Bayesian predictive densities as an interpretation of a class of Skew–Student $t$ distributions with application to medical data

Abdolnasser Sadeghkhani

Department of Mathematics and Statistics, Queen’s University, Kingston, ON, Canada
E-mail: a.sadeghkhani@queensu.ca

Abstract

This paper describes a new Bayesian interpretation of a class of skew–Student $t$ distributions. We consider a hierarchical normal model with unknown covariance matrix and show that by imposing different restrictions on the parameter space, corresponding Bayes predictive density estimators under Kullback-Leibler loss function embrace some well-known skew–Student $t$ distributions. We show that obtained estimators perform better in terms of frequentist risk function over regular Bayes predictive density estimators. We apply our proposed methods to estimate future densities of medical data: the leg-length discrepancy and effect of exercise on the age at which a child starts to walk.

AMS 2010 subject classifications: 60E05, 62F15, 62F30

Keywords and phrases: Bayesian predictive density, Constrained hierarchical model, Kullback–Leibler loss function, Skew–Student $t$ distribution.

1 Introduction

Normal density has been used to analyze data in many applications for decades. However, Roberts (1998) as an example of a weighted normal model introduced the skew–normal density for asymmetric data set, but he did not use the term “skew–normal” then. Azzalini (1985) formalized the skew normal distribution as a generalization of a normal distribution which can be used to model asymmetric data. His work inspired many statisticians to study different versions of the skew–normal distributions rapidly. However, for many applications, the skew-normal model fails to provide a good fit for skewed data due to the excessive kurtosis of the data such as long–tailed data which occurs in several cases such as in finance and insurance business. Hansen (1994) proposed a skew extension to the Student $t$ distribution for modeling financial returns. Since then, several other papers have studied different versions of skew Student $t$ distributions for financial and other applications, see e.g. Bauwens and Laurent (2005); Branco and Dey (2001); Fernandez and Steel (1998) and Patton (2004).

Following Azzalini and Valle (1996), a random variable $T$ is said to have a skew–normal distribution $\mathcal{SN}_p(\lambda, \xi, \tau)$, if the probability density function (pdf) is

$$
\frac{2}{\tau^p} \phi_p\left(\frac{t - \xi}{\tau}\right) \Phi\left(\lambda^T \frac{t - \xi}{\tau}\right), \quad t \in \mathbb{R}^p,
$$

where $\phi_p(\cdot)$ is the pdf of a $p$ variate normal distribution, $\Phi(\cdot)$ is the cumulative density function (cdf) of a univariate standard normal density and $\lambda \in \mathbb{R}^p$ is the shape parameter which determines the skewness and $\lambda^T$ is its transpose. Another version of density [1.1] introduced by Gupta et al. (2004) which denoted
by \( \mathbb{SN}_p(\alpha_0, \alpha_1, \xi, \tau) \) is given by

\[
\frac{1}{\tau^p} \phi_p\left( \frac{t - \xi}{\tau} \right) \frac{\Phi_p(\alpha_0 + \alpha_1 \frac{t - \xi}{\tau}; 0)}{\Phi_n\left( \frac{\alpha_0}{\sqrt{1+\alpha_1^2}}, \frac{\alpha_0}{\sqrt{1+\alpha_1^2}}; \frac{\alpha_1^2}{1+\alpha_1^2} \right)}, \quad t \in \mathbb{R}^p, \tag{1.2}
\]

where \( \Phi_p(\cdot; 0) \) and \( \Phi_n(\cdot; \rho) \) are cdf’s of a \( N_p(0, I_p) \) and \( N_n(0, \Lambda) \) distributions respectively and \( \Lambda = (1 - \rho) I_n + \rho I_n F_n^T \) is the covariance matrix.

Furthermore, a Student t distribution with degrees of freedom \( \nu > 0 \), location parameter \( \xi \) and scale parameter \( \tau \), denoted by \( T_p(\nu, \xi, \tau) \) has a density on \( \mathbb{R}^p \), is given by

\[
\frac{1}{\tau^p} \frac{\Gamma\left( \frac{\nu+p}{2} \right)}{\Gamma\left( \frac{\nu}{2} \right) (\pi \nu)^{\frac{p}{2}}} \left( 1 + \frac{\| t - \xi \|^2}{\nu \tau^2} \right)^{-\frac{\nu+p}{2}}, \quad t \in \mathbb{R}^p. \tag{1.3}
\]

**Definition 1.1 (Skew–Student t distribution).** A random variable follows a skew–Student t distribution, denoted by \( \mathcal{ST}_p(\nu, \alpha_0, \alpha_1, \xi, \tau) \) with location \( \xi \in \mathbb{R}^p \), scale \( \tau > 0 \), \( \nu \in \mathbb{R}_+ \), \( \alpha_1 \in \mathbb{R}^p \) and \( \alpha_0 \in \mathbb{R} \) if \( Z = (T - \xi)/\tau \) has pdf

\[
T_p(z; \xi, \tau, \nu) \frac{F_p\left( \nu + p, (\alpha_0 + \alpha_1^T z) \sqrt{\frac{\nu + p}{\nu + z^T z}} \right)}{F_p\left( \nu, \frac{\alpha_0}{\sqrt{1+\alpha_1^2}} \right)} - \frac{F_p\left( \nu + p, (\alpha_2 + \alpha_1^T z) \sqrt{\frac{\nu + p}{\nu + z^T z}} \right)}{F_p\left( \nu, \frac{\alpha_2}{\sqrt{1+\alpha_1^2}} \right)}, \tag{1.4}
\]

where \( T_p \) denotes pdf a Student t distribution in (1.3) and \( F_p(\nu, \cdot) \) is cdf of a standard \( p \)-variate Student t distribution. For \( p = 1 \) and \( \alpha_0 = 0 \), density (1.4) reduces to \( 2 t(z, \nu) F_1\left( \nu + 1, \left( \frac{\nu + 1}{\nu + z^2} \right)^{\frac{1}{2}} \alpha_1 z \right) \), and setting \( \alpha_0 = \alpha_1 = 0 \) reduces (1.4) to a standard Student t distribution.

