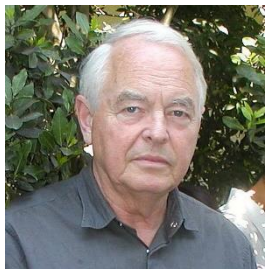
Submission to NSF 2026 Idea Machine Competition
https://www.nsf.gov/news/special_reports/nsf2026ideamachine/enter.jsp

How to Upgrade Safety Requirements to Prevent Technological Catastrophes?

Team Members: The idea was submitted on behalf of Co-PIs of the AFOSR project “Massively Parallel Approaches for Buffered Probability Optimization and Application” under award number FA9550-18-1-0391.

Prof. Stan Uryasev, University of Florida



Prof. Tyrrell Rockafellar, University of Washington and University of Florida

Prof. Michael Zabarankin, Stevens Institute of Technology

Dr. Drew Kouri, Sandia National Laboratories

Prof. Johannes O. Royset, Naval Postgraduate School

1. What is the compelling question or challenge?

How to upgrade safety requirements in various engineering areas to prevent major technological catastrophes? How to implement conservative control of both chances and magnitude of very large losses?

2. What do we know now about this Big Idea and what are the key research questions we need to address?

Why don't existing safety requirements (regulations) prevent major technological catastrophes? A simple, and to some extent surprising answer, is that the current safety requirements *are not designed* to prevent major catastrophes.

The problem is that safety requirements in a majority of engineering fields are optimistically designed and do not control the magnitude of very large outcomes. Some thresholds (lower bounds) are specified that should be exceeded with low probability. For instance, in nuclear safety, a level of release of radiation in the environment is specified, which may be exceeded, say, only once in 10,000 years. Nevertheless, the *magnitude of exceedance is not controlled*, even for extremely large outcomes. Safety regulations do not distinguish the magnitude of outcomes exceeding a high threshold corresponding to a major accident. For instance, safety requirements were not violated in the Fukushima accident. We can verify calculations of safety models for the Fukushima nuclear power plant and come to a conclusion that the plant worked as designed (no violation because the probability of radiation release is low, in spite of a very large magnitude of the release, which is not controlled). Similar concerns arise in defining financial ratings of companies, certification of materials (A/B basis), designing of dams, etc.

We suggest to investigate how to upgrade risk constraints in various engineering areas to take into account the magnitude of outcomes, in addition to chances, to make "illegal" major catastrophes. This is a huge diverse effort requiring: 1) statistical and mathematical analysis (theory of risk measures); 2) computer modelling to demonstrate that the upgraded requirements/regulations prevent past catastrophes; 3) collection of appropriate data; 4) setting risk regulations (legal issues); etc.

There are two equivalent variants of the "optimistic lower-bound risk management approach":

1) Fix a threshold, which is the lower bound of outcomes in tail of the distribution, and constrain the Probability of Exceedance (POE), also known as the Survival, Survivor, or Reliability function;

2) Fix the probability of the tail, and constrain the lower bound of the tail outcomes, called the quantile (or Value-at-Risk (VaR) in finance).

The majority of risk constraints for controlling low probability events in various engineering areas are currently set with POE.

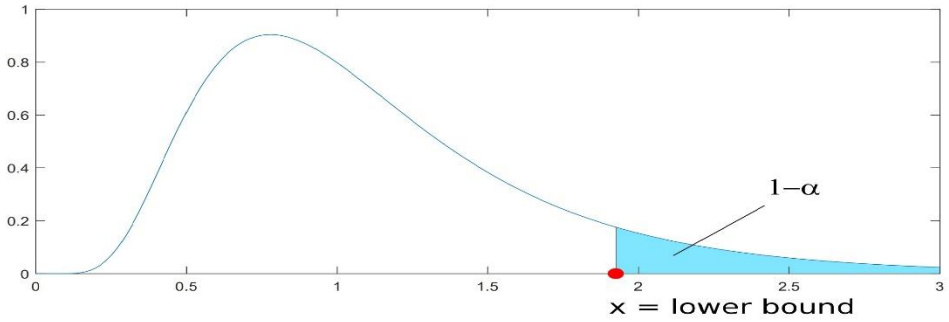
It has been recognized in finance that the second approach based on VaR needs to be improved. VaR has been supplemented with the Conditional Value-at-Risk (CVaR), also called Expected Shortfall, Average VaR, Tail VaR, and Superquantile. By definition, CVaR is the average value of outcomes in the tail with some specified probability. For instance, if the probability of the tail is 10%, than this is the average of the worst-case 10% outcomes. So, by construction, CVaR takes into account both probability

