The Geometry of Extreme Value Distributions–Part 1

William F. Shadwick
Omega Analysis Limited
Kyoto 28 January 2011

The goal of Extreme Value Theory (EVT) is to make statistical estimates of the likelihood and severity of 'random' events which have not been observed, based on observed data–for example:

The level of flooding that might be expected in north-eastern Australia 1 year in 50.

The loss on the S&P 500 Index that should only be exceeded 1 day in 100 and the average of losses in excess of this level.

This is a very ambitious goal but two approaches to this sort of problem are very tractable due to remarkable limit theorems analogous to the Central Limit Theorem.

These results can be unified, explained and extended in terms of differential invariants for the affine group on the line.

Before describing that, I want to illustrate the impact of good estimates of extremes in a context that everyone is familiar with because of recent events in financial markets around the world.

Daily returns (percentage gain or loss) on an investment (for instance in an equity market index fund) appear to be random.

If we assume that they are independent, identically distributed random variables, we can use EVT to make estimates of the risk of large losses.

This model is 'wrong', like all models (even in physics). What matters is the accuracy of its predictions.

Imagine that we have an accurate way of estimating 'risk' for both buyers and sellers in a stock market.

Suppose that the average daily return is positive. We should expect that the risk to a buyer will be less than the risk to a seller. While this remains the case the market should continue to rise.

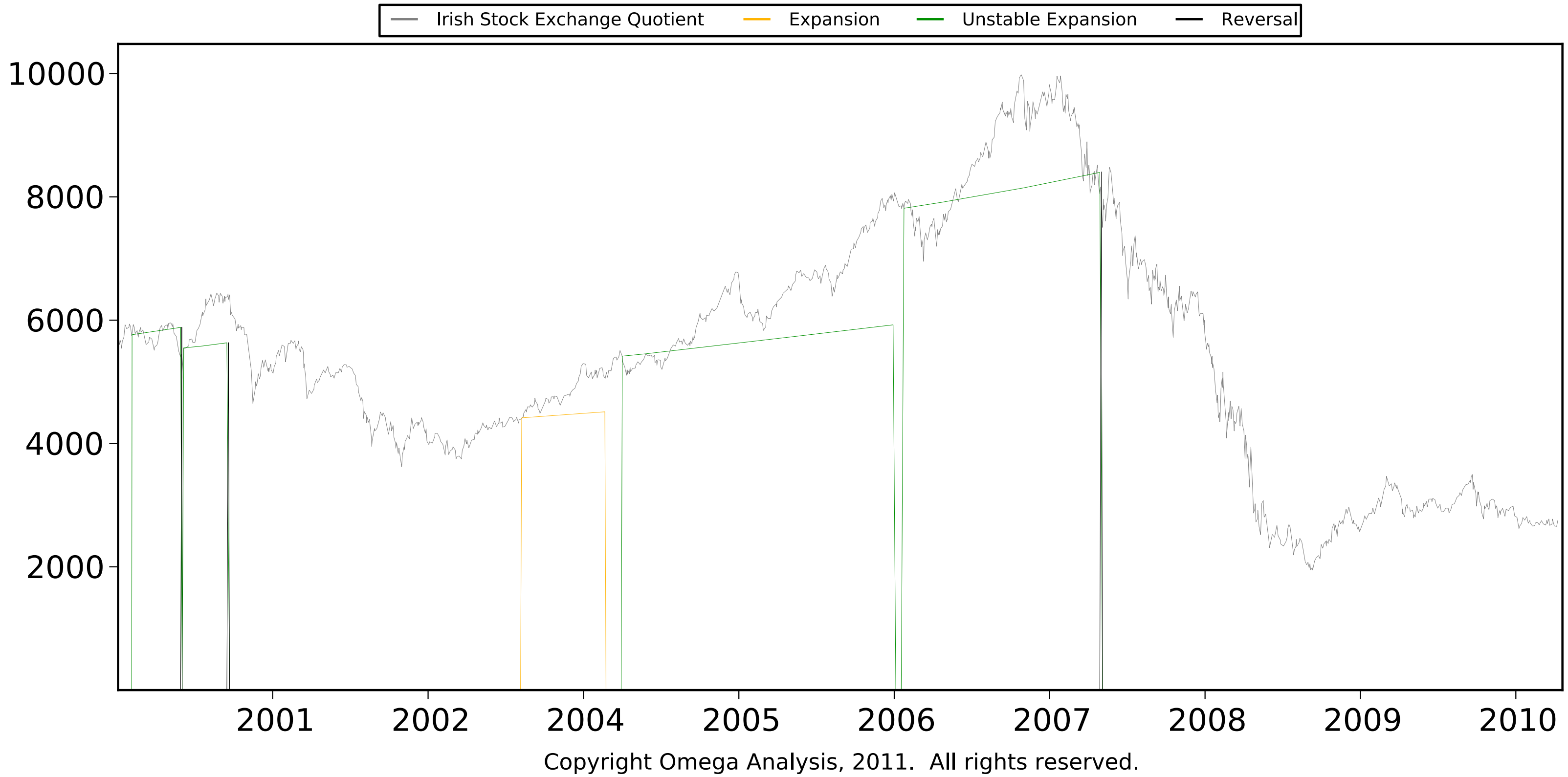
But if the buyer experiences higher risk than the seller, selling will gain momentum and the market may go from expansion to contraction.

