Web-Supplement for:

Variational Message Passing for Skew t Regression Algebraic and Numerical Details

BY LUCA MAESTRINI[a] AND MATT P. WAND[b]

[a]*Department of Statistical Sciences, University of Padova, Via Cesare Battisti 241, 35121 Padova PD, Italy*

[b]*School of Mathematical and Physical Sciences, University of Technology Sydney, PO Box 123, Broadway 2007, Australia*

S.1 Distributions

Here we describe the distribution functions mentioned in this article.

S.1.1 Skew t distribution

According to the formulation introduced by Azzalini & Capitanio (2003), a $d \times 1$ random vector $\boldsymbol{x}$ is distributed as a d-variate skew t distribution, written $\boldsymbol{x} \sim \text{Skew-}t_d\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu\right)$, if its probability density function is

$$p\left(\boldsymbol{x}\right) = 2t_d\left(\boldsymbol{x};\nu\right) T_1\left\{\boldsymbol{\lambda}^T\boldsymbol{\Omega}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\left(\frac{\nu+d}{Q_{\boldsymbol{x}}+\nu}\right)^{1/2};\nu+d\right\}$$

where

$$Q_{\boldsymbol{x}} = \left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right), \quad t_d\left(\boldsymbol{x};\nu\right) = \frac{\Gamma\left\{\left(\nu+d\right)/2\right\}}{\left|\boldsymbol{\Sigma}\right|^{1/2}\left(\pi\nu\right)^{d/2}\Gamma\left(\nu/2\right)}\left(1+\frac{Q_{\boldsymbol{x}}}{\nu}\right)^{-(\nu+d)/2}$$

is a d-dimensional t-variate density function with ν degrees of freedom and $T_1\left(y;\nu+d\right)$ indicates the scalar t distribution function with $\nu+d$ degrees of freedom. The vectors $\boldsymbol{\mu}, \boldsymbol{\lambda} \in \mathbb{R}^d$ are location and shape (skewness) parameter vectors respectively, while ν is the number of degrees of freedom. Also, $\boldsymbol{\Omega}$ is the diagonal matrix having the square root of the diagonal elements of the $d \times d$ full rank covariance matrix $\boldsymbol{\Sigma}$ on its main diagonal such that $\mathbf{R} = \boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Omega}^{-1}$ is the correlation matrix associated with $\boldsymbol{\Sigma}$. The skew t distribution approaches the skew normal distribution as $\nu \to \infty$.

In our univariate formulation, written as Skew-$t\left(\mu, \sigma^2, \lambda, \nu\right)$, the multivariate parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\lambda}$ respectively correspond to the univariate parameters μ, σ^2 and λ with $\boldsymbol{\Omega}$ corresponding to $\sqrt{\sigma^2}$.

S.1.2 Other distributions

The inverse square root Nadarajah, Moon Rock and Sea Sponge distributions are defined in the supplementary material of McLean and Wand (2018). The remaining distributions follow the convention of Wand (2017) and related supplementary material.

S.2 Derivation of Algorithm 2

We here derive the algorithm based on the product density restriction (9) and provided in this supplementary material, to which we refer as Algorithm 2.

It follows from (8) that the logarithm of the normal density term in the likelihood factor is

$$\log p\left(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2\right) = -\frac{1+\lambda^2}{2\sigma^2}\left(\boldsymbol{y}-\boldsymbol{A\theta}-\frac{\sigma\lambda\left|\boldsymbol{a}_1\right|\odot\sqrt{\boldsymbol{a}_2}}{\sqrt{1+\lambda^2}}\right)^T \mathrm{diag}\left(\boldsymbol{a}_2\right)^{-1}\left(\boldsymbol{y}-\boldsymbol{A\theta}-\frac{\sigma\lambda\left|\boldsymbol{a}_1\right|\odot\sqrt{\boldsymbol{a}_2}}{\sqrt{1+\lambda^2}}\right)$$
$$-\frac{1}{2}\boldsymbol{1}_n\log\left(\boldsymbol{a_2}\right)+\frac{n}{2}\left\{\log\left(1+\lambda^2\right)-\log\left(\sigma^2\right)\right\}+\text{const}$$

where ‘const’ indicates terms not depending on the likelihood parameters.

With simple applications of formulae (2)–(5) and steps similar to those given in Sections 4.1.5 and S.2.5.5 of Wand (2017) for the Gaussian likelihood fragment and in Sections S.3.2 and S.3.4 of McLean and Wand (2018) for the t likelihood and skew normal fragment updates we derive the message updates of the VMP algorithm. From hereafter we denote with ‘const’ terms that do not depend on the variable(s) of interest.

The message from $p\left(y|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2\right)$ to $\boldsymbol{\theta}$ is proportional to a multivariate normal density function with natural parameter update

$$\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\boldsymbol{\theta}} \longleftarrow \left(1+\mu_{q(\lambda^2)}\right)\mu_{q(1/\sigma^2)}\begin{bmatrix}\boldsymbol{A}^T\mathrm{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\boldsymbol{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{y}\\ -\frac{1}{2}\mathrm{vec}\left(\boldsymbol{A}^T\mathrm{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\boldsymbol{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{A}\right)\end{bmatrix}$$
$$-\mu_{q\left(\lambda\sqrt{1+\lambda^2}\right)}\mu_{q(1/\sigma)}\begin{bmatrix}\boldsymbol{A}^T\mathrm{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\boldsymbol{1}_n/\sqrt{\boldsymbol{a}_2}\right)\right\}E_{q(\boldsymbol{a}_1)}\left|\boldsymbol{a}_1\right|\\ \boldsymbol{0}\end{bmatrix},$$

where

$$\mu_{q\left(1/\sigma^k\right)} \equiv \int_0^\infty \left(1/\sigma^k\right)q^*\left(\sigma^2\right)d\sigma^2 \quad \text{for} \quad k=1,2,$$
$$\mu_{q(\lambda^2)} \equiv \int_{-\infty}^\infty \lambda^2 q^*\left(\lambda\right)d\lambda$$
$$\text{and} \quad \mu_{q\left(\lambda\sqrt{1+\lambda^2}\right)} \equiv \int_{-\infty}^\infty \lambda\sqrt{1+\lambda^2}q^*\left(\lambda\right)d\lambda,$$

with $q^*\left(\sigma^2\right)$ proportional to

$$m_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\sigma^2}\left(\sigma^2\right)m_{\sigma^2\to p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)}\left(\sigma^2\right)$$