We may conclude another extension as follows.

**Definition 1.2.** A random variable \( T \) follows a skew–Student t distribution, denoted by \( \mathcal{ST}_p(\nu, \alpha_0, \alpha_1, \alpha_2, \xi, \tau) \) with location \( \xi \in \mathbb{R}^p \), scale \( \tau > 0 \), \( \nu \in \mathbb{R}_+ \), \( \alpha_1 \in \mathbb{R}^p \), \( \alpha_0, \alpha_2 \in \mathbb{R} \), if \( Z = (T - \xi)/\tau \in \mathbb{R}^p \), has pdf

\[
T_p(z; \xi, \tau, \nu) \frac{F_p\left( \nu + p, (\alpha_0 + \alpha_1^T z) \sqrt{\frac{\nu + p}{\nu + z^T z}} \right)}{F_p\left( \nu, \frac{\alpha_0}{\sqrt{1+\alpha_1^2}} \right)} - \frac{F_p\left( \nu + p, (\alpha_2 + \alpha_1^T z) \sqrt{\frac{\nu + p}{\nu + z^T z}} \right)}{F_p\left( \nu, \frac{\alpha_2}{\sqrt{1+\alpha_1^2}} \right)}, \tag{1.5}
\]

where \( T_p \) and \( F_p \) are as in Definition 1.1.

The distribution (1.1), for \( p = 1 \), was formerly obtained by O’Hagan and Leonard (1976) via defining the following model.

\[
X \mid \theta \sim \mathcal{N}(\theta, \sigma^2), \quad \theta \mid \mu, \tau^2 \sim \mathcal{N}(\mu, \tau^2), \quad \theta \geq \mu,
\]

which the marginal distribution \( X \) follows \( \mathbb{SN}(\tau/\sigma, \mu, \sqrt{\sigma^2 + \tau^2}) \).

Liseo and Loperfido (2003) introduced another skew–normal density in a multivariate case by suppose that the covariance matrices \( \Sigma \) and \( \Omega \) are known and we have

\[
X \mid \theta \sim \mathcal{N}_p(\theta, \Sigma), \quad \theta \mid \mu \sim \mathcal{N}_p(\mu, \Omega), \quad C \theta + d \leq 0,
\]

where \( C \) is a \( k \times p \) full rank matrix and \( d \in \mathbb{R}^k \). They showed that the marginal density of \( X \) is given by

\[
\frac{1}{\Phi_k(C\mu + d, C\Omega C^T; 0)} \phi_p(x, \mu + \Sigma + \Omega) \Phi_k(C\Delta(\Sigma^{-1} x + \Omega^{-1} \mu) + d, C\Delta C^T; 0),
\]
\[ \Delta^{-1} = (\Sigma^{-1} + \Omega^{-1}), \quad \Phi_p(a, C; 0) \] is cdf of a \( p \)-variate normal density with mean vector \( a \) and covariance matrix \( C \).

Here, we use a hierarchical normal model with unknown covariance matrix in the presence of prior information on parameter spaces to obtain a class of skew–Student \( t \) densities. In contrast to the Liseo and Loperdido’s method which is based on marginal densities, we use another approach founded on posterior predictive density estimators to construct skew–Student \( t \) distributions. In fact, previous studies (see, e.g. Corcuera and Giummolè 1999), indicates that under Kullback–Leibler (KL) loss, as defined in below

\[
L_{KL}(\theta, \hat{q}(\cdot; x_1, x_2)) = \int_{\mathbb{R}^p} q_0(y_1) \log \frac{q_\theta(y_1)}{\hat{q}(y_1; x_1, x_2)} 
\]

the Bayes predictive density estimator can be obtained by

\[
\hat{q}_\pi(y_1; x_1, x_2) = \int_{\mathbb{R}^p} \phi\left(\frac{y_1 - \theta_1}{\sigma_y}\right) \pi(\theta | x_1, x_2) d\theta, 
\]

in other words, under the KL loss, the Bayes density coincides with the posterior predictive densities.

The rest of the paper is organized as follows: In Section 2 we introduce useful lemmas and sketch the problem set-up. Section 3 involves Bayes predictive densities and their interesting representations as skew–Student \( t \) distributions, as well as the KL risk performances. Section 4 contains two different medical problems and we try to illustrate our methodology in finding the predictive density estimators, while some concluding remarks will appear in Section 5.

