Statistics

Dyrehaugen Memo

2021-01-20

Statistics

Science uses statistics &, as per Popper, doesn't really "accept", just fails to reject at some significance. It's fundamentally disconfirmatory. Stat. "evidence" is inherently probabilistic and cannot be "degenerate" (i.e. provide certainties). (Taleb)

Fat Tails

Extremes

The field of Extreme Value Theory focuses on tail properties, not the mean or statistical inference.

It is vastly more effective to focus on being insulated from the harm of random events than try to figure them out in the required details (the inferential errors under fat tails are huge). So it is more solid, much wiser, more ethical, and more effective to focus on detection heuristics and policies rather than fabricate statistical properties.

Catastrophe Principle

Memo Taleb (DarwinCollege):

Where a Pareto distribution prevails (among many), and randomly select two people with combined wealth of £36 million. The most likely combination is not £18 million and £18 million. It is approximately £35,999,000 and £1,000. This highlights the crisp distinction between the two domains; for the class of subexponential distributions, *ruin is more likely to come from a single extreme event than from a series of bad episodes.* This logic underpins classical risk theory as outlined by Lundberg early in the 20 th Century and formalized by Cramer, but forgotten by economists in recent times. This indicates that insurance can only work in Medocristan; you should never write an uncapped insurance contract if there is a risk of catastrophe. The point is called the catastrophe principle.

Cramer showed insurance could not work outside what he called the Cramer condition, which excludes possible ruin from single shocks.

With fat tail distributions, extreme events away from the centre of the distribution play a very large role. Black swans are not more frequent, they are more consequential. The fattest tail distribution has just one very large extreme deviation, rather than many departures form the norm.

There are three types of fat tails based on mathematical properties.

First there are entry level fat tails. This is any distribution with fatter tails than the Gaussian i.e. with more observations within one sigma and with kurtosis (a function of the fourth central moment) higher than three.

Second, there are subexponential distributions.

LogNormal:

The subexponential class includes the lognormal, which is one of the strangest things on earth because sometimes it cheats and moves up to the top of the diagram. At low variance, it is thin-tailed, at high variance, it behaves like the very fat tailed. in-tailed, at high variance, it behaves like the very fat tailed.

Membership in the subexponential class satisfies the Cramer condition of possibility of insurance (losses are more likely to come from many events than a single one) Technically it means that the expectation of the exponential of the random variable exists.

Third, what is called by a variety of names, the power law, or slowly varying class, or "Pareto tails" class correspond to real fat tails.

The traditional statisticians approach to fat tails has been to assume a different distribution but keep doing business as usual, using same metrics, tests, and statements of significance. But this is not how it really works and they fall into logical inconsistencies.

Once we are outsaide the zone for which statistical techniques were designed, things no longer work as planned. Here are some consequences

1) The law of large numbers, when it works, works too slowly in the real world (this is more shocking than you think as it cancels most statistical estimators)

2) The mean of the distribution will not correspond to the sample mean. In fact, there is no fat tailed distribution in which the mean can be properly estimated directly from the sample mean, unless we have orders of magnitude more data than we do

3) Standard deviations and variance are not useable. They fail out of sample.

4) Beta, Sharpe Ratio and other common financial metrics are uninformative.

5) Robust statistics is not robust at all.

6) The so-called "empirical distribution" is not empirical (as it misrepresents the expected payoffs in the tails).

7) Linear regression doesn't work.

8) Maximum likelihood methods work for parameters (good news). We can have plug in estimators in some situations.

9) The gap between dis-confirmatory and confirmatory empiricism is wider than in situations covered by common statistics i.e. difference between absence of evidence and evidence of absence becomes larger.

10) Principal component analysis is likely to produce false factors.

11) Methods of moments fail to work. Higher moments are uninformative or do not exist.

12) There is no such thing as a typical large deviation: conditional on having a large move, such move is not defined.

13) The Gini coefficient ceases to be additive. It becomes super-additive. The Gini gives an illusion of large con- centrations of wealth. (In other words, inequality in a continent, say Europe, can be higher than the average inequality of its members).

