Bayesian inference for multivariate skew-normal and skew-t distributions

Brunero Liseo

Sapienza Università di Roma

Banff, May 2013

Outline

Joint research with

- Antonio Parisi (Roma Tor Vergata)

◇ 1. Inferential Problems for the SN_p model

Outline

Joint research with

- Antonio Parisi (Roma Tor Vergata)

◇ 1. Inferential Problems for the SN_p model

◇ 2. Two slides on existing solutions

Outline

Joint research with

- Antonio Parisi (Roma Tor Vergata)

◇ 1. Inferential Problems for the SN_p model

◇ 2. Two slides on existing solutions

◇ 3. Bayesian proposal based on
Latent structure representation + Adaptive Importance Sampling.

Outline

Joint research with

- Antonio Parisi (Roma Tor Vergata)

◇ 1. Inferential Problems for the SN_p model

◇ 2. Two slides on existing solutions

◇ 3. Bayesian proposal based on
Latent structure representation + Adaptive Importance Sampling.

◇ 4. Extension to multivariate skew-t family and . . . possibly
. . . to skew-t copula.

Outline

Joint research with

- Antonio Parisi (Roma Tor Vergata)

◇ 1. Inferential Problems for the SN_p model

◇ 2. Two slides on existing solutions

◇ 3. Bayesian proposal based on
Latent structure representation + Adaptive Importance Sampling.

◇ 4. Extension to multivariate skew-t family and . . . possibly
. . . to skew-t copula.

◇ 5. An example.

Multivariate SN

Azzalini e Dalla Valle, Biometrika (1996)

- Conditioning: if $\mathbf{X}$ has $N(0,1)$ marginals and $\boldsymbol{\Omega}$ is a correlation matrix,

$$\binom{Z}{\mathbf{X}} \sim N_{p+1}\left[\binom{0}{\mathbf{0}}, \begin{pmatrix} 1 & \boldsymbol{\delta}^T \\ \boldsymbol{\delta} & \boldsymbol{\Omega} \end{pmatrix}\right] \Rightarrow \mathbf{U} = \begin{cases} \mathbf{X} & Z > 0 \\ -\mathbf{X} & Z < 0 \end{cases} \sim SN_p(\boldsymbol{\Omega}, \mathbf{0}, \boldsymbol{\alpha})$$

with density

$$f\left(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}\right) = 2\varphi_p\left(\mathbf{x}; \boldsymbol{\Omega}\right) \cdot \Phi_1\left[\boldsymbol{\alpha}'\mathbf{x}\right] \qquad \mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\alpha} \in \mathbb{R}^p$$

with $\boldsymbol{\alpha} = (1 - \boldsymbol{\delta}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\delta})^{-\frac{1}{2}}\boldsymbol{\Omega}^{-1}\boldsymbol{\delta}$.

Adding location and scale parameters

Let $\boldsymbol{\xi}$ a p-dimensional vector and let

$$\boldsymbol{\omega} = \mathrm{diag}\left(\omega_{\mathbf{1}}, \ldots, \omega_{\mathbf{p}}\right)$$

be the "vector" of marginal scale parameters, that is $\boldsymbol{\Sigma} = \boldsymbol{\omega}\Omega\boldsymbol{\omega}$.
Then $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{X} \sim SN_p(\boldsymbol{\Sigma}, \boldsymbol{\xi}, \boldsymbol{\alpha})$ with density

$$f(\mathbf{y}; \boldsymbol{\xi}, \Sigma, \boldsymbol{\alpha}) = 2\varphi_p(\mathbf{y} - \boldsymbol{\xi}; \boldsymbol{\Sigma})\Phi_1\left[\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})\right]$$

Inference

The likelihood function for an i.i.d. sample is then

$$
\begin{aligned}
L(\boldsymbol{\Sigma}, \xi, \alpha; \mathbf{y}) \quad &\propto \quad |\, \boldsymbol{\Sigma} \,|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \left[(\mathbf{y}_i - \boldsymbol{\xi})' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\xi}) \right] \right\} \\
&\times \quad \prod_{i=1}^{n} \boldsymbol{\Phi}_1 \left(\boldsymbol{\alpha}' \boldsymbol{\omega}^{-1} (\mathbf{y}_i - \boldsymbol{\xi}) \right).
\end{aligned}
$$

Difficult to work with...(Azzalini & Capitanio, 1999, and many others...)

Inference

The likelihood function for an i.i.d. sample is then

$$L(\boldsymbol{\Sigma}, \xi, \alpha; \mathbf{y}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left[(\mathbf{y}_i - \boldsymbol{\xi})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\xi})\right]\right\} \times \prod_{i=1}^{n} \boldsymbol{\Phi}_1\left(\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\mathbf{y}_i - \boldsymbol{\xi})\right).$$

Difficult to work with...(Azzalini & Capitanio, 1999, and many others...)
Small simulation: 2K samples of size 30 from a $SN_2\left(\boldsymbol{\xi} = (0,0), \Sigma = I_2, \boldsymbol{\alpha} = (2,2)\right)$.

Inference

The likelihood function for an i.i.d. sample is then

$$L(\boldsymbol{\Sigma}, \boldsymbol{\xi}, \boldsymbol{\alpha}; \mathbf{y}) \propto |\,\boldsymbol{\Sigma}\,|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left[(\mathbf{y}_i - \boldsymbol{\xi})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\xi})\right]\right\} \times \prod_{i=1}^{n} \boldsymbol{\Phi}_1\left(\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\mathbf{y}_i - \boldsymbol{\xi})\right).$$

Difficult to work with...(Azzalini & Capitanio, 1999, and many others...)

