Evaluating Safety-Critical Systems: A (Conservative) Bayesian's View

Dr. Xingyu Zhao [phonetically, Sing-You]

@ TAS Node in Resilience

19-June-2023

Assistant Professor in Safety-Critical Systems,
V&V Group | Intelligent Vehicles | WMG | University of Warwick,
https://www.xzhao.me/

The talk covers joint works with…

- PhD and PDRAs
  - PhD (started in 2013): Centre for Software Reliability, City University of London
    - Probabilistic assessment of safety-critical software (Nuclear PPS)
  - PDRA: Heriot-Watt University
    - Probabilistic verification on Robotics and Autonomous Systems (RASs)
  - Programme Fellow: University of York
    - Assuring Autonomy International Programme (AAIP)
  - PDRA: University of Liverpool
    - ~~DL testing, XAI, safety analysis for Learning-Enabled Systems (LESs)~~
- Lecturer in AI at University of Liverpool since 2021
- Assistant Professor at Warwick in June 2023
- 11 related publications (listed at the end)

Why assessing Safety-Critical Systems is hard?

WARWICK
THE UNIVERSITY OF WARWICK

… compared to assessing the fairness of a coin

	Assessing A Coin	Assessing Safety-Critical Systems (SCSs)
Metrics	probability of seeing tail in the next toss, ~0.5	E.g. pfd (prob. of failure per demand), SIL4, ~10^-4
Amount of testing	A few trials of flipping the coin	Impractical number of tests needed, and expensive
Assumptions in the stochastic process	No doubts in assuming a Bernoulli Process	Complex and may have doubts in the assumptions
Prior knowledge (PK)	Easy to elicit and formalise	More careful/reluctant to express; limited PK; non-informative priors is misleading…
Conjugacy	Why not	Introducing implicit knowledge/assumptions
Application context	In a simple gambling game?	Complex, interactive, dynamic, .e.g., RAS missions

Why assessing Safety-Critical Systems is hard?

WARWICK
THE UNIVERSITY OF WARWICK

Correspondingly, 6 (correlated) questions:

1. How to practically assessing ultra-high reliability, with clear definitions of metrics?
2. How to effectively model failure-free/sparse-failure evidence?
3. How to incorporate doubts on assumptions in the stochastic failure process?
4. How to incorporate limited, partial/vague prior knowledge?
5. Can we get rid of conjugacy in the reasoning?
6. How to model SCSs in a more dynamic, interactive application context?

- Q1, Q2 and Q6 are specific to SCSs; Q3 is generic to any statistical inference; Q4 and Q5 are fundamental to any Bayesian methods;
- Have we solved them? ☺

``The RAND study''

[HTML] **Driving** to **safety**: How **many miles** of **driving** would it take to demonstrate autonomous vehicle reliability?

N Kalra, SM Paddock - Transportation Research Part A: Policy and Practice, 2016 - Elsevier

… of **miles** of **driving** that would be needed to provide clear statistical evidence of autonomous vehicle **safety**. … injuries are rare events compared to vehicle **miles** traveled, we show that fully …

Save Cite Cited by 1248 Related articles All 7 versions

- Context:
  - AVs tested on public roads in the US for years; millions of autonomous miles have been driven
- Metrics, inc.
  - *probability of fatality-event per driven mile (pfm)*
- Method: A common frequentist statistical inference model
  - For claiming AVs is XX times safer than human with levels of confidence
  - seeing evidence millions/billions of autonomous miles driven
- Conclusions: inc. Operational testing alone is impractical
  - E.g., to claim, with 95% conf., AVs are as safe as human, it needs 275 millions of fatality-free miles.

We agree with RAND, but…

- The main message is not new, while RAND nicely reformulated it for AVs

The **infeasibility** of **quantifying** the **reliability** of life-critical real-time software
RW Butler, GB Finelli - IEEE Transactions on Software …, 1993 - ieeexplore.ieee.org
… software **reliability**. Research efforts started with **reliability** growth models in the early 1970's. In recent years, an emphasis on developing methods that enable **reliability quantification** of …
Save Cite Cited by 526 Related articles All 24 versions

