Fakultät für Physik

UNIVERSITÄT DUISBURG ESSEN

Rare and Large Events: Examples from the Natural Sciences and Economics

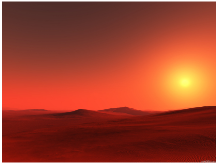
Thomas Guhr

rare events: optimal solutions and challenges — from charge transfer reactions to supervolcanoes

DPG spring meeting, Berlin, March 2014

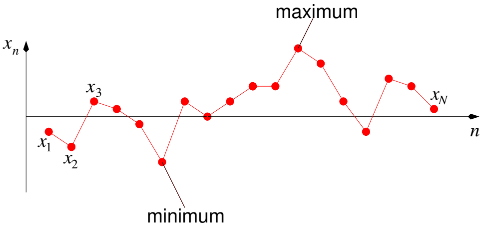
Outline

- some aspects of extreme, large and rare
- global warming and record statistics
- microwave experiments on freak waves
- price distributions and credit crisis



some aspects of extreme, large and rare

Central Limit Theorem



independent random variables
$x_n, \ n = 1, \ldots, N$
distribution $p(x_n)$

Central Limit Theorem:

scaled sample mean $\dfrac{1}{\sqrt{N}} \displaystyle\sum_{n=1}^{N} x_n$ converges to Gaussian

if second moment exists

What about the distributions of the minima and maxima ?

Extreme Value Theory

probability of maximum $P_N^{(\max)}(x) = (P(x))^N$

cumulative distribution $P(x) = \int_{-\infty}^{x} p(x')dx'$

large N limit: if properly rescaled $P_N^{(\max)}(x)$ converges, then to

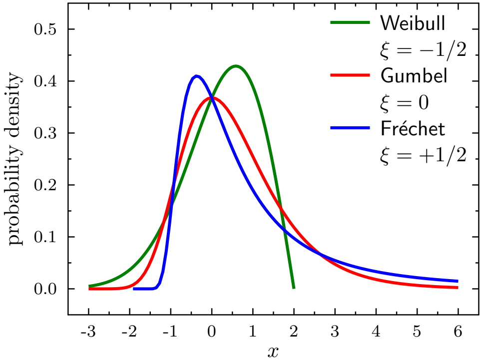
$$\exp\left(-(1+\xi x)^{-1/\xi}\right)$$

or

$$\exp\left(-\exp(-x)\right),\ \xi = 0$$

where ξ depends on $p(x)$

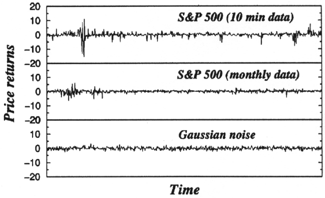
Fisher, Tippett (20's), Gnedenko (40's)

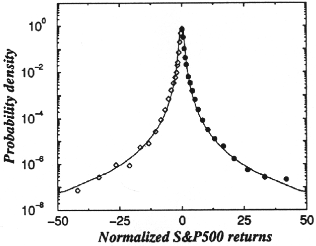


An Example from Finance

prices $S(t)$, returns

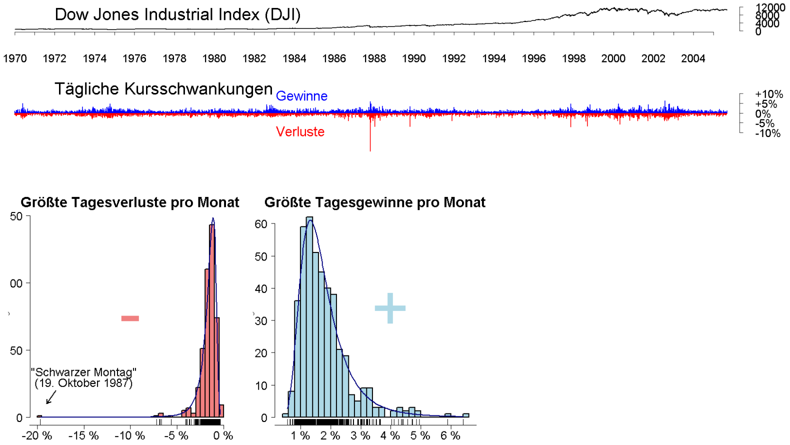
$$r_{\Delta t}(t) = \frac{S(t+\Delta t) - S(t)}{S(t)}$$





non–Gaussian, heavy tails! (Mantegna, Stanley, ..., 90's)

