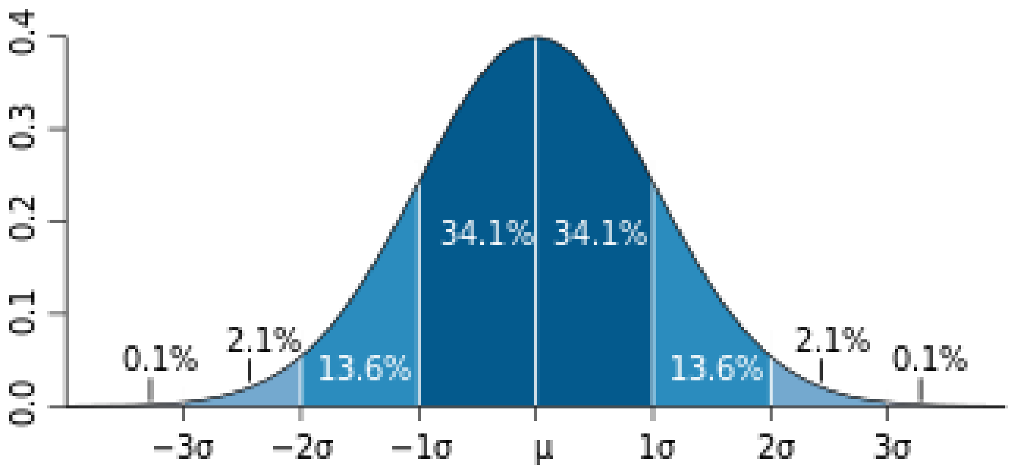
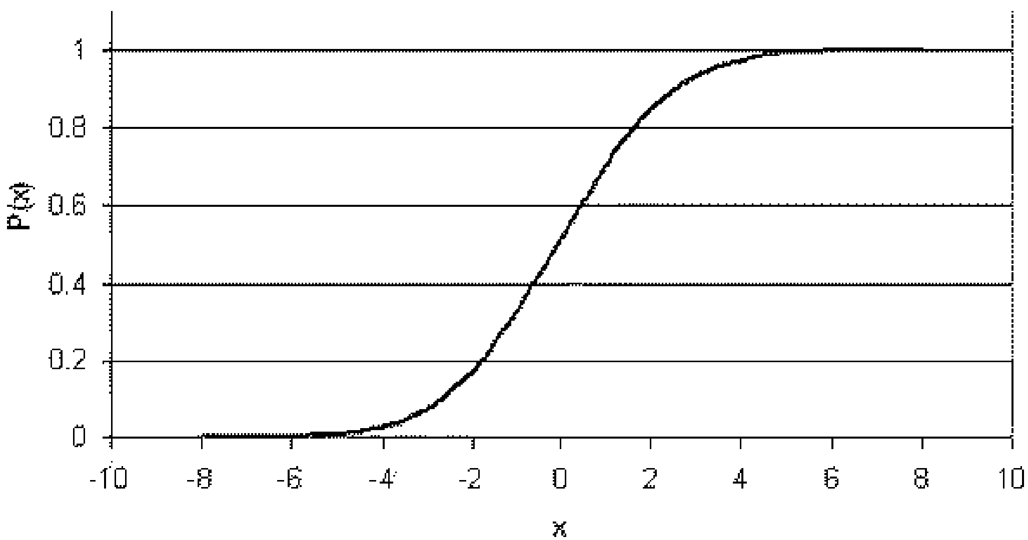
Statistics and clustering of extreme space climate events

Alexander Ruzmaikin

Statistics of Large Numbers: LLN, CLT, Gaussian Distribution



PDF



CDF

Stable distribution

Law of Small Numbers

Statistics of Extremes

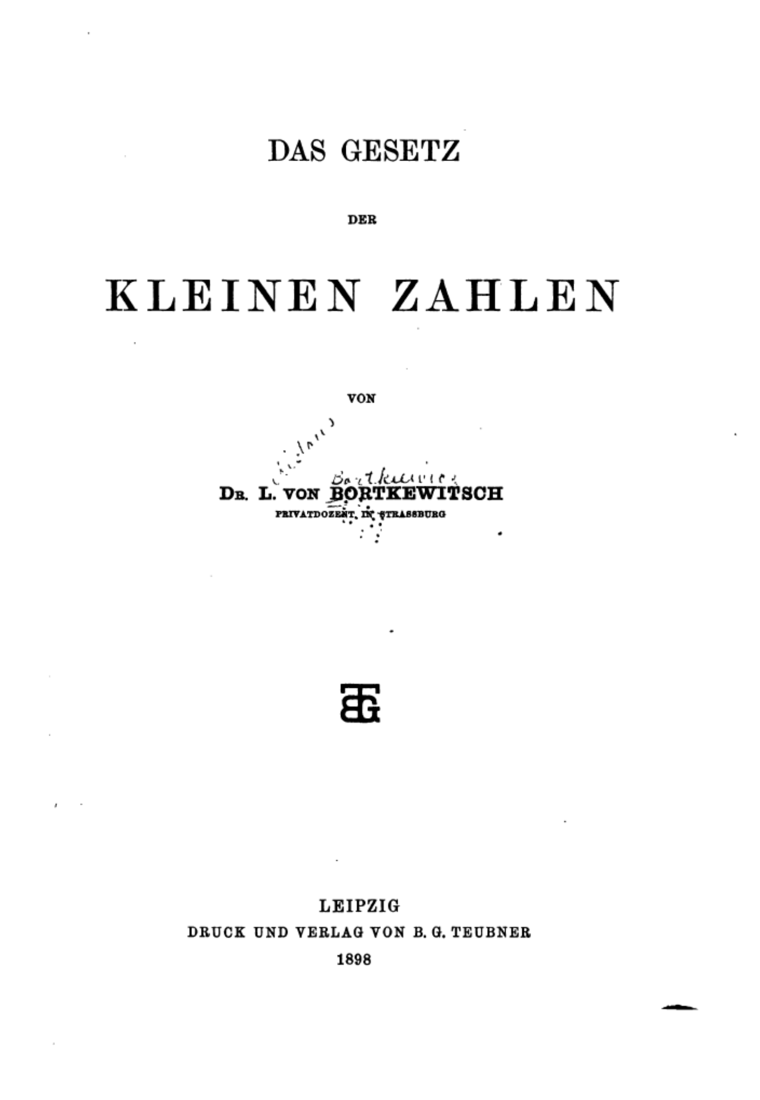
The Beginning: Nikolaus Bernoulli's problem of 1709