Small simulation: 2K samples of size 30 from a $SN_2\,(\boldsymbol{\xi} = (0,0), \Sigma = I_2, \boldsymbol{\alpha} = (2,2))$.

Estimates obtained with the **R** suite `sn`.

Inference

The likelihood function for an i.i.d. sample is then

$$L(\boldsymbol{\Sigma}, \boldsymbol{\xi}, \boldsymbol{\alpha}; \mathbf{y}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left[(\mathbf{y}_i - \boldsymbol{\xi})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\xi})\right]\right\} \times \prod_{i=1}^{n} \Phi_1\left(\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\mathbf{y}_i - \boldsymbol{\xi})\right).$$

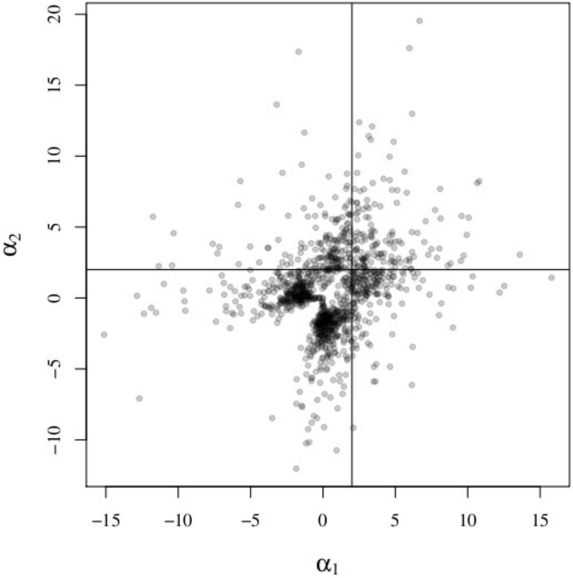
Difficult to work with...(Azzalini & Capitanio, 1999, and many others...)

Small simulation: 2K samples of size 30 from a $SN_2(\boldsymbol{\xi} = (0,0), \Sigma = I_2, \boldsymbol{\alpha} = (2,2))$.

Estimates obtained with the **R** suite `sn`.

38% of samples resulted in an infinite estimate for $\boldsymbol{\alpha}$.

Finite point estimates for α.



Existing solutions

In addition to the obvious MLE strategy

- Penalized Likelihood Arellano Valle & Azzalini (2013)
- Hellinger distance based Greco (2011)
- semiparametric local likelihood Ma & Hart, (2007)
- Bias Prevention, Sartori (2006)

The big problem

The likelihood can be multi-modal.

- Technical problem: MLE difficult to find -
  or . . .
  in a Bayesian setting, Gibbs sampling does not necessarily works . . .
- Statistical problem: nearly unidentifiability

An alternative:

Exploit the latent structure of the SN density in order to obtain an augmented likelihood.
This will hopefully help solving the former problem ...

The following result holds.

Proposition.

Let $\boldsymbol{\Omega}$ be a correlation matrix, $\boldsymbol{\delta}$ a p-dimensional vector and $\boldsymbol{\alpha} = (1 - \boldsymbol{\delta}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\delta})^{-\frac{1}{2}} \boldsymbol{\Omega}^{-1} \boldsymbol{\delta}$. Define

$$\begin{pmatrix} Z \\ \mathbf{X} \end{pmatrix} \sim N_{p+1}\left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} 1 & \boldsymbol{\delta}^T \\ \boldsymbol{\delta} & \boldsymbol{\Omega} \end{pmatrix}\right] \text{ and } \mathbf{U} = \begin{cases} \mathbf{X} & Z \geq 0 \\ -\mathbf{X} & Z < 0 \end{cases}.$$

Then, (a) the random vector $\mathbf{Y} = \boldsymbol{\omega}\mathbf{U} + \boldsymbol{\xi} \sim SN_p(\boldsymbol{\Sigma}, \boldsymbol{\xi}, \boldsymbol{\alpha})$, with $\boldsymbol{\Sigma} = \boldsymbol{\omega}\boldsymbol{\Omega}\boldsymbol{\omega}$, and (b) the joint density of $(\mathbf{Y}, Z)$ is given by

$$f_{p+1}(\mathbf{y}, z) = f_p(\mathbf{y} \mid z) f(z) = N_p\left(\boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\delta}|z|, \boldsymbol{\omega}(\boldsymbol{\Omega} - \boldsymbol{\delta}\boldsymbol{\delta}')\boldsymbol{\omega}\right) \times N_1(0, 1).$$

Also, write

$$\psi = \boldsymbol{\omega}\boldsymbol{\delta}; \quad \boldsymbol{\omega}(\boldsymbol{\Omega} - \boldsymbol{\delta}\boldsymbol{\delta}')\boldsymbol{\omega} = \boldsymbol{\Sigma} - \psi\psi' = \mathbf{G};$$

The parameter vector is then
$\boldsymbol{\theta}^* = (\boldsymbol{\delta}, \boldsymbol{\Sigma}, \boldsymbol{\xi})$ - more suitable for elicitation -
$\boldsymbol{\theta} = (\psi, \mathbf{G}, \boldsymbol{\xi})$ more suitable for computation.

