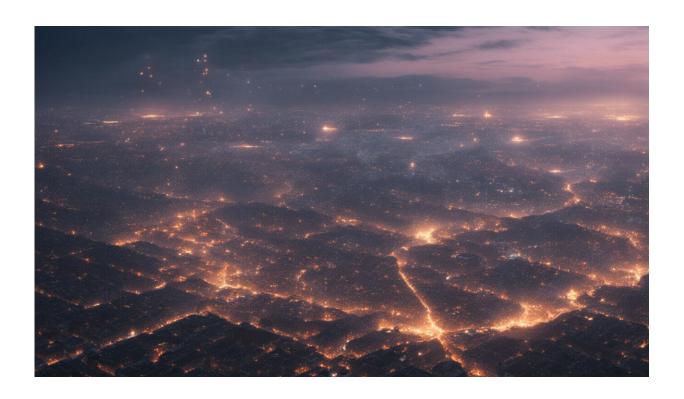# New method to analyze low-probability, high-risk events such as earthquakes, pandemics

November 18 2021, by Louise Lerner

Credit: AI-generated image ([disclaimer](#))

Quick—if you had to guess, what would you think is most likely to end all life on Earth: a meteor strike, climate change or a solar flare? (Choose carefully.)

A new [statistical method](#) could help accurately analyze the risk of very

worst (or best) case scenarios. Scientists have announced a new way to tease out information about events that are rare, but highly consequential—such as pandemics and insurance payouts.

The discovery helps statisticians use math to figure out the shape of the underlying distribution of a set of data. This can help everyone from investors to government officials make informed decisions—and is especially helpful when the data is sparse, as for major earthquakes.

"Though they are by definition rare, such events do occur and they matter; we hope this is a useful set of tools to understand and calculate these risks better," said mathematical biologist Joel Cohen, a co-author of a new study published Nov. 16 in the *Proceedings of the National Academy of Sciences*. A visiting scholar with the University of Chicago's statistics department, Cohen is a professor at the Rockefeller University and at the Earth Institute of Columbia University.

Varying the questions

Statistics is the science of using limited data to learn about the world—and the future. Its questions range from "When is the best time of year to spray pesticides on a field of crops?" to "How likely is it that a global pandemic will shut down large swaths of public life?"

At a century old, the statistical theory of rare-but-extreme events is a relatively new field, and scientists are still cataloging the best ways to crunch different kinds of data. Calculation methods can significantly affect conclusions, so researchers have to tune their approaches to the data carefully.

Two powerful tools in statistics are the average and the variance. You're probably familiar with the average; if one student scores 80 on a test and one student scores 82, their average score is 81. Variance, on the other

hand, measures how widely spread out those scores are: You'd get the same average if one student scored 62 and the other scored 100, but the classroom implications would be very different.

In most situations, both the average and the variance are finite numbers, like the situation above. But things get stranger when you look at events that are very rare, but enormously consequential when they do happen. In most years, there isn't a gigantic burst of activity from the sun's surface big enough to fry all of Earth's electronics—but if that happened this year, the results could be catastrophic. Similarly, although the vast majority of tech startups fizzle out, a Google or a Facebook occasionally comes along.

"There's a category where large events happen very rarely, but often enough to drive the average and/or the variance towards infinity," said Cohen.

These situations, where the average and variance approach infinity as more and more data is collected, require their own special tools. And understanding the risk of these types of events (known in statistical parlance as events with "heavy-tailed distribution") is important for many people. Government officials need to know how much effort and money they should invest in disaster preparation, and investors want to know how to maximize returns.

Cohen and his colleagues looked at a mathematical method recently used to calculate risk, which splits the variance in the middle and calculates the variance below the average, and above the average, which can give you more information about downside risks and upside risks. For example, a tech company may be much more likely to fail (that is, to wind up below the average) than to succeed (wind up above the average), which an investor might like to know as she's considering whether to invest. But the method had not been examined for distributions of low-

probability, very high-impact events with infinite mean and variance.

Running tests, the scientists found that standard ways to work with these numbers, called semi-variances, don't yield much information. But they found other ways that did work. For example, they could extract useful information by calculating the ratio of the log of the average to the log of the semi-variance. "Without the logs, you get less useful information," Cohen said. "But with the logs, the limiting behavior for large samples of data gives you information about the shape of the underlying distribution, which is very useful." Such information can help inform decision-making.

The researchers hope this lays the foundation for new and better exploration of risks.

"We think there are practical applications for financial mathematics, for agricultural economics, and potentially even epidemics, but since it's so new, we're not even sure what the most useful areas might be," Cohen said. "We just opened up this world. It's just at the beginning."

Provided by University of Chicago