Procjena repa i mjere rizika

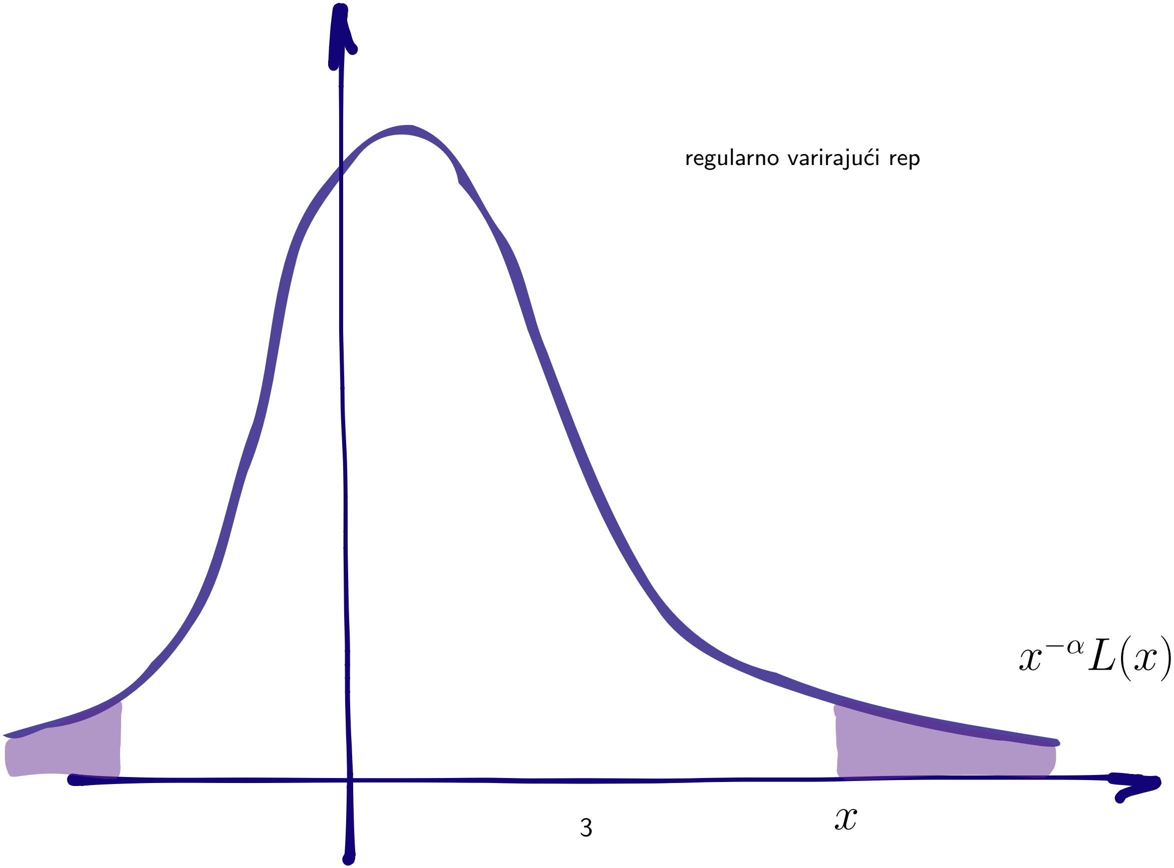
Bojan Basrak, PMF–MO Zagreb

Financijski praktikum

6. travnja 2016.

Teški repovi

Regularna varijacija



Rep razdiobe F zovemo **regularno varirajućim** ako vrijedi

$$\overline{F}(x) = P(X > x) = \frac{L(x)}{x^\alpha}, \quad x > 0\,,$$

gdje je $\alpha > 0$ tzv. repni indeks a L je sporo promjenjiva funkcija, odn. za nju vrijedi

$$\lim_{x\to\infty} \frac{L(cx)}{L(x)} = 1\,,\ \text{za sve } c > 0\,.$$

To su razdiobe vrlo teškog repa popularne u aktuarstvu i financijama, a vrlo važne i u teoriji vjerojatnosti. Poznato je za ovakvu sl. varijablu X ako je npr. nenegativna

$$EX^\delta = \begin{cases} +\infty & \delta > \alpha\,, \\ < \infty & \delta < \alpha\,. \end{cases}$$

Primjer

i) Pareto razdioba

$$\overline{F}(x) = \left(\frac{\kappa}{\kappa + x} \right)^{\alpha},$$

za $x > 0$ i neke parametre $\alpha, \kappa > 0$. Jednoparametarska Pareto razdioba je $\overline{F}(x) = x^{-\alpha},\ x > 1$ (to je zapravo log eksponencijalna razdioba).

ii) Fréchetova razdioba

$$\Phi(x) = e^{-x^{-\alpha}},$$

za $x > 0$.

