PRESIDENCY UNIVERSITY
KOLKATA

DISSERTATION THESIS ON

A HEURISTIC APPROACH TO FIT SHAPE PARAMETER OF SKEW-NORMAL DISTRIBUTION BASED ON A REGRESSION MODEL

SUBMITTED BY:

Ritwick Mondal

ROLL NO: STAT016

SEM-4; PG-2

REG. NO. : 18214110010

UNDER THE SUPERVISION OF:

Dr.Radhakanta Das

CONTENTS

Acknowledgements

I would take this opportunity to express my sincere gratitude and thanks to my advisor **Dr.Radhakanta Das,Presidency University, Kolkata** for giving me the opportunity with this wonderful project "**A Heuristic Approach to Fit Shape Parameter of Skew-Normal Distribution Based on a Regression Model**" .

I wish to thank the entire faculty of the Department of Statistics for enriching me with their vast knowledge in various fields of Statistics. I would like to thank the entire staff of the Department of Statistics who were always ready with their help.

The project has helped me to have a more detailed knowledge over interesting topic like estimation of shape parameter of skew normal distribution and topic related to this.It also helped me in doing a lot of research and I came to know about so many new things in finalizing this project within the limited time.It helped me increase my knowledge and skills.

Abstract-

The normal distribution is symmetric and enjoys many important properties. That is why it is widely used in practice.Asymmetry in data is a situation where the normality assumption is not valid.

In many practical applications it has been observed that real data sets are not symmetric,not unimodal.They exhibit some skewness, therefore do not conform exactly to the normal distribution, which is popular and easy to be handled.

Azzalini (**1985**) introduced a new class of distributions named the **skew normal distribution**,which is mathematically tractable and includes the normal distribution as a special case with skewness parameter being zero.The skew normal distribution family is well known for modeling and analyzing **skewed data**.

The idea behind skew normal distribution is that it incorporates a new parameter regulating shape and skewness on the symmetric normal distribution.It has three parameters: location, scale and shape.SN distribution is specifically useful for data sets having distribution that exhibits skewness other than zero.

Key Words-

Method of moment estimator (MME),Maximum Likelihood Estimator (MLE),Q-Penalty,Simulations.

1.Introduction-

Our main objective in this project is to give a a heuristic approach to Fit Shape parameter of Skew-Normal Distribution Based on a Regression Model .Therefore, it is very important to estimate the parameters of the skew normal distribution from both an analytical and a practical point of view.In most of the books I've seen the estimation procedure of location and scale parameter provided the skew/shape parameter is known.

So it is very important to estimate the skew parameter other wise we can't estimate the other two parameter when skewness parameter will be unknown to us.In this project,I'll be suggesting the estimation procedure of shape parameter of skew normal distribution. In the estimation procedure,I use the **method of moment** (MM) and **maximum likelihood** (ML) method based on the maximization of the likelihood function with respect to parameters of interest.

There will be two cases in the estimation problem of shape parameter:location,scale parameter is known or unknown.We'll consider both of the cases in this project.We'll see that Finding maximum likelihood estimator of shape parameter is non trivial due to its **non linear form** and it that situation we can't get any mathematical formula for this.

In the first part,we first see some important properties of the Skew-Normal Distribution which will help us in the estimation part and then based on a regression model We'll fit its shape parameter.

In the later part of this project we'll work with some simmulated data .

1.1 Historical Background:

- Azzalini $(1985, 1986)$, who studied many properties of SND :A Class of Distribution which includes the Normal Ones, Scan. J. Statist., $12, 171-178$.
- Henze (1986) :A probabilistic Representation of the Skew-Normal Distribution, Scand J. Stat., $13, 271-275$.
- Gupta and Brown (2001): Reliability Studies of the Skew Normal Distribution and Its Application to a Strength-Stress Model, Commun. Statist.-Theory Meth., $11, 2427-2445$.
- Gupta, Nguyen and Sanqui (2004): Characterization of Skew-Normal Distribution, AISM, $56, 351-360$.
- Arnold and Lin (2004): Characterizations of the Skew-normal and Generalized Chi Distributions, Sankhya: The Indian Journal of Stat., $66, 593-606$.
- Figueiredo and Gomes (2013) : The Skew-Normal Distribution in SPC, REVSTAT Statistical Journal, $11, 83-104$.
- Ngunkeng and Ning (2014) : Information Approach for the Change-Point Detection in the Skew Normal Distribution and Its Applications, Sequential Analysis,$33, 475-490$.

