City Research Online

City, University of London Institutional Repository

Citation: Salako, K. ORCID: 0000-0003-0394-7833, Strigini, L. ORCID: 0000-0002-4246-2866 and Zhao, X. (2021). Proofs of Conservative Confidence Bounds on PFD, Using Claims of Improved Reliability. London, UK: Centre for Software Reliability, City, University of London.

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/25905/

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

Proofs of Conservative Confidence Bounds on PFD, Using Claims of Improved Reliability

Kizito Salako, Lorenzo Strigini
Centre for Software Reliability
City, University of London
Northampton Square EC1V 0HB, U.K.
{k.o.salako,l.strigini}@city.ac.uk

Xingyu Zhao
Department of Computer Science
University of Liverpool
Ashton Street L69 3BX, U.K.
xingyu.zhao@liverpool.ac.uk

***Abstract*—We derive infima for posterior confidence in bounds on system *pfd*, subject to various constraints – called "*prior knowledge*" – on joint prior distributions. This concerns confidence bounds on the *pfd* for a system B, upon observing two systems, A and B, operate without failure. In particular, the results hold when evidence supports a claim of B being as reliable, or better, than A. The propositions proved in this technical report are motivated, explained and discussed in the paper "*Conservative Confidence Bounds in Safety, from Generalised Claims of Improvement & Statistical Evidence*", reported at the 51st IEEE/IFIP DSN conference.**

Introduction

This report derives conservative confidence bounds on the *probability of failure on demand* (*pfd*) for a system. Let X_A and X_B be the unknown *pfd*s of systems A and B. Consider the following 4 forms of "prior knowledge" that each constrain the joint prior distribution of $\langle X_A, X_B\rangle$. We refer to these constraints as PK 1, 2, 3 and 4 respectively.

Prior Knowledge 1. *certainty that the system pfd X is no better than some $p_l \geqslant 0$. That is, $P(X \geqslant p_l) = 1$.*

Prior Knowledge 2. *$\theta \times 100\%$ confidence that the system pfd X meets, or surpasses, a pfd ε. That is, $P(X \leqslant \varepsilon) = \theta$.*

Prior Knowledge 3. *confidence in version A's pfd being α or better, and in the B version being an improvement:*

$$P(X_B \leqslant X_A, X_A \leqslant \alpha) = \varphi \tag{1}$$

where $\varepsilon \leqslant \alpha \leqslant 1$ and $0 < \varphi < 1$. In particular, ϕ is defined as the value of φ when $\alpha = 1$.

Prior Knowledge 4. *confidence in version A's pfd falling within some range of values, and version B being an improvement: for some sub-interval I of $[0, 1]$, with ϕ as just defined,*

$$P(X_B \leqslant X_A, X_A \in I) = \frac{\phi}{1-\phi} P(X_A < X_B, X_A \in I) \tag{2}$$

In particular, we consider the case when such a requirement holds for the two intervals $[p_l \leqslant X_A \leqslant \varepsilon]$, $[\alpha \leqslant X_A \leqslant 1]$ and, thus (as probabilities must add up to 1), also holds for $[\varepsilon < X_A < \alpha]$.

Let two independent Bernoulli processes characterise the occurrence of failures for systems A and B. If systems A and B experience no failures, respectively, on a sequence of n_A and n_B independent demands, then the posterior probability that $[X_B \leqslant p]$ for some p is

$$P(X_B \leqslant p \mid n_A, n_B) = \frac{\mathbb{E}[\, L(X_A, X_B)\mathbf{1}_{X_B \leqslant p}\,]}{\mathbb{E}[\, L(X_A, X_B)\,]} \tag{3}$$

where $L(x, y) = (1-x)^{n_A}(1-y)^{n_B}$ is the likelihood function, and $\mathbf{1}_{\mathtt{S}}$ is an indicator function – it equals 1 when predicate $\mathtt{S}$ is true, and 0 otherwise.

Let $\mathcal{D}$ be the set of all probability distributions over the unit square. The following two propositions are constrained optimisation problems that give the infima (i.e. greatest lower bounds) for (3) under different circumstances. Solving these problems entails determining preferred joint prior distributions with $P(X_B < p \mid n_A, n_B)$ equal to the relevant infimum. Each optimisation problem is solved subject to PK constraints and certain parameter ranges of θ, φ and ϕ. The solutions illustrate how to solve analogous optimisation problems for other parameter ranges, using similar solution steps.