Distribution of Maxima and Minima in Return Series



(Keller–Ressel, Steiner, 2005)

Large Deviations Theory

example: coin–tossing with $x_n \in \{0, 1\},\ n = 1, \ldots, N$

for a given value z with $0.5 < z < 1$, the tail probability

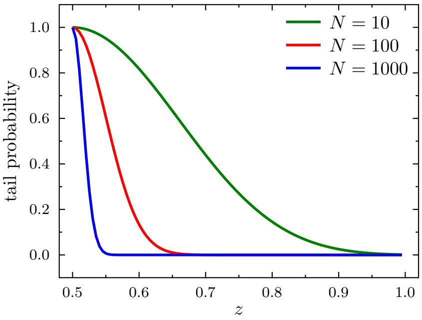
that sample mean $\frac{1}{N}\sum_{n=1}^{N} x_n$ is larger than z approaches

$\exp\left(-N I(z)\right)$

with "entropy"

$$I(z) = (1 - z)\ln(1 - z) + z\ln z + \ln 2$$

in the large N limit



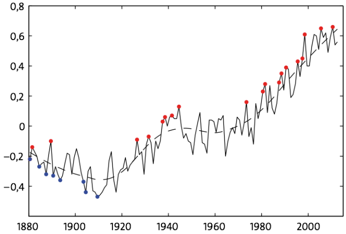
Cramér (30's), Gärtner (70's), Ellis (80's)

global warming and record statistics

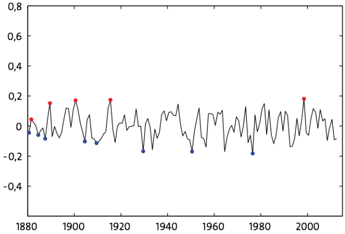
Temperature Records

a record is an event larger or smaller than everything before

yearly global temperatur deviations in degree Celsius
from long–term mean value



upwards trend



detrended

many more high temperature records due to trend !

(Wergen, Krug, Rahmsdorf, 2014)

Linear Drift Model

discrete time $t = 1, 2, 3, \ldots$, time series $x_t = \sigma\varepsilon_t$
independent random variables ε_t with symmetric distribution, standard deviation σ $\longrightarrow$
probability that x_t is a record is $q(t) = 1/t$

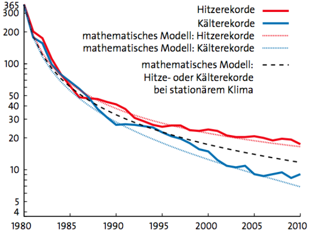
now: model with linear trend $x_t = ct + \sigma\varepsilon_t$ with
drift constant c $\longrightarrow$

$$q(t) \approx \frac{1}{t} + \frac{c}{\sigma} f(t)$$

with slowly varying function $f(t) = \dfrac{2\sqrt{\pi}}{e^2}\sqrt{\ln\left(\dfrac{t^2}{8\pi}\right)}$

(Wergen, Krug, 2010)

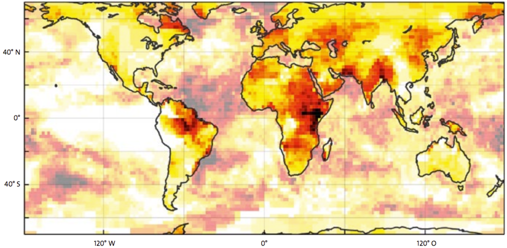
Model versus Data



European temperature records on daily basis, compared to same calendar days

$c/\sigma = 0.014/\text{year}$

more high, less low temperature records



monthly based deviations,
decade 2001 to 2010
black: 20 times over $1/t$

(Wergen, Krug, Rahmsdorf, 2014)