1.2 Basic definition and Some Important properties:

Let Y is random variable which follows Skew Normal Distribution denoted by $SND(\mu, \sigma, \lambda)$ and its pdf is given by
$f_Y(y) = \frac{2}{\sigma}\phi(\frac{(y-\mu)}{\sigma})\Phi(\frac{\lambda}{\sigma}(y-\mu)), y \in \mathbb{R}$

where,

$$\mu = \text{Location Parameter}, \mu \in \mathbb{R}$$

$$\sigma = \text{Scale Parameter}, \sigma > 0$$

$$\lambda = \text{Shape Parameter}, \lambda \in \mathbb{R}$$

$$\Phi(.) = \text{Standard Normal C.D.F}$$

$$\phi(.) = \text{Standard Normal P.D.F}$$

- If $Y \sim SND(\mu, \sigma, \lambda)$ then , $\frac{(Y-\mu)}{\sigma} \sim SND(0, 1, \lambda)$
- If $Y \sim SND(\mu, \sigma, \lambda)$ then , $-Y \sim SND(-\mu, \sigma, -\lambda)$
- The skewness of the distribution increases as the value of λ increases in its absolute value.
- If $\lambda > 0 \Rightarrow SND(\mu, \sigma, \lambda)$ is **positively skewed**, $\lambda < 0 \Rightarrow SND(\mu, \sigma, \lambda)$ is **negatively skewed.**
- If $\lambda = 0$, the distribution becomes the normal distribution i.e,$N(\mu, \sigma^2)$.
- If $Y \sim SND(0, 1, \lambda)$ (Standard Skew Normal) then $Y^2 \sim \chi^2_1$
- M.G.F of Standard Skew Normal distribution is given by $M_Y(t) = 2.e^{t^2/2}\Phi(\delta t)$ where $\delta = \frac{\lambda}{1+\lambda^2}$
- Mean and Variance of Y is given by, $E(Y) = \mu + \sigma\sqrt{\frac{2}{\pi}}\frac{\lambda}{\sqrt{1+\lambda^2}}, Var(Y) = \sigma^2(1 - \frac{2}{\pi}\frac{\lambda^2}{1+\lambda^2})$
- If U, V iid $\sim N(0, 1)$, then $(\frac{\lambda}{\sqrt{1+\lambda^2}})|U| + (\frac{1}{\sqrt{1+\lambda^2}})V \sim SND(\mu, \sigma, \lambda)$ where $U, V \sim N(0, 1)$ independently.

Let us plot the density function of SKN distribution for different values of λ:

```
package 'fGarch' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\LENOVO\AppData\Local\Temp\RtmpIxfpOc\downloaded_packages
```

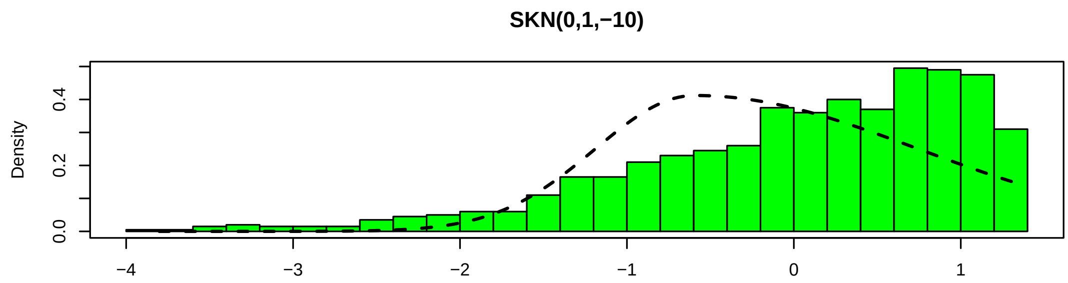


Fig.1.1:SKN(0,1,−10)

