Some mathematical musings about extreme events for geotechnical risk, without (much of) the math

Pete Quinn
BGC Engineering Inc., Mons, Belgium

ABSTRACT
Extreme events seem to occur with surprising regularity, at frequencies greatly exceeding those expected from a typical analysis of probabilities that assume well-behaved statistical characteristics. Many natural phenomena, including those involving geologic materials, have heavy tails that decay much slower than a normal distribution, often as a power law, with fractal geometric character and associated non-intuitive behavior, particularly for the rare events that represent the tails of the probability distributions. This paper exploits an example from an unrelated field – finance – to illustrate key behaviours of power law phenomena before returning to those of geologic materials, with discussion of important implications to geohazard risk assessment.

RÉSUMÉ
Les événements extrêmes semblent se produire avec une régularité surprenante, à des fréquences largement supérieures à celles attendues d'une analyse typique de probabilités reposant sur des caractéristiques statistiques bien agencées. De nombreux phénomènes naturels, y compris ceux impliquant des matériaux géologiques, ont des queues lourdes qui se désintègrent beaucoup plus lentement qu'une distribution normale, souvent sous forme de loi de puissance, avec un caractère géométrique fractal et un comportement associé non intuitif, en particulier pour les événements rares qui représentent les queues des distributions de probabilité. Cet article exploite un exemple tiré d'un domaine non lié - la finance - pour illustrer les comportements clés des phénomènes de loi de puissance avant de revenir à ceux des matériaux géologiques, en analysant les implications importantes pour l'évaluation des risques géologiques.

1 INTRODUCTION

This paper is about risk, which we may define as expected loss. A key notion about risk is that it is uncertain; the word "expected" conveys a best estimate, or perhaps most likely value, but is by no means the only possible outcome. Risk cannot be calculated deterministically, rather it is a probabilistic metric. Probability may be estimated statistically, using frequence notions. Alternatively, it may be estimated subjectively, using Bayesian thinking, as a measure of degree of belief. In either event, the outcome is a random variable with some degree of (perhaps unknown) variability.

This paper first examines the mental process involved in approaching the problem of probability estimation, and then shows how we commonly obtain extremely unconservative estimates of the risk from rare events, using an example from outside geotechnique. This finding is used to emphasize the importance of properly considering rare events in risk analysis, by either accurately estimating their likelihood, or ensuring the choice of risk management solution is robust against uncertainty in their true frequency. In the absence of a clear basis to do otherwise, a defensible first, default approach is to plan for the maximum credible event.

2 HOW ONE APPROACHES THE PROBLEM OF RISK ANALYSIS: PROBABILITY ESTIMATION AND RARE EVENTS

The following discussion represents the experience, observations and opinions of the author, and may not withstand serious academic scrutiny; nevertheless, they are honest representations of real observations and experience, and the main conclusions are defended by examples of real phenomena.

In geotechnical problems, risk is often estimated by geotechnical or geological engineers, who tend to develop practical applications for problem solving and design from past experience, or by reference to empiricism. Such experience, both positive (e.g., successful structures) and negative (e.g., failures), serves as a statistical basis or reference that supports confidence in future analysis. The use of empiricism usually includes the parallel development of mental models representing the applicable physical systems. These physical models are subsequently represented by mathematical models, which may be solved by simple equations when reasonably representative closed-form solutions exist. They may also be attacked by advanced numerical methods which solve the partial differential equations assumed to represent physical reality; usually when appropriate closed-form solutions are beyond reach.

Often implicit in either mathematical approach is an assumption of determinism; in other words, an equation or numerical model produces an answer which is, at least by many people or at first glance, assumed to be correct, accurate and precise. An engineer with proper training will recognize that the answer is uncertain, to some degree, and may remember to, for example, appropriately limit the number of significant digits. This nuance may be lost on some readers, for example laypeople decision-makers. Importantly, the distinction may sometimes be forgotten by the analyst; an engineer calculating allowable bearing capacity of a footing may assume values for friction angle, cohesion, unit weight and groundwater depth, and imagine that the calculated bearing capacity is meaningful to within, say, plus or minus 15 to 25%, perhaps forgetting that the safety factor applied in old working stress design approaches was typically 3, reflecting the considerable uncertainty in those assumed parameters, loads and problem geometry.

