Scaling for Skewness, with Spin-Offs and Insights

Minor

Chris Jones

The Open University, U.K.

Layout of Talk

1) PREAMBLE – a gentle stroll through some background ideas concerning three- and four-parameter distributions on ***R***
2) brief INTRODUCTION to kernel smoothing
   - what??: something apparently completely different!
3) the MAIN TALK on a particular family of distributions with “simple exponential tails”
   - with spin-offs!

PREAMBLE

- consider the univariate continuous one-sample situation for simplicity
- of course, classical statistical modelling involves the fitting of parametric distributions
- these parametric models might involve, say, four parameters: location, μ, scale, σ, and two shape parameters, *a* and *b*, say, accounting for skewness and tailweight IN SOME WAY

Model is of the form $$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}; a, b\right)$$

Options …

Option 1:

a controls skewness; b controls tailweight

a=0 => symmetry, tailweight changing

b=0 => asymmetry, tailweight as that of *g*

a=b=0 => *g* (a simple symmetric density)

“Obvious”, perhaps, but not always as easy as it may seem

For example, here’s a famous three-parameter (“b=0”) asymmetric family: $2\varphi(x)\Phi(ax)$. But *a* also affects tailweight: for a>0, the right-hand tail goes as $2\varphi(x)$, the left as $2\varphi(x)\varphi(ax)/(a|x|)$.

Options …

Option 2:

a controls left-tailweight
b controls right-tailweight
a=b => symmetry, tailweight changing
a=b=1 => *g*

Less obvious, perhaps, but easier to do and permeating almost all my efforts in this area!

Generalised density of order statistic:

$$\boxed{f(x) = \frac{g(x)G(x)^{a-1}(1-G(x))^{b-1}}{B(a,b)}}$$

(a,b>0 real)

$$X = G^{-1}(Beta(a,b))$$

(Jones, 2004, *Test*)

Roles of a and b

- a=b=1: f = g
- a=b: family of symmetric distributions
- a≠b: skew distributions
- a controls left-hand tail weight, b controls right
- the smaller a or b, the heavier the corresponding tail

Example 1: skew t

$$f(x) = \frac{\left(1+\frac{x}{\sqrt{a+b+x^2}}\right)^{a+1/2}\left(1-\frac{x}{\sqrt{a+b+x^2}}\right)^{b+1/2}}{B(a,b)\sqrt{a+b}\,2^{a+b-1}}$$

