Analysis methods of heavy-tailed data

Natalia Markovich

Institute of Control Sciences
Russian Academy of Sciences, Moscow, Russia

July 4, 2022

Modul 2

The Modul 2 contains

several rough tools for the detection of heavy-tails, the number of finite moments, independence and long-range dependence.

Modul 2: Lesson 1

Rough methods for the detection of heavy tails.

Rough methods for the detection of heavy tails and the number of finite moments.

1. **Ratio of the maximum to the sum**
   Let $X_1, X_2, ..., X_n$ be i.i.d. r.v.s. We define the following statistic

$$R_n(p) = M_n(p)/S_n(p), \qquad n \geq 1, \qquad p > 0, \qquad \textit{where}$$

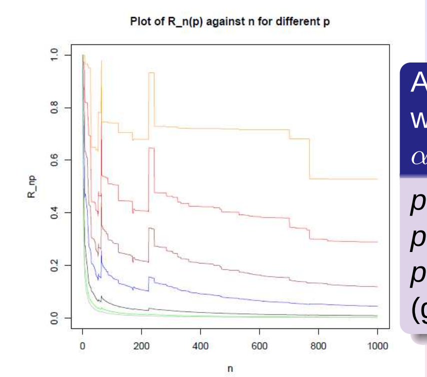
$$\begin{aligned} S_n(p) &= |X_1|^p + \ldots |X_n|^p, \\ M_n(p) &= \max\left(|X_1|^p, \ldots, |X_n|^p\right), \qquad n \geq 1 \end{aligned}$$

   to check the moment conditions of the data.

For different values of p the plot of $n \to R_n(p)$ gives

a preliminary information about the distribution $\mathbb{P}\{|X| > x\}$. Then $\mathbb{E}|X|^p < \infty$ follows, if $R_n(p)$ is small for large n: $R_n(p) \to^{a.s.} 0 \Leftrightarrow \mathbb{E}|X|^p < \infty$, $n \to \infty$. For large n a significant difference between $R_n(p)$ and zero indicates that the moment $\mathbb{E}|X|^p$ is infinite.

Plot of $R_n(p)$ against n for different p



A Frechet distributed sample with $\gamma = 0.5$ (tail index $\alpha = 1/\gamma = 2$) was generated.

$p = 1$ (black), $p = 2$ (blue), $p = 3$ (brown), $p = 5$ (red), $p = 10$ (orange), $p = 0.5$ (green), $p = 0.25$ (gray).

For $p \in \{0.25, 0.5, 1\}$ $R_n(p)$ goes to zero for increasing n.
For $p \in \{2, 3, 5, 10\}$ $R_n(p)$ seems to go to a positive constant for increasing n.

Conclusions: $E|X|^p < \infty$ for $p < 2$ only, $E|X|^p = \infty$ for $p \geq 2$.

Rough methods for the detection of heavy tails and the number of finite moments.

2. **QQ-plot** A QQ-plot (or "quantiles against quantiles"-plot) draws the dependence $\left\{\left(X_{(k)}, F^{\leftarrow}\left(\frac{n-k+1}{n+1}\right)\right) : k = 1, ..., n\right\}$, where $X_{(1)} \geq ... \geq X_{(n)}$ are the order statistics of the sample, and $F^{\leftarrow}$ is an inverse function of the DF F.

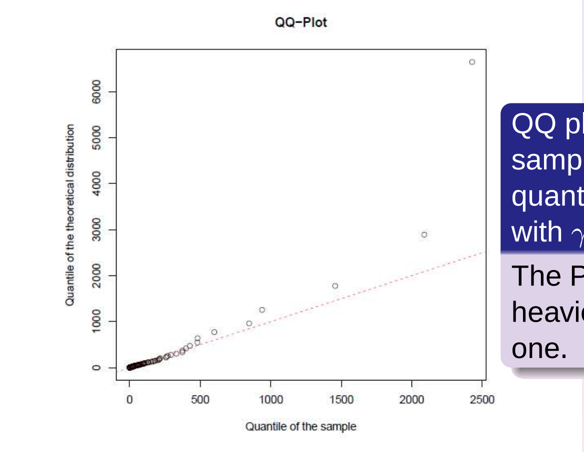
Background:

If X_1 has DF $F(x)$ then $F(X_1) = U_1$ is uniformly distributed.

Often, the QQ-plot is built as a dependence of exponential quantiles

against the order statistics of the underlying sample. Then $F^{\leftarrow}$ is an inverse function of the exponential DF. Then a linear QQ-plot corresponds to the exponential distribution.

QQ plots

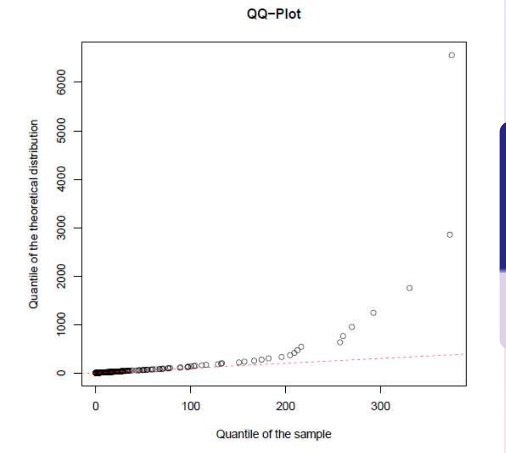


QQ plot for a Fréchet distributed sample with $\gamma = 0.5$ against the quantiles of a Pareto distribution with $\gamma = 1.2$ and $\sigma = 2$.

The Pareto distribution has a heavier tail than the Fréchet one.

If the model is true, points of the QQ plot lie at the line $y = x$.

QQ plots with removed largest observations

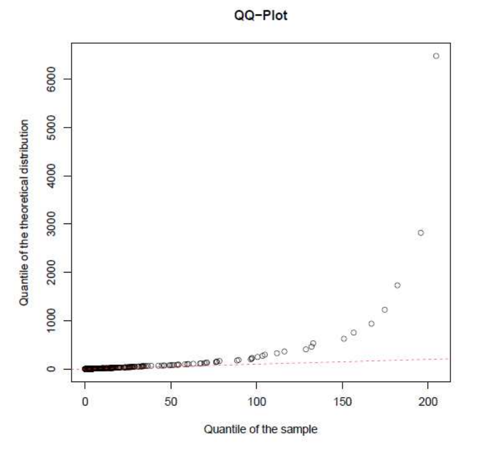


New outliers appear.