Bringing this discussion back to the subject of risk analysis, we have already asserted that risk estimation is a problem of probability estimation. In the author's experience, most people do not think probabilistically, at least not as a natural default. Notions of probability are conceptually difficult. People tend to first think in black and white, true or false, with little instinctive room for a middle setting on the mental switch. A first conceptual step toward probabilistic thinking is to allow a third, or middle setting, or "maybe" rather than yes or no. This transforms the binary range of possibilities to something closer to a spectrum, but with only one shade of grey, and may be adequate for dealing with a range of day to day problems but is surely inadequate for rigorous risk analysis.

A next natural conceptual step is to allow that "maybe" can have a range of values between certainly true (or white, or $P = 1$) and certainly false (or black, or $P = 0$). "Maybe" can be represented numerically as a probability ranging between 0 and 1, or by a descriptor that has some tangible meaning to the user (e.g., very likely).

In engineering and quantitative risk analysis, the first and most natural notion of probability is derived from frequence approaches, whereby the probability is calculated as the number of positive outcomes in a set of trials. This can be done in a mathematically precise way for flipping fair coins or rolling fair dice, or in a less precise, but equally valid way, philosophically, where a number of trials can be imagined in a credible, defensible, way. Often, one is forced to use a relatively sparse set of empirical data as the basis for generating probabilities. Taleb (2010) cautions us about the Ludic fallacy, whereby one assumes that calculated probabilities have the same meaning as those derived for games of chance, which is rarely the case in nature or life outside casinos.

With training in probability and statistics, and experience putting the concepts into practice, frequence notions become relatively natural. A next conceptual step, which is a critical fundamental concept, but perhaps (in the author's experience and observation at any rate) the most difficult to grasp, is to move from the frequence frame to the subjective, or Bayesian, where probability is a measure of degree of belief or uncertainty. The original mathematical formulation of the concept of probability, with primary roots in the 16th (Gerolamo Cardano) and 17th (Pierre de Fermat and Blaise Pascal) centuries, imposed no specific basis for its estimation, and both frequence and subjective probabilities have the same validity in a philosophical sense (Vick, 2002).

The last challenge – that of embracing subjective probability – is of eminent concern when working with qualitative risk. The balance of this paper will focus on specific aspects of quantitative risk assessment hence we return to frequence notions of probability.

Natural phenomena or characteristics are often observed to follow a Gaussian, or normal distribution, with the bulk of events occurring near some average, or mean value which is also – in the Gaussian case - the middle, or median - 50th percentile - and most common, or modal value. The normal distribution, or bell curve, is an easy, familiar concept, often assumed to represent reality, and perhaps an instinctive default for any first exploration of the statistics of a given problem. With a normally distributed phenomenon, one need only know the average value and some measure of variability (often the standard deviation) to have a complete working understanding of variability or uncertainty. While many natural phenomena are not normally distributed in a precise sense, the bulk of their behavior can often be well captured by the bell curve or by a simple transformation of it; for example, a log-normal distribution, wherein the logarithm of the measure is normally distributed.

When the behavior of a system is determined by the typical – or average - event, a Gaussian description is fully adequate, as no single occurrence has a measurable influence on overall behavior. As an example, consider the total net weight of a truckload of healthy free-range chickens on their way from farmer Bob to farmer Jane for better sunshine; if we know the average weight of a chicken is 2.5 kg, but the birds in the truck range from about 1 to 5 kg, we can be very confident that a truck with 1000 chickens will carry about 2500 kg of happy birds, plus or minus a tiny margin of error. There is no credible possibility that any of the chickens might weigh 200 kg, thus skewing the total weight well outside expectations.

The Law of Large Numbers (first proved by Jacob Bernoulli) tells us that after a large number of trials, the average value of all trials will converge on the expected value. However, when phenomena have "wild variability" (Mandelbrot and Hudson, 2004), they may not have a calculable mean value (or other statistical moments); in these cases, the condition of the overall system may be dominated by a small number of events. Sornette (2006) provides several examples of natural phenomena that are dominated by extreme events:

- In California, the maximum credible earthquake accounts for about one third of long-term energy released from the crust.
- For some watersheds, the century or millennium-scale flood has more landscape-changing effect than all other events combined. The 2014 Canmore, AB and 1996 Saguenay, QC flood events are two Canadian examples from recent memory.
- The largest volcanic eruptions cause major meteorological perturbations that may lead to important modifications of the biosphere.

