Journal of Mathematical Extension
Vol. 5, No. 2 (2), (2011), 37-50

Empirical Bayes Estimation in Multiple Linear Regression with Multivariate Skew-Normal Distribution as Prior

M. Khounsiavash*
Islamic Azad University-Science and Research Branch

T. Baghfalaki
Shahid Beheshti University

M. Ganjali
Shahid Beheshti University

Abstract. We develop a new empirical Bayes analysis in multiple regression models. In the present work we consider multivariate skew-normal as prior for coefficients of the model in a skew-normal population and give empirical Bayes estimation for parameters of the model. The marginal distribution of response is found to be a closed skew-normal distribution. The empirical Bayes estimator is found in a closed form and the model is applied on a data set.

AMS Subject Classification: 62C12.
Keywords and Phrases: Multivariate skew-normal distribution, multivariate closed skew-normal distribution; empirical bayes estimation; bayesian regression.

1. Introduction

The skew symmetric distributions have been used in many different ways in Bayesian regression analysis. For example, De la Cruz ([8]) discussed non-linear regression models for longitudinal data with errors that follow

Received January 2011; Final Revised September 2011
*Corresponding author

skew-elliptical distribution. Rodrigues and Bolfarine ([15]) used skew priors where the observations contain measurement errors and a positive explanatory variable that causes a strong asymmetry on the response variable. Ferreira and Steel ([9]) introduced multivariate skewed regression with fat tails by considering a linear regression structure with skewed and heavy-tailed error terms. In order to allow for heavy tails they used skewed versions of t-student distribution. Sahu et al. ([16]) developed a new class of distributions by introducing skewness in multivariate elliptical symmetric distributions and gave practical applications in Bayesian regression models. There are many other ideas in application of skew elliptical distributions. For example Mukhopadhyhy and Vidakovic ([14]) compared the performance of a linear Bayes rules in estimating a normal mean with that of corresponding Bayes rules. They used the family of standard skew normal distribution as prior. Here, in multiple regression model we shall use skew-normal distribution as prior for coefficient of linear regression models which can be empirically predicted. The main advantage of this article is the application of multivariate skew-normal distribution as prior for the regression coefficients with a marginal closed skew-normal response variable.
The plan of the remainder of this paper is as follows: In Section 2 we introduce the model and the prior distributions. In Section 3 we obtain the empirical Bayes estimates of the parameters of the model in a closed form and in Section 4 an example is analyzed based on the results obtained in the preceding sections.

2. Distributions, Model and Prior

2.1 Distributions

The skew-normal distribution introduced by Azzalini ([1]) refer to a parametric class of probability distribution which includes the standard normal as a special case, a random variable X has a standard skew-normal distribution, written by $X \sim SN(\lambda)$, if its probability density function (pdf) is given by

$$f_X(x) = 2\phi(x)\Phi(\lambda x) \quad , \lambda \in R, \quad , x \in R, \tag{1}$$

where ϕ and Φ are the pdf and cumulative distribution function (cdf) of a standard normal, respectively, and λ is the skewness parameter.
For definition of multivariate version of this distribution consider the following lemma (Azzalini, [1]).

Lemma 2.1. *If f_0 is the d-dimensional probability density function such that $f_0(x) = f_0(-x)$ for $x \in R^d$, and G is a one-dimensional differentiable distribution function such that G' is a density symmetric about 0, and w is a real-valued function such that $w(-x) = -w(x)$ for all $x \in R^d$, then*

$$f(x) = 2f_0(x)G\{w(x)\}, \quad x \in R^d, \tag{2}$$

is a density function on R^d.
The above formulation is in the form presented by Azzalini and Capitanio ([4]). Of course, the statement is more general but less operative than this lemma. This lemma also proved in Azzalini and Capitanio ([3]). Consider the case that $f_0(x)$ in lemma 1.1 is $\phi_d(x;\Omega)$ the density function of an $N_d(0,\Omega)$ variable, where Ω is a positive definite matrix, also take $G = \Phi$ (Where Φ is the cumulated distribution function (cdf) of a standard normal distribution) and w to be a linear function. The density function with location parameter ζ is