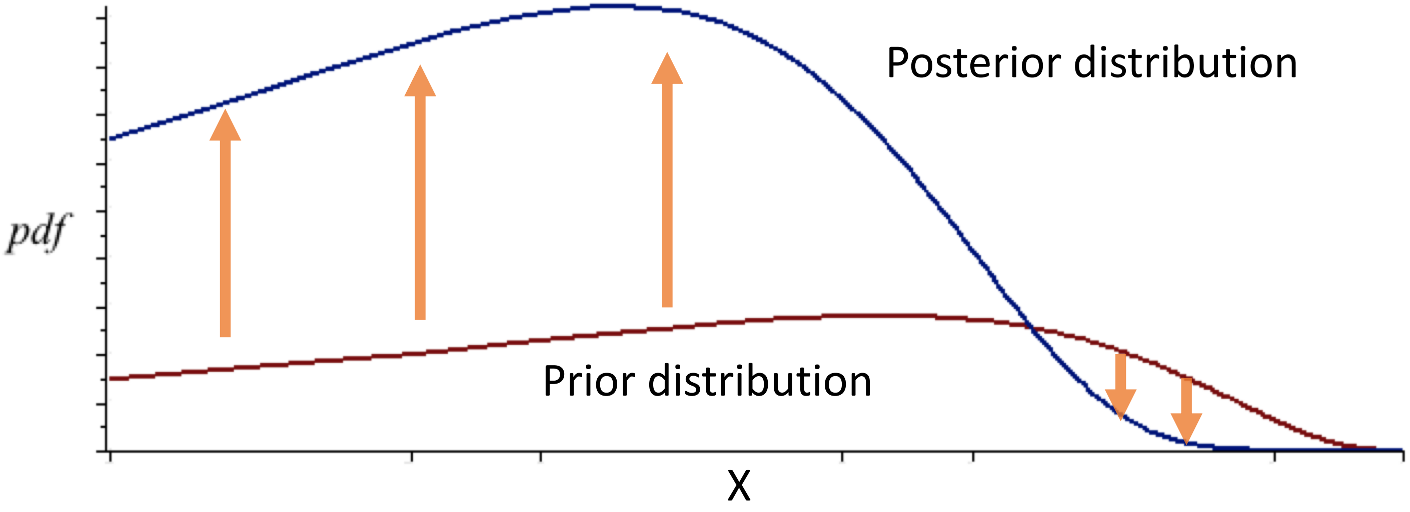
Validation of **ultra-high dependability** for software-based systems
B Littlewood, L Strigini - Communications of the ACM, 1993
… **dependability** required. This can be very difficult, as we shall see later; **validating**: gaining confidence that a certain **dependability** … levels of **dependability** that can currently be **validated**. …
Save Cite Cited by 420 Related articles All 23 versions

- No one puts SCSs in operational/statistical testing without strong prior confidence in safety
- So how to incorporate prior knowledge (PK) in safety in statistical inference?
  - In a statistical principled way
  - Bayesian inference seems to be a good answer

Bayesian inference, a reminder…

Seeing data, the prior distribution is ``scaled'' into a posterior distribution, according to the likelihoods.

- Where to get the priors?
  - Non-informative priors (for SCSs)?
- What is the Likelihood?
  - Poisson/Binomial/Bernoulli Process?
- What forms of posteriors are of practical interest?
  - A complete post. dist. is costly/luxury
  - Posterior mean
  - Posterior confidence bounds



$$f(x|\text{data}) \propto L(\text{data}|x)\, f(x)$$

With only limited, partial prior knowledge…

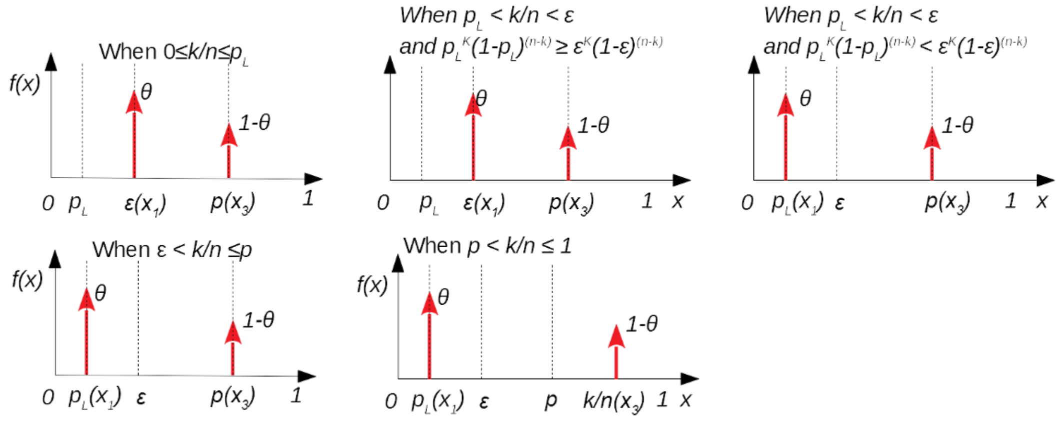
$$Pr(X \leqslant p \mid k\&n) \qquad (1)$$

