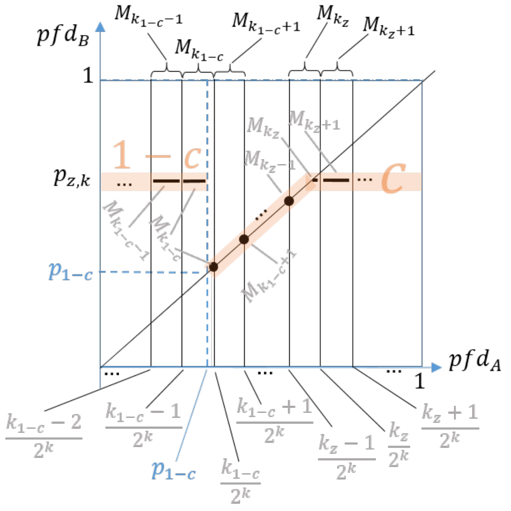

Figure G.20: Each dyadic partition has an associated worst-case joint prior distribution, with mass allocation given by 1) first, moving all of the NWTES probability mass c (assigned to the area underneath the diagonal) to lie on the right. And then, 2) ensuring that all of the mass in each vertical strip lies as close as possible to, or uniformly on, the $p_{z,k}$-line. The r-th strip is constrained by (G.1) to have associated mass M_r. Note that interval k_{1-c} is defined to contain the *pfd* value p_{1-c}. The mass for the k_{1-c}-th strip is split – some mass is allocated to the diagonal to ensure the mass for the region underneath the diagonal equals c, and the remaining mass for this strip is allocated to the area above the diagonal (still within the k_{1-c} strip and to the left of the p_{1-c} vertical line) to lie uniformly on the $p_{z,k}$-line. Also depicted is the interval k_z which, for the k-th partition, is defined to contain the *pfd* value $p_{z,k}$.

The optimisation can now begin. For the k-th partition, choose an arbitrary feasible prior $F \in \mathcal{D}$, and its allocation of constrained masses M_i, M_i^u, M_i^l (for $1 \leqslant i \leqslant 2^k$) to the strips of the partition. Using identical arguments to those in Appendix B and Appendix D, restrict the optimisation from $\mathcal{D}$ to a subset $\mathcal{D}^*$ (of discrete marginal pfd_B distributions), and bound the objective function over $\mathcal{D}^*$ as follows:

$$\mathbb{E}[pfd_B \mid B \text{ passes } n \text{ tests}] \leqslant 1 - \frac{\sum_{i=1}^{2^k}\left[\left(1-y_i^u\right)^{n+1} M_i^u + \left(1-y_i^l\right)^{n+1} M_i^l\right]}{\sum_{i=1}^{2^k}\left[\left(1-y_i^u\right)^{n} M_i^u + \left(1-y_i^l\right)^{n} M_i^l\right]} \tag{G.3}$$

where y_i^u lies in the maximum vertical range above the unit-square diagonal and within the rectangular strip for the ith interval. The range for y_i^l is similarly defined (see Fig. G.19).

The aim now is to bound the r.h.s. of (G.3). As in previous appendices[27], convexity analysis reveals the maximum posterior expected *pfd* $S_{k,M}^*$, as well as a conservative prior distribution $F_{k,M}^* \in \mathcal{D}^*$ that attains it, and an associated unique *pfd* value $p_{z,k,M}$, all of which satisfy

partitions have the added notational convenience that by stating "let $k \to \infty$" one invokes a countable sequence of partition refinements, as well as indicating that these refinements have mesh sizes $\frac{1}{2^k}$ exponentially tending to zero.

[27] The symbols k and M in the subscripts of $S_{k,M}^*$, $F_{k,M}^*$ and $p_{z,k,M}$ remind the reader that the arguments from previous appendices apply here to each partition k, and for each suitably constrained initial collection M of probability masses.

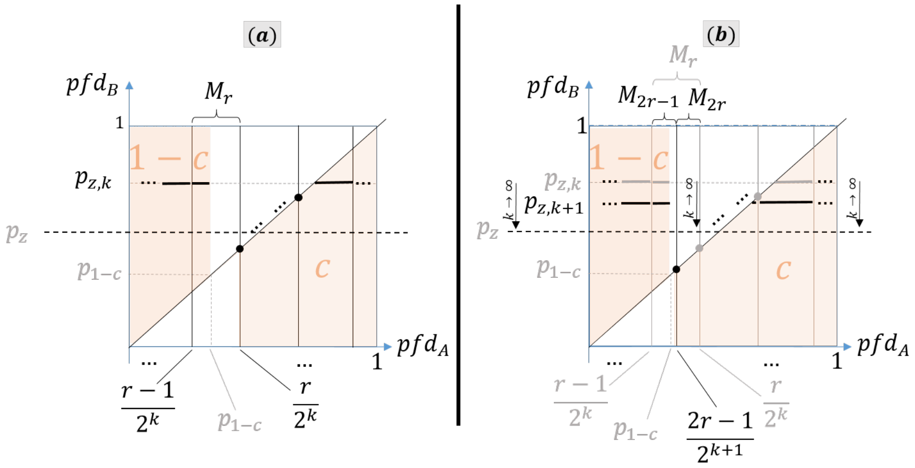

Figure G.21: Refinements of partitions can only reduce the value $p_{z,k}$. A refinement of **(a)** the k-th partition *forces* **(b)** more probability mass to lie further *below* the $p_{z,k}$-line. For example, the refinement forces part of the M_r mass – i.e. M_{2r-1} – to lie further below the $p_{z,k}$-line. Thus, as the dyadic partitions are refined (as $k \to \infty$), the $p_{z,k}$-line moves downwards, because with each refinement the unknown B-system *pfd* becomes only more likely to be more reliable. And the *completeness of the real numbers* guarantees that this bounded monotonically reducing sequence of probabilities $p_{z,k}$ converge to some p_z. That is, $p_{z,k} \to p_z$, and the horizontal $p_{z,k}$-line tends to the horizontal p_z-line.

$$p_{z,k,M} = 1 - \frac{n}{n+1}\left(1 - S^*_{k,M}\right) \qquad \text{(G.4)}$$

$$1 - S^*_{k,M} = \frac{\Phi(p_{z,k,M},\, n+1) + \Psi(p_{z,k,M},\, n+1)}{\Phi(p_{z,k,M},\, n) + \Psi(p_{z,k,M},\, n)} \qquad \text{(G.5)}$$

where

$$\Phi(p,n) := \sum_{i=1}^{k_z}\left[(1-p)^n M^u_i + (1-y_i)^n M^l_i\right],$$

$$\Psi(p,n) := \sum_{i=k_z+1}^{2^k}\left[(1-y_i)^n M^u_i + (1-p)^n\, M^l_i\right],$$

y_i is the B-system *pfd* value closest to $p_{z,k,M}$, on the diagonal in the i-th strip. In fact, $y_i = \frac{i}{2^k}$ when $1 \leqslant i \leqslant k_z$ and $y_i = \frac{i-1}{2^k}$ when $k_z + 1 \leqslant i \leqslant 2^k$. And the k_z-th interval contains $p_{z,k,M}$.