QQ plot for a Fréchet distributed sample with $\gamma = 0.5$ against the quantiles of a Pareto distribution with $\gamma = 1.2$ and $\sigma = 2$.

The 10 largest observations were removed.

QQ plots with removed largest observations



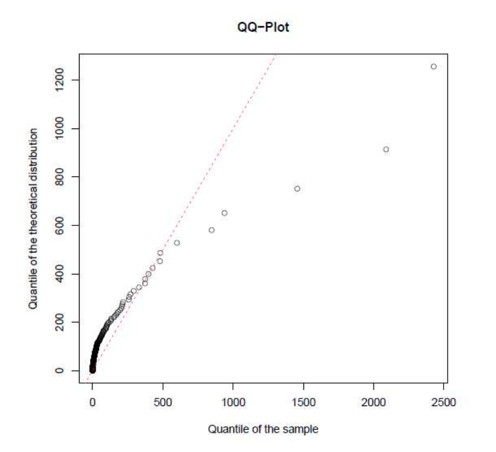
The outliers persist.

QQ plot for a Fréchet distributed sample with $\gamma = 0.5$ against the quantiles of a Pareto distribution with $\gamma = 1.2$ and $\sigma = 2$.

The 20 largest observations were removed.

This shows that the removal does not impact the heavy tails.

QQ plots



QQ plot for a Fréchet distributed sample with $\gamma = 0.5$ against the quantiles of an lognormal distribution with $\mu = 2.5$ and $\sigma = 1.5$.

The QQ plot is below the line $y = x$.

$F_{hyp}^{\leftarrow}(q) < F_n^{\leftarrow}(q)$, where $F_n(q)$ and $F_{hyp}(q)$ denote the empirical and model distribution functions, and $F^{\leftarrow}$ denotes the inverse function, or $F_{hyp}(x_q) > F_n(x_q)$, or $\overline{F}_{hyp}(x_q) < \overline{F}_n(x_q)$. The sample has a heavier tail than the model.

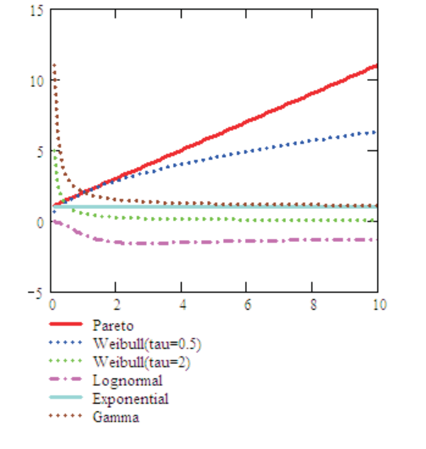
Rough methods for the detection of heavy tails and the number of finite moments.

3. **Plot of the mean excess function**

$$e_n(u) = \sum_{i=1}^{n}(X_i - u)\mathbf{1}\{X_i > u\} / \sum_{i=1}^{n}\mathbf{1}\{X_i > u\}$$

1. Heavy-tailed distributions: $e(u) \to \infty$, $u \to \infty$.
2. A Pareto distribution: a linear $e(u)$.
3. An exponential distribution: the constant $e(u) = 1/\lambda$.
4. Light-tailed distributions: $e(u) \to 0$, $u \to \infty$.

Rough methods for the detection of heavy tails and the number of finite moments.



Plots of function $e(u)$ for several distributions.

4. Hill's and other estimators of the tail index

The Hill estimate works bad if

1. the underlying DF does not have a regularly varying tail,
2. the tail index $\alpha = 1/\gamma$ is not positive,
3. the sample size is not large enough,
4. the tail is not heavy enough, i.e. γ is not big,
5. $F \subseteq \mathcal{R}_\alpha$, since it strongly depends on the slowly varying function $\ell(x)$ that is usually unknown.

The disadvantages of Hill's estimate show

that one has to apply several estimates of the tail index to deal with the complex analysis of data.

Modul 2: Lesson 2

Rough methods for the detection of independence.

Definition of the independence.

Definition

Random variables $\xi_1, \xi_2, ..., \xi_n$ are independent if for any $x_1, x_2, ..., x_n$ it holds

$$\begin{aligned}\mathbb{P}\{\xi_1 \leq x_1, \xi_2 \leq x_2 \quad &, \quad ..., \xi_n \leq x_n\} \\ &= \quad \mathbb{P}\{\xi_1 \leq x_1\}\mathbb{P}\{\xi_2 \leq x_2\}...\mathbb{P}\{\xi_n \leq x_n\}\end{aligned}$$

or in terms of cumulative distribution functions

$$F(x_1, x_2, ..., x_n) = F_1(x_1)F_2(x_2)...F_n(x_n),$$

where $F_k(x_k)$ is a distribution function of the r.v. ξ_k.

The opposite is also true.
It is not easy to build non-gaussian multivariate DFs.

Multivariate distribution for independent random variables.

Example

Let $\xi_1, \xi_2, ..., \xi_n$ be iid random variables with normal DF $F_k(x_k) = \frac{1}{\sigma_k\sqrt{2\pi}} \int_{-\infty}^{x_k} \exp\{-\frac{(z-a_k)^2}{2\sigma_k^2}\}dz$, $1 \leq k \leq n$ Then

$$F(x_1, ..., x_n) = (2\pi)^{-n/2} \sqcap_{k=1}^{n} \sigma_k^{-1} \int_{-\infty}^{x_k} \exp\{-\frac{(z-a_k)^2}{2\sigma_k^2}\}dz$$

If $\xi_1, \xi_2, ..., \xi_n$ have a probability density function (PDF) $f_k(x) = \frac{1}{\sigma_k\sqrt{2\pi}} \exp\{-\frac{(x-a_k)^2}{2\sigma_k^2}\}$ then multivariate PDF

$$f(x_1, ..., x_n) = \frac{(2\pi)^{-n/2}}{\sigma_1\sigma_2...\sigma_n} \exp\{-\frac{1}{2}\sum_{k=1}^{n}\frac{(x_k - a_k)^2}{\sigma_k^2}\}$$

Rough measures of the dependence.