Proposition 1. *Consider the optimisation problem*

$$\inf_{\mathcal{D}} P(X_B \leqslant p \mid n_A, n_B)$$

where $\varepsilon \leqslant p$, subject to systems A, B satisfying PK 1, 2, 3.

Fig. 1 shows a prior distribution that solves this problem when $\varphi > \theta > 1-\theta, \varepsilon < \alpha \leqslant p$. The infimum is

$$\frac{Num}{Num + L(p_l, p)(1-\varphi)} \tag{4}$$

where $Num := L(\alpha, \alpha)(\varphi - \theta) + L(\varepsilon, \varepsilon)(\varphi + \theta - 1) + L(\alpha, \varepsilon)(1-\varphi)$. This is the value of $P(X_B < p \mid n_A, n_B)$ computed using this prior distribution.

Proposition 2. *Consider the optimisation problem*

$$\inf_{\mathcal{D}} P(X_B \leqslant p \mid n_A, n_B)$$

where $\varepsilon \leqslant p$, subject to systems A, B satisfying PK 1, 2, 4.

Fig. 2 shows a prior distribution that solves this problem when $\phi > \theta$, $p < \alpha < 1$. The infimum is

$$\frac{L(\varepsilon, \varepsilon)\phi^2\theta}{L(\varepsilon, \varepsilon)\phi^2\theta + Den} \tag{5}$$

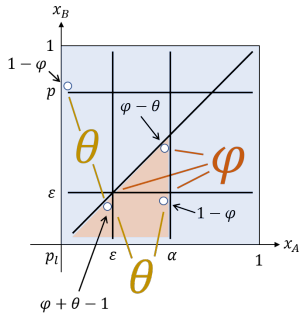


Fig. 1: A prior distribution that solves proposition 1, when $\varphi > \theta > 1-\theta$, $\varepsilon < \alpha \leqslant p$.

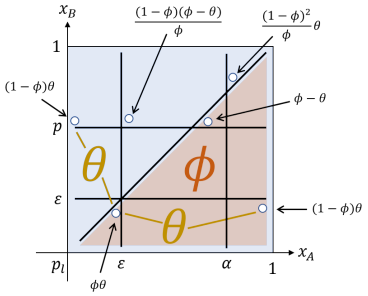


Fig. 2: A prior distribution that solves proposition 2 when $\phi > \theta$, $p < \alpha < 1$.

where $Den := L(p,p)\phi(\phi-\theta) + L(p_l,p)\phi(1-\phi)\theta + L(\varepsilon,p)(1-\phi)(\phi-\theta) + L(\alpha,\alpha)(1-\phi)^2\theta$. *This is the value of* $P(X_B < p \mid n_A, n_B)$ *computed using this prior distribution.*

The joint prior distributions in Fig.s 1 and 2 are depicted "from above", looking down on the distributions' domain. In each figure, the PKs partition the domain into the subsets shown, with associated masses for θ, φ and ϕ.

Consequences of Invariant Ratio Property In PK 4

Consider a probability distribution P *over the unit square. Suppose* $0 < \phi < 1$ *and, for all sub-intervals* I *of* $[0,1]$,

$$P(X_A \geqslant X_B, X_A \in I) = \frac{\phi}{1-\phi} P(X_A < X_B, X_A \in I) \quad (6)$$

Then

$$P(X_A < X_B) = P(X_A < X_B \mid X_A \in I) = 1-\phi \quad (7)$$

Moreover, $P(X_A < X_B \mid n_A) = 1-\phi$ *for all* $n_A \geqslant 0$.

