TOWARDS EFFICIENT WAYS OF ESTIMATING FAILURE PROBABILITY OF MECHANICAL STRUCTURES UNDER INTERVAL UNCERTAINTY

MICHAEL BEER[1], MARCO DE ANGELIS[1], and
VLADIK KREINOVICH[2]

[1]Inst. for Risk and Uncertainty, School of Engineering, Univ. of Liverpool, UK.
E-mail: mbeer@liverpool.ac.uk, Marco.De-Angelis@liverpool.ac.uk
[2]Department of Computer Science, University of Texas at El Paso, USA.
E-mail: vladik@utep.edu

The textbook approach to estimating the failure probability of mechanical structures assumes that (1) we know the probability distribution on the set of all possible values of the quantities $\theta = (\theta_1, \ldots, \theta_n)$ describing the structure and its environment; this distribution is usually described by the probability density function $f(\theta)$; and (2) we know which combinations of the quantities θ_i correspond to stability and which to failure; the corresponding set Ω_F is usually described by a *limit* function $g(\omega)$ such that stable states correspond to $g(\theta) > 0$ while failures correspond to $g(\theta) < 0$. Once we know this information, we can find the desired failure probability P as the integral $P = \int_{\Omega_F} f(\theta)\,d\theta$. In realistic situations, when the number of quantities is large, P can be computed by a (somewhat time-consuming) Monte-Carlo algorithm (MCA).

In practice, we often do not know the exact probability distribution, and we do not know the exact expression for the limit function. Usually, we know that the distribution belongs to a certain family (e.g., that it is normal), but we do not know the exact values of the parameters corresponding to the given distribution; at best, we know the *intervals* containing these parameters. Similarly, we know the general parametric expression for the limit function (e.g., we know that the function $g(\theta)$ is linear or quadratic), but we do not know the exact values of the corresponding parameters, we only know the intervals of possible values of these parameters. Let us list all the parameters corresponding to the probability distribution and to the limit function as $p_1, \ldots, p_m$. For each of these parameters p_i, let us denote the midpoint of its interval of possible values by $\widetilde{p}_i$, and its half-width by Δ_i; then, this interval takes the form $[\widetilde{p}_i - \Delta_i, \widetilde{p}_i - \Delta_i]$. Each value p_i from this interval has the form $\widetilde{p}_i + \Delta p_i$, where $|\Delta p_i| \le \Delta_i$.

Let $P(p_1, \ldots, p_m)$ denote the failure probability corresponding to parameters p_i. The values Δ_i are usually reasonable small, so we can expand $P(p_1, \ldots, p_m) = P(\widetilde{p}_1 + \Delta p_1, \ldots, \widetilde{p}_m + \Delta p_m)$ into Taylor series and keep only linear terms in this expansion: $P(\widetilde{p}_1 = \Delta p_1, \ldots) = \widetilde{P} + \sum\limits_{i=1}^{m} c_i \cdot \Delta p_i$, where $\widetilde{P} \stackrel{\text{def}}{=} P(\widetilde{p}_1, \ldots, \widetilde{p}_m)$ and $c_i \stackrel{\text{def}}{=} \dfrac{\partial P}{\partial p_i}$. We are interested in the range $[\underline{P}, \overline{P}]$ of this expression when $|\Delta p_i| \le \Delta_i$. One can easily check that $\underline{P} = \widetilde{P} - \Delta$ and $\overline{P} = \widetilde{P} + \Delta$, where $\Delta = \sum\limits_{i=1}^{m} |c_i| \cdot \Delta_i$; see, e.g.,[2,4].

In the ideal situation, when we know the exact values of $P(p_1, \ldots, p_n)$ for different p_i, then we could estimate the derivatives c_i and, thus, the value Δ by using numerical differentiation: $c_i \approx \dfrac{P_i - \widetilde{P}}{\Delta_i}$, where $P_i \stackrel{\text{def}}{=} P(\widetilde{p}_1, \ldots, \widetilde{p}_{i-1}, \widetilde{p}_i + \Delta_i, \widetilde{p}_{i+1}, \ldots, \widetilde{p}_m)$. As a result, we get[2]

Second International Conference on Vulnerability and Risk Analysis and Management (ICVRAM2014).
Edited by Michael Beer, Ivan S.K. Au & Jim W. Hall

ISBN: 981-973-0000-00-0 :: doi: 10.3850/981-973-0000-00-0_abstr13-12a

$\overline{P} = \widetilde{P} + \sum_{i=1}^{m} |P_i - \widetilde{P}|$; this estimate requires $m+1$ calls to MCA: to compute $\widetilde{P}$ and m values P_i. When the number of parameters m is large, it is computationally more efficient to use a special estimation technique based on the use of Cauchy distribution [2,3].