Correlation between two r.v.s X_1 and X_2 is defined by

$$\rho(X_1, X_2) = \frac{cov(X_1, X_2)}{\sqrt{var(X_1)var(X_2)}},$$

where

$$cov(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$$

is a covariance, and $var(x)$ is a variance.

1. If X_1 and X_2 are independent then $\rho(X_1, X_2) = 0$.
2. If $\rho(X_1, X_2) = 0$ then not necessary X_1 and X_2 are independent!
3. If Gaussian r.v.s X_1 and X_2 are not correlated then they are independent.
4. For non-Gaussian r.v.s it may be not true.

Empirical correlation

Empirical correlations of n observed pairs $(X_1, Y_1), ..., (X_n, Y_n)$ is calculated by formula

$$\rho(X,Y) = \frac{m_{xy}}{s_x s_y}, \qquad (1)$$

where

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2, \quad s_y^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2$$

are the empirical variances,

$$m_{xy} = \frac{1}{n}\left(\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i\right)$$

is empirical covariance.

Examples of dependent uncorrelated random variables

If X_1 and X_2 are independent then $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2)$

Example 1

Let $X_1 = \cos\varphi$ and $X_2 = \sin\varphi$, $\varphi \in U_{0,2\pi}$ be dependent random variables.
However, $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2)$ holds due to symmetry of their distributions regarding 0.

$$\mathbb{E}(X_1) = \int_0^{2\pi} \frac{1}{2\pi}\cos x dx = 0, \qquad \mathbb{E}(X_2) = \int_0^{2\pi} \frac{1}{2\pi}\sin x dx = 0,$$

$$\mathbb{E}(X_1X_2) = \int_0^{2\pi} \frac{1}{2\pi}\cos x \sin x dx = 0 = \mathbb{E}(X_1)\mathbb{E}(X_2)$$

RVs X_1 and X_2 are dependent but their covariance $cov(X, Y) = 0$.

Examples of dependent uncorrelated random variables

Example 2

Let X be symmetrical distributed regarding 0 and $\mathbb{E}(X) = 0$.

Let $Y = X^2 \Rightarrow \mathbb{E}(XY) = \mathbb{E}(X^3) = 0 = \mathbb{E}(X)\mathbb{E}(Y)$.

X and Y are dependent but their covariance $cov(X, Y) = 0$

Rough measures of the dependence.

Generally, the covariances and correlations cannot indicate the dependence.

1. They describe the degree to which two r.v.s. are linearly dependent:

$$\rho(X_1, X_2) \in [-1, 1].$$

2. $\rho(X_1, X_2) = \pm 1 \Leftrightarrow X_1$ and X_2 are perfectly linearly dependent, i.e. $X_2 = \alpha + \beta X_1$, for some $\alpha \in R$ and $\beta \neq 0$.

What tool could be an appropriate to indicate the dependence?

Mixing conditions as exact measures of the dependence.

Definition

(Rosenblatt, 1956) The strictly stationary ergodic sequence of random vectors X_t is strongly mixing with rate function ϕ_k for σ-field, if

$$\sup_{A\in\sigma(X_t,t\le 0),B\in\sigma(X_t,t>k)} |\mathbb{P}(A\cap B)-\mathbb{P}(A)\mathbb{P}(B)| = \phi_k \to 0, k \to \infty.^a$$

[a] $\sigma(X_t, t \le 0)$ implies σ-field of subsets X_t, $t \le 0$

The rate ϕ_k shows how fast the dependence between the past and the future decreases.
If ϕ_k decays to 0 at an exponential rate e^{-k} than (X_t) is said to be strongly mixing with geometric rate.

Other dependence measures

Another measure of dependence is given by the dependence index:

$$\beta_n = \sup_{x,y} \sum_{j=1}^{n} |f_j(x,y) - f(x)f(y)|,$$

where $f(x)$ is a marginal probability density function (PDF) of a stationary sequence $\{X_j, j = 1, 2, ...\}$,
$f_j(x,y)$ is a joint PDF of X_1 and X_{1+j}, $j = 1, 2, ...$.

1. for i.i.d. sequences $\beta_n = 0$ for all n,
2. for sequences with strong long range dependence β_n may tend to infinity, and
3. in between β_n may converge to a finite limit at various rates.

It is difficult to estimate such dependence measures by statistical tools.

The autocorrelation function (*ACF*) is

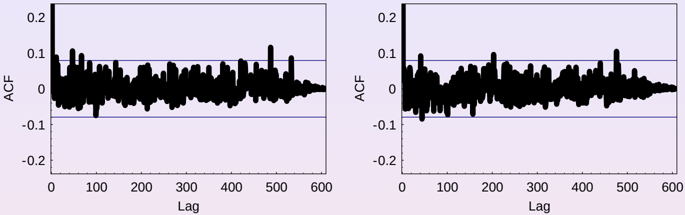
$$\rho_X(h) = \rho(X_t, X_{t+h}) = \mathbb{E}\left((X_t - \mathbb{E}(X_t))(X_{t+h} - \mathbb{E}(X_{t+h}))\right)/Var\,(X_t)$$

Let $\{X_t, t = 0, \pm 1, \pm 2, ...\}$ be a stationary sample series. The standard sample *ACF* at lag $h \in Z$ is

$$\rho_{n,X}(h) = \frac{\sum_{t=1}^{n-h}(X_t - \overline{X}_n)(X_{t+h} - \overline{X}_n)}{\sum_{t=1}^{n}(X_t - \overline{X}_n)^2},$$

where $\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$ represents the sample mean.

The accuracy of $\rho_{n,X}(h)$ may be poor if the sample size n is small or h is large. The relevance of $\rho_{n,X}(h)$ is determined by its rate of convergence to the real *ACF*. When the distribution of the X_t's is very heavy-tailed (in the sense that $\mathbb{E}X_t^4 = \infty$), this rate can be extremely slow.