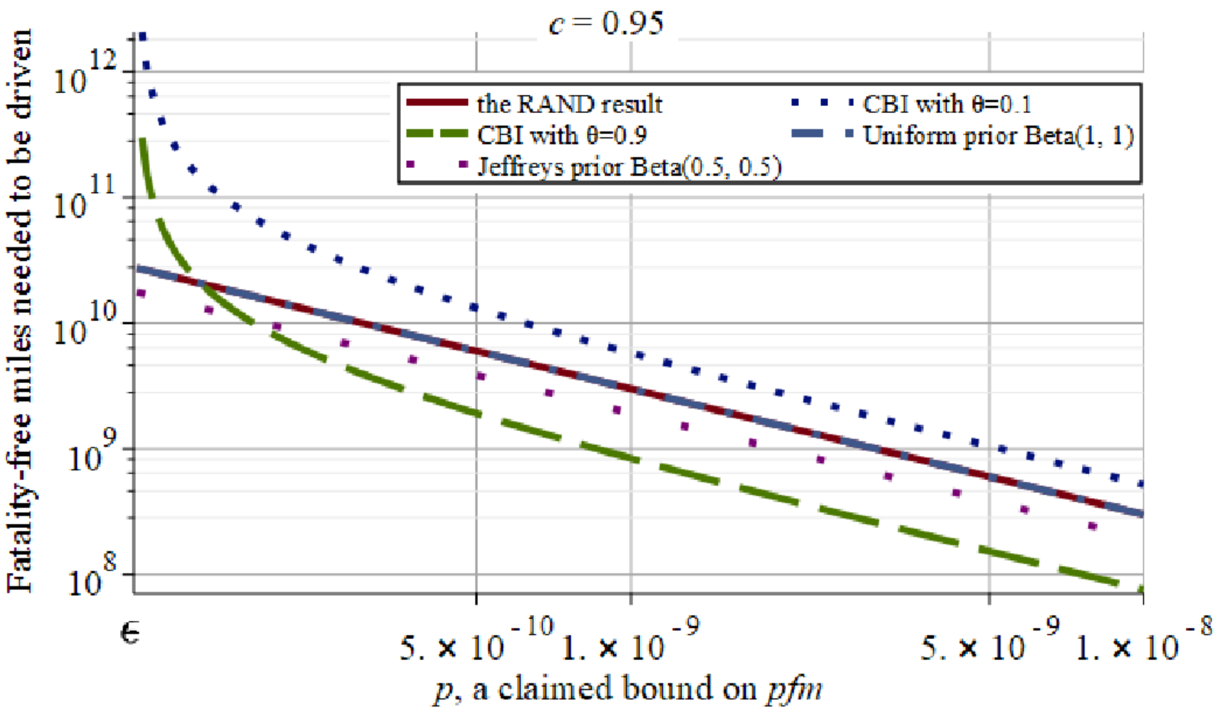
$$Pr(X \leqslant \epsilon) = \theta \qquad (2)$$

- Posteriors:
  - a posterior confidence bound in a required *pfm p*, after seeing *k* fatality-events in *n* driven miles, cf. Eq. (1).
- Examples of PK in Eq. (2)
  - far from being specific about a singe, complete *f(x)*.
  - an infinite set of distributions satisfying Eq. (2)
- Bayesian inference is a new optimisation problem
  - To minimise (1), subject to (2), what is the corresponding f(x)?

