City Research Online

City St George's, University of London

Citation: Bishop, P. G., Povyakalo, A. A. & Strigini, L. (2024). How to reason about Risk, given Inevitable Doubt on Arguments for High Dependability. Paper presented at the SAFECOMP 2024, 17-20 Sep 2024, Florence, Italy.

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/34211/

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, 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. For full details of reuse please refer to City Research Online policy.

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

How to reason about Risk, given Inevitable Doubt on Arguments for High Dependability

Peter Bishop, Andrey Povyakalo, Lorenzo Strigini
Centre for Software Reliability
City, University of London, London, U.K.
{P.Bishop, A.A.Povyakalo, L.Strigini}@city.ac.uk

***Abstract*—For highly critical systems, thorough and costly processes exist to verify that they are safe enough before they are allowed to operate. Yet any such *a priori* assessment is affected by uncertainty: it may be wrong. Examples like the Boeing 737 MAX and Fukushima underscore how badly they may, occasionally, be wrong. We argue that risk assessment should take into account, more explicitly than is now usual, this uncertainty. Basic quantitative reasoning shows how this would change how we describe the risk of operating a new system. This may set new priorities in safety assessment research. We identify some research directions that may help this community better to forecast and control risk.**

Keywords—Risk assessment, risk quantification, safety case, epistemic uncertainty.

I. Introduction

For many systems, society requires high confidence that they will function safely enough. For some of them, reliability and safety requirements (e.g. in civil aviation, nuclear power, railways, self-driving vehicles, and others) are so stringent that sufficient assurance of their satisfaction cannot be achieved purely by statistics of observed proper functioning before deployment [1]. Such systems are often complex combinations of hardware and software and people, and a dominant concern is that of subtle design faults. For instance, in civil aviation, documents AC 25.1309-1 and DO-178C [2] stipulate that for complex designs, assurance of the required dependability can be claimed on the basis of thoroughness of design and validation practices, rather than statistics. A complex regime of quality and verification practices is prescribed and enforced. Yet experience, with aircraft as well as other types of critical systems, shows that every now and then such a claim, despite all the effort to ensure it is correct, is proved in operation to have been wrong, by accidents, in the worst case, or by near misses, or by new analyses [3][7][8].

Interestingly, this issue of uncertainty about safety claims is seldom addressed in risk communication and is left mostly as a topic for the political aspects of debate, e.g., by anti-nuclear protest groups. Yet proper assessment of risk is fundamental for reasonable, ethical and economical decisions in engineering. Regulatory agencies and practitioners are of course aware of the uncertainty. Sensible precautions are prescribed, like monitoring of systems in operation, but we find hardly any technical work on understanding how exactly the uncertainty should be taken into account in risk assessment and related decisions.

II. Risk level for a new system

How should one quantify the risk incurred in, say, one flight of a newly certified aircraft type? A way of reasoning would be: the claim may be correct, i.e., the probability of catastrophic failure per flight (we will call this a *pfd*, probability of failure per demand) is at most the claimed upper bound, say q_L ("L" for "low"). On the other hand, the claim might be false (due for instance to some wrong assumption in the reasoning, some omission in hazard analysis, random errors in developing a safety argument). If so, it is difficult to tell how badly wrong it might be – how high the *pfd* could actually be. An assessor might want to say "1 is the only upper bound I can trust"; or consider that too implausible, and in practice be sure that the *pfd* can be no worse than q_H (H for high, with of course $0 < q_L < q_H < 1$).

Those stating and/or accepting the q_L claim will have some reasonably high confidence that it is true. They could think that it has a probability at least, say, p_L, of being true. In this case they can conclude that the probability of accident at the next demand is bounded above by

$$p_L \, q_L + (1 - p_L) \, q_H \qquad (1)$$

This is the simplest case. In some cases there are reasons for more complex claims, e.g. a sequence of fallback claims, $q_{L1}, q_{L2}, \ldots q_{Ln}$, with $q_L < q_{L_1} < \ldots < q_{Ln} < q_H < 1$, with associated probability of each one being true if the previous one turned out to be false. Indeed, that such more complex and "fault-tolerant" arguments are desirable is one of our conclusions. But it is useful here to explore the simplest scenario in (1).

