Predicting the Unpredictable

Ana Cascon and William F. Shadwick

August 16, 2024

1 The Geometry of Univariate Probability Distributions

Felix Klein taught us that transformation groups generate geometry and Elie Cartan taught us how to uncover that geometry through differential invariants. In 2008 we began to study the geometry of univariate probability distributions determined by the action of $\mathcal{A}$, the affine group on the line. The Cartan equivalence problem to be solved is the simplest example we know of the general case which Cartan developed for his classification of infinite Lie pseudogroups. The transformation group $\mathcal{A}$ must leave invariant a function f. When that function is the cumulative probability distribution, the exceptional constant curvature cases that we discover are the Extreme Value Theory Attractors.When the function is the Characteristic Function they are the Lévy Stable distributions and in particular in the symmetric case these are the Normal and Cauchy distributions.When the function is the Omega function of a distribution the result is a remarkable family of distributions that we describe here.[1] All of this geometry (with the exception of the explicit description of this new family of distributions) was presented as a short graduate course at Impa in June and July 2023, which we refer to as C-S 2023 where references are required in what follows.[2]

2 What's EVT?

Extreme Value Theory (EVT) is a remarkable collection of results in probability and statistics that deal with data in the 'tails' of probability distributions.[3] EVT is what you need to know if you want to make predictions about higher than 'expected' rainfall or temperatures, or the probability of a 20% drop in the US equity market by the end of the year, or how many hospital beds you need to cover admissions when a viral epidemic strikes.The subject of EVT began to be explored about a century ago. Today it has grown to a large collection of sophisticated theorems in probability and statistical techniques for applying the theory to data. Extracting information about what has not yet been observed from what has already been is the aim of statistics in general. But EVT is the specialist tool for 'seeing around a corner' or 'looking into the future'. There are many applications (for example in insurance, quality control and finance) in which its practitioners attempt to do just that.

In principle, EVT should have provided advance warning of the dramatic rise in market risk in the run up to September 2008. It might also have been expected to give accurate estimates of the growth in Covid-19 hospital admissions in 2020. But in fact, it did neither. Because it couldn't. The amount of data required to implement EVT statistics with existing technology made these applications completely impossible. Until now.

[1]Following Kronecker, one might say that the exceptional distributions were invented by God and that the rest are the work of man.

[2]Ana Cascon and William F. Shadwick, A short course on the geometry of probability distributions 2023, https://www.youtube.com/playlist?list=PLo4jXE-LdDTTOvPRBuSUcCJHyZOzpsVDn

[3]The subject began in 1928 with a discovery by Fisher and Tippett about the distributions of 'extremes'– the maxima of independent, identically distributed samples of random data as the sample size tends to infinity.They answered the question of what limiting distributions were possible for extremes. They found that if a limit existed, it could only be one of three 'types' which are now known as the Weibull, Gumbel and Fréchet distributions–the EVT attractors. Before long Richard von Mises showed that the three types were really a 1 parameter family of distributions with negative, zero and positive parameter values tracing out Fisher and Tippett's discoveries.In 1943 Gnedenko independently discovered the three types and showed that whether or not the extremes of a distribution F had one of the three possible limits depended only on the behaviour of F as it approached the upper limit of its support. The next major advance came with the work of Balkema and de Haan and, independently, Picklands in the mid 1970s with what became the statistical mainstay of the subject, Generalised Pareto distributions. This simple family of distributions had the same limits as the Extreme Value distributions and came with a natural way to 'attach' them as tails to empirical distributions of data.

3 No Longer Impossible

Accurate predictions from small data samples are not impossible any more. We have discovered a totally new approach to the problem of 'fitting' an Extreme Value distribution to data. It's completely algorithmic so it's ideally suited to automation, and the algorithm is something that even an AI can be given the ability to implement.

We've already shown (see OmegaAnalysis.com for examples from Covid-19 and from financial markets) that real world data from all sorts of sources has 'tails' that are very well predicted by the distributions we've discovered.[4] This appeared quite remarkable to us at first but the more of it we've seen the less surprising it seems. The phenomena we observe come from 'nature'. They are governed by processes that, however far from understandable through mathematical models they may be, are unlikely to be the result of some random series of Cosmic random number machines. Through the use of AI we should now be able to search out and test the predictive power in previously infeasible cases.

4 From Data Samples to Out of Sample Predictions

Our algorithm uses a tail measure that applies to any probability distribution that's bounded below and has a finite average. This function is invariant under changes of location and scale and is a natural measure of how fat the tail is. Its value is a real number $\alpha > 1$ and the smaller the value of α the fatter the tail.

Because any data sample can be thought of as an empirical distribution we can use this function to assign a 'tail parameter' to the data. The family of distributions we've discovered is parametrised by α so once this number has been calculated we just have to match it to the unique distribution in our family with the same α.Then we can start making predictions beyond or even far beyond any of the observations in our data set.