If n men of equal age die within t years, what is the expected life of the last man?

Find the solution

Statistics of Small Numbers?!

Ladislav von Bortkiewich
(1868-1931)



DAS GESETZ

DER

KLEINEN ZAHLEN

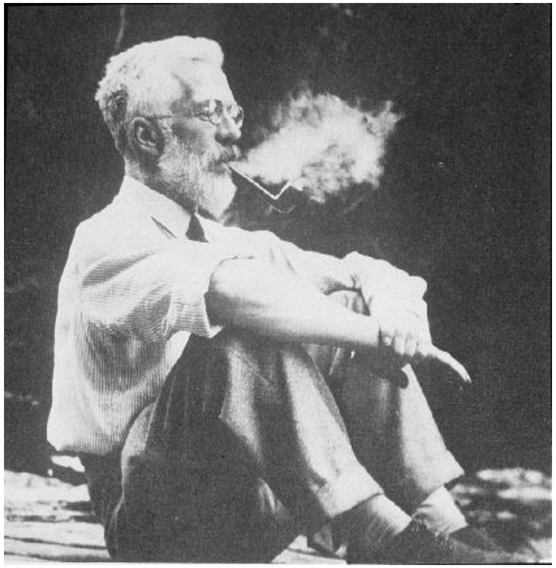
VON

DR. L. VON BORTKEWITSCH
PRIVATDOZENT IN STRASSBURG

LEIPZIG
DRUCK UND VERLAG VON B. G. TEUBNER
1898

Prussian army horse-kick problem

Fisher-Tippett-Gnedenko Theorem



Ronald Fisher
(1890-1962)
Statistical Methods for Research Workers

Leonard Tippett (1902-1985)
British Cotton Industry random number table (now random number generators)

Boris Gnedenko
(1912-1995)
Textile Institute, Ivanovo

The FTG Theorem

If e_1, e_2, …,e_n, …are iid random events and M_n = max(e_1, e_2, …,e_n)

Then Prob($M_n \leq x$), as n ➔ is

$$\text{GEV} = \exp\{-(1 + \gamma(x-\mu)/\sigma)^{-1/\gamma}\}, \quad 1 + \gamma(x-\mu)/\sigma) > 0,$$

where σ is scale, μ is location and γ is shape

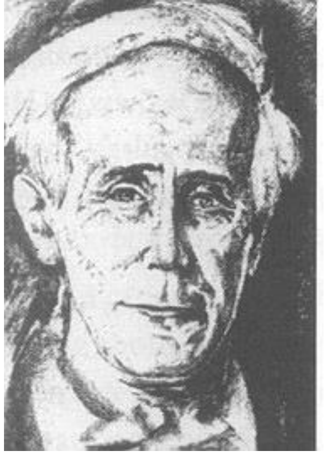
Predecessors

Dependent on $\gamma = 1/\alpha$, $y = (x-\mu)/\sigma$

$G = \exp\{-\exp(-y)\}$, $\gamma = 0$

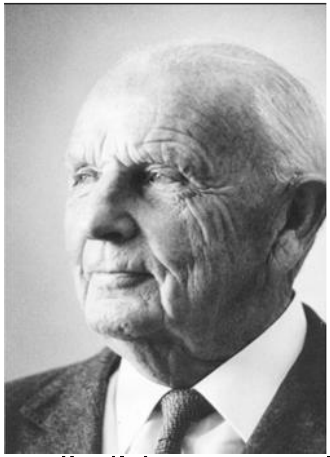
$F = \exp(-\ y^{\square\alpha}) \approx 1-\ y^{\square\alpha}$, $\gamma > 0$, $y > 0$ $(F = 0,\ y > 0)$

