Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsapl/070301

Bayesian Analysis of the Epsilon Skew Exponential Power Distribution

Michael Weldensea * *and Hassan Elsalloukh*

Department of Mathematics and Statistics, University of Arkansas at Little Rock, 2801 South University Avenue, Little rock, AR 72204-1099, U.S.A.

Received: 10 Oct. 2019, Revised: 23 Dec. 2019, Accepted: 28 Dec. 2019
Published online: 1 Sep. 2020

Abstract: The Epsilon Skew Exponential Power Distribution (ESEP) that was introduced by Elsalloukh [1] is an asymmetric distribution used for modeling asymmetric data. The ESEP includes Normal, Laplace, Epsilon Skew Normal (ESN), and Epsilon Skew Laplace (ESL) as particular cases, [1]. In the present study, since the ESEP distribution encompasses members with skewed and symmetric distributions, we perform and investigate the Bayesian analysis of this distribution using the methods of latent variables and uniform scale mixture for implementing the most common MCMC algorithm known as Gibbs sampling. Furthermore, we develop the posterior distributions and the full conditional distributions of each parameter of the ESEP using Jeffrey's non-informative and informative priors for each parameter. Finally, we provide examples to show the fitting accuracy and strength of the ESEP distribution compared to other distributions used in literature.

Keywords: Epsilon Skew exponential, Jeffrey's prior, Deviance Information Criteria, Uniform scale mixture, Gibbs sampler

1 Introduction

The most commonly used continuous probability distribution is the normal (or Gaussian) distribution. This distribution was discovered by De Moivre who was an 18^{th} century statistician and consultant, but he could not find the mathematical expression of the normal distribution. However, Adrain(1808) and Gauss(1809) independently developed the mathematical formula for the normal distribution for fitting errors due to measurements. It is also believed that this same distribution had been discovered by Laplace in 1778. According to Lane [2], Laplace also derived the famous central limit theorem. Laplace also introduced a distribution known as Laplace and used in many applications including data with heavier tails than normal tails. For example, Easterling [3] and Hsu [4] used Laplace distribution for modeling errors due to measurements and position, respectively.
The Exponential Power family (EP) incorporated both the normal distribution and Laplace distribution as special cases. However, the EP distribution is symmetrical and therefore can not be used for modeling skewed data. In real world, data do not always follow a normal distribution. As a result many researchers have been developing a parametric family of distributions that possesses skewness and kurtosis. Azzalini [5] developed a distribution, the Skew-Normal Distribution, by combining a standard normal distribution and its density function and studied some of its properties. Mudholkar and Hutson [6] introduced an asymmetric density, the Epsilon Skew Normal distribution(ESN). This probability density function is defined by

$$f(x) = \frac{1}{\sqrt{2\Pi}\sigma} \begin{cases} exp(\frac{(x-\theta)^2}{2(1-\varepsilon)^2\sigma^2}) & : x \geq \theta \\ exp(\frac{-(x-\theta)^2}{2(1+\varepsilon)^2\sigma^2}) & : x < \theta\,. \end{cases} \tag{1}$$

Researchers have further dug into devising a family of asymmetric distributions which can be solvable analytically. These asymmetric distributions are useful in financial and econometric sectors for modeling data that do not follow a normal distribution. Boris Choy and Chan [7] developed a technique for conducting statistical modeling using the two

* Corresponding author e-mail: mgweldensea@ualr.edu

scale mixture representation that is normal and uniform scale mixtures and provided a uniform scale mixture representation for the Generalized T-student density that can be implemented easily in Gibbs sampling. Besag and Green [8] reviewed how the introduction of auxiliary variables simplify the complexity of the model and the problem of multimodelity in the application of Gibbs sampler method in statistical physics and discussed an early development of MCMC. Damlen [9] presented the use of an auxiliary (or latent) variables for sampling non-standard densities as an alternative approach of sampling for rejection-based methods, metropolis-hasting algorithm in Bayesian analysis of non-conjugate and hierarchical methods. This method is helpful for developing a full conditional distribution very easily and makes the calculation of posterior distribution much easier. Naranjo [10] studied Bayesian analysis of Skewed Exponential Power (SEP) and Asymmetric Exponential Power (AEP) families [11] and computed the full conditional distribution of each parameter for implementing Gibbs sampling using the idea of mixture of uniform distribution [7] and latent variables for sampling non-standard densities [9]. Meleki and Nematollahi [12] studied the Bayesian approach under the informative and non-informative priors by exploiting the latent variables and stochastic representation of Epsilon skew normal (ESN) distribution in constructing the augmented likelihood function. Smith [13] created an R package called Bayesian Output Analysis (BOA). It helps us in assessing convergence and posterior inference for MCMC output and also discussed the main difference between the frequentist and Bayesian modeling. Elsalloukh [1] introduced a new class of asymmetric distribution called the Epsilon-Skewed-Exponential Power distribution (ESEP). This distribution includes ESN, EP family, Normal, and Laplace distributions as special cases. So, in this study we explore the Bayesian analysis for ESEP distribution using the idea of scale mixture uniform and the latent variables adopted by [10]. The probability density function of ESEP is depicted as follows:

$$f(x)=\frac{\alpha}{2\sqrt{2}\sigma\Gamma(\frac{1}{\alpha})}\begin{cases} exp(\frac{-(x-\theta)^{\alpha}}{2^{\frac{\alpha}{2}}(1-\varepsilon)^{\alpha}\sigma^{\alpha}}) & : x\geq\theta \\ exp(\frac{-(\theta-x)^{\alpha}}{2^{\frac{\alpha}{2}}(1+\varepsilon)^{\alpha}\sigma^{\alpha}}) & : x<\theta. \end{cases} \tag{2}$$

The epsilon-skew-exponential power distribution denoted by X $\sim$ ESEP$(\theta,\sigma,\varepsilon,\alpha)$ is a unimodal distribution at θ with the probability $\eta=(1+\varepsilon)/2$, $1-\eta=(1-\varepsilon)/2$ below and above the mode respectively.
The remaining of this paper is outlined as follows. In Section 2, we present the scale mixture of ESEP distribution and some of its main properties. The Bayesian analysis approaches with informative and non informative priors, the derivation of posterior distribution of the model based on the Likelihood of ESEP distribution and the full conditional distribution of each parameters are provided in Sections 3. In Section 4, a simulation study and real data set is presented to illustrate the performance of proposed ESEP distribution and its parameter estimates. In Section 5, conclusion is given.