Uvedimo

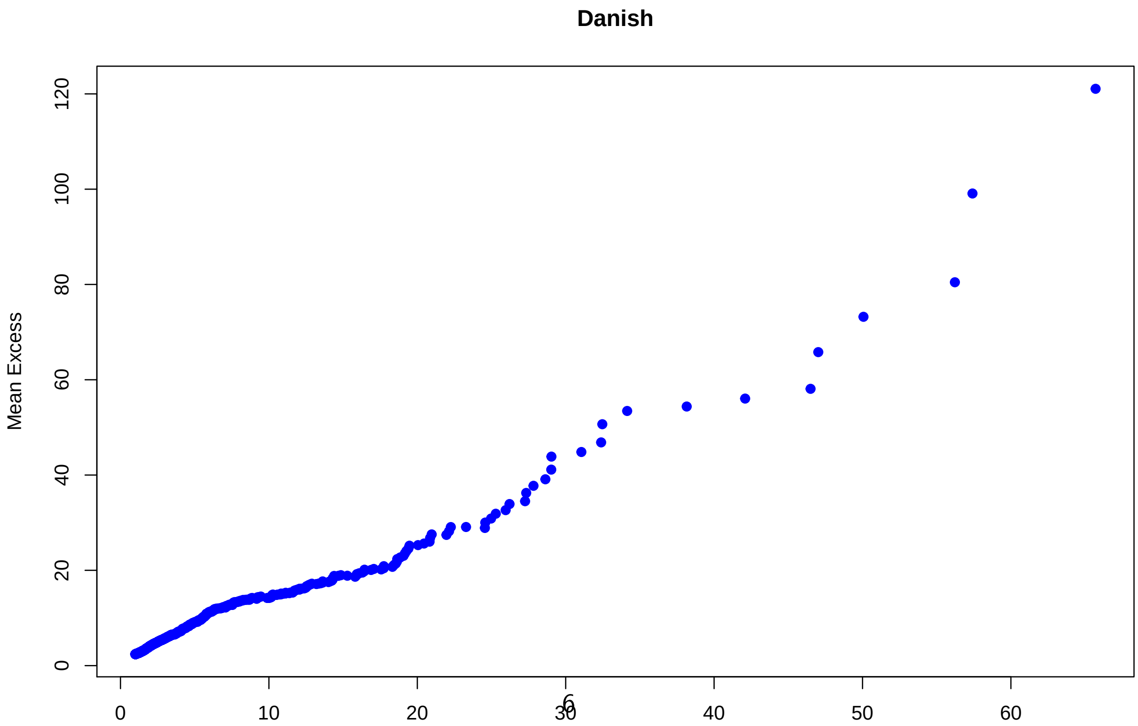
$$M_n = \max\{X_1, \dots, X_n\},$$

ako su X_i njd i reg. var. tada za $a_n = F^{\leftarrow}(1 - n^{-1})$ vrijedi

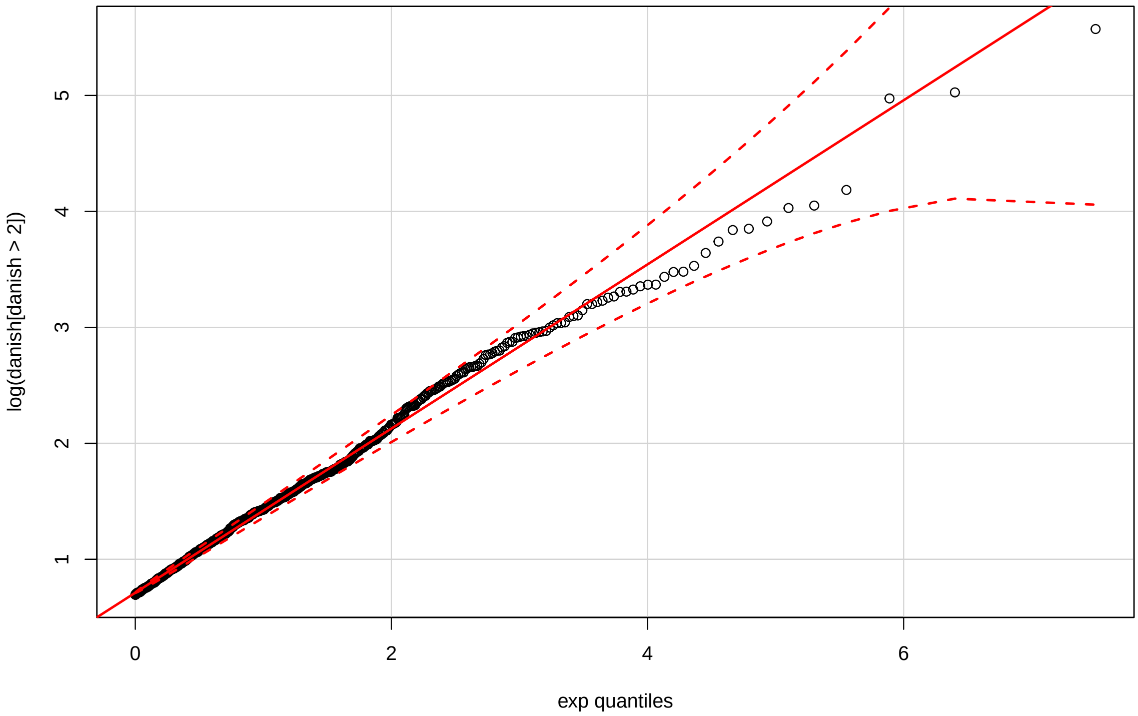
