Stochastic Episodes with Light and Heavy Tails: Models, Properties, and Testing

2019 Fall Symposium

Nevada Chapter of the American Statistical Association

Reno, 19 October, 2019

Anna K. Panorska, Tom Kozubowski, Marek Arendarczyk*, Fares Qeadan, and Alexander Gershunov*****

Department of Mathematics and Statistics, University of Nevada Reno, USA
***Mathematics Institute, University of Wroclaw, Poland**
****Department of internal Medicine (Biostatistics), University of Utah, USA**
*****Climate Research Division, Scripps Institution of Oceanography, UC San Diego, USA**

Our thanks go to

- **Organizers of the symposium**
- **National Science Foundation of the United States, ATM-0231781.**
- **Marie Curie International Research Staff Exchange Scheme Fellowship within the 7th European Community Framework Programme, RARE-318984.**
- **California Energy Commission's Public Interest Energy Research Program: California Climate Change Center.**
- **National Oceanic and Atmospheric Administration Office of Global Programs: California Applications Program.**

Outline

- **Motivation: Extreme hydroclimatic and weather events, e.g. drought, flood, heat waves; financial modeling, energy use.**

  "From the tropics to the arctic, climate and weather have powerful direct and indirect impacts on human life. Extremes of heat and cold can cause potentially fatal illnesses, ... Other weather extremes, such as heavy rains, floods, and hurricanes, also have severe impacts on health. Approximately 600 000 deaths occurred worldwide as a result of weather-related natural disasters in the 1990s."
  The WHO, 12 November, 2008

- **Stochastic Models: First joint models for duration N, magnitude X and maximum Y of events: (N, X, Y) in the exponential and heavy tail cases.**

- **Decision: light or heavy tailed model? A likelihood ratio test for exponential versus Pareto distribution.**
- **Examples**
- **Summary**

MOTIVATION/EXAMPLES

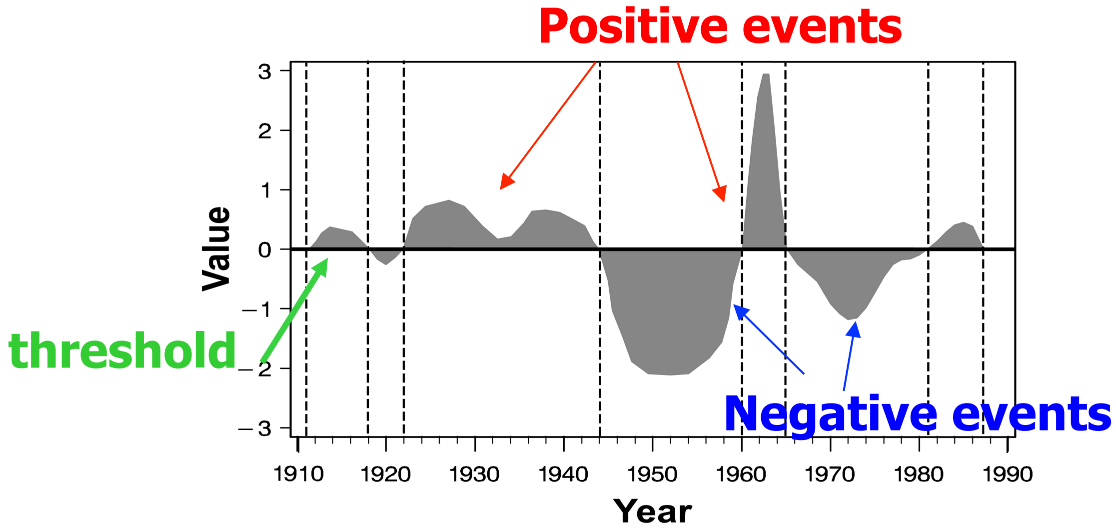
- Climate and hydrology – drought,
- Climate and weather – heat waves,
- Financial growth/decline episodes,
- Peak over Threshold theory

Start with a process: $X_1, X_2, \ldots, X_k,$

$$\text{Duration} = N$$

$$\text{Magnitude} = X = \sum_{i=1}^{N} X_i$$

$$\text{Max/Peak} = Y = \max_{i=1\ldots N} X_i$$



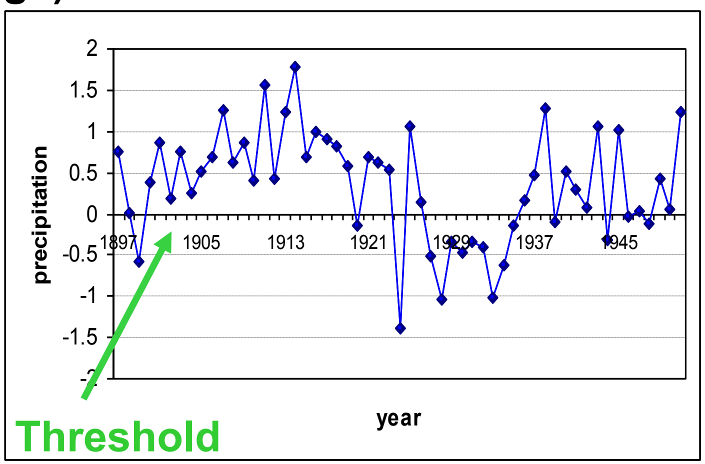
GOAL: Model for the random vector

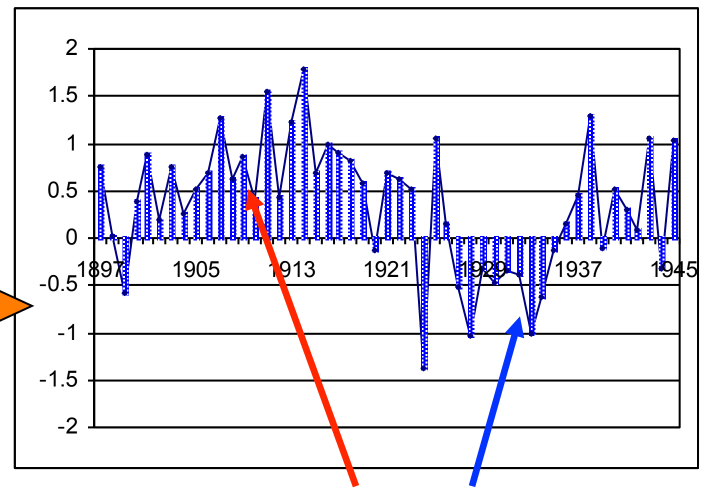
$$\left(N, \sum_{i=1}^{N} X_i, \max_{1 \le i \le N} X_i\right) = (N, X, Y)$$

Motivation: Climate and hydrology - drought

"DUST BOWL" - Great drought of the 1930s in the USA, the setting of John Steinbeck's "Grapes of Wrath".