and magnitude of the tail. CVaR is a so-called "coherent risk measure" having exceptional mathematical properties (a Google search for "coherent risk measure" shows more than 22.5 million hits).

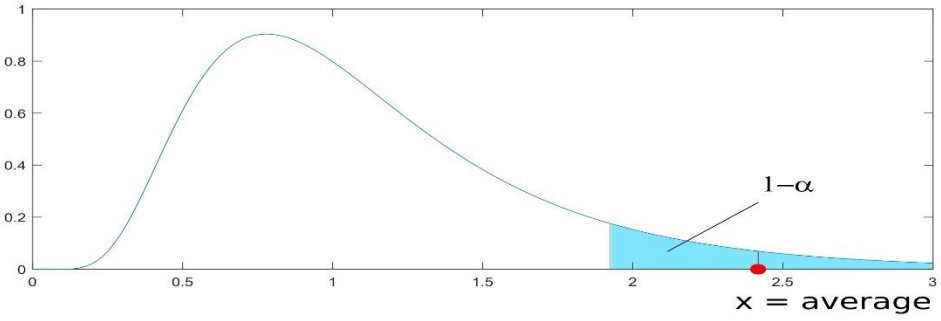
It was recently discovered that there is a conservative analog of POE, called Buffered POE (bPOE), which takes into account both the chances and magnitudes of losses. By definition, bPOE is a probability of the distribution tail with the known average value of this tail. So, bPOE is the inverse function of CVaR. The safety constraints with CVaR and bPOE are equivalent. Therefore, POE can be upgraded to bPOE, similar to how VaR was upgraded to CVaR. Recently, a general formula for bPOE was derived, and it was shown that, similar to CVaR, bPOE risk management/optimization can be done with convex and linear programming.

The bPOE risk management approach is one of the outcomes of the DARPA EQUiPS grant SNL 014150709. For more details, download the presentation "Risk Management with POE, VaR, CVaR, and bPOE" at http://www.ise.ufl.edu/uryasev/files/2018/06/risk_management_CVaR_bPOE8.pdf

- **Optimistic concept** based on **lower bound** of outcomes in the tail



- **Conservative concept** based on **average value** of the tail



3. Why does it matter?

Let us list some recent technological catastrophes: 1) Fukushima nuclear accident in Japan; 2) Financial crisis in 2008; 3) Space shuttle Columbia accident; 4) Hurricane Katrina flood. These catastrophes resulted in deaths of people and astronomical financial losses. However, in all of these cases safety requirements were not violated (at least from a legal point of view) and nobody was punished for the faulty design of engineering systems.

We anticipate that a conservative upgrade of safety requirements, which takes into account both magnitude as well as chances of losses, will benefit virtually all engineering fields.

For instance, nuclear safety regulations require extremely low "core melt frequency". However, they do not distinguish between a large and an extremely large radiation release (which may be poisoning the Earth). We think that this inconsistency can be fixed by using the bPOE risk function.

Another major example is the 2008 Financial Crisis. At least partially, it can be attributed to defaults of financial companies overloaded with exposure to derivative instruments. Derivatives may lead to very high losses with very low probability. This was the case for the AIG giant insurance company, which overloaded the trading book with Collateralized Debt Obligation (CDO) upper tranches. The chance of default was low and AIG kept an AAA rating in spite of an extremely high exposure. The U.S. Government had to bailout AIG for $85 Billion. We suggest to develop so called Buffered Ratings, which will take into account the exposure as well as chances of default. Imagine that we had done this upgrade of ratings before the 2008 Financial Crisis and therefore reduced the magnitude of losses. The 2008 Crisis cost tops $22 trillion according to some estimates.