stage 2) Now, $F^*_{k,M}$ gives a worse posterior expected B-system *pfd* than our initial choice of prior F, *while having the same upper (i.e. M^u_i) and lower (i.e. M^l_i) mass allocations as F for each strip*. We now seek a most conservative prior amongst all $F^*_{k,M}$ conservative priors – one that remains consistent with the allocation of masses M_i for the strips of the k-th dyadic partition. Appendix H proves that, given any two distinct intervals of the partition, a and b say ($a < b$), the constrained movement of mass allocated by a feasible prior distribution $F^*_{k,M}$ – from M^l_a, M^u_b to M^u_a, M^l_b while keeping all other probability masses fixed – increases $p_{z,k,M}$. And, by (G.4), increases $S^*_{k,M}$ as well. This is true for any such conservative prior $F^*_{k,M}$. Consequently, by the constrained movement to the right, of all of the probability mass c below the unit-square diagonal, one constructs the most conservative prior F^*_k consistent with the k-th partition (see Fig. G.20). F^*_k has an associated unique *pfd* value $p_{z,k}$ and maximum posterior expected *pfd*, S^*_k, that satisfy

$$p_{z,k} = 1 - \frac{n}{n+1}\left(1 - S^*_k\right) \qquad \text{(G.6)}$$

$$1 - S^*_k = \frac{\Phi(p_{z,k},\, n+1) + \Psi(p_{z,k},\, n+1)}{\Phi(p_{z,k},\, n) + \Psi(p_{z,k},\, n)} \qquad \text{(G.7)}$$

where

$$M^u_{k_{1-c}} := 1 - c - \sum_{i=1}^{k_{1-c}-1} M_i\,,$$

$$\Phi(p,n) := \sum_{i=1}^{k_{1-c}-1}(1-p)^n M_i + (1-p)^n M^u_{k_{1-c}}\,,$$

$$\Psi(p,n) := \sum_{i=k_{1-c}+1}^{k_z-1}(1 - {}^{i}\!/\!{}_{2^k})^n M_i + \sum_{i=k_z}^{2^k}(1-p)^n\, M_i$$
$$+\ (1 - {}^{k_{1-c}}\!/\!{}_{2^k})^n\left(M_{k_{1-c}} - M^u_{k_{1-c}}\right), \qquad \text{(G.8)}$$

and the k_{1-c}-th interval of the k-th partition contains the *pfd* p_{1-c}, while the k_z-th interval contains $p_{z,k}$. Note that, like k_{1-c}, the value of k_z varies between partitions.

stage 3) Using this sequence of priors F^*_k (i.e. one such prior for each partition), what follows are convergence arguments to obtain the worst-case posterior expected *pfd* S^*_{cmplt}, its associated p_z, and the prior F^* that attains it, all in the limit as $k \to \infty$.

3.a) The following observations will be useful for the convergence arguments. These observations are a consequence of the dyadic rationals in $[0, 1]$ being a dense subset.

- As $k \to \infty$, the width of successive intervals k_{1-c} tend to zero. That is, the sequence of left endpoints tends to the sequence of right endpoints – sandwiching the value p_{1-c} inbetween them (see Fig. G.22).

$$\frac{k_{1-c}-1}{2^k} \longrightarrow p_{1-c} \longleftarrow \frac{k_{1-c}}{2^k} \quad (\text{as } k \to \infty) \qquad \text{(G.9)}$$

- A partition refinement from k to $k + 1$ does not change the amount of probability mass allocated to the horizontal $p_{z,k}$-line. It does, however, move some mass on the diagonal further away from that line (see Fig. G.21). This

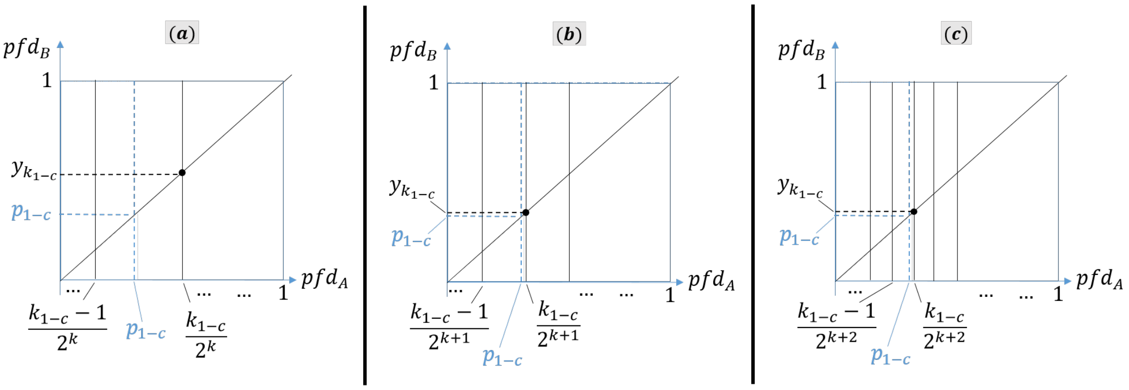

Figure G.22: As the partitions get refined from **(a)** through to **(c)**, the endpoints of the k_{1-c} intervals tend to one another, sandwiching p_{1-c} inbetween them. Thus, $\frac{k_{1-c}-1}{2^k} \longrightarrow p_{1-c} \longleftarrow \frac{k_{1-c}}{2^k}$ (as $k \to \infty$). The value of k_{1-c} varies across partitions, as implied by these pictures.

makes the B-system only more likely to be more reliable, but with the same $p_{z,k}$. Therefore, the expected B-system *pfd* reduces, thus reducing the posterior expectation $S^*_{k,z}$ of the B-system failing[28]. And hence, by (G.6), $p_{z,k}$ must reduce. Since the monotonically reducing sequence of probabilities $p_{z,k}$ cannot be smaller than zero, by the *completeness of the real numbers* there is a limiting probability p_z that, itself, cannot be smaller than zero. That is,

$$p_{z,k} \overset{k\to\infty}{\longrightarrow} p_z. \tag{G.10}$$

Moreover, interval widths uniformly tend to zero as $k \to \infty$, so the two sequences of endpoints of the k_z-th intervals tend to one another, sandwiching the converging $p_{z,k}$ inbetween them. From (G.10),

$$\frac{k_z - 1}{2^k} \longrightarrow p_z \longleftarrow \frac{k_z}{2^k} \quad (\text{as } k \to \infty) \tag{G.11}$$

We now proceed with the main convergence arguments.

3.b) We show that S^*_{cmplt} exists and $S^*_k \overset{k\to\infty}{\longrightarrow} S^*_{cmplt}$. Consider each partition k. By the *mean value theorem for integrals*, the continuity of f_A on $[0, 1]$ implies

$$M_i = \int_{(i-1)/2^k}^{i/2^k} f_A(x)\mathrm{d}x = \frac{1}{2^k} f_A(x^*)$$

for some $x^* \in [{}^{(i-1)}\!/_{2^k}, {}^{i}\!/_{2^k}]$. The continuity of f_A implies it is bounded on each interval $[{}^{(i-1)}\!/_{2^k}, {}^{i}\!/_{2^k}]$ of the partition and attains its lower and upper bounds, $f_A(x^i_{\min})$ and $f_A(x^i_{\max})$ say, for some $x^i_{\min}, x^i_{\max} \in [{}^{(i-1)}\!/_{2^k}, {}^{i}\!/_{2^k}]$ so that

$$\frac{1}{2^k} f_A(x^i_{\min}) \leqslant \frac{1}{2^k} f_A(x^*) \leqslant \frac{1}{2^k} f_A(x^i_{\max}) \tag{G.12}$$

Similarly, the continuous function $(1 - x)^n f_A(x)$ over $[0, 1]$ is bounded and attains its bounds, $(1 - x^i_{\min})^n f_A(x^i_{\min})$ and $(1 - x^i_{\max})^n f_A(x^i_{\max})$ say[29], on interval i of the partition.