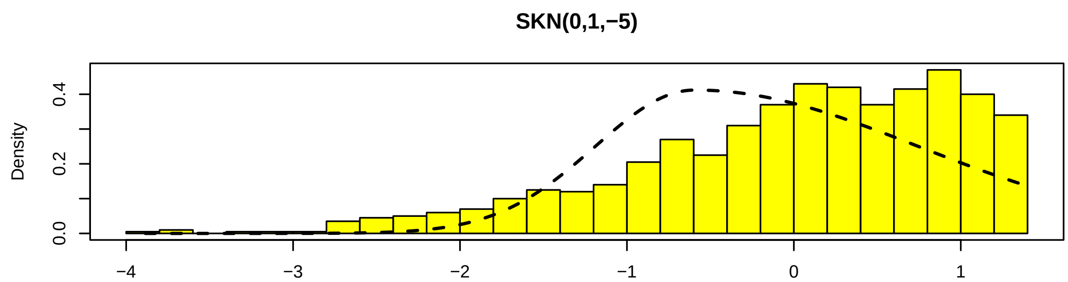


Fig.1.2:SKN(0,1,−5)

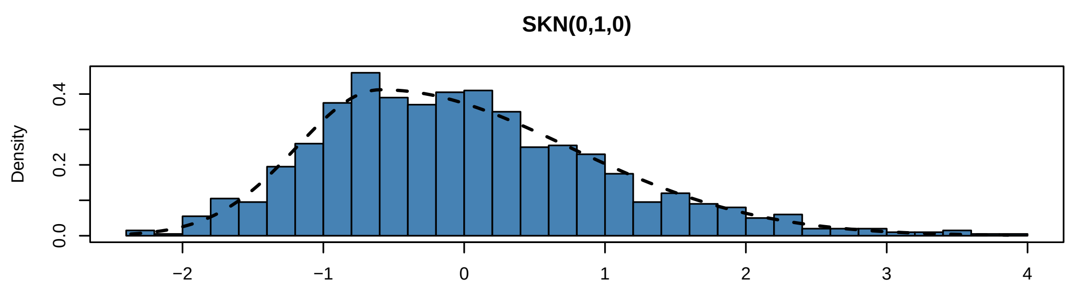


Fig.1.3:SKN(0,1,0)

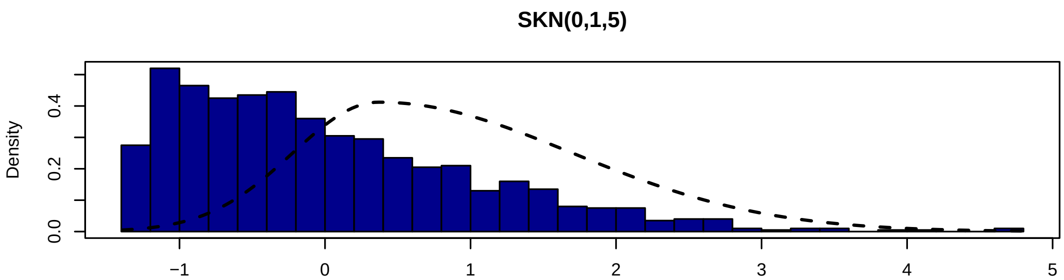


Fig.1.4:SKN(0,1,5)