Original precipitation data in standard deviation units plotted as difference from a threshold (source: western juniper tree rings).





Precipitation "events" or episodes"

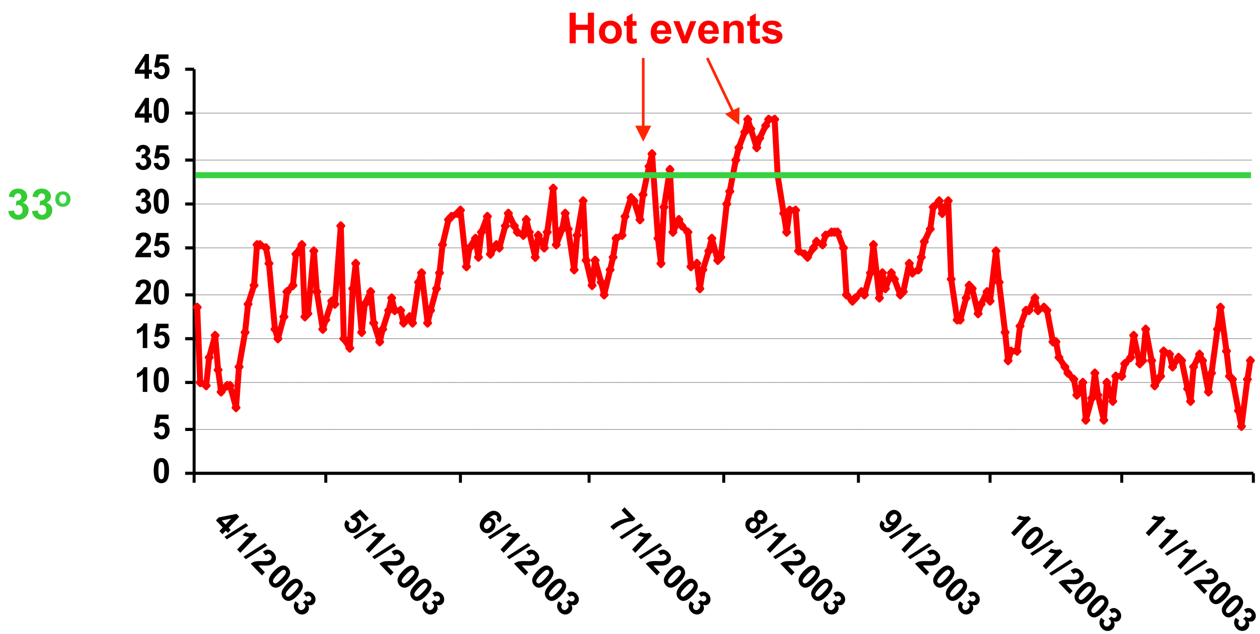
An episode is a period with the process staying consecutively above/below threshold: e.g. "dry", "wet" year, drought, flood, etc.

Threshold for "dry" or "wet" depends on the definition of the episode (e.g. drought).

(Kozubowski, Panorska (2005, 2007), Biondi et al. (2007), Saito et al. (2008))

Motivation: Climate and weather – heat waves

- In August 2003, France experienced an extreme heat wave, that resulted in an estimated 14,802 deaths*.
- Hot event: consecutive observations above the 33°C.



What are the chances that a large heat wave will happen again?

What are the chances of a hot event with duration equal to this one, 11 days?

What are the chances that a heat wave longer than 6 days and larger than current 98th percentile of heat waves' magnitudes?

Heat waves of 2019 and their public health toll are being studied now.

*Dhainaut et al. 2004.

Motivation: Climate and hydrology – floods

Example: Flood events in Reno, NV.

Main question: What is the size of hydrological or weather extremes, that is the size of high percentiles of hydroclimatic processes (e.g. precipitation, stream flow, temperature)?

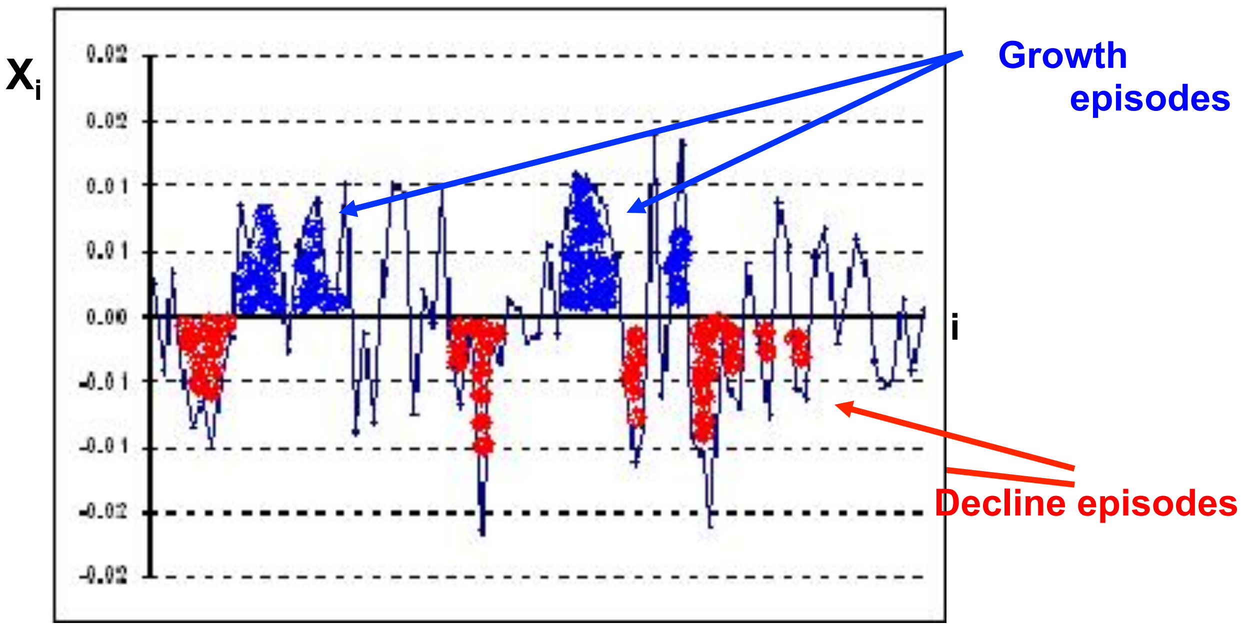
What are the chances of a flood of given magnitude?