Proof. For any sub-interval I of $[0,1]$, (6) implies

$$\begin{aligned} & P(X_A \in I) \\ = \;& P(X_A < X_B, X_A \in I) + P(X_A \geqslant X_B, X_A \in I) \\ = \;& P(X_A < X_B, X_A \in I) + \frac{\phi}{1-\phi} P(X_A < X_B, X_A \in I) \\ = \;& \frac{1}{1-\phi} P(X_A < X_B, X_A \in I) \end{aligned}$$

Thus, whenever I is such that $P(X_A \in I) \neq 0$,

$$P(X_A < X_B \mid X_A \in I) = \frac{P(X_A < X_B, X_A \in I)}{P(X_A \in I)} = 1-\phi$$

In particular, when $I = [0,1]$, $P(X_A < X_B) = 1-\phi$. Moreover, the following limiting, sandwiching argument proves that $P(X_A < X_B \mid n_A) = 1-\phi$.

Consider the following sequence $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \ldots$ of partitions of $(0,1]$, consisting of dyadic intervals:
$\mathcal{P}_1 = \{(0,1]\}$, $\mathcal{P}_2 = \{(0, 1/2], (1/2, 1]\}$, $\mathcal{P}_3 = \{(0, 1/4], (1/4, 1/2], (1/2, 3/4], (3/4, 1]\}$, ... and so on.

$\mathcal{P}_n$ contains 2^{n-1} disjoint intervals. For interval I in $\mathcal{P}_n$, let x^I_{inf} and x^I_{sup} be the infimum and supremum of this interval. Let $L(x) = (1-x)^{n_A}$. The monotonicity and continuity of L over $[0,1]$ guarantees L attains all values from $L(x^I_{sup})$ to $L(x^I_{inf})$. Consequently, (7) implies

$$\begin{aligned} & \frac{\sum_{I\in\mathcal{P}_n} L(x^I_{inf}) P(X_A < X_B, X_A \in I)}{\sum_{I\in\mathcal{P}_n} L(x^I_{sup}) P(X_A < X_B, X_A \in I)/(1-\phi)} \\ = \;& \frac{\sum_{I\in\mathcal{P}_n} L(x^I_{inf}) P(X_A < X_B, X_A \in I)}{\sum_{I\in\mathcal{P}_n} L(x^I_{sup}) P(X_A \in I)} \\ \geqslant \;& \frac{\sum\limits_{I\in\mathcal{P}_n} \mathbb{E}[\, L(X_A) \mathbf{1}_{X_A < X_B} \mathbf{1}_{X_A \in I}\,]}{\sum\limits_{I\in\mathcal{P}_n} \mathbb{E}[\, L(X_A) \mathbf{1}_{X_A \in I}\,]} \\ \geqslant \;& \frac{\sum_{I\in\mathcal{P}_n} L(x^I_{sup}) P(X_A < X_B, X_A \in I)}{\sum_{I\in\mathcal{P}_n} L(x^I_{inf}) P(X_A \in I)} \\ = \;& \frac{\sum_{I\in\mathcal{P}_n} L(x^I_{sup}) P(X_A < X_B, X_A \in I)}{\sum_{I\in\mathcal{P}_n} L(x^I_{inf}) P(X_A < X_B, X_A \in I)/(1-\phi)} . \end{aligned} \quad (8)$$

Again, by the monotonicity and continuity of L over $[0,1]$, the *monotone convergence theorem* [1] implies

$$\begin{aligned} & \lim_{n\to\infty} \sum_{I\in\mathcal{P}_n} L(x^I_{inf}) P(X_A < X_B, X_A \in I) \\ = \;& \lim_{n\to\infty} \sum_{I\in\mathcal{P}_n} L(x^I_{sup}) P(X_A < X_B, X_A \in I) \end{aligned}$$

Thus, upon using the following identity (for all n) in (8),

$$\frac{\sum\limits_{I\in\mathcal{P}_n} \mathbb{E}[\, L(X_A) \mathbf{1}_{X_A < X_B} \mathbf{1}_{X_A \in I}\,]}{\sum\limits_{I\in\mathcal{P}_n} \mathbb{E}[\, L(X_A) \mathbf{1}_{X_A \in I}\,]} = \frac{\mathbb{E}[\, L(X_A) \mathbf{1}_{X_A < X_B}\,]}{\mathbb{E}[\, L(X_A)\,]}$$

