System Design and Assessment Notes

Note 24

Reliability-Confidence Algorithms for Assessing Complex Systems

by

Chris Ashley

Air Force Weapons Laboratory

7 June 1978

Abstract

This paper proposes, describes, and discusses four levels of reliability assessments.

21Mar80

Reliability-Confidence Algorithms
for Assessing Complex Systems

INTRODUCTION

Consider the question of whether the threshold of failure of a given system is greater than a specified level of stress (the threat) expected in a certain harsh environment. Suppose it is proposed that the difference between the threshold and the threat is at least M dB (M for margin). It is reasonable to ask what is the probability that this proposition is true. Suppose it is further proposed that this probability is greater than or equal to a certain value R. Then we say the reliability of the first proposition is R. (If M = 0 then R is also the reliability of the system at that stress level.)

In statistical terms, [R,1] is a confidence interval for the binomial (or hypergeometric) parameter p. A further question is, what is the confidence level to be associated with this confidence interval? In the actual assessment of a complex system only incomplete data is ever available from which to estimate the reliability R. Moreover the data which is available is usually contaminated by error from several different sources. Consequently the confidence C(R) warranted by any proposed value of R > 0 must in actual practice always be less than unity. This confidence level is rigorously calculable as a function of the experimental data (and other supplementary information) available to support the estimates M and R. When this confidence is calculated, it is in practice often found to be sufficiently less than unity (typically .9 or less) to require that the confidence level accompany any advertising of the estimates M and R.

For these reasons the algorithms used to calculate C(R) from experimental data are very important in the assessment of systems. For complex systems, RC (reliability-confidence) algorithms have improved considerably in the past few years. However, they still contain theoretical weaknesses. Further work is required to eliminate these weaknesses.

This paper examines RC algorithms currently in use for EMP assessment of complex systems. It discusses the weaknesses of these algorithms, and describes further work needed in this area.

15F78

Reliability-Confidence Algorithms for Assessing Complex Systems

Table 1 lists a sequence of questions, the answers to which represent goals which historically have, one after the other, been accepted as legitimate or minimum acceptable products of an EMP assessment. Successively more difficult questions are to be associated with increasing *levels* of assessment sophistication.

Table 1: Some possible EMP test goals:

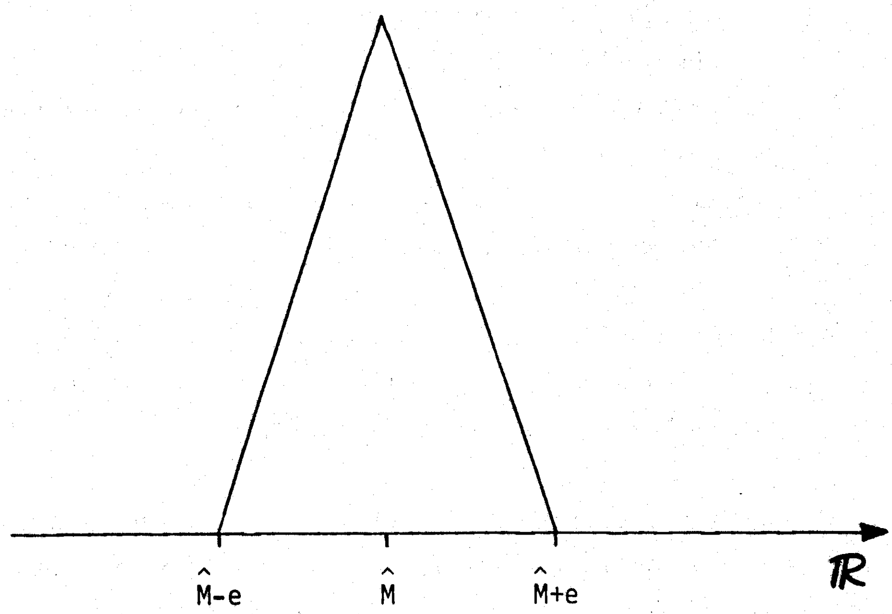
1. "Smoke" test: did the system keep working?
2. *Margin:* *how much* would the field level (or failure threshold) have to be increased (decreased) so as to just barely cause failure? (Worst case orientation, polarization, etc.)
3. If $\hat{M}$ is our best guess of the system margin, what is the *minimum probability* that the truth is at least that good? I.e., what is the *reliability* of that guess?
4. If $\hat{M}$ is proposed as the margin, and R is claimed as the reliability of that margin, what is the *confidence* level to be associated with that claim?

If the first goal seems a little simple-minded today, keep in mind that it represents the outlook used even today to interpret some old Pacific "EMP folklore" data: "The plane kept flying, and didn't seem to be affected in any significant way," or "The protective gaps did arc over." In fact, the first goal is still a primary tool in estimating upset vulnerability: "Upsets were observed on 2% of the pulses, so we estimate that there is vulnerability in 2% of the system state-times."