and $q^*\left(\lambda\right)$ similarly defined. $E_{q(\boldsymbol{a}_1)}$ denotes expectation with respect to $q^*\left(\boldsymbol{a}_1\right) \equiv \prod_{i=1}^n q^*\left(a_{1i}\right)$, where $q^*\left(a_{1i}\right)$ is proportional to

$$m_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to a_{1i}}\left(a_{1i}\right)m_{a_{1i}\to p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)}\left(a_{1i}\right) = m_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to a_{1i}}\left(a_{1i}\right)m_{p(\boldsymbol{a}_1)\to a_{1i}}\left(a_{1i}\right)$$

and $E_{q(\boldsymbol{a}_2)}$ is similarly defined with $q^*\left(a_{2i}\right)$ being proportional to

$$m_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to a_{2i}}\left(a_{2i}\right)m_{a_{2i}\to p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)} = m_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to a_{2i}}\left(a_{2i}\right)m_{p(\boldsymbol{a}_2|\nu)\to a_{2i}}\left(a_{2i}\right).$$

The message from $p\left(y|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2\right)$ to σ^2 is proportional to an inverse square root Nadarajah density function with natural parameter update

$$\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\sigma^2} \longleftarrow \begin{bmatrix} -n/2 \\ \\ \mu_{q\left(\lambda\sqrt{1+\lambda^2}\right)}\left\{\boldsymbol{y}-\boldsymbol{A}E_{q(\boldsymbol{\theta})}\left(\boldsymbol{\theta}\right)\right\}^T\mathrm{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\boldsymbol{1}_n/\sqrt{\boldsymbol{a}_2}\right)\right\}E_{q(\boldsymbol{a}_1)}\left|\boldsymbol{a}_1\right| \\ \\ \left(1+\mu_{q(\lambda^2)}\right)G_{\text{VMP}}\Big(\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\leftrightarrow\boldsymbol{\theta}};\boldsymbol{A}^T\mathrm{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\boldsymbol{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{A}, \\ \boldsymbol{A}^T\mathrm{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\boldsymbol{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{y},\boldsymbol{y}^T\mathrm{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\boldsymbol{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{y}\Big)\end{bmatrix},$$

where $E_{q(\boldsymbol{\theta})}$ denotes expectation with respect to the normalization of

$$m_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right) m_{\boldsymbol{\theta}\to p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)}\left(\boldsymbol{\theta}\right).$$

The treatment of $m_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\sigma^2}\left(\sigma^2\right)$ is analogous to that for the messages from the likelihood factor to σ^2 for the asymmetric Laplace distribution in Section S.3.3 of McLean and Wand (2018).

The message from $p\left(y|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2\right)$ to λ is proportional to density functions within the Sea Sponge exponential family with natural parameter update

$$\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\lambda} \longleftarrow \left[\begin{array}{c} n/2 \\ \\ \mu_{q(1/\sigma^2)}G_{\text{VMP}}\Big(\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\leftrightarrow\boldsymbol{\theta}}; \boldsymbol{A}^T\text{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\boldsymbol{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{A}, \\ \boldsymbol{A}^T\text{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\boldsymbol{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{y}, \boldsymbol{y}^T\text{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\boldsymbol{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{y}\Big) - \frac{1}{2}E_{q(\boldsymbol{a}_1)}\left\|\boldsymbol{a}_1\right\|^2 \\ \\ \mu_{q(1/\sigma)}\left\{\boldsymbol{y}-\boldsymbol{A}E_{q(\boldsymbol{\theta})}\left(\boldsymbol{\theta}\right)\right\}^T\text{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\boldsymbol{1}_n/\sqrt{\boldsymbol{a}_2}\right)\right\}E_{q(\boldsymbol{a}_1)}\left|\boldsymbol{a}_1\right| \end{array}\right],$$

where $\mu_{q\left(1/\boldsymbol{a}_2^k\right)} \equiv \int_0^\infty \left(1/\boldsymbol{a}_2^k\right) q^*\left(\boldsymbol{a}_2\right) d\boldsymbol{a}_2$, for $k = 1/2, 1$.

The messages from $p\left(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2\right)$ to the a_{1i}, $1\le i\le n$, variables are

$$m_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to a_{1i}}\left(a_{1i}\right) = \exp\left\{\left[\begin{array}{c}|a_{1i}| \\ a_{1i}^2\end{array}\right]^T \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to a_{1i}}\right\}$$

with natural parameter update

$$\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to a_{1i}} \longleftarrow \left[\begin{array}{c}\mu_{q(1/\sigma)}\mu_{q\left(\lambda\sqrt{1+\lambda^2}\right)}\left\{E_{q(\boldsymbol{a}_2)}\left(1/\sqrt{\boldsymbol{a}_2}\right)\right\}_i\left\{\boldsymbol{y}-\boldsymbol{A}E_{q(\boldsymbol{\theta})}\left(\boldsymbol{\theta}\right)\right\}_i \\ -\frac{1}{2}\mu_{q(\lambda^2)}\end{array}\right].$$

Messages from $p\left(\boldsymbol{a}_1\right)$ to a_{1i}, $1\le i\le n$, are

$$m_{p(\boldsymbol{a}_1)\to a_{1i}}\left(a_{1i}\right) = \exp\left(-\frac{1}{2}a_{1i}^2\right),$$

hence

$$q^*\left(a_{1i}\right) \propto \exp\left\{\left[\begin{array}{c}|a_{1i}| \\ a_{1i}^2\end{array}\right]^T\left[\begin{array}{c}\mu_{q(1/\sigma)}\mu_{q\left(\lambda\sqrt{1+\lambda^2}\right)}\left\{E_{q(\boldsymbol{a}_2)}\left(1/\sqrt{\boldsymbol{a}_2}\right)\right\}_i\left\{\boldsymbol{y}-\boldsymbol{A}E_{q(\boldsymbol{\theta})}\left(\boldsymbol{\theta}\right)\right\}_i \\ -\frac{1}{2}\left(1+\mu_{q(\lambda^2)}\right)\end{array}\right]\right\}.$$