5 The Tail Measure

Suppose that F is a distribution with mean μ defined on the interval $[A, B]$ where $|A| < \infty$. Let $\omega = \int_A^\mu F(x)dx$.It is easy to verify that $2\omega = E_F(|x - \mu|)$, so ω is a natural measure of spread around the mean:the larger ω the fatter the tail.[5] Our tail measure is defined by

$$\alpha(F) = \frac{\mu - A}{\omega}. \tag{1}$$

It is obvious that α is an affine invariant, that $\alpha > 1$ and that the larger the value of α, the lower the probability of finding the random variable with distribution F far above the mean.

6 Exceptional Omega Functions and the CSS Distributions

For the 2008 Robby Gardner memorial workshop at MSRI, we solved the equivalence problem for Omega functions[6] under $\mathcal{A}$, the affine group on the line. For an Omega function Ω on $[0, \infty)$ and any point $x_0 \in [0, \infty)$ let $\Omega_{x_0} = \frac{1}{\Omega(x_0)}\Omega$. The Omega functions $\Omega_F(x) = x^\lambda$(and their mirror images on the left half line) are the exceptional cases for which $[\Omega_{x_0}] = [\Omega]$ for all $x_0 \in [0, \infty)$, where $[\Omega]$ is the equivalence class of Ω under $\mathcal{A}$.

The CSS distributions[7] on the right half line are the distributions $CSS(\lambda)$ obtained by inverting the family of Omega functions $\Omega_F(x) = x^\lambda$. They all have mean 1 and $\omega = \lambda^{-1}$, so the family is parametrised by the tail measure α.

$$\alpha(CSS(x, \lambda)) = \lambda. \tag{2}$$

The distributions are given by the following formula[8]

$$CSS(x, \lambda) = \frac{x^{2\lambda} - (\lambda + 1)x^\lambda + \lambda x^{\lambda - 1}}{(x^\lambda - 1)^2}. \tag{3}$$

[4]In technical terms, that means that the data behave as though they come from a distribution in the domain of attraction of a Fréchet distribution $F(x, \alpha)$ with $\alpha > 1$.

[5]See C-S Lecture 2 for the alternative Markov Inequality that demonstrates this fact.

[6]See C-S 2023 Lecture 2. For any univariate distribution with finite first moment its Omega function is an alternative description to the cumulative or Characteristic function which has many useful properties.

[7]Together with B. A. Shadwick, we first discovered these previously unknown distributions in 2003 by solveing the inverse problem for Omega Functions. It was only after we solved the equivalence problem that we recognised their geometric significance.

[8]The apparent singularity at $x = 1$ is removeable.

7 Predicting the (formerly) Unpredictable

It is easy to check, either by using our theorems about the geometry of Extreme Value Distributions or by applying Gnedenko's limit tests (see C-S 2023), that $CSS(x, \lambda)$ is in the domain of attraction of the Fréchet distribution $F(x, \lambda)$. Therefore we have a canonical method of fitting a natural tail model to any data set with at least two distinct points. You only need to calculate the value of α for the empirical distribution given by the data sample S

$$\alpha(S) = \frac{\mu_S - S_{min}}{\omega_S} \tag{4}$$

and the unique affine transformation that maps S_{min} to 0 and μ_S to 1. This gives a consistent, canonical method for fitting the data by a standard CSS distribution. The results of calculations made with that distribution (such as average excess on a given value) can then be mapped back to the data space by the inverse affine transformation, extending that data set beyond the observations it contains.[9]

This is reproducible by anyone with the same data–because it's completely canonical. There are no choices, and there is no ambiguity. Unlike the case of standard EVT methods, our approach, which makes its fits using *all* the data, provides transparent reproducible results. These have proved reliable even with very small data samples. What is more, our approach, now available to everyone, is very well suited to in-sample testing of its reliability and to automation.

8 Infinite in All Directions

Would you like to know if the current solar cycle is producing abnormally low numbers of sunspots? How do do the observed numbers of tropical storms compare with what you would have expected if you'd had this technology in 1901? Do the gaps between consecutive prime numbers agree with the statistical predictions you'd have made by fitting the distribution of gaps for the primes with up to a million digits? After all the primes with up to five million digits?

We've already tested our new approach on these questions and many more.

What happens when you start with a simple collection of numbers and use this method to generate a dynamical system that moves points around? What new kinds of video games can you make using those dynamics? How far from 'normal' do the responses to a post on $\mathbb{X}$ have to be to provide evidence that an account is being 'throttled'?

And the next time there's a viral epidemic, maybe you'd like to make some accurate predictions about health care demands to replace the disastrously wrong ones the specialists provide.

If you have an AI you think is trustworthy, you can ask it to come up with as many more examples as you like–even if it doesn't know where the phrase 'Infinite in all directions' comes from.

[9] As used to be well known, there's no entirely free lunch. We have to assume that the data comes from a distribution with a finite first moment. So if the underlying process were driven by a Cauchy distribution for example, our fits would systematically underestimate the fatness of the tails. But the more data we had the closer the tail measure would be to $\alpha = 1$. It is instructive to check the rate of convergence to this 'red line' limiting value of α with larger and larger data sets generated by a Cauchy distribution or by a Fréchet distribution $F(x, \lambda)$ with $\lambda \leq 1$.