Now suppose the buyer's risk is greater than the seller's risk but the market does not change direction and the expansion continues.

This is an unstable expansion–a potential market bubble–which may end with a crash as the overdue contraction arrives.

If this simple model of 'market motion' were correct, we should see the evidence in historic data.

Here's what this model predicted in the case of Ireland's major stock market index from January 2000 to December 2010.



Predicted Expansions for the Irish Stock Index ISEQ

Omega Analysis

But if expansions work according to this model, contractions must as well.

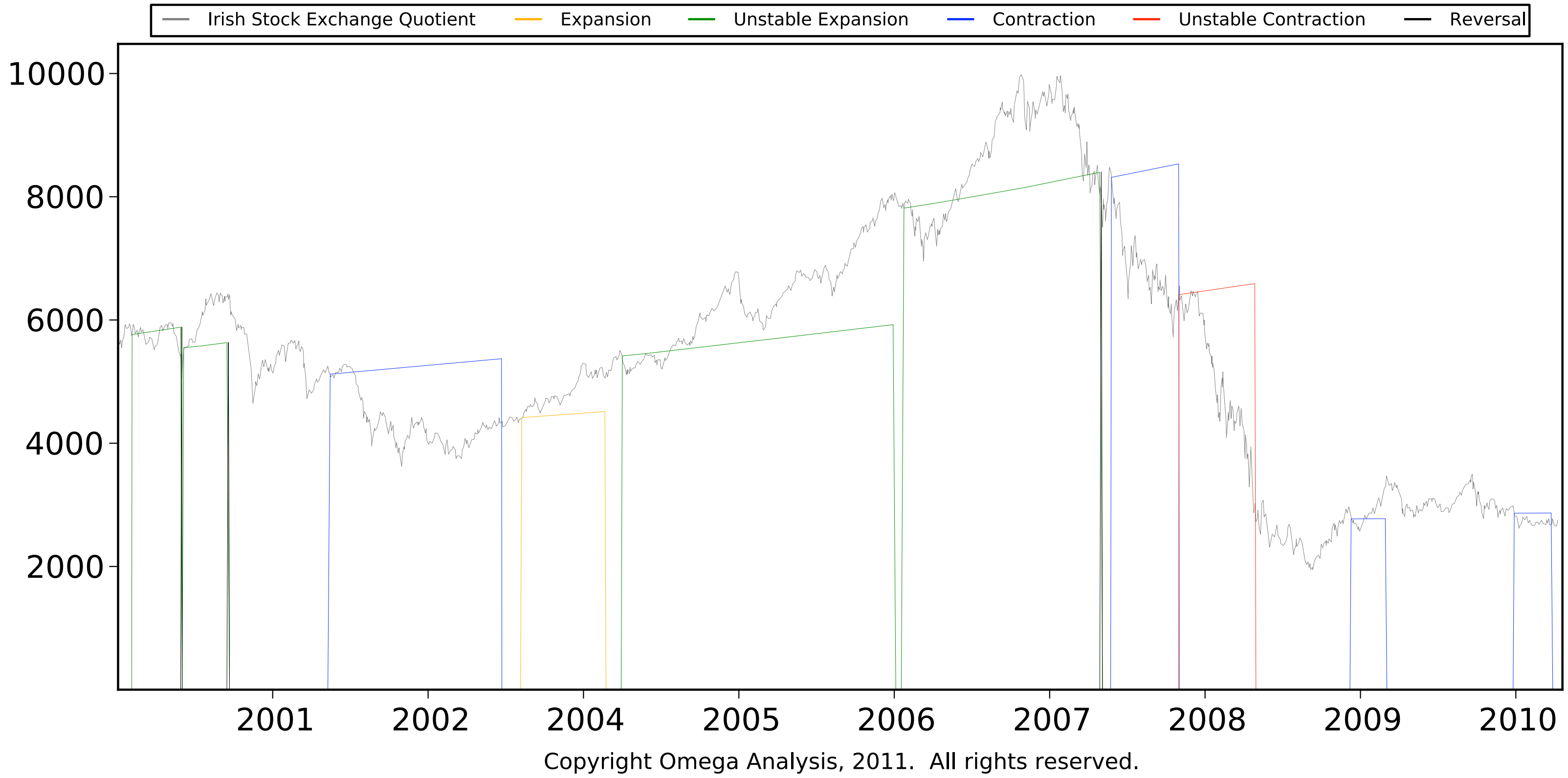
When the average return is negative, a seller's risk should be less than a buyer's.