Estimates of standard sample ACF of samples

of size 610 corresponding to real data: TCP-flow sizes (left), and durations (right). The dotted horizontal lines indicate 95% asymptotic confidence bounds $\pm 1.96/\sqrt{n}$.

The autocorrelation function for heavy-tailed data.

For heavy-tailed data with infinite variance it is better to use the modified sample *ACF*:

$$\widetilde{\rho}_{n,X}(h) = \frac{\sum_{t=1}^{n-h} X_t X_{t+h}}{\sum_{t=1}^{n} X_t^2},$$

i.e. without $\overline{X}_n$.

However, this estimate may behave in a very unpredictable way if one uses the class of non-linear processes in the sense that this sample *ACF* may converge in distribution to a non-degenerate r.v. depending on h.
For linear models it converges in distribution to a constant depending on h, Davis and Resnick (1985).

Confidence intervals of the sample ACF: linear processes.

The causal ARMA process (autoregressive moving average)

X_t has the representation

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \qquad t = 0, \pm 1, ..., \tag{2}$$

$Z_t \sim WN(0, \sigma^2)$ (white noise), $\{\psi_j\}$ is a sequence of real numbers providing the convergence of random series in (2), i.e., $\sum_{j=0}^{\infty} |\psi_j| < \infty$, Brockwell and Davis (1991).
$E|Z_t| \le \sigma \Rightarrow E|X_t| \le \sum_{j=-\infty}^{\infty} |\psi_j| E|Z_{t-j}| \le \sum_{j=-\infty}^{\infty} |\psi_j| \sigma < \infty$
The model ARMA(p,q) has the form
$X_t = \sum_{j=0}^{p} \theta_j X_{t-j} + \sum_{j=0}^{q} \psi_j Z_{t-j}$, $t = 1, ..., n$.

The model MA(q) has the form $X_t = \sum_{j=0}^{q} \psi_j Z_{t-j}$.

Confidence intervals of the sample ACF: linear processes.

Bartlett's formula.

If the process is linear (2), $\sum_{j=-\infty}^{\infty}|\psi_j| < \infty$ and $\mathbb{E}Z_t^4 < \infty$, then standard sample ACF $\rho_{n,X}(i)$ has the asymptotic joint normal distribution with mean $\rho_X(i)$ (i.e. ACF) and variance $var\left(\rho_{n,X}(i)\right) = c_{ii}/n$,

$$\begin{aligned} c_{ii} &= \sum_{k=-\infty}^{\infty}[\rho_X^2(k+i) + \rho_X(k-i)\rho_X(k+i) + 2\rho_X^2(i)\rho_X^2(k) \\ &- 4\rho_X(i)\rho_X(k)\rho_X(k+i)], \end{aligned} \tag{3}$$

as $n \to \infty$.

The Bartlett's formula (3) allows to check the hypothesis $\rho_{n,X}(i) = 0$.

Why Bartlett's bounds $\pm 1.96/\sqrt{n}$?

Normal distribution

$$\mathbb{P}\{-z \leq Z \leq z\} = 1 - \alpha = 0.95,$$

$$\Phi(z) = \mathbb{P}\{Z \leq z\} = 1 - \alpha/2 = 0.975, \quad z = \Phi^{-1}(0.975) = 1.96$$

Normalization $Z = \frac{\overline{x}-\mu}{\sigma/\sqrt{n}}$

$$\begin{aligned} 0.95 &= \mathbb{P}\{-z \leq Z \leq z\} = \mathbb{P}\{-1.96 \leq \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \leq 1.96\} \\ &= \mathbb{P}\{\overline{x} - 1.96\sigma/\sqrt{n} \leq \mu \leq \overline{x} + 1.96\sigma/\sqrt{n}\} \end{aligned} \tag{4}$$

Since sample ACF $\rho_{n,x}(i) \in N(\rho_x(i), c_{ii}/n)$

then $\overline{x}$ in (4) can be substituted by $\rho_{n,x}(i)$,
μ by $\rho_x(i)$ and σ by $\sqrt{c_{ii}/n}$.

Confidence intervals of the sample ACF.

Example

For i.i.d. white noise Z_t we have $\rho_Z(0) = 1$ and $\rho_Z(i) = 0$ for $i \neq 0$ (since Z_t and Z_{t+i} are independent) and $var\left(\rho_{n,Z}(i)\right) = 1/n$ by (3) since $c_{ii} = 1$.

For ARMA process driven by such a white noise Z_t

1. the sample ACF is approximately normally distributed with mean 0 and variance $1/n$ for sufficiently large n.
2. It provides 95% confidence interval with the bounds $\pm 1.96/\sqrt{n}$ for the sample ACF.
3. The hypothesis $\rho_{n,X}(i) = 0$ is accepted if $\rho_{n,X}(i)$ falls within this interval.

Exercise 11: Prove $c_{ii} = 1$ in Example.

Exercise 12: Prove $\mathbb{P}\{\rho_X - 1.96/\sqrt{n} < \rho_{n,X}(i) < \rho_X + 1.96/\sqrt{n}\}$.

Questions

What would happened if

1. the noise Z_t is not normal, and (or)
2. the process is not linear?

Confidence intervals of the heavy-tailed sample ACF $\widetilde{\rho}_{n,X}(h)$: linear processes.

ARMA process with i.i.d. regularly varying noise and tail index $0 < \alpha < 2$ (infinite variance).

$\widetilde{\rho}_{n,X}(h)$ estimates the quantity $\sum_j \psi_j \psi_{j+h} / \sum_j \psi_j^2$ that represents the autocorrelation $cov(X_0, X_h)$ in the case of a finite variance.

Illusion: heavy-tailed sample ACF $\widetilde{\rho}_{n,X}(h)$ can be applied to heavy-tailed processes without problem.

Recommendation (Resnick (2006), p.349):

- For $\alpha < 1$ use $\widetilde{\rho}_{n,X}(h)$;
- for $1 < \alpha < 2$ the classical sample ACF $\rho_{n,X}(h)$.