Augmented Likelihood Function

The above result allows to set up efficient MCMC and/or Population Monte Carlo algorithms. The augmented likelihood function is

$$
\begin{aligned}
L(\boldsymbol{\theta}; \mathbf{y}, \mathbf{z}) & \propto \prod_{i=1}^{n} \left\{ \varphi_p(\mathbf{y}_i - \boldsymbol{\xi} - \boldsymbol{\psi} \mid z_i \mid ; \boldsymbol{\Sigma} - \boldsymbol{\psi}\boldsymbol{\psi}') \times \varphi_1(z_i; 1) \right\} \\
& = \frac{1}{\mid \mathbf{G} \mid^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} z_i^2 \right) \\
& \times \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (\mathbf{y}_i - \boldsymbol{\xi} - \boldsymbol{\psi} \mid z_i \mid)' \mathbf{G}^{-1} (\mathbf{y}_i - \boldsymbol{\xi} - \boldsymbol{\psi} \mid z_i \mid) \right) .
\end{aligned}
$$

Warning! The matrix $\mathbf{G}$ must be positive definite $\Rightarrow$ constraint for the values of $\boldsymbol{\delta}$ and $\boldsymbol{\Omega}$ which must be taken into account when exploring the parameter space via simulation methods.
This issue seems to have been neglected in Bayesian literature.

Objective priors

The above formulation makes the SN model almost Gaussian ...
We set, as usual in Bayesian inference,

$$\pi(\boldsymbol{\xi}) \propto 1 \qquad \text{and} \qquad \mathbf{G} \sim IW_p(m, \boldsymbol{\Lambda})$$

[in the limiting - objective Bayes - case $m \to 0, \boldsymbol{\Lambda} \to \mathbf{0}$]

$$\pi(\boldsymbol{\xi}, \mathbf{G}) \propto \frac{1}{\mid G \mid^{\frac{p+1}{2}}}$$

The choice of the prior for $\boldsymbol{\delta}$ is much more delicate.
One must use a proper prior on $\boldsymbol{\delta}$ (or $\boldsymbol{\alpha}$)

- the Jeffreys' prior is improper
- the one-at-the-time reference prior is quite complicated to use but it is proper and it has the required coverage properties.
- a Beta prior (in the $\boldsymbol{\delta}$ parametrization) is a good compromise.

Objective priors for $\boldsymbol{\delta}$

Assume that $\boldsymbol{\Omega} = \mathrm{diag}\,(1,1)$

The Jeffreys' prior in the $\boldsymbol{\alpha}$ set-up is (up to an approximation)

$$\pi(\alpha_1, \alpha_2) \propto \frac{1}{1 + 2\eta^2(\alpha_1^2 + \alpha_2^2)}$$

with $\eta = \pi/2$

which is improper. In the $\boldsymbol{\delta}$ parametrization, for $\boldsymbol{\delta}'\boldsymbol{\delta} \leq 1$,

$$\left|\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\delta}}\right| = \frac{1}{\sqrt{1 - (\boldsymbol{\delta}'\boldsymbol{\delta})}}$$

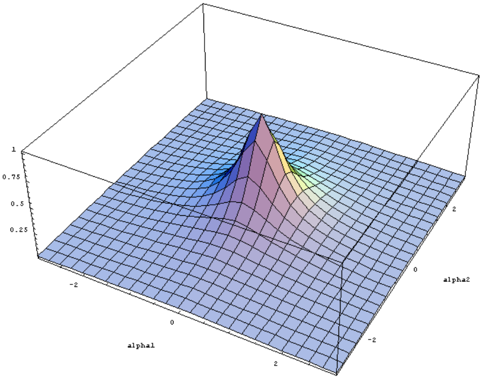
$$\pi(\delta_1, \delta_2) \propto \frac{\sqrt{1 - \delta_1^2 - \delta_2^2}}{1 + (2\eta^2 - 1)(\delta_1^2 + \delta_2^2)}$$

Reference prior

The *proper* reference prior when α_1 is the parameter of interest is

$$\pi_R(\alpha_2 \mid \alpha_1)\pi_R(\alpha_1) \propto \frac{(1+2\eta^2\alpha_1^2)^{1/4}}{(1+2\eta^2(\alpha_1^2+\alpha_2^2))^{3/4}} \frac{1}{\sqrt{(1+2\eta^2\alpha_1^2)}}$$
$$\exp\left(-\frac{1}{4}\int \log\left(1+2\eta^2(\alpha_1^2+\alpha_2^2)\right)\pi_R(\alpha_2 \mid \alpha_1)d\alpha_2\right)$$

Reference prior



The final prior for $\boldsymbol{\delta}$

In practice,

- the prior must depend on $\boldsymbol{\Omega}$;
- set $\beta_i = (1 + \delta_i)/2$
- set β_i's $\overset{\text{iid}}{\sim}$ Beta(.25, .25)

Then

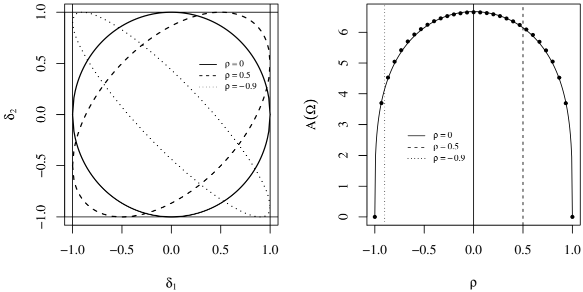
$$\pi(\boldsymbol{\delta}|\boldsymbol{\Omega}) = \frac{1}{A(\boldsymbol{\Omega})} \prod_{j=1}^{p} \left(1 - \delta_j^2)^{-3/4}\right)$$

$A(\boldsymbol{\Omega})$ is the normalizing constant, the ellipsoid of acceptable values for $\boldsymbol{\delta}$.