In such a case we should see stable contraction.

When the return is negative but the seller's risk is greater than the buyer's, we should see a change in market direction.

If such a change does not occur, we have an unstable contraction–a potential 'anti-bubble' of panic selling.

Here's what this model predicted in the case of Ireland–as the Irish economy was subjected to the wrong interest rates and a disastrous bank bail-out.



Expansion and Contraction Predictions for ISEQ

Omega Analysis

First Approach to Extremes: Sample Maxima

Suppose we make independent draws of samples of size N from a fixed probability distribution.

The maximum of each sample is another random variable. It can be distributed in only one of three ways as N tends to ∞.

This is the result of a remarkable limit theorem first stated by Fisher and Tippett in 1928.

Omega Analysis

We consider Probability Distributions

$$F : [\alpha(F), \omega(F)] \rightarrow [0, 1] \qquad (1)$$

where F is C^2 and $F^{'} > 0$.

We may have $\alpha(F)$=-∞ and/or $\omega(F)$=∞.

First Approach to Extremes: Sample Maxima

Let $X_1, \ldots, X_N$ be a sample of N independent, identically distributed random variables with distribution function F. Let X_{Max} be the sample maximum.

If $X_{Max} < r$ then all of the sample draws must be less than r and the probability of this is $F^N(r)$.

Thus the distribution of X_{Max} is F^N.

In 1928 Fisher and Tippett addressed the question:

Does there exist a sequence of ‘location-scale’ transformations $x \to a_N x + b_N$ and a distribution G such that

$$F^N(a_N x + b_N) \to G(x) \tag{2}$$

as N tends to ∞?

Intuition: Any such G should be the distribution of its own extremes. Fisher and Tippett proved that there are only three families of distributions with this 'stability property'.

$$\Phi(x,\alpha) = e^{-(-x)^{\alpha}}, \; x \in (-\infty, 0], \; \alpha > 0 \tag{3}$$

$$\Psi(x,\alpha) = e^{\frac{-1}{x^{\alpha}}}, \; x \in [0, \infty), \; \alpha > 0 \tag{4}$$

$$\Lambda(x) = e^{-e^{-x}}, \; x \in (-\infty, \infty). \tag{5}$$

(Weibull, Fréchet and Gumbel distributions respectively)

Omega Analysis

Fisher and Tippett showed that the sample maxima limit for the Normal distribution was Gumbel type. The convergence was extremely slow. They found that the Weibull type $\Phi(x, \alpha)$ was a better approximation, even though it was not the ultimate limit.