$$f(y) = 2\phi_d(y-\zeta;\Omega)\Phi(\alpha'\omega^{-1}(y-\zeta)), \quad y \in R^d, \tag{3}$$

where α is the shape (skewness) parameter vector ($\alpha \in R^d$) and ω is the diagonal matrix formed by the positive root of the diagonal elemnts of Ω. In this case Y has multivariate SN distribution and we write $Y \sim SN_d(\zeta,\Omega,\alpha)$.
In the following definition we review multivariate closed skew-normal distribution from Gonzlez-Faras et. al. ([11])

Definition 2.2. *Consider $p \geqslant 1$, $q \geqslant 1$, $\mu \in R^p, \nu \in R^q$, D an arbitrary $q \times p$ matrix and Σ and Δ positive definite matrices of dimensions $p \times p$ and $q \times q$, respectively. Then the probability density function (pdf) of a CSN distribution is given by:*

$$g_{p,q}(y) = C\phi_p(y;\mu,\Sigma)\Phi_q(D(y-\mu);\nu,\Delta), \quad y \in R^p \tag{4}$$

with:

$$C^{-1} = \Phi_q(0; \nu, \Delta + D\Sigma D')$$

where $\Phi_p(.; \eta, \Psi)$ *is the cdf of the p-dimensional normal distribution, with* $\eta \in R^p$ *as the mean and* Ψ *a* $p \times p$ *covariance matrix. We denote a p-dimensional random vector distributed according to a CSN distribution with parameters* q, μ, Σ, D, ν *and* Δ *by*

$$y \sim CSN_{p,q}(\mu, \Sigma, D, \nu, \Delta).$$

The multivariate skew-normal distribution have been introduced as a generalization of the normal distribution to model, in a natural way, skewness features in the multi dimensional distribution. This families also have properties similar to the normal distribution. However, two important properties have been absent: the closure for the joint distribution of independent members of the multivariate skew-normal family and the closure under linear combinations other than those given by non-singular matrices.

The CSN distribution, as defined in Definition 1.1, has more properties similar to the normal distribution than any other. Gonzlez-Faras et al. ([11]) show that for a random vector with the CSN distribution, all column (row) full rank linear transformations are in the family of CSN distributions. They also prove closure under sum of independent CSN random vectors and the closure for the joint distribution of independent CSN distributions, thus providing a result that characterizes the CSN distributions.

2.2 Model and Prior

We consider the following linear regression model that is

$$Y = X\beta + \epsilon \tag{5}$$

where Y is an $n \times 1$ vector of observation, X is an $n \times p$ matrix of known constants, β is a $p \times 1$ vector of unknown parameters and ϵ is an $n \times 1$ vector of unobservable random errors, both Y and ϵ are random vectors. We assume that $Y|\beta \sim N(X\beta, \sigma^2 I)$, $\epsilon|\beta \sim N(0, \sigma^2 I)$ also

$\beta \sim SN(\zeta, \Omega, \alpha)$.
This model has two good properties, (1) when we use multivariate skew-normal distribution as prior density for β, this includes multivariate normal distribution (distribution with zero skewness parameters) as a special case, and (2) with attention to marginal distribution of Y, Y has CSN distribution, (we prove this in the next section) therefore this model is a good descriptive for skew population.
In the following, we discuss the empirical Bayes for estimating parameters of the model, given in (5).