The calculation of confidence intervals in both cases is not easy.

Confidence intervals of the heavy-tailed sample ACF $\widetilde{\rho}_{n,X}(h)$: nonlinear processes.

GARCH(p,q) process (Mikosch (2002)):

$$X_t = \mu + \sigma_t Z_t, \qquad t \in Z,$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2,$$

- (Z_t) is an i.i.d. sequence,
- σ_t and Z_t are independent for a fixed t,
- the mean μ is estimated from the data, particularly $\mu = 0$,
- α_i and β_j are non-negative constants.

If the marginal distribution of the time series is very heavy-tailed, namely, the fourth moment is infinite, then the asymptotic normal confidence bounds for the sample ACF are not applicable anymore.

Modul 2: Lesson 4

Testing of long-range dependence.

Testing of long-range dependence (LRD).

Definition

A stationary process (X_t) is long range dependent if $\sum_{h=0}^{\infty} |\rho_X(h)| = \infty$, where $\rho_X(h)$ is the ACF at lag $h \in Z$, and short range dependent otherwise.

- This property implies that even though $\rho_X(h)$'s are individually small for large lags, their cumulative effect is important.
- To detect LRD by statistical procedures one replaces $\rho_X(h)$ by the sample ACF $\rho_{n,X}(h)$.
- The LRD effect is typical for long time series, e.g., several thousand points. One can look at lags $250, 300, 350$, etc.
- Usual short range dependent data sets would show a sample ACF dying after only a few lags and then persisting within the 95% Gaussian confidence window $\pm 1.96/\sqrt{n}$.

Testing of long-range dependence LRD: Hurst parameter

One can assume that for some constant $c_\rho > 0$

$\rho_X(h) \sim c_\rho h^{2(H-1)}$ for large h and $H \in (0.5, 1)$ (LRD case)
$\sum_{h=1}^{\infty} h^{2(H-1)}$ converges if $2(1-H) > 1$
$H = 0.5$ implies $\rho_X(h) = 0$ due to $c_\rho = 0$ (Short-range dependence)

- The constant $H \in (0.5, 1)$ is called the Hurst parameter.
- The closer H is to 1 the slower is the rate of $\rho_X(h)$ to zero as $h \to \infty$, i.e., the longer is the range of dependence in the time series.
- Kettani and Gubner (2002):

$$\hat{H}_n = 0.5\left(1 + \log_2(1 + \rho_{n,X}(1))\right)$$

if $\rho_X(h) = 0.5(|h+1|^{2H} - 2|h|^{2H} + |h-1|^{2H})$

Testing of dependence: Ljung-Box portmanteau test

It checks whether the ACF differs from zero:

Hypothesis H_0 : The data are independently distributed, i.e. $\rho_X(1) = \rho_X(2) = ... = \rho_X(h) = 0$ against
Hypothesis H_1 : The data are not independently distributed, i.e. $\rho_X(\tau) \neq 0$ for at least one $\tau \in \{1, ..., m\}$
We calculate the statistic

$$Q_{n,h} = n(n+2)\sum_{j=1}^{h} \widehat{\rho}(j)/(n-j),$$

where $\rho_{n,X}(j) = \widehat{\rho}(j)$ is sample ACF at lag j

The distribution of $Q_{n,h}$ may be approximated by

χ^2-distribution with h degree of freedom.

Testing of dependence: Ljung-Box portmanteau test

iid hypothesis is rejected at level η

if $Q_{n,h} > \chi^2_\eta(h)$, where $\chi^2_\eta(h)$ is $(1-\eta)$th quantile of χ^2-distribution with h degree of freedom, i.e.

$$\mathbb{P}\{\chi^2 > \chi^2_\eta(h)\} = \eta, \qquad 0 < \eta < 1.$$

Usually, $\eta = 0.05$.

Main assumptions

1. The variance is finite
2. The Ljung-Box test is commonly used in autoregressive integrated moving average (ARIMA) modeling

Testing of dependence: Runde's portmanteau test, Runde (1997)

Test statistic

$$Q_R = \left(\frac{n}{\ln n}\right)^{2/\alpha} \sum_{j=1}^{h} \widehat{\rho}^2(j)$$

Limit distribution of Q_R is stable, i.e. $\mathbb{P}\{Q_R > x\} \sim x^{-\alpha}$, $x > 0$

$$Q_R \to^d Q_h(\alpha)$$

Main assumptions

1. Infinite variance since $\mathbb{E}X_i^2 = \infty$: tail index $1 < \alpha < 2$
   The test is powerful for $\alpha \leq 1.6$
2. Only for symmetrically distributed rvs

Testing of dependence: Runde's portmanteau test

The hypothesis H_0 regarding independence of $\{X_i\}$ must be accepted at level η (e.g. $\eta = 0.05$) if

$$Q_R \leq Q_h(\eta)$$

Beforehand, the rv X_i must be symmetrized by $Y_i = s_i X_i$, where s_i is discrete rv that takes values 1 and -1 with probability $1/2$.

Table: Critical points of the limit distribution $Q_h(\alpha)$ for $\alpha = 1.5$, $\eta = 0.05$

Lag, h	$Q_h(0.05)$
2	13.53
3	16.32
4	18.28
5	19.17

Self-similarity

The statistical properties of time series often depend on the **time scale** of the measurement and are not self-similar over all time scales.

Definition

A continuous-time process $\{X_t, t \geq 0\}$ is self-similar with Hurst parameter H (and without loss of generality $H \in (0,1)$, called H-ss), if for any real $a > 0$ and $t \geq 0$ the scaled process $a^{-H}X_{at}$ is equal in distribution to the original process X_t, i.e. for any $n > 0$

$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} a^{-H}(X_{at_1}, \ldots, X_{at_n})$$

holds
(Beran (1994): Statistics for Long-Memory Processes, Chapman & Hall, New York).

Hurst parameter: Aggregated variance estimate

Let $\{X_i, i = 1, 2, ..., n\}$ be the original time series.

One plots $\log \widehat{Var} X^{(m)}$ versus $\log m$.

The straight line approximating the points has the slope $\beta = 2H - 2$, $-1 \leq \beta < 0$.