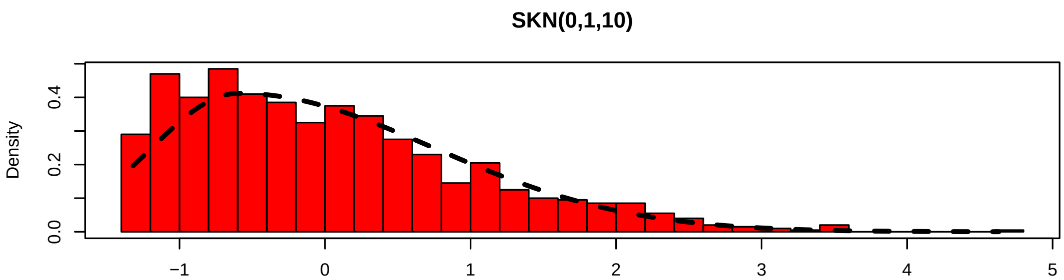


Fig.1.5:SKN(0,1,10)

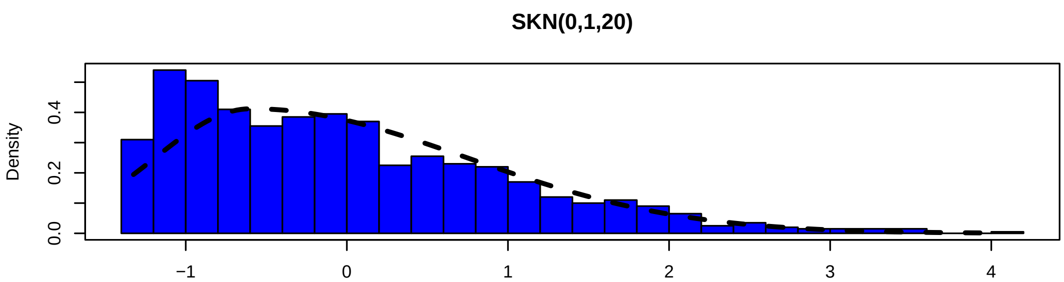


Fig.1.6:SKN(0,1,20)

2.Estimation Procedure-

We'll consider two cases in the estimation of shape parameter (λ):

- Both μ, σ are known
- Both μ, σ are unknown (there will be some computational challanges)

2.1 Method of Moment estimator (MME)

Suppose we have n samples $y_1, y_2, ..., y_n$ drawn from $SND(\mu, \sigma, \lambda)$,then we'll find the **Method of moments estimator(MME)** of λ

Case 1 (Both μ, σ are known):- From the property of SND, $E(Y_1) = \mu + \sigma\sqrt{\frac{2}{\pi}}\frac{\lambda}{\sqrt{1+\lambda^2}}$

By method of moment , $E(Y_1) = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$=> \mu + \sigma\sqrt{\frac{2}{\pi}}\frac{\lambda}{\sqrt{1+\lambda^2}} = \bar{y}$

$=> \frac{(\mu-\bar{y})}{\sigma\sqrt{\frac{2}{\pi}}} = \frac{\lambda}{\sqrt{1+\lambda^2}}$

$=> \frac{(\mu-\bar{y})^2}{\sigma^2\left(\frac{2}{\pi}\right)} = \frac{\lambda^2}{1+\lambda^2}$

$=> 1 - \frac{(\mu-\bar{y})^2}{\sigma^2\left(\frac{2}{\pi}\right)} = \frac{1}{1+\lambda^2}$

$=> 1 + \lambda^2 = \frac{\sigma^2\left(\frac{2}{\pi}\right)}{\sigma^2\left(\frac{2}{\pi}\right)-(\mu-\bar{y})^2}$

$=> \lambda^2 = \frac{(\mu-\bar{y})^2}{\sigma^2\left(\frac{2}{\pi}\right)-(\mu-\bar{y})^2}$

Since μ, σ are known, the quantity $\frac{(\mu-\bar{y})^2}{\sigma^2\left(\frac{2}{\pi}\right)-(\mu-\bar{y})^2}$ is free from unknown quantity .

So the method of moment estimator (MME) of λ is suggested by $\widehat{\lambda}^2 = \frac{(\mu-\bar{y})^2}{\sigma^2\left(\frac{2}{\pi}\right)-(\mu-\bar{y})^2}$.........................(2.1.1)

Since,$\frac{\widehat{\lambda}^2}{1+\widehat{\lambda}^2} = \left(\frac{\bar{y}-\mu}{\sigma\sqrt{\frac{2}{\pi}}}\right)^2 > 0$ so,the solution of $\widehat{\lambda}$ always exists.