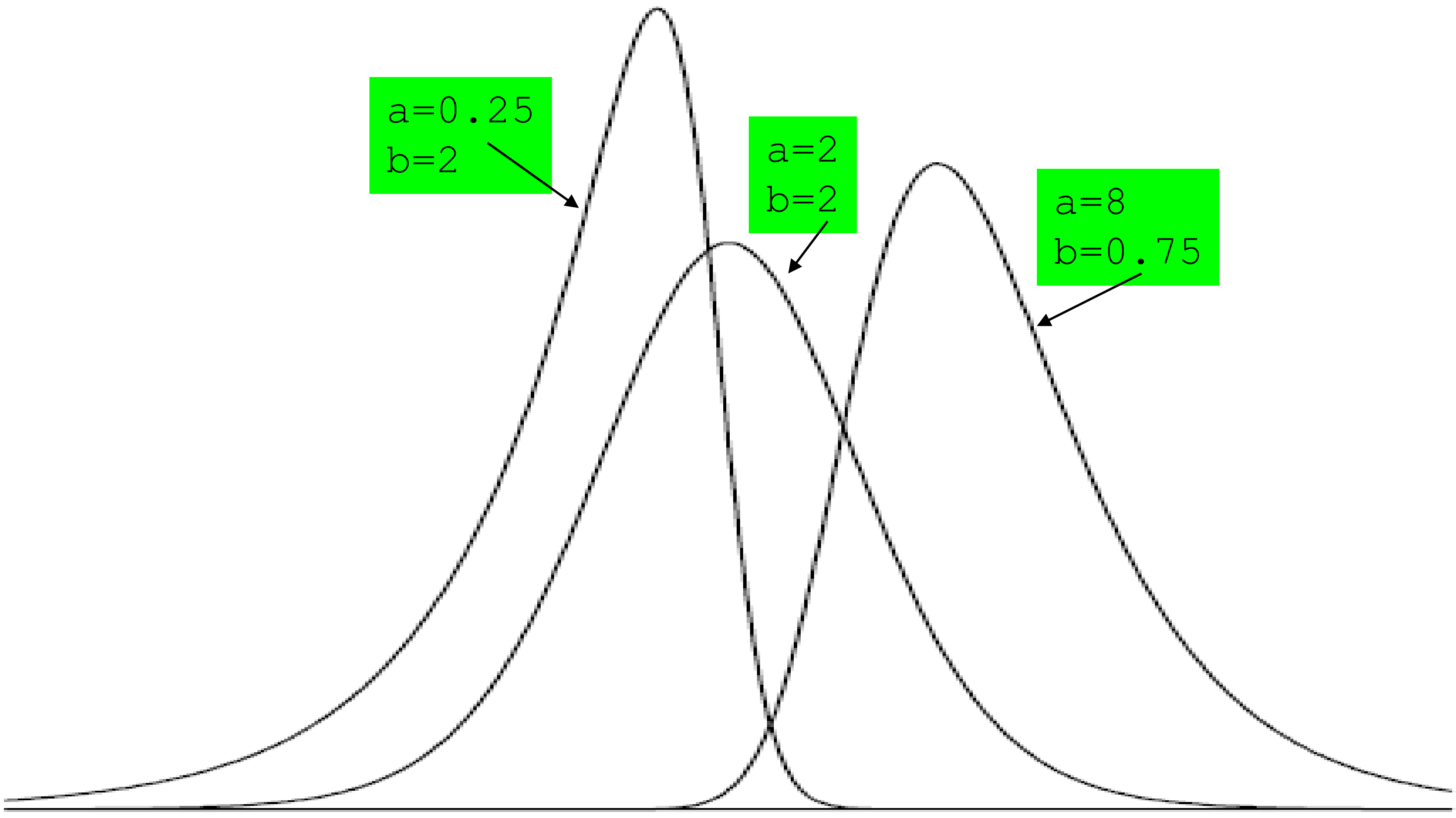
- when *a*=*b*, *f* is Student *t* density on 2*a* d.f.
- Jones & Faddy (2003, JRSSB)
- "order statistics" from *t* dist'n on 2 d.f.!
- density tails go as:
  - – left-hand tail: |x|^-(2a+1)
  - – right-hand tail: x^-(2b+1)

