

**Review of X. Xie, T.Y. Chen, F.-C. Kuo, B. Xu,  
“A theoretical analysis of the risk evaluation formulas for  
spectrum-based fault localization”, *ACM Transactions on  
Software Engineering and Methodology* 22, 4 (2013) 1–40**

T.H. Tse

Department of Computer Science  
The University of Hong Kong  
Pokfulam, Hong Kong

Spectrum-based fault localization is a popular technique in automatic program debugging. Researchers analyze the distribution of pass and fail cases in program testing using different risk evaluation formulas, and validate how their proposals are better than earlier work via empirical studies. In this paper, the authors propose a theoretical framework to compare 30 risk evaluation formulas in terms of the percentage of code examined before a fault is identified. They rank the formulas using “better” and “equivalent” relations. Only five formulas are proven to be the most efficient. Many of the best-known formulas are not among them.

There is an unhealthy tendency toward empirical studies in software testing and debugging research. Researchers use hypothesis testing to determine whether their proposal is better than that of their predecessors. Reviewers demand more subject programs and larger test pools for further validation. It is refreshing to see that the authors of this paper do not simply rely on empirical studies, but prove mathematically whether various proposals have hit their mark.

This paper is not the only example of the successful application of mathematical theory by Chen’s research group. Chen and Merkel prove in one paper [1] that no test case generation technique can be better than random testing by more than 50 percent. Hence, their proposed adaptive random testing technique is close to this theoretic limit. In another paper [2], Chen and Yu prove that their proposed proportional sampling strategy is the only partition testing strategy that ensures that the probability of finding at least one failure is no lower than random testing for any program. Understandably, some researchers are disgruntled because these theoretical results stop them from making further incremental proposals.

## References

- [1] T.Y. Chen, R. Merkel, “An upper bound on software testing effectiveness”, *ACM Transactions on Software Engineering and Methodology* 17, 3 (2008), 16:1–16:27.
- [2] T.Y. Chen, Y.T. Yu, “The universal safeness of test allocation strategies for partition testing”, *Information Sciences* 129, 1–4 (2000), 105–118.