Motivation: Financial growth/decline episodes

Daily exchange rates between Japanese Yen and British pound quoted in UK pounds, Jan. 2, 1980-May 21, 1996.

Process X_i: Daily log returns, X_i=log(Rate_day_i/Rate_day_i-1), n=4274.

Episodes: consecutive days when the exchange rates were growing/declining, i.e.: Growth $X_i > 0$, Decline $X_i < 0$.



N = length of a growth period, X=ΣXi=cumulative log return over a growth period, Y=maximum log return over the growth period.

Nolan 2001, Kozubowski and Podgorski 2001, Kozubowski and Panorska 2004, 2007, 2011

Random vector $(N, \sum_{i=1}^{N} X_i, \max_{1\le i\le N} X_i) = (N, X, Y)$

GOAL: Construct a mathematically natural model for the JOINT distribution of (duration, magnitude X and maximum Y) of events.

Notable properties of the random vector (N, X, Y):

- **All components are related/dependent, the joint behavior of X and Y is not trivial,**
- **The sum and maximum are of random number of random observations.**

Hierarchical approach:

1. **Specify distribution of N**
2. **Given N=n, find conditional distr.** $f(x, y \mid n)$ **of (X, Y|N=n) =** $(\sum_{i=1}^{n} X_i, \max_{i=1,\ldots,n} X_i)$

3. **Get the joint distribution of (N, X, Y) as**

$$f(n, x, y) = f(x, y \mid n) f_N(n)$$

Basic Model: Xi's iid exponential

Trivariate model: N-Geo(p), X_i's iid exp(β), independent of N. All distributions below have explicit representations in terms of pdf and/or cdf.

$$(X,Y,N)=\underbrace{\left(\sum_{i=1}^{N} X_i, \max_{1\le i\le N} X_i, N\right)}_{\text{TETLG}(\beta,p)}$$

Trivariate Distribution with exponential, truncated logistic and geometric marginals

Main component:

$$(X,Y)=\underbrace{\left(\sum_{i=1}^{n} X_i, \max_{1\le i\le n} X_i\right)}_{\text{BGGE}(\beta,n)}$$

BGGE: Bivariate Distribution with Gamma and Generalized Exponential Marginals

Other bivariate marginals:

$$(N,X)=\underbrace{\left(N, \sum_{i=1}^{N} X_i\right)}_{\text{BEG}(\beta,p)}$$

$$(N,Y)=\underbrace{\left(N, \max_{1\le i\le n} X_i\right)}_{\text{BTLG}(\beta,p)}$$

BEG: Bivariate distr. with exponential and geometric marginals

BTLG: Bivariate distr. with truncated logistic and geometric marginals

•Kozubowski, Panorska (2005, 2008), Biondi, Kozubowski, Panorska (2005, 2007 (a, b)), Saito, Biondi, Kozubowski, Panorska (2008), Qeadan, Kozubowski, Panorska (2008, 2011, 2012)

What if light tails are too restrictive? Data shows possibility of heavy tail?

We work with $(X_1, X_2, \ldots, X_n)$ with a multivariate Pareto type II, MP(α, β) (Lomax) distribution, Arnold 1983, with representation

$$(X_1, X_2, \ldots, X_n) \overset{d}{=} \left(\frac{E_1}{T}, \frac{E_2}{T}, \ldots, \frac{E_n}{T} \right),$$

where $(E_1, E_2, \ldots, E_n)$ are iid exponential β, and T is independent of Ei random variable with Gamma(1/ α, 1/ α) distribution.

T may be thought of as the "background" or "common" risk. Marginal distributions are univariate Pareto II with survival functions

$$S(x) = P(X_i > x) = \; = \left[\frac{1}{1 + \alpha\beta x} \right]^{1/\alpha}, \;\; x \geq 0.$$

Applications of vector $(X_1, X_2, \ldots, X_n)$: reliability, finance, insurance etc.
(see for example Asimit, Furman and Vernic (2010), Asimit, Vernic and Zitikis (2013), Cai and Tan (2007), Chiragiev and Landsman (2007), Langseth (2007), and the refs therein)

Heavy Tail Model: (X_1, …, X_n) Multivariate Pareto MP(α, β) vector

Trivariate model: N-Geo(p), (X_1, …, X_n) Multivariate Pareto(α, β) vector independent of N.

$$(X, Y, N) = \underbrace{\left(\sum_{i=1}^{N} X_i, \max_{1 \le i \le N} X_i, N \right)}_{\mathrm{GSMP}(\alpha, \beta, p)}$$

GSMP: Geometric Sum and Maximum of Pareto variables

- **1-d Marginal distributions: N- geo(p); X – Pareto II(α, βp);**
  Y – known but not explicit formula for pdf