We shall return to the question of this ‘Penultimate Approximation’.

They gave no method for determining for a given distribution F whether or not it had a limit and if so, what the limit was.

It took 15 years to fill this gap.

Omega Analysis

In 1943 Gnedenko provided an independent derivation of the ‘three types’ theorem as well as necessary and sufficient conditions for convergence (without the C^2 assumption).

He showed that Domains of Attraction (the collection of distributions which converges to a given type) depended only on the limiting shape of the distribution as $x \to \omega(F)$

Gnedenko's necessary and sufficient condition for F to be in the domain of attraction of the Fréchet distribution $F(x, \alpha)$ is

$$\lim_{x \to \infty} \frac{1 - F(x)}{1 - F(tx)} = t^{\alpha} \qquad (6)$$

for all $t > 0$.

For the Weibull distribution $W(x, \alpha)$ $\omega(F) = B$ must be finite and

$$\lim_{x \to 0_-} \frac{1 - F(tx + B)}{1 - F(x + B)} = t^{\alpha} \qquad (7)$$

for all $t > 0$.

Gnedenko gave a variety of necessary and sufficient conditions for F to be in the domain of attraction of the Gumbel distribution.

He was not satisfied that any of them were either definitive or practical.

We provide a remedy for this in the C^2 case.

The ‘three types’ of distributions Fisher and Tippett discovered are really a one parameter family (up to affine transformations).

They are more conveniently denoted on variable domains depending on α as follows:

$$E_\alpha(x) = exp(\frac{-1}{(1+\frac{x}{\alpha})^\alpha}), \quad \alpha \neq 0 \tag{8}$$

This is the Weibull type, defined on $[-\infty, -\alpha]$ when $\alpha < 0$ and is the Fréchet type defined on $[-\alpha, \infty,]$ when $\alpha > 0$.

As $|\alpha| \to \infty$, both types have the Gumbel distribution $E_\infty = e^{-e^{-x}}$ as their limit.

Omega Analysis

Second Approach to Extremes: 'Peaks over threshold'

Does the distribution of a random variable X, conditional on X exceeding a threshold T, tend to a limit as T tends to $\omega(F)$ up to location scale transformations?

In this case we say F has a PoT limit.

In 1975 Picklands showed that there was a significant connection between PoT limits and Domains of Attraction of extreme value distributions.

F is in the domain of attraction of E_α if and only if the PoT limit of F is,up to a location-scale transformation, a Generalized Pareto distribution G_α as $x \to \omega(F)$ where

$$G_\alpha(x) = 1 - \frac{1}{(1 + \frac{x}{\alpha})^\alpha}, \quad \alpha \neq 0 \tag{9}$$

and

$$G_\infty(x) = 1 - e^{-x} \tag{10}$$

G_α is defined on $[0, -\alpha]$ when $\alpha < 0$ and on $[-\alpha, \infty]$ when $\alpha > 0$. As $|\alpha| \to \infty$, both types have the exponential distribution $G_\infty = 1 - e^{-x}$ as their limit.

But now we have a mystery.

Where do these distributions come from?

Everything in the Domain of Attraction of E_α is converging to everything else in that Domain of Attraction.

What, if anything, is special about the Generalized Pareto distributions?

The resolution of this mystery will be left for Part 2. It has significant implications for statistical modelling of extremes.

The Geometry of Extreme Value Distributions is the information is invariant under the proper affine group on the line $\mathcal{A}$.

A distribution F is in the domain of attraction of G if and only if there is a 1-parameter family $g_\lambda \in \mathcal{A}$ such that

$$[\lim_{\lambda\to\infty} F^\lambda \circ g_\lambda] = [G], \tag{11}$$

where $[G]$ denotes the $\mathcal{A}$-equivalence class of G.

It will be convenient to replace F by $I = log(F)$ which is particularly well suited to studying the equivalence class of powers of F.