An example with $p = 2$

Different shapes of the ellipsoid and an approximation of $A(\boldsymbol{\Omega})$

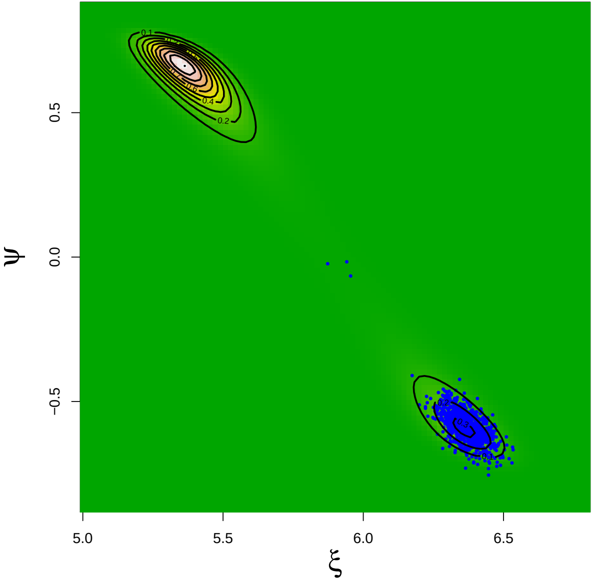
$$A(\boldsymbol{\Omega}) \approx a(1 - \rho^2)^b.$$



Bayesian calculation

- When full conditional distributions are easy to sample from, the Bayesian analysis of latent structure models is usually implemented via Gibbs sampling
- In the SN_p model, all the parameters have full conditional distribution which are (more or less..) simple to sample from.
- However, if the posterior surface is not sufficiently smooth, Markov chain based algorithms risk to be trapped into small portions of the parameter space
- On the other hand, the use of simple importance sampling strategies is complicated by (the crucial!) choice of the importance density.

A simple example of disastrous Gibbs sampling in the scalar SN case (ω known)



Population MonteCarlo algorithm

- PMC algorithms can overcome the above problems, still retaining the efficiency of the full conditional distributions (Celeux, Marin, Robert (2006, CSDA))

PMC

PMC algorithms are essentially Iterated Sampling Importance Resampling (Rubin, 1988) algorithms where, at each iteration, a population of particles is drawn from one (or more than one . . . proposal density (in this case, the full conditional distributions.) Then the particles are re-sampled according to a multinomial scheme with probabilities proportional to the importance weights.

PMC algorithm in detail

Suppose your parameter vector is $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$, $\pi(\boldsymbol{\theta}, \mathbf{z}|\mathbf{x})$ is the un-normalised posterior, and $q(\boldsymbol{\theta}, \mathbf{z})$ is the joint proposal density. For $t = 1, \ldots T$, and $i = 1, \ldots n$,

(a) Select the proposal distribution $q_{it}(\cdot)$

(b) Generate $(\boldsymbol{\theta}_i^{(t)}, \mathbf{z}_i^{(t)}) \sim q_{it}(\cdot)$

(c) Compute $\rho_i^{(t)} = \pi(\boldsymbol{\theta}_i^{(t)}|\mathbf{x})/q_{it}(\boldsymbol{\theta}_i^{(t)})$ and normalise weights so that $\sum_i = \rho_i^{(t)} = 1$.

(d) Resample n values from the $\boldsymbol{\theta}_i^{(t)}$ with replacement, using the weights $\rho_i^{(t)}$, to create the posterior sample at iteration t.

Population MonteCarlo algorithm (2)

- PMC makes use of the full conditional distributions (a usual MCMC device) in a MC perspective, avoiding convergence issues.

Population MonteCarlo algorithm (2)

- PMC makes use of the full conditional distributions (a usual MCMC device) in a MC perspective, avoiding convergence issues.
- The algorithm is replicated several times to guarantee better exploration of the multimodal posterior surface.

Population MonteCarlo algorithm (2)

- PMC makes use of the full conditional distributions (a usual MCMC device) in a MC perspective, avoiding convergence issues.
- The algorithm is replicated several times to guarantee better exploration of the multimodal posterior surface.
- In a sense PMC brings us beyond Importance Sampling and MCMC methods

Practical Implementation

The practical use of the algorithm is too complicate to be illustrated in the general p-dimensional case.
From now on we explicitly consider the case $p = 2$.

Practical Implementation

The practical use of the algorithm is too complicate to be illustrated in the general p-dimensional case.
From now on we explicitly consider the case $p = 2$.
In the PMC approach (and in the presence of latent structure) it is reasonable to use proposal distributions which resemble the full conditionals (Celeux, Marin and Robert, CSDA06).

Practical Implementation

The practical use of the algorithm is too complicate to be illustrated in the general p-dimensional case.
From now on we explicitly consider the case $p = 2$.
In the PMC approach (and in the presence of latent structure) it is reasonable to use proposal distributions which resemble the full conditionals (Celeux, Marin and Robert, CSDA06).
In the $(G, \boldsymbol{\xi}, \boldsymbol{\delta})$-parametrization, the augmented likelihood is

$$\mathcal{L}(G, \xi, \delta; y, z) \propto \frac{1}{|G|^{n/2}} \exp(-\tfrac{1}{2} z' z)$$

$$\exp\{-\tfrac{1}{2} \textstyle\sum_{i=1}^{n} (y_i - \xi - \delta |z_i|)' G^{-1} (y_i - \xi - \delta |z_i|)\}$$

Objective Priors when $p = 2$

$$\pi(\boldsymbol{\xi}, \boldsymbol{\delta}, \mathbf{G}) = \pi(\boldsymbol{\xi})\pi(\boldsymbol{\delta}, \mathbf{G})$$

- $\pi(\boldsymbol{\xi}) \propto 1$
- $\pi(\boldsymbol{\delta} \mid \mathbf{G}) \propto \prod_{j=1}^{p} \left(1 - \delta_j^2\right)^{-3/4}\right) \mathbb{I}_{A(\mathbf{G})}(\boldsymbol{\delta})$
- $\pi(\mathbf{G})$ is such that $\pi(\boldsymbol{\Sigma}) \propto (\omega_{1,1}^2 \omega_{2,2}^2 (1 - \rho^2))^{-1}$

where $A(\mathbf{G}) = \left\{\boldsymbol{\delta} : \left|\boldsymbol{\Sigma} - \boldsymbol{\delta}\boldsymbol{\delta}'\right| > 0\right\}$

which is equivalent to $\boldsymbol{\Sigma} \sim \mathcal{IW}(m = 0, W = \mathbf{0})$ and

$$\omega_{11}^2 = \frac{G_{11}}{1 - \delta_1^2}; \qquad \omega_{22}^2 = \frac{G_{22}}{1 - \delta_2^2}$$

$$\omega_{12} = \rho = \frac{G_{12}}{\omega_{11}\omega_{22}} + \delta_1\delta_2$$

Full conditionals / 1

Then it is easy to derive the full conditional distributions.

