Analysis methods of heavy-tailed data

Natalia Markovich

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

Exercises

Theoretical exercises to Modul 1, Lesson 1

Exercise 1:

1. Prove that normal distribution is not heavy-tailed.
2. Prove that uniform distribution is not heavy-tailed.
3. Prove that exponential distribution is not heavy-tailed.
4. Prove that lognormal distribution is heavy-tailed.
5. Check when Weibull distribution is heavy-tailed or light-tailed.

Exercise 2:

If $X > 0$ has a Φ_α distribution

- $\log X^\alpha$ has distribution Λ
- $-X^{-1}$ has distribution Ψ_α

Prove these transformations

Theoretical exercises to Modul 1, Lesson 2

Exercise 3:

1. Prove that $\ell(x) = \ln(\ln x)$ is slowly varying function
2. Is $\ell(x) = -1$ slowly varying function or not?
3. Is $\ell(x) = 0$ slowly varying function or not?
4. Is $\ell(x) = x$, $x > 0$ slowly varying function or not?
5. Is $\ell(x) = 1/x$, $x > 0$ slowly varying function or not?
6. Let $\ell_1(x)$ and $\ell_2(x)$ vary slowly. Are $\ell_1(x)\ell_2(x)$, $\ell_1(x) + \ell_2(x)$ slowly varying or not?

Exercise 4:

For heavy-tailed distributions the following relation

$$\overline{F(x)} \gg f(x) \gg f'(x) \gg ...$$

is valid. For light-tailed distributions $\overline{F(x)}, f(x), f'(x), ...$ have the same magnitude.
Compare values $\overline{F(x)}, f(x), f'(x), ...$

1. for Pareto distribution;
2. for exponential distribution;
3. for normal distribution.

$$\overline{F}(x) = 1 - F(x) = x^{-1/\sqrt{\log x}}, \qquad x > 1 \tag{1}$$

$$\overline{F}(x) = (\log x)^{-\beta}, \qquad x \geq e,\ \beta > 0 \tag{2}$$

Exercise 5:

Prove that (1) and (2) are super-heavy-tailed, i.e. $\lim_{t\to\infty} \overline{F}(tx)/\overline{F}(t) = 1$ for any $x > 0$ is fulfilled.

Exercise 6:

Find other examples of super-heavy tailed distributions.

Exercise 7:

1. Let X be Weibull distributed with shape parameter $\beta \geq 1$. Will $Y = \exp X$ be light-, heavy- or super-heavy-tailed distributed?
2. Let X be exponential distributed. Will $Y = \exp X$ be light-, heavy- or super-heavy-tailed distributed?
3. Let X be Pareto distributed. Will $Y = \exp X$ be light-, heavy- or super-heavy-tailed distributed?
4. Let X be uniform distributed. Will $Y = \exp X$ be light-, heavy- or super-heavy-tailed distributed?
5. Let X be lognormal distributed. Will $Y = \ln X$ be light-, heavy- or super-heavy-tailed distributed?
6. Let X be Frechet distributed. Will $Y = \ln X$ be light-, heavy- or super-heavy-tailed distributed?

Exercise 8:

If X is Weibull distributed with shape parameter $\tau \geq 1$. Will $Y = \exp X$ be heavy- or super-heavy-tailed distributed?

Theoretical exercises to Modul 1, Lesson 3

Exercise 9: confidence interval of group estimate

Using (3) obtain the confidence interval of α.

$$l\left(l^{-1}\sum_{i=1}^{l}k_{l,i}-\alpha(1+\alpha)^{-1}\right)\left(\sum_{i=1}^{l}k_{l,i}^2-l^{-1}\sum_{i=1}^{l}k_{l,i}\right)^{-1/2}\to^d N(0,1) \quad (3)$$

Exercise 10:

Recursiveness of group estimate $z_l = (1/l)\sum_{i=1}^{l} k_{l,i}$.

Assuming the process is strongly stationary (i.e. expectation $Ek_{j,i} = c$ for any i, j) prove that the bias of z_{l+i} is the same as for z_l but variance is less for uncorrelated $\{k_{l,i}\}$.

$$bias(z_{l+i}) = bias(z_l), \qquad var(z_{l+i}) = var(z_l)l/(l+i), \tag{4}$$

$$var(z_{l+i}) < var(z_l) \qquad \text{for} \qquad \forall i > 0$$

Prove (4).

Theoretical exercises to Modul 1, Lesson 4

Exercise 11:

Mean excess function is determined by

$$e(u) = \mathbb{E}(X - u | X > u), \qquad 0 \le u < x_F \le \infty,$$

x_F is end-point of the distribution.
If X is a positive r.v. with df $F(x)$ and $EX < \infty$, then

$$e(u) \;=\; \int_u^{x_F} (x - u) dF(x) / \overline{F}(u) = \frac{\int_u^{x_F} \overline{F}(x) dx}{\overline{F}(u)}$$

$F_u(x) = P\{X - u \le x | X > u\}$, $x \ge 0$ is the excess df of the r.v. X over the threshold u.

1. Find $e(u)$ for normal distribution.
2. Find $e(u)$ for exponential distribution.
3. Find $e(u)$ for Pareto distribution.
4. Find $e(u)$ for Weibull distribution.

Exercise 12:

Prove the formula for the mean squared error

$$MSE(\hat{\gamma}) = bias^2 + variance$$

for any estimate $\hat{\gamma}$.