3. Empirical Bayes Estimation

Assume that the multiple linear regression model is in the form of (5) where Y is a random vector with

$$Y|\beta \sim N(X\beta, \sigma^2 I).$$

The unknown parameters in the model are the set of coefficients $\beta = (\beta_1, \ldots, \beta_p)^t$, we assume that σ^2 is known until the end of the section where we will discuss how to handle the case of unknown σ^2.
For the Bayesian multiple linear regression model we consider multivariate skew-normal distribution as appropriate prior for β, that is $\beta \sim SN(\zeta, \Omega, \alpha)$, where ζ, Ω and α are vector of hyperparameters. Following Gonzalez-Farias et al. ([12])

$$\beta \sim CSN_{p\times 1}(\zeta, \Omega, \alpha'\omega^{-1}, 0, 1),$$

and

$$\epsilon \sim CSN_{n\times 1}(0, \sigma^2 I, 0, 0, 1).$$

Therefore

$$X\beta \sim SSN_{n\times 1}(X\zeta, X\Omega X', \alpha'\omega^{-1}(X'X)^{-1}X', 0, 1)$$

where SSN means singular skew-normal, and finally,

$$Y \sim CSN_{n\times 2}(X\zeta, X\Omega X' + \sigma^2 I, D^*, V^*, \Delta^*)$$

where

$$V^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$D^* = \begin{pmatrix} D_1 \\ 0 \end{pmatrix}_{2\times n};\ D_1 = \alpha'\omega^{-1}\Omega X'[X\Omega X' + \sigma^2 I]^{-1},$$

$$\Delta^* = \begin{pmatrix} A_{11} & 0 \\ 0 & 1 \end{pmatrix};$$

where,

$$\begin{aligned} A_{11} &= 1 + \alpha'\omega^{-1}\Omega\omega^{-1}\alpha - \alpha'\omega^{-1}\Omega X'(X\Omega X' + \sigma^2 I)^{-1} X\Omega\omega^{-1}\alpha \\ &= 1 + \alpha'\omega^{-1}(I + \Omega X'(X\Omega X' + \sigma^2 I)^{-1} X)\Omega\omega^{-1}\alpha. \end{aligned}$$

Therefore the marginal distribution of Y is multivariate closed skew-normal (CSN) distribution.
As in equation (4), $V = 0$, the parameters of the multivariate closed skew-normal distribution can be identifiable (Flecher et. al., ([10]), pp. 1979). Because of this property in our model, all of the parameters are identifiable.
In empirical Bayes methodology we estimate hyperparameters of prior distribution, by maximization of marginal distribution of Y with respect to $\theta = (\zeta, \Omega, \alpha)$, therefore numerical maximization of $ln[m(Y|X, \theta]$ (logarithm of marginal distribution) is required. One may use "R" software to find this maximum point. For this purpose, at first we have to define the logarithm of marginal distribution to "R" as a function of its parameters, then we have to use "*nlminb*" routine for optimizing (minimizing) the function.
Now let we get this numerical estimates and recall them $\hat{\theta} = (\hat{\zeta}, \hat{\Omega}, \hat{\alpha})$.We now present the empirical Bayes estimation for β, following Lehmann ([13], pp. 241).

$$\begin{aligned} f(y|X, \beta) &= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}(y'y - 2\beta' X' y + \beta' X' X\beta)} \\ &= \frac{e^{-\frac{1}{2\sigma^2} y'y}}{(2\pi\sigma^2)^{n/2}} e^{\frac{1}{\sigma^2}\beta' X' y} e^{-\frac{1}{2\sigma^2}\beta' X' X\beta}. \end{aligned}$$

Therefore

$$E[\frac{\partial}{\partial Y}\frac{\beta' X' Y}{\sigma^2}|Y,X] = \frac{1}{m(y|\hat{\theta},X)}\int(\frac{\partial}{\partial Y}\frac{\beta' X' Y}{\sigma^2})f(Y|\beta)\pi(\beta)d\beta$$

$$= \frac{1}{m(y|\hat{\theta},X)}\int(\frac{\partial}{\partial Y}\frac{\beta' X' Y}{\sigma^2})e^{\frac{1}{\sigma^2}\beta' X' Y}e^{-\frac{1}{2\sigma^2}\beta' X' X\beta}h(Y)\pi(\beta)d\beta$$