$W = \exp\{-(-y^{\alpha})\}$, $\gamma < 0$, $y < 0$ $(W = 1,\ y \geq 0)$



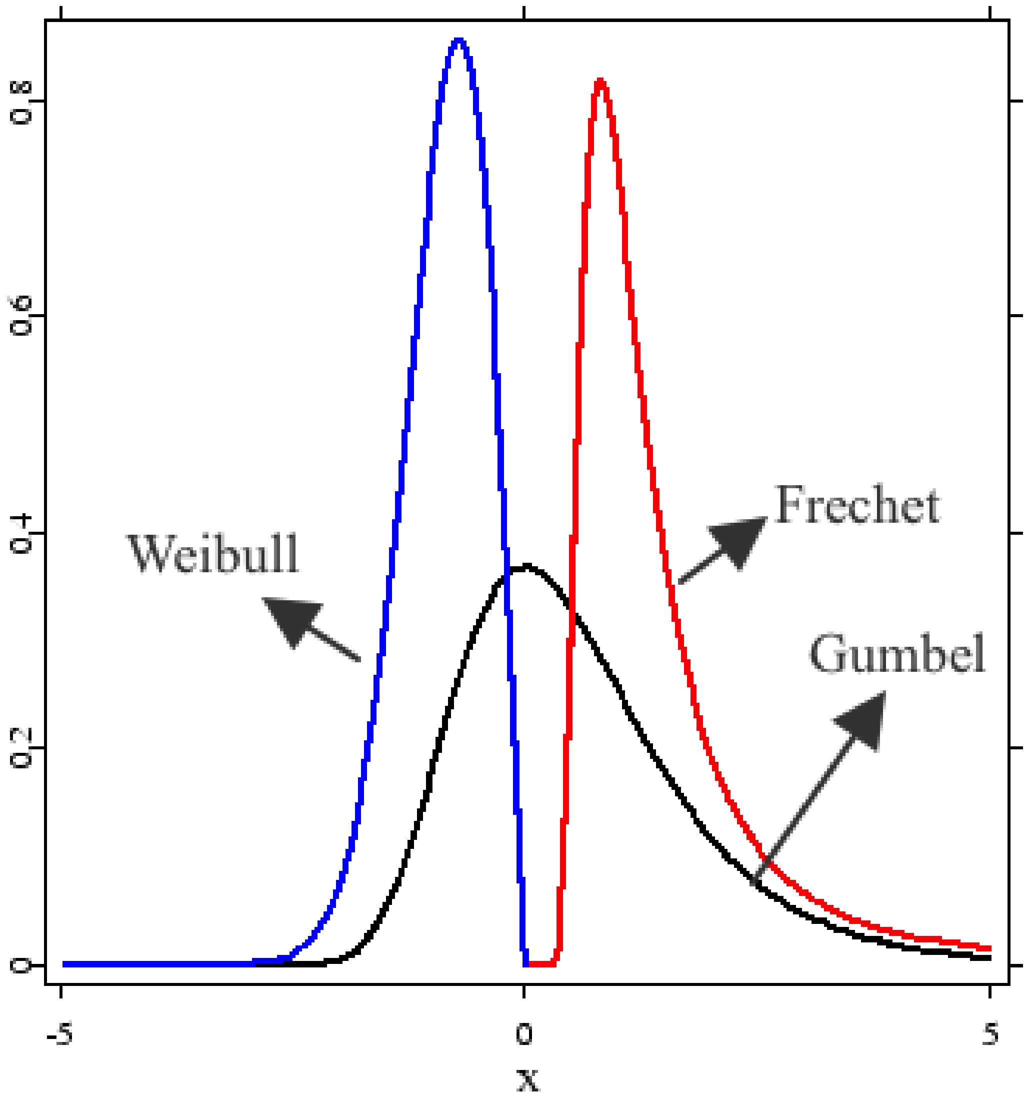
Gumbel (1891-1966)

Frechet(1878-1973)



Weibull (1887-1979)

Probability Densities (PDFs)



The Theorem: Details

$M_n = \max(e_1, e_2, \ldots, e_n)$, $Prob(e_i < x) = F(x)$

$Prob(M_n \leq x) = Prob(e_1 \leq x) \ldots Prob(e_n \leq x) = F(x)^n$

Max Stability: $F(x)^n = F(a_n x + b_n)$, as $n \to$ (Frechet, 1927)

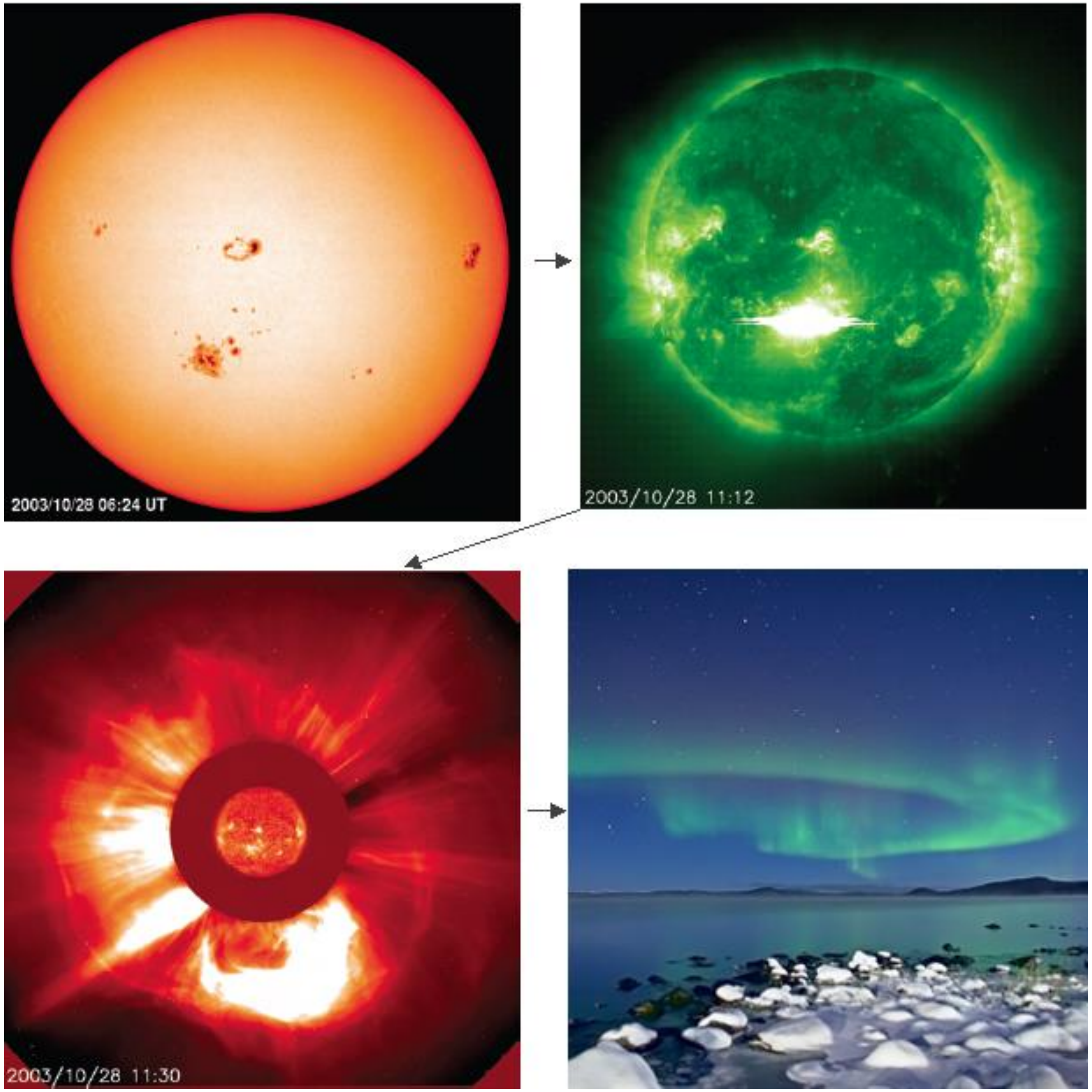
Gumbel: $a_n = 1$, $b_n = -\sigma \log(n)$