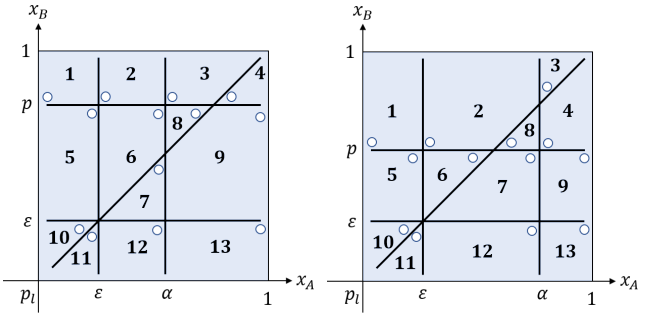
and then taking limits as $n \to \infty$, the *monotone convergence theorem* implies

$$\begin{aligned} & (1-\phi) \lim_{n\to\infty} \frac{\sum_{I\in\mathcal{P}_n} L(x^I_{inf}) P(X_A < X_B, X_A \in I)}{\sum_{I\in\mathcal{P}_n} L(x^I_{sup}) P(X_A < X_B, X_A \in I)} \\ = \;& (1-\phi) \frac{\lim\limits_{n\to\infty} \sum_{I\in\mathcal{P}_n} L(x^I_{inf}) P(X_A < X_B, X_A \in I)}{\lim\limits_{n\to\infty} \sum_{I\in\mathcal{P}_n} L(x^I_{sup}) P(X_A < X_B, X_A \in I)} \\ = \;& 1-\phi \geqslant \frac{\mathbb{E}[L(X_A) \mathbf{1}_{X_A < X_B}\,]}{\mathbb{E}[L(X_A)\,]} \geqslant 1-\phi \\ = \;& (1-\phi) \frac{\lim\limits_{n\to\infty} \sum_{I\in\mathcal{P}_n} L(x^I_{sup}) P(X_A < X_B, X_A \in I)}{\lim\limits_{n\to\infty} \sum_{I\in\mathcal{P}_n} L(x^I_{inf}) P(X_A < X_B, X_A \in I)} \\ = \;& (1-\phi) \lim_{n\to\infty} \frac{\sum_{I\in\mathcal{P}_n} L(x^I_{sup}) P(X_A < X_B, X_A \in I)}{\sum_{I\in\mathcal{P}_n} L(x^I_{inf}) P(X_A < X_B, X_A \in I)} \end{aligned} \quad (9)$$

Since $P(X_A < X_B \mid n_A) = \frac{\mathbb{E}[(1-X_A)^{n_A}\mathbf{1}_{X_A<X_B}]}{\mathbb{E}[(1-X_A)^{n_A}]}$ by definition, (9) shows that $P(X_A < X_B \mid n_A) = 1 - \phi$. □

Remark : $P(n_A \text{ successes} \mid X_A < X_B) = P(n_A \text{ successes})$.

PROOFS OF PROPOSITIONS 1 AND 2



(a) $\varepsilon \leqslant \alpha \leqslant p$ (b) $p \leqslant \alpha \leqslant 1$

Fig. 3: Priors that solve propositions 1 or 2 will have one of these discrete distribution forms, depending on the value of α. There are 13 regions, where region i has probability mass M_i. The white circles approximately indicate a corner within each region i. When mass M_i is assigned to a point within the region, the closer the point is to the corner, the smaller $P(X_B < p \mid n_A, n_B)$ becomes.

In what follows, propositions 1 and 2 are proved. In either proposition, we seek a prior distribution – from the set of all probability distributions that are justified by the PK for the respective proposition – that gives the greatest lower bound on the value of $P(X_B \leqslant p \mid n_A, n_B)$.

PK 1, 2, 3 and 4, along with the bound p on X_B, partition a subset of the unit square into 13 subsets. Two such partitions are illustrated in Fig. 3, with each subset having an associated probability mass that will be determined by solving the constrained optimisation problems in the propositions.

So, consider a finite partition of the unit square into n regions $R_1, \ldots, R_n$, and let $\mathcal{D}$ be the set of all probability distributions over the unit square. Each of these distributions assigns probabilities – generically labelled $M_1, M_2, \ldots, M_n$ – to the respective regions. Denote, by $\mathcal{D}^*$, that subset of $\mathcal{D}$ consisting of discrete distributions that assign each probability M_i to a single point within region R_i. The following lemma guarantees that extremising over $\mathcal{D}^*$ is the same as extremising over $\mathcal{D}$. Its proof is reproduced from [2].