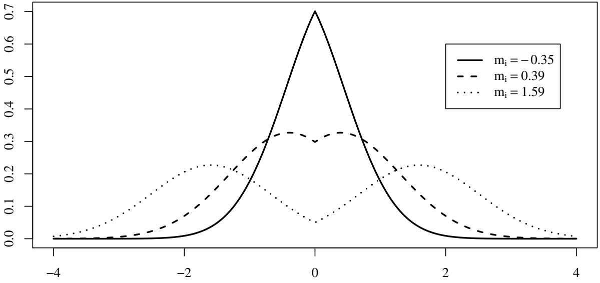
The $z_i's$ are conditionally (on $\boldsymbol{\theta}$) i.i.d. with density

$$f(z_i|\mathbf{y},\boldsymbol{\theta}) = \begin{cases} \mathcal{N}^+(m_i, v) & z_i \geq 0 \\ \mathcal{N}^-(-m_i, v) & z_i < 0 \end{cases}$$

where

$$\begin{aligned} \mathbf{m} &= v\left[(\mathbf{y} - \mathbf{1}_n \otimes \xi)' \, \mathbf{G}^{-1}\psi\right] \\ v &= \left(1 + \psi'\mathbf{G}^{-1}\psi\right)^{-1} \end{aligned}$$

Full conditionals for z_i's for different values of m_i



Full conditionals / 2

$$
\begin{aligned}
\boldsymbol{\xi}|\mathbf{y},\cdots &\sim \mathcal{N}_p\left(\bar{\mathbf{y}} - \psi|\bar{\mathbf{z}}|, \tfrac{1}{n}\mathbf{G}\right) \\
\psi|\mathbf{y},\cdots &\sim \pi(\psi|\mathbf{G})\varphi_p\left(\psi - \frac{\sum_i |z_i|(\mathbf{y}_i - \boldsymbol{\xi})}{\sum_i z_i^2}; \frac{G}{\sum_i z_i^2}\right) \\
\mathbf{G}|\mathbf{y},\cdots &\sim \pi(\mathbf{G})\, IW_p(n+m, W_\star)
\end{aligned}
$$

where

$$W_\star = W + \sum_{i=1}^{n}(\mathbf{y}_i - \psi|z_i| - \boldsymbol{\xi})(\mathbf{y}_i - \psi|z_i| - \boldsymbol{\xi})'$$

PMC-SN Algorithm

0 Initialization: For $t = 0$, choose $\left(\theta_0^{(1)}, \theta_0^{(2)}, \cdots, \theta_0^{(M)}\right)$

1 For $t = 1, \cdots T$, and for $j = 1, \cdots, M$

- For $i = 1, \cdots, n$
  { generate $z_{i,t}^{(j)}$ from $k(\cdot \mid \mathbf{y}_i, \boldsymbol{\theta}_{t-1}^{(j)})$ }
- Generate $\boldsymbol{\theta}_t^{(j)}$ from $\pi(\cdot \mid \mathbf{y}, \mathbf{z}_t^{(j)})$
- Compute

$$n_t^{(j)} = \frac{1}{M} \sum_{l=1}^{M} \frac{\pi(\boldsymbol{\theta}_t^{(j)} \mid \mathbf{y}, \mathbf{z}_t^{(l)})}{k(\mathbf{z}_t^{(l)} \mid \mathbf{y}, \boldsymbol{\theta}_{t-1}^{j})}$$

$$d_t^{(j)} = \frac{1}{M} \sum_{l=1}^{M} \frac{k(\mathbf{z}_t^{(l)} \mid \mathbf{y}, \boldsymbol{\theta}_{t-1}^{j})\pi(\boldsymbol{\theta}_t^{(j)} \mid \mathbf{y}, \mathbf{z}_t^{(l)})}{k(\mathbf{z}_t^{(l)} \mid \mathbf{y}, \boldsymbol{\theta}_{t-1}^{l})}$$

  and $r_t^{(j)} = n_t^{(j)} / d_t^{(j)}$, $\rho_t^{(j)} = r_t^{(j)} / \sum_{h=1}^{M} r_t^{(h)}$

- Resample with replacement from $\left(\theta_t^{(1)}, \theta_t^{(2)}, \cdots, \theta_t^{(M)}\right)$ with weights equal to $\left(\rho_t^{(1)}, \rho_t^{(2)}, \cdots, \rho_t^{(M)}\right)$

Model Selection

Typical problem: comparing two nested models:

Normal vs. Skew-Normal

$$H_0 : \mathbf{Y} \sim N_p(\xi, \boldsymbol{\Sigma}) \text{ vs. } H_1 : \mathbf{Y} \sim SN_p(\xi, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$$