The implications of (1) are stark: during early operation of a new critical system, a reasonable upper bound on the probability of accident per demand appears to be dominated by the second summand: the probability of the claim being wrong, times the worst-case *pfd* if it *is* wrong. For example, when $q_L = 10^{-9}$, p_L =99%, $q_H = 10^{-3}$, (1) evaluates to

$$0.99 \times 10^{-9} + 0.01 \times 10^{-3} \approx 10^{-5} \qquad (2)$$

Of course, the real value of the *pfd* is a specific number between 0 and 1, but is unknown. Assessors can say that, to the best of their knowledge, the probability of accident at the next demand is no more than shown in (1). This is the "expected value" of the *pfd*. More precisely, as q_H is an upper bound, it is the worst-case value of this expected *pfd*, within the constraints set by the parameters stated.

III. Value of operational experience

If the system is allowed to start operation, observing accident-free operation will gradually disprove the worst case assumption underlying (1): the worst-case estimate of risk, stated there, can be updated. Monitoring of operation generally aims to detect any unsafe system behaviours – not just accidents but also "near-misses" or "incidents" [7] so that improvements can be made [8][9]. But in these two pages we aim just to illustrate how observing zero undesired events may help assurance, and the limits to how much it helps.

An appropriate method was published in 2011 for quantifying this improvement [5]. According to Bayes' rule, accident-free operation increases the probability that the safety claim q_L was actually correct. This updated probability is what one should use in deciding whether to use or operate this

Position paper presented at SAFECOMP 2024, Florence, Sept. 2024
Copyright of City St George's, University of London

system further. A subtle issue arises: as we observe more and more safe operation, assuming q_H as the worst-case *pfd* is no longer conservative (because it quickly becomes unbelievable): to calculate the upper bound on risk, the *pfd* assumed for the case that the claimed bound q_L is wrong needs to shift accordingly. All this has been solved [5], and we can ignore the mathematical detail: what matters is that the worst-case assessment of risk gets more favourable as operation accumulates without accidents (or incidents or near misses, if, as usual, the monitoring regime aims to detect all of them).

Fig. 1 illustrates this effect. The x axis represents the number of demands completed in operation with no accidents: the amount of favourable evidence from operation. The y axis gives the worst-case probability of accident at the next demand. However this updated upper bound will only approach q_L asymptotically, as the amount of operation observed tends to infinity. And this trend may feel quite slow: e.g., with initial strong confidence of 80% in a claim of $pfd \leq 10^{-5}$, to believe a bound of 10^{-4} on expected *pfd*, taking into account that 20% doubt, requires waiting for 1000 accident-free demands. It must be noted that demonstrating an even lower q_L, with the same p_L, would not reduce worst-case risk in early operation.

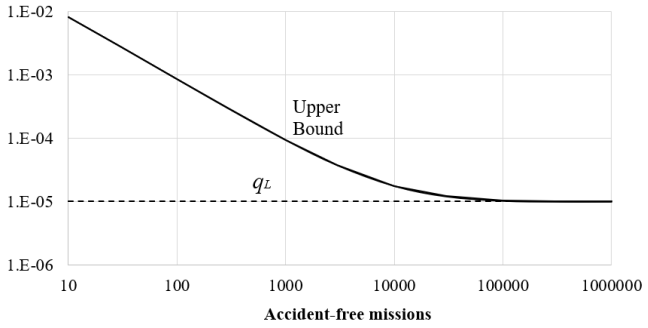

Fig. 1. Upper bound probability of accident per mission versus number of observed accident-free missions, for $q_L = 10^{-5}$ with confidence $p_L = 80\%$.

IV. Discussion

Epistemic uncertainty about safety claims is an "elephant in the room" in current practice. If it is taken into account, the assessment of worst-case risk may be orders of magnitude worse than is claimed in a conventional safety argument (however carefully produced). We have proposed a possible formalisation of reasoning about the fact that carefully demonstrated safety claims may be wrong, that they seem usually not to be *very dangerously* wrong, and that cautious operation allows one to believe in progressively improving upper bounds on risk. A "reasonable" bound on the probability of accident per demand in early operation is practically unaffected by the very low *pfd* formally claimed, and accepted by regulatory authorities, but depends heavily on the probability of errors in those claims.