2 ESEP distribution

In this paper, we consider the ESEP distribution proposed by [1] and defined as, a random variable X has an ESEP distribution if there exist a shape parameter $\alpha>0$, a location parameter $\theta\in\Re$, a scale parameter $\sigma>0$, and a skewness parameter $-1<\varepsilon<1$ and its probability density function given by equation (2). Note that

a)When α = 1, equation (2) reduces to Epsilon-Skew-Laplace density function (ESL), which was defined, by [14] with a pdf

$$f(x)=\frac{1}{2\sqrt{2}\sigma}\begin{cases} exp(\frac{-(x-\theta)}{2^{\frac{1}{2}}(1-\varepsilon)\sigma}) & : x\geq\theta \\ exp(\frac{-(\theta-x)}{2^{\frac{1}{2}}(1+\varepsilon)\sigma}) & : x<\theta. \end{cases}$$

b)When α = 2, equation (2) reduces to Epsilon-Skew-Normal density function (ESN), which was defined by [6] wih a pdf give by (1).
c)When $\alpha=2,$ and $\varepsilon=0$, equation (2) reduces to Normal density function
d)When $\alpha=1,$ and $\varepsilon=0$, equation (2) reduces to Laplace density function.

The graphs below are the graph of the ESEP for $\varepsilon=0$ and $\varepsilon=-0.7$ respectively, keeping other variables fixed:

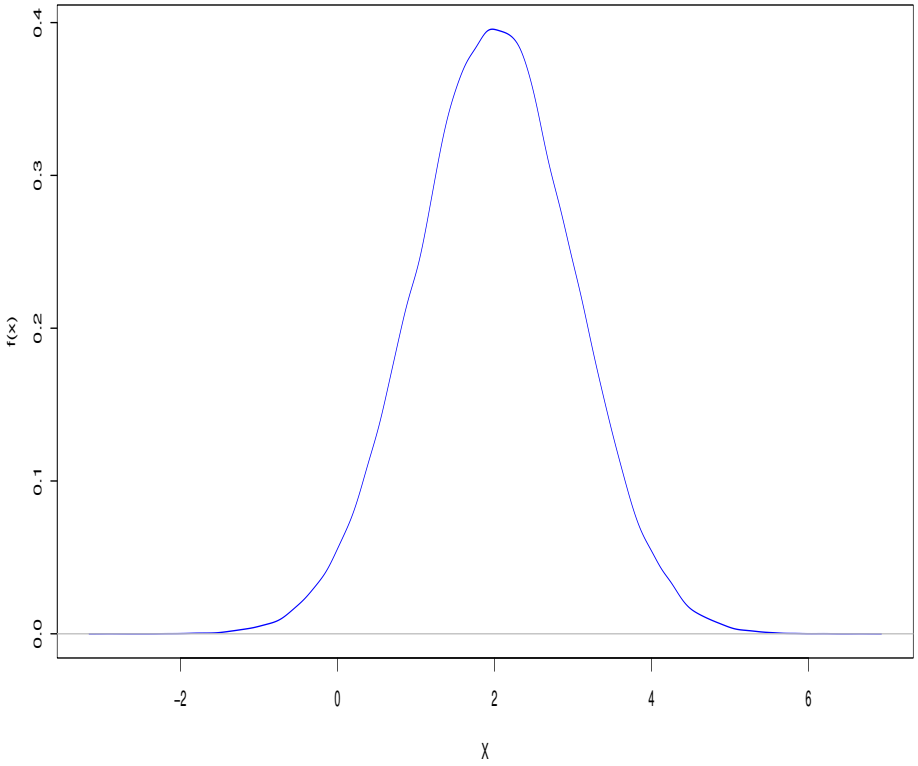

Fig. 1: epsilon = 0

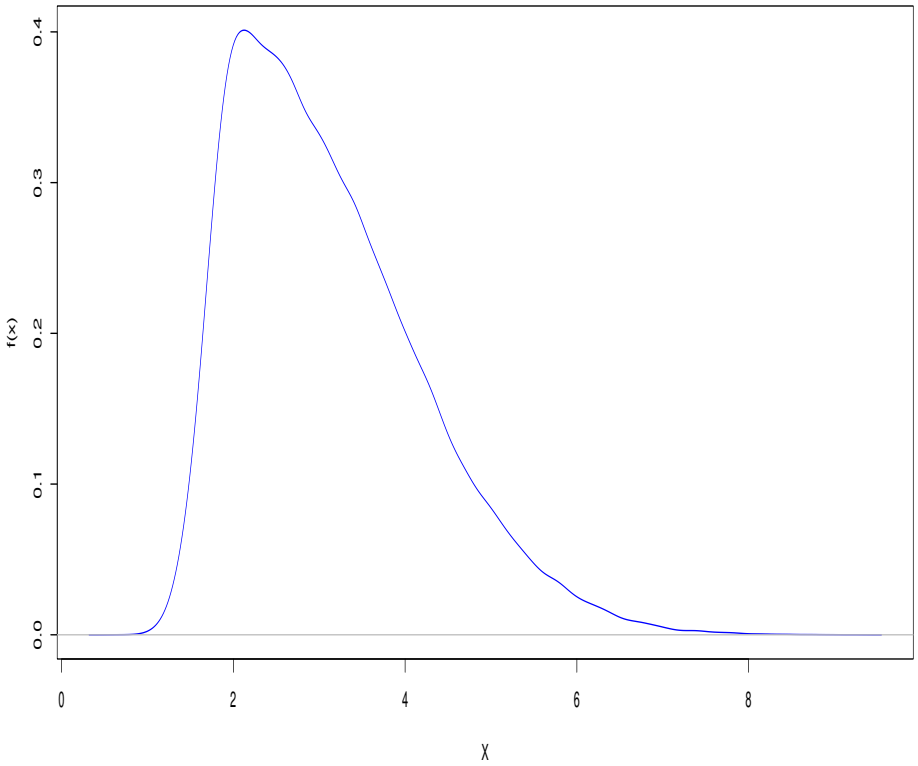

Fig. 2: epsilon = -0.7

Many researchers used the scale mixtures of normal and uniform distributions for representing several distributions and statistical models mainly for Bayesian analysis such as, Boris Choy and Chan [7] used a uniform scale mixture representation of Generalized T-distribution (GT) density, and Naranjo [11] used scale mixture uniform of Skew Exponential Power (SEP) and Asymmetric Exponential Power (AEP) distributions for performing Bayesian analysis. Qin [15] stated that if a random variable X has a uni-modal and skewed distribution with location θ, then it may be represented as a scale mixture uniform (SMU). Here, we exploit the mixture uniform representation of EP distribution defined by [15] for ESEP distribution.