- The largest earthquake, hurricane, tsunami or other natural disaster may have an important, possibly debilitating effect on a region's or country's economy and social fabric. The 2004 Indian Ocean tsunami and 2005 Hurricane Katrina are both good examples.

Such phenomena have "wild variability," or "heavy tails" (Taleb 2010; Mandelbrot and Hudson 2004). Their probability distributions may be relatively well described by normal distributions for typical values close to their "average" values (e.g., mean, median, modal, geometric mean), but the frequency of rare events is usually underpredicted. The probability distribution of extreme events, or more precisely of the tails of the overall probability distributions, may follow power laws, but in any event decay more slowly than the bell curve, which decays exponentially. When the exponent of the power law is appropriate, the overall behavior may be dominated by the rare events, whose frequency may be difficult to estimate, given their rarity.

Risk analysis begins with an estimation of the temporal frequency of some potentially damaging event. If extreme events are the most damaging and make a significant contribution to total estimated potential damage, it is critical to estimate their frequency accurately; however, as the following section illustrates, rare events governed by power laws may occur much more often than intuitively expected.

3 EXAMPLE OF A PHENOMENON WITH "WILD" VARIABILITY – STOCK MARKET INDEX PRICE CHANGES

The following section illustrates some of the difficulties in evaluating risk from extreme events, using an example from outside geotechnique: daily price changes in a financial market, specifically the Standard and Poor's 500 (S&P 500) index, which tracks the value of stocks for 500 of the largest US companies by market capitalization. This example was chosen due to the easy availability of long time-series datasets. While the use of such an example may seem incongruent for examining geohazard risk, the key similarity – power law behavior of rare events – is very much like that for many natural phenomena, and as such renders the example meaningful for discussion.

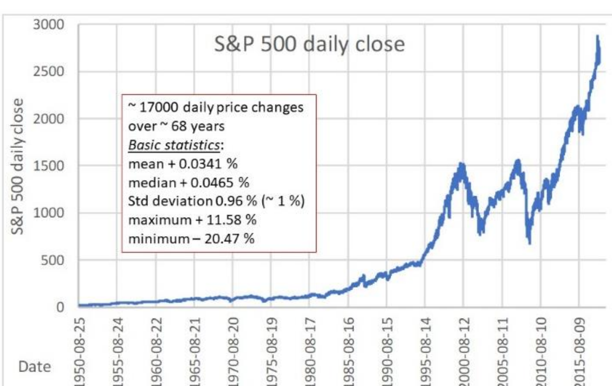

Figure 1. S&P 500 index values from 1950 to 2018.

Figure 1 shows index values from 1950 to 2018 for the S&P 500, with some basic statistics. The average (mean or median) daily change is very close to zero, but slightly positive, leading to an overall increase in index value over 68 years. These simple statistics are further illustrated in Figure 2 and Figure 3, which show the daily price changes plotted with the normal distribution that best fits the data. Figure 3, which shows frequency on a logarithmic scale, gives a reasonable fit to the normal distribution around the middle region, for the most common daily changes. This plot also emphasizes the frequency of outlier – or rare, extreme – events.

If we examine the right side of Figure 3, there are 17 daily price changes of 6% or greater – roughly 6 standard deviations distant from the mean value. The best fit normal distribution predicts zero events of this magnitude (more precisely, we expect 0.000006 events, which is so much closer to 0 than 1 as to suggest no events being credible); therefore, the actual occurrence of events of at least mean value plus six standard deviations is roughly 2.7 million times more frequent than expected, if we presume the Gaussian, or normal distribution. The most extreme (largest) event – a daily decline of just over 20% - should not occur in theory; its single occurrence is 1.2 x 10^{91} times more frequent than expected, rendering it entirely implausible, if we were to assume validity of the best fit normal distribution.