With only partial priors, solutions

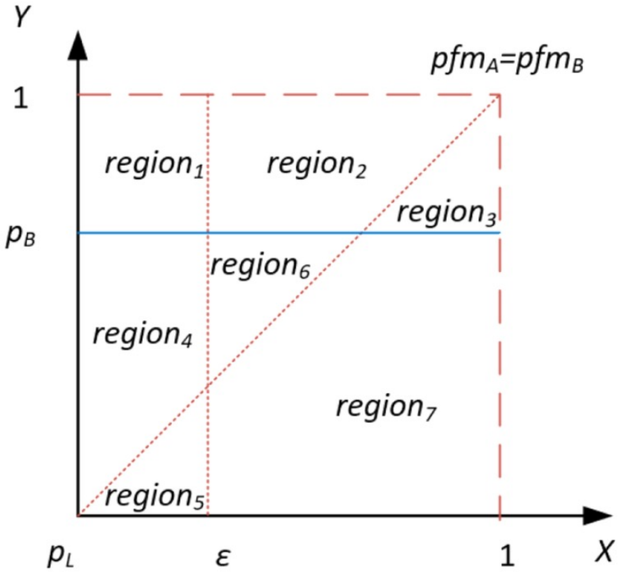
- Optimisation can be analytically solved.
  - formal guarantees on conservatism
- NB, no parametric families, no conjugacy, being different to:
  - Robust Bayesian inference
  - Imprecise probabilities
- References [2,4,5,10,11]
  - for different PKs/posteriors/ observations/applications





More PK, on versions of SCSs

- Safety regulation principles
  - GALE: globally at least equivalent (French Railway, US FDA medical devices)
  - A high confidence that "the new system should be no unsafe than existing systems"
- Formalise such knowledge as PK $Pr(Y \leqslant X) = \phi,$
  - A joint prior dist. of failure probabilities of two versions
  - Probability mass M_i in different region i encodes PKs, e.g.,
    - $M5 + M7 + M3 = \phi$ (New B is no unsafe than old A)
    - $M1 + M4 + M5 = \theta$ (Marginal conf. bound. on old A)
  - Again, constraints on prior distributions.



More PK, on versions of SCSs

- Similarly, another optimisation problem, but 2D:
  - what is the worst-case joint prior distinction, that

$$\begin{aligned} \underset{\mathcal{D}}{\text{minimise}} \quad & Pr(Y \leqslant p_B | n_A, n_B) \\ \text{subject to} \quad & Pr(X \leqslant \epsilon) = \theta, \\ & Pr(Y \leqslant X) = \phi, \end{aligned}$$

- Refs [2, 3, 6]
  - Various forms of priors/posterior
  - More interesting scenarios/RQs

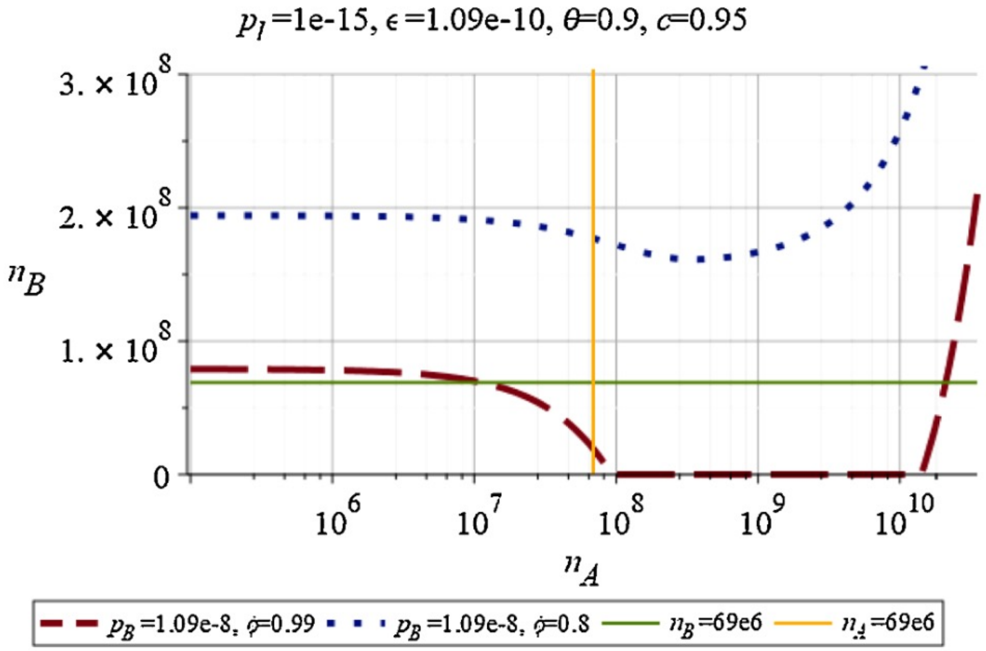


Fig. 7. Fatality-free miles that need to be driven in City-B (by Version-B), given that n_A fatality-free miles have been driven in City-A (by Version-A) in scenarios Q4 and Q5. The straight horizontal and vertical lines show the amount of road testing that would yield the target confidence $c = 95\%$ in the required bound $pfm \leqslant 1.09e-8$ in the single-version, single-city scenario of Q1.