Standard manipulations involving the standard normal density provide expressions for the expectations with respect to $E_{q(\boldsymbol{a}_1)}$

$$E_{q(\boldsymbol{a}_1)}\left|\boldsymbol{a}_1\right| = \frac{\boldsymbol{\omega}_3+\zeta'\left(\boldsymbol{\omega}_3\right)}{\sqrt{1+\mu_{q(\lambda^2)}}} \quad \text{and} \quad E_{q(\boldsymbol{a}_1)}\left\|\boldsymbol{a}_1\right\|^2 = \frac{n+\boldsymbol{1}_n^T\left[\boldsymbol{\omega}_3\odot\left\{\boldsymbol{\omega}_3+\zeta'\left(\boldsymbol{\omega}_3\right)\right\}\right]}{1+\mu_{q(\lambda^2)}},$$

where $\boldsymbol{\omega}_3$ is defined in Algorithm 2. The previous expressions involve the first derivative of $\zeta(x) \equiv \log\left(2\Phi\left(x\right)\right)$, that is, $\zeta'\left(x\right) \equiv \phi\left(x\right)/\Phi\left(x\right)$ with ϕ and Φ distribution density function and cumulative distribution function of the standard normal respectively. The function `zeta()` within the `R` package `sn` Azzalini (2017) supports stable computation of ζ'.

The messages from $p\left(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2\right)$ to a_{2i}, $1\le i\le n$, are

$$m_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to a_{2i}}\left(a_{2i}\right) = \exp\left\{\left[\begin{array}{c}\log\left(a_{2i}\right) \\ 1/\sqrt{a_{2i}} \\ 1/a_{2i}\end{array}\right]^T\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to a_{2i}}\right\},$$

with natural parameter update

$$\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to a_{2i}} \longleftarrow \begin{bmatrix} -1/2 \\ \mu_{q(1/\sigma)}\mu_{q\left(\lambda\sqrt{1+\lambda^2}\right)} \left\{\boldsymbol{y}-\boldsymbol{A}E_{q(\boldsymbol{\theta})}\left(\boldsymbol{\theta}\right)\right\}_i \left(E_{q(\boldsymbol{a}_1)}\left|\boldsymbol{a}_1\right|\right)_i \\ -\frac{1}{2}\mu_{q(1/\sigma^2)}\left(1+\mu_{q(\lambda^2)}\right)E_{q(\boldsymbol{\theta})}\left\{\left(\boldsymbol{y}-\boldsymbol{A\theta}\right)_i\right\}^2 \end{bmatrix},$$

which is within the inverse square root Nadarajah family, that provides expressions for expectations with respect to $E_{q(\boldsymbol{a}_2)}$ included in the algorithm by conjugacy with

$$m_{p(\boldsymbol{a}_{2i}|\nu)\to a_{2i}}\left(a_{2i}\right) = \exp\left\{\begin{bmatrix} \log\left(a_{2i}\right) \\ 1/a_{2i} \end{bmatrix}^T \begin{bmatrix} -\frac{1}{2}\mu_{q(\nu)}-1 \\ -\frac{1}{2}\mu_{q(\nu)} \end{bmatrix}\right\}$$

from the inverse chi-squared family.

The natural parameter vector of the message from $p\left(\boldsymbol{a}_2|\nu\right)$ to ν involves expectation with respect to $\boldsymbol{a}_2$ only and has form

$$\boldsymbol{\eta}_{p(\boldsymbol{a}_2|\nu)\to\nu} \longleftarrow \begin{bmatrix} n \\ -\boldsymbol{1}_n^T E_{q(\boldsymbol{a}_2)}\left\{\log\left(\boldsymbol{a}_2\right)+\boldsymbol{1}_n/\boldsymbol{a}_2\right\} \end{bmatrix},$$

which is proportional to a factor of 2 rescaling of a Moon Rock density function. Imposing conjugacy, the message $m_{\nu\to p(\boldsymbol{a}_2|\nu)}\left(\nu\right)$ is proportional to the same exponential family. Therefore

$$q^*\left(\nu\right) \propto \exp\left\{\begin{bmatrix} (\nu/2)\log\left(\nu/2\right)-\log\left(\Gamma\left(\nu/2\right)\right) \\ (\nu/2) \end{bmatrix}^T \boldsymbol{\eta}_{p(\boldsymbol{a}_2|\nu)\leftrightarrow\nu}\right\},$$

which leads to

$$\mu_{q(\nu)} \equiv \int_0^\infty \nu q^*\left(\nu\right)d\nu.$$

Note that the algorithm requires initialization of one of the vectors of expectations involving the two auxiliary variables $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$. In Algorithm 2 we choose to initialize $E_{q(\boldsymbol{a}_2)}\left(\boldsymbol{1}_n/\sqrt{\boldsymbol{a}_2}\right)$.

S.3 Derivation of Algorithm 1

We now derive the algorithm based on product density restriction (11). The main implications in terms of algebra passing from assumption (9) to assumption (11) concern the auxiliary variables. Expectations with respect to $E_{q(\boldsymbol{a}_1)}$ and $E_{q(\boldsymbol{a}_2)}$ are replaced by the joint expectation $E_{q(\boldsymbol{a}_1,\boldsymbol{a}_2)}$. Following steps similar to those for Algorithm 2 we obtain

$$q^*\left(a_{1i},a_{2i}\right) \propto \exp\left\{\begin{bmatrix} a_{1i}^2 \\ \left|a_{1i}\right|/\sqrt{a_{2i}} \\ 1/a_{2i} \\ \log\left(a_{2i}\right) \end{bmatrix}^T \boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right\} = \exp\left\{\begin{bmatrix} a_{1i}^2 \\ \left|a_{1i}\right|/\sqrt{a_{2i}} \\ 1/a_{2i} \\ \log\left(a_{2i}\right) \end{bmatrix}^T \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix}\right\},$$

where we use the shorthand:

$$\begin{aligned} \eta_1 &\equiv \left(\boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right)_1 = -\frac{1}{2}\left(1+\mu_{q(\lambda^2)}\right), \\ \eta_2 &\equiv \left(\boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right)_2 = \mu_{q\left(\lambda\sqrt{1+\lambda}\right)}\mu_{q(1/\sigma)}\boldsymbol{\tau}_1, \\ \eta_3 &\equiv \left(\boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right)_3 = \left(1+\mu_{q(\lambda^2)}\right)\mu_{q(1/\sigma^2)}\boldsymbol{\tau}_2-\frac{1}{2}\mu_{q(\nu)} \\ \text{and} \quad \eta_4 &\equiv \left(\boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right)_4 = -\frac{1}{2}\left(3+\mu_{q(\nu)}\right). \end{aligned}$$

Algorithm 2 The inputs, updates and outputs of the skew t likelihood fragment assuming $q\left(\boldsymbol{\theta},\sigma^2,\lambda,\nu,\boldsymbol{a}_1,\boldsymbol{a}_2\right)=q\left(\boldsymbol{\theta}\right)q\left(\sigma^2\right)q\left(\lambda\right)q\left(\nu\right)\prod_{i=1}^n q\left(a_{1i}\right)q\left(a_{2i}\right)$.