So, consider the following upper bounds on (G.8),

$$\begin{aligned}
\Phi^u(p, n) &:= \sum_{i=1}^{k_{1-c}-1} (1-p)^n \frac{1}{2^k} f_A(x^i_{\max}) + (1-p)^n M^u_{k_{1-c}} \\
\Psi^u(p, n) &:= \sum_{i=k_{1-c}+1}^{k_z-1} \left(1 - x^i_{\max}\right)^n \frac{1}{2^k} f_A(x^i_{\max}) + \sum_{i=k_z}^{2^k} (1-p)^n \frac{1}{2^k} f_A(x^i_{\max}) \\
&\quad + \left(1 - x^{k_{1-c}}_{\max}\right)^n \left(\frac{1}{2^k} f_A(x^{k_{1-c}}_{\max}) - M^u_{k_{1-c}}\right)
\end{aligned} \tag{G.13}$$

and lower bounds,

$$\begin{aligned}
\Phi^l(p, n) &:= \sum_{i=1}^{k_{1-c}-1} (1-p)^n \frac{1}{2^k} f_A(x^i_{\min}) + (1-p)^n M^u_{k_{1-c}} \\
\Psi^l(p, n) &:= \sum_{i=k_{1-c}+1}^{k_z-1} \left(1 - x^i_{\min}\right)^n \frac{1}{2^k} f_A(x^i_{\min}) + \sum_{i=k_z}^{2^k} (1-p)^n \frac{1}{2^k} f_A(x^i_{\min}) \\
&\quad + \left(1 - x^{k_{1-c}}_{\min}\right)^n \left(\frac{1}{2^k} f_A(x^{k_{1-c}}_{\min}) - M^u_{k_{1-c}}\right)
\end{aligned} \tag{G.14}$$

Using these bounds we may bound (G.7) for each partition, and take limits as $k \to \infty$, so

$$\begin{aligned}
&\lim_{k\to\infty} \frac{\Phi^l(p_{z,k},\, n+1) + \Psi^l(p_{z,k},\, n+1)}{\Phi^u(p_{z,k},\, n) + \Psi^u(p_{z,k},\, n)} \\
&\leqslant \lim_{k\to\infty} 1 - S^*_k \\
&\leqslant \lim_{k\to\infty} \frac{\Phi^u(p_{z,k},\, n+1) + \Psi^u(p_{z,k},\, n+1)}{\Phi^l(p_{z,k},\, n) + \Psi^l(p_{z,k},\, n)}
\end{aligned} \tag{G.15}$$

Each of the limits in the bounds in (G.15) converge:

[28] *Claim*:

$$\mathbb{E}[\, pfd_B |\, B \text{ passes } n \text{ tests}] \leqslant \mathbb{E}[pfd_B]$$

Proof: Apply *Jensen's inequality* twice to the convex function f, defined over $[0, 1]$ as $f(x) = x^r$ (for $r > 1$), as follows:

$$\begin{aligned}
\mathbb{E}[\, pfd_B |\, B \text{ passes } n \text{ tests}] &= \frac{\mathbb{E}[Y(1-Y)^n]}{\mathbb{E}[(1-Y)^n]} = 1 - \frac{\mathbb{E}[(1-Y)^{n+1}]}{\mathbb{E}[(1-Y)^n]} \\
&= 1 - \frac{\mathbb{E}[((1-Y)^n)^{\frac{n+1}{n}}]}{\mathbb{E}[(1-Y)^n]} \leqslant 1 - \frac{(\mathbb{E}[(1-Y)^n])^{1+\frac{1}{n}}}{\mathbb{E}[(1-Y)^n]} \\
&= 1 - (\mathbb{E}[(1-Y)^n])^{\frac{1}{n}} \leqslant 1 - ((\mathbb{E}[1-Y])^n)^{\frac{1}{n}} \\
&= 1 - (1 - \mathbb{E}[Y]) = \mathbb{E}[Y] = \mathbb{E}[pfd_B] \qquad \blacksquare
\end{aligned}$$

[29] These $x^i_{\min}$ and $x^i_{\max}$ are not, in general, the ones in (G.12), nor do they necessarily minimise/maximise the function $(1 - x)^{n+1} f_A(x)$ over $[0, 1]$. However, it is notationally convenient to use these symbols to indicate those *pfd* values at which the respective continuous functions they appear in attain their extrema.

- $M^{u}_{k_{1-c}} \xrightarrow{k\to\infty} 0$, since $0 \leqslant M^{u}_{k_{1-c}} \leqslant \frac{1}{2^k} f_A(x^{k_{1-c}}_{\max})$ and, by (G.9), $x^{k_{1-c}}_{\max} \xrightarrow{k\to\infty} p_{1-c}$;
- the integrability of f_A, the continuity of $(1-x)^n$, (G.9) and (G.10), altogether imply that

  $$(1-p_{z,k})^n \sum_{i=1}^{k_{1-c}-1} \frac{1}{2^k} f_A(x^i_{\min}) \xrightarrow{k\to\infty} (1-p_z)^n \int_0^{p_{1-c}} f_A(x)\mathrm{d}x$$

  And $(1-p_{z,k})^n \sum_{i=1}^{k_{1-c}-1} \frac{1}{2^k} f_A(x^i_{\max})$ also tends to the same limit[30]. Similarly,

  $$(1-p_{z,k})^n \sum_{i=k_z}^{2^k} \frac{1}{2^k} f_A(x^i_{\min}) \xrightarrow{k\to\infty} (1-p_z)^n \int_{p_z}^{1} f_A(x)\mathrm{d}x\,,$$

  and $(1-p_{z,k})^n \sum_{i=k_z}^{2^k} \frac{1}{2^k} f_A(x^i_{\max})$ has the same limit;
- the continuity of $(1-x)^n f_A(x)$ over $[0,1]$ and (G.11), together ensure the convergence

  $$\sum_{i=k_{1-c}}^{k_z-1} \left(1 - x^i_{\min}\right)^n \frac{1}{2^k} f_A(x^i_{\min}) \xrightarrow{k\to\infty} \int_{p_{1-c}}^{p_z} (1-x)^n f_A(x)\mathrm{d}x\,,$$

  and that $\sum_{i=k_{1-c}}^{k_z-1} \left(1 - x^i_{\max}\right)^n \frac{1}{2^k} f_A(x^i_{\max})$ has the same limit;
- since $x^{k_{1-c}}_{\min} \in [{}^{(k_{1-c}\,-\,1)}/_{2^k}, {}^{k_{1-c}}/_{2^k}]$ then (G.9) justifies both

  $$x^{k_{1-c}}_{\min} \xrightarrow{k\to\infty} p_{1-c} \quad \text{and} \quad (1 - x^{k_{1-c}}_{\min})^n \xrightarrow{k\to\infty} (1-p_{1-c})^n.$$

  The limit $(1 - x^{k_{1-c}}_{\max})^n \xrightarrow{k\to\infty} (1-p_{1-c})^n$ also holds.

Consequently, using all of these limits in (G.15), the lower and upper bounds both converge to the limit $1 - S^*_{cmplt}$, defined as

$$\frac{(1-p_z)^{n+1}(1-c+\int_{p_z}^1 f_A(x)\,\mathrm{d}x) + \int_{p_{1-c}}^{p_z}(1-x)^{n+1} f_A(x)\,\mathrm{d}x}{(1-p_z)^{n}(1-c+\int_{p_z}^1 f_A(x)\,\mathrm{d}x) + \int_{p_{1-c}}^{p_z}(1-x)^{n} f_A(x)\,\mathrm{d}x} \qquad \text{(G.16)}$$

This proves S^*_{cmplt} exists and $S^*_k \xrightarrow{k\to\infty} S^*_{cmplt}$.

3.c) Notice that (17) holds, from (G.16). Also, by taking the limits of both sides of (G.6) and using the convergence of both S^*_k and $p_{z,k}$, we have

$$p_z \xleftarrow{k\to\infty} p_{z,k} = 1 - \frac{n}{n+1}(1 - S^*_k) \xrightarrow{k\to\infty} 1 - \frac{n}{n+1}\left(1 - S^*_{cmplt}\right) \qquad \text{(G.17)}$$

That is, $p_z = 1 - \frac{n}{n+1}\left(1 - S^*_{cmplt}\right)$, so (18) holds.