### 2 Problem set-up

Consider the following (canonical) normal model

\[ X_i \sim N_p(\theta_i, \sigma^2 I_p), \quad i = 1, 2, \quad Y_1 \sim N_p(\theta_1, \sigma^2 I_p), \quad S^2 \sim \sigma^2 \chi^2_k, \] are independent,

with \( k \geq 2, \: \theta_1 \in \mathbb{R}^p, \: \theta_2 \in \mathbb{R}^p, \: \sigma^2 \in \mathbb{R}_+, \) and \( \theta_1 - \theta_2 \in A \subseteq \mathbb{R}^p \). We consider that \( \theta = (\theta_1, \theta_2) \) and \( \sigma^2 \) are unknown, and the objective is to obtain a predictive density estimate for \( Y_1 \).

It can be verified that the joint density of \( (X, S^2) \) in model (2.7), supported on \( \mathbb{R}^{2p} \times \mathbb{R}_+ \), is given by

\[
p_{\theta, \sigma^2}(x, s^2) = \frac{(s^2)^{k/2-1}}{(2\pi \sigma^2)^p} \exp \left\{ -\frac{1}{2\sigma^2} \left( \|x_1 - \theta_1\|^2 + \|x_2 - \theta_2\|^2 + s^2 \right) \right\}, \quad (2.8)
\]

We assume the prior density

\[
\pi_A(\theta, \sigma^2) = \frac{1}{\sigma^2} \mathbb{I}_A(\theta_1 - \theta_2), \quad (2.9)
\]

on the restricted parameter space \( \theta_1 - \theta_2 \in A \), where \( A \) is a subset of \( \mathbb{R}^p \)

**Lemma 2.1.** For model (2.7), and prior in (2.9), we have

(a) the posterior density \( \pi(\theta | x, s^2) \) is proportional to

\[
(1 + \frac{\|x_2 - \theta_2\|^2}{s^2 + \|x_1 - \theta_1\|^2})^{-(p+k/2)} \left(1 + \frac{\|x_1 - \theta_1\|^2}{s^2}\right)^{-(p+k/2)}, \quad (2.10)
\]
(b) the marginal posterior density is given by

\[ \pi(\theta_1 | x, s^2) \propto T_p(\nu = k, \xi = x_1, \tau = \frac{s}{\sqrt{k}}) \mathbb{P}(V \in A), \]

where

\[ V \sim T_p \left( \nu = k + p, \xi = \theta_1 - x_2, \tau = \sqrt{\frac{s^2 + \|x_1 - \theta_1\|^2}{p + k}} \right). \tag{2.11} \]

**Proof.** See Appendix A.1. \qed

For example, for \( p = 1 \) and \( A = [0, \infty) \), the probability as defined in Lemma 2.1 is equivalent to

\[ F_1 \left( k + 1, \frac{\theta_1 - x_2}{\sqrt{\frac{s^2 + \|x_1 - \theta_1\|^2}{p + k}}} \right), \]

where \( F_1(\nu, \cdot) \) is cdf of a standard Student \( t \) distribution with degrees of freedom \( \nu > 1 \).

Next lemma was introduced by Aitchison (1975), gives posterior predictive density, with respect to \( \pi_0(\theta_1, \sigma^2) = \frac{1}{\sigma^2} \) and based on \((X_1, S^2)\). For the first step, we need the definition of the scale inverse chi squared density. A random variable \( X \) is said to have a scale inverse chi squared density whenever for all degrees of freedom \( \nu > 0 \) and scale parameter \( \tau > 0 \), the pdf is

\[ \frac{(\tau^2 \nu/2)^{\nu/2} \exp\left[-\frac{\nu \tau^2}{2x}\right]}{\Gamma(\nu/2) \ 1 + \nu/2}, \ x \in \mathbb{R}, \tag{2.12} \]

and we denoted by \( SIv - \chi^2(\nu, \tau) \).

**Lemma 2.2.** For model \( (2.7) \), the posterior predictive density of \( Y_1 \) (Bayes predictive density estimator) associated with the non-informative prior density \( \pi_0(\theta_1, \sigma^2) = \frac{1}{\sigma^2} \)

\[ T_p \left( \nu = k, \xi = x_1, \tau = \sqrt{\frac{2s_1^2}{k}} \right). \]

**Proof.** For the non-informative prior \( \pi_0(\theta_1, \sigma^2) = \frac{1}{\sigma^2} \) we have

\[ \pi(\theta_1, \sigma^2 | x_1, s^2) \propto (\sigma^2)^{-\frac{\nu+k}{2} - 1} \exp\left\{ -\frac{t'}{2\sigma^2} \right\}, \]

where \( t' = \|x_1 - \theta_1\|^2 + s_1^2 \). This gives \( \pi(\sigma^2 | x_1) \propto (\sigma^2)^{-(k/2+1)} \exp\left\{ -\frac{s_1^2}{2\sigma^2} \right\} \). It is recognized as a kernel of
scale inverse chi–square $SI_{\nu} = \chi^2(k, \sqrt{\frac{2s^2}{k}})$. (equation 2.12). Therefore we have

$$q(y_1; x, s^2)_{\pi_0} = \int_0^\infty q(y_1 \mid x, s^2) \pi(\sigma^2 | x, s^2) d\sigma^2$$

$$\propto \int_0^\infty (\sigma^2)^{-p/2} \exp\{ -\| y_1 - x_1 \|^2 / 2\sigma^2 \} (\sigma^2)^{-k/2-1} \exp\{ -s^2 / 2\sigma^2 \} d\sigma^2$$

$$\propto \int_0^\infty (\sigma^2)^{-p+k/2-1} \exp\{ -\frac{1}{2\sigma^2} (\| y_1 - X_1 \|^2 + \frac{s^2}{2}) \} d\sigma^2$$

$$\propto t''^{p+k/2} \int_0^\infty z^{p+k/2-1} \exp\{ -z \} dz, \text{ with } t'' = \| y_1 - x_1 \|^2 + \frac{s^2}{2}$$

$$\propto \left( 1 + \frac{2k\| y_1 - x_1 \|^2}{k s^2} \right)^\left( -\frac{p+k}{2} \right).$$

This is the kernel of $T_p(\nu = k, \xi = x_1, \tau = \sqrt{\frac{2s^2}{k}})$ and hence the proof.