$$= \frac{1}{m(y|\hat{\theta},X)}\int[(\frac{\partial}{\partial Y}e^{\frac{1}{\sigma^2}\beta' X' Y}h(Y))$$

$$-(\frac{\partial h(Y)}{\partial Y})e^{\frac{1}{\sigma^2}\beta' X' Y}]e^{-\frac{1}{2\sigma^2}\beta' X' X\beta}\pi(\beta)d\beta$$

$$= \frac{\partial\ln(m(y|\hat{\theta},X))}{\partial Y} - \frac{\partial ln(h(Y))}{\partial Y}.$$

We have $\frac{\partial ln(h(Y))}{\partial Y} = -\frac{y}{\sigma^2}$, for finding $\frac{\partial\ln(m(y|\hat{\theta},X))}{\partial Y}$ as $m(y|\hat{\theta},X)$ has multivariate closed skew-normal density $(CSN(\hat{\mu},\hat{\Sigma},\hat{D}^*,V^*,\hat{\Delta}^*))$ with $\hat{\mu} = X\hat{\zeta},\ \hat{\Sigma} = X\hat{\Omega}X' + \sigma^2 I$, we have

$$\begin{aligned} m(y|\hat{\theta},X) &= \phi_n(y;\hat{\mu},\hat{\Sigma})\frac{\Phi_2(\hat{D}^*(y-\hat{\mu}),V^*,\hat{\Delta}^*)}{\Phi_2(0,V^*,\hat{\Delta}^*+\hat{D}^*\Sigma\hat{D}^{*t})} \\ &= \frac{\phi_n(y;\hat{\mu},\hat{\Sigma})}{\Phi_2(0,V^*,\hat{\Delta}^*+\hat{D}^*\hat{\Sigma}\hat{D}^{*\prime})}\times\frac{1}{2}\times\Phi(\frac{\hat{D}_1(\mathbf{y}-\hat{\mu})}{\sqrt{\hat{A}_{11}}}). \end{aligned}$$

Since

$$\begin{aligned} \Phi_2(\hat{D}^*(y-\hat{\mu}),V^*,\hat{\Delta}^*) &= \int_{-\infty}^{\hat{D}_1(y-\hat{\mu})}\int_{-\infty}^{0}\phi_2((t_1,t_2),V^*,\hat{\Delta}^*)\,dt_1\,dt_2 \\ &= \int_{-\infty}^{\hat{D}_1(y-\hat{\mu})}\int_{-\infty}^{0}\frac{e^{-\frac{1}{2\hat{A}_{11}}t_1^2}}{\sqrt{2\pi\hat{A}_{11}}}\frac{e^{-\frac{1}{2}t_2^2}}{\sqrt{2\pi}}dt_1\,dt_2 \\ &= \frac{1}{2}\int_{-\infty}^{\hat{D}_1(y-\hat{\mu})}\frac{e^{-\frac{1}{2\hat{A}_{11}}t_1^2}}{\sqrt{2\pi\hat{A}_{11}}}dt_1 \\ &= \frac{1}{2}\times\Phi(\frac{\hat{D}_1(y-\hat{\mu})}{\sqrt{\hat{A}_{11}}}), \end{aligned}$$