Obviously the second goal represents an improvement, though, since it gives one a handle on "how much hardening" one needs to put into the system to prevent failure (and possibly where or how to put it in), or how much one has that one doesn't need. Since all one wants is that the system be hard enough to assure it will work, what more could one ask? Why mess around with those third and fourth goals, since they involve statistics and other such "murky", "slimy" ideas?

Well, goal 2, a margin assessment, would be fine if one could find out the margin M exactly, for sure; but, sad to say, the world of experiment usually just isn't that nice. After a long test and a lot of difficult analysis one is more likely to have a "best guess" of M, but have to admit too that a somewhat larger or smaller value may be the real value instead. In fact, the truth might have some distribution about one's guess like that shown in this figure.

Given the estimate $\hat{M}$, what is the true value of the margin M?



This happens to be a symmetric triangular distribution; let's use it as a simple example. So in this case the guessed value $\hat{M}$ of M is the mean of the uncertainty distribution, the expected value; it is the mode, or maximum likelihood estimate, too; it is even the median. What more could a reasonable person ask for? Unfortunately, there is still a difficulty. If we produce a guess of M, as in that second goal, but say nothing about the error e, then even if the symmetric triangular form of the uncertainty distribution is guaranteed (which in real life it never is), we're still dead. This is because no matter what $\hat{M}$ is (say 10 dB, or 30 or 100!), if you'll tell me how close to 50% you want to be then there is an "error" e which will make the probability of failure (or success) of the system lie within that much of 50%. That is, system failure will be an even bet. A little simple algebra, to keep using this triangular distribution as an example, shows that the value of e given in this easy theorem will put the probability of failure within ε of 50%. (Notice that for small ε this value of e is approximately $\frac{\hat{M}}{\varepsilon}$.)

Easy Theorem: For the symmetric triangular distribution, given $\hat{M}$ and $\varepsilon > 0$,

$$|.5-P(M \geq 0)| \leq \varepsilon$$

if and only if

$$e \geq \frac{\hat{M}}{2\varepsilon} (1 + \sqrt{1-2\varepsilon}) \quad .$$

(Since the "theorem" really is easy, we won't clutter up this paper by including the proof.) The following claim is a generalization of the foregoing easy theorem.

Claim: For any continuous distribution (family), given location parameter $\hat{M}$ and $\varepsilon > 0$, there exists a scale parameter value e such that

$$|.5-P(M \geq 0)| \leq \varepsilon \quad .$$

So, no matter what margin $\hat{M}$ is proposed, realistically the EMP assessor also has to find out something about the distribution of uncertainty about that value or he will not even reach the *first* goal in the list. He won't even know whether the system will or won't work in the field, in an EMP environment, let alone with how much margin.

This means we have to try for goal 3, a reliability assessment, at least.

Now, all of this has been obvious for some while. In September 1972 Bob Parker and others at Sandia Labs and AFWL produced a procedure which might be called an RSS&CLT (root sum square and Central Limit Theorem) procedure for estimating and using the uncertainty distribution of the margin. To quote Reference [R1], "What one desires to find is the probability that a deployed system will survive under the defined conditions. One can state this probability as: $Pr(W_0 \geq W_c)$ where W_0 has a probability density function (PDF) depending on the system [and measurement] variables, and W_c has a PDF depending on the variables associated with the system threshold." [Of course, their probability is equal to $Pr(\frac{W_0}{W_c} \geq 1)$, which is equal to $Pr(\log \frac{W_0}{W_c} \geq 0)$.]

In September 1975 Vince Jones and Dr Higgins, of Boeing, put out a DNA document entitled "SV Safety Margin Assessment" (Reference [R2]). In their report they say, "The statistical nature of the problem requires probabilistic statements of survivability in place of the binary determination 'sure-safe' or 'sure-kill'." That is, they advocate jumping from assessment level 1 directly to assessment level 3. The means which they offer for doing this is a quantity which they call the Data Quality. (In general the Data Quality increases as the quality of the data decreases, and vice versa. This seems to me a peculiar way to use words, so I usually refer to their variable by the more descriptive term "data unquality".)

Now, there is one rather strong assumption in the Jones-Higgins approach. This assumption appears in their statement, "With one exception, the data quality [or data unquality] is defined ... [to] = 3σ " (p. 8 of Reference [R2]). Confirming this assumption is not easy. Essentially it is precisely the problem of estimating a parameter (in this case disguised as three times the parameter) of a distribution. Of course, it is also more or less necessary to know what the distribution is. The latter problem was solved three years earlier by Parker et al. when they suggested depending upon the Central Limit Theorem to assure the final margin error distribution was normal. But neither the Parker group nor Jones-Higgins faced up to the penalties in the estimation of the scale parameter of that distribution.