For this purpose

we calculate the averages within each block of the data with number $k = 1, 2, ..., [n/m]$ of size m

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i,$$

$$\widehat{Var} X^{(m)} = \frac{m}{n} \sum_{k=1}^{n/m} \left(X^{(m)}(k) \right)^2 - \left(\frac{m}{n} \sum_{k=1}^{n/m} X^{(m)}(k) \right)^2$$

The estimate of the Hurst parameter

is a slope of the plot $\log(R(l_i, r)/S(l_i, r))$ against $\log(r)$, where $i = 1, ...K$ and r denotes a range.

For this purpose, we divide the time series X^n in K intervals

of length $[n/K]$. We compute $R(l_i, r)/S(l_i, r)$ by formula

$$Q(l, r) = R(l, r)/S(l, r) = \frac{1}{S(l, r)}\left(\max_{0 \le i \le l} \mu_i(l, r) - \min_{0 \le i \le l} \mu_i(l, r)\right),$$

where

$$S(l, r) = \left(\frac{1}{r}\sum_{i=l+1}^{l+r}\left(X_i - \overline{X_{l,r}}\right)^2\right)^{1/2},$$

Continuation

$$\mu_i(l,r) = \sum_{j=1}^{i} \left(X_{l+j} - \overline{X_{l,r}} \right), \qquad \overline{X_{l,r}} = \frac{1}{r} \sum_{i=l+1}^{l+r} X_i,$$

for each lag r, starting at points $l_i = i[n/K] + 1$ such that $l_i + r \le n$.
We take the average values of the R/S statistics,

$$Q(r) = \overline{Q(l_i, r)}, \qquad i = 1, ..., K.$$

Test of self-similarity: Higuchi's method

Using a given time series $X_1, X_2, ..., X_n$,

one constructs a new time series X_k^m:

$$X_k^m : X_m, X_{m+k}, X_{m+2k}, ..., X_{m+[(n-m)/k]k},$$

$m = 1, 2, ..., k$.
Then one calculates

$$L_m(k) = \frac{n-1}{k^2[(n-m)/k]} \sum_{i=1}^{[(n-m)/k]} |X_{m+ik} - X_{m+(i-1)k}|,$$

and computes a log-log plot of the statistic $\overline{L(k)}$ (that is the average value over k sets of $L_m(k)$) versus k.
A **constant slope** D in $\overline{L(k)} \propto k^{-D}$ indicates self-similarity.

Example: Higuchi's estimate by Skype data

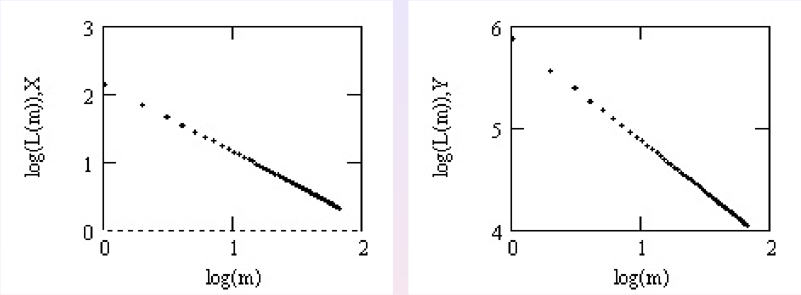


Figure: Testing the self-similarity by Higuchi's method using $\log \overline{L(k)}$ versus $\log k$ for Skype inter-arrival times (left) and packet lengths (right).

Example: Aggregated variance estimate by Skype data

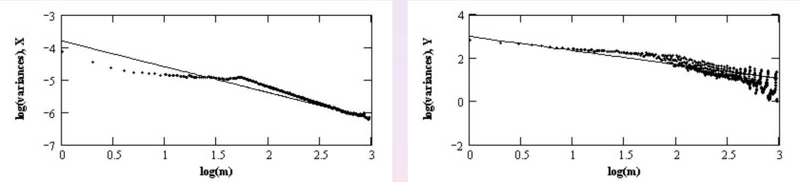
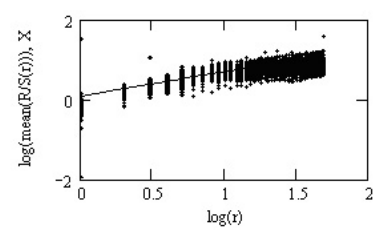


Figure: Estimation of the Hurst parameter of inter-arrival times and packets lengths of a Skype flow by the aggregated variance method (left) and (right).

Example: R/S estimate by Skype data



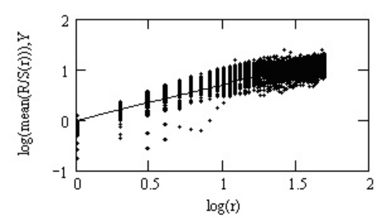


Figure: Estimation of the Hurst parameter of inter-arrival times and packets lengths of a Skype flow by the R/S method (left) and (right).

Table: Estimation of the Hurst parameter by the Skype data.

	Estimation methods		
r.v.	R/S	Higuchi's	Aggregated variance
Inter-arrival times (sec)	0.7	0.811	0.8
Packet lengths (bytes)	0.68	0.958	0.75

Modul 2: Lesson 7

Dependence detection by bivariate data.

Dependence detection by bivariate data

Measures

- Rank coefficient Kendall's τ can be estimated by the sample:

$$\rho_\tau = \frac{2S_\tau}{n(n-1)}, \qquad \text{where} \qquad S_\tau = \sum_{i=1}^{n} \sum_{j=i+1}^{n} sign\left(r_j - r_i\right),$$

  where r_i is the order number of the individuum by the second property with the number i by the first property.

- Rank coefficient Spearman's ρ can be estimated by the sample:

$$\rho_S = 1 - \frac{6S_\rho}{n^3 - n}, \qquad \text{where} \qquad S_\rho = \sum_{i=1}^{n} (r_i - i)^2$$

- Pickands dependence function $A(t)$.

ρ_τ and ρ_S can be represented by means of $A(t)$.