OK, what about the likelihood?

- Allowing doubts in fundamental assumptions behind a likelihood

- How to formalise?
  - Using Klotz's model to relax i.i.d.
    - while i.i.d. is a special case ($x = \lambda$)
  - Doubts in iid, combinations of x, λ
  - ``instead of a single likelihood function, we introduce a set of likelihoods allowing doubts''
    - Loosely speaking only
- Optimisation, over a set of likelihoods

Statistical inference in Bernoulli trials with dependence
J Klotz - The Annals of **statistics**, 1973 - JSTOR
A model for **Bernoulli trials** with Markov dependence is developed which possesses the usual frequency parameter p = P[X i = 1] and an additional dependence parameter λ = P[X i = 1 …
☆ Save 99 Cite Cited by 148 Related articles All 4 versions

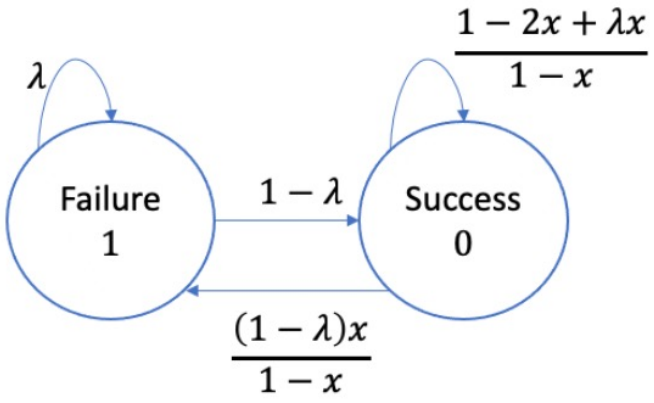
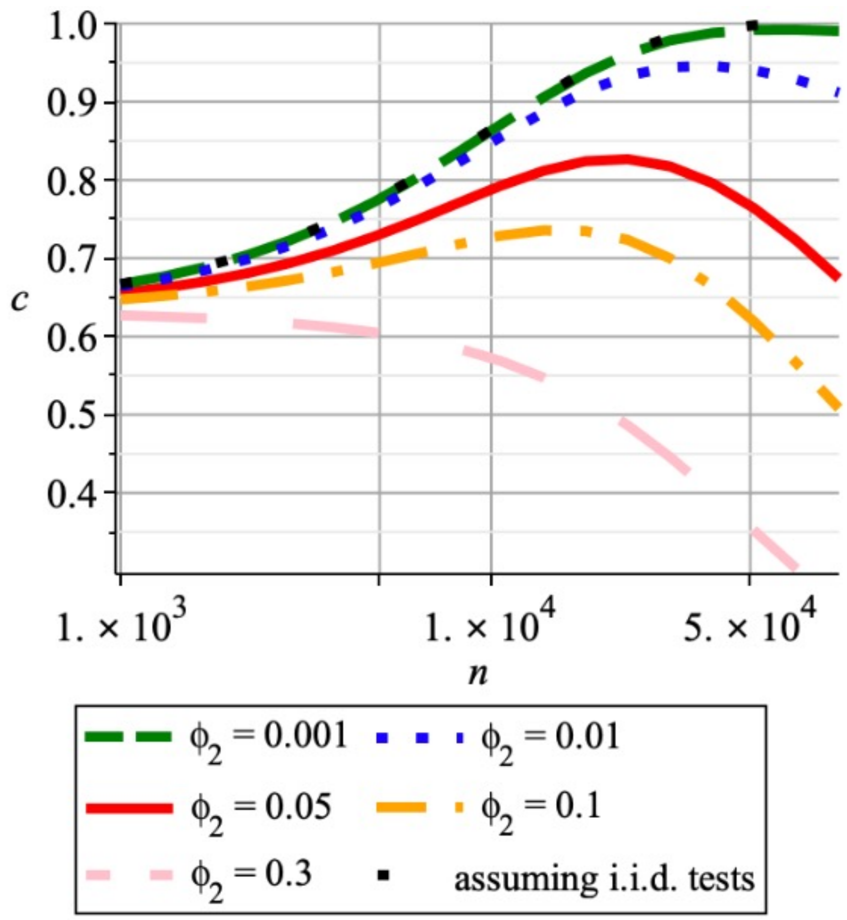


Fig. 1: The Klotz model with dependent Bernoulli trials [17].