Main component:

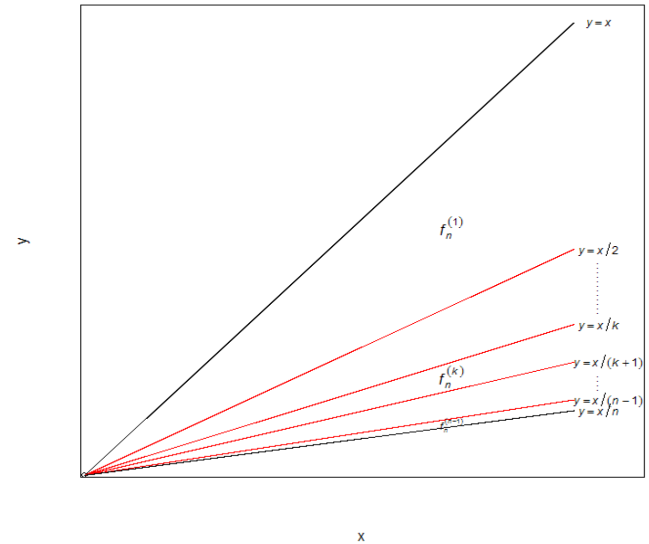
$$(X, Y) = \underbrace{\left(\sum_{i=1}^{n} X_i, \max_{1 \le i \le n} X_i \right)}_{\mathrm{SMP}(\alpha, \beta, n)}$$

SMP: Sum and Maximum of Pareto variables

Distributions of all other marginals and conditionals (pdf, cdf, estimation) is known exactly.

Arendarczyk, Kozubowski, Panorska (2018), J of Multivariate Analysis

PDF of Geometric Sum, Maximum, and Duration of Pareto variables



- **Domain of the pdf x/n ≤ y ≤ x**
- **PDF formula $f_n^{(k)}$ will depend on the sector k of the plane.**

$$f\,(x,y\,|\,n)=\frac{\Gamma(n+1/\alpha)}{\Gamma(1/\alpha)}\;(\alpha\beta)^{\mathrm{n}}\left[\frac{1}{1+\alpha\beta x}\right]^{n+1/\alpha}H(x,y,n),\quad \text{where}$$

$$\mathrm{H(x,y,n)}=\sum_{\mathrm{s=1}}^{\mathrm{k}}\frac{\mathrm{n(n\text{-}1)}}{\mathrm{(s\text{-}1)!(n\text{-}s)!}}(x-sy)^{n-2}(-1)^{s+1},\ \text{and}\ \ (x,y)\in S_k=\left\{(x,y):0\le\frac{1}{k+1}x\le y\le\frac{1}{k}x\right\}.$$

- **We obtain exact trivariate pdf:** $f(n,x,y)=f(x,y\,|\,n)f_N(n)$

•NOTE: As $\alpha\to 0$, we obtain the pdf for the model with exponential distribution

A Note on Estimation of Parameters

- Maximum likelihood estimators always exist for all parameters
- Get MLEs using standard numerical methods.
- Interesting statistical point: Computation of MLEs uses only data on X and N, not on Y. However, we will see very reasonable fit of the marginal distributions to Y.

A Practical Question: Which model should we use: exponential or Pareto?

QUESTION: **Are the observations of (X, Y, N) coming from a distribution with light or heavy tail?**

NEED **a decision tool/test differentiating between Pareto and exponential distributions.**

Motivation: **Why bother with designing a test? Why tails matter?**

Why tails matter?
An example: flood

Why important?

What is the size of hydrological or weather extremes, that is the size of high percentiles of hydroclimatic processes (e.g. precipitation, stream flow, temperature)?

It influences

- **the size of climatic and hydrological risk (insurance and safety);**
- **safe engineering design standards (USGS, US Corps of Engineers);**
- **water management policies and procedures;**
- **insurance policies.**

Extremes change in the changing global climate.

NEED: Study the nature of daily precipitation distribution tails to find realistic estimates of high precipitation percentiles.

Started with precipitation- it is not influenced by people and drives river flow, floods, and droughts.

Likelihood Ratio Test for exponential versus Pareto*

Decide if the observations are from a distribution with a light or heavy tail.

Used the following model for the observations from Pareto II, survival function:

$$S(x) = \left(\frac{1}{1 + \omega x / s} \right)^{1/\omega}, \quad x \geq 0, s > 0, \omega \geq 0.$$

This model encompasses Pareto II : ($\omega > 0$) and exponential : ($\omega=0$)

(with the understanding that $\omega=0$ corresponds to the limiting exponential case $\omega \to 0$) .

- ***We just finished work on a new and more general test for tail index of a Generalized Pareto using Greenwood statistic – stay tuned.**

Deciding about exp versus heavy tail: Likelihood Ratio Test of Exponentiality versus Pareto

The Problem: Test $H_0 : \omega = 0$ (exponentiality) vs. $H_1 : \omega > 0$ (Pareto II)

Asymptotic Distribution of the Test Statistic

Proposition. **Under the null hypothesis, the asymptotic distribution of the deviance statistic $-2\log\lambda_n$, where λ_n is the test statistic, is the same as that of IW, where I and W are two independent random variables having a Bernoulli distribution with parameter ½ and a chi-square distribution with one degree of freedom, respectively.**

$$\text{Deviance} = -2\log\lambda_n \overset{d}{=} \underbrace{\overset{\text{Bern}(0.5)}{I}\;\overset{\chi^2(1)}{W}}_{\text{Mixed distribution}}$$