Lemma 1. *Let f and g be non-negative functions with finite non-zero expectations, and let each function be defined over the unit square. Let $\mathcal{D}$ and $\mathcal{D}^*$ be as already defined. Then,*

$$\inf_{\mathcal{D}^*} \frac{\sum_{i=1}^n f(u_i)M_i}{\sum_{i=1}^n g(u_i)M_i} = \inf_{\mathcal{D}} \frac{\mathbb{E}[f]}{\mathbb{E}[g]} \quad (\textit{where } u_i \in R_i) \tag{10}$$

Proof. Clearly, $\mathcal{D}^* \subset \mathcal{D}$ implies the l.h.s. of (10) is greater than, or equal to, the r.h.s.. The converse is also true, as the following "proof by contradiction" shows. For any distribution in $\mathcal{D}$, assume

$$\frac{\mathbb{E}[f]}{\mathbb{E}[g]} = \frac{\sum_{i=1}^n \mathbb{E}[f \mid R_i]M_i}{\sum_{i=1}^n \mathbb{E}[g \mid R_i]M_i} < \frac{\sum_{i=1}^n f(u_i)M_i}{\sum_{i=1}^n g(u_i)M_i} \tag{11}$$

for all $u_i \in R_i$. Rearranging gives

$$\sum_{i=1}^n \mathbb{E}[f \mid R_i]M_i \sum_{i=1}^n g(u_i)M_i < \sum_{i=1}^n \mathbb{E}[g \mid R_i]M_i \sum_{i=1}^n f(u_i)M_i$$

for all $u_i \in R_i$. An equivalent way of writing this, using indicator functions, is

$$\sum_{i=1}^n \mathbb{E}[f \mid R_i]M_i \sum_{i=1}^n g(w_i)\mathbf{1}_{R_i}M_i < \sum_{i=1}^n \mathbb{E}[g \mid R_i]M_i \sum_{i=1}^n f(w_i)\mathbf{1}_{R_i}M_i$$

where the w_i are dummy variables. Taking conditional expectations $\mathbb{E}[\,\cdot \mid R_j]$ for $j = 1, \ldots, n$, on both sides of this inequality, and summing these, yields the contradiction

$$\sum_{i=1}^n \mathbb{E}[f \mid R_i]M_i \sum_{i=1}^n \mathbb{E}[g \mid R_i]M_i < \sum_{i=1}^n \mathbb{E}[g \mid R_i]M_i \sum_{i=1}^n \mathbb{E}[f \mid R_i]M_i$$

Consequently, (11) must be false for all distributions in $\mathcal{D}$. Hence, the r.h.s. of (10) is greater than, or equal to, the l.h.s.. □

So, for each proposition, (3) and lemma 1 imply

$$\begin{aligned}
\inf_{\mathcal{D}} P(X_B \leqslant p \mid n_A, n_B) &= \inf_{\mathcal{D}} \frac{\mathbb{E}[\,L(X_A, X_B)\mathbf{1}_{X_B \leqslant p}\,]}{\mathbb{E}[\,L(X_A, X_B)\,]} \\
&= \inf_{\mathcal{D}^*} \frac{\sum_{i=1}^{i=13} L(x_i, y_i)M_i\mathbf{1}_{y_i \leqslant p}}{\sum_{i=1}^{i=13} L(x_i, y_i)M_i} \\
&= \inf_{\mathcal{D}^*} \frac{1}{1 + \frac{\sum_{i=1}^{i=13} L(x_i, y_i)M_i\mathbf{1}_{y_i > p}}{\sum_{i=1}^{i=13} L(x_i, y_i)M_i\mathbf{1}_{y_i < p}}} \\
&= \frac{1}{1 + \sup_{\mathcal{D}^*} \frac{\sum_{i=1}^{i=13} L(x_i, y_i)M_i\mathbf{1}_{y_i > p}}{\sum_{i=1}^{i=13} L(x_i, y_i)M_i\mathbf{1}_{y_i < p}}}
\end{aligned} \tag{12}$$

Above, note that the events $[y_i \leqslant p]$ have been replaced by the events $[y_i < p]$. This does not change the infimum. However, now, the priors that give the infimum value for $P(X_B \leqslant p \mid n_A, n_B)$ will do so by their value for the posterior probability $P(X_B < p \mid n_A, n_B)$.