Pickands dependence function

Let $(X_1, Y_1), ..., (X_n, Y_n)$ be a bivariate i.i.d. random sample

with a bivariate extreme value distribution $G(x, y)$.

Example: X_1 is a TCP-flow file size, Y_1 is a duration of its transmission.

Similarly to univariate case it implies, that there exist normalizing constants $a_{j,n} > 0$ and $b_{j,n} \in R$, $j = 1, 2$ such that as $n \to \infty$,

$$\begin{aligned} & \mathbb{P}\{\left(M_{1,n} - b_{1,n}\right)/a_{1,n} \leq x, \left(M_{2,n} - b_{2,n}\right)/a_{2,n} \leq y\} \quad (5) \\ = \; & F^n\left(a_{1,n}x + b_{1,n}, a_{2,n}y + b_{2,n}\right) \to G(x, y), \end{aligned}$$

where $M_{1,n} = \max(X_1, ..., X_n)$, $M_{2,n} = \max(Y_1, ..., Y_n)$ are the component-wise maxima.
The vector $(M_{1,n}, M_{2,n})$ will in general not be presence in the original data.

Pickands dependence function.

Definition

- Let $F_1(x)$ and $F_2(y)$ be the DFs of X and Y.
- Let $G_j(x)$, $j = 1, 2$ be a univariate extreme value DF and $F_j(x)$ is in its domain of attraction.
- $G(x, y)$ may be determined by margins $G_1(x)$ and $G_2(y)$ by the representation

$$G(x,y) = \exp\left(\log\left(G_1(x)G_2(y)\right) A\left(\frac{\log\left(G_2(y)\right)}{\log\left(G_1(x)G_2(y)\right)}\right)\right),$$

where $A(t)$, $t \in [0, 1]$, is the Pickands dependence function.

Pickands dependence function: properties.

In the bivariate case the function $A(t)$ satisfies two properties:

1. $\max\{(1-t),t\} \le A(t) \le 1$, $t \in [0,1]$, i.e. $A(0)=A(1)=1$ and lies inside the triangle determined by points $(0,1)$, $(1,1)$ and $(0.5,0.5)$;
2. $A(t)$ is continuous and convex.

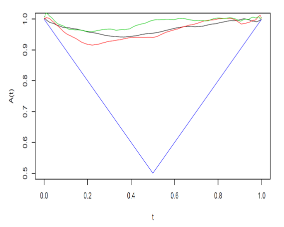
Case $A(t) \equiv 1$ corresponds to a total independence

and $A(t) = \max\{(1-t),t\} = (1-t) \vee t$ corresponds to a total dependence.

Partition $X_1,...,X_n$ and $Y_1,...,Y_n$ into r blocks.

Find block maxima within each block: $X_1^*,...,X_r^*$ and $Y_1^*,...,Y_r^*$.

Example of Pickands $A-$function and its estimation



Estimation of $A-$function

$X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_n$, $n = 10000$ are partition into r blocks of equal size $m \in \{20, 50, 100\}$, $r = [n/m]$.
Block-maxima $\{X_1^*, ..., X_r^*\}$, $\{Y_1^*, ..., Y_r^*\}$ are calculated.
Pickands A-function is estimated by Hall and Tajvidi (2000)

Fréchet distributed rvs $\{X_i\}$ and lognormal distributed rvs $\{Y_i\}$: The black ($m = 20$), red ($m = 50$), green ($m = 100$).

Fréchet distributed rvs $\{X_i\}$, $\{Y_i = 2 \cdot X_i\}$: the blue ($m = 20$) shows total dependence.

Capéraà et al. (1997):

$$\log \widehat{A}_n^C(t) = \frac{1}{r}\sum_{i=1}^{r} \log \max\left(t\hat{\xi}_i, (1-t)\hat{\eta}_i\right)$$

$$-t\frac{1}{r}\sum_{i=1}^{r} \log \hat{\xi}_i - (1-t)\frac{1}{r}\sum_{i=1}^{r} \log \hat{\eta}_i .$$

Here $\hat{\xi}_i = -\log \widehat{G}_1(X_i^*)$ and $\hat{\eta}_i = -\log \widehat{G}_2(Y_i^*)$, $i = 1, ..., n$,
$\overline{\xi}_r = r^{-1}\sum_{i=1}^{r} \hat{\xi}_i$, $\overline{\eta}_r = r^{-1}\sum_{i=1}^{r} \hat{\eta}_i$.

Hall and Tajvidi (2000):

$$\widehat{A}_n^{HT}(t) = \left((1/r)\sum_{i=1}^{r} \min\left(\frac{\hat{\xi}_i/\overline{\xi}_r}{1-t}, \frac{\hat{\eta}_i/\overline{\eta}_r}{t}\right)\right)^{-1}$$

A-estimators. Continuation.

Here, $\{X_1^*, ..., X_r^*\}$ and $\{Y_1^*, ..., Y_r^*\}$

are block-maxima of samples $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_n$, i.e. samples are partitioned into r blocks of equal size and the maxima are taken over each block.

Transformations

$-\log \widehat{G}_1(X_i^*)$, $-\log \widehat{G}_2(Y_i^*)$ lead to exponential distributed r.v.s $\hat{\xi}_i$, $\hat{\eta}_i$:

$$\mathbb{P}\{-\log G(X) \le x\} = \mathbb{P}\{G(X) \ge e^{-x}\} = 1 - e^{-x}$$

In $\widehat{A}_n^C$-estimator

$\widehat{G}_1(X_i*)$ and $\widehat{G}_2(Y_i*)$ cannot be equal to 1.

Problems of A-estimators

1. The estimators are not convex. They may be improved by taking a convex hull.
2. The margins $G_1(x)$ and $G_2(x)$ are unknown. One has to replace them by their estimates $\widehat{G}_1(x)$ and $\widehat{G}_2(x)$, e.g., by empirical DFs constructed by component-wise maxima over blocks of data:

$$\widehat{G}_1(x) = 1/r\sum_{i=1}^{r}\theta(x - X_i^*), \qquad \widehat{G}_2(y) = 1/r\sum_{i=1}^{r}\theta(y - Y_i^*),$$

where $\theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0. \end{cases}$ The amount of these maxima may be very moderate that may reflect on the accuracy of an empirical DF.