3.d) The final convergence argument we outline shows the existence and uniqueness of a limiting prior distribution F^* that attains the worst-case posterior expected B-system *pfd* S^*_{cmplt}. The argument relies on results from measure theory. Consider the *measurable space* consisting of the unit square and the *Borel sigma-algebra*, $\sigma(\mathcal{R})$, generated by the half-open rectangles[31] of the unit square, $\mathcal{R}$. All distributions in $\mathcal{D}$ are defined with respect to this measurable space[32]. If, on $\mathcal{R}$, the sequence of conservative priors F^*_k has a limit that agrees with an identified pre-(probability)measure F^*, then *Carathéodory's extension theorem* ([37], pages 39–40, Theorem 6.1) guarantees F^* can be extended to a unique[33] probability measure on $\sigma(\mathcal{R})$. F^*'s extension lies in $\mathcal{D}$ and *is* the conservative prior we seek.

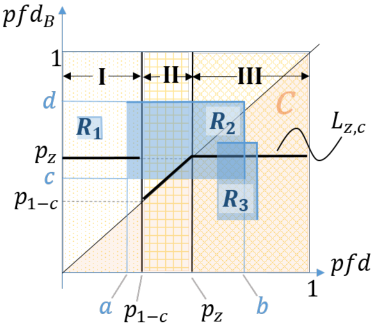

Figure G.23: Half-open rectangles and the three regions defined by $L_{z,c}$. The probability that F^* – a pre-(probability)measure – assigns to a half-open rectangle, is computed by summing the probability contributions from the rectangle's intersection with each of regions **I, II, III** and $L_{z,c}$. On each rectangle, the difference between F^*_k and F^* can be made vanishingly small as $k \to \infty$. We show that this convergence is uniform on the set $\mathcal{R}$ of half-open rectangles.

To show this, begin by defining a candidate F^* on $\mathcal{R}$ as follows. Appreciate that the unit square consists of three regions, I, II and III, each defined by the *pfd* values p_z and p_{1-c} (see Fig. G.23). Region I contains the horizontal line-segment $L^{I}_{z,c}$, of points (pfd_A, pfd_B) such that $pfd_A \in [0, p_{1-c})$ and $pfd_B = p_z$. Region II contains the diagonal line-segment $L^{II}_{z,c}$, of points with $pfd_A = pfd_B$ and $pfd_A \in [p_{1-c}, p_z)$. And, region III contains the horizontal line-segment $L^{III}_{z,c}$, of points with $pfd_A \in [p_z, 1]$ and $pfd_B = p_z$. Together, we denote these 3 line-segments as the set $L_{z,c}$. This is a Borel set (i.e. it is in $\sigma(\mathcal{R})$), since it is the countable limit of set operations involving open sets on the unit square. So, the intersection of $L_{z,c}$ with any half-open rectangle R is also Borel. Denote the projection of such an intersection onto the horizontal axis of the unit square as $\pi_A(L_{z,c} \cap R)$. The *measurable projection theorem* ([39], page 498, Theorem 13.2.7) guarantees this projection is a Lebesgue measurable set. Hence, using the (Lebesgue) integrability of f_A, define a pre-(probability)measure F^*:

$$F^*(R) := \int_{[0,1]} \mathbf{1}_{\pi_A(L_{z,c}\cap R)} f_A(x)\mathrm{d}x \quad \text{for } R \in \mathcal{R}$$

By the disjoint union $L_{z,c} = L^{I}_{z,c} \cup L^{II}_{z,c} \cup L^{III}_{z,c}$, and the non-

[30] Since $\int_0^{p_{1-c}} f_A(x)\mathrm{d}x = 1 - c$, this limit is simply $(1-p_z)^n(1-c)$.

[31] A half-open rectangle is $(a, b] \times (c, d]$ where $a < b, c < d$.

[32] One may distinguish between the distribution (function) F, typically defined on a *semi-algebra* ($\mathcal{R}$, in the present case), and its unique extension μ_F, where μ_F is a probability measure defined on the *sigma-algebra* generated by $\mathcal{R}$, denoted $\sigma(\mathcal{R})$. However, we will blur this distinction and write F for both notions. Which of these is meant will be clear from the context.

[33] So, the limit of the conservative priors F^*_k on $\sigma(\mathcal{R})$ must also agree with this unique extension of F^*. That the priors F^*_k converge weakly to a limiting probability measure on $\sigma(\mathcal{R})$ follows from an argument given by Prokhorov (e.g. see [38], page 64, corollary 2.4.5)

overlapping of the projections involved, we expand $F^*(R)$ as

$$F^*(R) \;=\; \int_{[0,1]} \mathbf{1}_{\pi_A(L^{I}_{z,c}\cap R)} f_A(x)\mathrm{d}x \;+\; \int_{[0,1]} \mathbf{1}_{\pi_A(L^{II}_{z,c}\cap R)} f_A(x)\mathrm{d}x \;+\; \int_{[0,1]} \mathbf{1}_{\pi_A(L^{III}_{z,c}\cap R)} f_A(x)\mathrm{d}x \tag{G.18}$$

Appendix I shows F^* is a pre-(probability)measure. Notice, F^* satisfies the NWTES and f_A constraints on the probability masses it allocates to rectangles. So its unique extension to $\sigma(\mathcal{R})$ – also denoted F^* – necessarily satisfies the constraints as well. Moreover, this unique extension is a prior distribution that attains the worst-case posterior expected B-system *pfd* S^*_{cmplt}.[34]

The restriction of each F^*_k to the set $\mathcal{R}$ may be similarly defined. Let k_* denote the interval in the k-th partition containing p_z. Using F^*_k's associated $p_{z,k}$, define the following 4 sets of (pfd_A, pfd_B) points in regions I, II and III, analogous to the line-segments in the $L_{z,c}$ set:

(a) the horizontal line-segment where $pfd_A \in [0, {}^{1}/_{2^k}) \cup \ldots \cup [\frac{(k_{1-c}-1)}{2^k}, p_{1-c})$ and $pfd_B = p_{z,k}$ – a subset of region I;

(b) the set of points where $pfd_A = pfd_B$ and $pfd_B \in \{{}^{k_{1-c}}/_{2^k}, {}^{(k_{1-c}+1)}/_{2^k}, \ldots, {}^{(k_*-1)}/_{2^k}\}$ – a subset of region II;

(c) the set of points where $pfd_A = pfd_B$ and $pfd_B \in \{{}^{k_*}/_{2^k}, {}^{(k_*+1)}/_{2^k}, \ldots, {}^{(k_z-1)}/_{2^k}\}$, as well as the horizontal line-segment that is $pfd_A \in (p_{z,k}, {}^{k_z}/_{2^k}] \cup \ldots \cup (\frac{2^k-1}{2^k}, 1]$ and $pfd_B = p_{z,k}$ – altogether, a subset of region III.