2.1 Proposition 1

Let X be a random variable and $U1$ and $U2$ be auxiliary variables.

If $\mathrm{X}|U_1 = u_1 \sim \mathrm{U}(\theta, \theta + \sqrt{2}(1-\varepsilon)\sigma u_1^{\frac{1}{\alpha}})$, $u_1 \sim \mathrm{Ga}(1+\frac{1}{\alpha},1)$,

$\mathrm{X}|U_2 = u_2 \sim \mathrm{U}(\theta - \sqrt{2}(1+\varepsilon)\sigma u_2^{\frac{1}{\alpha}}, \theta)$, and $u_2 \sim \mathrm{Ga}(1+\frac{1}{\alpha},1)$, then X has an ESEP$(\theta, \sigma, \alpha, \varepsilon)$.

This proposition represents the uniform scale mixture of ESEP distribution and its proof is as follows :
We first start with

$\int$ f(x| u_1) f(u_1) $I(x \geq \theta)$ with probability of $\frac{1-\varepsilon}{2}$

$$
\begin{aligned}
&= \int \frac{1}{\sqrt{2}(1-\varepsilon)\sigma u_1^{\frac{1}{\alpha}}\Gamma(1+\frac{1}{\alpha})} u_1^{\frac{1}{\alpha}} exp(-u_1) I[x \geq \theta] du_1 \\
&= \int \frac{1}{\sqrt{2}(1-\varepsilon)\sigma\Gamma(1+\frac{1}{\alpha})} exp(-u_1) I[x \geq \theta] du_1 \\
&= \frac{1}{\sqrt{2}(1-\varepsilon)\sigma\Gamma(1+\frac{1}{\alpha})} \int exp(-u_1) I[x \geq \theta] du_1 \\
&= \frac{1}{\sqrt{2}(1-\varepsilon)\sigma\Gamma(1+\frac{1}{\alpha})} [-e^{-u_1}] I[x \geq \theta] \, for \, u_1 > \left[\frac{(X-\theta)}{\sqrt{2}(1-\varepsilon)\sigma}\right]^{\alpha}. \qquad (3)
\end{aligned}
$$

When substituting U_1 in(3), we get

$$
\int f(x|u_1) f(u_1) = \frac{1}{\sqrt{2}(1-\varepsilon)\sigma\Gamma(1+\frac{1}{\alpha})} e^{-\left[\frac{(X-\theta)}{\sqrt{2}(1-\varepsilon)\sigma}\right]^{\alpha}} I[x \geq \theta]. \qquad (4)
$$

Similarly, we integrate
$\int$ f(x| u_2) f(u_2) $I(x < \theta)$ with probability of $\frac{\varepsilon+1}{2}$

$$
\begin{aligned}
&= \int \frac{1}{\sqrt{2}(1+\varepsilon)\sigma u_2^{\frac{1}{\alpha}}\Gamma(1+\frac{1}{\alpha})} u_2^{\frac{1}{\alpha}} exp(-u_2) I[x < \theta] du_2 \\
&= \int \frac{1}{\sqrt{2}(1+\varepsilon)\sigma\Gamma(1+\frac{1}{\alpha})} exp(-u_2) I[x < \theta] du_1 \\
&= \frac{1}{\sqrt{2}(1+\varepsilon)\sigma\Gamma(1+\frac{1}{\alpha})} \int exp(-u_2) I[x < \theta] du_1 \\
&= \frac{1}{\sqrt{2}(1+\varepsilon)\sigma\Gamma(1+\frac{1}{\alpha})} - e^{-u_2} I[x < \theta] \, for \, u_2 > \left[\frac{(\theta-X)}{\sqrt{2}(1+\varepsilon)\sigma}\right]^{\alpha}. \qquad (5)
\end{aligned}
$$

Substituting U_1 in (5), we get

$$
\int f(x|u_2) f(u_2) = \frac{1}{\sqrt{2}(1+\varepsilon)\sigma\Gamma(1+\frac{1}{\alpha})} e^{-\left[\frac{(\theta-X)}{\sqrt{2}(1+\varepsilon)\sigma}\right]^{\alpha}} I[x < \theta]. \qquad (6)
$$

Combining (4) and (6), we get f(x) = $\frac{\varepsilon-1}{2}\int$ f(x| u_1) f(u_1) + $\frac{\varepsilon+1}{2}\int$ f(x| u_2) f(u_2), which is the Pdf of the distribution of ESEP($\mu,\sigma,\alpha,\varepsilon$).

3 Bayesian Analysis of ESEP

In this section, we use Proposition 1 for Bayesian approach in implementing the MCMC technique for ESEP distribution and develop the posterior distribution of a model where the ESEP distribution is considered for the likelihood. Let X be an ESEP random variable then the likelihood of a variable X and the latent vector of the mixing parameters of U_{1i},U_{2i} is given by

$$L(\theta,\sigma,\varepsilon,\alpha|X,U_1,U_2)\propto\prod_{i=1}^{n}\frac{\alpha}{2\sqrt{2}\sigma\Gamma(\frac{1}{\alpha})}\begin{cases}exp(-U_{1i})I[\theta<X_i\leq\theta+\sqrt{2}(1-\varepsilon)\sigma u_1^1/\alpha]\\ for: x\geq\theta\\ exp(-U_{2i})I[\theta-\sqrt{2}(1+\varepsilon)\sigma u_2^1/\alpha<X_i\leq\theta]\\ for: x<\theta.\end{cases}\tag{7}$$

Then the joint posterior distribution of the unobserved parameters θ, σ, ε, α, U_1, and U_2 is the combination of the likelihood function and the prior distribution of each parameter and is given by

$$K(\theta,\sigma,\varepsilon,\alpha|U_1,U_2)\propto\pi(\theta)\pi(\sigma)\pi(\varepsilon)\pi(\alpha)L(\theta,\sigma,\varepsilon,\alpha|X,U_1,U_2),\tag{8}$$

where $\pi(\theta),\pi(\sigma),\pi(\varepsilon)$, and $\pi(\alpha)$ are the prior distributions of θ, σ, ε, and α respectively.