OK, what about the likelihood?

- X axis: number of tests
- Y axis: poster confidence in a bound
- Curves representing levels of doubts (ϕ_2) in i.i.d.
  - i.i.d. is the special case (black dotted)
- More interesting results
  - E.g., i.i.d. is not always optimistic
  - In some cases, posteriors not sensitive to doubts at all.
  - Ref. TSE [1], QRE (under review)



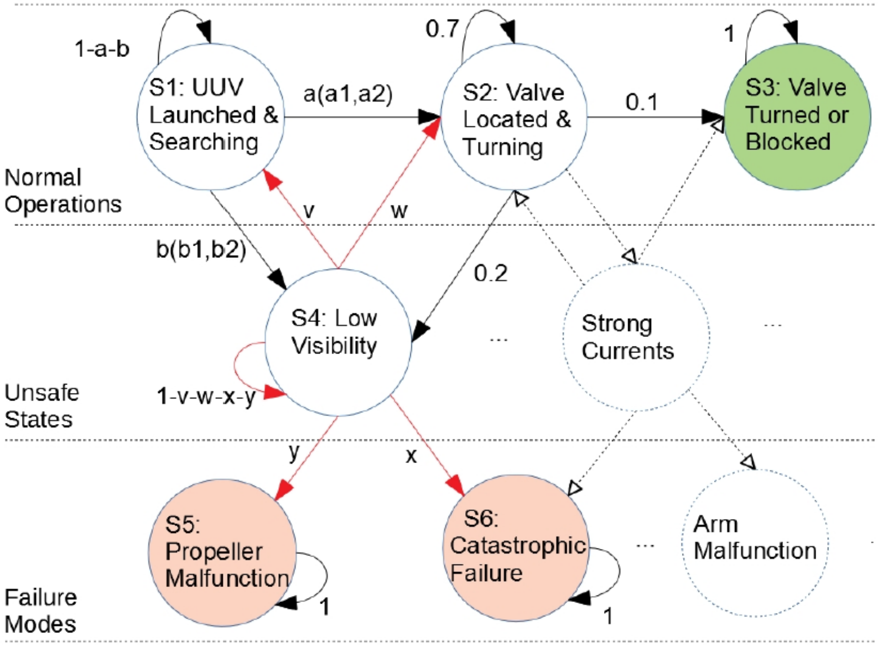
A quick summary, so far….

- A set of priors
  - representing limited, partial PK
- A set of likelihoods
  - encoding doubts in assumptions behind
- Guaranteed conservatism, for different forms of posteriors
- Bayesian inference as an optimisation
  - Finding the worst-case combination of priors and likelihood, for the given posters
  - No assumptions on parametric families/conjugacy
  - Analytical solutions
    - For runtime Bayesian estimators, next…

$$f(x|\text{data}) \propto L(\text{data}|x)f(x)$$

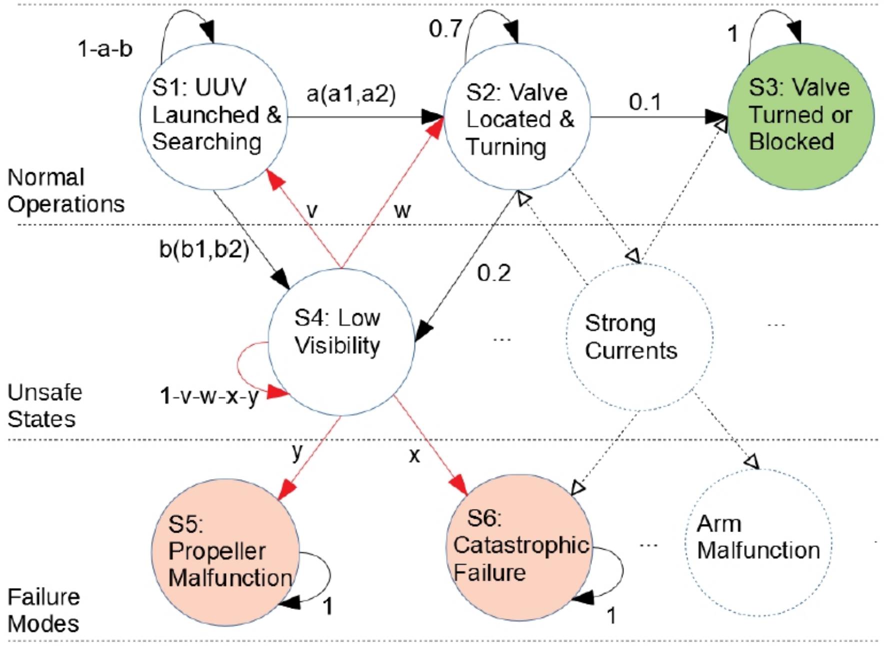
Bayesian estimators in Prob. Model Checking