The equivalence problem for an invariant function I under the proper affine group is a simple one.

We will show that equivalence classes are determined by one functional relation $J = H(I)$ where

$$J = \frac{I_{xx}}{I_x^2}. \tag{12}$$

The Equivalence Problem

Coframe for the affine group

$$\theta_1 = y dx \tag{13}$$

$$\theta_2 = \frac{dy}{y} \tag{14}$$

Coframe adapted to the (invariant) $I = log(F)$

$$\omega_1 = dI = \frac{F_x}{F} dx \tag{15}$$

$$\omega_2 = \frac{dy}{y} \tag{16}$$

We have

$$\omega_1 = K\theta_1 \tag{17}$$

where

$$K = \frac{F_x}{yF} \tag{18}$$

Any diffeomorphism that preserves both co-frames also preserves K so we have a second functionally independent invariant.

Now

$$dI = \omega_1 \tag{19}$$

and

$$\frac{dK}{K} = J\omega_1 - \omega_2, \tag{20}$$

where

$$J = \frac{F_{xx}F}{F_x^2} - 1 = \frac{I_{xx}}{I_x^2}. \tag{21}$$

But J depends only on x so it must be a function of I.

It follows that the remaining information is in the functional dependence of J on I.

This dependence relation determines equivalence classes of F.

Suppose that $[F^\lambda] = [F]$ for all $\lambda > 0$.

Because the relation $J = H(I)$ determines equivalence classes we must have $J_{F^\lambda} = H(I_{F^\lambda})$

From the definitions of I_F and J_F we have $I_{F^\lambda} = \lambda I_F$ and $J_{F^\lambda} = \frac{1}{\lambda} J_F$ so $\frac{1}{\lambda} H(I) = H(\lambda I)$ for all $\lambda > 0$.

Differentiating with respect to λ and evaluating at $\lambda = 1$ shows that $[F^\lambda] = [F]$ if and only if there is a constant c such that

$$H(I) = \frac{c}{I} \tag{22}$$

Thus $[F^\lambda] = [F]$ if and only if $J = \frac{c}{I}$ for constant c.

Each value of c determines a distinct equivalence class.

It is easy to see that the Extreme Value distributions provide normal forms for these equivalence classes.

The equivalence class of E_α is given by $c = 1 + \frac{1}{\alpha}$ for $\alpha \neq 0$ and E_∞ is given by $c = 1$.

Omega Analysis

The Geometry of Domains of Attraction

The invariants I and J also provide us with a new affine invariant of the 1-parameter family of equivalence classes $[F^\lambda]$ because

$$I_{F^\lambda} J_{F^\lambda} = I_F J_F. \tag{23}$$

for all $\lambda \in (0, \infty)$.

Thus the product of these two invariants is independent of λ. This result is the key to our characterisation of domains of attraction.

Theorem 1: If F is a C^2 distribution on $[\alpha(F), \omega(F)]$, a necessary and sufficient condition for F to be in the domain of attraction of E_α is that

$$\lim_{x \to \omega(F)_-} I_F J_F = 1 + \frac{1}{\alpha}. \tag{24}$$

A necessary and sufficient condition for F to be in the domain of attraction of E_∞ is that

$$\lim_{x \to \omega(F)_-} I_F J_F = 1. \tag{25}$$

Omega Analysis

It is easy to check that all of the standard probability distributions belong to the Domain of Attraction of the Extreme Value distributions E_α and E_∞.

It is also easy to verify that for each α the Generalized Pareto distribution G_α is in the Domain of Attraction of E_α and that G_∞ is in the Domain of Attraction of E_∞.

