

Conditional Density Estimation under Non-Normality Assumption.

Julie Carreau

Yoshua Bengio

June 9th

Goals

Tackle regression modelling when the noise distribution is **unknown, asymmetric** and **heavy-tailed**.

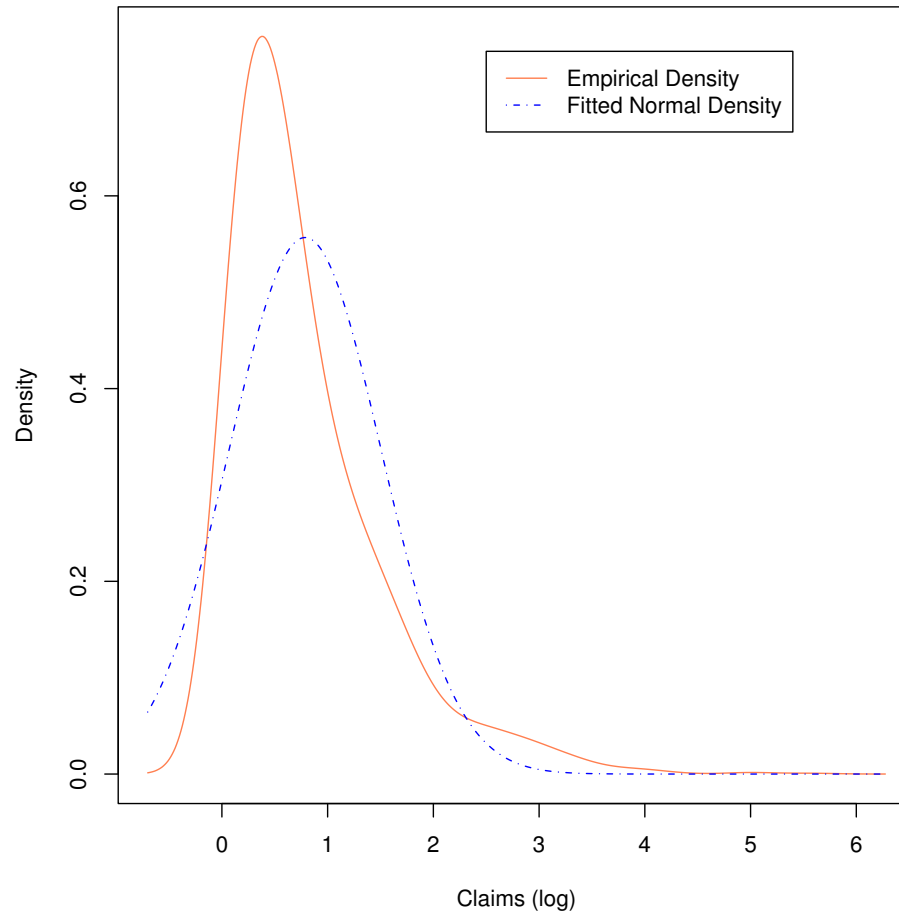
- We **cannot throw away** extreme observations since they were generated by the underlying process.
- Robust regression methods **downweight extreme observations** symmetrically ; this introduces a bias in the case of asymmetry.

Normality Assumption

In the regression model, $Y = f(X, \theta) + \epsilon$, the assumption $\epsilon \sim \mathcal{N}(0, \sigma^2)$ entails that the distribution of ϵ :

- is **symmetric**
- has **support = \mathbb{R}**
- has **exponential tails**.

★ In some applications such as **insurance** and **finance**, the **normality assumption is inadequate**.