Denote these 4 sets as $L^{I}_{z,c,k}$, $L^{II}_{z,c,k}$, $L^{III,1}_{z,c,k}$ and $L^{III,2}_{z,c,k}$ respectively, and collectively as $L_{z,c,k}$. Examples of $L_{z,c,k}$ are depicted as horizontal bars on the $p_{z,k}$-line and dots on the diagonal in Fig.s G.21 and G.24. For any $R \in \mathcal{R}$, we have (see Fig. G.20)

$$\begin{aligned} F^*_k(R) = & \sum_{i=1}^{k_{1-c}-1} 2^k M_{k,i} \int_{[0,1]} \mathbf{1}_{[\frac{i-1}{2^k}, \frac{i}{2^k}) \cap \pi_A(L^{I}_{z,c,k}\cap R)} \mathrm{d}x \\ & + \frac{M_{k_{1-c}} - M_{k,k_{1-c}}}{p_{1-c} - \frac{k_{1-c}-1}{2^k}} \int_{[0,1]} \mathbf{1}_{[\frac{k_{1-c}-1}{2^k},\, p_{1-c}) \cap \pi_A(L^{I}_{z,c,k}\cap R)} \mathrm{d}x \end{aligned}$$

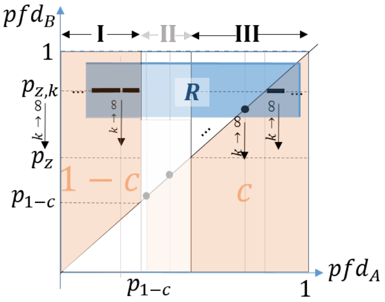

Figure G.24: Focusing on regions **I** and **III**, while the half-open rectangle ***R*** intersects the horizontal $p_{z,k}$-line for some k, it does not intersect any part of the horizontal p_z-line in either region **I** or **III**. Consequently, since $p_{z,k} \to p_z$ as $k \to \infty$ (i.e. over a countable number of partition refinements), the probability of the intersection of the rectangle with either region **I** or **III** tends to zero.

$$\begin{aligned} & + \sum_{{}^{i}/_{2^k} \in \pi_A(L^{II}_{z,c,k}\cap R)} M_{k,i} \;+\; \sum_{{}^{i}/_{2^k} \in \pi_A(L^{III,1}_{z,c,k}\cap R)} M_{k,i} \\ & + \frac{M_{k,k_z}}{{}^{k_z}/_{2^k} - p_{z,k}} \int_{[0,1]} \mathbf{1}_{(p_{z,k},\, \frac{k_z}{2^k}] \cap \pi_A(L^{III,2}_{z,c,k}\cap R)} \mathrm{d}x \\ & + \sum_{i=k_z}^{2^k} 2^k M_{k,i} \int_{[0,1]} \mathbf{1}_{(\frac{i-1}{2^k}, \frac{i}{2^k}] \cap \pi_A(L^{III,2}_{z,c,k}\cap R)} \mathrm{d}x \end{aligned} \tag{G.19}$$

where i in “$M_{k,i}$” denotes an interval in the k-th partition that contains points in either $\pi_A(L^{II}_{z,c,k} \cap R)$ or $\pi_A(L^{III,1}_{z,c,k} \cap R)$. And $M_{k,i}$ is the mass M_i that f_A assigns to these intervals (see (G.1)), with one exception: $i = k_{1-c}$ has $M_{k,k_{1-c}} := c - \sum_{j=k_{1-c}+1}^{2^k} M_j$, because F^*_k splits the mass for the k_{1-c} interval (see Fig. G.20).

We can now show that the sequence $\{F^*_k\}_{k\geqslant 1}$ converges to F^* uniformly on $\mathcal{R}$. That is, $\sup_{R\in\mathcal{R}} \left|F^*_k(R) - F^*(R)\right| \xrightarrow{k\to\infty} 0$.

For an arbitrarily chosen rectangle R, Fig. G.23 illustrates how the rectangle can either intersect $L_{z,c}$ or not. Suppose R does not intersect $L_{z,c}$, for instance as depicted in Fig. G.24. Then there exists some $K > 0$ such that no intersections $L_{z,c,k} \cap R$ occur for all partitions with $k \geqslant K$. That this must be true, in particular for the $p_{z,k}$ horizontal line-segments in $L_{z,c,k}$, follows from the convergence of $p_{z,k}$ to p_z (and, therefore, their respective line-segments converge) and our assumption that R does not intersect $L_{z,c}$. Also, no intersections occur between R and any of the diagonal points of $L_{z,c,k}$ for k greater than suitably large K; again, due to the convergence of $p_{z,k}$, and the fact that *if* the $p_{z,k}$ horizontal line-segment falls below the points in R for some $K > 0$, *then* so do any points on the diagonal of $L_{z,c,k}$ for all $k \geqslant K$ (because the largest of the pfd_B values for these points, $(k_z - 1)/2^k$, itself, converges to p_z by (G.11)). Consequently, from (G.18) and (G.19), $F^*_k(R) = 0 = F^*(R)$ for all $k \geqslant K$. This shows the convergence of the F^*_k to F^* on those R such that $L_{z,c} \cap R = \varnothing$. That is, for these R and all k large enough,

$$\left|F^*_k(R) - F^*(R)\right| = 0 \tag{G.20}$$

We now turn our attention to those R that intersect $L_{z,c}$. Divide the task of showing convergence on each such R into 3 cases: *a*) showing convergence on that part of the intersection

[34] This follows by definition since, from (G.16), the prior F^* gives:

$$\begin{aligned} 1 - \mathbb{E}[pfd_B \,|\, B \text{ passes } n \text{ tests}] &= \frac{\int_{[0,1]\times[0,1]} (1 - pfd_B)^{n+1} \mathrm{d}F^*}{\int_{[0,1]\times[0,1]} (1 - pfd_B)^{n} \mathrm{d}F^*} \\ &= \frac{(1-p_z)^{n+1}\left[F^*(L^{I}_{z,c}) + F^*(L^{III}_{z,c})\right] + \int\limits_{u\in L^{II}_{z,c}} (1-\pi_B(u))^{n+1} \mathrm{d}F^*(u)}{(1-p_z)^{n}\left[F^*(L^{I}_{z,c}) + F^*(L^{III}_{z,c})\right] + \int\limits_{u\in L^{II}_{z,c}} (1-\pi_B(u))^{n} \mathrm{d}F^*(u)} \\ &= \frac{(1-p_z)^{n+1}(1 - c + \int_{p_z}^{1} f_A(x)\,\mathrm{d}x) + \int_{p_{1-c}}^{p_z} (1-x)^{n+1} f_A(x)\,\mathrm{d}x}{(1-p_z)^{n}(1 - c + \int_{p_z}^{1} f_A(x)\,\mathrm{d}x) + \int_{p_{1-c}}^{p_z} (1-x)^{n} f_A(x)\,\mathrm{d}x} \\ &= 1 - S^*_{cmplt}\,. \end{aligned}$$

Note, as countable intersections of open rectangles, $L^{I}_{z,c}$, $L^{II}_{z,c}$ and $L^{III}_{z,c}$ are contained in $\sigma(\mathcal{R})$. Consequently, the extension of F^* to $\sigma(\mathcal{R})$ is defined on them, with definition given by (G.18) and use of the *monotone convergence theorem*.

lying in region I (denoted $L^{I}_{z,c} \cap R$), *b*) that part lying in region II (denoted $L^{II}_{z,c} \cap R$), and *c*) the part in region III (denoted $L^{III}_{z,c} \cap R$).

a) Consider region I and the subset $L^{I}_{z,c} \cap R$ (as exemplified in Fig. G.25). Since $p_{z,k}$ tends to p_z, this implies $L^{I}_{z,c,k}$ tends to $L^{I}_{z,c}$. So, if R intersects $L^{I}_{z,c}$, then R intersects $L^{I}_{z,c,k}$ as well, for all sufficiently large k. Suppose this for the k-th partition, and let l and r denote, respectively, the left-most and right-most of those intervals of $[0, 1]$ whose vertical strips contain $R \cap L^{I}_{z,c}$. The vertical strip for interval l has associated probability mass M_l, as determined by f_A (recall (G.1)). And probability M_r is associated with the strip for interval r. So, by the definitions of F^* and F^*_k on $\mathcal{R}$ (see (G.18) and (G.19)), bound the difference between $F^*_k(R)$ and $F^*(R)$ in region I as

$$\left| \begin{array}{l} \displaystyle\sum_{i=1}^{k_{1-c}-1} 2^k M_{k,i} \int_{[0,1]} \mathbf{1}_{\left[\frac{i-1}{2^k}, \frac{i}{2^k}\right) \cap \pi_A(L^{I}_{z,c,k} \cap R)} \mathrm{d}x \\ \displaystyle + \quad \frac{M_{k_{1-c}} - M_{k,k_{1-c}}}{p_{1-c} - \frac{k_{1-c}-1}{2^k}} \int_{[0,1]} \mathbf{1}_{\left[\frac{k_{1-c}-1}{2^k},\, p_{1-c}\right) \cap \pi_A(L^{I}_{z,c,k} \cap R)} \mathrm{d}x \\ \displaystyle - \quad \int_{[0,1]} \mathbf{1}_{\pi_A(L^{I}_{z,c} \cap R)} f_A(x) \mathrm{d}x \end{array} \right| < M_l + M_r \tag{G.21}$$