### 3 Predictive densities and their representations

Given the prior $\pi_A(\theta, \sigma^2)$ in (2.9), the posterior predictive density is given by

$$q_{\pi,A}(y_1 \mid x, s^2) = \int_{\mathbb{R}_+^p} \int_{\mathbb{R}_+^p} q_{\theta_1, \sigma^2}(y_1) \pi_A(\theta, \sigma^2 \mid x, s^2) d\theta_1 d\sigma^2,$$

where

$$\pi_A(\theta_1, \sigma^2 \mid x, s^2) \propto e^{-\frac{s^2}{2\sigma^2} (\sigma^2)^{-\left(\frac{p+k}{2}+1\right)} \phi\left(\frac{\theta_1 - x_1}{\sigma}\right)} \int_{\{\theta_2: \theta_1 - \theta_2 \in A\}} (\sigma^2)^{-\frac{p}{2}} \phi\left(\frac{\theta_2 - x_2}{\sigma}\right) d\theta_2.$$

$$\propto e^{-\frac{s^2}{2\sigma^2} (\sigma^2)^{-\left(\frac{p+k}{2}+1\right)}} \mathbb{P}(W \in A), \quad (3.13)$$

and $\ln \sim N_p(\theta_1 - x_2, \sigma^2 I_p)$.

**Lemma 3.3.** For model (2.7), uniform prior (2.9) on $A = \mathbb{R}_+^p$, by setting $\eta = \frac{1}{\sigma^2}$, we have

$$\theta_1 \mid \eta, x, s^2 \sim SN_{\nu} \left( \alpha_0 = (x_1 - x_2)\sqrt{\frac{\eta}{2}}, \alpha_1 = 1, \xi = 1, \tau = \frac{1}{\sqrt{\eta}} \right), \quad (3.14)$$

and,

$$\eta \mid x, s^2 \sim \pi_{U,\nu}(\eta \mid x, s^2) = e^{-\frac{\eta}{2} - 1} e^{-\frac{\eta}{2}} \frac{p^2}{\Gamma\left(\frac{p}{2}\right)} \Phi_p\left( \left( x_1 - x_2 \right) \sqrt{\frac{\eta}{2}}, 0 \right) \frac{F_p\left( k, \frac{\eta - x_2^2}{\sqrt{2s^2/k}} \right)}{\Gamma\left(\frac{p}{2}\right)(\frac{2\pi}{s^2})^{k/2}}, \quad (3.15)$$

where $SN_{\nu}$ is defined in (1.3) and $F_p(\nu, \cdot)$ is cdf of a standard $p$–variate Student $t$ distribution with degrees of freedom $\nu$. 

5
Proof. The probability in (3.13) can be replaced by $\Phi_p\left(\frac{\eta - x_2}{\alpha}\right)$. Thus

$$
\pi_{U_A}(\theta_1, \sigma^2 \mid x, s^2) \propto \frac{e^{-\frac{x_1 - x_2}{\sigma^2}}}{(\sigma^2)^{k/2+1}} \frac{1}{\sqrt{2\pi}} \phi_p\left(\frac{\theta_1 - x_1}{\sigma}\right) \Phi_p\left(\frac{\theta_1 - x_1}{\sigma}; 0\right)
$$

$$
\propto \Phi_p\left(\frac{x_1 - x_2}{\sqrt{2}\tau}; 0\right) \frac{e^{-\frac{x_1 - x_2}{2\tau}}}{(\sigma^2)^{k/2+1}} \frac{1}{\sqrt{2\pi}} \phi_p\left(\frac{\theta_1 - x_1}{\sigma}\right) \frac{1}{\Phi_p\left(\frac{\alpha_0 + \alpha_1 \eta - \xi}{\alpha}; 0\right)},
$$

Changing the variable $\eta = 1/\sigma^2$ proves (3.14).

In addition, we can write

$$
\pi_{U_A}(\eta \mid x, s^2) \propto \frac{\eta^{k/2-1} e^{-s^2\eta/2}}{\Gamma\left(\frac{k}{2}\right)\left(\frac{2}{s}\right)^{k/2}} \Phi_p\left(\frac{x_1 - x_2}{\sqrt{\eta}}; 0\right).
$$

(3.16)

Now, using the identity provided by Azzalini and Capitanio (2003) as follows

$$
E[\Phi_p(c \sqrt{\eta}; 0)] = F_p\left(2a, c \sqrt{\frac{a}{b}}\right), \text{ for } \eta \sim \text{Gamma}(a, b), a > 0, b > 0, c > 0,
$$

and choosing $a = k/2, b = s^2/2$ in it, (3.16) can be written as

$$
\pi_{U_A}(\eta \mid x, s^2) = \frac{\eta^{k/2-1} e^{-s^2\eta/2}}{\Gamma\left(\frac{k}{2}\right)\left(\frac{2}{s}\right)^{k/2}} \frac{\Phi_p\left(\frac{x_1 - x_2}{\sqrt{\eta}}; 0\right)}{F_p\left(k, \sqrt{\frac{s_2}{s^2}}\right)}.
$$