Example 2: log F

$$f(x) = \frac{1}{B(a,b)} \frac{e^{ax}}{(1+e^{x})^{a+b}}$$

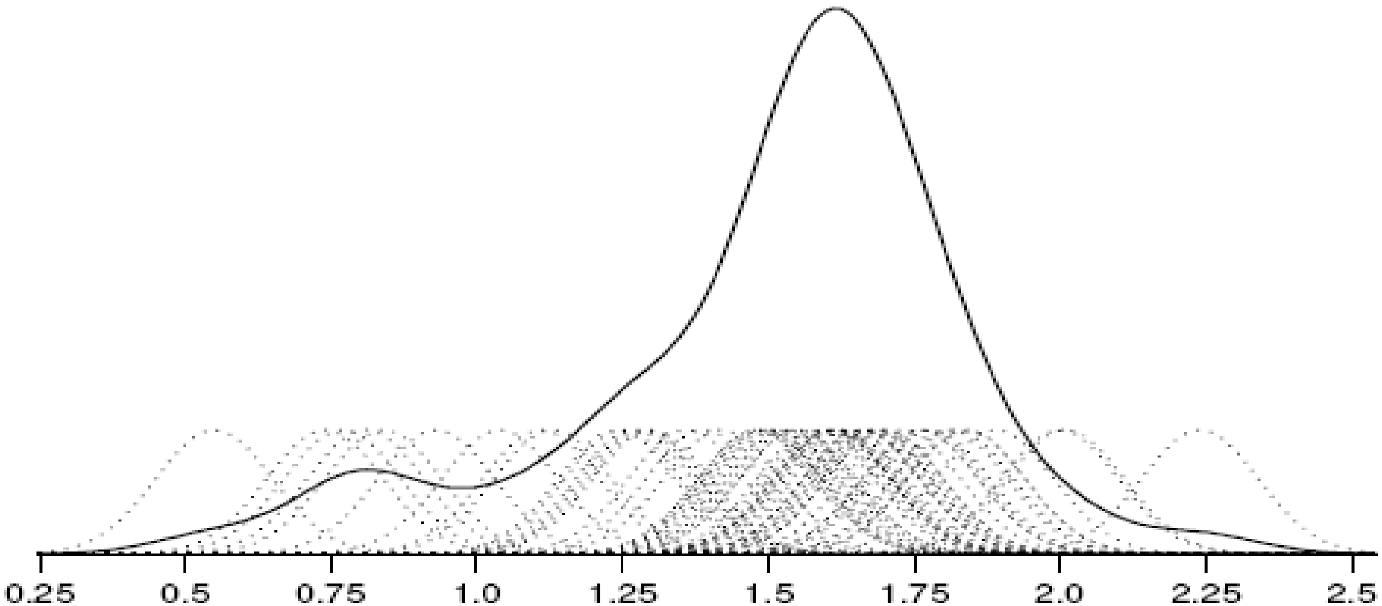
- density of log(Y) where Y~F
- goes back to R.A. Fisher in the 1920s
- “order statistics” from logistic distribution
- density tails go as:
  - left-hand tail: exp(ax)
  - right-hand tail: exp(-bx)

Some log F densities



(normalised to unit variance)

BRIEF INTERLUDE: INTRODUCTION TO KERNEL SMOOTHING



Kernel Density Estimation

estimate density f by

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)$$

The (bandwidth) parameter h controls the degree of smoothing and is (i) difficult and (ii) important to specify well.

Associated with the kernel **density** estimate is the kernel **distribution function** estimate

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}\left(\frac{x - X_i}{h}\right)$$

(where $\mathcal{K}(x) = \int^{x} K(y)dy$) ...

... and with this a kernel **quantile** estimate obtained by solving

$$p = \hat{F}(x).$$

MAIN TALK!

The log F density again

$$f(x) = \frac{1}{B(a,b)} \frac{e^{ax}}{(1+e^{x})^{a+b}}$$

This has the property of simple exponential tails:

$$x \to -\infty \Rightarrow f(x) \sim e^{ax}$$

$$x \to \infty \Rightarrow f(x) \sim e^{-bx}$$

The simple exponential tail property is shared by:

- the log F distribution
- the asymmetric Laplace distribution

$$f(x) = \frac{ab}{a+b}\exp\{axI(x<0) - bxI(x \geq 0)\}$$

- the hyperbolic distribution

$$f(x) \propto \exp\left\{\left(\frac{a-b}{2}\right)x - \left(\frac{a+b}{2}\right)\sqrt{1+x^2}\right\}$$

Is there a general form for such distributions?

A general family of distributions with simple exponential tails

Starting point: simple symmetric g with distribution function G and

$$G^{[2]}(x) = \int_{-\infty}^{x} G(t)\,dt.$$

General form for density is:

$$\boxed{f(x) \propto \exp\left\{ax - (a+b)G^{[2]}(x)\right\}}$$

(Jones, to appear, Statistica Sinica)