Additionally, because f_A attains all the values between its bounds on intervals l and r, and f_A attains its maximum on $[0, 1]$ (all due to f_A being continuous on $[0, 1]$), the *mean value theorem for integrals* applied to (G.1) justifies

$$M_l + M_r \leqslant \frac{1}{2^k} \left(\max_{x \in [\frac{l-1}{2^k}, \frac{l}{2^k}]} f_A(x) + \max_{x \in [\frac{r-1}{2^k}, \frac{r}{2^k}]} f_A(x) \right) \leqslant \frac{1}{2^{k-1}} \max_{x \in [0,1]} f_A(x)$$

So, the bound (G.21) becomes

$$\left| \begin{array}{l} \displaystyle\sum_{i=1}^{k_{1-c}-1} 2^k M_{k,i} \int_{[0,1]} \mathbf{1}_{\left[\frac{i-1}{2^k}, \frac{i}{2^k}\right) \cap \pi_A(L^{I}_{z,c,k} \cap R)} \mathrm{d}x \\ \displaystyle + \quad \frac{M_{k_{1-c}} - M_{k,k_{1-c}}}{p_{1-c} - \frac{k_{1-c}-1}{2^k}} \int_{[0,1]} \mathbf{1}_{\left[\frac{k_{1-c}-1}{2^k},\, p_{1-c}\right) \cap \pi_A(L^{I}_{z,c,k} \cap R)} \mathrm{d}x \\ \displaystyle - \quad \int_{[0,1]} \mathbf{1}_{\pi_A(L^{I}_{z,c} \cap R)} f_A(x) \mathrm{d}x \end{array} \right| < \frac{1}{2^{k-1}} \max_{x \in [0,1]} f_A(x) \tag{G.22}$$

We will use this bound shortly.

b) Next, consider region II with its subset $L^{II}_{z,c} \cap R$ (exemplified in Fig. G.26). Since the mesh sizes of the partitions tend to zero, if R intersects $L^{II}_{z,c}$ then it intersects $L^{II}_{z,c,k}$ for all sufficiently large k (e.g. for all k such that $^1/_{2^k}$ is smaller than the size of the interval $\pi_A(L^{II}_{z,c} \cap R)$). Let the intervals l and r be defined in a similar fashion to what was done for the region I case. So, l denotes the interval for the leftmost of the strips that contain $L^{II}_{z,c} \cap R$, and r the rightmost of these intervals. Again, using identical arguments given for region I, we obtain the following bound on the difference between $F^*_k(R)$ and $F^*(R)$ in region II:

$$\left| \sum_{^i/_{2^k} \in \pi_A(R \cap L^{II}_{z,c,k})} M_{k,i} - \int_{[0,1]} \mathbf{1}_{\pi_A(L^{II}_{z,c} \cap R)} f_A(x) \mathrm{d}x \right| < \frac{1}{2^{k-1}} \max_{x \in [0,1]} f_A(x) \tag{G.23}$$

We will use this bound shortly.

c) The final consideration is for region III, and its subsets $L^{III,1}_{z,c} \cap R$ and $L^{III,2}_{z,c} \cap R$. Once again, the convergence of $p_{z,k}$ implies that $L^{III,2}_{z,c,k}$ tends to $L^{III}_{z,c}$, but also that $L^{III,1}_{z,c,k}$ tends to the point (p_z, p_z). So, if R intersects $L^{III}_{z,c}$, then R intersects $L^{III,2}_{z,c,k}$ (and possibly $L^{III,1}_{z,c,k}$) for all sufficiently large k. The leftmost and rightmost of those intervals with strips containing $L^{III}_{z,c} \cap R$ are denoted l and r, as before. And the following bound can be deduced, as was done for the other regions, by bounding the difference between $F^*_k(R)$ and $F^*(R)$ in region III with $M_l + M_r$:

$$\left| \begin{array}{l} \displaystyle\sum_{^i/_{2^k} \in \pi_A(L^{III,1}_{z,c,k} \cap R)} M_{k,i} - \int_{[0,1]} \mathbf{1}_{\pi_A(L^{III}_{z,c} \cap R)} f_A(x) \mathrm{d}x \\ \displaystyle + \quad \frac{M_{k,k_z}}{^{k_z}/_{2^k} - p_{z,k}} \int_{[0,1]} \mathbf{1}_{\left(p_{z,k},\, \frac{k_z}{2^k}\right] \cap \pi_A(L^{III,2}_{z,c,k} \cap R)} \mathrm{d}x \\ \displaystyle + \quad \sum_{i=k_z}^{2^k} 2^k M_{k,i} \int_{[0,1]} \mathbf{1}_{\left(\frac{i-1}{2^k}, \frac{i}{2^k}\right] \cap \pi_A(L^{III,2}_{z,c,k} \cap R)} \mathrm{d}x \end{array} \right| < \frac{1}{2^{k-1}} \max_{x \in [0,1]} f_A(x) \tag{G.24}$$

So, using (G.18) and (G.19), the bounds (G.22), (G.23) and (G.24) imply the following uniform bound, valid for sufficiently large k and all $R \in \mathcal{R}$ that intersect $L_{z,c}$:

$$0 < \left| F^*_k(R) - F^*(R) \right| < \frac{3}{2^{k-1}} \max_{x \in [0,1]} f_A(x) \tag{G.25}$$

Together, (G.20) and (G.25) show that for all large k,

$$0 < \sup_{R \in \mathcal{R}} \left| F^*_k(R) - F^*(R) \right| < \frac{3}{2^{k-1}} \max_{x \in [0,1]} f_A(x) \tag{G.26}$$

That is, $\sup_{R \in \mathcal{R}} \left| F^*_k(R) - F^*(R) \right| \xrightarrow{k \to \infty} 0$. So the sequence of probability measures $\{F^*_k\}_{k \geqslant 1}$ converges uniformly to F^* on $\mathcal{R}$. With this, *Carathéodory's extension theorem* guarantees both the existence and uniqueness of F^*'s extension to $\sigma(\mathcal{R})$. And this extension must agree with the F^*_k on $\sigma(\mathcal{R})$, in the limit of large k. Hence, this F^* extension lies in $\mathcal{D}$, it satisfies both the NWTES and f_A constraints on its probability masses, and it attains the worst-case posterior expected B-system *pfd*, S^*_{cmplt}. This completes *stage 3* of the proof, and thus completes the proof. ■

A Remark: Can $p_z < p_{1-c}$ for the conservative prior F^* (as in Fig. G.27)? Such a measure appears consistent with the NWTES and f_A constraints, and consistent with our optimisations through the constrained movement of probability mass. However, it is in fact inconsistent, since this F^* satisfies both

$$1 - S^*_{cmplt} = \frac{(1 - p_z)^{n+1}(c + \int_0^{p_z} f_A(x)\,\mathrm{d}x) + \int_{p_z}^{p_{1-c}} (1 - x)^{n+1} f_A(x)\,\mathrm{d}x}{(1 - p_z)^n(1 - c + \int_{p_z}^1 f_A(x)\,\mathrm{d}x) + \int_{p_{1-c}}^{p_z} (1 - x)^n f_A(x)\,\mathrm{d}x}$$

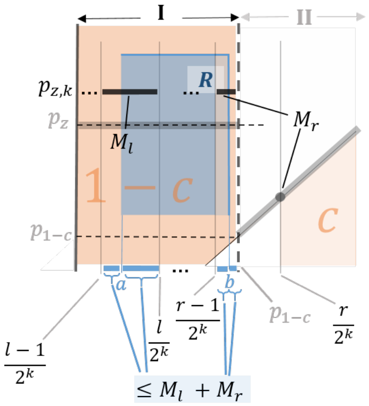

Figure G.25: For all large enough k, the difference between F_k^* and F^* on that part of rectangle $\boldsymbol{R}$ in region **I** is smaller than $M_l + M_r$.