thus

$$\begin{aligned}\ln(m(y|\hat{\theta}, X)) &= \ln(\phi_n(y; \hat{\mu}, \hat{\Sigma})) - \ln(\Phi_2(0, V^*, \hat{\Delta}^* + \hat{D}^*\hat{\Sigma}\hat{D}^{*'})) \\ &\quad - \ln(2) + \ln(\Phi(\frac{\hat{D}_1(y-\hat{\mu})}{\sqrt{\hat{A}_{11}}}))\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \ln(m(y|\hat{\theta}, X))}{\partial y} &= \frac{\partial \ln(\phi_n(y; \hat{\mu}, \hat{\Sigma}))}{\partial y} + \frac{\frac{\hat{D}_1'}{\sqrt{\hat{A}_{11}}}\phi(\frac{\hat{D}_1(y-\hat{\mu})}{\sqrt{\hat{A}_{11}}})}{\Phi(\frac{\hat{D}_1(y-\hat{\mu})}{\sqrt{\hat{A}_{11}}})} \\ &= -\hat{\Sigma}^{-1}(y-\hat{\mu}) + \frac{\frac{\hat{D}_1'}{\sqrt{\hat{A}_{11}}}\phi(\frac{\hat{D}_1(y-\hat{\mu})}{\sqrt{\hat{A}_{11}}})}{\Phi(\frac{\hat{D}_1(y-\hat{\mu})}{\sqrt{\hat{A}_{11}}})}.\end{aligned}$$

Therefore empirical Bayes estimation is given by

$$\begin{aligned}E(\frac{X\beta}{\sigma^2}|Y) &= \frac{y}{\sigma^2} - \Sigma^{-1}(y-\hat{\mu}) + \frac{\frac{\hat{D}_1'}{\sqrt{\hat{A}_{11}}}\phi(\frac{\hat{D}_1(y-\hat{\mu})}{\sqrt{\hat{A}_{11}}})}{\Phi(\frac{\hat{D}_1(y-\hat{\mu})}{\sqrt{\hat{A}_{11}}})}, \\ XE(\beta|Y) &= y - \sigma^2\hat{\Sigma}^{-1}(y-\hat{\mu}) + \sigma^2\frac{\frac{\hat{D}_1'}{\sqrt{\hat{A}_{11}}}\phi(\frac{\hat{D}_1(y-\hat{\mu})}{\sqrt{\hat{A}_{11}}})}{\Phi(\frac{\hat{D}_1(y-\hat{\mu})}{\sqrt{\hat{A}_{11}}})},\end{aligned}$$

thus

$$\begin{aligned}E(\beta|Y) &= (X'X)^{-1}X'y - \sigma^2(X'X)^{-1}X'\hat{\Sigma}^{-1}(y-\hat{\mu}) \\ &+ \sigma^2(X'X)^{-1}X'\frac{\frac{\hat{D}_1'}{\sqrt{\hat{A}_{11}}}\phi(\frac{\hat{D}_1(y-\hat{\mu})}{\sqrt{\hat{A}_{11}}})}{\Phi(\frac{D_1(y-\hat{\mu})}{\sqrt{\hat{A}_{11}}})}. \qquad (6)\end{aligned}$$

It's clear that, this estimator is an extension of the empirical Bayes estimation for multiple regression when we select multivariate normal distribution ($\alpha = 0$ then $\hat{D}_1 = 0$) as prior distribution for β.
For computing $MSE(\beta, \hat{\beta})$, we have

$$MSE(\beta, \hat{\beta}) = (X'X)^{-1}X'MSE(X\beta, X\hat{\beta})X(X'X)^{-1}$$

which $MSE(X\beta, X\hat{\beta})$ given by:

$$
\begin{aligned}
R(X\beta, E[X\beta|Y]) &= E\left[(X\beta - E[X\beta|Y])\,(X\beta - E[X\beta|Y])'\right] \\
&= E\left[\left(X\beta - \sigma^2\left(\frac{\partial \ln(m(Y|\hat{\theta},X))}{\partial Y} - \frac{\partial \ln(h(Y))}{\partial Y}\right)\right)\right. \\
&\times \left.\left(X\beta - \sigma^2\left(\frac{\partial \ln(m(Y|\hat{\theta},X))}{\partial Y} - \frac{\partial \ln(h(Y))}{\partial Y}\right)\right)'\right] \\
&= E\left[\left(X\beta - Y - \sigma^2\left(\frac{\partial \ln(m(Y|\hat{\theta},X))}{\partial Y}\right)\right)\right. \\
&\times \left.\left(X\beta - Y - \sigma^2\left(\frac{\partial \ln(m(Y|\hat{\theta},X))}{\partial Y}\right)\right)'\right] \\
&= E\left[(X\beta - Y)\,(X\beta - Y)'\right] \\
&- 2\sigma^2 E\left[(X\beta - Y)\left(\frac{\partial \ln(m(Y|\hat{\theta},X))}{\partial Y}\right)'\right] \\
&+ \sigma^4 E\left[\left(\frac{\partial \ln(m(Y|\hat{\theta},X))}{\partial Y}\right)\left(\frac{\partial \ln(m(Y|\hat{\theta},X))}{\partial Y}\right)'\right] \\
&\overset{*}{=} E\left[(X\beta - Y)\,(X\beta - Y)'\right] \\
&- 2\sigma^4 E\left[\frac{\partial^2 \ln(m(Y|\hat{\theta},X))}{\partial Y \partial Y'}\right] \\
&+ \sigma^4 E\left[\left(\frac{\partial \ln(m(Y|\hat{\theta},X))}{\partial Y}\right)\left(\frac{\partial \ln(m(Y|\hat{\theta},X))}{\partial Y}\right)'\right] \\
&= E[YY'] - 2EY]\beta' X' + X\beta\beta' X' + 2\sigma^4 \Sigma^{-1} \\
&+ \sigma^4 \Sigma^{-1}\left\{EYY'] - 2EY]\beta' X' + X\beta\beta' X'\right\}\Sigma^{-1} \\
&- 2\sigma^4 \times \frac{{D_1}' D_1}{A_{11}} \times E\left(\frac{\phi\left(\frac{D_1(Y-\mu)}{\sqrt{A_{11}}}\right)}{\Phi\left(\frac{D_1(Y-\mu)}{\sqrt{A_{11}}}\right)}\right) \\
&+ 2\sigma^4 \times \frac{{D_1}'}{\sqrt{A_{11}}} E\left((Y-\mu)\,\frac{\phi\left(\frac{D_1(Y-\mu)}{\sqrt{A_{11}}}\right)}{\Phi\left(\frac{D_1(Y-\mu)}{\sqrt{A_{11}}}\right)}\right) \\
&+ 3\sigma^4 \times \frac{{D_1}' D_1}{A_{11}} \times E\left(\frac{\phi\left(\frac{D_1(Y-\mu)}{\sqrt{A_{11}}}\right)}{\Phi\left(\frac{D_1(Y-\mu)}{\sqrt{A_{11}}}\right)}\right)^2
\end{aligned}
$$

which * *equality* is a straight consequence of Stein's identity. Also

$E[YY']$ and $E[Y]$ had given by equation $(4)-(5)$ from Flecher et. al. [10], the other expectation have to calculate by an approximation method, such as Monte-Carlo integration.

3.1 Unknown σ^2 Cases

When σ^2 is unknown, we can define model with two approaches. In Bayesian methodology we assign prior on every unknown parameter, therefore we can use inverse-gamma distribution or Jeffryes prior for σ^2 in Bayesian model. Also, Berger ([6], pp. 172) suggested a method for estimating σ^2 based on repeated ys. In multiple linear regression responses are given by

$$y_i = \beta' \mathbf{x_i} + \epsilon_\mathbf{i}; \quad \mathbf{i : 1, \ldots, n}$$

such that

$$\epsilon_i|\beta \sim \mathbf{N}(\mathbf{0}, \sigma^\mathbf{2}),$$

that is

$$y_i|\beta \sim \mathbf{N}(\beta' \mathbf{x_i}, \sigma^\mathbf{2}); \;\; \mathbf{i : 1, \ldots, n}.$$