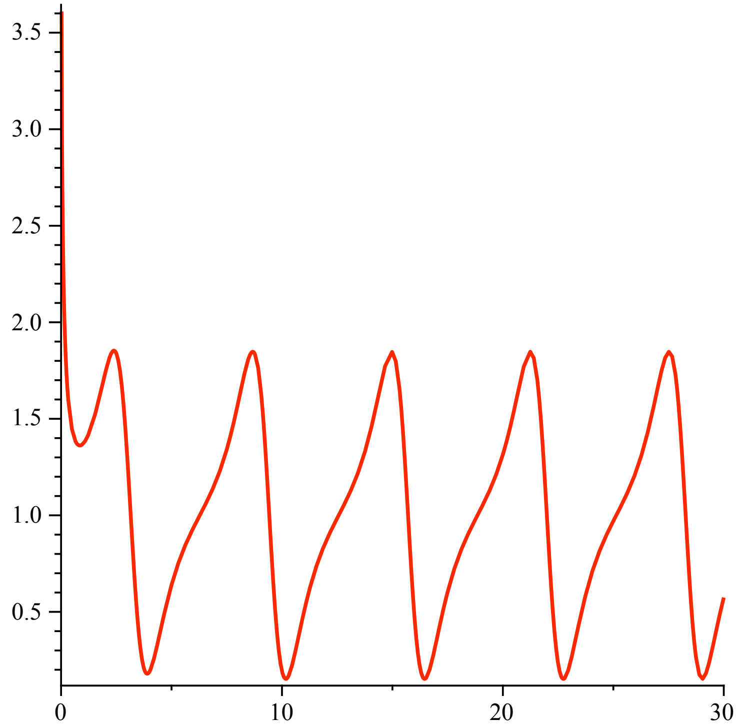
The following example, due to Von Mises, shows that smooth distributions need not have an Extreme Value limit.

If F is defined on $[0, \infty]$ by

$$F(x) = 1 - exp(-x - \frac{sin(x)}{2}). \qquad (26)$$

then $I_F J_F$ has no limit as $x \to \infty$.

Omega Analysis



$I_F J_F$ has no limit as $x \to \infty$

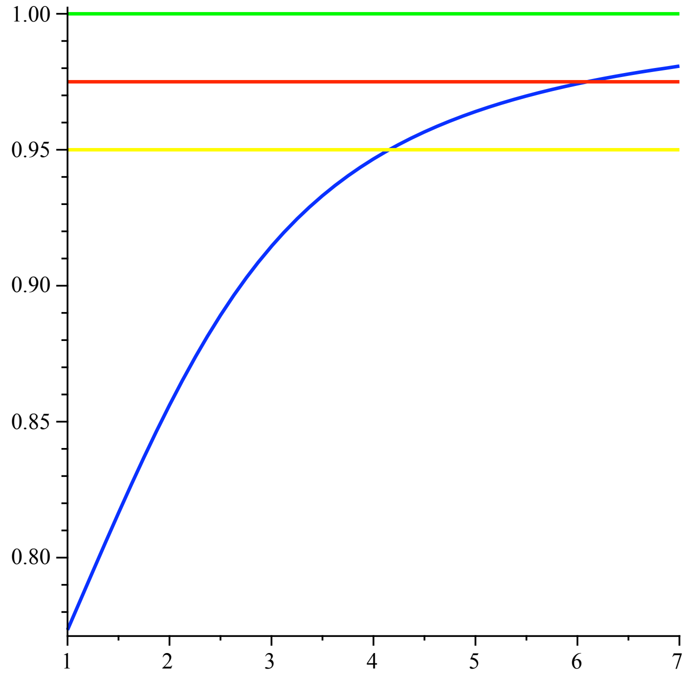
Omega Analysis

The ‘Penultimate Approximation’ for Normal Extremes

The Normal distribution’s IJ limit is 1 but it converges to 1 from below and very slowly. This illustrates the utility of Fisher and Tippett’s Penultimate Approximation.

For example, E_{-20} has $IJ = 1 - \frac{1}{20} = 0.95$. The Normal IJ remains closer to this value than it is to 1 until x is over 6 standard deviations above the mean.

