Practical Statistical Evaluation of Critical Software

Peter Bernard Ladkin

University of Bielefeld and Causalis Limited

2 February 2016

Prolegomena

- The work reported in the paper is joint with

  Bev Littlewood of City University London

- This is my talk. I haven't asked Bev if he approves of it
- It's a speech. There will be a White Paper to follow, when I have accumulated comments

Motivation: IEC 61508-7 Annex D

- IEC 61508-7:2010 Annex D, First paragraph:
  This annex provides initial guidelines on the use of a probabilistic approach to determining software safety integrity for pre-developed software based on operational experience. This approach is considered particularly appropriate as part of the qualification of operating systems, library modules, compilers and other system software. The annex provides an indication of what is possible, but the techniques should be used only by those who are competent in statistical analysis.
- The first and last sentences are accurate. The second woefully misleading.
- There are technical things wrong also: it mixes up Bernoulli and Poisson process models
- It has been that way for nearly twenty years (unchanged since the 1997 edition)

Options and Progress

- Make minimal corrections to Annex D, then stop.
- Make corrections and reduce scope; convene new IEC committee to produce a Technical Report/Specification with more detail.
- The German National Committee has formed a Working Group, with international guests, whose output will feed in to the IEC 61508-3 Maintenance Team (November 2015).

Statistics in General

There are two ways to approach statistics.

- (Many math and engineering undergrads; all social scientists):
  Formulas, manipulate;
  numbers, plug in;
  significance tests and confidence levels: plug in.

- (Tukey and sensible people):
  I've got some numbers.
  What are they telling me?
  How can I get them to talk?
  Is what they say what I want to hear?

Statistics of Critical Systems

Is this history/are these numbers telling me that this system is adequately safe to use?

(Weaker) Are the numbers supporting such an argument substantially?

There are two answers:

- The aerospace answer
- The IEC 61508 answer

The Aerospace Answer

- Statistical reasoning allows you to draw conclusions to high confidence but not certainty.
- The regulatory performance requirement is absolute.
- That would be equivalent to 100% confidence.
- Statistical reasoning generally cannot give you that.
- However, the AMC allows you to show adequate safety-reliability short of certainty, and have this accepted as demonstration of compliance.
- An anecdote: the 'highly-reliable" autopilot
  - Punchline: PBL gets thrown off a mailing list

Another anecdote: Ariane 5

- Lots of lessons. We'll just learn one.
- A SW routine threw up a run-time exception, on all channels.
- These run-time exceptions were neither anticipated nor handled.
- The kit was thereby lost.
- It happened because one input parameter had different values from previous experience.
- To reason from previous to future experience, one (just one) strict requirement is that **all** inputs must have values seen in the history
- Another requirement for Bernoulli/Poisson reasoning: with the same frequency
- Lesson: you can't just muck about with "close enough"

Aerospace and Statistics

Many aerospace engineers say: don't touch statistics, for three broad reasons

1. The precise record-keeping needed to render the statistical reasoning valid often (mostly) does not exist;
2. The reasoning rests on impractically precise assumptions, which (see above) must be rigorously fulfilled;
3. Such reasoning is particularly vulnerable to misuse.

All these are true!

And not just in aerospace.

Other People

There are other customers for critical kit besides commercial airlines.

An anecdote. There are lots of such anecdotes.

Within two months of founding the committee, two groups have approached me with an example.

The (Non-Aerospace) Need

- Lots of people have legacy kit which historically is very reliable
- Selling that working kit to new customers often requires satisfying the conditions of IEC 61508
- But when the kit was built, the documentation requirements were not as stringent as now, and the needed documentation does not exist

A Reminder of the Concepts Used in IEC 61508

A Decision

Suppose you are riding in a lift in a mine shaft. The lift safety systems are software-controlled. Two scenarios.

- The engineer says “We have used this SW-based safety kit for fifteen years and it’s never failed”
- The engineer says “We used some SW for fifteen years. It never failed, but we didn’t have the documentation to qualify this software according to IEC 61508:2010.

  So we reimplemented all the SW and this is its first trip!

  I can assure you, THE DOCUMENTATION IS PERFECT! ”

Summary: What Works I

On-demand safety functions, such as a nuclear-reactor SCRAM, which are to be reliable say 999 times out of 1,000 (expected rate of SCRAMS is less than one per year)

Summary: What Does Not Work I

Commercial-aircraft flight control systems for which you need a fleet-operational-lifetime's worth of data before you can satisfy the AMC (Littlewood-Strigini, Butler-Finelli, both 1993)

Summary: What Works II

- Detailed operation logs with all relevant parameters
- With no failures
- And no masked failures!
- Bernoulli- resp. Poisson-process reasoning; some limit theorems

Summary: What Does Not Work II

- Plugging some numbers into some formula, such as in Table D.1 of IEC 61508-7:2010 Annex D
- Mixing up Bernoulli and Poisson processes in mathematical formulas, as in Annex D

Finish

All done

Fertig

Thth thth-at's all, folks!

Thanks

Fini

so much

Danke

Merci

!