A New Class of Multivariate Skew Distributions.

Sujit Sahu [a]

University of Southampton

http://www.maths.soton.ac.uk/staff/sahu/

Sheffield: March 2001.

[a] Joint work with Dipak Dey and M.D. Branco

Multivariate Skew Distributions

Outline:

- Motivation
- Elliptical distributions
- Skewing the elliptical distributions
- Examples: Skew normal and t
- Regression models and posteriors
- MCMC: HOWTO
- Numerical examples
- Discussion

Why consider skew distributions?

These provide alternatives to

- transforming the data
- symmetric error distributions
- non-parametric approaches

The distributions are

- quite flexible and general
- for modelling heavy tail and skewness
- tractable and parametric
- tools for robustness studies.

Corresponding symmetric error distributions are special cases.

Multivariate Skew Distributions

Outline:

- Motivation
- Elliptical distributions
- Skewing the elliptical distributions
- Examples: Skew normal and t
- Regression models and posteriors
- MCMC: HowTo
- Numerical examples
- Discussion

Consider two densities:

Multivariate Normal:

$$C\,|\Omega|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\theta})^T \Omega^{-1} (\mathbf{x}-\boldsymbol{\theta})\right]$$

Multivariate t:

$$C\,|\Omega|^{-\frac{1}{2}} \left[1 + \frac{[\mathbf{x}-\boldsymbol{\theta}]^T \Omega^{-1} [\mathbf{x}-\boldsymbol{\theta}]}{\nu}\right]^{-\frac{\nu+k}{2}}$$

Notes:

- the const C does not involve the location and scale parameters.
- the kernels are scalar functions of $[\mathbf{x}-\boldsymbol{\theta}]^T \Omega^{-1} [\mathbf{x}-\boldsymbol{\theta}]$.

There are other distributions

- Laplace
- Pearson Type II
- Pearson Type VII

which belong to the same broad class called

`Elliptically Symmetric Distributions`.

See Kelker (1970) and/or a text book

Fang, Kotz and Ng (1990) *Symmetric Multivariate and Related Distributions.*

Notations and identities:

- $g(u;k,...) =$ a non-increasing function, $\mathbb{R}^+$ to $\mathbb{R}^+$.
  - ◇ $u = [\mathbf{x} - \boldsymbol{\theta}]^T \Omega^{-1} [\mathbf{x} - \boldsymbol{\theta}]$.
  - ◇ $k =$ dimension
  - ◇ ...=Additional parameters, like ν=df in the t distribution.
- $g^{(k)}(u) = \frac{\Gamma(\frac{k}{2})}{\pi^{\frac{k}{2}}} \frac{g(u;k,...)}{\int_0^\infty r^{\frac{k}{2}-1} g(r;k,...)dr}$
  - ◇ Called the density generator
- $f\left(\mathbf{x}|\boldsymbol{\theta}, \Omega; g^{(k)}\right) = |\Omega|^{-\frac{1}{2}} g^{(k)}\left[(\mathbf{x} - \boldsymbol{\theta})^T \Omega^{-1} (\mathbf{x} - \boldsymbol{\theta})\right], \ \mathbf{x} \in \mathbb{R}^k$
  - ◇ is the multivariate density

For multivariate normal:

- take $g(u;k) = \exp(-\frac{u}{2})$,
- then $g^{(k)}(u) = \frac{\exp(-\frac{u}{2})}{(2\pi)^{\frac{k}{2}}}$.

For multivariate t:

- take $g(u;k,\nu) = \left[1 + \frac{u}{\nu}\right]^{-\frac{\nu+k}{2}}, \nu > 0.$,
- then $g^{(k)}(u) = \frac{\Gamma\left(\frac{\nu+k}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\frac{k}{2}}} g(u;k,\nu)$.