Frechet: $a_n = n^{-1/\alpha}$, $b_n = m(1 - n^{-1/\alpha})$

(Gnedenko, 1943)

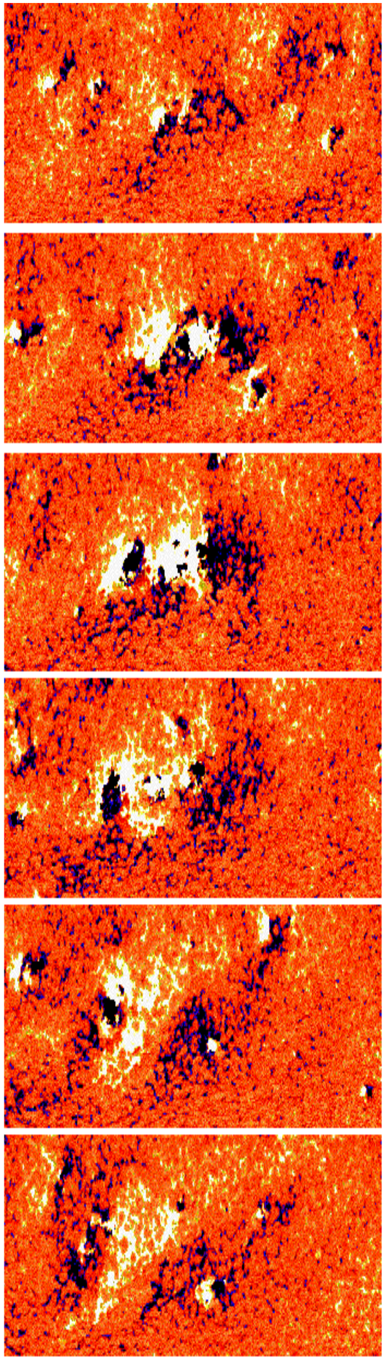
Application to
Solar Coronal Mass Ejections

Coronal Mass Ejections (CMEs)

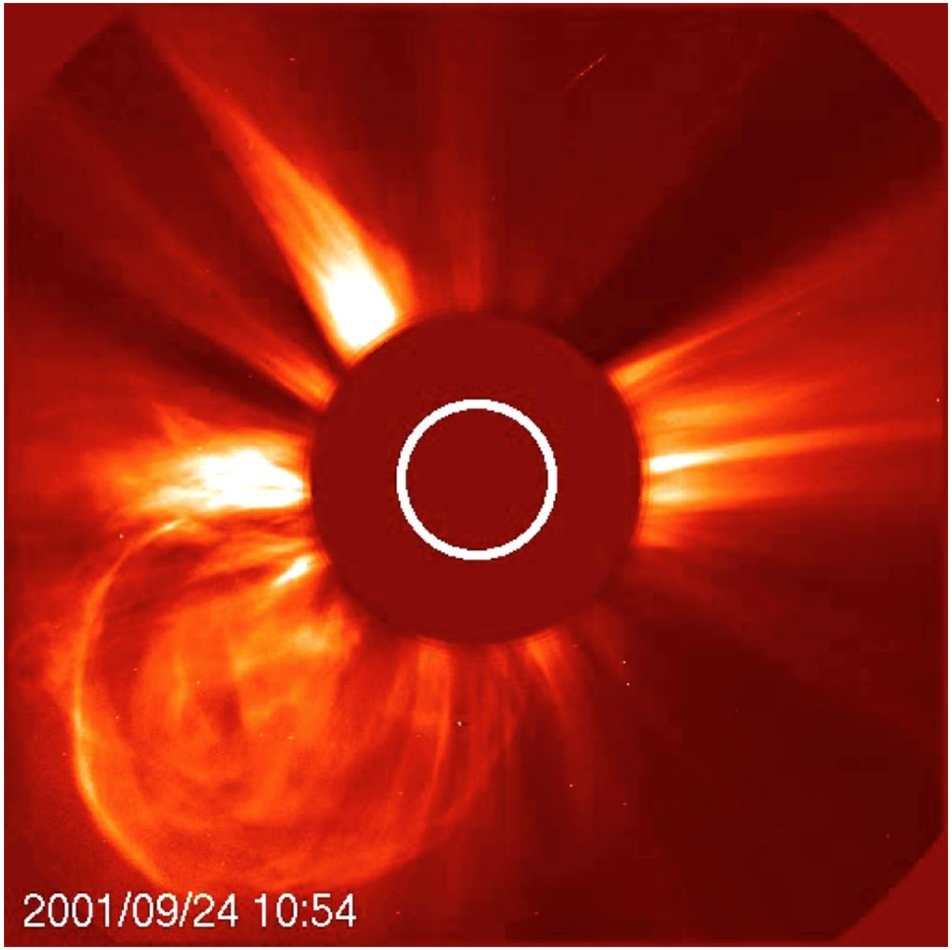


CMEs are drivers of Space Weather since they generate energetic particles and disturb the Earth magnetosphere triggering geomagnetic storms