This completes the proof of (3.15). \hfill \Box

**Theorem 3.1.** For model (2.7), the Bayes predictive density estimator based on additional prior $\hat{q}_{\pi_A}(y_1; x, s^2)$ associated with a uniform prior (2.9) on $A = \mathbb{R}^P_+$, is given by

$$
\text{ST}_p\left(\nu = k, \alpha_0 = \sqrt{\frac{2}{3}} \frac{x_1 - x_2}{\sqrt{s^2/k}}; \alpha_1 = 1/\sqrt{3}, \xi = x_1, \tau = \sqrt{\frac{2s^2}{k}}\right),
$$

where $\text{ST}_p$ is skew–Student t distribution, as defined in Definition 1.1. Equivalently, we can write

$$
T_p\left(\nu = k, \xi = x_1, \tau = \sqrt{\frac{2s^2}{k}}\right) = \frac{F_p\left(k + p, \left(\sqrt{\frac{2}{3}}\frac{x_1 - x_2}{\sqrt{s^2/k}}\right) + \frac{y_1 - x_1}{\sqrt{3}}\right) \sqrt{\frac{k + 1}{2s^2 + (y_1 - x_1)^2}}}{F_p\left(k, \sqrt{\frac{x_1 - x_2}{2s^2/k}}\right)}.
$$

(3.18)

**Proof.** See Appendix A.2. \hfill \Box

**Lemma 3.4.** For model (2.7) and uniform prior (2.9) on $A = [-m, m]^P$, the marginal posterior distribution $\pi(\theta_1 \mid \eta, x, s^2)$ is given by

$$
\text{SN}_p\left(\alpha_0 = (x_1 - x_2 + m) \sqrt{\frac{\eta}{2}}; \alpha_1 = 1, \xi = 1, \tau = \frac{1}{\sqrt{\eta}}\right).
$$

and also,

$$
\pi(\eta \mid x, s^2) = \frac{\eta^{k/2-1} e^{-s^2\eta/2}}{\Gamma\left(\frac{k}{2}\right)\left(\frac{2}{s}\right)^{k/2}} \frac{\Phi_p\left(\frac{(x_1 - x_2 + m) \sqrt{\eta/2} - 0}{\sqrt{x_1 - x_2 + m}}\right) - \Phi_p\left(\frac{(x_1 - x_2 - m) \sqrt{\eta/2} - 0}{\sqrt{x_1 - x_2 - m}}\right)}{F_p\left(k, \sqrt{\frac{x_1 - x_2 + m}{2s^2/k}}\right) - F_p\left(k, \sqrt{\frac{x_1 - x_2 - m}{2s^2/k}}\right)}.
$$

(3.19)
Proof. The proof may be easily derived from a similar analysis to Lemma (3.3) with the probability
\[ P(W \in A) = \Phi_p\left(\frac{\theta_1 - x_2 + m}{\sigma}; 0\right) - \Phi_p\left(\frac{\theta_1 - x_2 - m}{\sigma}; 0\right) \] in (3.13).

Theorem 3.2. For model (2.7), the Bayes predictive density estimator based on additional prior informa-
tion \( \hat{q}_{\pi,A}(y_1; x, s^2), A = [-m, m]^p \) with \( m > 0 \), and a uniform prior (2.9), is given by
\[
ST_p\left(\alpha_0 = \sqrt{\frac{2}{3}} \sqrt{\frac{x_1 - x_2 + m}{2s/k}}, \alpha_1 = \sqrt{\frac{2}{3}} \sqrt{\frac{x_1 - x_2 - m}{2s/k}}, \xi = x_1, \tau = \sqrt{\frac{2s^2}{k}}\right), \tag{3.19}
\]
In other words,
\[
T_p\left(\nu = k, \xi = x_1, \tau = \sqrt{\frac{2s^2}{k}}\right) \frac{F_p(k + 1, L_1(x, s^2)) - F_p(1, \sqrt{\frac{x_1 - x_2 + m}{2s/k}})}{F_p(k + 1, L_2(x, s^2)) - F_p(1, \sqrt{\frac{x_1 - x_2 - m}{2s/k}})}, \tag{3.20}
\]
where \( L_1(x, s^2) = \sqrt{\frac{2}{3}}(x_1 - x_2 + m) + \frac{y_1 - x_1}{\sqrt{3}} \sqrt{\frac{k+1}{2s^2 + \|y_1 - x_1\|^2}}, L_2(x, s^2) = \sqrt{\frac{2}{3}}(x_1 - x_2 - m) + \frac{y_1 - x_1}{\sqrt{3}} \sqrt{\frac{k+1}{2s^2 + \|y_1 - x_1\|^2}}. \)

Proof. The proof is straightforward and analogous to Theorem 3.1.