3.1 Prior distribution

A prior probability distribution of an uncertain quantity P is the probability distribution that would state one's uncertainty about P before some evidence is taken into account. Even though there are different types of prior distributions, in this study we use only the non-informative and informative priors.

I) The non-informative Jeffrey's prior

In situations where there is no strong prior belief, it is advisable to use the non-informative prior along with the data for making inference. Sir Harold Jeffrey(1946,1961) defined an invariant non-informative prior distribution with respect to transformation of the parameters, as proportional to the square root of the determinant Fishers' Information matrix of the parameters. That is $\pi(\theta)\propto\sqrt{|I(\theta)|}$, where I($\theta$) is the Fisher information for θ [16].

The Fisher information matrix derived by Elsalloukh [17] for each parameters of ESEP distribution is presented as follows:

$$I(\theta,\sigma,\varepsilon,\alpha)=\begin{bmatrix}\frac{(\alpha-1)\Gamma(1-\frac{1}{\alpha})}{2\sigma^2(\Gamma(1+\frac{1}{\alpha}))} & 0 & \frac{-\alpha\varepsilon}{\sqrt{2}\sigma(\Gamma(1+\frac{1}{\alpha}))} & \frac{-1}{\sqrt{2}\sigma(\Gamma(1+\alpha)}\\ 0 & \frac{\alpha}{\sigma^2} & 0 & \frac{\varepsilon}{2}-\psi(1+\frac{1}{\alpha})\\ \frac{-\alpha\varepsilon}{\sqrt{2}\sigma(\Gamma(1+\frac{1}{\alpha}))} & 0 & \frac{1+\alpha}{\varepsilon^2-1} & 0\\ \frac{-1}{\sqrt{2}\sigma(\Gamma(1+\alpha)} & \frac{\varepsilon}{2}-\psi(1+\frac{1}{\alpha}) & 0 & \frac{4(\alpha^2+\alpha)\psi(1+\frac{1}{\alpha})+(2\alpha+1)\psi'(1+\frac{1}{\alpha})+K}{\alpha^5}\end{bmatrix},$$

where $K=\psi(1+\frac{1}{\alpha})^2+2\alpha^2$ and $\psi(x)$ is a digamma function given by $\psi(x)=\frac{\Gamma'(x)}{\Gamma(x)}$.

From the Fisher information matrix, the corresponding non-informative prior distribution of the parameters is given by

$$\pi(\theta)\propto 1,\pi(\sigma)\propto\frac{1}{\sigma},\pi(\varepsilon)\propto(\eta)^{-1/2}(1-\eta)^{-1/2},\pi(\alpha)\propto\frac{1}{\alpha}.\tag{9}$$

II) The informative priors

In situations where we know the prior distribution of the parameter of our interest, it is good to use the informative prior distribution for making inference. Here, we choose the informative prior distribution for each parameter in such a way that the posterior distribution is conjugate to the prior distribution. Thus, the informative priors that we use for Bayesian analysis of ESEP distribution are given as follows:

$$\pi(\theta)\propto N(a,b),\pi(\sigma)\propto G(\alpha,\beta),\pi(\varepsilon)\propto Beta(\alpha,\beta),\pi(\alpha)\propto G(\alpha,\beta),\tag{10}$$

where G denotes a gamma distribution. There are different methods of choosing the value of the hyper-parameter. We choose the value of the hyper-parameter that assures the behavior of the posteriors.

3.2 The Full Conditional Distribution of the Parameters and Unobserved Variables

Generally from (8), we develop the full distribution of each parameter by eliminating all the terms that are independent of individual parameters as follows:

- The full conditional distribution of θ given the other variables, can be obtained from (8) by eliminating all the terms that are independent of θ, that is

$$k(\theta) \propto \pi(\theta) I[\underline{\theta}, \overline{\theta}], \tag{11}$$

where $\underline{\theta} = Max\left\{ \max_{u_{1i}>0} \left\{X_i - \sqrt{2}(1-\varepsilon)\sigma u_2^1/\alpha\right\}, \max_{u_{1i}>0} \left\{X_i\right\} \right\}$,

and $\overline{\theta} = Min\left\{ \min_{u_{2i}>0} \left\{X_i + \sqrt{2}(1+\varepsilon)\sigma u_2^1/\alpha\right\}, \min_{u_{2i}>0} \left\{X_i\right\} \right\}$,

- The full conditional distribution of ε given the other variables, can be obtained from (8) by eliminating all the terms that are independent of ε, that is

$$k(\varepsilon) \propto \pi(\varepsilon) I[\underline{\varepsilon}, \overline{\varepsilon}], \tag{12}$$

where $\underline{\varepsilon} = Max\left\{ -1, \max_{u_2>0} \left\{ \frac{(\theta - X_i)}{\sqrt{2}\sigma u_{2i}^1/\alpha} - 1 \right\} \right\}$

and $\overline{\varepsilon} = Min\left\{ \max_{u_1>0} \left\{ 1 - \frac{(X_i - \theta)}{\sqrt{2}\sigma u_{1i}^1/\alpha}, 1 \right\} \right\}$,

- The full conditional distribution of σ given the other variables, can be obtained from (8) by eliminating all the terms that are independent of sigma

$\sigma > \left\{ \frac{(X_i-\theta)}{\sqrt{2}u_{1i}^{\frac{1}{\alpha}}(1-\varepsilon)} \right\}$ and $\sigma > \left\{ \frac{-(X_i-\theta)}{\sqrt{2}u_{2i}^{\frac{1}{\alpha}}(1+\varepsilon)} \right\}$, then we take the maximum of these two as $\underline{\sigma}$ that is $\underline{\sigma}$ = Max ($\sigma > \left\{ \frac{(X_i-\theta)}{\sqrt{2}u_{1i}^{\frac{1}{\alpha}}(1-\varepsilon)} \right\}$, $\sigma > \left\{ \frac{-(X_i-\theta)}{\sqrt{2}u_{2i}^{\frac{1}{\alpha}}(1+\varepsilon)} \right\}$) , then the full distribution of sigma is

$$k(\sigma) \propto \pi(\sigma) \frac{1}{\sigma^n} I\{\sigma > \underline{\sigma}\}, \tag{13}$$