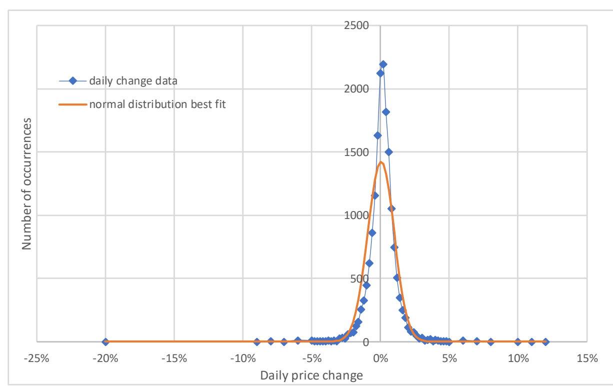

Figure 2. Distribution of daily index value changes: S&P 500.

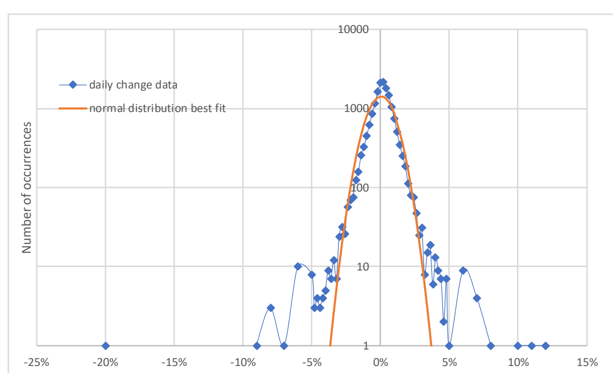

Figure 3. Distribution of daily index value changes: S&P 500. Frequency plotted on logarithmic scale.

Further insight can be gained by plotting the daily price changes in rank order. Figure 4 shows a log-log plot of the absolute values of daily price changes, with the largest daily decline and gain being the first and second plotted values. The rank order plot can be broadly subdivided into two main regions, with a flatter nearly linear portion followed by a steep downturn. The vast majority of daily price changes – ranging between about – 0.5% to + 0.5% - are in this latter region. Figure 5 shows the 500 largest daily price changes, with a best fit power law trend. One can see that the power law gives a very good fit to the data, with increasing variability or uncertainty (i.e., variation from the best fit trend) at lower rank order (i.e., for the rarest events).

Power law phenomena have fractal characteristics (Mandelbrot 1982) and as such are scale-invariant. This suggests that the maximum observed, or expected, value depends on the scale – or duration - of observation. Figure 6 illustrates this idea: within a year, or about 200 trading days, we expect to see a maximum daily change of +/- 3 to 5%, roughly (i.e., 3 to 5 standard deviations for this data set), with nearly all daily changes less than +/- 0.5%. By contrast, if we follow the market for ten years or more, we expect to see at least one daily change on the order of about +/- 10%, or ten standard deviations from the mean.

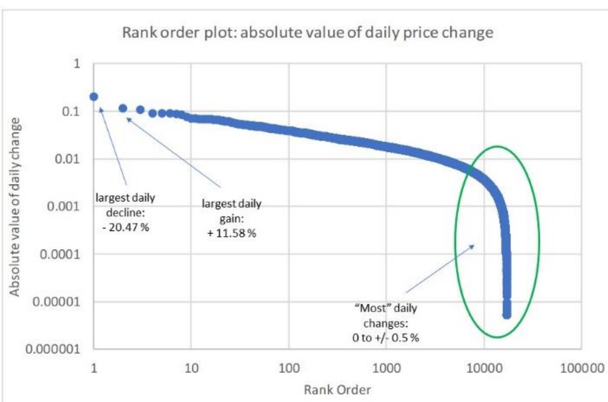

Figure 4. Rank order plot of daily index value changes: S&P 500.

A key characteristic of power law phenomena is that the rare events may dominate the long-term trends. This is illustrated for the S&P 500 in Figure 7: if we remove the ten largest daily gains from the record, today's index value would be roughly half its actual value. In other words, 10 out of about 17,000 events account for half of the overall behavior of the system, and the largest of these events – a single day decline – accounts for about one fifth (!!) of the overall behavior.