While it takes 30 observations in the Gaussian to stabilize the mean up to a given level, it takes 10^{11} observations in the Pareto to bring the sample error down by the same amount (assuming the mean exists). You cannot make claims about the stability of the sample mean with a fat tailed distribution. There are other ways to do this, but not from observations on the sample mean.

We have known at least since Sextus Empiricus that we cannot rule out degeneracy but there are situations in which we can rule out non-degeneracy. If I see a distribution that has no randomness, I cannot say it is not random. That is, we cannot say there are no black swans. Let us now add one observation. I can now see it is random, and I can rule out degeneracy. I can say it is not not random. On the right hand side we have seen a black swan, therefore the statement that there are no black swans is wrong. This is the negative empiricism that underpins Western science. As we gather information, we can rule things out. If we see a 20 sigma event, we can rule out that the distribution is thin-tailed.

Pareto - Scalability

The intuition behind the Pareto Law. It is simply defined as: say X is a random variable. For x sufficently large, the probability of exceeding 2x divided by the probability of exceeding x is no different from the probability of exceeding 4x divided by the probability of exceeding 2x, and so forth.

So if we have a Pareto (or Pareto-style) distribution, the ratio of people with £16 million compared to £8 million is the same as the ratio of people with £2 million and £1 million. There is a constant inequality.

This distribution has no characteristic scale which makes it very easy to understand. Although this distribution often has no mean and no standard deviation we still understand it. But because it has no mean we have to ditch the statistical textbooks and do something more solid, more rigorous.

A Pareto distribution has no higher moments: moments either do not exist or become statistically more and more unstable.

In 2009 I took 55 years of data and looked at how much of the kurtosis (a function of the fourth moment) came from the largest observation. For a Gaussian the maximum contribution over the same time span should be around $.008 \pm .0028$. For the S&P 500 it was about 80 per cent. This tells us that we dont know anything about kurtosis. Its sample error is huge; or it may not exist so the measurement is heavily sample dependent. If we dont know anything about the fourth moment, we know nothing about the stability of the second moment. It means we are not in a class of distribution that allows us to work with the variance, even if it exists. This is finance.

We cannot use standard statistical methods with financial data.

Financial data, debunks all the college textbooks we are currently using Econometrics that deals with squares goes out of the window. The variance of the squares is analogous to the fourth moment. The variance of the squares is analogous to the fourth moment. We do not know the variance. But we can work very easily with Pareto distributions. They give us less information, but nevertheless, it is more rigorous if the data are uncapped or if there are any open variables.

Principal component analysis is a dimension reduction method for big data and it works beautifully with thin tails. But if there is not enough data there is an illusion of a structure. As we increase the data (the n variables), the structure becomes flat.

Lessons:

Once we know something is fat-tailed, we can use heuristics to see how an exposure there reacts to random events: how much is a given unit harmed by them. It is vastly more effective to focus on being insulated from the harm of random events than try to figure them out in the required details (as we saw the inferential errors under fat tails are huge). So it is more solid, much wiser, more ethical, and more effective to focus on detection heuristics and policies rather than fabricate statistical properties.

The beautiful thing we discovered is that everything that is fragile has to present a concave exposure similar –if not identical –to the payoff of a short option, that is, a negative exposure to volatility. It is nonlinear, necessarily. It has to have harm that accelerates with intensity, up to the point of breaking. If I jump 10m I am harmed more than 10 times than if I jump one metre. That is a necessary property of fragility.

We just need to look at acceleration in the tails. We have built effective stress testing heuristics based on such an option-like property.

In the real world we want simple things that work; we want to impress our accountant and not our peers. (My argument in the latest instalment of the Incerto, Skin in the Game is that systems judged by peers and not evolution rot from overcomplication). To survive we need to have clear techniques that map to our procedural intuitions.

The new focus is on how to detect and measure convexity and concavity. This is much, much simpler than probability.

Taleb (2017) Darwin Colleges(pdf)

Statistical Consequences of Fat Tails

> Conventional statistics fail to cover fat tails; physicists who use power laws do not usually produce statistical estimators.