Will use:

- pdf = $f\left(\mathbf{x}|\boldsymbol{\theta}, \Omega; g^{(k)}\right)$
- cdf = $F\left(\mathbf{x}|\boldsymbol{\theta}, \Omega; g^{(k)}\right)$

Partition $\mathbf{X}, \boldsymbol{\theta}, \Omega$ into

$$\mathbf{X}_{k} = \begin{pmatrix} \mathbf{X}_{k_1}^{(1)} \\ \mathbf{X}_{k_2}^{(2)} \end{pmatrix}, \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}^{(1)} \\ \boldsymbol{\theta}^{(2)} \end{pmatrix},$$

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$$

$$\begin{aligned} \boldsymbol{\theta}_{1.2} &= \boldsymbol{\theta}^{(1)} + \Omega_{12}\Omega_{22}^{-1}\left(\mathbf{x}^{(2)} - \boldsymbol{\theta}^{(2)}\right) \\ \Omega_{11.2} &= \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21} \end{aligned}$$

Also define:

$$q(\mathbf{x}^{(2)}) = \left(\mathbf{x}^{(2)} - \boldsymbol{\theta}^{(2)}\right)^T \Omega_{22}^{-1} \left(\mathbf{x}^{(2)} - \boldsymbol{\theta}^{(2)}\right).$$

For conditional densities we need to define for $a > 0$ the density generator:

$$g_a^{(k_1)}(u) = \frac{\Gamma(\frac{k_1}{2})}{\pi^{\frac{k_1}{2}}} \frac{g(a+u;k)}{\int_0^\infty r^{\frac{k_1}{2}-1} g(a+r;k)dr}$$

It is slightly different than what we had

$$g^{(k)}(u) = \frac{\Gamma(\frac{k}{2})}{\pi^{\frac{k}{2}}} \frac{g(u;k)}{\int_0^\infty r^{\frac{k}{2}-1} g(r;k)dr}.$$

The result is: if $\mathbf{X} \sim f(\mathbf{x}|\boldsymbol{\theta}, \Omega; g^{(k)})$, then

$$\mathbf{X}^{(1)}|\mathbf{x}^{(2)} \sim f\left(\mathbf{x}_1|\boldsymbol{\theta}_{1.2}, \Omega_{11.2}; g_{q(\mathbf{x}^{(2)})}^{(k_1)}\right).$$

Examples of conditionals

Normal:

$$g_a^{(k_1)}(u) = g^{(k_1)}(u)$$

since $g(u, k) = \exp(-\frac{u}{2})$, i.e. a cancels in the ratio.

t:

After some manipulation the conditional density of $\mathbf{X}^{(1)}|\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$ will be:

$$t_{k_1,\nu+k_2}\left(\boldsymbol{\theta}_{1.2}, \frac{\nu + q(\mathbf{x}^{(2)})}{\nu + k_2}\,\Omega_{11.2}\right).$$

- dimension $= k_1$,
- df $= \nu + k_2$.

Multivariate Skew Distributions

Outline:

- Motivation
- Elliptical distributions
- Skewing the elliptical distributions
- Examples: Skew normal and t
- Regression models and posteriors
- MCMC: HowTo
- Numerical examples
- Discussion

Suppose

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{0} \end{pmatrix}, \ \Omega = \begin{pmatrix} \Sigma & 0 \\ 0 & I \end{pmatrix}$$

and

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{\epsilon} \\ \mathbf{Z} \end{pmatrix} \sim f\left(\mathbf{x}|\boldsymbol{\theta}, \Omega; g^{(2m)}\right),$$

Consider

$$\mathbf{Y} = D\mathbf{Z} + \boldsymbol{\epsilon}, \qquad (1)$$

where $D = \text{diag}(\delta_1, \ldots, \delta_m)$.

The multivariate class is developed by considering the random variable

$$[\mathbf{Y}|\mathbf{Z} > \mathbf{0}]$$

Theorem 1 *Let $\mathbf{y}_* = \mathbf{y} - \boldsymbol{\mu}$. Then the pdf of $\mathbf{Y}|\mathbf{Z} > 0$ is given by*

$$f\left(\mathbf{y}|\boldsymbol{\mu}, \Sigma, D; g^{(m)}\right) =$$

$$2^m f_{\mathbf{Y}}\left(\mathbf{y}|\boldsymbol{\mu}, \Sigma + D^2; g^{(m)}\right)$$
$$F\left(\left[I - D(\Sigma + D^2)^{-1}D\right]^{-\frac{1}{2}} D(\Sigma + D^2)^{-1}\mathbf{y}_*|\mathbf{0}, I; g^{(m)}_{q(\mathbf{y}_*)}\right),$$

where

$$q(\mathbf{y}_*) = \mathbf{y}_*^T(\Sigma + D^2)^{-1}\mathbf{y}_*.$$

Looks complicated because the formula is quite general!