- Previously, modelling SCSs at a very high-level
  - Only one variable of reliability for each SCS…
- In Model-driven Engineering,
  - System behaviours, e.g., DTMC/CTMC/MDP
  - Properties, e.g., PCTL/CSL
  - Verification, e.g., PMC (offline or at runtime)



``… only as good as the formal model''

- How to get an accurate Markov model?
  - E.g., the key transition parameters.
- What if the formal model is subject to changes?
  - How to do accurate change-point detection?
  - What is the new formal model after the change?
- Formulated as statistical inference problems
  - runtime Bayesian estimators with "fresh data"
  - aforementioned ideas for fundamental problems
  - efficient enough for runtime verification



Case studies of UUVs

- A video demo
  - https://drive.google.com/file/d/1fLZ3Bip8Y0KRiaWfMOMRZStPbPpCdHqy/view?usp=sharing
- Refs [7,9]

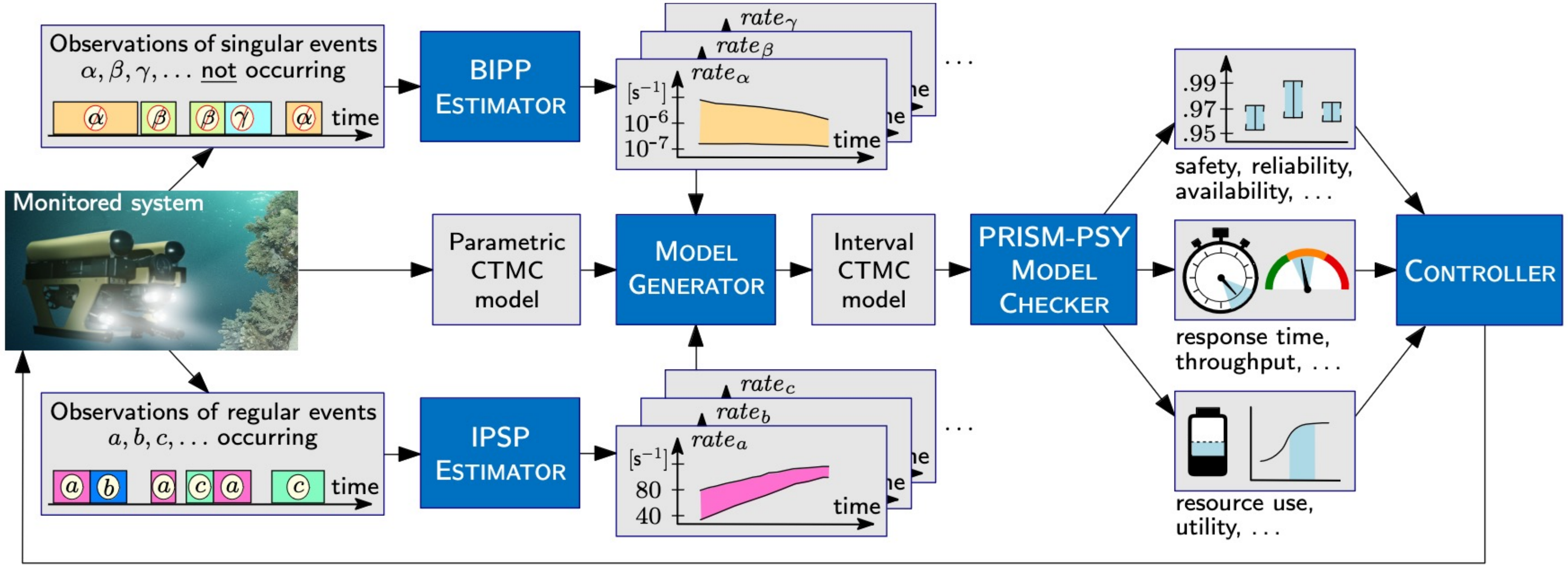


Figure 1: The integration of BIPP and IPSP Bayesian inference with interval CTMC model checking supports the online robust quantitative verification and reconfiguration of autonomous systems under parametric uncertainty.