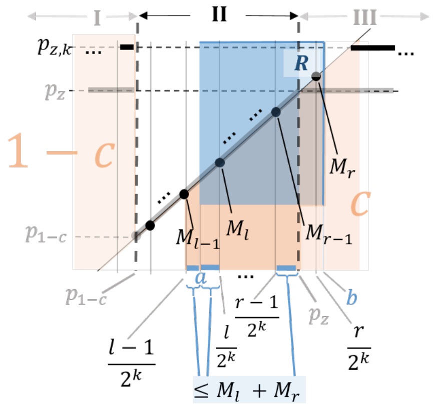

Figure G.26: For all large enough k, the difference between F_k^* and F^* on that part of rectangle $\boldsymbol{R}$ in region **II** is smaller than $M_l + M_r$.

and $p_z = 1 - \frac{n}{n+1}(1 - S^*_{cmplt})$. Together, these imply

$$\frac{n+1}{n}(1-p_z) = \frac{(1-p_z)^{n+1}(c + \int_0^{p_z} f_A(x)\,\mathrm{d}x) + \int_{p_z}^{p_{1-c}}(1-x)^{n+1} f_A(x)\,\mathrm{d}x}{(1-p_z)^n(1-c+\int_{p_z}^1 f_A(x)\,\mathrm{d}x) + \int_{p_{1-c}}^{p_z}(1-x)^n f_A(x)\,\mathrm{d}x}$$

Hence, since $\frac{n+1}{n} > 1$, we must have

$$\frac{(1-p_z)^{n+1}(c + \int_0^{p_z} f_A(x)\,\mathrm{d}x) + \int_{p_z}^{p_{1-c}}(1-x)^{n+1} f_A(x)\,\mathrm{d}x}{(1-p_z)^{n+1}(c+\int_0^{p_z} f_A(x)\,\mathrm{d}x) + (1-p_z)\int_{p_z}^{p_{1-c}}(1-x)^n f_A(x)\,\mathrm{d}x} > 1$$

Or, upon simplifying, we get $\int_{p_z}^{p_{1-c}}(1-x)^n(p_z - x) f_A(x)\,\mathrm{d}x > 0$. This is a contradiction since this integral cannot be positive (i.e. $p_z < x$ for all $x \in (p_z, p_{1-c}]$, by assumption).

Appendix H.

Lemma 3. *Consider the k-th dyadic partition of the unit square, as in Appendix G, and the masses M_i (for $1 \leqslant i \leqslant 2^k$) defined by f_A – a continuous marginal distribution for the A-system pfd. Furthermore, let the M_i, M_i^u, M_i^l masses, k_z, and $p_{z,k,M}$, be defined and constrained as in Appendix G. Then, for $a < b$, the constrained movement of probability mass from M_a^l, M_b^u to M_a^u, M_b^l – keeping all other masses fixed – increases $p_{z,k,M}$.*

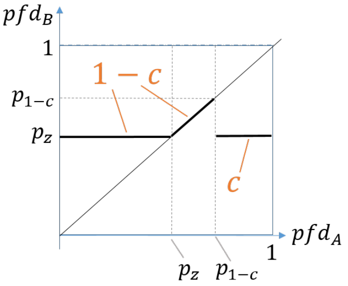

Figure G.27: F^* with $p_z < p_{1-c}$, as depicted, is inconsistent.

Proof. We will prove this by deducing that, under the constraints on the probability masses, $p_{z,k,M}$ is a continuously differentiable function of M_a^l and $\frac{\partial p_{z,k,M}}{\partial M_a^l} < 0$. There are three cases to consider here: i) $a < k_z < b$, ii) $a < b < k_z$ and iii) $k_z < a < b$.

Case (i) : Assume $a < k_z < b$. From Appendix G, note that the masses satisfy $M_i = M_i^u + M_i^l$, and (for $n \geqslant 1$)

$$\frac{n+1}{n}(1 - p_{z,k,M}) = \frac{Nu}{De} \tag{H.1}$$

where Nu is defined as

$$\begin{aligned} Nu = &\sum_{\substack{i=1\\ i\neq a,b}}^{k_z}\Big[(1-p_{z,k,M})^{n+1}\Big(M_i - M_i^l\Big) + (1-y_i)^{n+1}M_i^l\Big] + \\ &\sum_{\substack{i=k_z+1\\ i\neq a,b}}^{2^k}\Big[(1-y_i)^{n+1}\Big(M_i - M_i^l\Big) + (1-p_{z,k,M})^{n+1}M_i^l\Big] + \\ &(1-p_{z,k,M})^{n+1}\Big(M_a - M_a^l\Big) + (1-y_a)^{n+1}M_a^l \quad + \\ &(1-y_b)^{n+1}\Big(M_b - M_b^l\Big) + (1-p_{z,k,M})^{n+1}M_b^l \end{aligned}$$

and De is similarly defined, but with "$n+1$" replaced by "n".

As in Appendix C and Appendix E, the identity (H.1) holds for any set of M_i^l masses and their associated $p_{z,k,M}$ value, where the M_i^l satisfy the constraints of Appendix G. In particular, using both the NWTES constraint $\sum_{\substack{i=1\\ i\neq a,b}}^{2^k} M_i^l + M_a^l + M_b^l = c$ and the constraint $\sum_{i=1}^{2^k} M_i = 1$, we can couple the "a" and "b" interval contributions in (H.1) by eliminating explicit reference to M_b^l and M_b in the expressions for Nu and De, resulting in

$$\begin{aligned} Nu = &\sum_{\substack{i=1\\ i\neq a,b}}^{k_z}\Big[(1-p_{z,k,M})^{n+1}\Big(M_i - M_i^l\Big) + (1-y_i)^{n+1}M_i^l\Big] \quad + \\ &\sum_{\substack{i=k_z+1\\ i\neq a,b}}^{2^k}\Big[(1-y_i)^{n+1}\Big(M_i - M_i^l\Big) + (1-p_{z,k,M})^{n+1}M_i^l\Big] + \end{aligned}$$

$$\begin{aligned}&(1-p_{z,k,M})^{n+1}\left(M_a-M_a^l\right)+(1-y_a)^{n+1}M_a^l \quad + \\ &(1-y_b)^{n+1}\left(1-\sum_{\substack{i=1\\i\neq b}}^{2^k}M_i-c+\sum_{\substack{i=1\\i\neq a,b}}^{2^k}M_i^l+M_a^l\right) \quad + \\ &(1-p_{z,k,M})^{n+1}\left(c-\sum_{\substack{i=1\\i\neq a,b}}^{2^k}M_i^l-M_a^l\right)\end{aligned} \tag{H.2}$$

and a similar expression for De, with "$n+1$" replaced by "n".