Multivariate Skew Distributions

Outline:

- Motivation
- Elliptical distributions
- Skewing the elliptical distributions
- Examples: Skew normal and t
- Regression models and posteriors
- MCMC: HOWTO
- Numerical examples
- Discussion

A new multivariate skew normal distribution:

$$f(\mathbf{y}|\boldsymbol{\mu}, \Sigma, D) = 2^m |\Sigma + D^2|^{-\frac{1}{2}} \phi_m \left[(\Sigma + D^2)^{-\frac{1}{2}} (\mathbf{y} - \boldsymbol{\mu}) \right] \Phi_m \left[(I - D(\Sigma + D^2)^{-1} D)^{-\frac{1}{2}} D(\Sigma + D^2)^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right],$$

where

- ϕ_m = the pdf
- Φ_m = the cdf

of the m dimensional standard normal distribution.

Special cases:

- $D = 0$ gives the standard MVN.

- $\Sigma = \sigma^2 I$ and $D = \delta I$ gives iid marginals, with pdf

$$\frac{2}{\sqrt{\sigma^2+\delta^2}}\phi\left(\frac{y_i-\mu_i}{\sqrt{\sigma^2+\delta^2}}\right)\Phi\left(\frac{\delta}{\sigma}\frac{y_i-\mu_i}{\sqrt{\sigma^2+\delta^2}}\right)$$

- Which is the univariate version of Azzalini's skew normal distribution.
- Our multivariate version is different
- we condition on the same number of random variables
- he conditions on exactly one random variable.
- That is why we get iid skew marginals.

Moments and Skewness:

mgf is given by:

$$M_{\mathbf{Y}}(\mathbf{t}) = 2^m e^{\mathbf{t}^T\boldsymbol{\mu}+\mathbf{t}^T(\Sigma+D^2)\mathbf{t}/2}\Phi_m(D\mathbf{t}).$$

The mean and variance of skew normal $(\boldsymbol{\mu}, \Sigma, D)$ are given by,

$$E(\mathbf{Y}) = \boldsymbol{\mu} + \left(\frac{2}{\pi}\right)^{1/2}\boldsymbol{\delta},$$

$$\text{Cov}(\mathbf{Y}) = \Sigma + \left(1-\frac{2}{\pi}\right)D^2.$$

Comparing with Azzalini's:

Our version has pdf:

$$f(\mathbf{y}) = 2^2\,\phi(y_1)\phi(y_2)\Phi(\delta y_1)\Phi(\delta y_2),$$

Azzalini and Dalla Valle (1996) version,

$$f(\mathbf{y}) = 2\,\phi(y_1)\phi(y_2)\Phi\left[\delta(y_1 + y_2)\right].$$

Ours is much easier to analyse and think!

Use the skewness measure $\beta_{1,2}$ introduced by Mardia (1970).

Ours:

$$\beta_{1,2} = 4(4-\pi)^2 \left\{ \frac{\delta^2}{\delta^2(\pi-2)+\pi} \right\}^3$$

Theirs:

$$\beta'_{1,2} = 16(4-\pi)^2 \left\{ \frac{\delta^2}{2\delta^2(\pi-2)+\pi} \right\}^3.$$

Skew normal densities:

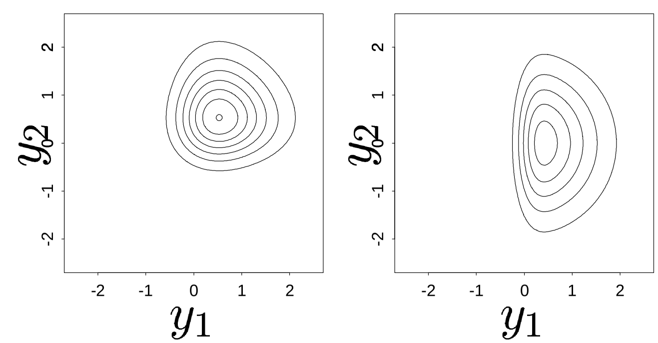


Figure 1: Contour plots of bivariate skew normal distributions. Left panel for our pdf and right panel for their pdf.