One more example is a non-conservative Material Certification process (A and B basis). B basis is a 95% lower confidence bound on the tenth percentile and the A basis is a 95% lower confidence bound of the first percentile. Let us consider that 1,000 coupons of a ceramic material were tested and all coupons passed the test (i.e., the material was certified with A and B basis). Now, suppose that there was a "sabotage" and 9 coupons were replaced by defective coupons with zero strength. The chance of getting a defective coupon is 0.9%. The A-Basis and B-Basis requirements would still be valid because they do not take into account the magnitude of outcomes below the threshold. Engineers design good materials in spite of non-conservative safety requirements. However, requirements need to be upgraded to be more conservative. The ceramic plates covering the Columbia Shuttle passed A-Basis and B-Basis strength requirements.

One final example is the 2005 Levee Failure in New Orleans following the passage of Hurricane Katrina. Levee designers have not taken into account the magnitude of possible losses of the failure. Safety margins should take into account the magnitude of possible losses.

4. If we invest in this area, what would success look like?

The ultimate success is an upgrade of currently "overly optimistic" safety requirements across all engineering fields and supplementing them with more conservative measures. The idea is that extremely large loss (such as a financial loss at an insurance company or radiation release at a nuclear power plant) would trigger the violation of requirements and introduce a "negative feedback" in the control loop. In engineering practice, a Bayesian statistical approach is frequently used to update values of tracked characteristics. It is needed to setup characteristics sensitive to very large outcomes.

The task of upgrading safety requirements across all engineering fields is too challenging, complicated, and expensive to be address in full by the NSF. However, the NSF can play a leading role in identifying that it is an important problem and attracting the attention of the engineering community to these issues.

NSF can play a leading role in the development of risk management methodology to upgrade safety requirements. At this time there is NO clear theoretical basis showing the shortcomings of the POE safety criteria and the significant role that this criteria played in major catastrophes. Safety constraints for POE are equivalent to safety constraints for VaR. It has been demonstrated that such safety constraints have shortcomings and that they should be upgraded with CVaR (coherent measures of risk). However, for POE a similar statement has not yet been investigated even from a theoretical point of view.

The suggested upgrade could be applied to past "historical" major catastrophes/losses to check if the tighter requirements would be violated. Case studies should be conducted by teams of researchers in various areas to confirm the importance of the upgrades.

5. Why is this the right time to invest in this area?

The significant number of technological catastrophes in recent years demonstrated a pressing need for an upgrade of safety requirements in many engineering areas. Recent advances allow us to address these issues:

1) Internet provided access to various sources of information, such as databases and datasets describing events of interest.

2) High speed parallel computational capabilities are in place for modeling and evaluating millions of low probability scenarios.

3) Significant improvements in the theory of risk that include (a) the development of axiomatic theory for risk measures, (b) the definition of the new bPOE risk measure, and (c) the development of statistical methods, such as linear regression for CVaR estimation.

4) Numerical methods and software for efficient risk management (team submitted the topic is conducting AFOSR grant FA9550-18-1-0391 on massively parallel algorithms for bPOE modeling and optimization).

5) Funding agencies value challenging interdisciplinary projects.

6. Three key words describing big idea

Risk, safety, catastrophe

7. Three publication references

1. Rockafellar, R.T. and S. Uryasev. Conditional Value-at-Risk for General Loss Distributions. Journal of Banking and Finance, 26/7, 2002, 1443-1471 http://www.ise.ufl.edu/uryasev/files/2011/11/cvar2_jbf.pdf

2. Rockafellar, R.T. and J.O. Royset. On Buffered Failure Probability in Design and Optimization of Structures. Reliability Engineering & System Safety, 95(5), 2010, 499-510 https://sites.math.washington.edu/~rtr/papers/rtr211-BufferedProb.pdf

3. Mafusalov, A. and S. Uryasev. Buffered Probability of Exceedance: Mathematical Properties and Optimization. SIAM Journal on Optimization 28(2), 2018, 1077-1103 http://www.ise.ufl.edu/uryasev/files/2018/05/Buffered_probability_of_exceedance.pdf