Mixed distribution

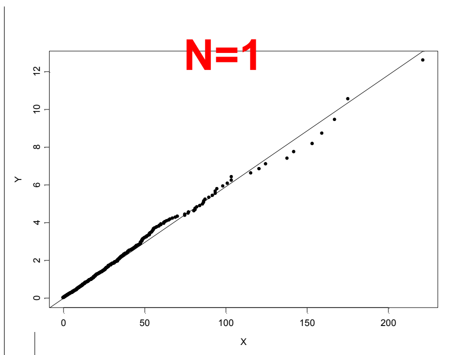
Bern(0.5) $\chi^2(1)$

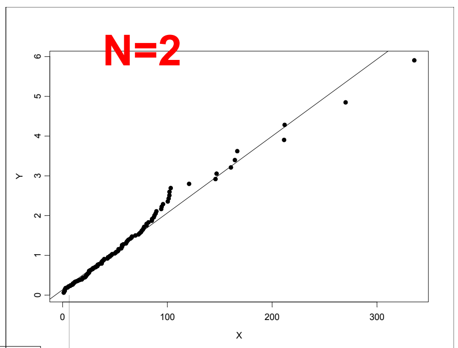
Kozubowski, et al. (2009). Testing exponentiality versus Pareto distribution via likelihood ratio. *Communications in Statistics: Simulation and Computation.* 38 (1), pp. 118-139.

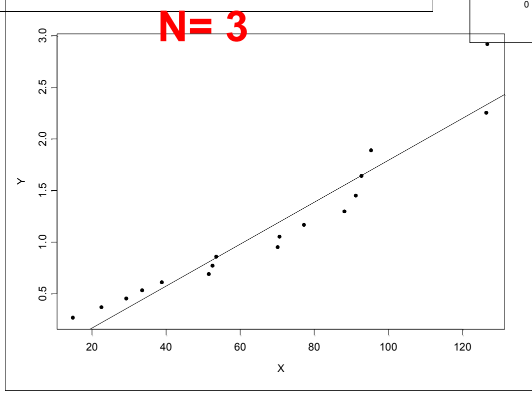
Example 1: Pareto case
PRECIPITATION in Miami, Fl, USA

DATA: Total daily precipitation, January 1, 1950 to Dec 31, 2013, (23741 obs). Used excesses over the 95th percentile.

