Robust inference based on flexible parametric families of distributions

Adelchi Azzalini

(Università di Padova, Italia)

ICORS, Parma, June 2009

Outline of the talk

- skew-symmetric families of distributions
- flexible likelihood for robust inference
- some numerical comparison

Skew-symmetric distributions — Introduction

A generator of distributions

- context: families of continuous distributions on $\mathbb{R}^d$
- start from a density f_0 symmetric around 0,

$$f_0(x) = f_0(-x) \qquad\qquad (x \in \mathbb{R}^d)$$

- choose a real-valued $w(x)$ such that $w(-x) = -w(x)$
- choose a scalar cdf $G(\cdot)$ with symmetric pdf $G'(\cdot)$
- then

$$f(x) = 2\, f_0(x)\, G\{w(x)\}$$

  is a *skew-symmetric* pdf

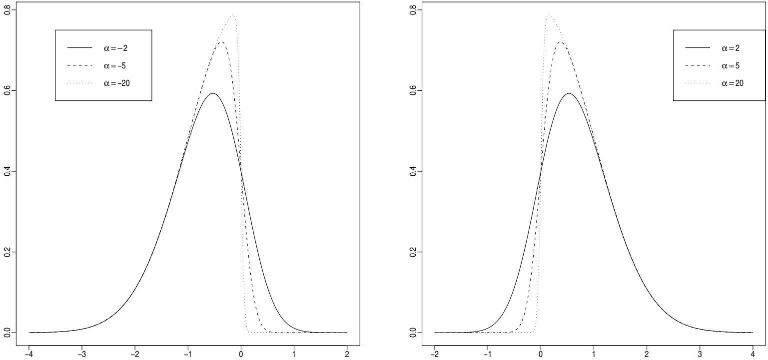
Basic case: skew-normal distribution ($d = 1$)

Choose $N(0,1)$ ingredients:

$$f_0(x) = \varphi(x), \qquad G = \Phi, \qquad w(x) = \alpha\, x$$

and get

$$f(x) = 2\,\varphi(x)\,\Phi(\alpha x)$$



Regulate both skewness and kurtosis

Select f_0 from a symmetric family with adjustable tails.

Interesting cases:

- Exponential power (Subbotin, 1923):

$$f_0(x) \propto \exp\left(-\frac{\|x\|_\Omega^\nu}{\nu}\right)$$

- Student's t:

$$f_0(x) \propto \left(1 + \frac{\|x\|_\Omega^2}{\nu}\right)^{-\frac{\nu+d}{2}}$$

In both cases ν regulates the tail thickness

Various options for the skewing factor

Skew-t distribution (case $d = 1$)

- let $Z \sim$ Skew-normal(α)
- then a natural form of skew-t (ST) variate is

$$X = \frac{Z}{\sqrt{\chi^2_\nu/\nu}}$$

- density is

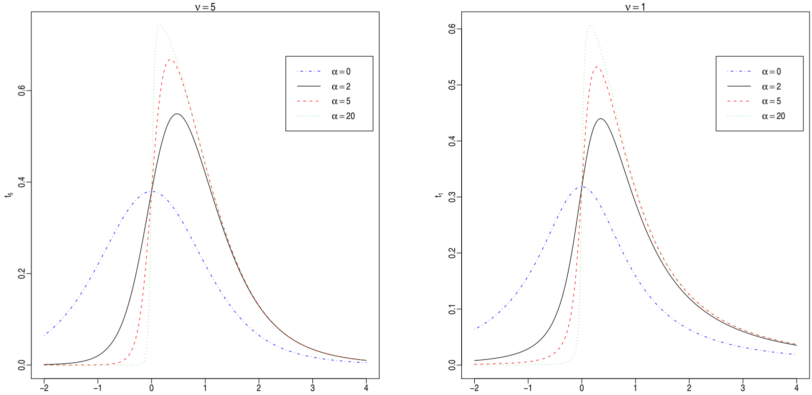
$$f(x) = 2\, t_\nu(x)\, T_{\nu+1}\{w(x)\}$$