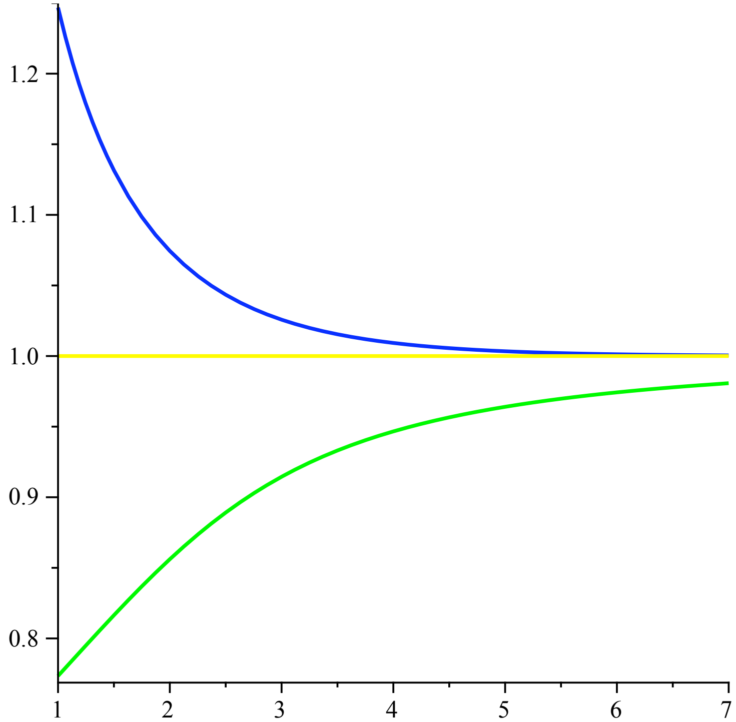
At this point the Normal distribution differs from 1 only in the 9^{th} decimal place. In almost any statistical application this difference is irrelevant.



E_{-20} is a better approximation than E_∞ for all $x < 6.09$

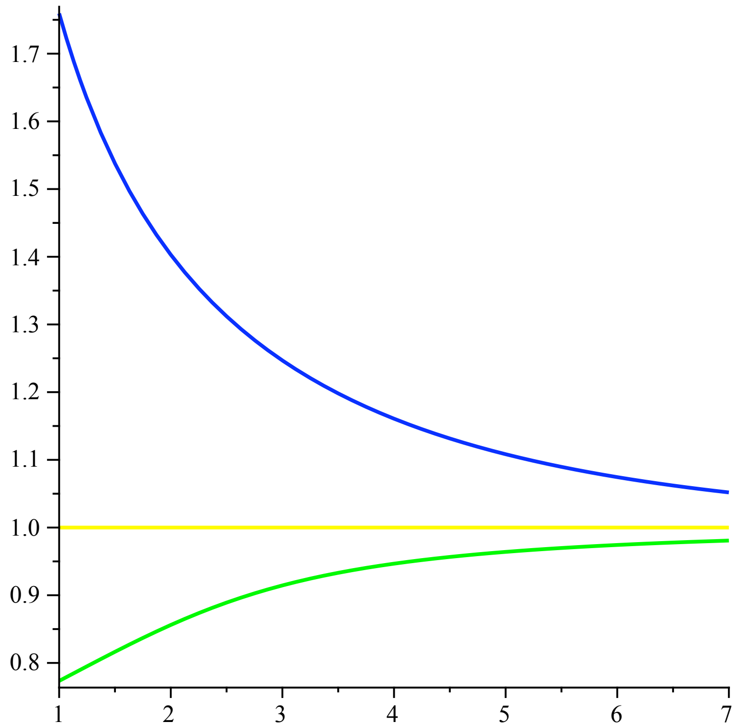
For $x > 6.09$ the Normal distribution differs from 1 by only one part in 10^9.

We can see that the Normal distribution converges to E_∞ much more slowly than G_∞ does.



IJ_{Normal} in green, IJ_{G_∞} in blue

Or can we? Here's the picture for another representative of $[G_\infty]$



IJ_{Normal} in green, IJ_G in blue

Does it make sense to compare the value of IJ at the same point x for these distributions?

We'll show in Part 2 that the invariant IJ provides us with an intrinsic means of making this comparison and illustrate the importance of this for statistical estimation.

We'll also reveal the geometry behind the relation between the Sample Maxima and PoT approaches.

But that's for Monday...