The main tool in Bayesian inference is the Bayes factor

$$B_{01} = \frac{p(\mathbf{y} \mid H_0)}{p(\mathbf{y} \mid H_1)} = \frac{\int_{\boldsymbol{\Sigma}} \int_{\xi} L_0(\xi, \boldsymbol{\Sigma}; \mathbf{y}) \pi_0(\xi, \boldsymbol{\Sigma}) d\xi d\boldsymbol{\Sigma}}{\int_{\boldsymbol{\alpha}} \int_{\boldsymbol{\Sigma}} \int_{\xi} L_1(\xi, \boldsymbol{\Sigma}, \boldsymbol{\alpha}; \mathbf{y}) \pi_1(\xi, \boldsymbol{\Sigma}, \boldsymbol{\alpha}) d\xi d\boldsymbol{\Sigma} d\boldsymbol{\alpha}}$$

- B_{01} is well defined with proper priors
- Improper priors can be used only for those parameters which appears on both the models

$\Rightarrow \pi_1(\boldsymbol{\alpha})$ must be proper.
In this case, $p(\mathbf{y} \mid H_0)$ has a closed form expression,

$\Rightarrow \pi_1(\boldsymbol{\alpha})$ must be proper.
In this case, $p(\mathbf{y} \mid H_0)$ has a closed form expression,
one needs to evaluate $p(\mathbf{y} \mid H_1)$ only!
Expressions for $p(\mathbf{y} \mid H_1)$ are remarkably simple with PMC.

$$p(\mathbf{y} \mid H_1) \approx \frac{\sum_{t=1}^{T} H_t \sum_{j=1}^{N} \tilde{\rho}_j^{(t)}}{N \sum_{t=1}^{T} H_t}.$$

where the $\tilde{\rho}_j$'s are the un-normalised weights, and

$$H_t = -\sum_{i=1}^{N} \rho_i^{(t)} \log(\rho_i^{(t)})$$

is an entropy measure of performance of the t-th iteration of the algorithm. H_t takes high values when the normalised weights of the particles in the t-th iteration. are concentrated around $1/N$.

Chib's Method

Alternatively one use the identity

$$\log p(\mathbf{y} \mid H_1) = \log p_1(\mathbf{y}; \boldsymbol{\theta}) + \log \pi_1(\boldsymbol{\theta}) - \log \pi_1(\boldsymbol{\theta} \mid \mathbf{y}). \qquad (1)$$

which is valid for all $\boldsymbol{\theta}$.
While the first two components of the sum are easy to evaluate, the last one needs to be estimated using the simulation for the vector $\mathbf{z}$.

$$\hat{\pi}_1(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{1}{M} \sum_{j=1}^{M} \pi(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{z}_{(j)}) \qquad (2)$$

This method, when applied in the SN model, requires additional simulations.

Extensions to multivariate skew-t model

Easy, from Dickey's (1968) representation theorem of a t density as a scale mixture of normal densities. Let

$$Z \mid W \sim SN_p\left(\boldsymbol{\psi}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}\right), \qquad W \sim \chi^2_\nu/\nu$$

then,

$$Z \sim \text{Skew-}t_{\nu,p}(\boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\Sigma})$$

with density

$$\begin{aligned} f(z) &= 2\,\frac{\Gamma((\nu+p)/2)\nu^{\nu/2}}{(\pi\nu)^{p/2}\Gamma(\nu/2) \mid \boldsymbol{\Sigma} \mid^{1/2}}\left(1+\frac{Q(\mathbf{z})}{\nu}\right)^{-(\nu+p)/2} \\ &\times \quad P\left\{T_{\nu+p} \leq \boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\mathbf{z}-\boldsymbol{\xi})\sqrt{\frac{\nu+p}{\nu+Q(\mathbf{z})}}\right\} \end{aligned}$$

with

$$Q(\mathbf{z}) = (\mathbf{z}-\boldsymbol{\psi})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\psi})$$

Remarks

- The completion idea is used again at a low computational cost

Remarks

- The completion idea is used again at a low computational cost
- the density of a multivariate Skew-t can be written as the product of
  - a scalar normal density
  - a multivariate normal density
  - a chi squared density

Skew-t model

Given n observations from a p-variate Skew-t

$$\mathbf{y}_i \sim \text{Skew-}t_{\nu,p}(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Sigma}) \quad i = 1, \ldots, n$$

the augmented likelihood is proportional to

$$\mathcal{L}(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Sigma}, \nu; \mathbf{y}, \mathbf{z}, \mathbf{v}) \propto |\Sigma|^{-n/2} \exp\left(-\tfrac{1}{2}\mathbf{z}'\mathbf{z}\right)$$
$$\exp\left\{-\frac{1}{2}\sum_{i=1}^{n} v_i(\mathbf{y}_i - \boldsymbol{\xi} - \boldsymbol{\omega}\boldsymbol{\delta}\frac{|z|}{\sqrt{v}})'\Sigma^{-1}(\mathbf{y}_i - \boldsymbol{\xi} - \boldsymbol{\omega}\boldsymbol{\delta}\frac{|z|}{\sqrt{v}})\right\}$$
$$\frac{(\nu/2)^{(n\nu/2)}}{(\Gamma(\nu/2))^n}(\prod_{i=1}^{n} v_i)^{\nu/2-1} \exp\{-\nu/2\sum_{i=1}^{n} v_i\}$$

with $\omega_j = \Sigma_{jj}^{1/2}$, $j = 1, \ldots, p$ and $\boldsymbol{\omega} = \mathrm{diag}\,(\omega_1, \ldots, \omega_p)$.