We can contrast this power law behavior with an example of tame variability. Thinking back to our free-range chickens in Section 2; the largest chicken, which at 5 kg is twice the average weight, accounts for about 0.2% of the total weight of the 1000 chickens. It is nearly impossible to imagine a single chicken being added to the truck that would weigh even as much as 1% of the group (that would require a 25 kg chicken), and fully impossible to imagine 20% (a 500 kg chicken). However, such is the possible effect when a phenomenon's rare events are governed by a power law; when the power law has appropriate characteristics, the largest (rarest) event can contribute a finite (non-vanishing) proportion of the overall behavior, or in the terminology of Sornette (2006), be a "king."

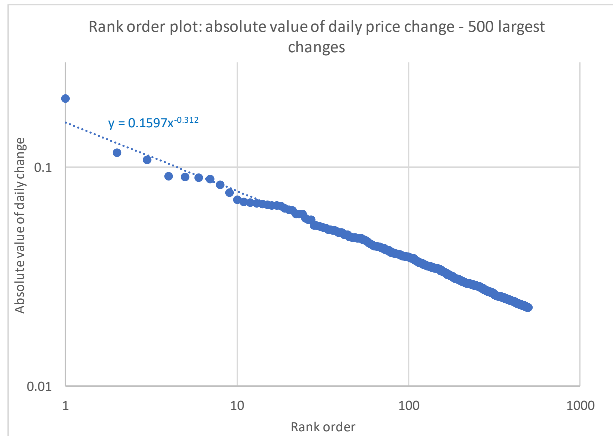

Figure 5. Rank order plot of the largest 500 daily index value changes: S&P 500.

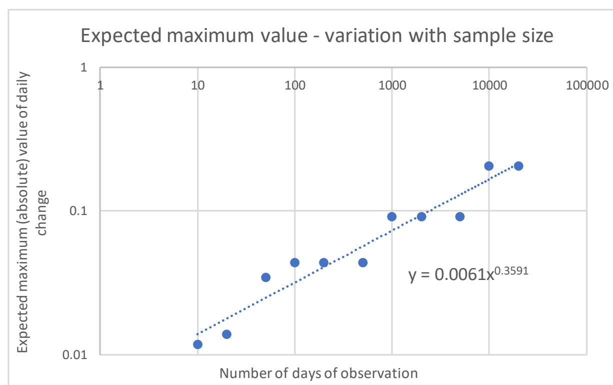

Figure 6. Expected maximum daily price change as a function of period of observation.

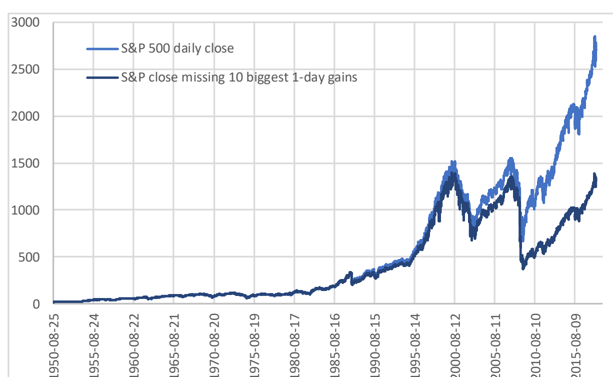

Figure 7. S&P 500 since 1950 to 2018, also showing the hypothetical chart if the top ten daily gains were missing.

5 DISCUSSION

The financial example discussed in the previous section was used mainly because it is derived from a large dataset that is widely available and serves to easily illustrate many important concepts in the statistics of extreme events for phenomena with “wild” variability.

Social and physical phenomena commonly have fractal characteristics, with power law tails for their statistical distributions. The literature on this topic is very rich, and includes, for example, Becerra et al. (2006), Benguigui and Marinov (2015) and Hergarten (2004). The latter author also discusses risk assessment for natural hazards affected by power laws, distinguishing between those where expected loss is governed by the typical event and those governed by the extreme event. Therefore, a key point in this paper – that risk assessment for power law phenomena requires careful consideration of the tails of the distribution – is already established.

The main thesis in this paper is, however, that in practice it may be conceptually difficult to properly consider rare events in risk analysis, and thus they tend to be systematically and substantially underestimated. First, one most recognize and accept that rare events, which may never have been experienced or encountered in practice by the analyst, are important and need to be well understood. Second, one must resist the instinct to defer to first order, simple, deterministic analysis and make the mental leap to more nuanced probabilistic thinking, particularly considering the probability of rare events. Third, generating a statistical description of rare events is very difficult, since the available data (i.e., past occurrences) are, by definition, infrequent or possibly unobserved.