$$\frac{M_n}{a_n} \xrightarrow{d} \Phi_\alpha .$$

```r
library(evir)
data(danish)
help(danish)

meplot(danish,main="Danish",pch=19,col=4)
```

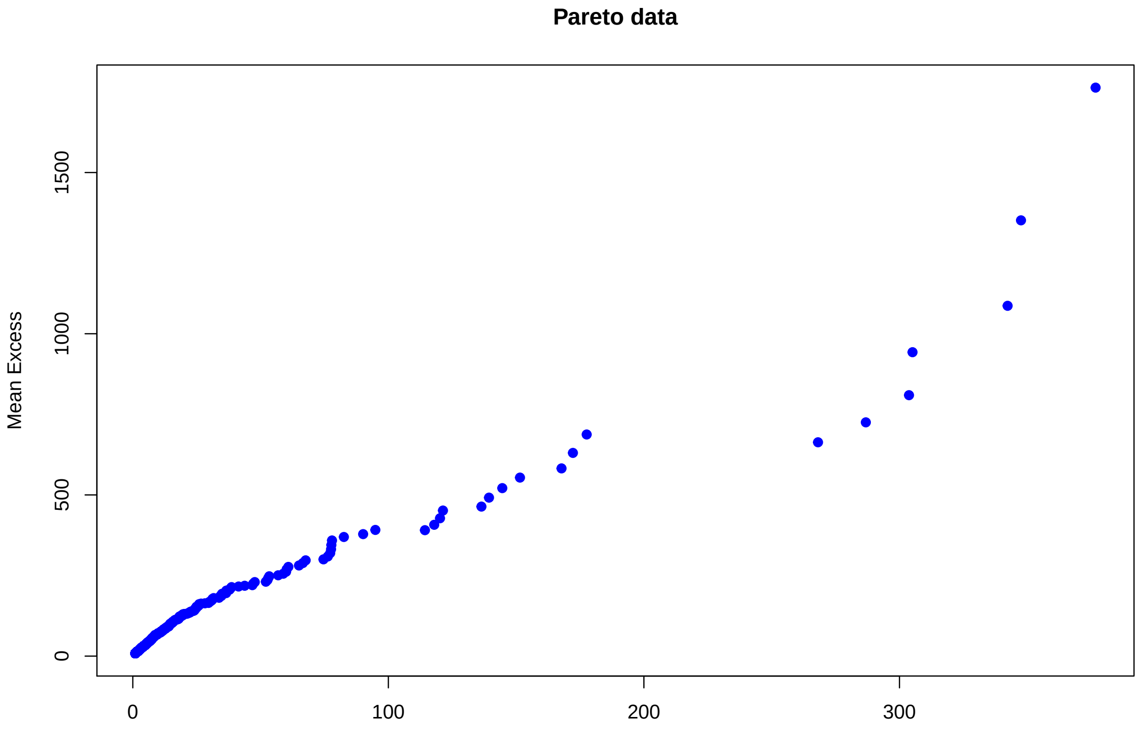


```r
library(car)
qqPlot(log(danish[danish>2]), dist="exp")
```