Data Inputs: $\boldsymbol{y}$, $\boldsymbol{A}$.

Parameter Inputs: $\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\boldsymbol{\theta}}$, $\boldsymbol{\eta}_{\boldsymbol{\theta}\to p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)}$, $\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\sigma^2}$, $\boldsymbol{\eta}_{\sigma^2\to p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)}$, $\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\lambda}$, $\boldsymbol{\eta}_{\lambda\to p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)}$, $\boldsymbol{\eta}_{p(\boldsymbol{a}_2|\nu)\to\nu}$, $\boldsymbol{\eta}_{\nu\to p(\boldsymbol{a}_2|\nu)}$, $E_{q(\boldsymbol{a}_2)}\left(\mathbf{1}_n/\sqrt{\boldsymbol{a}_2}\right)$.

Updates:

$$\mu_{q(1/\sigma)}\longleftarrow(E\boldsymbol{T})_2^{\text{ISRN}}\left(\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\leftrightarrow\sigma^2}\right)$$

$$\mu_{q(1/\sigma^2)}\longleftarrow(E\boldsymbol{T})_3^{\text{ISRN}}\left(\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\leftrightarrow\sigma^2}\right)$$

$$\mu_{q(\lambda^2)}\longleftarrow(E\boldsymbol{T})_2^{\text{SS}}\left(\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\leftrightarrow\lambda}\right)$$

$$\mu_{q\left(\lambda\sqrt{1+\lambda^2}\right)}\longleftarrow(E\boldsymbol{T})_3^{\text{SS}}\left(\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\leftrightarrow\lambda}\right)$$

$$\mu_{q(\nu)}\longleftarrow 2\,(E\boldsymbol{T})_2^{\text{MR}}\left(\boldsymbol{\eta}_{p(\boldsymbol{a}_2|\nu)\leftrightarrow\nu}\right)$$

$$\boldsymbol{\omega}_1\longleftarrow\boldsymbol{y}+\frac{1}{2}\boldsymbol{A}\left\{\text{vec}^{-1}\left(\left(\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\leftrightarrow\boldsymbol{\theta}}\right)_2\right)\right\}^{-1}\left(\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\leftrightarrow\boldsymbol{\theta}}\right)_1$$

$$\boldsymbol{\omega}_2\longleftarrow E_{q(\boldsymbol{a}_2)}\left(\frac{\mathbf{1}_n}{\sqrt{\boldsymbol{a}_2}}\right)\odot\boldsymbol{\omega}_1$$

$$\boldsymbol{\omega}_3\longleftarrow\frac{\mu_{q(1/\sigma)}\mu_{q\left(\lambda\sqrt{1+\lambda^2}\right)}\boldsymbol{\omega}_2}{\sqrt{1+\mu_{q(\lambda^2)}}}$$

$$E_{q(\boldsymbol{a}_1)}\left|\boldsymbol{a}_1\right|\longleftarrow\frac{\boldsymbol{\omega}_3+\zeta'\left(\boldsymbol{\omega}_3\right)}{\sqrt{1+\mu_{q(\lambda^2)}}}$$

$$E_{q(\boldsymbol{a}_1)}\left\|\boldsymbol{a}_1\right\|^2\longleftarrow\frac{n+\mathbf{1}_n^T\left[\boldsymbol{\omega}_3\odot\left\{\boldsymbol{\omega}_3+\zeta'\left(\boldsymbol{\omega}_3\right)\right\}\right]}{1+\mu_{q(\lambda^2)}}$$

$$\boldsymbol{\omega}_4\longleftarrow\left[G_{\text{VMP}}\left(\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\leftrightarrow\boldsymbol{\theta}};\boldsymbol{A}^T\boldsymbol{e}_i\boldsymbol{e}_i^T\boldsymbol{A},\boldsymbol{A}^T\boldsymbol{e}_i\boldsymbol{e}_i^T\boldsymbol{y},y_i^2\right)\right]_{1\le i\le n}$$

$$\boldsymbol{\eta}_{q(\boldsymbol{a}_2)}\longleftarrow\begin{bmatrix}-\frac{1}{2}\mu_{q(\nu)}-\frac{3}{2}\\ \mu_{q(1/\sigma)}\mu_{q\left(\lambda\sqrt{1+\lambda^2}\right)}\boldsymbol{\omega}_1\odot E_{q(\boldsymbol{a}_1)}\left|\boldsymbol{a}_1\right|\\ \mu_{q(1/\sigma^2)}\left(1+\mu_{q(\lambda^2)}\right)\boldsymbol{\omega}_4-\frac{1}{2}\mu_{q(\nu)}\end{bmatrix}$$

$$E_{q(\boldsymbol{a}_2)}\left(\log\left(\boldsymbol{a}_2\right)\right)\longleftarrow(E\boldsymbol{T})_1^{\text{ISRN}}\left(\boldsymbol{\eta}_{q(\boldsymbol{a}_2)}\right)$$

$$E_{q(\boldsymbol{a}_2)}\left(\mathbf{1}_n/\sqrt{\boldsymbol{a}_2}\right)\longleftarrow(E\boldsymbol{T})_2^{\text{ISRN}}\left(\boldsymbol{\eta}_{q(\boldsymbol{a}_2)}\right)$$

$$E_{q(\boldsymbol{a}_2)}\left(\mathbf{1}_n/\boldsymbol{a}_2\right)\longleftarrow(E\boldsymbol{T})_3^{\text{ISRN}}\left(\boldsymbol{\eta}_{q(\boldsymbol{a}_2)}\right)$$

$$\begin{aligned}\boldsymbol{\omega}_5\longleftarrow G_{\text{VMP}}\Big(&\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\leftrightarrow\boldsymbol{\theta}};\boldsymbol{A}^T\text{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\mathbf{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{A},\boldsymbol{A}^T\text{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\mathbf{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{y},\\ &\boldsymbol{y}^T\text{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\mathbf{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{y}\Big)\end{aligned}$$