Special Cases

- G is point mass at zero, $G^{[2]}=xI(x>0)$

☺ f is asymmetric Laplace

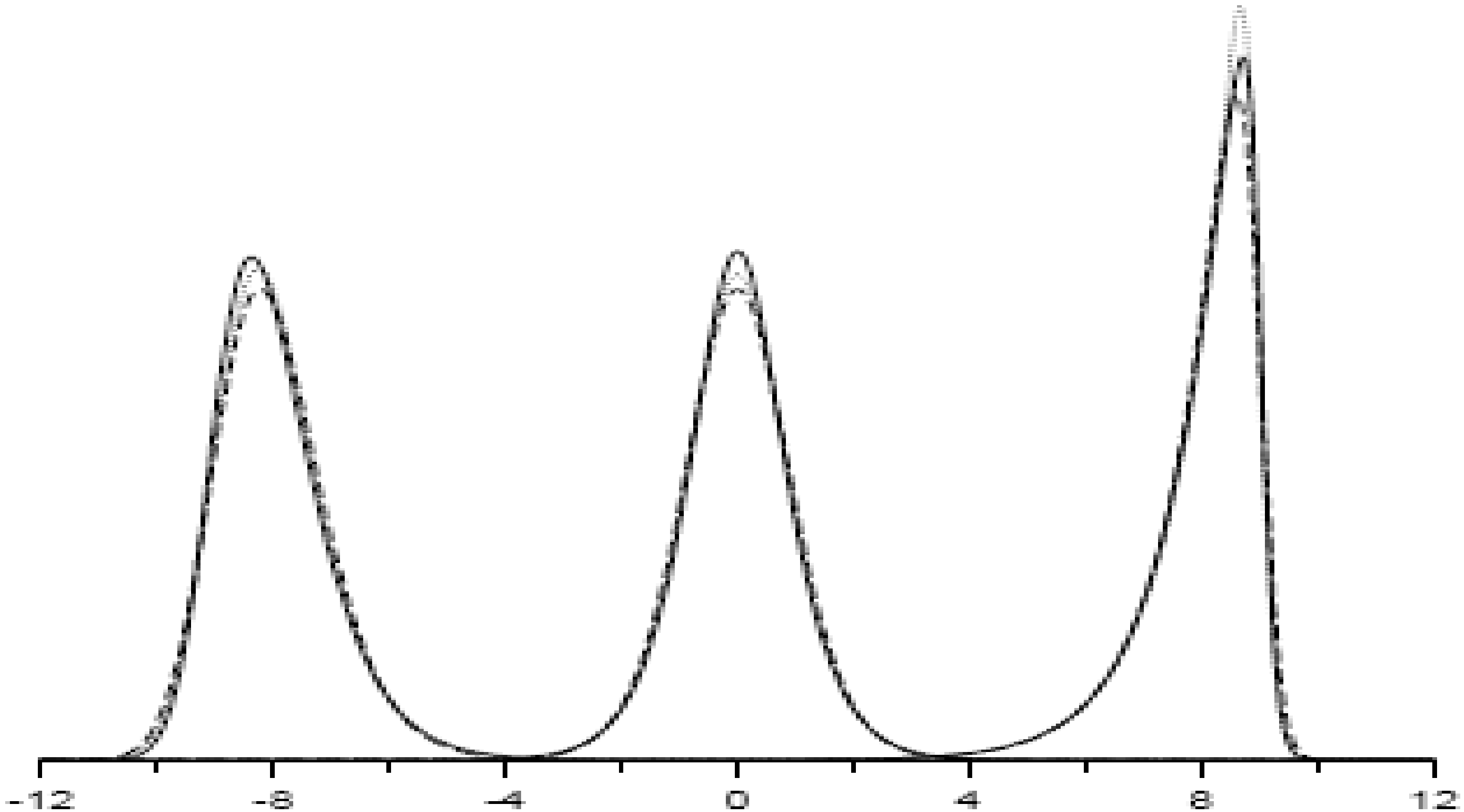
- G is logistic, $G^{[2]}=\log(1+\exp(x))$