Taleb's Research Site

> Take nothing for granted - *It is what it is.* Another 300 years of data is required to test a statistical hypothesis. A dataset has no variance. A distribution's standard deviation will not converge in a lifetime's worth of data.

Fat tailed random variables challenge our conceptions of mean and standard deviation. Linear regression also breaks under fat tails. The convincing case is made that power law distributions should be the default for modeling data rather than the thin-tailed Normal distribution.

Any distribution with more density in the tails than the Normal distribution is said to have thick tails. This corresponds to raw kurtosis > 3. The tail density needs to decay slower than Normal, $\frac{-x^2}{e^{2\sigma^2}}$.

Fat tailed distributions are the thickest tailed distributions. The power law is an example of this - they're the distributions with so much additional density in their tails that moments $E[X^p]$ are no longer finite.

Power Law Distributions

Pareto Distribution Pareto discovered that 20% percent of taxpayers had 80% of the income across countries in Europe. One parameter of the Pareto power law distribution is , which is known as the *tail index.* Pareto's 80-20 example corresponds to $= 1.16$. The tail index describes the behavior of density decay in the tail, as its name implies.

The strange thing about power law distributions is that, depending on the tail index , some of its moments may not exist or be infinite. There is no finite mean if < 1, and there is no finite variance if < 2. The same applies for skewness at < 3 and kurtosis when < 4, and so on. The tails get thicker as gets smaller.

Pseudo-convergence: A tail index less than 2 doesn't mean that we can't compute the sample variance of dataset. Rather, betting on the stability of the variance is unwise because this sample variance will never converge, and can in fact "spike" at any time. Furthermore, if the 4th moment (kurtosis) doesn't exist, this may imply unbearably slow convergence of the 2nd moment (variance).

The Central Limit Theorem, which is typically very useful for sums and averages, requires a finite variance, so tail indices < 2 do not obey. The assumption for the analytic Black-Scholes-Merton price for a financial option - that the random walk sum of movements converges to the Normal distribution - is also violated, so that breaks too. If the tail index is slightly over 2, it will converge to the Normal in the limit, but very slowly.

Tail events - the unlikely events of the atypically large magnitude - are the most indicative of the tail behavior. But these tail events are rare. Without a deep understanding of the underlying process which has generated these samples, it can be tough to rule out that the data was generated by a power law. In this sense, we might consider that "most" processes are fat tailed by default - or, we should at least assume they are until we have enough quantitative or qualitative data to prove otherwise.

Review of Taleb (Gelman)

Pandemic Risk Management

Non-Ergodic

Paranoia or Nothing

Taleb and collegues have som very interesting methodological remarks in the early stages of the COVID-19 outbreak:

> Clearly, we are dealing with an extreme fat-tailed process owing to an increased connectivity, which increases the spreading in a nonlinear way. Fat tailed processes have special attributes, making conventional risk-management approaches inadequate

The general (non-naive) precautionary principle delin- eates conditions where actions must be taken to reduce risk of ruin, and traditional cost-benefit analyses must not be used. These are ruin problems where, over time, exposure to tail events leads to a certain eventual extinction. While there is a very high probability for humanity surviving a single such event, over time, there is eventually zero probability of surviving repeated exposures to such events. While repeated risks can be taken by individuals with a limited life expectancy, ruin exposures must never be taken at the systemic and collective level. In technical terms, the precautionary principle applies when traditional statistical averages are invalid because risks are not ergodic.

Historically based estimates of spreading rates for pandemics in general, and for the current one in particular, underestimate the rate of spread because of the rapid increases in transportation connectivity over recent years. This means that expectations of the extent of harm are under- estimates both because events are inherently fat tailed, and because the tail is becoming fatter as connectivity increases

Estimates of the virus's reproductive ratio R_0 —the number of cases one case generates on average over the course of its infectious period in an otherwise uninfected population—are biased downwards. This property comes from fat-tailedness due to individual 'superspreader' events. Simply,R_0 is estimated from an average which takes longer to converge as it is itself a fat-tailed variable.

Norman/Bar-Yam/Taleb Note (pdf)