The rational function in the denominator of (12) can be made as large as possible, if the M_i probabilities within each region R_i are allocated to points (x_i, y_i) that, asymptotically, make terms like $L(x_i, y_i)M_i\mathbf{1}_{y_i<p}$ as small as possible, and terms like $L(x_i, y_i)M_i\mathbf{1}_{y_i>p}$ as large as possible. These (x_i, y_i) locations tend toward certain "corners" of each region, as approximately depicted in Fig. 3. These unique corners within each region are *limit points* of the respective regions – they are arbitrarily well-approximated[1] by countable sequences of points within the respective region [3,4].

Given these limit points at which probability mass may lie for optimality, and given the PK in each proposition, a prior that gives the supremum of the expression in the denominator of (12) can be solved for, either numerically or analytically.

Numerical Solutions: To numerically obtain the conservative priors, cast the problem in the denominator of (12) as a *linear-fractional programming* problem,

[1]Using the "open balls" topology associated with 2D Euclidean space.

$$\underset{M_i}{\text{maximise}} \ \frac{\boldsymbol{c}^T\boldsymbol{M}}{\boldsymbol{d}^T\boldsymbol{M}}, \text{ subject to } \boldsymbol{AM} \leqslant \boldsymbol{b} \tag{13}$$

where the thirteen M_is are variables (denoted as a column vector $\boldsymbol{M}$), the row vectors $\boldsymbol{c}^T$ and $\boldsymbol{d}^T$ each have 13 components – consisting of the $L(x_i, y_i)$ polynomials and zeroes – and the matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$ ensure that the inequality $\boldsymbol{AM} \leqslant \boldsymbol{b}$ constrains the $\boldsymbol{M}$ to satisfy the PK in each proposition. This linear fractional programming problem can be translated into a more convenient, equivalent linear programming problem via the *Charnes-Cooper transformation* [5] and solved.

Analytical Solutions: Alternatively, these worst-case priors may be deduced by systematically placing masses at the preferred "corners" within the regions, in a manner that achieves conservatism and is consistent with prior knowledge. We illustrate this by deriving the worst case priors in Fig.s 1 and 2 – respectively, these solve propositions 1 and 2.

1) proof of proposition 1: to derive prior Fig. 1, consider the PK for proposition 1, when $\varphi > \theta > 1-\theta > 1-\varphi$ and $\alpha \leqslant p$. Lemma 1 implies the conservative prior must conform to Fig. 3a. The parameter inequalities determine the probability masses at the 13 regional points as follows (see Fig. 4).

In Fig. 3a, because of PK 3, the complement of the triangular "φ" region contains probability $1-\varphi$. To be conservative, all of this probability mass should be allocated to the point (p_l, p), where X_A is the most reliable it can be (i.e. p_l) while X_B is not reliable enough, but only just[2] (see Fig. 4a). Since PK 1 and 2 mean $P(X_A \leqslant \varepsilon) = \theta$, then probability mass $\varphi + \theta - 1$ must be assigned to $(\varepsilon, \varepsilon)$ (see Fig.s 3a and 4b). And, since PK 1 and 2 also apply to version B (i.e. $P(X_B \leqslant \varepsilon) = \theta$), mass $1-\varphi$ must be located at (α, ε) (see Fig. 4c). Finally, PK 3 implies the triangular region must have mass φ; therefore, mass $\varphi - \theta$ must be assigned to (α, α) (see Fig. 4d). Thus, we obtain conservative prior Fig. 1.

This distribution gives the infimum in (4) by computing $P(X_B < p \mid n_A, n_B)$ for this distribution. Using the standard expression for $P(X_B < p \mid n_A, n_B)$, we have

$$\begin{aligned} P(X_B < p \mid n_A, n_B) &= \frac{P(X_B < p,\ n_A \,\&\, n_B \text{ successes})}{P(n_A \,\&\, n_B \text{ successes})} \\ &= \frac{\mathbb{E}[\, L(X_A, X_B)\mathbf{1}_{X_B<p}\,]}{\mathbb{E}[\, L(X_A, X_B)\,]} \\ &= \frac{Num}{Num + L(p_l, p)(1-\varphi)} \end{aligned}$$

where $Num := L(\alpha,\alpha)(\varphi - \theta) + L(\varepsilon,\varepsilon)(\varphi + \theta - 1) + L(\alpha,\varepsilon)(1-\varphi)$.