$$\begin{aligned}\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\boldsymbol{\theta}}\longleftarrow&\left(1+\mu_{q(\lambda^2)}\right)\mu_{q(1/\sigma^2)}\begin{bmatrix}\boldsymbol{A}^T\text{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\mathbf{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{y}\\ -\frac{1}{2}\text{vec}\left(\boldsymbol{A}^T\text{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\mathbf{1}_n/\boldsymbol{a}_2\right)\right\}\boldsymbol{A}\right)\end{bmatrix}\\ &-\mu_{q\left(\lambda\sqrt{1+\lambda^2}\right)}\mu_{q(1/\sigma)}\begin{bmatrix}\boldsymbol{A}^T\text{diag}\left\{E_{q(\boldsymbol{a}_2)}\left(\mathbf{1}_n/\sqrt{\boldsymbol{a}_2}\right)\right\}E_{q(\boldsymbol{a}_1)}\left|\boldsymbol{a}_1\right|\\ \mathbf{0}\end{bmatrix}\end{aligned}$$

$$\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\sigma^2}\longleftarrow\begin{bmatrix}-n/2\\ \mu_{q\left(\lambda\sqrt{1+\lambda^2}\right)}\boldsymbol{\omega}_2^T E_{q(\boldsymbol{a}_1)}\left|\boldsymbol{a}_1\right|\\ \left(1+\mu_{q(\lambda^2)}\right)\boldsymbol{\omega}_5\end{bmatrix}$$

Continued overleaf...

Continuing...

$$\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\lambda} \longleftarrow \begin{bmatrix} n/2 \\ \mu_{q(1/\sigma^2)}\boldsymbol{\omega}_5 - \frac{1}{2}E_{q(\boldsymbol{a}_1)}\left\|\boldsymbol{a}_1\right\|^2 \\ \mu_{q(1/\sigma)}\boldsymbol{\omega}_2^T E_{q(\boldsymbol{a}_1)}\left|\boldsymbol{a}_1\right| \end{bmatrix}$$

$$\boldsymbol{\eta}_{p(\boldsymbol{a}_2|\nu)\to\nu} \longleftarrow \begin{bmatrix} n \\ -\boldsymbol{1}_n^T E_{q(\boldsymbol{a}_2)}\left\{\log\left(\boldsymbol{a}_2\right) + \boldsymbol{1}_n/\boldsymbol{a}_2\right\} \end{bmatrix}.$$

Parameter Outputs: $\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\boldsymbol{\theta}}$, $\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\sigma^2}$, $\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\theta},\sigma^2,\lambda,\boldsymbol{a}_1,\boldsymbol{a}_2)\to\lambda}$, $\boldsymbol{\eta}_{p(\boldsymbol{a}_2|\nu)\to\nu}$.

Expressions for $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are given in Algorithm 1.

Algorithm 1 updates include the following sufficient statistic expectations of a bivariate exponential family identified by the superscript *MW*

$$\begin{aligned}
(E\boldsymbol{T})_1^{\text{MW}}\left(\boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right) &\equiv E_{q(a_{1i},a_{2i})}\left(a_{1i}^2\right) = \frac{N_1}{D},\\
(E\boldsymbol{T})_2^{\text{MW}}\left(\boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right) &\equiv E_{q(a_{1i},a_{2i})}\left(\left|a_{1i}\right|/\sqrt{a_{2i}}\right) = \frac{N_2}{D},\\
(E\boldsymbol{T})_3^{\text{MW}}\left(\boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right) &\equiv E_{q(a_{1i},a_{2i})}\left(1/a_{2i}\right) = \frac{N_3}{D}\\
\text{and}\quad (E\boldsymbol{T})_4^{\text{MW}}\left(\boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right) &\equiv E_{q(a_{1i},a_{2i})}\left\{\log\left(a_{2i}\right)\right\} = \frac{N_4}{D}
\end{aligned}$$

where

$$\begin{aligned}
D &\equiv \int_0^\infty\int_{-\infty}^\infty a_{2i}^{\eta_4}\exp\left(\eta_1 a_{1i}^2 + \eta_2\frac{|a_{1i}|}{\sqrt{a_{2i}}} + \frac{\eta_3}{a_{2i}}\right)da_{1i}da_{2i},\\
N_1 &\equiv \int_0^\infty\int_{-\infty}^\infty a_{1i}^2 a_{2i}^{\eta_4}\exp\left(\eta_1 a_{1i}^2 + \eta_2\frac{|a_{1i}|}{\sqrt{a_{2i}}} + \frac{\eta_3}{a_{2i}}\right)da_{1i}da_{2i},\\
N_2 &\equiv \int_0^\infty\int_{-\infty}^\infty |a_{1i}|\, a_{2i}^{\eta_4-\frac{1}{2}}\exp\left(\eta_1 a_{1i}^2 + \eta_2\frac{|a_{1i}|}{\sqrt{a_{2i}}} + \frac{\eta_3}{a_{2i}}\right)da_{1i}da_{2i},\\
N_3 &\equiv \int_0^\infty\int_{-\infty}^\infty a_{2i}^{\eta_4-1}\exp\left(\eta_1 a_{1i}^2 + \eta_2\frac{|a_{1i}|}{\sqrt{a_{2i}}} + \frac{\eta_3}{a_{2i}}\right)da_{1i}da_{2i}\\
\text{and}\quad N_4 &\equiv \int_0^\infty\int_{-\infty}^\infty a_{2i}^{\eta_4}\log\left(a_{2i}\right)\exp\left(\eta_1 a_{1i}^2 + \eta_2\frac{|a_{1i}|}{\sqrt{a_{2i}}} + \frac{\eta_3}{a_{2i}}\right)da_{1i}da_{2i}.
\end{aligned}$$