Natural disasters and failure of some man-made structures seem to occur with remarkable regularity, and often with surprising force. Mine tailings storage areas, for example, have suffered about 4 to 5 failures per year, on average, since 1978 (ICOLD 2001). More than ten disasters around the world since 1900 – including floods, cyclones, earthquakes and dam failures, but excluding technological accidents, epidemics and armed conflict – have each caused more than 100,000 fatalities (Wikipedia 2019), with the largest, a China flood in 1931, causing between 1 and 4 million fatalities. If we limit our concern to risk from geohazards other than earthquakes and floods, landslides, snow avalanches and volcanic eruptions have often caused thousands of casualties.

We conjecture here that extreme geohazard events occur more frequently than expected mainly because expectations, being largely subjective and not objective, are shaped by an inappropriate mental model of probability. The conceptual difficulty in understanding the statistics of rare events, combined with the absence of their past experience, leads to the naïve but natural (and erroneous) expectation that rare events are much less common than in reality. People are usually surprised by the most damaging events, thinking of them as “acts of god” rather than the normal phenomena they are, occurring with regularity that can be expected from proper study.

Proper risk analysis for geohazard events, which very often have power law characteristics, demands careful consideration of the true frequency of rare events. Where this cannot be estimated with high confidence, a safe approach may be to choose a course of action that is robust against uncertainty in frequency estimation, or in other words to have the maximum credible event factor heavily in the management approach.

ACKNOWLEDGEMENTS

This work was first formalized for an internal BGC workshop on mining risk in late 2018. The ideas in this paper have evolved in the author’s thinking over the past 30 years, but as they say there are no new ideas; perhaps just new ways of framing old ideas. The author was recently encouraged by colleagues to read certain authors, and he worked through recent books by Daniel Kahneman, Philip Tetlock, Malcolm Gladwell and Nassim Taleb, and the paper takes some inspiration from each of those writers. On reading the Black Swan in early 2019, the author found a graph very similar to that in Figure 17, which the author had previously believed to be original; it was obviously not, having been much earlier published by Taleb. The chicken example is also directly inspired by Taleb’s writings.

The general philosophical arc took significant inspiration from many lengthy discussions about mathematics, and in particular about complexity and fractal geometry, with the author’s late son, John.

REFERENCES

Becerra, O., N. Johnson, P. Meier, J. Restrepo, and M. Spagat, 2006. Natural Disasters, Casualties and Power Laws: A comparative analysis with Armed Conflict. Annual Meeting of the American Political Science Association, Philadelphia, PA.

Benguigui, L. and M. Marinov. 2015. A classification of natural and social distributions Part one: the descriptions. Downloaded from: [https://arxiv.org/ftp/arxiv/papers/1507/1507.03408.pdf] on 27 May 2019.

Hergarten, S. 2004. Aspects of risk assessment in power-law distributed natural hazards. Natural Hazards and Earth Science Systems. (2004) 4: 309-313.

ICOLD Committee on Tailings Dams and Waste Lagoons (1995-2001). 2001. Tailings Dams Risk of Dangerous Occurrences – Lessons Learnt from practical experiences.

Mandelbrot, B. and R.L. Hudson, 2004. The (Mis)behavior of markets; A fractal view of financial turbulence. Basic Books, New York.

Mandelbrot, B. 1982. The Fractal Geometry of Nature. W.H. Freeman and Company.

Schroeder, M. 1991. Fractals, Chaos, Power Laws; Minutes from an infinite paradise. Dover Publications, Inc., Mineola, New York.

Sornette, D. 2006. Critical Phenomena in Natural Sciences; Chaos, Fractals, Selforganization and Disorder: Concepts and Tools, Second Edition. Springer.

Taleb, N.N., 20010. The Black Swan: Second Edition. Random House, New York.

Vick, S.G., 2002. Degrees of Belief: Subjective Probability and Engineering Judgment. ASCE Press, Reston, Virginia.

Wikipedia, 2019. List of natural disasters by death toll. Accessed on 27 May 2019 at: [https://en.wikipedia.org/wiki/List_of_natural_disasters_by_death_toll]