3. The component-wise maxima may be not observable together, i.e. there are such pairs of maxima which do not presence in the sample.

Independence of maxima and initial random variables

The corresponding dependence (or independence)

of the initial r.v.s X and Y follows from the dependence (or independence) of their maxima assuming that corresponding sequence of pairs (X_i, Y_i) is independent.

Really, suppose the maxima X^* and Y^* are independent. Then

$$P\{X^* \leq x, Y^* \leq y\} = P\{X^* \leq x\}P\{Y^* \leq y\}.$$

Then for i.i.d. pairs (X_i, Y_i), $i = 1, ..., n$ we have

$$P\{X^* \leq x\}P\{Y^* \leq y\} = P^n\{X_i \leq x\}P^n\{Y_i \leq y\} = P^n\{X_i \leq x, Y_i \leq y$$

From here the independence of X_i and Y_i follows.
Suppose now that the maxima X^* and Y^* of r.v.s X and Y are dependent. Then evidently X and Y cannot be independent.

Bivariate quantile curve

Both estimates of $A(t)$ allow to get the estimate $\widehat{G}(x,y)$ and construct bivariate quantile curves

$$Q(\hat{G},p) = \{(x,y) : \hat{G}(x,y) = p\}, \qquad 0 < p < 1, \tag{6}$$

$$\hat{G}(x,y) \approx \exp\left(\log\left(\hat{G}_1(x)\hat{G}_2(y)\right)\hat{A}\left(\frac{\log\left(\hat{G}_2(y)\right)}{\log\left(\hat{G}_1(x)\hat{G}_2(y)\right)}\right)\right),$$

assuming that $\widehat{G}_1(x) = p^{(1-w)/\hat{A}(w)}$ and $\widehat{G}_2(x) = p^{w/\hat{A}(w)}$ for some $w \in [0,1]$ in order to get $\widehat{G}(x,y) = p$, Beirlant et al. (2004).
The quantile curve consists of the points

$$Q(\hat{G},p) = \{\left(\widehat{G}_1^{-1}(p^{(1-w)/\hat{A}(w)}), \widehat{G}_2^{-1}(p^{w/\hat{A}(w)})\right) : w \in [0,1]\}.$$

Why we need to detect dependence? Examples.

Example 1:

Maximum likelihood method can be used only if stationary random variables $X_1, ..., X_n$ are independent.
Then one can use $\prod_{i=1}^{n} f(X_i)$, where $f(X_i)$ is a density at X_i.

Otherwise $f(X_1, ..., X_n) \neq f(X_1) \cdot ... \cdot f(X_n)$.

Example 2:

Empirical distribution function $F_n(x) = 1/n \sum_{i=1}^{n} \theta(x - X_i)$, $\theta(x) = 1$ if $x \geq 0$ and $\theta(x) = 0$ if $x < 0$ is consistent unbiased estimate of the distribution function $F(x)$, i.e.

$$\mathbb{E}F_n(x) = F(x), VarF_n(x) = 1/nF(x)(1-F(x)) \to 0, \qquad n \to \infty$$

if $X_1, ..., X_n$ are independent.

Why we need to detect heavy tails? Example.

Estimation of expectation if $\mathbb{E}X^2 < \infty$

How to estimate the expectation $\mathbb{E}X$ by average?

Chebyshev's inequality requires a finite variance σ^2 of random variable X_i:

$$\mathbb{P}\{|1/n\sum_{i=1}^{n} X_i - \mathbb{E}X| \geq \frac{\sigma\chi}{\sqrt{n}}\} \leq 1/\chi^2$$

Central Limit Theorem (CLT): Sequence

$$\{\sqrt{n}(\overline{X_n} - \mu)/\sigma, n \geq 1\} \to^P N(0,1),$$

if $\mathbb{E}X^2 < \infty$. If $\alpha \in (1,2)$, i.e. $\mathbb{E}X_1 < \infty$, but $\mathbb{E}X_1^j = \infty, j \geq 2$ then CLT is not valid.

Why we need to detect heavy tails? Example.

Estimation of expectation if $\mathbb{E}X^2 = \infty$

Expectation $\mu = \int_0^\infty x dF(x)$

Alternative: $\mu = \int_0^1 Q(1-s)ds$,

where $Q(s) = \inf\{x : F(x) \geq s\}$, $0 < s < 1$ is quantile function

Sample estimate of $Q(s)$:

$$Q_n(s) = \inf\{x \in R : F_n(x) \geq s\},$$

$$\int_0^1 Q_n(1-s)ds = 1/n \sum_{i=1}^{n} X_i = \overline{X_n}$$

Estimation of quantile function

$$\widehat{Q_n}(1-s) \quad = \quad \begin{cases} Q_n^W(1-s), & \text{for} \qquad 0 < s < k/n; \\ Q_n(1-s), & \text{for} \qquad k/n \le s < 1, \end{cases}$$

Q_n^W is some estimate of high quantile.

Example (Peng 2001): Weissman's estimator

$$Q_n^W(1-s) = (\frac{k}{n})^{1/\widehat{\alpha}^H} X_{(n-k)} s^{-1/\widehat{\alpha}^H},$$

$s \searrow 0$,

$\widehat{\alpha}^H$ is Hill's estimator

Estimation of expectation if $\mathbb{E}X^2 = \infty$. Continuation.

Estimation of expectation

$$
\begin{aligned}
\widehat{\mu}_n^P &= \int_0^{k/n} Q_n^W(1-s)ds + \int_{k/n}^1 Q_n(1-s)ds \\
&= \frac{k}{n} \cdot \frac{\widehat{\alpha}^H}{\widehat{\alpha}^H - 1} \cdot X_{(n-k)} + \frac{1}{n} \sum_{i=k+1}^{n} X_{(n-i+1)}
\end{aligned}
$$

Then

$$
\frac{\sqrt{n}(\widehat{\mu}_n^P - \mu)}{\sqrt{k/n}X_{(n-k)}} \to^d N(0, \sigma^2(\alpha))
$$

as $n \to \infty$.