Large fire insurance claims in Denmark

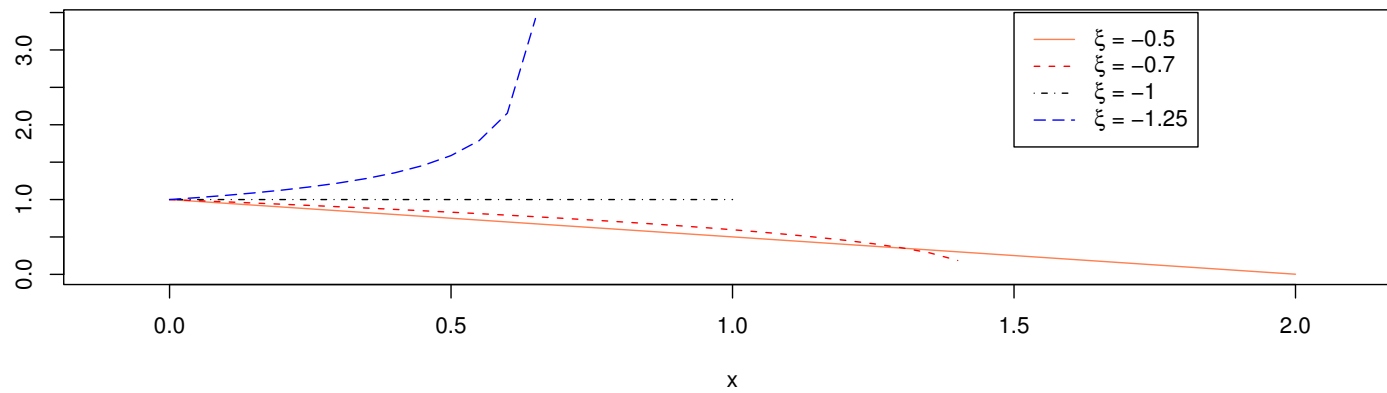
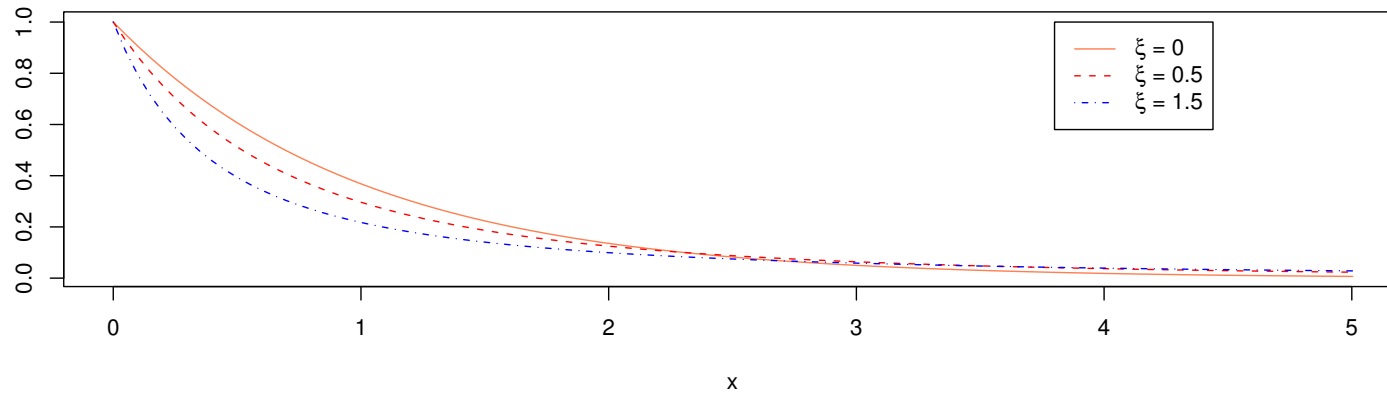
Proposal

Develop methods of conditional density estimation based on **Pareto-type tail** distributions.

Generalised Pareto Distribution (GPD) :

It fits the tail **more accurately** since it uses **only extreme observations** to build the model.

★ ξ is the tail index of the GPD : it controls the **thickness** of the tail.



$\xi > 0$ in the upper panel, and with $\xi < 0$ in the lower panel.

Threshold issue : bias-variance compromise

The GPD is fitted to **exceedances** over a given threshold u .

★ When $u \nearrow$, bias \searrow ; whereas when $u \searrow$, variance \searrow .

Random threshold

- Inconditional density : $(k + 1)$ -th order statistic.

$$X_{1,n} \geq \cdots \geq X_{k,n} \geq X_{k+1,n} \geq \cdots \geq X_{n,n}$$

- Conditional density : conditional quantile of level q

$$f_q(x) = \inf\{y \in \mathbb{R} : F_{Y|X=x}(y) \geq q\}.$$

★ q is chosen through cross-validation.

Methodology

For any conditional distribution function $F_{Y|X=x}$, and any $u(X)$, we define the **conditional excess distribution function** $F_{u(X)|X=x}$ as :

$$F_{u(X)|X=x}(y) = P(Y - u(X) \leq y | Y > u(X), X = x), \quad y \geq 0.$$

Then we can write :

$$F_{Y|X=x}(y) = \begin{cases} (1 - F_{Y|X=x}(u(x)))F_{u(X)|X=x}(y - u(x)) \\ \quad + F_{Y|X=x}(u(x)) & \text{if } Y > u(X), \\ F_{Y|X=x}(y) & \text{otherwise.} \end{cases}$$