- The full conditional distribution of α given the other variables, can be obtained from (8) by eliminating all the terms that are independent of α, that is

$$k(\alpha) \propto \frac{\pi(\alpha)}{(\Gamma(1+\frac{1}{\alpha}))^n} I[\alpha < \underline{\alpha}], \tag{14}$$

where $\underline{\alpha} = \{\ \alpha : \frac{(X_i-\theta)}{\sigma\sqrt{2}(1-\varepsilon)} < u_{1i}^{\frac{1}{\alpha}}, \frac{(\theta-X_i)}{\sigma\sqrt{2}(1+\varepsilon)} < u_{2i}^{\frac{1}{\alpha}}\ \}$,

- Finally the joint full conditional distribution of U_{1i} and U_{2i} can be obtained from (8) by eliminating all the terms that are independent of U_{1i} and U_{2i}, that is

$$\begin{aligned} k(U_{1i}, U_{2i}) \propto & Exp(-U_{1i}) I\left[U_{1i} > \left[\frac{(X_i-\theta)}{\sqrt{2}(1-\varepsilon)\sigma}\right]^{\alpha}\right] \\ & + Exp(-U_{2i}) I\left[U_{2i} > \left[\frac{(\theta-X_i)}{\sqrt{2}(1+\varepsilon)\sigma}\right]^{\alpha}\right]. \end{aligned} \tag{15}$$

When the priors of the posterior distribution of ESEP parameters are replaced by (9), non-informative prior distributions, (11), (12), and (13) become a standard uniform distribution and similarly when the priors of the posterior distribution of ESEP parameters are replaced by (10), informative prior distributions, (11), (12), and (13) become a truncated normal, truncated beta, and truncated gamma. Since the full conditional distributions of (11), (12), and (13) are standard known distributions, we can generate data from them. However, the full conditional distribution of (14) is not standard, so we generate data from it using the Metropolis-Hasting algorithm.

4 Numerical Studies

In this section, we present a simulation study in order to show the performance of ESEP distribution and its Bayesian analysis. We perform all our code and program on R version 3.4.3 and we use different R packages for carrying out our tasks. In Bayesian analysis, using uniform scale mixture, we generate data sets from ESEP with the parameter values $\theta = 1, \sigma = 2, \varepsilon = 0.7, , \alpha = 1$, and sample size n = 400. Then using the Gibbs sampler and MH algorithm (MCMC technique), a total of 100,000 iterations were generated from the posterior distribution of the parameters and a burn-in of 20,000 is considered for the two types of prior distributions.
Non-informative priors are given by $\pi(\theta) \propto 1, \pi(\sigma) \propto \frac{1}{\sigma}, \pi(\varepsilon) \propto (\eta)^{-1/2}(1-\eta)^{-1/2}, \pi(\alpha) \propto \frac{1}{\alpha}$, and informative priors are given by $\pi(\theta) \propto N(1,15), \pi(\sigma) \propto G(170,150), \pi(\varepsilon) \propto Beta(7,3), and\, \pi(\alpha) \propto G(2,3)$.

4.1 Convergence

After we simulated observations from ESEP distribution, we checked for the convergence of the MCMC chain using different initial values and chains converged to the stationary distribution. Before that we performed all the necessary procedures for getting the summary statistics and the convergence test. We used the coda and BOA packages in R for analysis and convergence diagnosis.
The summary statistics and the convergence diagnostics for both informative and non-informative priors are depicted below:
SUMMARY STATISTICS:
The summary statistics that we obtain from the posterior sample of the parameters for both informative and non-informative prior are depicted in Tables 1 and 2. The estimates are close to true values and their posterior standard deviations are small. As the batch ACF is small for all parameters, the autocorrelation between them is small. There are many ways to asses convergence visually and numerically. Here we are going to discuss some of the diagnostics for convergence.

Table 1: The True Value and the Estimate Posterior Mean of the Parameters for Informative Prior.

Informative Prior							
Parameters	True	Mean	SD	Naive SE	MC Error	Batch SE	Batch ACF
θ	1	1.0146	0.0119	3.779e-05	1.046e-04	9.779e-04	0.0291
σ	2	1.99618	0.0061	1.9498e-05	3.701e-05	3.7789e-05	0.02993
ε	0.7	0.6972	0.0229	7.267e-05	8.6838e-04	4.5636e-04	0.6172
α	1	1.2011	0.2936	9.2864e-04	8.576e-04	9.449e-04	0.0352

Table 2: The True Value and the Estimate Posterior Mean of the Parameters for Non-Informative Prior.

Non-Informative Prior							
Parameters	True	Mean	SD	Naive SE	MC Error	Batch SE	Batch ACF
θ	1	0.9846	0.01475	4.666e-057	1.264e-04	1.1861e-04	0.0553
σ	2	2.2356 8	0.0068	2.1396e-05	5.263e-05	4.234e-05	0.00273
ε	0.7	0.7092	0.0253	8.0229e-05	9.0733e-04	4.876e-04	0.5507
α	1	1.1986	0.2918	9.2277e-04	8.248e-04	9.049e-04	0.03922

TRACE PLOT
Another way of checking for convergence is using a trace plot. As we can see in Figure 3 and Figure 4, the chain is exploring the distribution by traversing to areas where its density is very low and the distribution of points is not changing as the chain progresses. This indicates that our chain mixes good for each parameter in both informative and non-informative priors. However, performing trace plot only does not guarantee that the chain is converged to the stationary. So, we provide another method of assessing convergence that is Gelman and Rubin diagnostic below.