☺ f is log F

- G is t_2, $G^{[2]}=\frac{1}{2}(x+\sqrt{1+x^2})$

☺ f is hyperbolic

- G is normal, $G^{[2]}= x\Phi(x)+\varphi(x)$
- G uniform, $G^{[2]}=\frac{1}{2}(1+x)I(-1<x<1)+I(x>1)$



solid line: log F

dashed line: hyperbolic

dotted line: normal-based

Practical Point 1

- the asymmetric Laplace is a three parameter distribution; other members of family have four;
- fourth parameter is redundant in practice: (asymptotic) correlations between ML estimates of σ and either of a or b are very near 1;
- reason: σ, a and b are all scale parameters, yet you only need two such parameters to describe main scale-related aspects of distribution [either (i) a left-scale and a right-scale or (ii) an overall scale and a left-right comparer]

Practical Point 2

Parametrise by μ, σ, a=1-p, b=p.
Then, score equation for μ reads:

$$p = \frac{1}{n}\sum_{i=1}^{n} G\left(\frac{\mu - X_i}{\sigma}\right)$$

This is kernel quantile estimation,
with kernel G and bandwidth σ

Includes bandwidth selection by choosing σ to solve the second score equation:

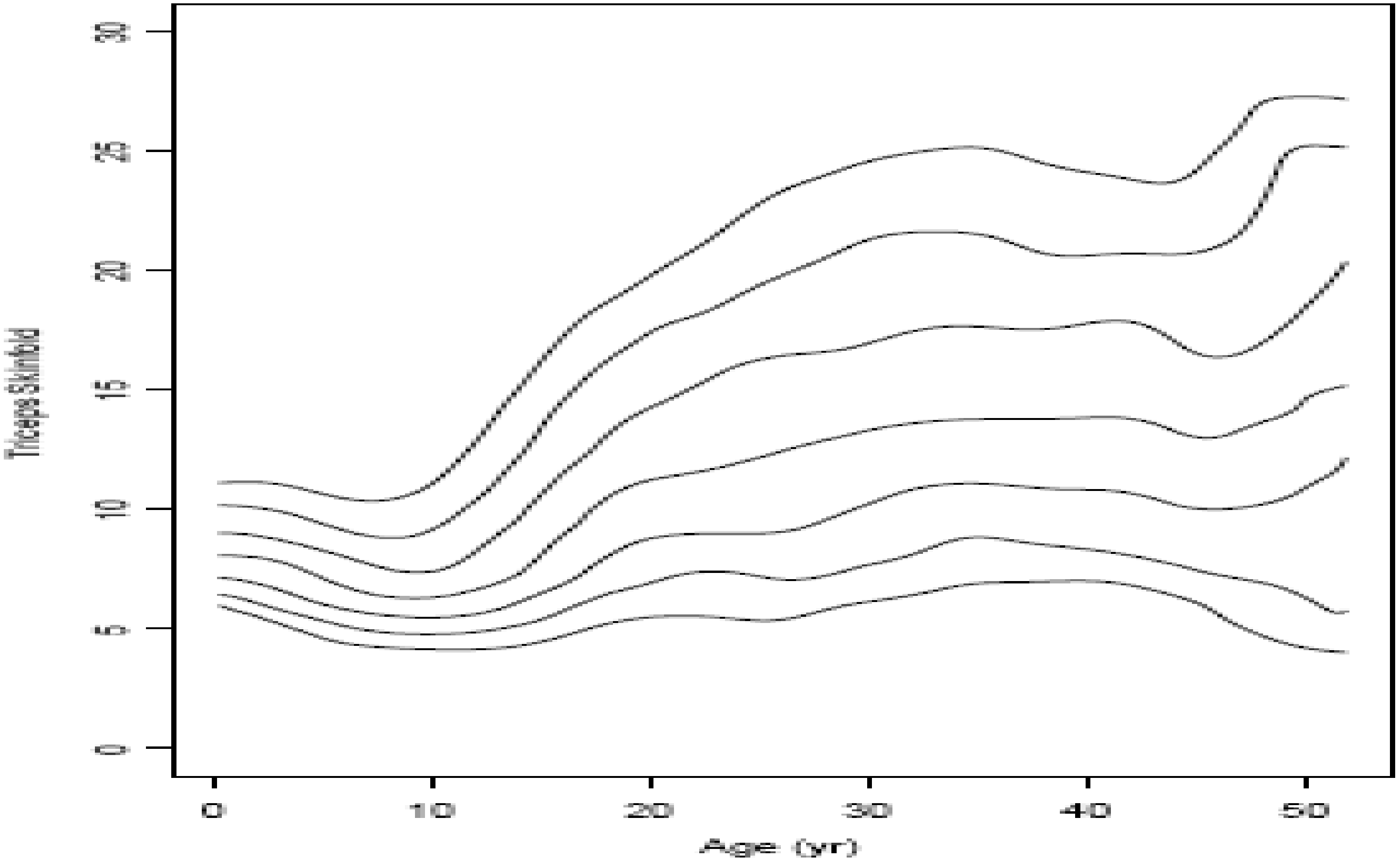
$$\sigma = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\left\{ p - G\left(\frac{\mu - X_i}{\sigma}\right)\right\}$$