Now, the fact is that the error distribution can only be *estimated*, based on a finite amount of experimental data. Assumptions about the values of those parameters are *necessarily*, and *unavoidably*, and (almost) *always*, in error. And the smaller the amount of experimental data upon which those parameter estimates are based, the greater the error is likely to be in assuming the estimates are correct. There have been multi-year, multi-million-dollar EMP assessment-test programs which, when the dust settled, had to estimate error parameters using fewer than a dozen experimental measurements. -- Now, assessment goal 4 is simply to take account of the inaccuracy in parameter estimation. Or, to put it another way, a level 4 assessment is one which makes public the degree of uncertainty in the estimation of the parameters of the reliability distribution, and does this quantitatively, and conscientiously.

One might ask ... in fact, sometimes one *does* ask ... is a level 4 assessment possible? Is goal 4 attainable? And, if so, how?

Well, the answer to the first part is "yes". In fact a level 4 assessment of a real, complex system has been completed in the past year. In April 1978 (this year) there was finally published a set of volumes collectively entitled "EC-135 EMP Assessment Program Final Report", by Jim Locasso and others of Autonetics Division of Rockwell International. One of these volumes is subtitled "Assessment Error Analysis". This volume is also known as AFWL TR 77-254. (This volume, Reference [R3], is unclassified.) In this volume there appear such charts as this.

TABLE 2

SUMMARY OF ERROR ANALYSES RESULTS FOR METHOD NO 2

Section	Error Type	e_w (90-90)	Comments
III.3	Analytic Wire Currents	±25 dB	Log-normal distribution by Central Limit Theorem. (Includes Impedance Ratio error).
III.3	Impedance Ratio Calculations	±13.5 dB	(1) Log-normal distribution. Acceptable per χ^2 goodness-of-fit test. (2) Impedance ratio method used to obtain wire current from bulk current or from other wire current.
III.5 and III.6	Measurement Error	±5.3 dB	(1) Includes simulator variability, calibration, and machine processing. (2) Machine processing may be log normal; others are not.
III.4	Threshold	±13.3 dB	(1) Log-normal distribution. Acceptable per χ^2 goodness-of-fit test. (2) 12 dB conservative bias with respect to mean.
III.8	Power On - Off	±9.7 dB	
III.1	Simulation Error	±12.4 dB	For orientation 1
		±6.7 dB	For orientation 2
		±10.6 dB	For orientation 3
		±11.5 dB	For orientation 4

Of course, this chart of experimental values provides a check for the reasonableness of future experimental determinations, and also of some of the gedanken values offered in the Parker volume. But much more importantly, notice the column heading "90-90". The first 90 asserts that at least 90% of the particular error is captured by the stated interval, and the second 90 says, essentially, that the size of the sample upon which this assertion is based is such that the confidence level to be associated with the capture assertion is only 90%.

For each major mission critical subsystem the appropriate numbers were then taken from this table and RSSed according to the prescription of Parker et al.

At this point let us digress for a moment, to make two points about language. The first is that the result of this process is a *reliability*, not a probability and not a credibility. Reliability is a *lower bound* on the true but unknown probability of success, and therefore the two words do *not* mean the same thing. Similarly, we are using the word "confidence" in a strict statistical sense; so that word is *not* "visceral probability" or "warm feeling" or anything like these. To keep the confusion down, I suggest we use the word "credibility" to denote "visceral" things, such as "warm feelings about the assessment" and so on. One increases credibility by being smart, conscientious, and thorough, for example by doing a level 4 assessment rather than, say, just level 3. One changes confidence level, in contrast, by changing sample size. Either one can be very large while the other is negligible. So let's not try to make the terms synonymous.

Table 3.

Reliability $\triangleq$ *lower bound* on probability of success
Reliability $\neq$ Confidence
Reliability $\neq$ credibility
Confidence $\neq$ credibility

In particular, the phrase "high confidence assessment" is usually meaningless. We should say, "a high credibility assessment", or "a highly credible assessment".

Second, the probabilities which these reliabilities lower bound are *conditional* probabilities. (So the reliabilities might be called conditional reliabilities.) The conditions include the physical assumptions of the assessment. For example, it might be assumed that a current overload on any single pin of a box is sufficient to put that box out of commission (at least until reset or repair is done). Similarly, all boxes in a subsystem might be assumed to be necessary to proper operation of that subsystem. For a third example, it might be *assumed* that time domain peak current or power is the controlling failure parameter. This is not to say that these things *should* be assumed. Rather we are saying that *if* one or more of them *is* assumed then the resulting reliabilities are *conditional* reliabilities, and the confidence levels are on those *conditional* reliabilities. I suggest it is important to keep a list of the conditions actually used in an assessment highly visible in the final report of the results of that assessment.