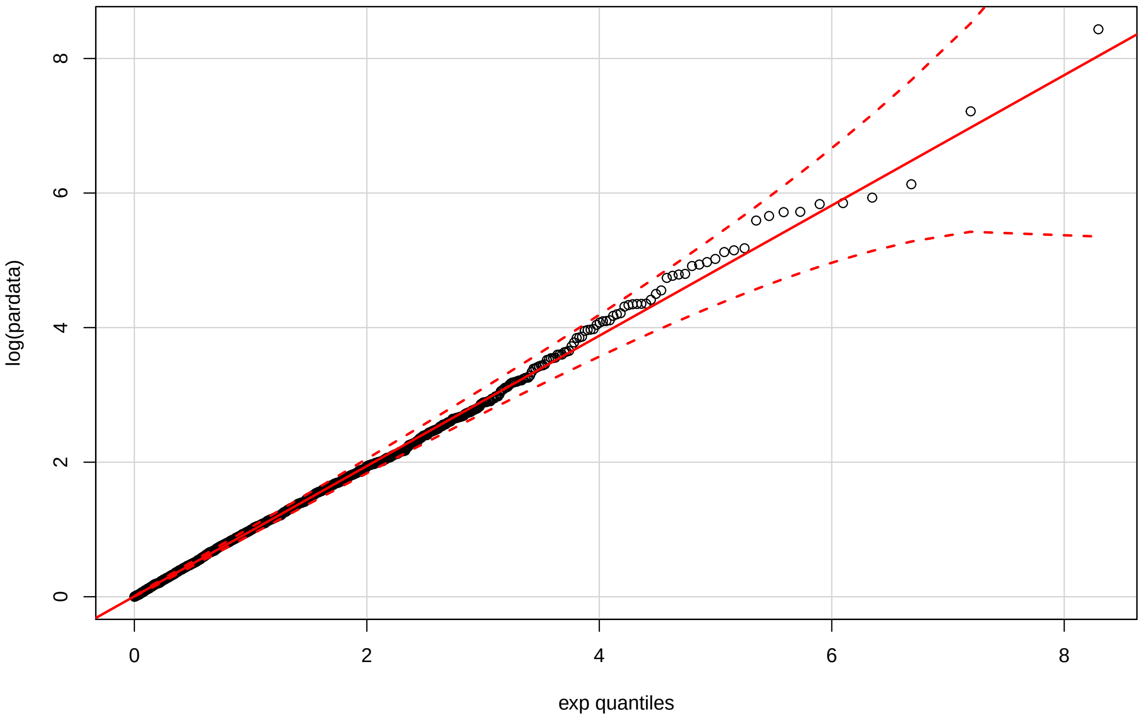


***

```r
U <- runif(2000)
pardata <- 1/(1-U) # ~ F(x) = 1-x^{-1}, x > 1
meplot(pardata,main="Pareto data",pch=19,col=4)
```



```r
qqPlot(log(pardata), dist="exp")
```



Reg. var. razdiobe mogu imati neobičan limes parcijalnih suma $S_n = X_1 + \cdots + X_n$, posebno c.g.t. ne vrijedi ako je $\alpha < 2$.

Asimptotsko repno ponašanje ukupne sume (npr. gubitaka ili šteta) određuje najveća, tj. za $x \to \infty$

$$P(S_n > x) = n\overline{F}(x)(1 + o(1))\,.$$

Primjer (nastavak)

iii) Burrova razdioba

$$\overline{F}(x) = \left(\frac{\kappa}{\kappa + x^\tau}\right)^\alpha\,,$$

za $x > 0$ i neke parametre $\alpha, \kappa, \tau > 0$.

iv) log-gamma, $X = e^Y$, $Y \sim \Gamma(\alpha, \beta)$

Subeksponencijalne razdiobe

Subeksponencijalne razdiobe su razdiobe koje za njd niz (X_i) zadovoljavaju

$$P(S_n > x) = P(M_n > x)(1 + o(1)) \,,$$

za $x \to \infty$ i sve $n \geq 2$ (dovoljno je zapravo i samo $n = 2$).

Uočite

$$P(M_n > x) = n\overline{F}(x)(1 + o(1)) \,,$$

što dokazuje da su sve reg.var. razdiobe subeksponecijalne.

Za subeksponencijalne razdiobe vrijedi i

$$\lim_{x\to\infty} \frac{\overline{F}(x+y)}{\overline{F}(x)} = 1$$

za sve y.