But its simulation performance is variable:

Table 1: *Mean squared errors associated with the estimation of normal quantiles from samples of size $n = 50$ for specified p and four estimation methods. The logistic kernel was used in the kernel methods.* 50,000 *replications*

p	Sample quantile	Harrell-Davis	Kernel; rule-of-thumb bandwidth	Kernel; bandwidth via (11)
0.50	0.032	0.027	0.032	0.022
0.75	0.037	0.032	0.031	0.035
0.9	0.063	0.049	0.047	0.049
0.95	0.086	0.076	0.068	0.075

And so to Quantile *Regression*



The usual (regression) log-likelihood,

$$-n\log\sigma+\sum_{i=1}^{n}\left\{(1-p)\left(\frac{Y_i-\alpha-\beta X_i}{\sigma}\right)-G^{[2]}\left(\frac{Y_i-\alpha-\beta X_i}{\sigma}\right)\right\},$$

is kernel localised to point x by

$$\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)\left\{-n\log\sigma+(1-p)\left(\frac{Y_i-\mu-\mu_1(X_i-x)}{\sigma}\right)-G^{[2]}\left(\frac{Y_i-\mu-\mu_1(X_i-x)}{\sigma}\right)\right\}$$

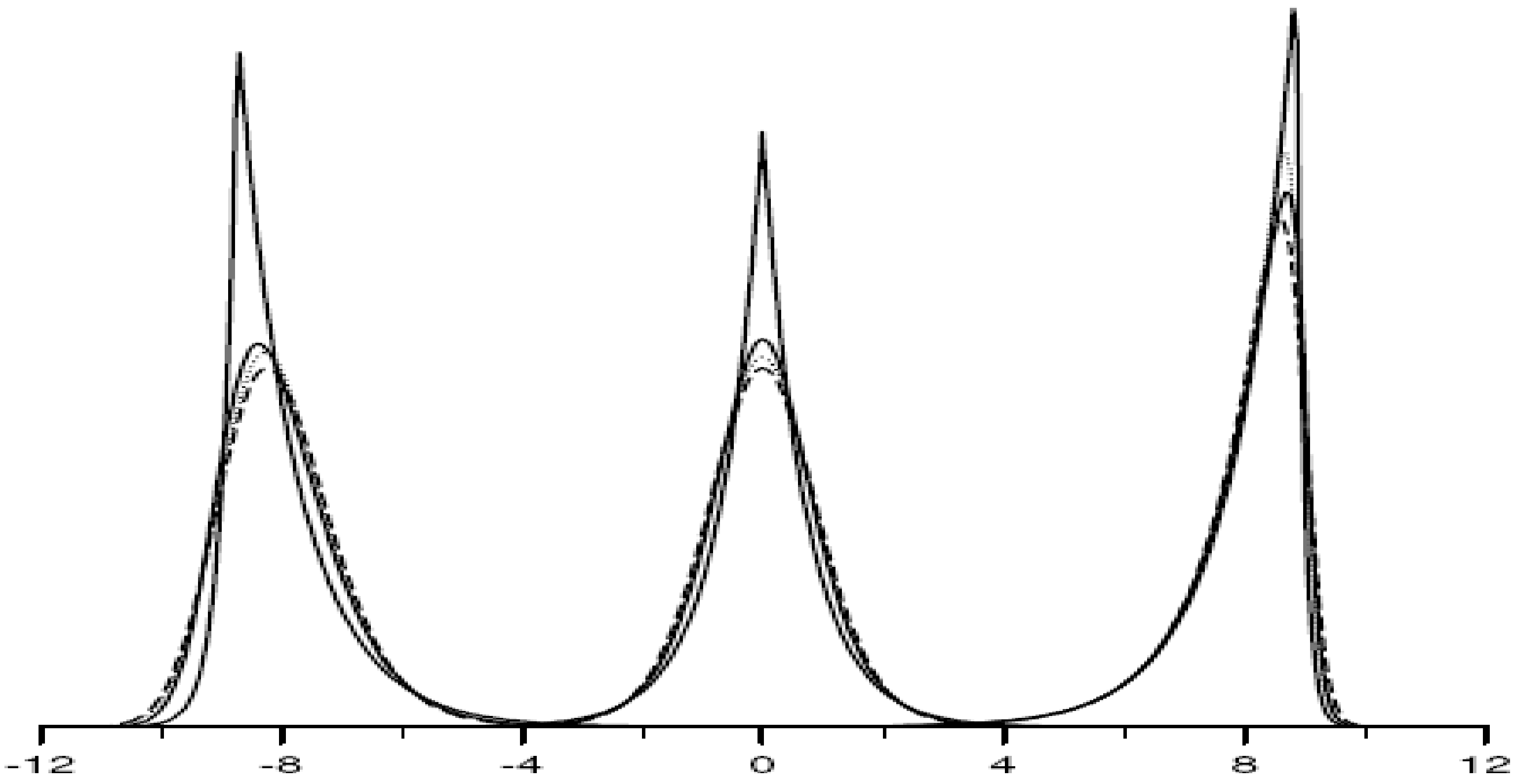
This new version of DOUBLE KERNEL LOCAL LINEAR QUANTILE REGRESSION compares favourably with the earlier, quite widely cited but more ad hoc, version of DKLLQR due to Yu & Jones (1998, JASA).

In simulations, the new method consistently outperforms the old method, if sometimes by only a small amount.

Based on theoretical and (admittedly somewhat limited) simulation evidence, we have:

- A clear recommendation:
  - replace Yu & Jones (1998) DKLLQR method by new version (Statistical Modelling, 2007)
- An unclear non-recommendation:
  - use new bandwidth selection?