Issues Beyond the Previous Examples

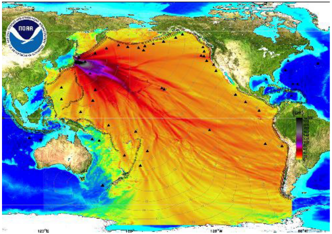
several other challenges are often encountered in systems of different kinds, to name but a few:

- finite sample size or short time series
- non–stationarity or non–locality
- correlations

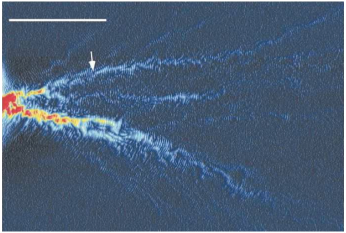
microwave experiments on freak waves

Freak Waves — Caustics and Branching

monster wave



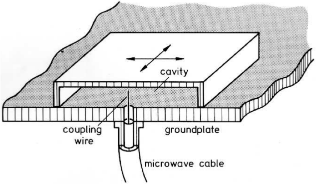
tsunami



flow in 2d electron gas

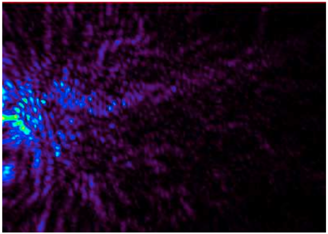
caustics in pool

Microwave Experiments

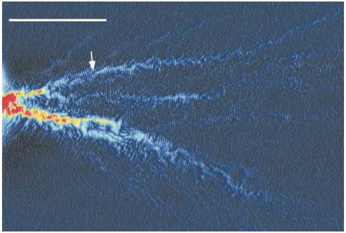


experimental setup

disorder potential



field distribution at 31GHz



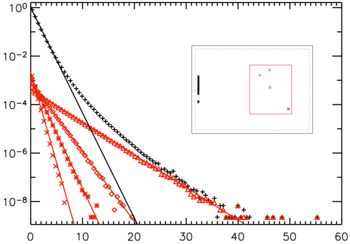
flow in 2d electron gas

(Barkhofen, Kuhl, Stöckmann et al., 2010)

Intensity Distribution — Fluctuating Variance

intensities I at each position
Rayleigh distributed

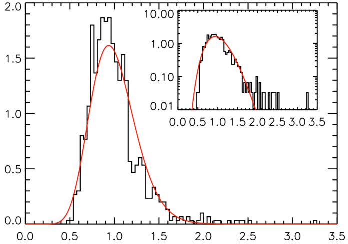
$$p(I|z) = \frac{1}{z}\exp\left(-\frac{I}{z}\right)$$



but with spatially fluctuating variance z, data fit

$$\chi^2_\nu(z) \sim z^{\nu/2-1}\exp\left(-\frac{z}{2}\right)$$

χ^2_ν distribution with $\nu = 30 \ldots 50$ degrees of freedom



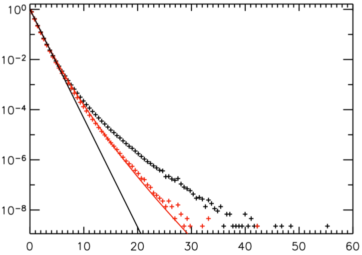
Compounding the Global Distribution

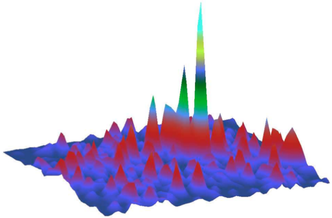
average Rayleigh distribution over variances

$$\langle p\rangle(I) = \int_0^\infty p(I|z)\chi_\nu^2(z)dz \sim I^{(\nu-2)/4}\mathcal{K}_{(\nu-2)/2}\left(\sqrt{2\nu I}\right)$$

compounding, mixture (mathematics), super statistics (physics)

