High-Impact Low-Probability Events Are Even More Important Than It Is Usually Assumed

Aaron Velasco[1], Olga Kosheleva[2], and Vladik Kreinovich[3]
Departments of [1]Earth, Environmental, and Resource Sciences,
[2]Teacher Education, [3]Computer Science
University of Texas at El Paso, El Paso, Texas 79968, USA
aavelasco@utep.edu, olgak@utep.edu, vladik@utep.edu

1. High-probability vs. low-probability events

- Many undesirable events happen all the time.
- It is important to prepare for these events, to mitigate the future damage as much as possible.
- Some of these events have a reasonable high probability; for example:
  - earthquakes regularly happen in California,
  - hurricanes regularly happen in Florida and other states,
  - tornadoes regularly happen in North Texas,
  - floods regularly happen around big rivers, etc.
- Such events happen all the time with varying strength.
- Everyone understands that we need to be prepared for situations when the strength will becomes high, causing disastrous consequences.

2. High-probability vs. low-probability events (cont-d)

- Undesirable events are not confined to zones where such events have a reasonably high frequency.
- For example:
  - while the majority of major earthquakes occur in seismic zones, where earthquakes are a common occurrence,
  - some major earthquakes occur in areas where strong earthquakes are very rare, where the last serious earthquake may have occurred thousands of years ago.
- The only reason we know about past events is indirectly, by the effect of this strong past earthquake on the geology of the region.

3. Low-probability events are important

- The problem with high-impact low-probability events is that:
  - in contrast to events in high-probability zones where most people are prepared,
  - people in low-probability events are mostly unprepared.
- For example, in California, since medium-size earthquakes happen there all the time:
  - building codes require that buildings be resistant to (at least medium-strength) earthquakes, and
  - within each building, most shelves are attached to the walls, so that they do not cause extra damage when the earthquake hits.

4. Low-probability events are important (cont-d)

- In contrast, in low-probability zones, none of these measures are implemented; as a result:
  - when the undesirable event happens – with even medium strength,
  - it causes much more damage than a similar-strength event in high-probability zones.

5. How should we allocate resources: current approach

- The fact that we need to take into account high-impact low-probability events is well understood.
- Of course, it is not realistically possible to take into account all possible low-probability events.
- So we need to allocate resources to the most important events.
- Traditional way to decide on the importance of an event is to multiply its probability by the damage it may cause.
- This idea is in perfect accordance with the decision theory.

6. Remaining problem

- The problem of the current approach to resource allocation is how to estimate the corresponding probability.
- For high-probability events, events that occur reasonably frequently, we can estimate it as the frequency of observed events.
- For example, if a major flood happens, on average, every 10 years, we estimate the yearly probability of this flood as $1/10$.
- For high-probability events, this estimate is based on a large number of observations and is, therefore, reasonably accurate.
- In principle, we can apply the same approach to low probability events – and this is exactly how such events are analyzed now.
- However, when events are rare, the sample of such events is very small.

7. Remaining problem (cont-d)

- It is known that for small samples, the difference between observed frequency and actual probability can be large.
- So, the natural questions are:
  - How can we take this difference into account? and
  - If we do, what will be the consequences?

8. What we do in this paper

- We show how the above difference can be taken into account.
- We also show that:
  - if we take this difference into account,
  - then high-impact low-frequency events become even more important than it is usually assumed.

9. Current way of allocating resources: justification

- There is a whole science of rational decision making, known as *decision theory*.
- According to decision theory, preferences of a rational person can be described by assigning:
  - to each possible alternative x,
  - a real number $u(x)$ called *utility*.
- Then, the decision maker prefers alternative x to alternative y if and only if the alternative x has higher utility: $u(x) > u(y)$.
- In situations when we have different outcomes with different utilities u_i and different probability p_i:
  - the equivalent utility u of the corresponding situation
  - is equal to the expected value of utility: $u = p_1 \cdot u_1 + p_2 \cdot u_2 + \ldots$

10. Current way of allocating resources (cont-d)

- In particular:
  - if we consider a disaster with potential damage d (and thus, utility $-d$) and probability p,
  - then the utility of not taking this potential disaster into account is equal to $p \cdot (-d) = -p \cdot d$.
- According to the above-mentioned notion of utility, this means that we need to primarily allocate resources:
  - to situations in which this negative utility is the worst,
  - i.e., in which the product $p \cdot d$ is the largest.

11. Description

- The above analysis leads to the way resources are allocated now.
- For each possible zone, we compute the product $p \cdot d$ of the probability p of the undesirable event and the damage d that would be caused by this event.
- The larger this product, the higher the priority of this zone.
- This way, many low-probability zones get funding:
  - in these zones, the probability p is lower than in the high-probability zones,
  - but, as we explained, the potential damage can be much higher,
  - since such zones are usually unprepared for the undesirable event (or at least much less prepared).

12. How the corresponding probabilities are estimated

- To use the usual techniques, we need to estimate, for each zone, the probability p of the undesirable event.
- In statistics, the usual way to estimate probability is take the frequency with which this even happened in the past.
- If in 200 years of record, the major Spring flood occurred 20 times, we estimate the probability of the flood as $20/200 = 0.1 = 10\%$.
- If in some other area, a similar flood happened only twice during the 200 years, we estimate the probability of flooding in this areas as
$$2/200 = 0.01 = 1\%.$$

13. How the probabilities are estimated (cont-d)

- As we have mentioned:
  - the main limitation of this approach is that
  - it does not take into account that the frequency is only as approximate value of the probability.
- It is therefore desirable to come up with a more adequate technique, a technique that would take this difference into account.