  where

$$w(x) = \alpha x \sqrt{\frac{\nu+1}{\nu+x^2}}$$

- Note: $f(x)$ is of skew-symmetric type
- Note: a multivariate version exists

Skew-t distribution: example of densities



A flexible distribution

- Consider ST has a general-purpose tool for statistical modelling
- Combines high flexibility for skewness and for the tails:
  α regulates skewness ($\alpha \in \mathbb{R}^d$),
  ν regulates the tail thickness ($\nu > 0$)
- Make use of the tail parameter to accomodate "outliers", possibly non-symmetrically distributed
- (Ideal in *d*-dimensional case: a tail parameter for each component)

Regression models with ST errors

- fitted model:

$$y = x^\top \beta + \varepsilon, \qquad \varepsilon \sim \text{(scale factor)} \times \text{ST}$$

- estimate parameters via MLE
  (or Bayesian approach, according to taste)
- adjust intercept because $\mathbb{E}\{\text{ST}\} \neq 0$
  various options:
  - intercept $= \hat{\beta}_0 + \mathbb{E}\{\varepsilon\}$ ...needs $\hat{\nu} > 1$
  - intercept $= \hat{\beta}_0 + \mathit{median}(\varepsilon)$ ...use this
  - others...

Flexible distribution approach vs M-estimation

- M-estimates converge to solution of non-linear equation:

$$\lambda(\theta) := \mathbb{E}\{\psi(X, \theta)\} = 0$$

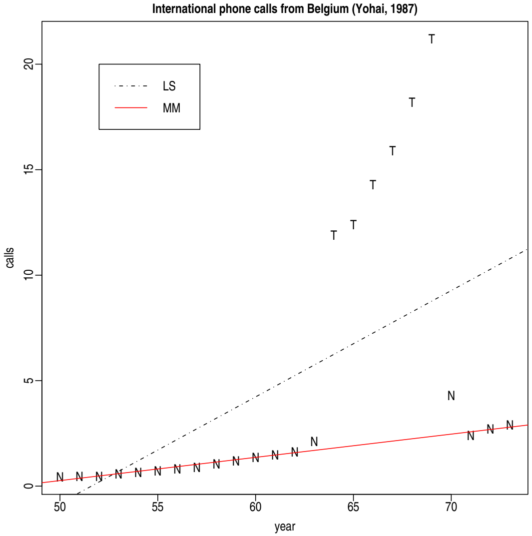
- In simple location case

$$\lambda(\theta) := \mathbb{E}\{\psi(X - \theta)\} = 0$$

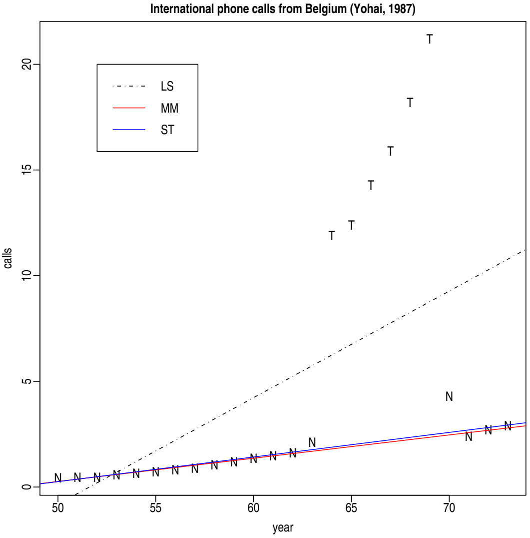
- What are we estimating?
- If the error distribution is not symmetric, no explicit solution

  In the "robust likelihood" approach we estimate the parameters of the error distribution
- Note:
  empirical evidence that real data have asymmetric outliers

A simple regression example (Yohai, 1987)



A simple regression example (Yohai, 1987)



A classical benchmark: stackloss data

$$\text{(loss function)} = \sum_{i=1}^{n} |y_i - \hat{y}_i|^p$$

p	0.5	1	2
LS	30.1	49.7	**178.8**
MM	27.1	45.3	222.8
LTS	25.9	44.7	241.7
ST	**25.0**	**43.4**	240.0

($n = 21$ with 3 covariates)