Take home message…

- When doing statistical testing for SCSs, no one starting from nothing…
  - too risky; too expensive; how to incorporate PK (in a principled way)?—Bayes
- Bayesian inference is hard to apply…
  - Elicit and formalise a prior distribution—at least, something as limited as a conf. bound;
    - Don't simply use uniform/non-informal priors for SCSs.
  - Doubts in assumptions of the likelihood—model the doubts!
    - Two use cases: conservative claims, model validation (i.e., claims should be insensitive to doubts)
  - Conjugacy/parametric-families—we don't need it!
- Versatile and efficient at different abstraction levels
  - Estimate some single reliability metric
  - Bayesian estimators for the underlying models in model-driven engineering

Publications related to this talk

Journal Publications

[1] K. Salako and **Zhao, X.**, "The Unnecessity of Assuming Statistically Independent Tests in Bayesian Software Reliability Assessments," *IEEE Tran. on Software Engineering*, vol. 49, no. 4, pp. 2829–2838, 2023

[2] **Zhao, X.**, K. Salako, L. Strigini, V. Robu, and D. Flynn, "Assessing safety-critical systems from operational testing: A study on autonomous vehicles," *Information and Software Technology*, vol. 128, p. 106393, 2020

[3] B. Littlewood, K. Salako, L. Strigini, and **Zhao, X.**, "On reliability assessment when a software-based system is replaced by a thought-to-be-better one," *Reliability Engineering & System Safety*, vol. 197, p. 106752, 2020

[4] **Zhao, X.**, 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

[5] **Zhao, X.**, 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

Conference Publications

[6] K. Salako, L. Strigini, and **Zhao, X.**, "Conservative confidence bounds in safety, from generalised claims of improvement & statistical evidence," in *51st Annual IEEE/IFIP Int. Conf. on Dependable Systems and Networks (DSN'21)*, pp. 451–462, IEEE, 2021

[7] **Zhao, X.**, R. Calinescu, S. Gerasimou, V. Robu, and D. Flynn, "Interval change-point detection for runtime probabilistic model checking," in *Proc. of the 35th IEEE/ACM Int. Conf. on Automated Software Engineering (ASE'20)*, pp. 163–174, ACM, 2020

[8] **Zhao, X.**, A. Banks, J. Sharp, V. Robu, D. Flynn, M. Fisher, and X. Huang, "A safety framework for critical systems utilising deep neural networks," in *Computer Safety, Reliability, and Security (SafeComp'20)*, vol. 12234 of *LNCS*, pp. 244–259, Springer, 2020

[9] **Zhao, X.**, V. Robu, D. Flynn, F. Dinmohammadi, M. Fisher, and M. Webster, "Probabilistic model checking of robots deployed in extreme environments," in *Proc. of the 33rd AAAI Conference on Artificial Intelligence (AAAI'19)*, vol. 33, (Honolulu, Hawaii, USA), pp. 8076–8084, 2019

[10] **Zhao, X.**, V. Robu, D. Flynn, K. Salako, and L. Strigini, "Assessing the safety and reliability of autonomous vehicles from road testing," in *Proc. of the 30th Int. Symp. on Software Reliability Engineering (ISSRE'19)*, (Berlin, Germany), pp. 13–23, IEEE, 2019. **(Best Paper Nominee: 3/134)**

[11] **Zhao, X.**, B. Littlewood, A. Povyakalo, and D. Wright, "Conservative claims about the probability of perfection of software-based systems," in *Proc. of the 26th IEEE Int. Symp. on Software Reliability Engineering (ISSRE'15)*, (Gaithersbury, MD, USA), pp. 130–140, IEEE, 2015

Thank you

- xingyu.zhao@warwick.ac.uk
- www.xzhao.me