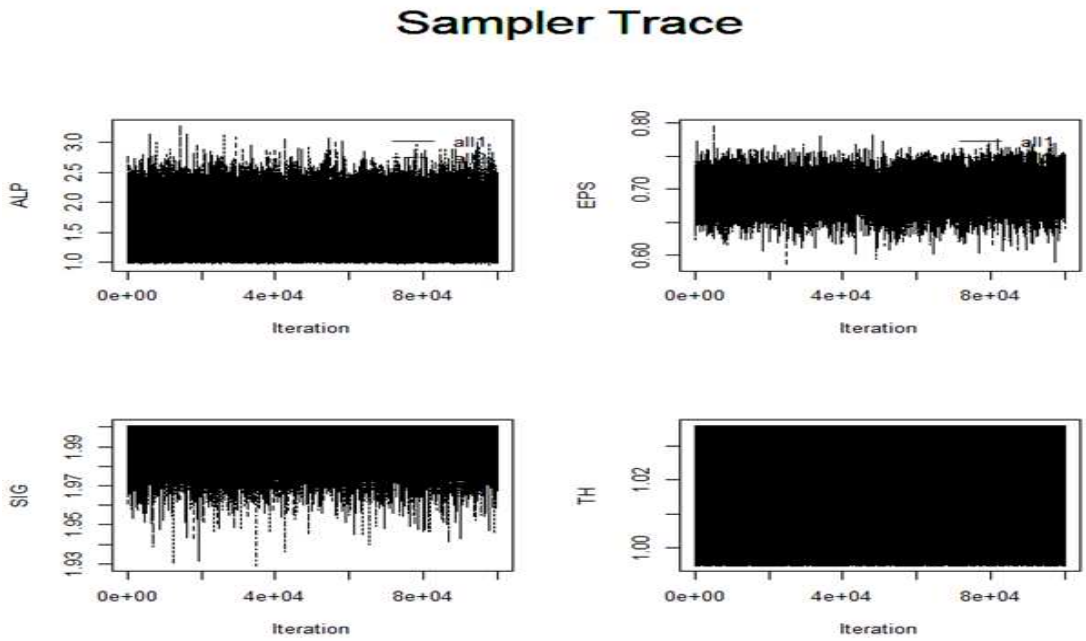

Fig. 3: *Trace Plot of Each of the Parameters for Informative Prior.*

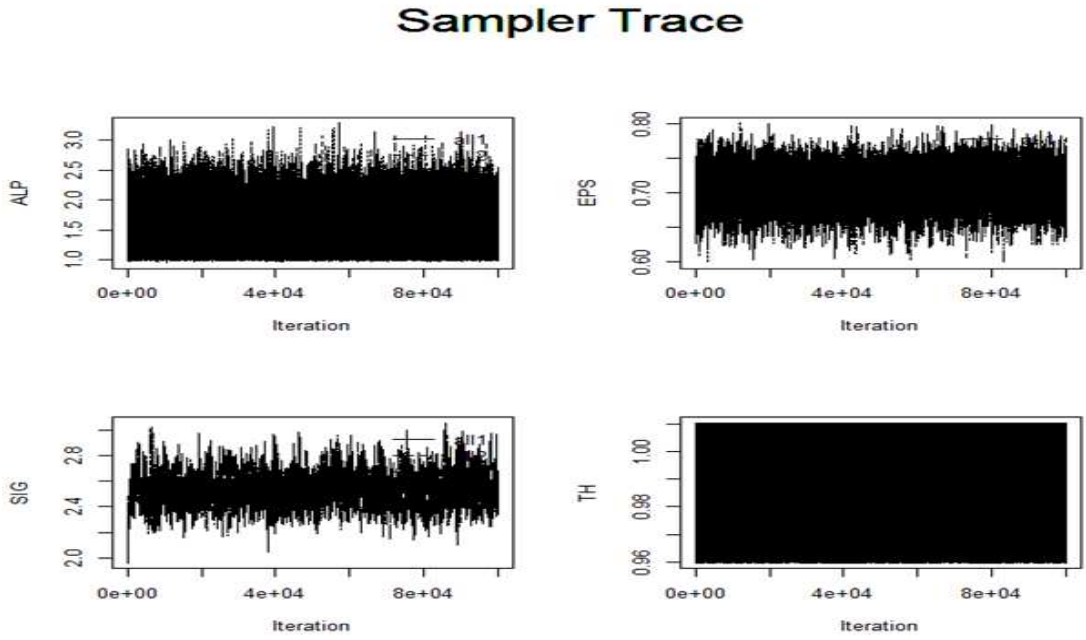

Fig. 4: *Trace Plot of Each of the Parameters for Non-Informative Prior.*

The most commonly used for assessing convergence is the Gelman and Rubin diagnostic. According to Gelman [18], if the potential reduction factor ($\widehat{R}$) for the estimands are near one (especially $\widehat{R}$ <1.1 or <1.2), there is potential

convergence of the MCMC chain. For informative prior, the $\widehat{R}$ for θ, σ, ε, and α is 1.0003, 0.9999923 ,1.00022, and 0.999999 respectively and for non-informative prior, the $\widehat{R}$ for θ, σ, ε, and α is 1.0000102 1.0100390 1.0006988 0.9999916 respectively. In both case $\widehat{R}$ is less than 1.1. Moreover we can see visually how the $\widehat{R}$ changes through the iteration for both informative and non-informative in Figure 5 and Figure 6, respectively. These plots show us the chain reduction is stable over time.