Regression with contaminated normal errors

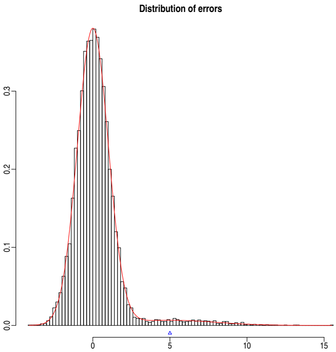
Simulate data from model:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

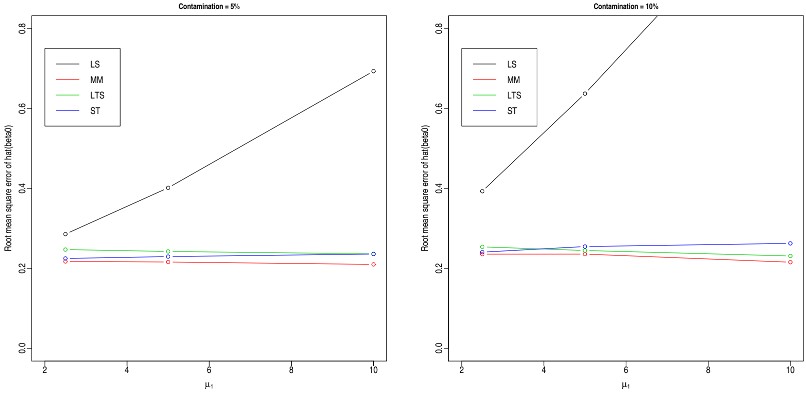
where

$$\begin{aligned} \varepsilon &\sim (1-\pi)\,\mathrm{N}(0,1) + \pi\,\mathrm{N}(\mu_1, 3) \\ \beta_0 &= 0 \\ \beta_1 &= 2 \\ \pi &= 0.05, \quad 0.10 \\ \mu_1 &= 2.5, \quad 5, \quad 10 \end{aligned}$$

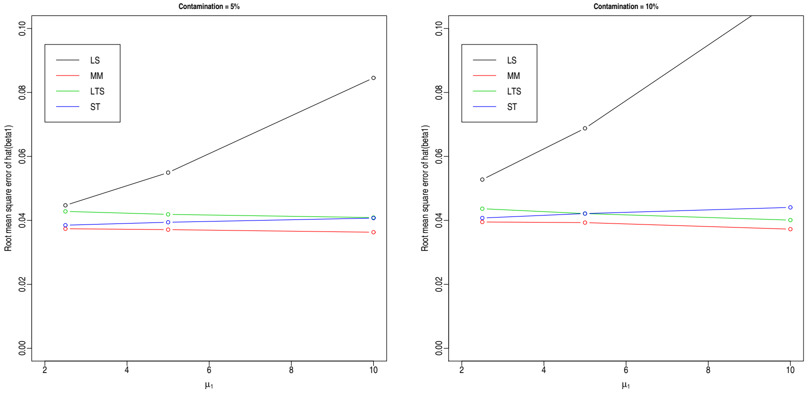
replicates: 10^4 in each case



Simulation: Root Mean Square Error for β_0



Simulation: Root Mean Square Error for β_1



Summary

- ST and other flexible families of distributions allow regulation of skewness and kurtosis
- corresponding likelihood inference appears reliable even when used outside the parametric class
- advantages are:
  - a probability model is fitted to the data
  - the quantities being estimated are explicitly known

References & resources

- Genton, M. G. (2004, *Skew-elliptical distributions. . .*) edited volume
- Azzalini, A. (2005, *Scand J. Stat.*, vol.32) Review paper with discussion

- Resources:
  `http://azzalini.stat.unipd.it/SN/`

- A. Azzalini & M. G. Genton (2008). Robust likelihood methods based on the skew-*t* and related distributions. *Int. Statist. Rev., 76, 106–129*