By the *implicit function theorem*, $p_{z,k,M}$ is a continuously differentiable function of M_a^l for $De > 0$.[35] So, differentiating (H.1) w.r.t. M_a^l gives

$$\frac{n+1}{n}\left(-\frac{\partial p_{z,k,M}}{\partial M_a^l}\right)=\frac{1}{De}\left(\frac{\partial Nu}{\partial M_a^l}-\frac{Nu}{De}\frac{\partial De}{\partial M_a^l}\right) \tag{H.3}$$

(H.3) shows that the sign of the derivative $\frac{\partial p_{z,k,M}}{\partial M_a^l}$ is the "negative" of the sign of $\frac{\partial Nu}{\partial M_a^l}-\frac{Nu}{De}\frac{\partial De}{\partial M_a^l}$. To determine the sign of the r.h.s. of (H.3), first observe that, using the expressions in (H.2), we may evaluate $\frac{\partial Nu}{\partial M_a^l}$ and $\frac{\partial De}{\partial M_a^l}$ as

$$\begin{aligned}\frac{\partial Nu}{\partial M_a^l} &= (n+1)\left(-\frac{\partial p_{z,k,M}}{\partial M_a^l}\right)\left(c-\sum_{\substack{i=1\\i\neq a,b}}^{2^k}M_i^l-M_a^l\right)(1-p_{z,k,M})^n \\ &+(n+1)\left(-\frac{\partial p_{z,k,M}}{\partial M_a^l}\right)\left(M_a-M_a^l\right)(1-p_{z,k,M})^n \\ &+(1-y_a)^{n+1}+(1-y_b)^{n+1} \\ &-(1-p_{z,k,M})^{n+1}-(1-p_{z,k,M})^{n+1}\end{aligned}$$

and $\frac{\partial De}{\partial M_a^l}$ has a similar expression, but with "$n-1$" replacing "n". Using these partial derivatives and (H.1), we can rewrite the expression within the brackets on the r.h.s. of (H.3) to obtain

$$\frac{\partial Nu}{\partial M_a^l}-\frac{Nu}{De}\frac{\partial De}{\partial M_a^l} = g(y_a)-g(p_{z,k,M})+g(y_b)-g(p_{z,k,M})$$

where g is the auxilliary function defined in (C.4) (with $p_{z,k,M}$ replacing p_z there). From the properties of g (Fig. C.14, with $S^*_{k,M}$ replacing S^*), its global minimum occurs at $p_{z,k,M}$ so that

$$g(y_i)-g(p_{k,z,M})>0 \quad \text{for all } i\neq k_z.$$

Hence, $\frac{\partial Nu}{\partial M_a^l}-\frac{Nu}{De}\frac{\partial De}{\partial M_a^l}>0$ and, by (H.3), one deduces $\frac{\partial p_{z,k,M}}{\partial M_a^l}<0$. That is, $p_{z,k,M}$ is a decreasing function of M_a^l and M_b^u together[36]. This completes the proof for when $a<k_z<b$.

Case (ii) : Assuming $a<b<k_z$, the proof follows an almost identical argument to that of case (i). One merely replaces the "a" and "b" expressions in Nu of (H.2) by

$$\begin{aligned}&(1-p_{z,k,M})^{n+1}\left(M_a-M_a^l\right)+(1-y_a)^{n+1}M_a^l+(1-y_b)^{n+1}\left(c-\sum_{\substack{i=1\\i\neq a,b}}^{2^k}M_i^l-M_a^l\right) \\ &+(1-p_{z,k,M})^{n+1}\left(1-\sum_{\substack{i=1\\i\neq b}}^{2^k}M_i-c+\sum_{\substack{i=1\\i\neq a,b}}^{2^k}M_i^l+M_a^l\right)\end{aligned} \tag{H.4}$$

and similarly for De. Then, like case (i), $\frac{\partial p_{z,k,M}}{\partial M_a^l}<0$ can be deduced; since $y_a<y_b<p_z$ and, from the properties of the g function, $g(y_a)-g(y_b)>0$ (this is the sign of the r.h.s. of (H.3)).

Case (iii) : Finally, assume $k_z<a<b$. Then, replace the "a" and "b" expressions in Nu of (H.2) with

$$\begin{aligned}&(1-p_{z,k,M})^{n+1}M_a^l+(1-y_a)^{n+1}\left(M_a-M_a^l\right)+(1-p_{z,k,M})^{n+1}\left(c-\sum_{\substack{i=1\\i\neq a,b}}^{2^k}M_i^l-M_a^l\right) \\ &+(1-y_b)^{n+1}\left(1-\sum_{\substack{i=1\\i\neq b}}^{2^k}M_i-c+\sum_{\substack{i=1\\i\neq a,b}}^{2^k}M_i^l+M_a^l\right)\end{aligned} \tag{H.5}$$

and similarly for De. Again, like case (i), one shows $\frac{\partial p_{z,k,M}}{\partial M_a^l}<0$; since $p_z<y_a<y_b$ and, by the properties of the g function, $g(y_b)-g(y_a)>0$ (this is the sign of the r.h.s. of (H.3)). ■

Appendix I.

Claim: F^* is a pre-(probability)measure on $\mathcal{R}$.

Proof. Note that $F^*(\varnothing)=0$ by definition, as the projection of the empty set is empty. Also by definition of $L_{z,c}$, and f_A being a probability density function,

$$F^*([0,1]\times[0,1])=\int_{[0,1]}\mathbf{1}_{\pi_A(L_{z,c})}f_A(x)\mathrm{d}x=\int_{[0,1]}f_A(x)\mathrm{d}x=1\,.$$

Furthermore, disjoint rectangles R_1 and R_2 have disjoint projections $\pi_A(L_{z,c}\cap R_1)$ and $\pi_A(L_{z,c}\cap R_2)$; since $L_{z,c}$ can be viewed as a function $L_{z,c}:[0,1]\to[0,1]$ and, consequently, if $x\in\pi_A(L_{z,c}\cap R_1)$ and $x\in\pi_A(L_{z,c}\cap R_2)$ then $(x,L_{z,c}(x))\in R_1$ and $(x,L_{z,c}(x))\in R_2$, contradicting the disjointness of R_1 and R_2.

So, if $\{R_i\}_{i\geqslant 1}$ are pair-wise disjoint rectangles, then the *monotone convergence theorem* and the fact that projection "π_A" commutes with countable set unions, justify

$$\begin{aligned}F^*(\cup_{i=1}^\infty R_i) &= \int_{[0,1]}\mathbf{1}_{\pi_A(L_{z,c}\cap(\cup_{i=1}^\infty R_i))}f_A(x)\mathrm{d}x \\ &= \int_{[0,1]}\mathbf{1}_{\pi_A(\cup_{i=1}^\infty(L_{z,c}\cap R_i))}f_A(x)\mathrm{d}x \\ &= \int_{[0,1]}\mathbf{1}_{\cup_{i=1}^\infty\pi_A(L_{z,c}\cap R_i)}f_A(x)\mathrm{d}x \\ &= \int_{[0,1]}\sum_{i=1}^\infty\mathbf{1}_{\pi_A(L_{z,c}\cap R_i)}f_A(x)\mathrm{d}x \\ &= \sum_{i=1}^\infty\int_{[0,1]}\mathbf{1}_{\pi_A(L_{z,c}\cap R_i)}f_A(x)\mathrm{d}x \;=\; \sum_{i=1}^\infty F^*(R_i)\,.\end{aligned}$$

■

[35] Note, the feasibility of the prior distributions – that these satisfy the constraints on the M_i, M_i^l masses – and the definition of $p_{z,k,M}$, altogether ensure $De>0$. Otherwise, the optimisation of Appendix G has a possibly unbounded objective function that is not a posterior, expected probability.

[36] Because M_b^u increases with M_a^l, since $M_b^u=1-\sum_{\substack{i=1\\i\neq b}}^{2^k}M_i-c+\sum_{\substack{i=1\\i\neq a,b}}^{2^k}M_i^l+M_a^l$