With appropriate choices of the g: generator function

$$f(\mathbf{y}|\boldsymbol{\mu}, \Sigma, D, \nu) =$$

$$2^m\, t_{m,\nu}(\mathbf{y}|\boldsymbol{\mu}, \Sigma + D^2)$$
$$T_{m,\nu+m}\Bigg[\left(\tfrac{\nu+q(\mathbf{y}_*)}{\nu+m}\right)^{-\frac{1}{2}}$$
$$\left(I - D(\Sigma + D^2)^{-1}D\right)^{-\frac{1}{2}}$$
$$D(\Sigma + D^2)^{-1}\mathbf{y}_*\Bigg].$$

- Its a new distribution.
- As expected $\Sigma = I$ and $D = \delta I$ does not produce iid marginals.

Moments and Skewness:

Can obtain the mgf as an integral of mgfs of scale mixture of skew normals.

because

Thorem: The proposed t distribution is a scale mixture of skew normals.

$$E(\mathbf{Y}) = \boldsymbol{\mu} + \left(\frac{\nu}{\pi}\right)^{1/2} \frac{\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)}\boldsymbol{\delta},$$

$$\text{Cov}(\mathbf{Y}) = (\Sigma + D^2)\frac{\nu}{\nu-2} - (\cdots)(\cdots)^T.$$

Measure of skewness Mardia (1970) does not have a closed form.

Consider the univariate version and compare with other skew t distributions.

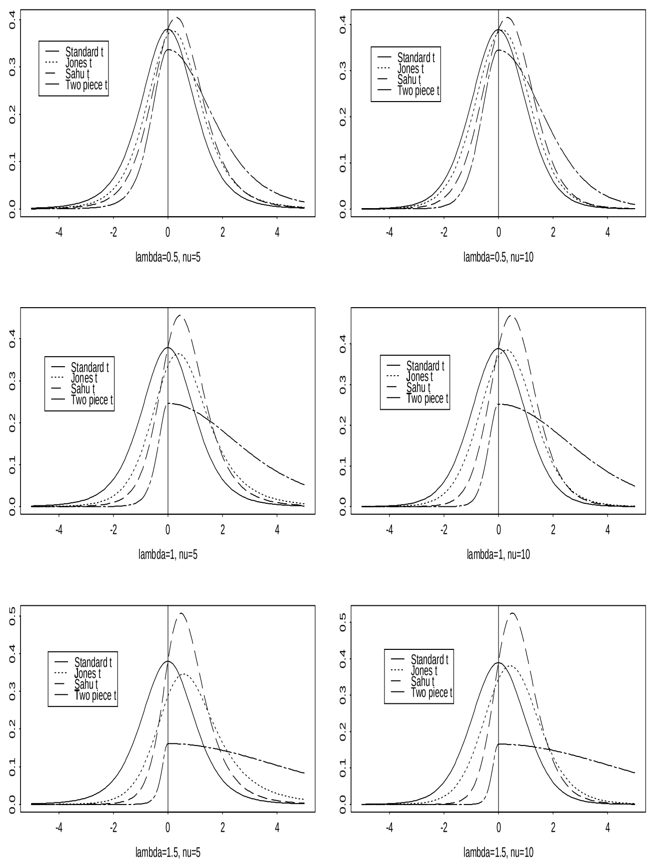


Figure 2: Plots of the density functions of skew t distributions.

Other skewness ideas

First two from Chris Jones:

1. $f(t|a,b) = C(a,b)^{-1}$

$$\left(1+\frac{t}{(a+b+t^2)^{1/2}}\right)^{a+1/2}$$

$$\left(1-\frac{t}{(a+b+t^2)^{1/2}}\right)^{b+1/2}$$

where

$$C(a,b) = B(a,b)(a+b)^{1/2}2^{a+b-1}$$

- When $a = b$, f is the standard symmetric t density with $2a$ df.
- $a < b$, f is negatively skewed.
- $a > b$, f is positively skewed.