In the presence of replication of y, suppose, $\left\{y_i^j, i : 1, 2, \ldots, n\right\}$ are the replicated observations for each j $(j : 1, \ldots, K)$ and

$$y_i^j \sim N(\beta' \mathbf{x_i}, \sigma^\mathbf{2}), \;\; \beta \sim \mathbf{SN}(\xi, \mathbf{\Omega}, \alpha)$$

then

$$\bar{y}^j|\beta = \frac{\mathbf{1}}{\mathbf{n}} \sum\nolimits_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{y_i^j}, \quad \mathbf{S^2} = \frac{\mathbf{1}}{\mathbf{K}} \sum_{\mathbf{j=1}}^{\mathbf{K}} \mathbf{S_j^2}$$

where

$$S_j^2 = \frac{1}{n-1} \sum\nolimits_{i=1}^{n} (y_i^j - \bar{y}^j)^2,$$

and this is a sufficient statistics. As suggested by Berger [6], for the normal population, the method discussed above can be used by replacing σ^2 by s^2, where s^2 is a realization of S^2.

4. An Application

To illustrate the approaches developed in the previous sections, we use a data set from the Australian institute of sport. In paticular, we consider the body fat percentage (%Bfat) as dependent variable. The data were collected from 202 athletes at the Australian institute of sport and are described in Cook and Weisberg (1994). Besides a constant term we use information on four covariates: sex (=1, for female and =0, for male), red cell plasma (RCC), white cell plasma (WCC), plasma ferritin concentration (Ferr). These data, in multivariate situation, have been used previously in the context of skewed distribution, Azzalini and Capitanio [4] and Ferreira and Steel [9].

In previous sections, we let $y_i|\beta \sim N(x_i'\beta, \sigma^2)$, $\beta \sim SN(\zeta, \Omega, \alpha)$, thus by using this model if we observe skewness in distribution of population, this skewness belongs to prior distribution of mean parameter that shifted to population distribution, Bansal et al. ([5]).

As we have discussed in section 3.1, for unknown σ^2 case, Berger ([6]) had proposed a methodology for estimating σ^2 . His method has an straight linkage to the Bootstrap method. In consequence when we don't have any repeated samples, one way for estimating σ^2 can be resampling from origin sample, calculating s_i^2 for each samples and then estimating σ^2 with $\hat{\sigma^2} = \frac{1}{K}\sum_{i=1}^{K} s_i^2$. Using this method for our data set, with $K = 10000$, $\hat{\sigma^2} = 8.157$. Also by maximization of the marginal distribution estimates of hyper parameters for skew-normal scenario are:

$$\zeta = \begin{pmatrix} 13.104 \\ 1.598 \\ 0.305 \\ -0.898 \\ 0.008 \end{pmatrix}, \quad \alpha = \begin{pmatrix} 4.396 \\ 0.692 \\ 1.683 \\ 3.544 \\ -0.018 \end{pmatrix},$$

$$\Omega = \begin{pmatrix} 63.625 & -11.958 & -1.768 & 0.207 & 0.063 \\ & 2.247 & 0.332 & -0.039 & -0.012 \\ & & 0.049 & -0.006 & -0.002 \\ & & & 0.001 & 0.0002 \\ & & & & 0.623 \times 10^4 \end{pmatrix}.$$

These for normal scenario are:

$$\zeta = \begin{pmatrix} 16.031 \\ 1.072 \\ 0.204 \\ -0.800 \\ 0.011 \end{pmatrix},$$

$$\Omega = \begin{pmatrix} 42.056 & -5.732 & -2.248 & -2.484 & 0.002 \\ & 0.782 & 0.306 & 0.339 & -2.969 \times 10^{-3} \\ & & 0.120 & 0.132 & -1.177 \times 10^{-3} \\ & & & 0.147 & -1.287 \times 10^{-3} \\ & & & & 1.273 \times 10^{-5} \end{pmatrix}.$$

We obtain empirical Bayes estimation for these data by (6), the results for these methods are given in Table 1, where a Bootstrap approach is used to find the standard error of the parameter estimates. This table has two important results, (1) standard error of parameters in skew-normal scenario is smaller than those of normal one, (2) The important result about sex variable, this parameter is significant in skew-normal scenario while is not significant under normal one. Table 1 shows that the body fat for males is higher than that for females, also the more RCC (WCC) the more body fat and the more plasma ferritin concentration the more body fat.