Fast CMEs

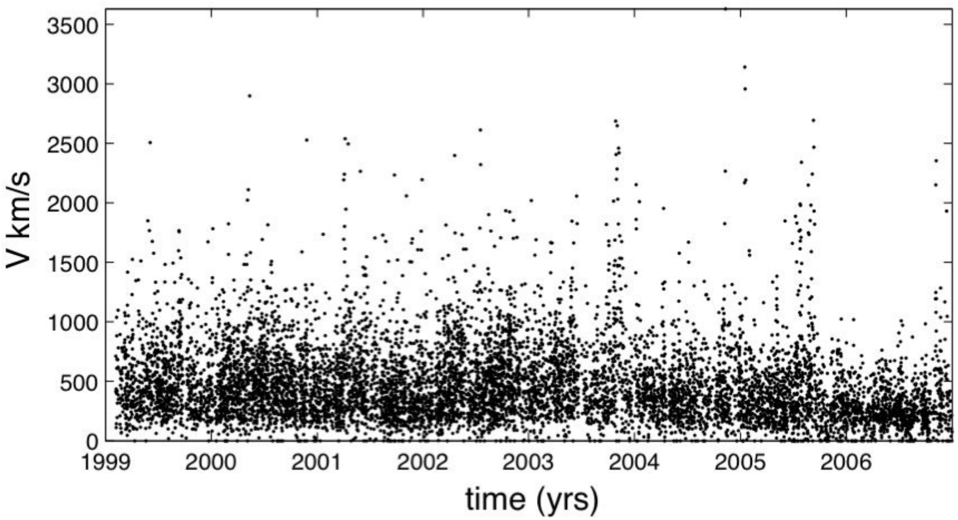


6 rotations, S. Hemisphere (180-360º)

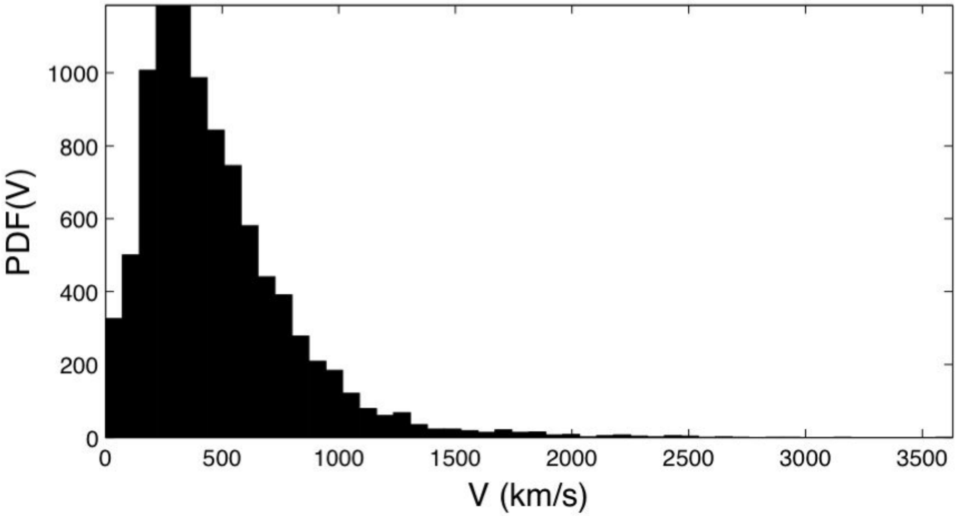


Observed Sep 24, 2001.
Speed 2,508 km/sec.

Distribution Function of CME speeds



9,408 CMEs detected by SOHO LASCO in 1999-2006



Non-Gaussian PDF

Vmean = 472 km/s

Extremes:

18% V > 700 km/s

6.2% V > 1,000 km/s

0.5% V > 2,000 km/s.

We can directly fit maxima to GEV distribution

Great!

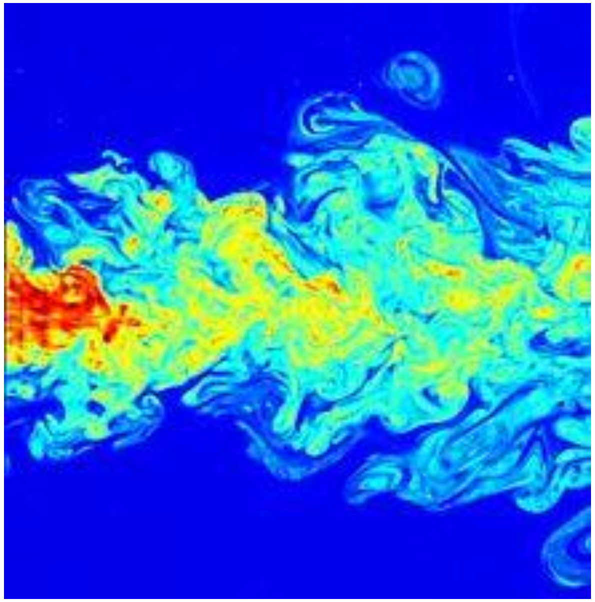
But there are no mathematically justified procedure
for curve fitting.

It depends on:

data sample,
adjustable parameters,
skill of a researcher.