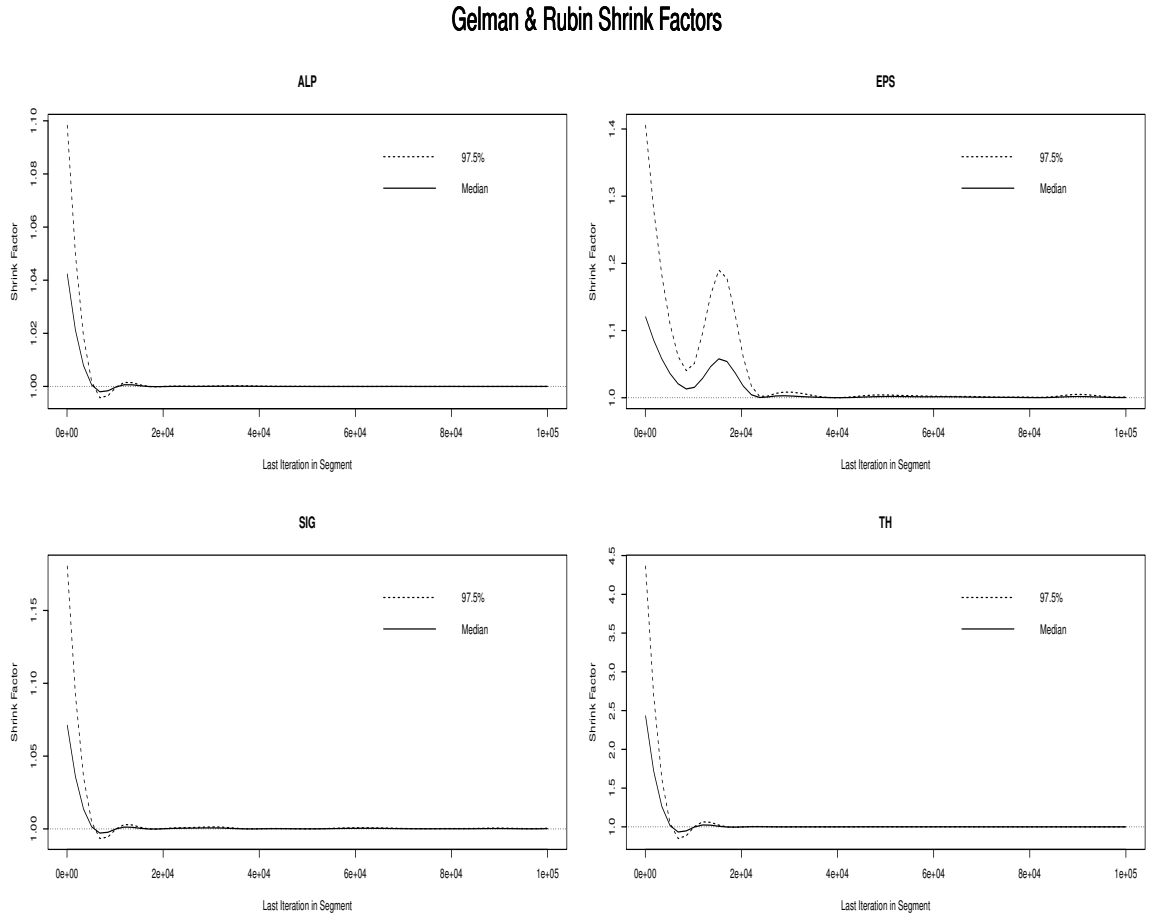

Fig. 5: *Gelman and Rubin of Each of the Parameters for Informative Prior.*

Brooks and Gelman [19] show graphically how the convergence of the chains are attained. They plot the Potential scale reduction factor $\hat{R}_c(k)$ against the number of iterations (K) and also they introduce another important diagnostic by plotting the two scale factor $V^{\frac{1}{2}}(K)$ and $W^{\frac{1}{2}}(K)$ as a function of the number of iterations K together on the same plot. Approximate convergence is not obtained until both lines are stabilized. Once the lines are stabilized at the same value, we say that the Markov chains are converged. For informative prior distribution, as we can see in Figure 5, the plot of the shrink factor against of the number of iterations for each parameter of our interest. This indicates that the convergence of the chains are attained after 10,000 iterations for the α, σ, θ and for ε after 20,000. Similarly, for non-informative prior distribution, as we can see in Figure 6 the approximate convergence of the Markov chains are obtained after around 5,000 iterations for ε, θ, and α respectively. Whereas for σ after 40,000 iterations.

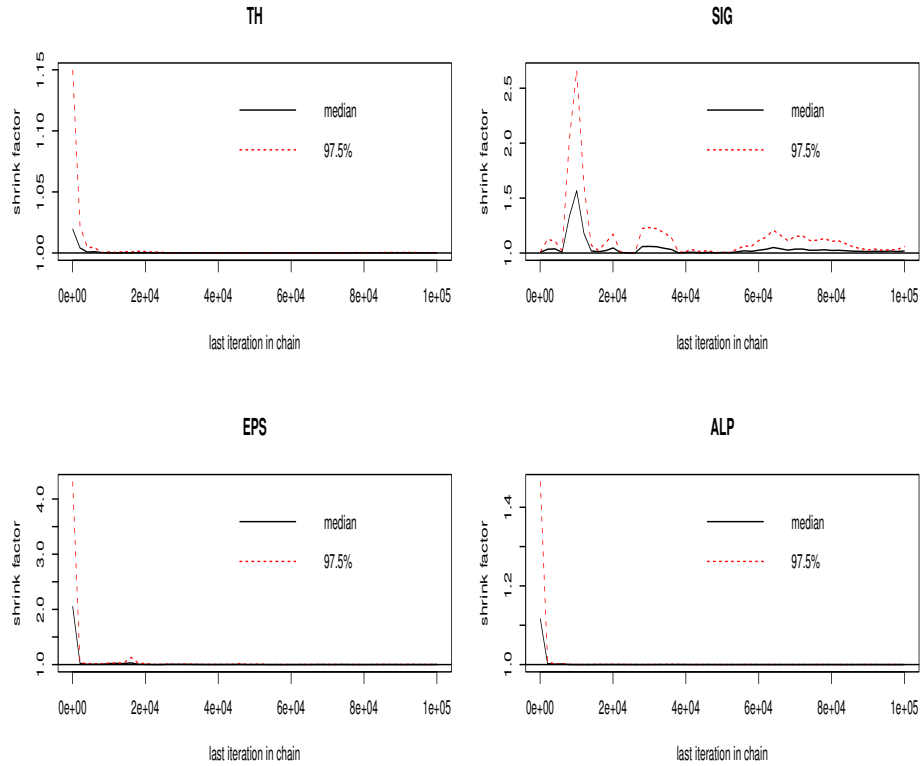

Fig. 6: *Gelman and Rubin of Each of the Parameters for Non-Informative Prior.*

4.2 Model Comparison

Here, we show the flexibility and the model strength of ESEP distribution for fitting data sets. The comparison technique used for comparing the ESEP distribution with other skewed distributions in literature is the Deviance Information Criterion (DIC).
The real data is taken from a data set collected at the Australian Institute of Sport [20]. From the several variables recorded on the 202 athletes, we use the height of 100 females for comparison purposes as other researchers used. Naranjo [10] used these data sets for comparing the skew exponential power distribution (SEP) with others. Table 3 presents the DICs, the posterior mean of Deviance $\bar{D}$ and the effective number of parameters $\hat{pD}$ of some distributions fitted to the data. From the table we observe that the DIC for ESEP distribution is small comparing to the other distribution. So, the model with small DIC value is a best model fit. As a result the ESEP distribution performs better than the other distributions. The reason can be, ESEP can be adjusted very easily to symmetric and asymmetric distribution by varying the value of skewness and shape parameters. Figure 2.5 shows the histograms and the predictive density of the estimated distribution of ESEP.

Table 3: Estimated DICs for Different Models.