2) proof of proposition 2: to derive conservative prior Fig. 2, consider the PK supporting proposition 2 when $\phi > \theta$ and $p < \alpha < 1$. Lemma 1, again, implies the conservative prior must conform to Fig. 3b. The stated constraints can now be used to determine the probability masses at the 13 regional points as follows (see Fig. 4).

[2]The consequence of this allocation is that no mass, from the complement of the "φ" region, contributes to the denominator of the "sup" expression in (12). Instead, all of this mass contributes to the numerator.

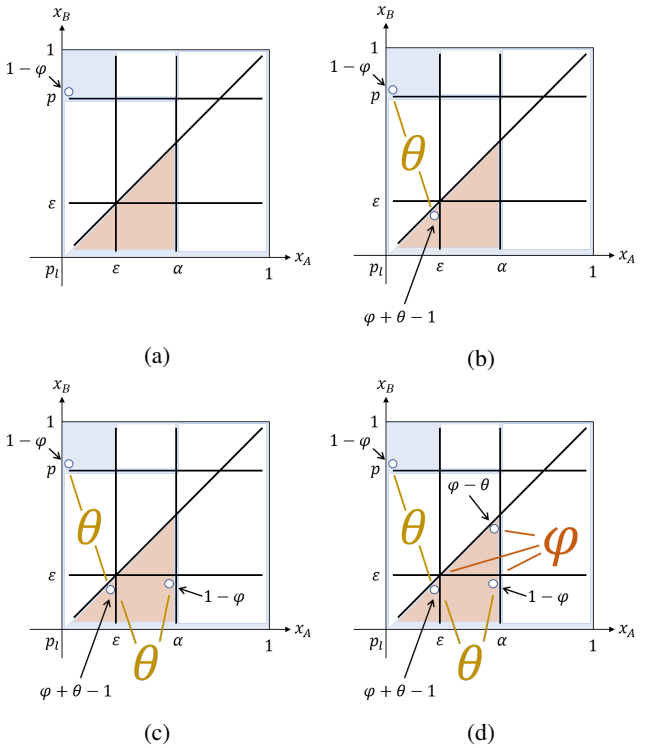


Fig. 4: A sequence of probability mass allocations that result in prior Fig. 1, which solves proposition 1. After each mass allocation, the white areas indicate regions where mass *cannot* lie.

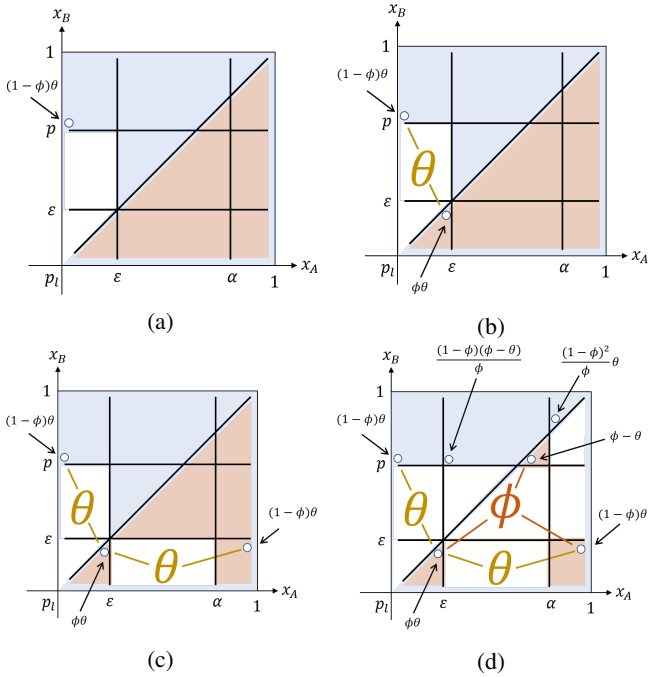


Fig. 5: A sequence of probability mass allocations that result in prior Fig. 2, which solves proposition 2. As in Fig. 4, white areas indicate regions where mass *cannot* lie.