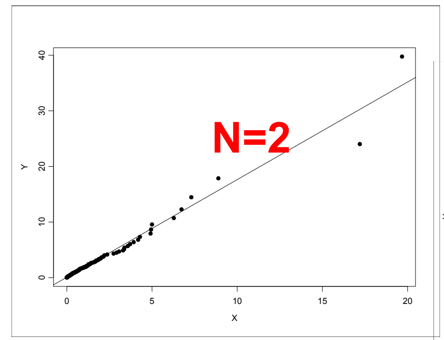
Fit of conditionals: X|N

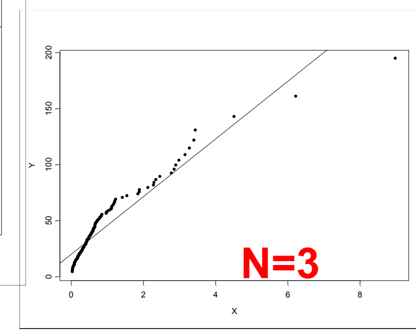




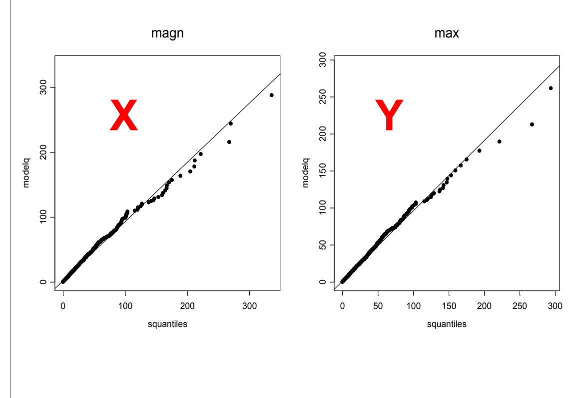


Fit of conditionals: Y|N



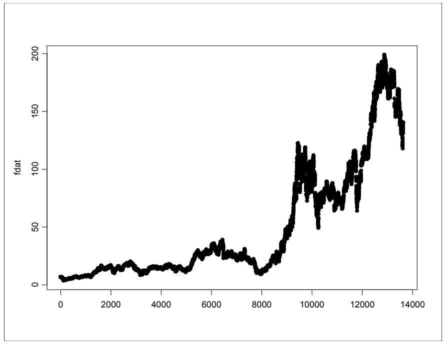


Fit of marginals: X and Y

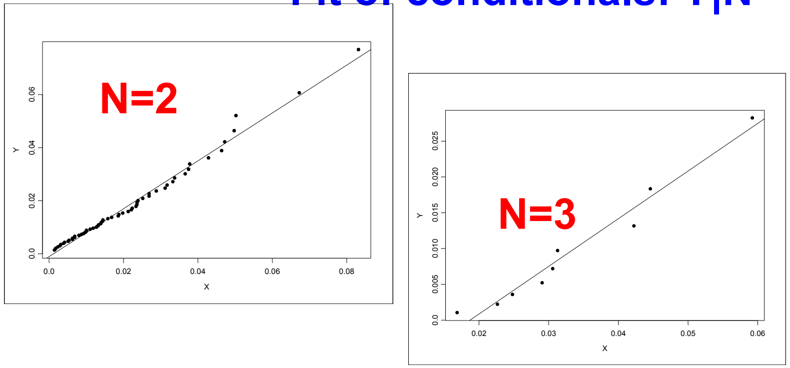


Example 2: Pareto case
Returns on IBM stock: Jan 2, 1962 to March 10, 2016

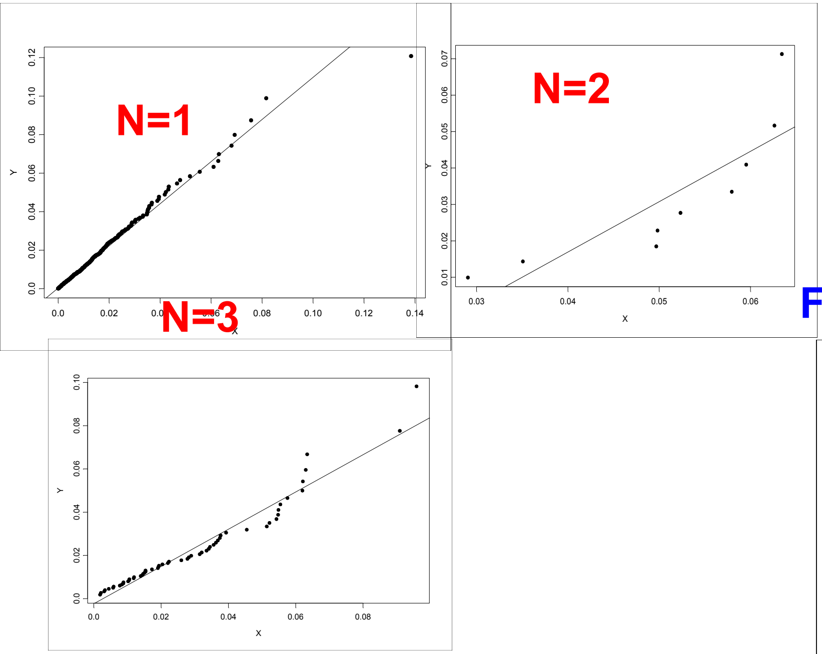
DATA: 13,635 returns, used (absolute) excesses below 5th percentile. Test stat= 3.346, .



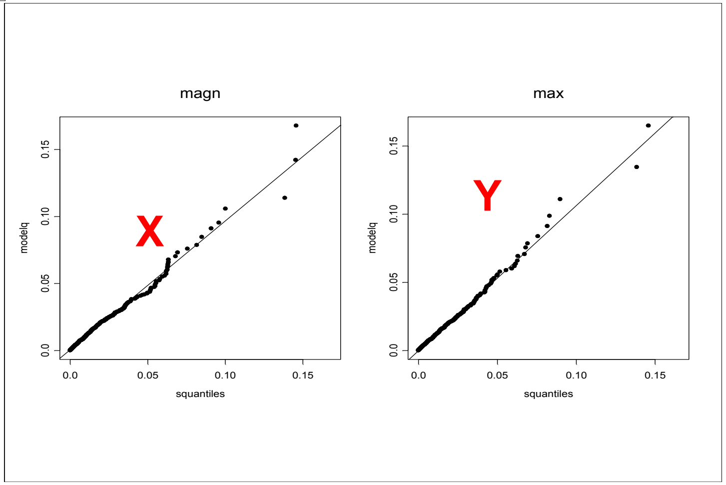
Fit of conditionals: Y|N



Fit of conditionals: X|N



Fit of marginals



EXAMPLE: Pareto case -Tail Measures of Risk, Tail Conditional Expectation*

- **Consider** $(X_1, X_2, \ldots, X_n)$ **- MP random risks/losses of a financial institution (e.g. lines of business, portfolio allocations, etc.)**

- **Let X= total risk, Y= max risk.**
- ***Existing work: Tail Conditional Expectations* (TCE)**

$$K(s)=E(X|X>s),\ K_{max} = E(Y|Y>s)$$

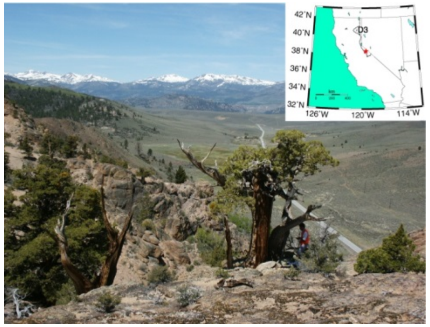
- **New work: We compute mixed measures of risk: E(X|Y>t) and E(Y|X>s)- need joint distribution of (X, Y) to compute them (focus: influence of max on the porfolio)**

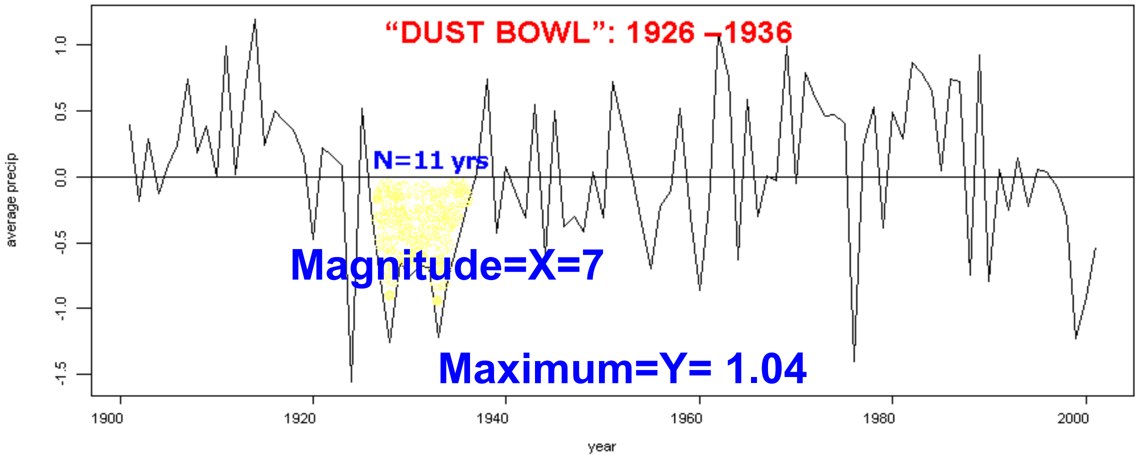
- **Both can be computed for our model, and result in closed (albeit long and messy) form formulas.**

Note: Hua and Joe (2011, ‘12, ‘14) considered E(X|Y>t) and E(Y|X>s) asymptotically, when s, t approach infinity (focus: influence of one asset on the portfolio)

•References: Arendarczyk, Kozubowski, Panorska (2018), *Vernic (2011), Chiragiev and Landsman (2007), Forman and Zitikis (2008), Hua and Joe (2011, 2012, 2014) and the refs therein.