2. We all know:

$$f(x,y) = f(y|x)f_X(x)$$

But define:

$$f_1(x,y) = \frac{g(x)f(x,y)}{f_X(x)}$$

Then the X marginal of f_1 will be $g(x)$ which we can choose to be skew modulo support considerations.

- Preserves the conditional
- Y marginal will be different.
- Can have one skew marginal and another symmetric marginal.

Let f and F be the pdf and cdf of any symmetric distribution on the real line.

3. Azzalini's:

$$g(x|\lambda) = 2f(x)F(\lambda x).$$

- Can replace F by H where H is a cdf.
- λ controls skewness.
- Can extend to multivariate.

4. For example: Fernandez and Steel

For $\tau > 0$ define $g(x|\tau) =$

$$\frac{2\tau}{1+\tau}\left\{f(x)I(x \geq 0) + f(\tau x)I(x < 0)\right\}.$$

- $\tau = 1$ corresponds to symmetric.
- $\tau > 1$ positively skew.

Multivariate Skew Distributions

Outline:

- Motivation
- Elliptical distributions
- Skewing the elliptical distributions
- Examples: Skew normal and t
- Regression models and posteriors
- MCMC: HowTo
- Numerical examples
- Discussion

Univariate response:

- General: $\boldsymbol{\mu} = X\boldsymbol{\beta}$.
- Example:
  $y_i = \alpha + \sum_{j=1}^{p} \beta_j x_{ij} + \delta z_i + \epsilon_i$.
- Assume normal prior for δ.
- ϵ_i is the skew elliptical distribution with scale σ^2.

Multivariate response:

- $\boldsymbol{\mu}_i = X_i\boldsymbol{\beta}$,
- $\Sigma^{-1} = Q \sim W_m(2r, 2\kappa)$: Wishart
- Assume suitable priors for other parameters.

We have two results for the posterior distribution in the univariate case:

Theorem 2 *Suppose that π_δ and π_ν are proper distributions. Then the posterior is proper under the skew normal or skew t model if $n > p$.*

Theorem 3 *Suppose that π_δ and π_ν are proper distributions. Then $E[(\sigma^2)^k|\mathbf{y}]$ exists under the skew normal or skew t model if $n - p > 2k$.*

Multivariate Skew Distributions

Outline:

- Motivation
- Elliptical distributions
- Skewing the elliptical distributions
- Examples: Skew normal and t
- Regression models and posteriors
- MCMC: Howto
- Numerical examples
- Discussion

Use a hierarchical setup:

$$\mathbf{Y}|\mathbf{Z}=\mathbf{z}\sim El\left(\boldsymbol{\mu}+D\mathbf{z},\Sigma;g_{q(z)}^{(m)}\right)$$

Examples:

- Normal: mean=$\boldsymbol{\mu}+D\mathbf{z}$ and covariance matrix Σ
- t case: $\mathbf{Y}|\mathbf{Z}=\mathbf{z}\sim t_{m,\nu+m}\left(\boldsymbol{\mu}+D\mathbf{z},\frac{\nu+\mathbf{z}^T\mathbf{z}}{\nu+m}\Sigma\right)$.

Then specification for Z:

- Normal: mean=0 and covariance matrix I.
- t case: $\mathbf{Z}\sim t_{m,\nu}\left(0,I\right)$.

Can obtain BUGS code for univariate response from my homepage.

For multivariate skew t errors:

$$\begin{aligned}
\mathbf{Y}_i|\cdots &\sim N_m\left(X_i^T\boldsymbol{\beta}+D\mathbf{z}_i,\frac{\Sigma}{w_i}\right)\\
\mathbf{Z}_i &\sim N_m(\mathbf{0},I)I(\mathbf{z}>\mathbf{0})\\
\boldsymbol{\beta} &\sim N_p(\boldsymbol{\beta}_0,\Lambda)\\
Q=\Sigma^{-1} &\sim W_m(2r,2\kappa)\\
\boldsymbol{\delta} &\sim N_m(\mathbf{0},\Gamma)\\
w_i &\sim \Gamma(\nu/2,\nu/2)\\
\nu &\sim \Gamma(1,0.1)I(\nu>3),
\end{aligned}$$