14. What do we know about the difference between probability and frequency

- According to statistics:
  - if we estimate the probability based on n observations, then, for large n,
  - the difference between the frequency f and probability p is normally distributed with mean $\mu = 0$ and standard deviation

$$\sigma = \sqrt{\frac{p \cdot (1-p)}{n}}.$$

- We do not know the exact probability p, we only know its approximate value f.
- By using this approximate value instead of p, we can estimate the above standard deviation as $\sigma \approx \sqrt{\frac{f \cdot (1-f)}{n}}$.

15. What do we know about the difference between probability and frequency (cont-d)

- In general, for a normal distribution, with confidence 95% all random values are located within the 2-sigma interval $[\mu - 2\sigma, \mu + 2\sigma]$.
- In our case, this means that the actual probability can be somewhere in the interval

$$\left[f - 2 \cdot \sqrt{\frac{f \cdot (1 - f)}{n}}, f + 2 \cdot \sqrt{\frac{f \cdot (1 - f)}{n}}\right].$$

16. How to take this difference into account when making a decision

- As we have mentioned, all we know about the probability p is that it is located somewhere on the interval.
- This probability may be smaller than f, it may be larger that the frequency f.
- Disaster preparedness means preparing for the worst possible scenario.
- So, it makes sense to consider the worst-case probability:

$$\overline{p} = f + 2 \cdot \sqrt{\frac{f \cdot (1 - f)}{n}}.$$

- So, the idea is to use this higher probability instead of the frequency when comparing the importance of different zones.
- In other words, instead of using the products $p \cdot d$ for $p = f$ (as in the traditional approach), we need to use the products $\overline{p} \cdot d$.

17. How this will affect our ranking of high-impact low-probability events

- Let us first illustrate, on the above two examples, how our estimate for probability will change if we use the new technique.
- In the first example, $\sqrt{\dfrac{f\cdot(1-f)}{n}}=\sqrt{\dfrac{0.1\cdot 0.9}{200}}=\sqrt{0.00045}\approx 0.02.$
- Thus, $\overline{p}\approx 0.1+2\cdot 0.02=0.14.$
- This probability is somewhat larger than the frequency (actually, 40% larger).
- However, it is in the same range as the frequency.

18. How this will affect our ranking of high-impact low-probability events (cont-d)

- In the second example, $\sqrt{\dfrac{f\cdot(1-f)}{n}}=\sqrt{\dfrac{0.01\cdot 0.99}{200}}=\sqrt{0.00005}\approx 0.007$.
- Thus, $\overline{p}\approx 0.01+2\cdot 0.007=0.024$.
- This probability is more than twice larger than the frequency.
- Actually, here, the 2-sigma term is larger than the original frequency.

19. Analysis and the resulting quantitative recommendations

- An important difference between these two examples is as follows.
- In the first example, the frequency is larger than the 2-sigma term, so the estimate $\overline{p}$ is of the same order as frequency.
- In the second example, the 2-sigma term is larger than the original frequency, so the estimate $\overline{p}$ is of the same order as the 2-sigma term.
- The borderline between these two cases if when the two terms are equal, i.e., when $f = 2 \cdot \sqrt{\dfrac{f \cdot (1-f)}{n}}$.
- Here, the frequency f is much smaller than 1, so $1 - f \approx 1$.
- Thus, the above formula takes the simplified form: $f = 2 \cdot \sqrt{\dfrac{f}{n}}$.
- By squaring both sides, we get $f^2 = 4 \cdot \dfrac{f}{n}$, i.e., equivalently, $f = \dfrac{4}{n}$.

20. Analysis and the resulting quantitative recommendations (cont-d)

- So, if we have more than 4 past events of this type, then the estimate $\overline{p}$ is of the same order as frequency.
- In this case, the usual estimate for the product $p \cdot d$ works OK.
- However, if we have fewer than 4 past events of this type, then the estimate $\overline{p}$ is much larger than the frequency.
- Thus, the usual estimate for the product leads to a drastic underestimation of this event's importance.

21. Quantitative consequences

- he new estimate $\overline{p}$ for the probability is, in general, larger than the usual estimate f.
- Thus, the new value $\overline{p} \cdot d$ is larger than the value $f \cdot d$ estimated by the traditional method.
- How larger? The ratio between these two products is equal to

$$\frac{\overline{p} \cdot d}{f \cdot d} = \frac{\overline{p}}{f} = \frac{f + 2 \cdot \sqrt{\dfrac{f \cdot (1 - f)}{n}}}{f} = 1 + 2 \cdot \sqrt{\frac{1 - f}{f \cdot n}} = 1 + \frac{2}{\sqrt{n}} \cdot \sqrt{\frac{1}{f} - 1}.$$

- We can see that as the frequency f decreases, this ratio grows – and this ratio tends to infinity as f tends to 0.
- So, the use of this new techniques increase the low-probability product much higher than high-probability one.
- Thus, it makes high-impact low-probability events even more important.

22. Acknowledgments

This work was supported in part by:

- National Science Foundation grants 1623190, HRD-1834620, HRD-2034030, and EAR-2225395;
- AT&T Fellowship in Information Technology;
- program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
- a grant from the Hungarian National Research, Development and Innovation Office (NRDI).