Scaling Approach

A. N. Kolmogorov



$$\delta \mathbf{u}(r) = \mathbf{u}(\mathbf{x}+\mathbf{r}) - \mathbf{u}(\mathbf{x})$$

$$\langle [\delta \mathbf{u}(r)]^n \rangle = C_n \varepsilon^{n/3} r^{n/3}$$

Scaling Approach to Extremes

Stoev et al., 2006

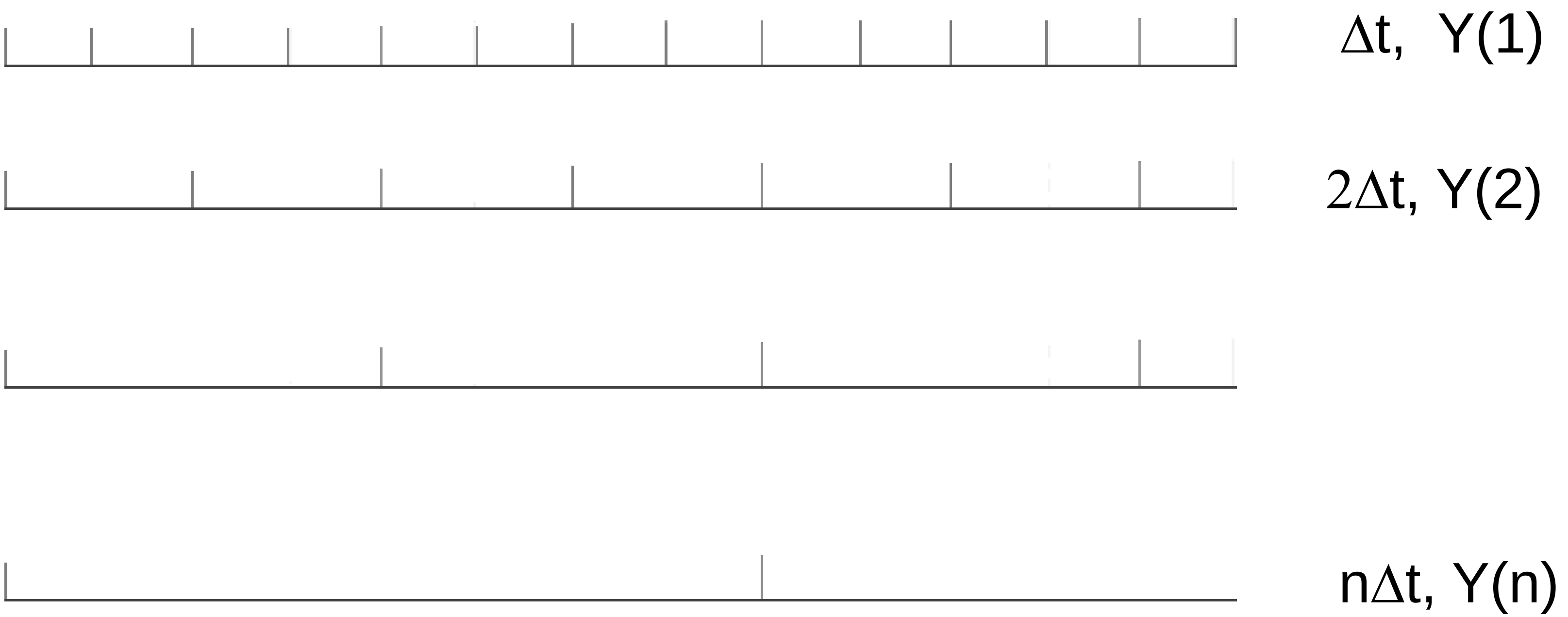
Consider time series of CME speeds, V(t).

Divide time axis into progressively increasing blocks: $\Delta t = 2^j$, j =1,2 ,3,...

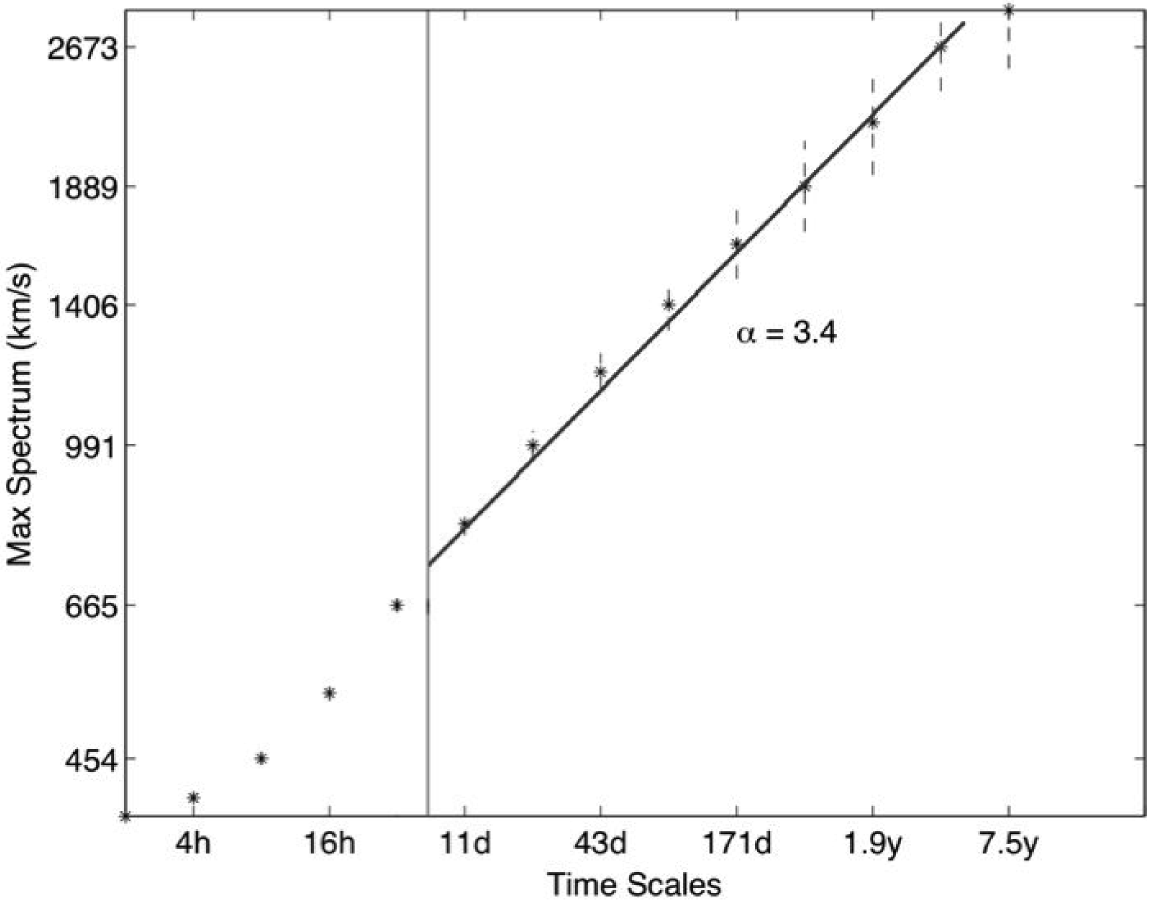
Find maxima M = max V(j) at each time scale.