Z_i and $\boldsymbol{\delta}$ needs bit of work.

$$\begin{aligned}
\mathbf{Z}_i|\cdots &\sim N_m(A_i^{-1}\mathbf{a}_i,A_i^{-1})I(\mathbf{z}_i>\mathbf{0})\\
\boldsymbol{\delta}|\cdots &\sim N_m(B^{-1}\mathbf{b},B^{-1})
\end{aligned}$$

where

$$A_i=I+w_iDQD,\ \ \mathbf{a}_i=w_iDQ(\mathbf{y}_i-X_i^T\boldsymbol{\beta}).$$

Multivariate Skew Distributions

Outline:

- Motivation
- Elliptical distributions
- Skewing the elliptical distributions
- Examples: Skew normal and t
- Regression models and posteriors
- MCMC: HowTo
- Numerical examples
- Discussion

Interview data:

For 335 applicants:

- interview scores
- gender
- locality

Objective is to see if local female candidates score better.

Regression model has four parameters:

- intercept $= \alpha$
- β_1 for gender (male=1).
- β_2 for locality (local=1).
- β_3 for interaction.

Summary plots: Histogram

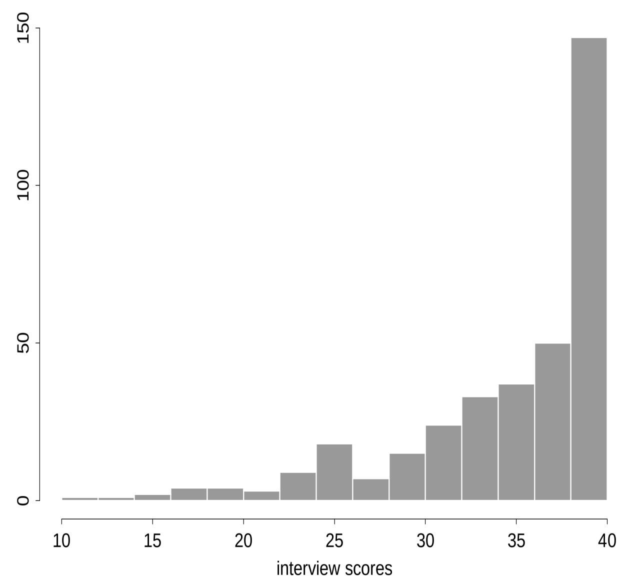


Figure 3: Histogram of interview scores.

Summary plots: Boxplot

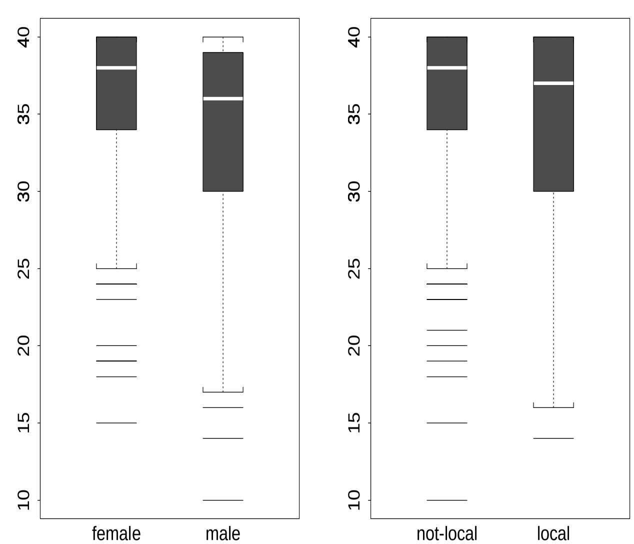


Figure 4: Box plot of interview scores.

	α	β_1	β_2	β_3	δ	ν	σ^2
N	35.9	-3.9	0.3	2.4			35.1
	0.7	0.9	0.8	1.2			2.7
S-N	31.7	-2.4	1.9	1.0	3.1		33.7
	1.0	1.0	0.9	1.2	0.8		2.9
t	37.4	-3.6	0.04	2.0		3.6	16.1
	0.6	0.9	0.7	1.1		0.6	2.1
S-t	35.9	-3.1	0.1	1.7	0.9	3.3	16.7
	0.6	0.9	0.6	1.1	0.2	0.3	1.9

Table 1: Parameter estimates for the interview data example. The standard deviations (not standard errors) are given in the second row.