EXAMPLE: Exponential Case: Climate and Hydrology - THE “DUST BOWL”



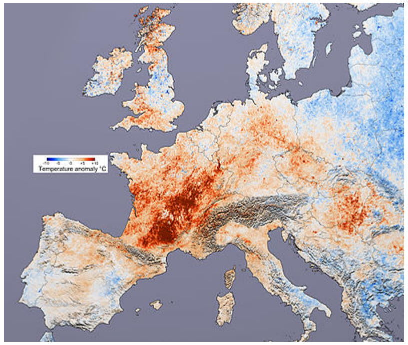


The data: Dendroclimatic (western juniper) reconstruction of precipitation from 300 BC to AD 2001 in the Walker River watershed (California/Nevada). California Climate Division 3.

Examples of answers to practical questions:

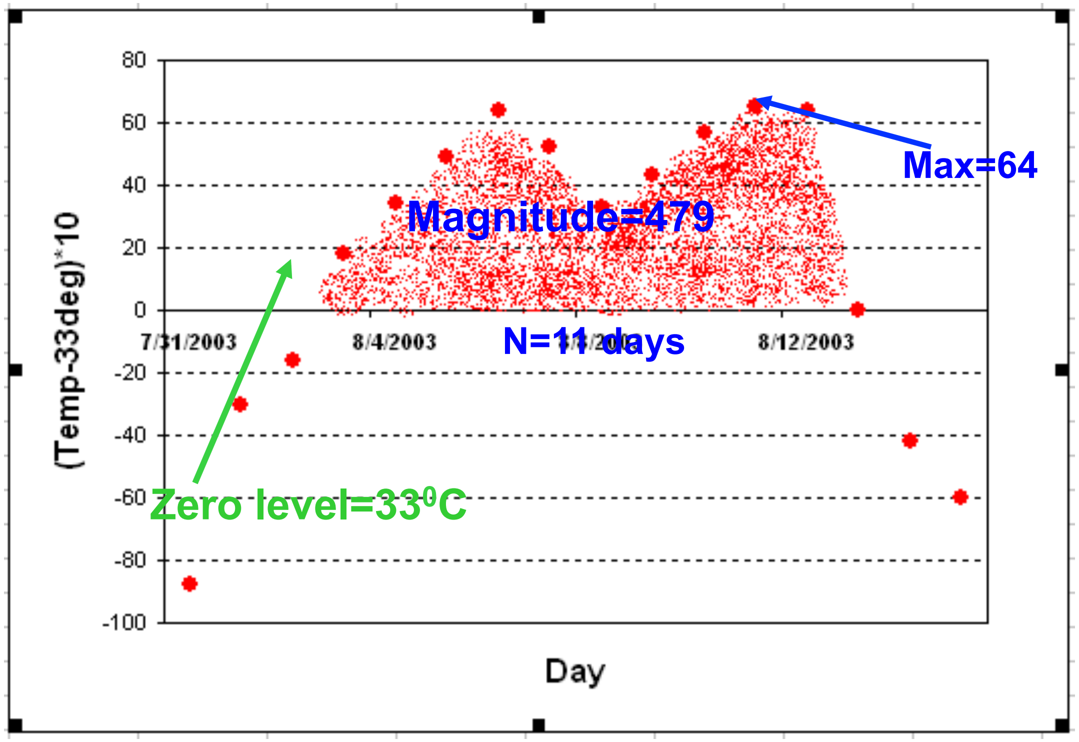
- Probability of a drought longer or larger than the 'Dust Bowl' is 0.08;
- Probability of a drought longer and larger than 'Dust Bowl' is 0.06.
- Conditional probability of a drought with at least 'Dust Bowl'’s magnitude given that duration is 11 years is 0.46

EXAMPLE: Exponential case: Paris heat wave of the 2003



In August 2003, France experienced an extreme heat wave, that resulted in an estimated 14,802 deaths*.

Definition of a hot event: consecutive observations above the 33°C.



- **Probability of a hot event with the same duration of 11 days is 0.000075;**
- **Conditional probability of a heat wave with at least that magnitude given that duration is 11 days is about 5.5e-4.**
- **Probability of a heat wave longer than 6 days and larger than 100 (98th percentile of magnitudes) is 0.005.**

Note: Maximum Y = 64deg C*10 (really 39.4oC)

***Dhainaut et al. 2004., Data from http://eca.knmi.nl/, station ID 104969**

Examples: Exponential Case- Financial Data

For the exchange rates data (Japanese Yen and GBP in GBP, Jan 2, 1980 – May 21, 1996), we constructed data set of episodes.

Growth Xi >0, decline Xi <0

For analysis, considered episodes of growth: Xi >0.

Our data were 1902 triples (X, Y, N) of growth episodes. We checked that:

- **Positive log-returns come from exponential distribution;**
- **Magnitudes of growth periods also come from exponential distribution;**
- **The fit of all bivariate marginals and conditional distributions is quite reasonable;**

OVERALL CONCLUSION: REASONABLE FIT of all the trivariate and marginal (bivariate) models (for the growth episodes)

Nolan 2001, Kozubowski and Podgorski 2001, Kozubowski and Panorska 2004, 2008, 2011.

In the matter of tails…..an Illustration

Precipitation: Diversity drives Volatility – North American Data

Q: Are extreme precipitation events of exponential or Pareto type?

The heaviest tails: regions with large variety of the synoptic systems producing precipitation.

Exponential tails: only in regions with similar synoptic systems causing precipitation.