much heavier tails (linear theory!), hot spots still outside

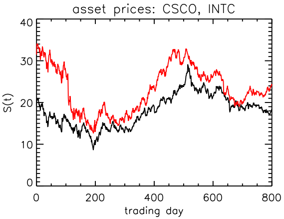




(Höhmann, Kuhl, Stöckmann, Kaplan, Heller, 2010)

price distributions and credit crisis

Financial Correlations are Non–Stationary

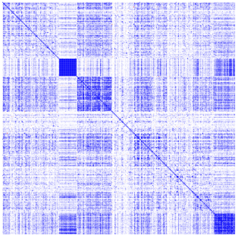


returns $r_k(t),\ r_l(t),\ t = 1, \ldots, T$

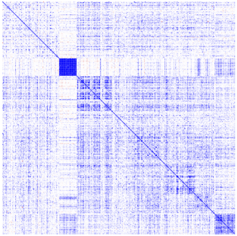
Pearson correlation coefficient

$$C_{kl} = \frac{1}{T} \sum_{t=1}^{T} r_k(t) r_l(t)$$

for K stocks, C is $K \times K$ matrix



fourth quarter '05

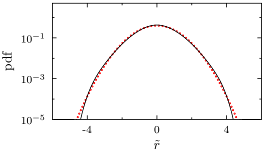


first quarter '06

Average over Ensemble of Correlation Matrices

multivariate distribution of returns $r = (r_1, \ldots, r_K)$ is Gaussian

$$g(r|C_s) \sim \exp\left(-\frac{1}{2} r^\dagger C_s^{-1} r\right)$$



if analyzed in short time intervals with correlation matrix C_s

idea to handle long time intervals:
replace $C_s \longrightarrow WW^\dagger$ with random correlation matrix $WW^\dagger$

$$\langle g \rangle(r|C_0) = \int g(r|WW^\dagger) f(W|C_0) d[W]$$

where $f(W|C_0)$ is Gaussian, such that $WW^\dagger$ fluctuate around empirical C_0 measured in whole, long time interval

(Schmitt, Chetalova, Schäfer, Guhr, 2013)

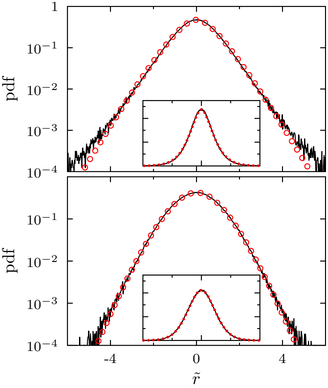
A Surprising Observation

$$\langle g \rangle (r|C_0) = \int_0^\infty g(r|zC_0)\chi_N^2(z)dz \sim \frac{\mathcal{K}_{(K-N)/2}\left(\sqrt{Nr^\dagger C_0^{-1} r}\right)}{\sqrt{Nr^\dagger C_0^{-1} r}^{(K-N)/2}}$$

where N measures strength of fluctuations around C_0

same form as compounded distribution of intensities in microwave experiments

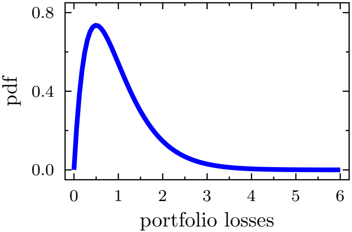
compounding traced back to fluctuating correlations



Credit Risk and Instability of Financial Markets

many companies borrow money from bank,
stock prices reveal the companies' ability to pay back
$\longrightarrow$ distribution of losses for the bank depends on returns

idea: use ensemble averaged $\langle g \rangle (r|C_0)$ for quantitative study



generic result: diversification does not work!

(Schmitt, Chetalova, Schäfer, Guhr, 2014)

Summary

- universal features of extreme values and large deviations
- linear drift model quantitatively models global warming
- non–stationarity and correlations often important
- microwave measurements of freak waves,
  fluctuating variances handled with compounding
- compounding traced back to fluctuating correlations
- model for multivariate returns distribution,
  generic features of credit risk