With standard manipulations involving the standard normal distribution density and cumulative distribution functions the previous expressions simplify as follows:

$$\begin{aligned}
(E\boldsymbol{T})_1^{\text{MW}}\left(\boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right) &= \frac{\eta_2}{4I_1}\left\{\frac{I_2}{\sqrt{\pi}\left(-\eta_1\right)^{3/2}} + \frac{\eta_2 I_3}{\eta_1^2}\right\} - \frac{1}{2\eta_1},\\
(E\boldsymbol{T})_2^{\text{MW}}\left(\boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right) &= \frac{1}{2I_1}\left(\frac{I_2}{\sqrt{-\pi\eta_1}} - \frac{\eta_2 I_3}{\eta_1}\right),\\
(E\boldsymbol{T})_3^{\text{MW}}\left(\boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right) &= \frac{I_3}{I_1}\\
\text{and}\quad (E\boldsymbol{T})_4^{\text{MW}}\left(\boldsymbol{\eta}_{q(a_{1i},a_{2i})}\right) &= \frac{I_4}{I_1}.
\end{aligned}$$

where

$$I_i = \int_0^\infty \left\{\log\left(x\right)\right\}^{p_i} x^{q_i} e^{r_i/x}\Phi\left(\frac{s_i}{\sqrt{x}}\right)dx, \quad i = 1,\ldots,4$$

and

$$p_1 = 0, \quad q_1 = \eta_4 < -\frac{3}{2}, \quad r_1 = \eta_3 - \frac{\eta_2^2}{4\eta_1} < 0, \quad s_1 = \frac{\eta_2}{\sqrt{-2\eta_1}} \in \mathbb{R},$$
$$p_2 = 0, \quad q_2 = \eta_4 - \frac{1}{2} < -2, \quad r_2 = \eta_3 < 0, \quad s_2 = \infty,$$
$$p_3 = 0, \quad q_3 = \eta_4 - 1 < -\frac{5}{2}, \quad r_3 = \eta_3 - \frac{\eta_2^2}{4\eta_1} < 0, \quad s_3 = \frac{\eta_2}{\sqrt{-2\eta_1}} \in \mathbb{R},$$
$$p_4 = 1, \quad q_4 = \eta_4 < -\frac{3}{2}, \quad r_4 = \eta_3 - \frac{\eta_2^2}{4\eta_1} < 0 \quad \text{and} \quad s_4 = \frac{\eta_2}{\sqrt{-2\eta_1}} \in \mathbb{R}.$$

Integral I_2 has the following simple closed form:

$$I_2 = (-r_2)^{q_2+1}\,\Gamma(-q_2-1).$$

The integrals I_1, I_3 and I_4 are expressible in closed form in terms of Gaussian hypergeometric functions ${}_2F_1(a,b;c;z)$, making use of results 4.3.8 and 4.3.9 in Ng & Geller (1969). For $i = 1, 3, 4$,

$$I_i = \frac{1}{2}(-r_i)^{q_i+1}\,\Gamma(-q_i-1) + \frac{s}{\sqrt{2\pi}}(-r_i)^{q_i+\frac{1}{2}}\,\Gamma\left(-q_i-\frac{1}{2}\right){}_2F_1\left(\frac{1}{2}, -q_i-\frac{1}{2}; \frac{3}{2}; \frac{s_i^2}{2r_i}\right) \quad \text{if} \quad \left|\frac{s_i^2}{2r_i}\right| < 1 \quad \text{and}$$
$$I_i = \frac{1}{2}(-r_i)^{q_i+1}\,\Gamma(-q_i-1) + \frac{s}{\sqrt{2\pi}}(-r_i)^{q_i+\frac{1}{2}}\,\Gamma\left(-q_i-\frac{1}{2}\right){}_2F_1\left(-q_i-1, -q_i-\frac{1}{2}; -q_i; \frac{2r_i}{s_i^2}\right) \quad \text{if} \quad \left|\frac{s_i^2}{2r_i}\right| > 1.$$

Evaluation of the Gaussian hypergeometric function is supported by the function `hyperg_2F1` in the `R` package `gsl` (Hankin, 2006). However, evaluation of ${}_2F_1(a,b;c;z)$ for argument values close to 1 is cumbersome in practical implementations. Therefore numerical integration is necessary when the argument $z = \left|s_i^2/(2r_i)\right|$ of the Gaussian hypergeometric function is close to 1.

Efficient numerical integration can be performed via the simple trapezoidal rule (see Appendix B of Wand et al., 2011). Working on the log-scale is strongly recommended to avoid underflow and overflow.

Derivation of Algorithm 1 is then analogous to that of Algorithm 2 apart for replacement of the components involving auxiliary variables with the previous results on the joint distribution of auxiliary variables.

S.4 Proof of Theorem 1

First note that

$$\text{Corr}(|a_1|, 1/\sqrt{a_2}|x) = \frac{E\left(|a_1|/\sqrt{a_2}|x\right) - E(|a_1||x)\,E\left(1/\sqrt{a_2}|x\right)}{\sqrt{E\left(a_1^2|x\right) - E(|a_1||x)^2}\sqrt{E(1/a_2|x) - E\left(1/\sqrt{a_2}|x\right)^2}}. \tag{S.1}$$

We then study single components of the previous expression.

Term $E(|a_1||x)$

Note that

$$E(|a_1||x) = \int_{-\infty}^{\infty} |a_1|\,p(a_1|x)\,da_1 = \frac{1}{p(x)}\int_{-\infty}^{\infty} p(a_2)\int_0^{\infty} |a_1|\,p(x|a_1,a_2)\,p(a_1)\,da_1 da_2$$

and consider the inner integral. With standard manipulations involving the standard normal distribution density function and the cumulative distribution function we can write

$$\int_{-\infty}^{\infty} |a_1|\,p(x|a_1,a_2)\,p(a_1)\,da_1 = \frac{1}{\sqrt{\pi}\sigma_0\sqrt{1+\lambda_0^2}\sqrt{a_2}}\left\{\frac{1}{\sqrt{\pi}}K_1 + \frac{\sqrt{2}\lambda_0(x-\mu_0)}{\sigma_0\sqrt{a_2}}K_2\right\} \tag{S.2}$$

where

$$K_1 = \exp\left\{-\frac{\left(1+\lambda_0^2\right)\left(x-\mu_0\right)^2}{2\sigma_0^2 a_2}\right\} \quad \text{and} \quad K_2 = \exp\left\{-\frac{\left(x-\mu_0\right)^2}{2\sigma_0^2 a_2}\right\} \Phi\left\{\frac{\lambda_0\left(x-\mu_0\right)}{\sigma_0\sqrt{a_2}}\right\}.$$

Hence,

$$E\left(\left|a_1\right| | x\right) = \frac{1}{\sqrt{\pi}\sigma_0\sqrt{1+\lambda_0^2} p\left(x\right)}\left\{\frac{1}{\sqrt{\pi}}\int_0^\infty \frac{1}{\sqrt{a_2}} p\left(a_2\right) K_1 da_2 + \frac{\sqrt{2}\lambda_0\left(x-\mu_0\right)}{\sigma_0}\int_0^\infty \frac{1}{a_2} p\left(a_2\right) K_2 da_2\right\}. \quad \text{(S.3)}$$