	Skew-t	t	Skew-N	Normal
Skew-t	1	5.2	7.1	12.3
t	–	1	1.80	7.0
Skew-N	–	–	1	5.2
Normal	–	–	–	1

Table 2: The log of the Bayes factors using the Laplace-bridge method. Each entry in the table is the log of the Bayes factor in favor of the model in the row compared to the model in the column.

A small data example:

Have $x =$% of body fatness, $y =$ skin thickness and density for 24 women.

Experiment with:

- informative: prior variance 1,
- non-informative: prior variance 100,

for the regression co-efficient.

	α	β	δ	ν	σ^2
Normal	28.41	0.86			19.1
	(0.66)	(0.14)			(6.2)
Skew-N	27.91	0.86	1.1		17.7
	(1.05)	(0.14)	(1.81)		(6.6)
t	28.35	0.87		14.3	16.6
	(0.66)	(0.13)		(8.8)	(6.1)
Skew-t	27.84	0.87	1.0	14.9	15.1
	(2.9)	(0.13)	(1.7)	(10.5)	(6.4)

Table 3: Parameter estimates. Prior variances of α and β are 1 each.

	α	β	δ	ν	σ^2
Normal	28.68	0.86			19.4
	(0.86)	(0.14)			(6.3)
Skew-N	26.64	0.85	2.70		12.3
	(3.02)	(0.14)	(3.9)		(8.2)
t	28.58	0.87		14.6	16.9
	(0.87)	(0.13)		(10.6)	(6.4)
Skew-t	26.89	0.86	2.21	15.0	11.0
	(2.9)	(0.14)	(3.7)	(10.9)	(7.2)

Table 4: Parameter estimates. Prior variances of α and β are 100 each.

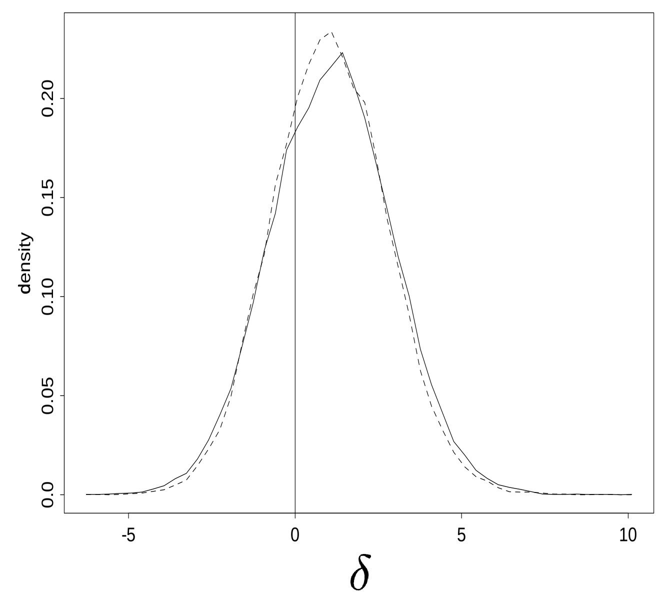


Figure 5: Posterior density of, δ, the skewness parameter. Solid line is for the skewed normal distribution and dotted line is for the skewed t-distribution. Prior variances of α and β are 1 each.