Table 1. Empirical Bayes estimation of parameters using Sport data set.

	Skew-Normal Scenario		Normal Scenario	
parameter	Estimate	S.D.	Estimate	S.D.
constant	14.164	2.650	10.046	2.740
RCC	1.385	0.096	2.616	0.531
WCC	0.151	0.021	-0.018	0.106
Sex	-0.353	0.084	-0.013	0.485
Ferr	0.016	0.001	0.009	0.004

References

[1] A. Azzalini, A class of distribution which includes the normal ones, *Scand. J. Statist.*, 12 (1985), 171-178.

[2] A. Azzalini, The skew-normal distribution and related multivariate families, *Scand. J. Statist.*, (2005), 159-188.

[3] A. Azzalini and A. Capitanio, Statistical applications of the multivariate skew-normal distributions, *J. Roy. Statist. Soc. B.*, (1999), 579-602.

[4] A. Azzalini and A. Capitanio, Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *J. Roy. Statist. Soc. B.*, (2003), 367-389.

[5] K. Bansal and M. Maadooliat, Xiaowei Wang, Empirical Bayes and hierarchical Bayes estimation of skew-normal populations, *Commun. Statist. Theory and Methods*, 37 (2008), 1024-1037.

[6] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, NewYork, 1985.

[7] R. Cook and S. Weisberg, *An introduction to regression graphics*, John Wiley and sons, INC., 1994.

[8] R. D. l. Cruz, Bayesian non-linear regression models with skew-elliptical errors: Applications to the classification of longitudinal profiles, *Computational Statistics and Data Analysis*, 53 (2008), 436-449.

[9] S. Ferreira and F. J. Steel, A new class of skewed multivariate distributions with applications to regression analysis, *Statistica Sinica*, 17 (2007), 99-113.

[10] C. Flecher, P. Naveaua, and D. Allard, Estimating the closed skew-normal distribution parameters using weighted moments, *Statistics and Probability Letters*, 79 (2009), 1977-1984.

[11] G. Gonzlez-Faras, J. A. Domnguez-Molina, and A. K. Gupta, The closed skewnormal distribution. In: Genton M. G., ed. Skew-Elliptical Distributions and Their Applications: *A Journey Beyond Normality*, Boca Raton, FL: Chapman & Hall, (2004), 25-42.

[12] G. Gonzlez-Faras, J. A. Domnguez-Molina, and A. K. Gupta, Additive properties of skew-normal random vectors, *J. Statist. Plann.*, 126 (2004), 521-534.

[13] E. Lehmann, *Theory of Point Estimation*. Springer-Verlag, NewYork, 1998.

[14] S. Mukhopadhyay and B. Vidakovic, Efficiency of linear Bayes rules for a normal mean: skewed priors class, *The Statistician*, 44 (1995), 389-397.

[15] J. Rodrigues and H. Bolfarine, Bayesian inference for an extended simple regression measurement error model using skewed priors, *Bayesian Analysis*, 2 (2007), 349-364.

[16] S. K. Sahu, D. K. Dey, and M. Branco, A new class of multivariate skew distributions with applications to Bayesian regression models, *Canad. J. Statist*, 2 (2003), 129-150.

Mohsen Khounsiavash
Department of Statistics
Ph.D Student
Islamic Azad University-Science and Research Branch
Tehran, Iran
E-mail: siavash@qiau.ac.ir

Taban Baghfalaki
Department of Statistics
Ph.D Student
Shahid Beheshti University
Tehran, Iran
E-mail: t.baghfaliki@gmail.com

Mojtaba Ganjali
Department of Statistics
Professor
Shahid Beheshti University
Tehran, Iran
E-mail: m-ganjali@sbu.ac.ir