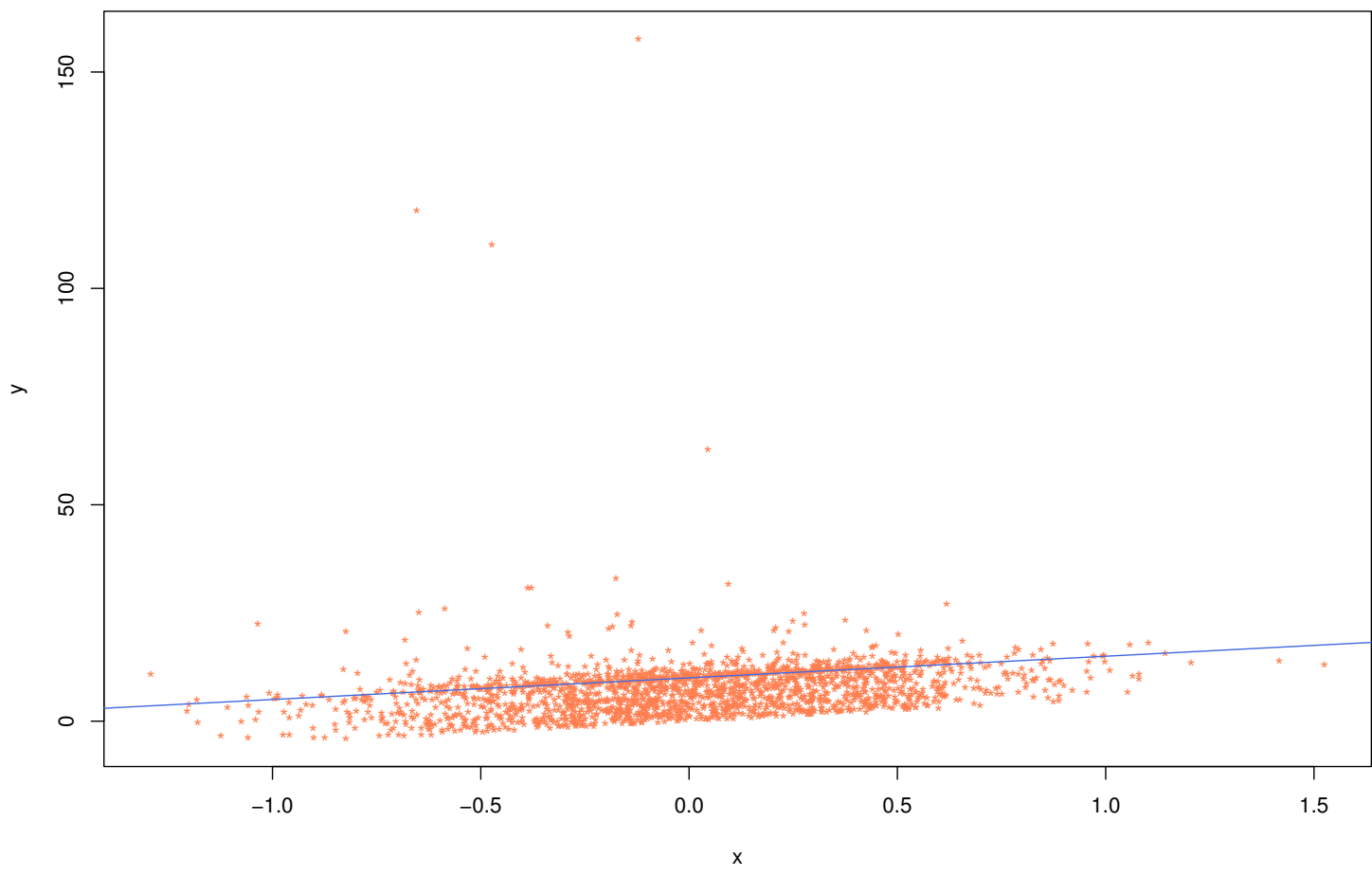
This suggest the use of a **non-parametric estimator** when $Y \leq u(X)$ and the use of the **GPD** when $Y > u(X)$.

Preliminary experiments

Artificial data set : dependent data with fat upper tail.

Uniform distribution below the threshold function and **conditional GPD** above.

GPD parameters are learned on a simulated data set with the above methodology.



Further work

Develop a **mixture of distributions** which includes a **fat-tailed component**.

- Add a **generalised Pareto component** in a mixture of Gaussian distribution; the resulting density is **discontinuous**.
- Build a **smoothed version of the GPD**; an hybrid between a normal distribution and the GPD.