Take log and average over number of intervals (k):

$$Y(j) = N^{-1} \sum_k \log_2 M(j,k) \quad \text{-- Max Spectrum}$$

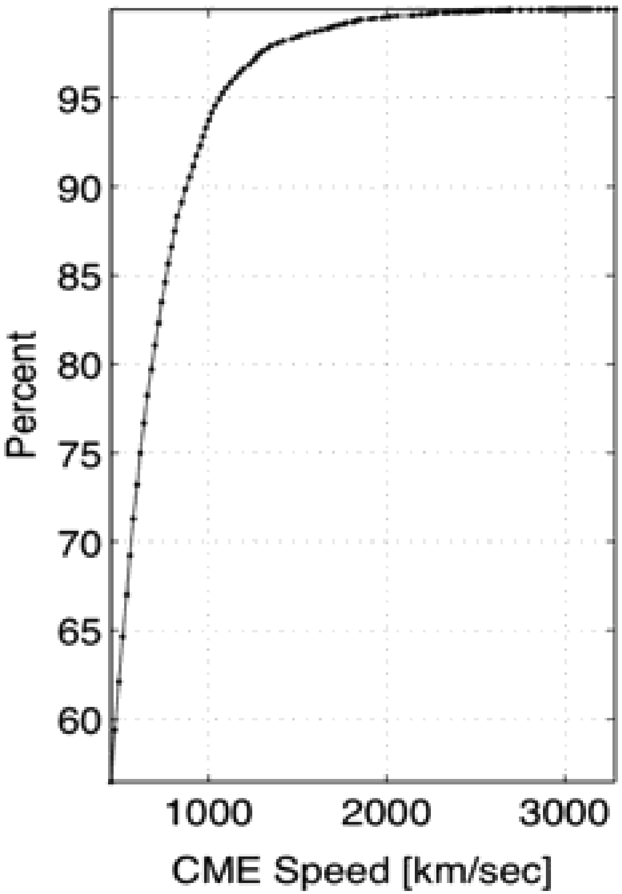


Max Spectrum



Range of speeds limited by linear fit gives a definition of “fast” CMEs.

CDF Tail of Fast CMEs



The slope of Max Spectrum ($1/\alpha$) is a heavy–tail exponent of extreme value probability density

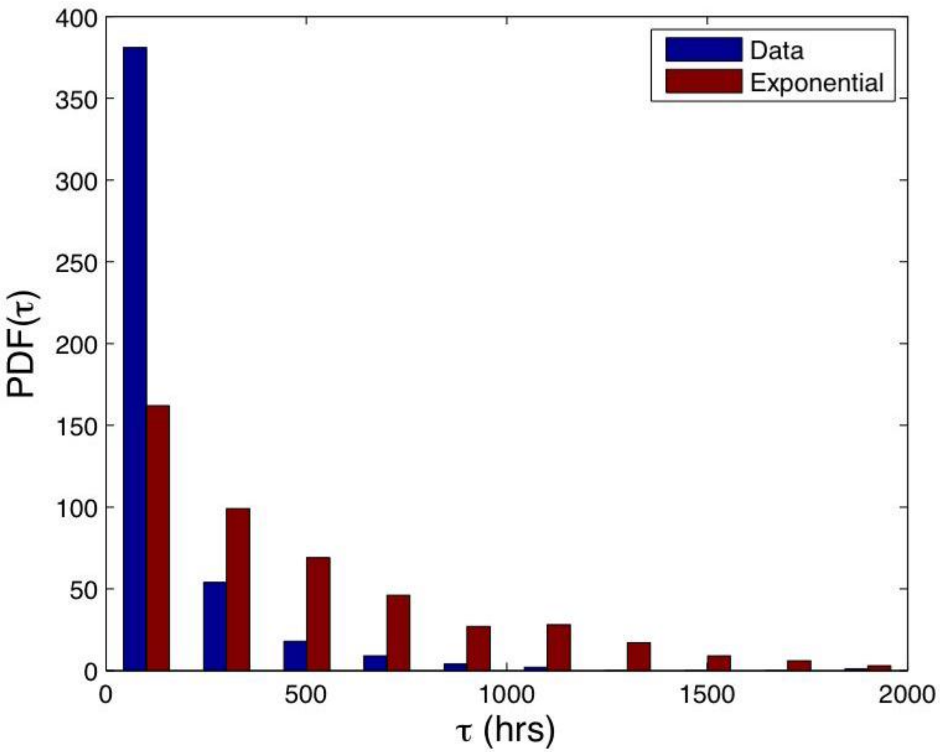
$$P = \exp(-Cx^{-\alpha}) \sim 1 - x^{-\alpha}, \text{ as } x \to \infty.$$

Cumulative distribution function.
Its high-speed tail is $1 - V^{-\alpha}$.