In practice, the corresponding finite-parametric families provide only an approximate description of the actual probability distribution and limit function. As a result of this model uncertainty, even if we know the exact values of the corresponding parameters, the computed failure probability C may be different from its actual value P. In some cases, the only information that we have about the model uncertainty $C - P$ is the upper bound δ, for which $|C - P| \le \delta$. Thus, when after (somewhat time-consuming) Monte-Carlo simulations, we get the value C, the only conclusion that we can make about the actual failure probability P is that $P \in [C - \delta, C + \delta]$. As a result, when we apply the above numerical computation techniques and get an estimate $\overline{C} = \widetilde{C} + \sum_{i=1}^{m} |C_i - \widetilde{C}|$, then, due to the differences $|P_i - C_i| \le \delta$ and $|\widetilde{P} - \widetilde{C}| \le \delta$, we can only conclude that $|\overline{C} - \overline{P}| \le (2m+1) \cdot \delta$. When the number m of parameters is large, this error becomes significant. How can we decrease this error?

In principle, we can use the fact that the linear function $\widetilde{P} + \sum c_i \cdot \Delta p_i$ always attains its maximum on the box $\prod[-\Delta_i, \Delta_i]$ at one of the vertices; so, we can compute all 2^m values $C(\varepsilon) = C(\widetilde{p}_1 + \varepsilon_1 \cdot \Delta_1, \ldots)$ corresponding to all combinations of $\varepsilon_i \in \{-1, 1\}$, and find the largest of these values; this largest value is δ-close to $\overline{P}$. This method, however, requires 2^m calls to MCA, and even when m is reasonable (e.g., 10), 2^m calls is too much. It is known that if we want to find $\widetilde{P}$ with accuracy δ, we cannot use fewer than exponentially many calls[1]. How can we decrease the uncertainty in $\overline{P}$ without increasing the number of calls too much?

In this paper, we propose several such methods. For example, the use of an estimate $\overline{C} = \widetilde{C} + \frac{1}{2} \cdot \sum_{i=1}^{m} |C_i - C_{-i}|$, where $C_{-i} \stackrel{\text{def}}{=} C(\widetilde{p}_1, \ldots, \widetilde{p}_{i-1}, \widetilde{p}_i - \Delta_i, \widetilde{p}_{i+1}, \ldots, \widetilde{p}_m)$ leads to accuracy $(m+1)\delta$ with $2m+1$ calls to MCA. We can decrease the uncertainty even more if we mark all the indices i for which $C_i - C_{-i} \ge 2\delta$ or $C_i - C_{-i} \le -2\delta$; for these indices, we can fix $p_i = \widetilde{p}_i + \Delta_i$ or, correspondingly, $p_i = \widetilde{p}_i - \Delta_i$, and only change other values p_j. If we have s such indices, we use $2m + 1 + 2(m - s)$ calls to MCA and get accuracy $(m - s + 1)\delta$. We can also divide all the remaining indices into pairs; for each pair, we try all 4 combinations of $\pm\Delta_i$; this will lead to $2m+1+4(m-s)$ calls and accuracy $\delta + \frac{1}{2}(m-s)\delta$. If instead of pairs, we divide into groups of k, we get $2m + 1 + 2^k(m - s)$ calls and accuracy $\delta + \frac{1}{k}(m - s)\delta$.

We also show how to modify the Cauchy deviate technique from[2,3] to take into account the upper bound δ on the model accuracy.

Keywords: failure probability, interval uncertainty, model uncertainty, Monte-Carlo methods, Cauchy deviate technique.

References

1. V. Kreinovich, *Neural, Parallel, and Scientific Computations* **2**(2), 225–234 (1994).
2. V. Kreinovich, In: F. Pavese and A. B. Forbes (eds.), *Data Modeling for Metrology and Testing in Measurement Science*, (Birkhauser-Springer, Boston, 2009).
3. V. Kreinovich and S. Ferson, *Reliability Engineering and Systems Safety*, **85**(1–3), 267–279 (2004)
4. S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice* (American Institute of Physics, New York, 2005).