Having made these points, we can return to the main thread of the discussion and begin to consider some of the difficulties peculiar to the level 4 assessment today.

First, it might occur to one to wonder, when one RSS's these 90% confidence conditional reliabilities, what is the confidence level to be associated with the resulting RSS? Well, about a year ago I would have said that RSSing N reliabilities with confidence level $C < 1$ each would depress the confidence in the result to C^N. But! about that time Jim Locasso, of Autonetics, produced this conjecture.

Conjecture: Let $x_i : N(0,\sigma_i^2)$, where $i \in \{1,\ldots,N\}$.

Let $[-e_i, e_i]$ be an R-reliable interval for x_i, at confidence level $\tilde{C}$.

Then the RSS of these intervals will produce an R-reliable interval for $X = \sum^N x_i$ at confidence level $C \geq \tilde{C}$.

In fact, Locasso has almost made a theorem of it; that is, I think he has "almost proved" it. He has also generated some numerical evidence for it as well. Moreover, although he considered only the normal case, I suspect his demonstrations can be generalized to the unimodal case. Be that as it may, Locasso's work on this non-decreasing-confidence conjecture is contained in the April 1978 EC-135 EMP Assessment Error Analysis volume, Reference [R3], which was cited above. It has also been published separately; cf. Reference [R4]. In that level 4 assessment it was assumed that the proof could be completed; that assessment took the confidence in the RSS to be not less than the 90% which had been demonstrated for each of the summands.

I submit that the EMP community badly needs to complete the proof of that conjecture.

By the way, Locasso also points out (p. 12, Reference [R4]) that there exist distributions for which his conjecture conclusion would be false, and for which C^N is indeed the right answer ... notably some multimodal distributions. So, though we need this proof, we will also need to be careful even after we have it; it will not guarantee starting at a high confidence level will *always* ensure that we finish at a high confidence level when RSSing errors.

Another problem in level 4 assessments is mixing different kinds of confidences, for example Neyman-Pearson and Bayesian.

Another is how to estimate variance using a nonparametric confidence interval, when the distribution is wholly unknown.

Table 4: Some problems in Level 4 (Confidence) Assessments:

1. Complete the proof of the non-decreasing-confidence conjecture for normal distributions (at least).
2. How should a mixture of classical and nonparametric confidence intervals be handled?
3. What is the best way to estimate variance from a nonparametric confidence interval on the original data?
4. Given that L of N elements of an (unstructured) population have been assessed at $C_I(R_I)$, where $0 < R_I \Rightarrow C_I < 1 \;\forall\; I \in \{1,...,L\}$, how does one calculate C(R) for the population?

(Incidentally, the EC-135 level 4 assessment cited above was conscious of these problems, and incorporated approprate and reasonable safeguards to protect the conclusions of the assessment from each. Cf. Reference [R5].)

In conclusion, it should be a reflex action, whenever one hears a reliability (e.g., EMP reliability) estimate for a system, to ask, "At what confidence level?" This is because any estimate of reliability has associated with it some confidence level (whether or not that level has been calculated), which depends on the *amount* (as well as the kind) of data supporting that estimate. And in "real world" assessments that confidence level, for any reliability estimate greater than zero, is always *less* (often a great deal less) than 1. Or, to put it another way, the reliability estimate for *any* system, given a set of data on that system, can correctly be said to be .999 ... at *some* confidence level. This insistence that any level 3 assessment be upgraded to level 4 should be maintained despite the fact that no one claims a level 4 assessment is easy (as shown by Table 4, above).

References

[R1] *EMP Handbook for Missiles and Aircraft in Flight*, prepared by Sandia Laboratories for the Air Force Weapons Laboratory, edited by D.E. Merewether, J.A. Cooper, and R.L. Parker. Air Force Weapons Laboratory TR 73-68; September 1972. (This Reference sometimes bears the title *EMP Interaction 1-1.)*

[R2] *Survivability/Vulnerability Safety Margin Assessment*, by V.K. Jones and Dr T.P. Higgins. DNA 3859Z; 30 September 1975.

[R3] *EC-135 EMP Assessment Program Assessment Error Analysis*, by J.V. Locasso, C.F. Juster, W.H. Cordova, J.F. Wagner, and B.A. Daiken. Air Force Weapons Laboratory TR 77-254; April 1978.

[R4] *The Confidence in Combinations of Imperfectly Known Variances*, by James V. Locasso. EMP Probability and Statistics Note 9; 8 July 1977.

[R5] *A Brief Presentation and Discussion of the Algorithm Used to Determine EC-135 EMP Margin Reliability-Confidence*, by Chris Ashley and James V. Locasso. EMP System Design and Assessment Note 23; 27 September 1977.