Stoev et al. 2006

Frequency of Occurrence of Fast CMEs

For a pure random (Poisson) process the times between events $\tau = t(i+1) - t(i)$ are independent and exponentially distributed: $\exp(-\tau/\tau_0)$.
Observed fast CMEs (blue) are correlated (clustered) in time.

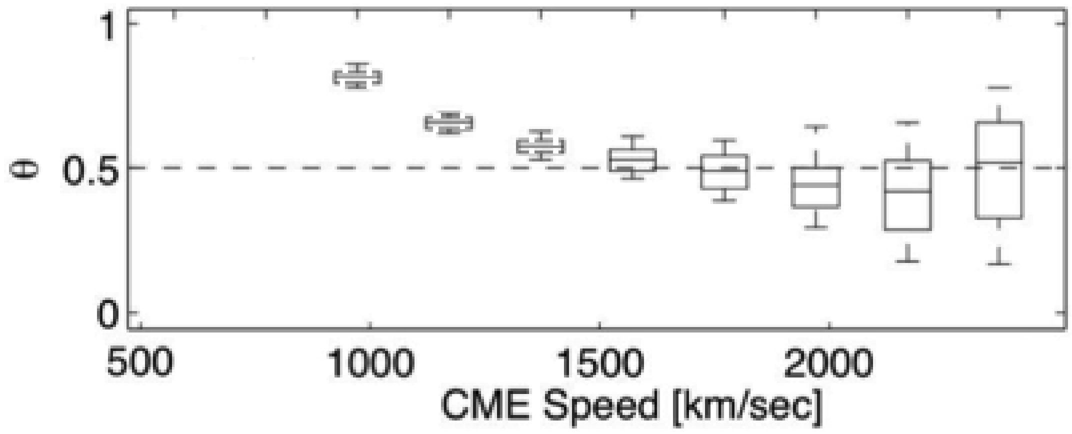


CMEs with V > 1000km/s

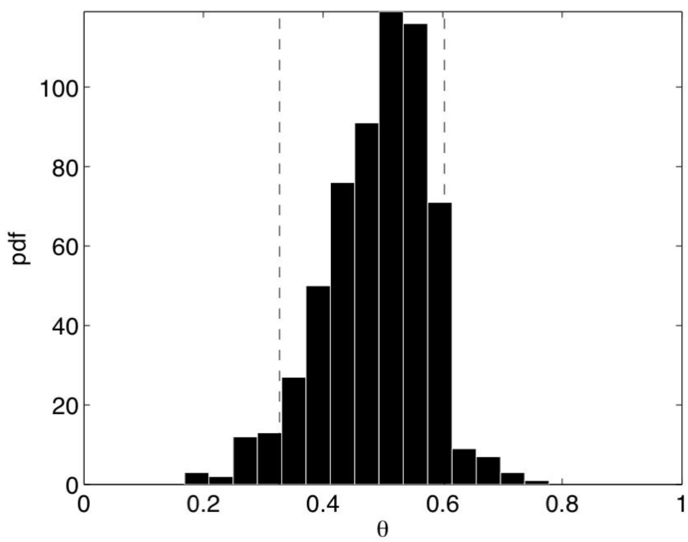
Clustering of extremes is characterized by the index $0 < \theta \leq 1$: $\exp(-\theta\, \tau/\tau_0)$.

Troubles Never Come Alone

Extremal Index of Fast CMEs



Determined by comparing observed Max Spectrum with Max Spectrum simulated from randomized data



$$\langle\theta\rangle = 0.5 \quad (0.3 - 0.6)$$

Fast CMEs with speeds 1,000-2,000 km/s arrive in clusters, *on average* 2-3 events closely spaced in time.

Fast CME Clusters

Size	N of Clusters	N of CMEs in Clusters	Proportion %	Mean Duration (hrs)
1	177	177	61	–
2	53	106	18	20
3	18	54	6	40
4	20	80	7	57
5	7	35	2	70
>5	17	169	6	108

$<\theta> = 0.5$, and speeds > 1000 km/s

Summary of the Data Analysis

- ✓ The Max Spectrum defines two exponents of extreme events:
  α (tail exponent) and θ (extremal index)

- ✓ The cumulative distribution of fast CMEs speeds follows a power law with α H3.4 (Fréchet extremes). This exponent defines *the fast CMEs.*

- ✓ The fast CMEs (and extreme SEPs associated with them) come in clusters. If one fast CME occurs it is followed on average by one or two other fast CME in a relatively short time. The mean time between CMEs with speeds exceeding 1,000 km/s is 42 hrs.

References

Stoev, S. A., G. Michailidis, and M.S. Taqqu (2006), Estimating heavy-tail exponents through max self-similarity, *Tech. Reports 445, 447, Dep. Statistics, Univ. Michigan.*

Hamidieh, K., S. Stoev, and G. Michailidis (2009), On the estimation of the extremal index based on scaling and resampling, *J. Comput. Graph. Statistics*, 18, 731--755, doi10.1198/cgs.2009.08065.

Ruzmaikin, A., J. Feynman, and S. Stoev (2011), Distribution and clustering of fast coronal mass ejections, *J. Geophys. Res*.,116, doi:10.1029/2010JA016247.

Questions?