3.1 Risk performance

It would be interesting to compare the frequentist risk performance of the Bayes predictive density estimator
\( \hat{q}_{\pi,A} \) (the Bayes estimator with considering additional information from Theorem 3.1 or 3.2 depending on
\( A \)) and \( \hat{q}_{\pi_0} \) (the Bayes estimator without considering additional information from Lemma 2.2). For model
(2.7) the KL risk function is given by
\[
R_{KL}(\theta, \hat{q}) = \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} L_{KL}(\theta, \hat{q}(\cdot; x)) p_\theta(x_1, x_2) \, dx_1 \, dx_2.
\]
where \( p_\theta(x_1, x_2) = \frac{1}{\sigma_{X_1} \sigma_{X_2}} \phi\left(\frac{x_1 - \theta_1}{\sigma_{X_1}}\right) \phi\left(\frac{x_2 - \theta_2}{\sigma_{X_2}}\right). \)

Figure 1 and 2 present the relative efficiency (risk ratio) of \( \hat{q}_{\pi,A} \) over \( \hat{q}_{\pi_0} \), for \( p = 1, k = 3 \) and \( \Delta = (\theta_1 - \theta_2)/\sigma \) based on restricted parameter \( A = [0, \infty) \) and \( A = [-6, 6] \) respectively. For both graphs, we have about 12% improvement in the KL risk function.

Figure 1: Risk ratio of the Bayes predictive density estimators for \( p = 1, k = 3 \) and \( A = [0, +\infty) \).
4 Examples

In this section, we apply the proposed methods in order to construct Bayes predictive density estimators through two well-known medical data.

**Example 1 (Leg-length discrepancy predictive density estimation)**

Leg-length Discrepancy (LLD), the difference between the lengths of two legs, is a topic that seemingly has been exhaustively examined. The LLD may be caused by trauma or mild developmental abnormalities, with onset in birth or childhood. In fact, it causes several conditions, including low back pain; osteoarthritis of the hip and knee; knee pain and running injuries, such as Achilles rupture. Harvey et al. (2010) used radiography to evaluate leg length in 3,026 adults. After following participants for 30 months they conducted exploratory analyses to determine whether there was an important threshold value of the LLD above which knee osteoarthritis was more likely. They did this by stratifying the LLD into four categories: less than 0.5 cm (reference group), 0.5 cm to less than 1 cm, 1 cm to less than 2 cm, and 2 cm or more. Their result showed leg-length inequality of 1 cm or more to be associated with prevalent, incident, symptomatic and progressive knee osteoarthritis that was strongest in the shorter leg.

Table 1 shows the body mass of participants grouped with the LLD (defined as inequality of 1 cm or more).

| Body mass index | sample size | mean | sd |
|-----------------|-------------|------|----|
| LLD greater or equal than 1 cm (group 1) | 429 | 31 | 5.7 |
| LLD less than 1 cm (group 2) | 2535 | 30.4 | 5.7 |

Table 1: Patient body mass

Also it is statistically significant at the 0.05 level the mean of body mass index in the group with LDD $\geq 1$ is greater than group with LDD $< 1$. Suppose random variable $X_1$, the body mass index with LDD $\geq 1$, follows $N(\theta_1, \sigma_1^2)$, is independent of $X_2$, the body mass index with LDD $< 1$ which is distributed as $N(\theta_2, \sigma_2^2)$, when their means are subject to the order restriction $\theta_1 \geq \theta_2$ and variances $\sigma_1^2$ and $\sigma_2^2$ are unknown.

Table 2 contains the predictive density estimators $\hat{q}_{\pi,A}$, and the Bayes estimators $q_{\pi_0}$ (predictive density
estimators without and with considering the additional information respectively) along with their means, 10th, 50th and 90th percentiles for the future density $Y_1$ of the body mass index of patients with $\text{LLD} \geq 1$ cm, based on the data from Table 1.

| Predictive Density estimation | Estimator | PDF | $y_1$, $P_{0.1}$, $P_{0.5}$, $P_{0.9}$ |
|------------------------------|-----------|-----|-------------------------------------|
| Bayes without additional information | $T_1(n_1 - 1, \bar{x}_1, \sqrt{2s^2_{n_1-1}})$ | $T_1(428,31,0.39)$ | 31, 30.5, 31, 31.5 |
| Bayes with additional information | $ST_1(n_1 - 1, \bar{x}_1 - \bar{x}_2, \sqrt{\frac{2s^2_{n_1-1}}{n_1-1}})$ | $ST_1(428,1.26,31,0.39)$ | 31.02, 30.52, 31.02, 31.52 |

Table 2: Predictive density estimators of future density of the body mass index of patients with $\text{LLD} \geq 1$ cm, along with their means, 10th, 50th and 90th percentiles.

**Example 2 (Child’s first walk)**

An experiment was conducted to evaluate the effect of exercise on the age at which a child starts to walk (see Silvapulle and Sen, 2005). Let $X$ denote the age (in months) at which a child starts to walk.

| Group 1 | 11 | 10 | 10 | 11.75 | 10.5 | 15 | $\bar{x}_1 = 11.37$, $s_1 = 1.44$ |
|---------|----|----|----|-------|------|----|-----------------------------------|
| Group 2 | 9  | 9.5| 9.75| 10    | 13   | 9.5| $\bar{x}_2 = 10.12$, $s_2 = 1.9$  |

Table 3: The age at which a child first walks

The first group performed daily exercises but not the special walking exercises while the second group performed a special walking exercise for 12 minutes per day beginning at age 1 week and lasting 7 weeks. (the original experiment consists of other groups, however, here we consider only two of them.)