The DIC of Skewed distribution			
Model	DIC	$\bar{D}$	$\hat{pD}$
Skew Normal	706.894	703.740	3.154
Skew t (v r.v)	704.673	701.568	3.105
Skew Laplace	705.698	701.735	3.962
SEP	704.340	700.152	4.187
ESEP	703.890	699.970	3.92

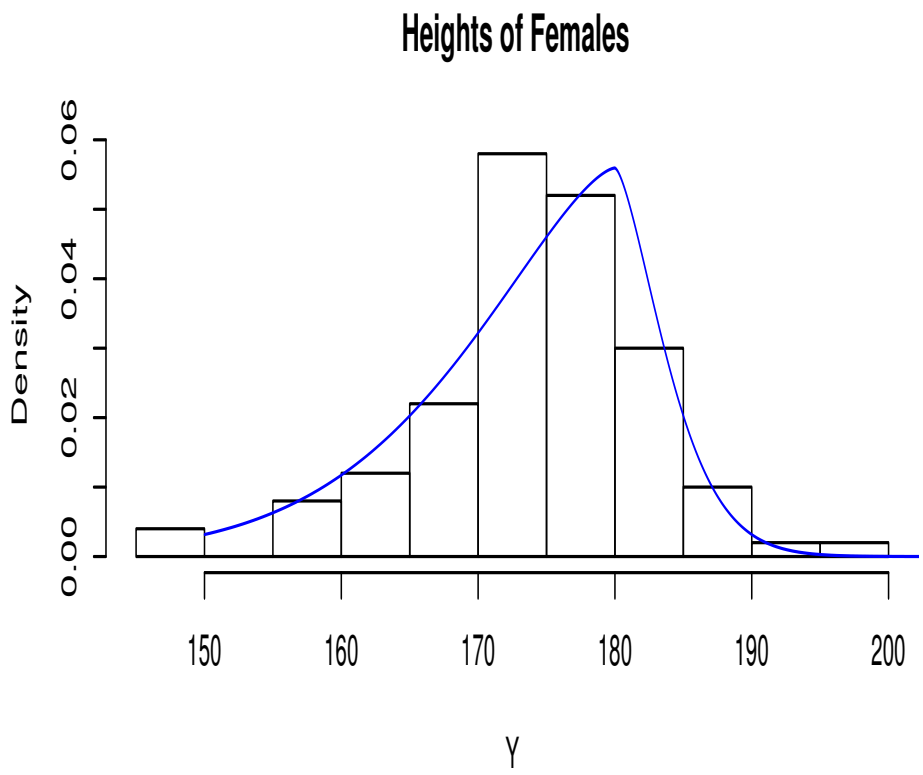

Fig. 7: *Histograms and Predictive Density of ESEP*

5 Conclusion

The Epsilon Skew Exponential Power distribution has been analyzed from Bayesian perspective view. The ESEP distribution encompasses the normal, skew normal, Laplace, and skew Laplace distributions as a particular case. By the technique of scale mixture of uniform, Gibbs sampling and Metropolis-Hasting algorithm, Bayesian analysis and inference are performed. The ESEP distribution can provide a flexible fit to both symmetric and asymmetric data because the skewness parameter and the shape parameter can be adjusted to fit to both symmetric and asymmetric data by varying their values simultaneously. Therefore, we can use ESEP distribution for fitting symmetric and asymmetric data.

Acknowledgement

The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article

References

[1] Elsalloukh, H., Guardiola, J. H., & Young, M. (2005). The epsilon-skew exponential power distribution family. Far East Journal of Theoretical Statistics, 17(1), 97.

[2] Lane, D. (2003). Online statistics education: A multimedia course of study (pp. 1317-1320). Association for the Advancement of Computing in Education (AACE).

[3] Easterling, R. G. (1978). Exponential responses with double exponential measurement error-A model for steam generator inspection. In Proceedings of the DOE Statistical Symposium, US Department of Energy (pp. 90-110).
[4] Hsu, D. A. (1979). Long-tailed distributions for position errors in navigation. Journal of the Royal Statistical Society: Series C (Applied Statistics), 28(1), 62-72.
[5] Azzalini, A. (1985). A class of distributions which includes the normal ones. Scandinavian journal of statistics, 171-178.
[6] Mudholkar, G. S., & Hutson, A. D. (2000). The epsilon–skew–normal distribution for analyzing near-normal data. Journal of Statistical Planning and Inference, 83(2), 291-309.
[7] Boris Choy, S. T., & Chan, J. S. (2008). Scale mixtures distributions in statistical modelling. Australian & New Zealand Journal of Statistics, 50(2), 135-146.
[8] Besag, J., & Green, P. J. (1993). Spatial statistics and Bayesian computation. Journal of the Royal Statistical Society: Series B (Methodological), 55(1), 25-37.
[9] Damlen, P., Wakefield, J., & Walker, S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(2), 331-344.
[10] Naranjo, L., Pérez, C. J., & Martín, J. (2012). Bayesian analysis of a skewed exponential power distribution. In Proceedings of COMPSTAT (pp. 641-652).
[11] Naranjo, L., Pérez, C. J., & Martín, J. (2015). Bayesian analysis of some models that use the asymmetric exponential power distribution. Statistics and Computing, 25(3), 497-514.
[12] Maleki, M., & Nematollahi, A. R. (2017). Bayesian approach to epsilon-skew-normal family. Communications in Statistics-Theory and Methods, 46(15), 7546-7561.
[13] Smith, B. J. (2007). boa: an R package for MCMC output convergence assessment and posterior inference. Journal of Statistical Software, 21(11), 1-37.
[14] Elsalloukh, H. (2008, August). The epsilon-skew Laplace distribution. In Proceedings of the JSM American Statistical Association Conference, Denver, CO, USA (pp. 3-7).
[15] Zhaohui S Qin, Paul Damien, and Stephen Walker. Scale mixture models with applications to Bayesian inference. In AIP Conference Proceedings (Vol. 690, No. 1, pp. 394-395). AIP.
[16] George EP Box and George C Tiao, Bayesian inference in statistical analysis, **40**, John Wiley & Sons, (2011).
[17] Elsalloukh, H. (2004). The epsilon-skew-exponential power distribution (Doctoral dissertation, Baylor University).
[18] Gelman, A., Carlin, J. B., Stren, H. S., & Dunson, D. B. (2013). Vehtari A., and DB Rubin. Bayesian Data Analysis.
[19] Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. Journal of computational and graphical statistics, 7(4), 434-455.
[20] Telford, Richard D., and Ross B. Cunningham. "Sex, sport, and body-size dependency of hematology in highly trained athletes." Medicine and science in sports and exercise 23.7 (1991): 788-79

Michael Weldensea is a Senior Statistician, Arkansas Department of Health, 4815 W. Markham, Little Rock, AR 72205

Hassan Elsalloukh is a Professor of Statistics, Department of Mathematics and Statistics, University of Arkansas at Little Rock, 2801 South University Avenue, ETAS 479, Little Rock, Arkansas, 72204