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

Aaron Velasco, Olga Kosheleva, and Vladik Kreinovich

Abstract A large proportion of undesirable events like earthquakes, floods, tornados occur in zones where these events are frequent. However, a significant number of such events occur in other zones, where such events are rare. For example, while most major earthquakes occur in a vicinity of major faults, i.e., on the border between two tectonic plates, some strong earthquakes also occur inside plates. We want to mitigate all undesirable events, but our resources are limited. So, to allocate these resources, we need to decide which ones are more important. For this decision, a natural idea is to use the product of the probability of the undesirable event and possible damage caused by this event. A natural way to estimate probability is to use the frequency of such events in the past. This works well for high-probability events like earthquakes in a seismic zone near a fault. However, for high-impact low-probability events the frequency is small and, as a result, the actual probability may be very different from the observed frequency. In this paper, we show how to take this difference between frequency and probability into account. We also show that if we do take this difference into account, then high-impact low-probability events turn out to be even more important than it is usually assumed.

Aaron Velasco
Department of Earth, Environmental, and Resource Sciences, University of Texas at El Paso
500 W. University, El Paso, Texas 79968, USA, e-mail: aavelasco@utep.edu

Olga Kosheleva
Department of Teacher Education, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: olgak@utep.edu

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: vladik@utep.edu

1 Introduction

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, and everyone understands that we need to be prepared for situations when the strength will becomes high, causing potentially disastrous consequences.

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 – and the only reason we know about it is indirectly, by the effect of this strong past earthquake on the geology of the region.

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. 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.

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.

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: e.g., 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, and 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?

What we do in this paper. 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.

2 Current Way of Allocating Resources: Justification, Description, and Limitations

Justification: let us use decision theory. There is a whole science of rational decision making, known as *decision theory*; see, e.g., [1, 2, 3, 4, 5, 6, 7]. 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*, so that 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:

$$u = p_1 \cdot u_1 + p_2 \cdot u_2 + \ldots$$

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.

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.

Comment. 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 in the previous section, the potential damage can be much higher, since such zones are usually unprepared for the undesirable event (or at least much less prepared).

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.

Example 1. If in 200 years of record, the major Spring flood occurred 20 times, we estimate the probability of the flood as

$$20/200 = 0.1 = 10\%.$$

Example 2. 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

$$2/200 = 0.01 = 1\%.$$

Limitations. 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.

3 Coming Up with a More Adequate Technique for Allocating Resources

What do we know about the difference between probability and frequency. According to statistics (see, e.g., [8]), 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

$$\sigma \approx \sqrt{\frac{f \cdot (1-f)}{n}}.$$

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

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

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 (1). 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:

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

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 $\overline{p} \cdot d$, where $\overline{p}$ is determined by the formula (2).

4 How This Will Affect Our Ranking of High-Impact Low-Probability Events

Examples. Before providing a general analysis, let us first illustrate, on the above two examples, how our estimate for probability will change if we use the new technique.

Example 1. In this case,

$$\sqrt{\frac{f \cdot (1-f)}{n}} = \sqrt{\frac{0.1 \cdot 0.9}{200}} = \sqrt{0.00045} \approx 0.02.$$

Thus, $\overline{p} \approx 0.1 + 2 \cdot 0.02 = 0.14$. This probability is somewhat larger than the frequency (actually, 40% larger) but it is in the same range as the frequency.

Example 2. In this case,

$$\sqrt{\frac{f \cdot (1-f)}{n}} = \sqrt{\frac{0.01 \cdot 0.99}{200}} = \sqrt{0.00005} \approx 0.007.$$

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.

Analysis and the resulting quantitative recommendations. An important difference between these two examples is that:

- 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, while
- 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

$$f = 2 \cdot \sqrt{\frac{f \cdot (1-f)}{n}}. \qquad (3)$$

Here, the frequency f is much smaller than 1, so $1 - f \approx 1$ and thus, the formula (3) takes the following simplified form:

$$f = 2 \cdot \sqrt{\frac{f}{n}}. \qquad (4)$$

By squaring both sides, we get

$$f^2 = 4 \cdot \frac{f}{n},$$

i.e., equivalently,

$$f = \frac{4}{n}.$$

Thus:

- 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 and thus, the usual estimate for the product leads to a drastic underestimation of this event's importance.

Quantitative consequences. The 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

$$\frac{\overline{p} \cdot d}{f \cdot d} = \frac{\overline{p}}{f} = \frac{f + 2 \cdot \sqrt{\frac{f \cdot (1-f)}{n}}}{f} = 1 + 2 \cdot \sqrt{\frac{1-f}{f \cdot n}} = 1 + \frac{2}{\sqrt{n}} \cdot \sqrt{\frac{1}{f} - 1}. \qquad (5)$$

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

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), HRD-1834620 and HRD-2034030 (CAHSI Includes), EAR-2225395, and by the AT&T Fellowship in Information Technology.

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

References

1. P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.
2. P. C. Fishburn, *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, Maryland, 1988.
3. V. Kreinovich, "Decision making under interval uncertainty (and beyond)", In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models for Social Sciences*, Springer Verlag, 2014, pp. 163–193.
4. R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.
5. H. T. Nguyen, O. Kosheleva, and V. Kreinovich, "Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction", *International Journal of Intelligent Systems*, 2009, Vol. 24, No. 1, pp. 27–47.
6. H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.
7. H. Raiffa, *Decision Analysis*, McGraw-Hill, Columbus, Ohio, 1997.
8. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.