For groups $i (i = 1, 2)$ let, $\theta_i$ be the mean age (in months) at which a child starts to walk. However, suppose that the researcher was prepared to assume that the walking exercises would not have negative effect of increasing the mean age at which a child starts to walk, and it was desired that this additional information be incorporated to improve on the statistical analysis. In this case, we have that $\theta_1 \geq \theta_2$. this can be considered as two univariate normal distributions $X_1$, $X_2$, when their means are subject to the order restriction $\theta_1 \geq \theta_2$ and variances $\sigma^2_1$ and $\sigma^2_2$ are different and unknown. Analogous to example 1, Table 4 can be similarly obtained.

| Predictive Density estimation | Estimator | PDF | $y_1$, $P_{0.1}$, $P_{0.5}$, $P_{0.9}$ |
|------------------------------|-----------|-----|-------------------------------------|
| Bayes without additional information | $T_1(n_1 - 1, \bar{x}_1, \sqrt{2s^2_{n_1-1}})$ | $T_1(5,11.37,1.2)$ | 11.37, 9.6, 11.37, 13.14 |
| Bayes with additional information | $ST_1(n_1 - 1, \bar{x}_1 - \bar{x}_2, \sqrt{\frac{2s^2_{n_1-1}}{n_1-1}})$ | $ST_1(5,0.85,11.37,1.2)$ | 11.45, 11.2, 11.44, 12.37 |

Table 4: Predictive density estimators of future density of child first walks in group 1 along with their means, 10th, 50th and 90th percentiles.

Figure 3, helps to visualize different predictive density estimators and the corresponding means and percentiles in Table 4.
5 Concluding remarks

This paper extends the line of work which seeks to find Bayesian interpretations of the skew–normal densities to skew–Student $t$ distributions. We have shown that different kind of constraints on the parameter space in a hierarchical normal model, yield the Bayesian predictive densities belong to a class of weighted Student $t$ distributions. More specifically we studied the restrictions $\theta_1 - \theta_2 \in \mathbb{R}_+^p$, and $\theta_1 - \theta_2 \in [-m, m]^p$, in model (2.7), which provides two different skew–Student $t$ distributions based on Definitions 1.1 and 1.2 respectively. Results suggest Bayes predictive density estimators based on additional information performs better than the Bayes predictive density without considering additional information in term of KL risk function. Finally, some numerical comparison and important examples were done to support the results.

Acknowledgement

The author thanks Éric Marchand (Université de Sherbrooke) for his very helpful comments on the manuscript.

Appendix

(A.1) Proof of Lemma 2.1

Proof. (a) We have

\[
\pi(\theta, \sigma^2 | x, s^2) \propto (\sigma^2)^{-(p+k/2+1)} \exp \left\{ \frac{-t}{2\sigma^2} \right\} \mathbb{I}_A(\theta_1 - \theta_2), \tag{5.21}
\]
where $t = \|x_1 - \theta_1\|^2 + \|x_2 - \theta_2\|^2 + s^2$. Now, letting $z = \frac{t}{2\sigma^2}$, by integrating out $\sigma^2$, we have
\[
\pi(\theta | x, s^2) \propto \int_0^{\infty} (\sigma^2)^{-(p+k/2)+1} \exp\left\{-\frac{t}{2\sigma^2}\right\} \mathbb{I}_A(\theta_1 - \theta_2) \, d\sigma^2 \\
\propto t^{-(p+k/2)} \mathbb{I}_A(\theta_1 - \theta_2) \int_0^{\infty} z^{p+k/2-1} \exp\{-z\} \, dz \\
\propto t^{-(p+k/2)} \mathbb{I}_A(\theta_1 - \theta_2) \\
\propto (\|x_1 - \theta_1\|^2 + \|x_2 - \theta_2\|^2 + s^2)^{-(p+k/2)} \mathbb{I}_A(\theta_1 - \theta_2) \\
\propto \left(1 + \frac{\|x_2 - \theta_2\|^2}{s^2 + \|x_1 - \theta_1\|^2}\right)^{-(p+k/2)} \left(1 + \frac{\|x_1 - \theta_1\|^2}{s^2}\right)^{-(p+k/2)} \mathbb{I}_A(\theta_1 - \theta_2).
\]
(b) We have
\[
\pi(\theta_1 | x, s^2) = \int \pi(\theta | x, s^2) \, d\theta_2 \\
\propto \left(1 + \frac{\|x_1 - \theta_1\|^2}{s^2}\right)^{-(p+k/2)} \int_{\{|\theta_2| < 2\pi\}} \left(1 + \frac{\|x_2 - \theta_2\|^2}{s^2 + \|x_1 - \theta_1\|^2}\right)^{-(p+k/2)} \, d\theta_2, \\
\propto \left(1 + \frac{\|\theta_1 - x_1\|^2}{k(s^2/k)}\right)^{-(p+k/2)} \int_A \left(1 + \frac{\|t - (\theta_1 - x_2)\|^2}{s^2 + \|x_1 - \theta_1\|^2}\right)^{-(p+k/2)} \, dt,
\]
by the change of variable $t = \theta_1 - \theta_2$. 