Current regulatory and industrial practice for safety critical systems in practice assumes that initial safety claims may be wrong, and includes wise precautions for that case, e.g., monitoring of operation to revise the analyses and calculations behind those claims, and cautious deployment of new systems in small numbers, to "bootstrap" the confidence that can be had in their safety [9]. However, not acknowledging the existence of epistemic uncertainty [12] in formal quantitative claims deprives decision makers of an important tool for insight and for basic error checking on critical decisions. Formal mathematical description of the reasoning in these decisions may help to clarify what risks are being taken, and what decisions in data collection and in design practice might help to control them. Last but not least, the public debate about controversial, potentially beneficial but risky technologies [3][4], might benefit from a transition, however difficult, to more explicit reasoning about the epistemic uncertainty involved in claims of extreme safety.

We and our colleagues have been publishing examples of the style of formal reasoning proposed here (which we call "conservative Bayesian inference": e.g., [9][11]); new and more applied research is needed, though. A company or regulator will want to study, e.g., how often safety claims have been wrong, and how badly, to get an idea of the parameters in (1) and similar formulas. System design with layers of defence can mitigate the risk from excessive claims made for the first layer; likewise, we expect that data collection (on the history of an industrial sector, on the trustworthiness of various forms of assurance techniques and of arguments based on them) can both inform the use of (1) and support more complex reasoning based on a chain of claims, q_{L1}, q_{L2}, … q_{Ln} as outlined earlier.

References

[1] B. Littlewood and L. Strigini, "Validation of ultra-high dependability for software-based systems," CACM, vol. 36, pp. 69–80, 1993.

[2] Requirements and Technical Concepts for Aviation (RTCA), DO-178C: Software Considerations in Airborne Systems and Equipment Certification, 2011.

[3] S. Wheatley, B. Sovacool, and D. Sornette, "Of disasters and dragon kings: A statistical analysis of nuclear power incidents & accidents," 2015. [Online]. Available: https://arxiv.org/abs/1504.02380

[4] J. Downer, "Disowning Fukushima: Managing the credibility of nuclear reliability assessment in the wake of disaster," Regulation & Governance, vol. 8, no. 3, pp. 287–309, 2014.

[5] 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 Trans. Software Eng., vol. 37, no. 5, pp. 708–717, 2011.

[6] P. Bishop, "Does software have to be ultra reliable in safety critical systems?" in Computer Safety, Reliability, and Security, F. Bitsch, J. Guiochet, M. Kaaniche, Eds. Springer, 2013, pp. 118–129.

[7] Federal Aviation Administration (FAA), Accident and Incident Data. https://www.faa.gov/data_research/accident_incident

[8] Federal Aviation Administration (FAA), Airworthiness Directives. https://www.faa.gov/regulations_policies/airworthiness_directives

[9] M. Shooman, "Avionics software problem occurrence rates," in ISSRE'96, Seventh International Symposium on Software Reliability Engineering. White Plains, New York, U.S.A.: IEEE Computer Society Press, 1996, pp. 55–64.

[10] P. Bishop, A. Povyakalo, and L. Strigini, "Bootstrapping confidence in future safety from past safe operation," 33rd International Symposium on Software Reliability Engineering (ISSRE). IEEE, 2022, pp. 97–108.

[11] L. Strigini and A. Povyakalo, "Software fault-freeness and reliability predictions," in Computer Safety, Reliability, and Security, F. Bitsch, J. Guiochet, and M. Kaaniche, Eds., Springer, 2013, pp. 106–117.

[12] C. Fox and G. Ülkümen "Distinguishing Two Dimensions of Uncertainty" in Brun, W., Keren, G., Kirkebøen, G., & Montgomery, H. (2011). Perspectives on Thinking, Judging, and Decision Making. Oslo: Universitetsforlaget.