Za sve $\varepsilon > 0$ i $x \to \infty$ vrijedi

$$e^{\varepsilon x}\overline{F}(x) \to \infty\,.$$

Tako da je za nenegativnu subeksponencijalnu X funkcija izvodnica momenata

$$m_X(t) = Ee^{tX} = \infty\,, \quad \text{za sve } t > 0\,.$$

Primjer

Weibullova razdioba

$$\overline{F}(x) = e^{-cx^{\tau}}\,,$$

za $x > 0$ i neke parametre $c, \tau > 0$ je subeksponencijalna za $\tau \in (0, 1)$ iako nije reg. varirajuća.

Procjena repa i kvantila

Neparametarska procjena

Pretpostavimo da želimo odrediti rep razdiobe

$$P(X > x) \quad \text{za "velike" } x \quad \text{ili kvantile}$$

$$F^{\leftarrow}(1-p) \quad \text{za "male" } p\,.$$

I rep i kvantile možemo bez dodatnih pretpostavki procijeniti iz empirijske funkcije distribucije (no to je vrlo problematično na rubu empirijske razdiobe).

Intervale pouzdanosti možemo napraviti koristeći asimptotsku normalnost, ili što je realnije u praksi bootstrap metode.

```r
quantile(danish,0.99)
```

```
##      99%
## 26.04253
```

```r
bootThetaQuantile <- function(x,i) {
quantile(x[i], probs=.99)  ### moze i drugi kvantil
}
library(boot)
```

```
##
## Attaching package: 'boot'
```

```
## The following object is masked from 'package:car':
##
##     logit
```

```r
nboot <- 4000
theta.boot.VaR <- boot(danish, bootThetaQuantile, R=nboot)
```

```r
boot.ci(theta.boot.VaR, conf=0.95,type=c("norm","basic","perc","
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 4000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = theta.boot.VaR, conf = 0.95, type = c("nor
##     "basic", "perc", "bca"))
##
## Intervals :
## Level      Normal              Basic
## 95%   (21.12, 31.14 )   (20.58, 31.22 )
##
## Level     Percentile            BCa
## 95%   (20.86, 31.51 )   (20.90, 31.51 )
## Calculations and Intervals on Original Scale
```

```r
quantile(pardata,0.99)
```

```
##      99%
## 114.3508
```

```r
bootThetaQuantile <- function(x,i) {
quantile(x[i], probs=.99)  ### moze i drugi kvantil
}
library(boot)
nboot <- 4000
theta.boot.VaR <- boot(pardata, bootThetaQuantile, R=nboot)
```

```r
boot.ci(theta.boot.VaR, conf=0.95,type=c("norm","basic","perc","
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 4000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = theta.boot.VaR, conf = 0.95, type = c("nor
##     "basic", "perc", "bca"))
##
## Intervals :
## Level      Normal              Basic
## 95%   ( 69.0, 174.4 )   ( 60.9, 161.0 )
##
## Level     Percentile            BCa
## 95%   ( 67.7, 167.8 )   ( 67.7, 167.8 )
## Calculations and Intervals on Original Scale
```

Parametarska procjena

Ako možemo opravdati pretpostavku (teorijski ili na osnovu grafa kvantila npr.) da podaci dolaze iz neke parametarske familije razdioba, možemo prvo procijeniti parametre nepoznate razdiobe i nakon toga odrediti kvantile odn. rep.

Intervale pouzdanosti možemo odrediti tzv. *parametarskom bootstrap* metodom: ako razdioba $F = F_\theta$ ovisi o nepoznatom parametru θ kojeg znamo procijeniti recimo sa $\hat{\theta}$, bootstrap uzorke za $i = 1, 2, 3, \ldots, B$

$$X^*_{i,1}, \ldots, X^*_{i,n}\,,$$

simuliramo iz razdiobe $F_{\hat{\theta}}$. A zatim postupimo kao i u običnom boostrapu.

```r
n <- length(pardata)
procjena <- mean(log(pardata))
procjena
```

```
## [1] 0.9795324
```

```r
proc.kvant<-function(y) ### pretpostavka podaci su Pareto
  { lamhat<- mean(log(y))
  out<-qexp(0.99,lamhat)
  exp(out)}
x <- exp(rexp(n*1000, 1/procjena))
nboot<- 1000
bootstrap.uzorak = matrix(x, nrow=n, ncol=nboot)
vars.boot<-apply(bootstrap.uzorak, 1, proc.kvant)
proc.kvant(pardata)
```

```
## [1] 110.1008
```

```r
quantile(vars.boot,c(0.025,0.975)) ### percentilni CI
```

```
##      2.5%     97.5%
##  84.18468 153.09568
```