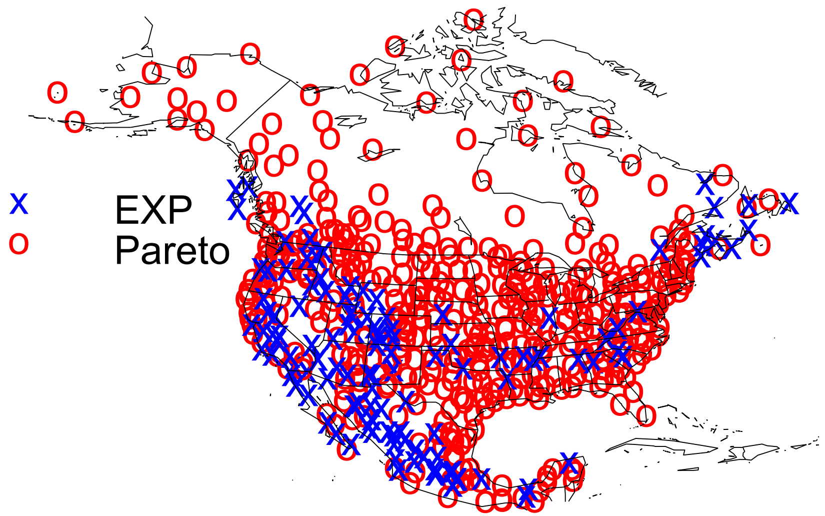
Diversity in causes/types of precipitation

⬇

Enhance volatility in precipitation

⬇

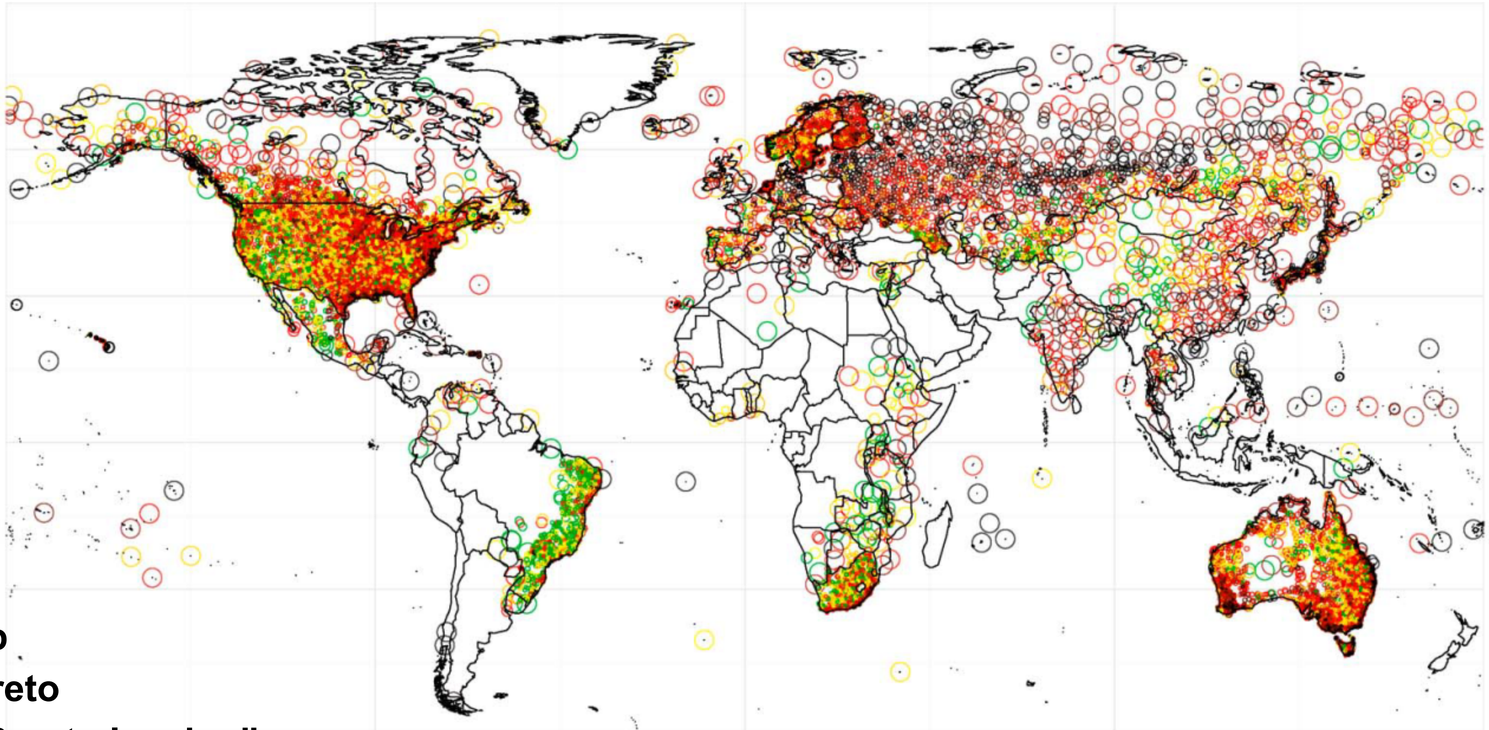
Heavy tailed excesses.



Spatial distribution of the decision (exp. or Pareto) regarding the tails of the excesses. 81% classified as Pareto on 5% significance level.

Data: Daily total precip. from 560 meteo. Stations: Canada, U.S., and Mexico; Jan. 1, 1950 to Dec. 31, 2001. Tested the distribution of excesses over local 75th percentile threshold.

Precipitation: Diversity Drives Volatility - Global Data



● exp
● Pareto
● “Pareto leaning”

About 65% Pareto, about 91% Pareto leaning or Pareto

Data: Daily total precip., 22,000 best quality stations from Global Historical Climatology Network Daily; Jan. 1, 1950 to Dec. 31, 2013. Tested the distribution of excesses over local 90th percentile threshold.

Cavanaugh, N. R., A. Gershunov, A. K., Panorska, and T. J. Kozubowski (2015),The probability distribution of intense daily precipitation, Geophys. Res. Lett.,42, 1560–1567, doi: 10.1002/2015GL063238.

Current and Future research

- Ilaria Vinci and Francesco Zuniga (PhD students): spatial and regression models for heat waves;
- Questions of goodness of fit for stochastic models
- Heavy tails in applications, tests and confidence intervals;
- Truncated distributions.

Thank You