Priors

Same as before . . .

$$\pi(\boldsymbol{\xi}, \boldsymbol{\delta}, \Sigma) = \pi(\boldsymbol{\xi})\pi(\boldsymbol{\delta} \mid \Sigma)\pi(\Sigma)\pi(\nu)$$

$$\begin{array}{rcl} \pi(\boldsymbol{\xi}) & \propto & 1 \\ \Sigma & \sim & \mathcal{IW}_p(\rightarrow \mathbf{0}, \rightarrow 0) \\ \pi(\boldsymbol{\delta} \mid \Sigma) & \propto & \prod_{j=1}^{p}(1-\delta_j^2)^{-3/4}\mathbb{I}_{\boldsymbol{\delta}}(\Delta) \\ & & \text{where } \Delta \text{ is the region of admissible values of } \delta|\Sigma. \\ \nu & \sim & \mathrm{Exp}(n_\nu) \end{array}$$

$(\boldsymbol{\xi}, \boldsymbol{\psi}, G, \nu)$ - parametrization

$$\left\{ \begin{array}{rcl} \boldsymbol{\xi} & = & \boldsymbol{\xi} \\ \boldsymbol{\psi} & = & \boldsymbol{\omega\delta} \\ G & = & \Sigma - \boldsymbol{\omega\delta\delta'\omega} \\ \nu & = & \nu \end{array} \right. \quad \Rightarrow \quad \left\{ \begin{array}{rcl} \boldsymbol{\xi} & = & \boldsymbol{\xi} \\ \boldsymbol{\delta}_j & = & (G_{jj} + \psi_j^2)^{-1/2}\boldsymbol{\psi} \\ \Sigma & = & G + \boldsymbol{\psi\psi'} \\ \nu & = & \nu \end{array} \right.$$

Augmented Likelihood Function

$$\mathbf{y}_i \overset{\text{iid}}{\sim} S\text{-}t_{\nu,p}(\boldsymbol{\xi}, \boldsymbol{\psi}, \mathbf{G}) \quad i = 1, \ldots, n,$$

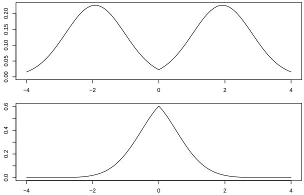
$$\mathcal{L}(\boldsymbol{\xi}, \boldsymbol{\delta}, G, \nu \mid \mathbf{y}, \mathbf{z}, \mathbf{v}) \propto |G|^{-n/2} \exp\left\{-\tfrac{1}{2}\mathbf{z}'\mathbf{z}\right\}$$
$$\exp\left\{ \underbrace{-\tfrac{1}{2}\sum_{i=1}^{n} v_i(\mathbf{y}_i - \boldsymbol{\xi} - \boldsymbol{\psi}\frac{|z|}{\sqrt{v}})' G^{-1}(\mathbf{y}_i - \boldsymbol{\xi} - \boldsymbol{\psi}\frac{|z|}{\sqrt{v}})}_{\zeta} \right\}$$
$$\frac{(\nu/2)^{(n\nu/2)}}{(\Gamma(\nu/2))^n}(\prod_{i=1}^{n} v_i)^{\nu/2-1} \exp\{-\nu/2 \sum_{i=1}^{n} v_i\}$$

where

$$\zeta = \mathbf{tr}\left(G^{-1} \underbrace{\sum_{i=1}^{n} v_i(\mathbf{y}_i - \boldsymbol{\xi} - \boldsymbol{\psi}\frac{|z|}{\sqrt{v}})(\mathbf{y}_i - \boldsymbol{\xi} - \boldsymbol{\psi}\frac{|z|}{\sqrt{v}})'}_{\Lambda} \right).$$

Full conditionals / 1

$$f(z_i|\cdots) = \begin{cases} \phi^+(m_i, v_\theta) & z_i \geq 0 \\ \phi^-(-m_i, v_\theta) & z_i < 0 \end{cases}$$



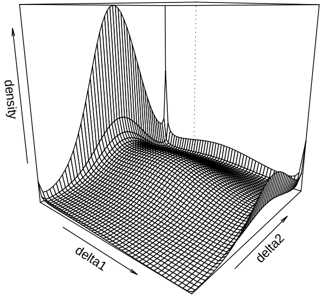
where

$$\begin{aligned} v_\theta &= (1 + \boldsymbol{\psi}' G^{-1} \boldsymbol{\psi})^{-1} \\ m_i &= \sqrt{v_i}\, v_\theta (\boldsymbol{\psi}' G^{-1} (\mathbf{y}_i - \boldsymbol{\xi})) \end{aligned}$$

Full conditionals / 2

$$f(\psi \mid \cdots) \propto \prod_{j=1}^{p} (1-\delta_j^2)^{-3/4} A(G) \text{ (parte della priori, trascurata)}$$

$$\phi_p\left(\psi \,\middle|\, \frac{\sum_{i=1}^n |z_i| \sqrt{v_i}(\mathbf{y}_i - \boldsymbol{\xi})}{\sum_{i=1}^n z_i^2}, \frac{G}{\sum_{i=1}^n z_i^2}\right) \mathbb{I}_{\psi}(\Psi)$$



Full conditionals / 3

$$
\begin{aligned}
\xi | \cdots &\sim \mathcal{N}_p\left(\frac{\overline{\mathbf{yv}}}{\bar{\mathbf{v}}} - \psi\frac{\overline{|\mathbf{z}|\sqrt{\mathbf{v}}}}{\bar{\mathbf{v}}}, \frac{1}{n\bar{\mathbf{v}}}\,G\right) \\
f(G|\cdots) &\propto \pi(G)|G|^{-n/2}\exp(-\tfrac{1}{2}\mathbf{tr}(G^{-1}\Lambda)) = \\
&= \pi(G)\,\mathcal{IW}(n-p-1, \Lambda)
\end{aligned}
$$