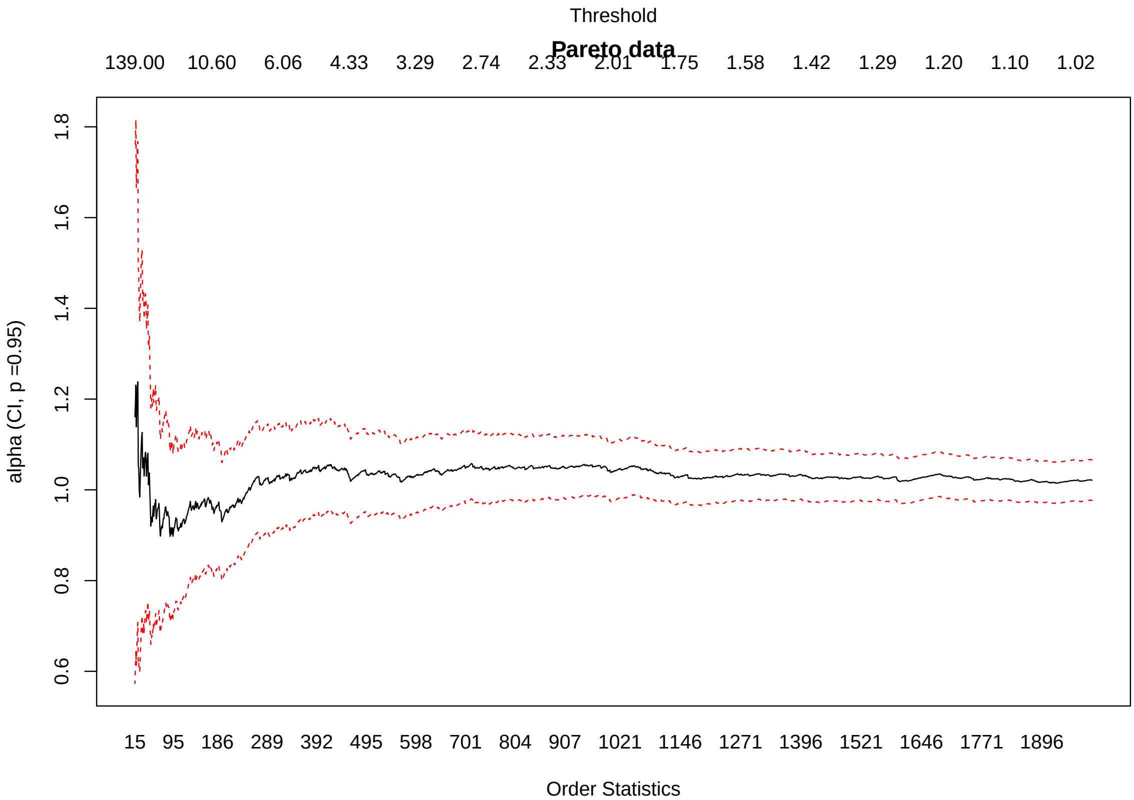
Procjena repa za regularno varirajuće razdiobe

Pretpostavka o reg.varijaciji može se smatrati poluparametarskom jer uz razdiobu određuju parametar α i nepoznata funkcija L.