Practical Point 3



A further advantage of our general family is that we can test for the appropriateness of the asymmetric (or indeed the symmetric) Laplace distribution

Such a test can be based on parametrising the (four-parameter with location 0) general family as:

$$\boxed{f(x) \propto \exp\left\{ax - (a+b)\sigma G^{[2]}(\sigma^{-1}x)\right\}}$$

and observing that the asymmetric Laplace distribution corresponds to $\sigma \to 0$

This is work at an early stage of progress, in collaboration with Karim Anaya-Izquierdo

POSTAMBLE

- focussed attention on tailweights when stipulating *a* and *b*
- if *a* and *b* are both scale parameters, can only really introduce skewness (i.e. 3 parameters)
- away from simple exponential tails, can afford for one of *a* or *b* to be a left- or right-scale parameter …
- open question: is it better to employ one or two "true" tailweight parameters (e.g. powers) in the presence of one, resp. no, left- or right-scale parameters?

References

(with K. Yu) **Local linear quantile regression.** *J. Amer. Statist. Assoc.* 93, 228–37, 1998.

(with M.J. Faddy) **A skew extension of the t distribution, with applications.** *J. Roy. Statist. Soc. Ser. B* 65, 159–74, 2003.

Families of distributions arising from distributions of order statistics (with discussion). *Test* 13, 1–43, 2004.

(with K. Yu) **Improved double kernel local linear quantile regression.** *Statistical Modelling* 7, 377–89, 2007.

On a class of distributions with simple exponential tails. *Statistica Sinica*, to appear.