Term $E\left(\left|a_1\right|/\sqrt{a_2}|x\right)$

Note that

$$E\left(\frac{\left|a_1\right|}{\sqrt{a_2}}\middle| x\right) = \int_0^\infty\int_{-\infty}^\infty \frac{\left|a_1\right|}{\sqrt{a_2}} p\left(a_1, a_2|x\right) da_1 da_2 = \frac{1}{p\left(x\right)}\int_0^\infty \frac{1}{\sqrt{a_2}} p\left(a_2\right)\int_{-\infty}^\infty \left|a_1\right| p\left(x|a_1, a_2\right) p\left(a_1\right) da_1 da_2.$$

Using result (S.2) we get

$$E\left(\frac{\left|a_1\right|}{\sqrt{a_2}}\middle| x\right) = \frac{1}{\sqrt{\pi}\sigma_0\sqrt{1+\lambda_0^2} p\left(x\right)}\left\{\frac{1}{\sqrt{\pi}}\int_0^\infty \frac{1}{a_2} p\left(a_2\right) K_1 da_2 + \frac{\sqrt{2}\lambda_0\left(x-\mu_0\right)}{\sigma_0}\int_0^\infty \frac{1}{a_2^{3/2}} p\left(a_2\right) K_2 da_2\right\}. \quad \text{(S.4)}$$

Term $E\left(a_1^2|x\right)$

Note that

$$E\left(a_1^2|x\right) = \int_{-\infty}^\infty a_1^2 p\left(a_1|x\right) da_1 = \frac{1}{p\left(x\right)}\int_{-\infty}^\infty p\left(a_2\right)\int_0^\infty a_1^2 p\left(x|a_1, a_2\right) p\left(a_1\right) da_1 da_2.$$

With standard manipulations involving the standard normal distribution density and cumulative distribution functions the inner integral in the previous expression becomes

$$\int_{-\infty}^\infty a_1^2 p\left(x|a_1, a_2\right) p\left(a_1\right) da_1 = \frac{1}{\sqrt{\pi}\sigma_0^2\left(1+\lambda_0^2\right)}\left[\frac{\lambda_0\left(x-\mu_0\right)}{\sqrt{\pi} a_2} K_1 + \frac{\sqrt{2}\left\{\lambda_0^2\left(x-\mu_0\right)^2 + \sigma_0^2 a_2\right\}}{\sigma_0 a_2^{3/2}} K_2\right].$$

Finally,

$$\begin{aligned} E\left(a_1^2|x\right) =& \frac{1}{\sqrt{\pi}\sigma_0^2\left(1+\lambda_0^2\right) p\left(x\right)}\left[\frac{\lambda_0\left(x-\mu_0\right)}{\sqrt{\pi}}\int_0^\infty \frac{1}{a_2} p\left(a_2\right) K_1 da_2\right. \\ &\left. + \frac{\sqrt{2}}{\sigma_0}\int_0^\infty\left\{\frac{\lambda_0^2\left(x-\mu_0\right)^2}{a_2^{3/2}} + \frac{\sigma_0^2}{\sqrt{a_2}}\right\} p\left(a_2\right) K_2 da_2\right]. \end{aligned} \quad \text{(S.5)}$$

Term $E\left(1/\sqrt{a_2}|x\right)$

Note that

$$E\left(\frac{1}{\sqrt{a_2}}\middle| x\right) = \int_0^\infty \frac{1}{\sqrt{a_2}} p\left(a_2|x\right) da_2 = \frac{1}{p\left(x\right)}\int_{-\infty}^\infty \frac{1}{\sqrt{a_2}} p\left(a_2\right)\int_0^\infty p\left(x|a_1, a_2\right) p\left(a_1\right) da_1 da_2.$$

With standard manipulations involving the standard normal distribution density and cumulative distribution functions the inner integral in the previous expression becomes

$$\int_0^\infty p\left(x|a_1, a_2\right) p\left(a_1\right) da_1 = \frac{\sqrt{2}}{\sqrt{\pi}\sigma_0\sqrt{a_2}} K_2.$$

It follows that

$$E\left(\frac{1}{\sqrt{a_2}}\middle| x\right) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma_0 p\left(x\right)}\int_0^\infty \frac{1}{a_2} p\left(a_2\right) K_2 da_2. \quad \text{(S.6)}$$

Term $E\left(1/a_2|x\right)$

Similarly to term $E\left(1/\sqrt{a_2}|x\right)$ we get

$$E\left(\frac{1}{a_2}\bigg|x\right)=\frac{\sqrt{2}}{\sqrt{\pi}\sigma_0 p\left(x\right)}\int_0^{\infty}\frac{1}{a_2^{3/2}}p\left(a_2\right)K_2 da_2. \tag{S.7}$$

Consider expression (S.1) again. Substituting expressions (S.4)–(S.7) in it and dividing numerator and denominator by $\left\{\sqrt{2}\lambda_0\left(x-\mu_0\right)\right\}\left\{\sqrt{\pi}\sigma_0 p\left(x\right)\right\}^{-1}\int_0^{\infty}\frac{1}{a_2^{3/2}}p\left(a_2\right)K_2 da_2$ we get

$$\begin{aligned}\operatorname{Corr}\left(\left|a_1\right|,1/\sqrt{a_2}|x=x_0\right)=&\left[1+\frac{\sigma_0}{\sqrt{2\pi}\lambda_0\left(x-\mu_0\right)}\frac{C_2}{G_3}-\frac{1}{p\left(x\right)}\left\{\frac{1}{\pi\lambda_0\left(x-\mu_0\right)}\frac{C_1G_2}{G_3}+\frac{\sqrt{2}}{\sqrt{\pi}\sigma_0}\frac{G_2^2}{G_3}\right\}\right]\\&\times\left\{1-\frac{\sqrt{2}}{\sqrt{\pi}\sigma_0 p\left(x\right)}\frac{G_2^2}{G_3}\right\}^{-1/2}\times\left[1+\frac{\sigma_0}{\sqrt{2\pi}\lambda_0\left(x-\mu_0\right)}\frac{C_2}{G_3}+\frac{\sigma_0^2}{\lambda_0^2\left(x-\mu_0\right)^2}\frac{G_1}{G_3}\right.\\&\left.-\frac{1}{p\left(x\right)}\left\{\frac{\sigma_0}{\sqrt{2}\pi^{3/2}\lambda_0^2\left(x-\mu_0\right)^2}\frac{C_1^2}{G_3}+\frac{2}{\pi\lambda_0\left(x-\mu_0\right)}\frac{C_1G_2}{G_3}+\frac{\sqrt{2}}{\sqrt{\pi}\sigma_0}\frac{G_2^2}{G_3}\right\}\right]^{-1/2}\end{aligned} \tag{S.8}$$