Full conditionals / 4

$$
\begin{aligned}
f(v_i|\cdots) &\propto v_i^{(\nu+p-2)/2} \exp\left\{-\frac{A_i}{2} v_i - B_i\sqrt{v_i}\right\} \\
&\text{or} \\
f(\eta_i|\cdots) &= \eta_i^{\nu+p-1} \exp\left\{-\tfrac{1}{2} A_i(\eta_i - A_i^{-1}B_i)^2\right\}
\end{aligned}
$$

Full conditionals / 4

$$
\begin{aligned}
f(v_i|\cdots) &\propto v_i^{(\nu+p-2)/2} \exp\left\{-\frac{A_i}{2} v_i - B_i\sqrt{v_i}\right\} \\
&\text{or} \\
f(\eta_i|\cdots) &= \eta_i^{\nu+p-1} \exp\left\{-\tfrac{1}{2} A_i(\eta_i - A_i^{-1}B_i)^2\right\}
\end{aligned}
$$

where $\eta_i = \sqrt{v_i}$,
$A_i = \nu + (y_i - \boldsymbol{\xi})' G^{-1}(y_i - \boldsymbol{\xi})$
$B_i = (y_i - \xi)' G^{-1} \psi |z_i|$.

Full conditionals / 4

$$
\begin{array}{rcl}
f(v_i|\cdots) & \propto & v_i^{(\nu+p-2)/2}\exp\left\{-\frac{A_i}{2}v_i - B_i\sqrt{v_i}\right\} \\
 & \text{or} & \\
f(\eta_i|\cdots) & = & \eta_i^{\nu+p-1}\exp\left\{-\frac{1}{2}A_i(\eta_i - A_i^{-1}B_i)^2\right\}
\end{array}
$$

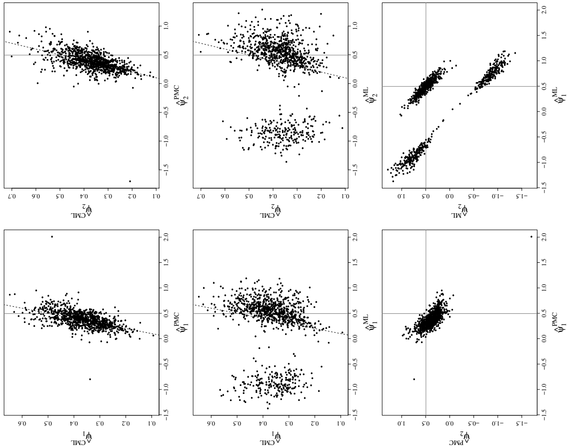
where $\eta_i = \sqrt{v_i}$,
$A_i = \nu + (y_i - \boldsymbol{\xi})' G^{-1}(y_i - \boldsymbol{\xi})$
$B_i = (y_i - \xi)' G^{-1}\psi |z_i|$.

Both densities can be sampled using a slice sampler. Finally

$$f(\nu|\cdots) \propto (\nu/2)^{n\nu/2}(\Gamma(\nu/2))^{-n}\exp\{-g\nu\}$$

with $g = \frac{1}{2}\sum_i \log(v_i) + \frac{1}{2}v_i^{-1} + n_\nu$.
Geweke (1992) provides an algorithm to sample from this.

SN_2 case: comparison of estimation methods through simulation: 10^4 samples, $\rho = -.5$, $\boldsymbol{\omega} = (1,1)'$, $\psi = (.495, .495)'$. It corresponds to $\boldsymbol{\alpha} \approx (7.02, 7.02)'$.



An illustration

240 obsv'n of monthly returns on *ABM Industries Inc.* and *The Boeing Co.* (Oct. '92 –Oct. 2012)

An illustration

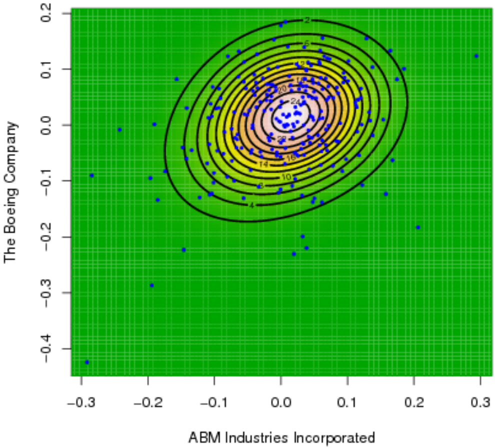
- 240 obsv'n of monthly returns on *ABM Industries Inc.* and *The Boeing Co.* (Oct. '92 –Oct. 2012)
- PMC algorithm ($T = 25, n = 3 \times 10^4$), objective priors

An illustration

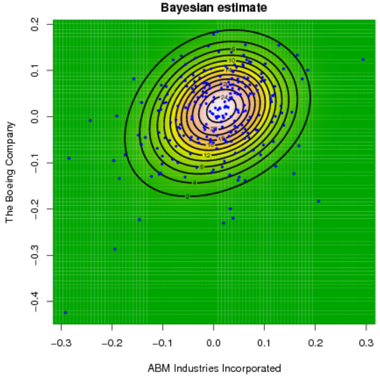
240 obsv'n of monthly returns on *ABM Industries Inc.* and *The Boeing Co.* (Oct. '92 –Oct. 2012)

PMC algorithm ($T = 25, n = 3 \times 10^4$), objective priors

Observed values and estimated density:



Comparison with ML approach



References

- Liseo, B. & Parisi, A. (2013) Bayesian inference for the multivariate skew-normal model: a Population MonteCarlo approach. *Computational Statistics and Data Analysis*, 63, pp. 125–138.
- Liseo, B. & Parisi, A. (2013) Adaptive Importance Sampling Methods for the multivariate Skew-Student distribution and skew-t copula. *Manuscript under preparation*