In Fig. 5a, because of PK 1 and 2 (concerning the engineering goal for version A), the region $[p_l \leqslant x_A \leqslant \varepsilon,\, p_l \leqslant x_B \leqslant 1]$ must contain mass θ. So, by PK 4, there must be probability mass $(1-\phi)\theta$ within this region but above the main diagonal, and mass $\phi\theta$ within this region but below the main diagonal. The location (p_l, p) gives the largest "L" value in the numerator of the "sup" expression in (12) – so, to be conservative, one must assign $(1-\phi)\theta$ to this location (see Fig. 5a). And, by Fig. 3b, mass $\phi\theta$ must then be assigned to $(\varepsilon,\, \varepsilon)$ (see Fig. 5b).

PK 1 and 2 (concerning the engineering goal for version B) now imply there must be mass $(1-\phi)\theta$ in the region $[\varepsilon < x_A \leqslant 1,\, p_l \leqslant x_B \leqslant \varepsilon]$. The location $(1,\, \varepsilon)$ within this region has an "L" value of zero – the smallest "L" value in the denominator of the "sup" expression in (12). So, assign mass $(1-\phi)\theta$ here (see Fig. 5c).

Since the total mass below the main diagonal is ϕ, the remaining unallocated mass below the main diagonal must be $\phi - \theta$. The location $(p,\, p)$ gives the largest "L" value (amongst locations below the main diagonal) in the numerator of the "sup" expression – so, assign all of $\phi - \theta$ to this location. Doing this also ensures that the "sup" expression's denominator is kept as small as possible, since none of this mass will contribute to the denominator.

And finally, within the regions $[\varepsilon < x_A \leqslant \alpha,\, p_l \leqslant x_B \leqslant 1]$ and $[\alpha \leqslant x_A \leqslant 1,\, p_l \leqslant x_B \leqslant 1]$, the respective locations $(\varepsilon,\, p)$ and $(\alpha,\, \alpha)$ give the largest "L" values (amongst locations in their respective regions that lie above the main diagonal) in the numerator of the "sup" expression. Because of PK 4, masses $\frac{(1-\phi)(\phi-\theta)}{\phi}$ and $\frac{(1-\phi)^2}{\phi}\theta$ must be allocated to these locations (see Fig. 5d). Thus, we obtain conservative prior Fig. 2.

This distribution gives the infimum in (5) by computing $P(X_B < p \mid n_A, n_B)$ for this distribution. Using the standard expression for $P(X_B < p \mid n_A, n_B)$, we have

$$\begin{aligned} P(X_B < p \mid n_A, n_B) &= \frac{P(X_B < p,\, n_A \,\&\, n_B \text{ successes})}{P(n_A \,\&\, n_B \text{ successes})} \\ &= \frac{\mathbb{E}[\, L(X_A, X_B)\mathbf{1}_{X_B<p}\,]}{\mathbb{E}[\, L(X_A, X_B)\,]} \\ &= \frac{L(\varepsilon,\varepsilon)\phi^2\theta}{L(\varepsilon,\varepsilon)\phi^2\theta + Den} \end{aligned}$$

where $Den := L(p,p)\phi(\phi-\theta) + L(p_l,p)\phi(1-\phi)\theta + L(\varepsilon,p)(1-\phi)(\phi-\theta) + L(\alpha,\alpha)(1-\phi)^2\theta$.

References

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

[2] E. Moreno and J. A. Cano, "Robust bayesian analysis with ε-contaminations partially known," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 53, no. 1, pp. 143–155, 1991. [Online]. Available: http://www.jstor.org/stable/2345731

[3] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., ser. International series in pure and applied mathematics. McGraw-Hill, 1976.

[4] E. T. Copson, *Metric Spaces*, ser. Cambridge Tracts in Mathematics. Cambridge University Press, 1968.

[5] A. Charnes and W. W. Cooper, "Programming with linear fractional functionals," *Naval Research Logistics Quarterly*, vol. 9, no. 3-4, pp. 181–186, 1962.