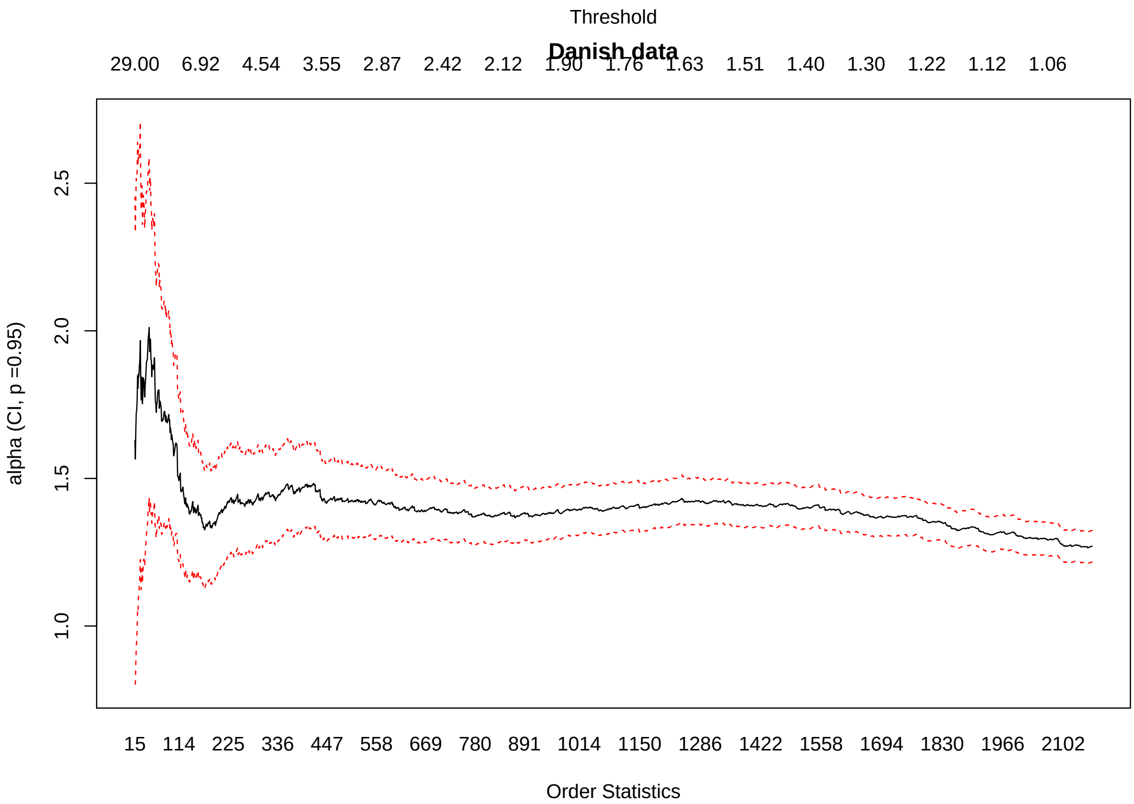
Prisjetmo se uređajnih statistika $X_{(1)} \leq \cdots \leq X_{(n)}$. Koristeći k najvećih vrijednosti u uzorku možemo procijeniti repni indeks α tzv. **Hillovim procjeniteljem**

$$\hat{\alpha}_{n,k} = \left(\frac{1}{k} \sum_{j=1}^{k} \log \frac{X_{(n-j+1)}}{X_{(n-k+1)}} \right)^{-1}$$

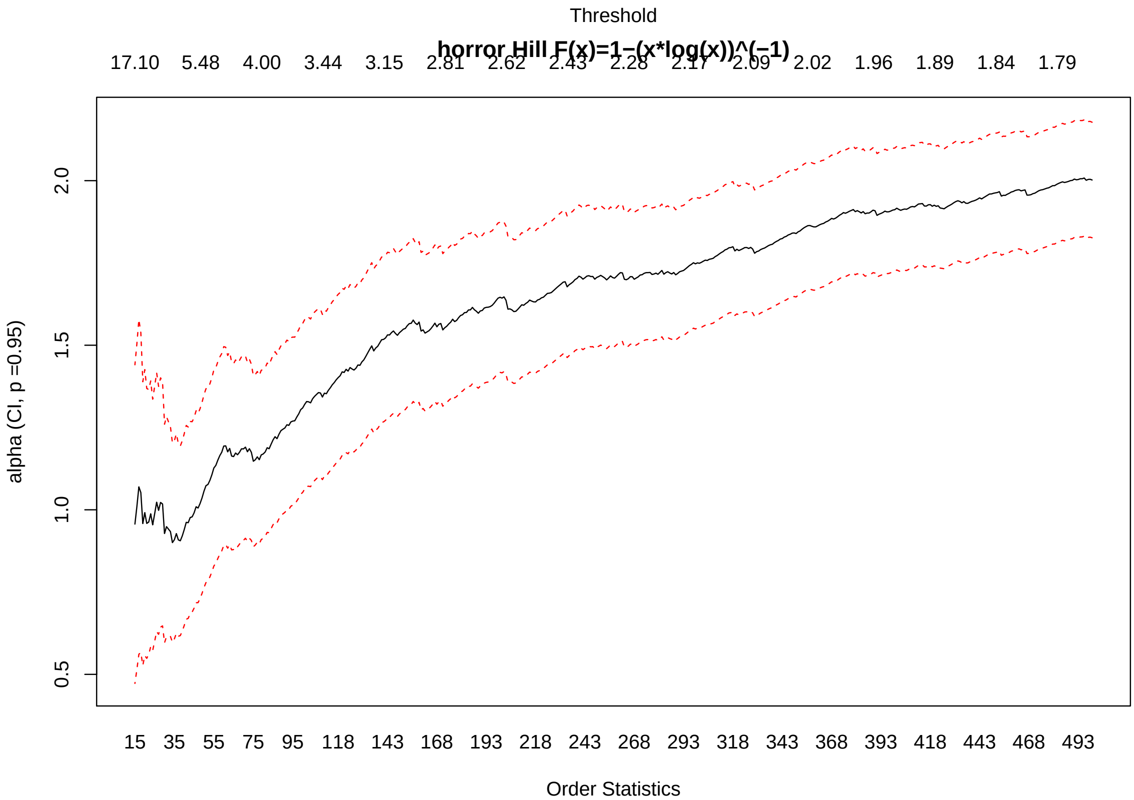
```r
hill(pardata,main="Pareto data")
```



```r
hill(danish,main="Danish data")
```



```r
hill(somedata,main="horror Hill F(x)=1-(x*log(x))^(-1)")
```