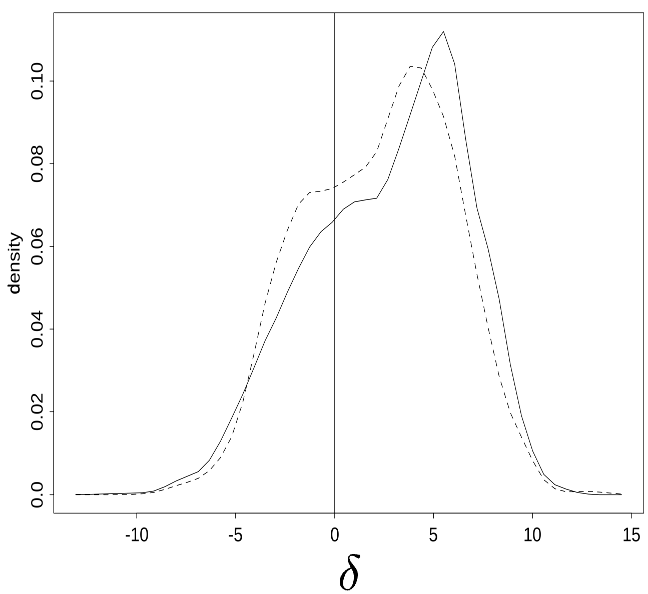


Figure 6: Posterior density of, δ, the skewness parameter. Solid line is for the skewed normal distribution and dotted line is for the skewed t-distribution. Prior variances of α and β are 100 each.

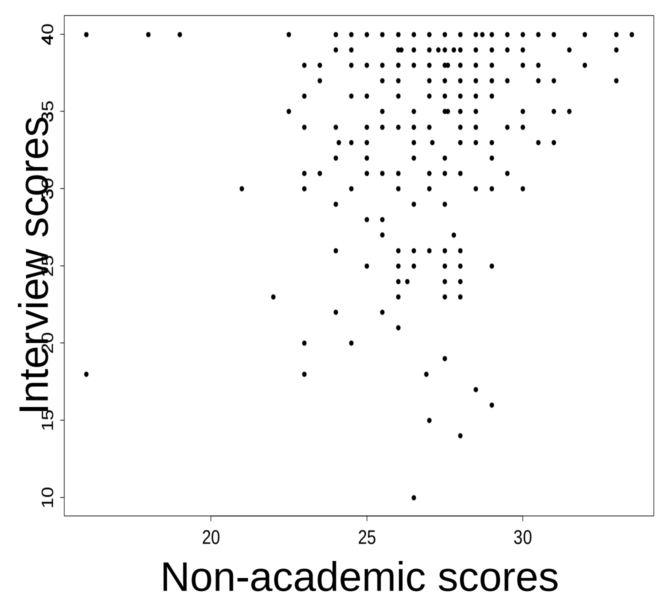


Figure 7: Scatter plot of the bivariate data used in model fitting.

Multivariate illustration: interview data

- Non-academic score
- interview score

No co-variates considered. Trying to estimate the means.

Normal	Skew Normal	t	Skew-t
-1776.0	-1671.7	-1722.5	-1631.8

Table 5: Marginal likelihood for the bivariate example.

Skew t model seems to be the best.

Example: Strength of glass fibre. Data from Smith and Naylor (1987).

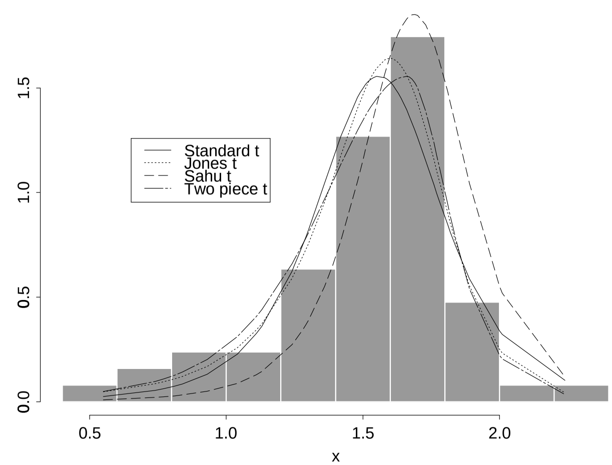
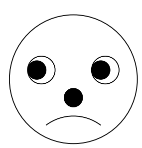


Figure 8: Histogram of the data and densities of the fitted skew t distributions for the glass fibre data.

Did not talk!

- Distribution theory, e.g.
- marginal and conditional distributions
- transformations and inter-relationships,
- What happens to the Bayes estimators?
- predictive distributions?
- And so on...



Nevertheless

- Transform the distribution rather than the data!
- Allows robust inference.
- All do-able because of MCMC.
- BUGS can do our univariate models.