(A.2) Proof of Theorem 3.1

By setting $U = \eta(\theta_1 - x_1)$, and $\eta = \frac{1}{\sigma^2}$, one can write the joint density of $(U, \eta) | x, s^2$ as multiplication of equations (3.14) and (3.15) in Lemma 3.3. So we have
\[
\tilde{q}_{\pi,A}(y_1; x, s^2) = \mathbb{E}(U, \eta | x, s^2) q(y_1 | x_1 + \frac{U}{\sqrt{\eta}}, \frac{1}{\sqrt{\eta}}) \\
= \int_{\mathbb{R}_+^p} \left(\int_{\mathbb{R}_+^p} \left(\frac{\eta}{2\pi}\right)^{\frac{p-k}{2}} e^{-\frac{1}{2\eta} (y_1 - \frac{\eta}{2})^2 \Phi_p(u) \Phi_p(\alpha_0 + \alpha_1 u; 0)} \frac{\Phi_p(\frac{\alpha_0}{\sqrt{1+\alpha_1^2}}; 0)}{(2\pi)^{\frac{p}{2}}} \, du\right) \pi(y_1 | x, s^2) \, d\eta \\
= \int_{\mathbb{R}_+^p} \frac{\eta^{1-k/2}}{(2\pi)^{\frac{p}{2}}} F_p(k; 2x_1^2) \Gamma(\frac{1}{2}) \left(\frac{2}{s^2}\right)^{k/2} \int_{\mathbb{R}_+^p} e^{-\frac{1}{2}u^2 + \sqrt{\eta}u^2} \Phi_p(u) \Phi_p(\alpha_0 + \alpha_1 u; 0) \, du \, d\eta \\
= \frac{\Gamma(k+p/2)}{\Gamma(k/2) \left(\frac{2}{s^2}\right)^{k/2} (2\pi)^{\frac{p}{2}}} F_p(k; 2x_1^2) \left(\frac{s^2}{2} + \frac{y_1 - x_1)^2}{4}\right)^{-k+p/2} \\
\times \int_{\mathbb{R}_+^p} \eta^{1-k/2} e^{-\frac{1}{2} \eta \left(\frac{x_1^2}{2} + \frac{(y_1 - x_1)^2}{4}\right)} \frac{1}{\sqrt{2}} \Phi_p\left(\sqrt{\eta} \left(\frac{x_1 - x_2}{\sqrt{3}} + \frac{y_1 - x_1}{\sqrt{6}}; 0\right)\right) \, d\eta \\
= \frac{\Gamma(k+p/2)}{\Gamma(k/2) \left(\frac{2}{s^2}\right)^{k/2} (2\pi)^{\frac{p}{2}}} F_p(k; 2x_1^2) \left(\sqrt{2} \mathbb{E}_{\eta} \Phi_p\left(\sqrt{\eta} \left(\frac{x_1 - x_2}{\sqrt{3}} + \frac{y_1 - x_1}{\sqrt{6}}; 0\right)\right)\right),
where $\eta \mid x, s^2 \sim Gamma\left(\frac{k+1}{2}, \frac{s^2}{2} + \frac{\|y_1-x_1\|^2}{4}\right)$, applying the identity (3.17) to above expectation completes the proof.

References

[1] Aitchison, J. (1975). Goodness of prediction fit. Biometrika, 62(3), 547-554.

[2] Azzalini, A. (1985). A class of distributions which includes the normal ones. Scandinavian Journal of Statistics, 12, 171-178.

[3] Azzalini, A., and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew $t$ distribution. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65(2), 367-389.

[4] Azzalini, A. and Valle, A. D. (1996). The multivariate skew–normal distribution. Biometrika, 83(4), 715-726.

[5] Bauwens, L., and Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. Journal of Business & Economic Statistics, 23(3), 346-354.

[6] Branco, M. D. and Dey D. K. (2001). A general class of multivariate skew–elliptical distributions. Journal of Multivariate Analysis, 79, 99-113.

[7] Capitanio, A., Azzalini, A., and Stanghellini, E. (2003). Graphical Models for Skew- Normal Variates. Scandinavian Journal of Statistics, 30, 129-144.

[8] Corcuera, J. M., and Giummolè, F. (1999). A generalized Bayes rule for prediction. Scandinavian Journal of Statistics, 26(2), 265-279.

[9] Fernandez, C. and Steel M. (1998). On Bayesian modelling of fat tails and skewness. Journal of the American Statistical Association 93, 359–371.

[10] Gupta, R. C., and Gupta, R. D. (2004). Generalized skew normal model. Test, 13(2), 501-524.

[11] Hansen, B. (1994). Autoregressive conditional density estimation. International Economic Review, 35, 705-730.

[12] Harvey, W. F., Yang, M., Cooke, T. D., Segal, N. A., Lane, N., Lewis, C. E., & Felson, D. T. (2010). Association of leg-length inequality with knee osteoarthritis: a cohort study. Annals of internal medicine, 152(5), 287-295.

[13] Liseo, B., and Loperfido, N. (2003). A Bayesian interpretation of the multivariate skew–normal distribution. Statistics & Probability Letters, 61(4), 395-401.

[14] O'Hagan, A., Leonard, T., (1976). Bayes estimation subject to uncertainty about parameter constraints. Biometrika 63(1), 201-203.

[15] Patton, A. (2004). On the out–of–sample importance of skewness and asymmetric dependence for asset allocation. Journal of Financial Econometric 2(1), 130–168.
[16] Roberts, C. (1998). A correlation model useful in the study of twins. *Journal of the American Statistical Association, 61*, 1184-1190.

[17] Silvapulle, M. J., and Sen, P. K. (2011). *Constrained statistical inference: Order, inequality, and shape constraints* 912. John Wiley & Sons.