with

$$C_1=\int_0^{\infty}\frac{1}{\sqrt{a_2}}p\left(a_2\right)K_1 da_2=\left(\frac{\nu_0}{2}\right)^{-\frac{1}{2}}\frac{\Gamma\left\{\left(\nu_0+1\right)/2\right\}}{\Gamma\left(\nu_0/2\right)}\left\{1+\frac{\left(1+\lambda_0^2\right)\left(x-\mu_0\right)^2}{\nu_0\sigma_0^2}\right\}^{-\frac{\nu_0+1}{2}}, \tag{S.9}$$

$$C_2=\int_0^{\infty}\frac{1}{a_2}p\left(a_2\right)K_1 da_2=\left\{1+\frac{\left(1+\lambda_0^2\right)\left(x-\mu_0\right)^2}{\nu_0\sigma_0^2}\right\}^{-\frac{\nu_0}{2}-1}, \tag{S.10}$$

$$G_1=\int_0^{\infty}\frac{1}{\sqrt{a_2}}p\left(a_2\right)K_2 da_2<\left(\frac{\nu_0}{2}\right)^{-\frac{1}{2}}\frac{\Gamma\left\{\left(\nu_0+1\right)/2\right\}}{\Gamma\left(\nu_0/2\right)}\left\{1+\frac{\left(x-\mu_0\right)^2}{\nu_0\sigma_0^2}\right\}^{-\frac{\nu_0+1}{2}}, \tag{S.11}$$

$$G_2=\int_0^{\infty}\frac{1}{a_2}p\left(a_2\right)K_2 da_2<\left\{\left(\frac{\nu_0}{2}\right)^{\frac{\nu_0}{2}}\Big/\Gamma\left(\frac{\nu_0}{2}\right)\right\}\left\{1+\frac{\left(x-\mu_0\right)^2}{\nu_0\sigma_0^2}\right\}^{-\frac{\nu_0}{2}-1}, \tag{S.12}$$

and

$$\begin{aligned}G_3=\int_0^{\infty}\frac{1}{a_2^{3/2}}p\left(a_2\right)K_2 da_2>&\left(\frac{\nu_0}{2}\right)^{-\frac{3}{2}}\frac{\Gamma\left\{\left(\nu_0+3\right)/2\right\}}{\Gamma\left(\nu_0/2\right)}\left\{1+\frac{\left(x-\mu_0\right)^2}{\nu_0\sigma_0^2}\right\}^{-\frac{\nu_0+3}{2}}\\&-\frac{\sigma_0}{\sqrt{2\pi}\lambda_0\left(x-\mu_0\right)}\left\{1+\frac{\left(1+\lambda_0^2\right)\left(x-\mu_0\right)^2}{\nu_0\sigma_0^2}\right\}^{-\frac{\nu_0}{2}-1}.\end{aligned} \tag{S.13}$$

The expressions and inequalities in (S.9)–(S.13) are obtained with standard algebra and integration involving the gamma function. In particular, the upper bound for G_1 and G_2 in (S.11) and (S.12) are derived using the fact that $\Phi\left(t\right)>1-\left(2\pi\right)^{-1/2}t^{-1}e^{-t/2}$, $\forall t\in\mathbb{R}$. The lower bound for G_3 in (S.13) is derived making use of $\Phi\left(t\right)<1$, $\forall t\in\mathbb{R}$.

We can then study the behavior of single components of (S.8) when $\left|\lambda_0\right|\to\infty$ using simplifications

(S.9)–(S.13) and setting $x = x_0 \in \mathbb{R}$. We then have, for instance,

$$\lim_{|\lambda_0|\to\infty} \frac{\sigma_0}{\sqrt{2\pi}\lambda_0 (x_0 - \mu_0)} \frac{C_2}{G_3} \leq \lim_{|\lambda_0|\to\infty} \frac{\sigma_0}{\sqrt{2\pi}\lambda_0 (x_0 - \mu_0)} \left\{1 + \frac{\left(1+\lambda_0^2\right)(x_0-\mu_0)^2}{\nu_0\sigma_0^2}\right\}^{-\frac{\nu_0}{2}-1}$$
$$\times \left[\left(\frac{\nu_0}{2}\right)^{-\frac{3}{2}} \frac{\Gamma\left\{(\nu_0+3)/2\right\}}{\Gamma(\nu_0/2)} \left\{1 + \frac{(x_0-\mu_0)^2}{\nu_0\sigma_0^2}\right\}^{-\frac{\nu_0+3}{2}} - \frac{\sigma_0}{\sqrt{2\pi}\lambda_0 (x_0 - \mu_0)} \left\{1 + \frac{\left(1+\lambda_0^2\right)(x_0-\mu_0)^2}{\nu_0\sigma_0^2}\right\}^{-\frac{\nu_0}{2}-1}\right]^{-1} = 0.$$

Similar arguments lead to the final expression

$$\lim_{|\lambda_0|\to\infty} \mathrm{Corr}\left(|a_1|, 1/\sqrt{a_2}|x = x_0\right) = \lim_{|\lambda_0|\to\infty} \frac{1 - \frac{\sqrt{2}}{\sqrt{\pi}\sigma_0 p(x_0)} \frac{G_2^2}{G_3}}{\sqrt{1 - \frac{\sqrt{2}}{\sqrt{\pi}\sigma_0 p(x_0)} \frac{G_2^2}{G_3}} \sqrt{1 - \frac{\sqrt{2}}{\sqrt{\pi}\sigma_0 p(x_0)} \frac{G_2^2}{G_3}}} = 1.$$

Additional References

Azzalini, A. (2017). The R package 'sn': the skew-normal and related distributions such as the skew-t. R package version 1.5.1. http://azzalini.stat.unipd.it/SN.

Hankin, R. K. S. (2006). Special functions in R: introducing the gsl package. *R News*, **6**, 24–26.

Ng, E. W. and Geller, M. (1969). A table of integrals of the error functions. *Journal of Research of the National Bureau of Standards B*, **73**, 1–20.