Procjena repa je sad

$$(\overline{F}(x))^{\wedge} = \frac{k}{n}\left(X_{(n-k+1)}\right)^{\hat{\alpha}} x^{-\hat{\alpha}}$$

A kvantile možemo procijeniti koristeći

$$(F^{\leftarrow}(1-p))^{\wedge} = \left(\frac{n}{k}p\right)^{-1/\hat{\alpha}} X_{(n-k+1)}$$

Odrediti optimalni k je iznimno teško čak i za njd podatke. Obično biramo takav k u čijoj se blizini procjenitelj $\hat{\alpha}_{n,k}$ ponaša "stabilno".

Uz dodatne tehničke uvjete, procjenitelj $\hat{\alpha}_{n,k}$ je asimptotski normalan, što omogućava npr. i izradu intervala pouzdanosti za repni indeks, repne vjerojatnosti i kvantile, no i njih u praksi treba koristiti oprezno.

```
## Graph may look strange !!
##
## Suggestion 1: Increase `p' above 0.99308
## Suggestion 2: Increase `start' above  22
```

```
## Graph may look strange !!
##
## Suggestion 1: Increase `p' above 0.9925
## Suggestion 2: Increase `start' above  21
```

Mjere rizika

Pretpostavimo da sl. varijabla Y modelira gubitak osiguravatelja ili gubitak investicijskog fonda u nekom periodu.
Razumno je pretpostaviti $EY < 0$. Pitanje je kako mjeriti rizik vezan uz Y. Postoji više mogućnosti:

Varijanca ili alternativno stand. devijacija - imaju niz nedostataka: ne mogu se definirati za neke razdiobe, simetrično tretiraju repove, ne detektiraju diversifikaciju na pravilan način,...

VaR (Value at Risk) - postoji uvijek, definira se jednostavno, za dani $\alpha \in (0,1)$ definira se kao

$$\mathrm{VaR}_\alpha(Y) = F_Y^{\leftarrow}(1-\alpha)$$

To je najčešće korištena mjera rizika, što objašnjava važnost procjene kvantila u financijskoj praksi. No i ona ima nedostataka, npr. moguće je

$$\mathrm{VaR}_\alpha(Y_1 + Y_2) > \mathrm{VaR}_\alpha(Y_1) + \mathrm{VaR}_\alpha(Y_2)\,,$$

dakle diverzifikacijom dobijemo veći rizik nego podijelom portfelja – problem za regulatora i risk managera. Dakle VaR nije uvijek razumna mjera rizika.

Očekivani manjak (expected shortfall ili mean excess loss) je treća opcija koja nema problema s diverzifikacijom i predstavlja sve češće korišten izbor u praksi.

$$e_Y(u_\alpha) = E(Y - u_\alpha | Y > u_\alpha) \quad \text{gdje } u_\alpha = \mathrm{VaR}_\alpha(Y)\,.$$

Bez obzira na korištenu mjeru rizika, u praksi je nužno provjeriti i kvalitetu metoda za njihovu procjenu – u tu svrhu obično koristimo *backtesting* procedure. .

Vrlo složen model s puno parametara može čak i opisati podatke vrlo dobro, no takav rezultat nije uvijek moguće proširiti na nove/nekorištene podatke, naime veliki broj parametara izaziva problem koji se naziva *overfitting*, a teorijski se može objasniti uobičajenim rastavom statističke greške na pristranost i varijancu.

John von Neumann:

With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.