So we'll always get a MME estimator (2.1.1) of shape parameter when the other two parameters are known.

Since λ is the shape parameter so we'll get its sign (λ) by the sign of sample $3rd$ order central moment ,$m_3 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^3$

- $m_3 > 0 => \widehat{\lambda} > 0$
- $m_3 < 0 => \widehat{\lambda} < 0$

So,An MME estimator of λ is suggested as $\widehat{\lambda_{MME}} = \widehat{\lambda} \times sign(m_3)$.

Case 2 (Both μ,σ are unknown):- When μ,σ are unknown then we can't use 2.1.1 as an estimator of λ as the quantity involves both the unknown parameter.

Let us derive the MME of the 3 unknown parameters by equating the first three moments of the population:

$$E(Y_1)=\mu+\sigma\sqrt{\tfrac{2}{\pi}}\tfrac{\lambda}{\sqrt{1+\lambda^2}}=\bar{y}=m_1$$

$$Var(Y_1)=\sigma^2(1-\tfrac{2}{\pi}\tfrac{\lambda^2}{1+\lambda^2})=\tfrac{1}{n}\textstyle\sum_{i=1}^{n}(y_i-\bar{y})^2=m_2$$

$$E(Y-E(Y))^3=\sigma^3\sqrt{\tfrac{2}{\pi}}(\tfrac{4}{\pi}-1)(\tfrac{\lambda}{\sqrt{1+\lambda^2}})^3=\tfrac{1}{n}\textstyle\sum_{i=1}^{n}(y_i-\bar{y})^3=m_3$$

Let us solve the 3 equations.

$$\frac{m_3^2}{m_2^3}=\frac{\left(\sigma^3\sqrt{\frac{2}{\pi}}(\frac{4}{\pi}-1)(\frac{\lambda}{\sqrt{1+\lambda^2}})^3\right)^2}{\left(\sigma^2(1-\frac{2}{\pi}\frac{\lambda^2}{1+\lambda^2})\right)^3}$$
$$=>\left(\frac{m_3^2}{m_2^3}\right)^{1/3}=\frac{\left(\frac{2}{\pi}(\frac{4}{\pi}-1)^2\right)^{1/3}(\frac{\lambda^2}{1+\lambda^2})}{(1-\frac{2}{\pi}\frac{\lambda^2}{1+\lambda^2})}$$
$$=>\left(\frac{m_3^2}{m_2^3}\right)^{1/3}=\frac{\left(\frac{2}{\pi}(\frac{4}{\pi}-1)^2\right)^{1/3}}{(\frac{1+\lambda^2}{\lambda^2}-\frac{2}{\pi})}$$
$$=>\left(\frac{m_2^3}{m_3^2}\right)^{1/3}=\frac{\frac{1+\lambda^2}{\lambda^2}-\frac{2}{\pi}}{\left((\frac{2}{\pi})(\frac{4}{\pi}-1)^2\right)^{1/3}}$$
$$=>\frac{1+\lambda^2}{\lambda^2}-\frac{2}{\pi}=\left((\frac{2}{\pi})(\frac{4}{\pi}-1)^2\frac{m_2^3}{m_3^2}\right)^{1/3}$$
$$=>\frac{1+\lambda^2}{\lambda^2}=\frac{2}{\pi}+\left((\frac{2}{\pi})(\frac{4}{\pi}-1)^2\frac{m_2^3}{m_3^2}\right)^{1/3}$$
$$=>\frac{1}{\lambda^2}+1=\frac{2}{\pi}+\left((\frac{2}{\pi})(\frac{4}{\pi}-1)^2\frac{m_2^3}{m_3^2}\right)^{1/3}$$
$$=>\frac{1}{\lambda^2}=\frac{2}{\pi}+\left((\frac{2}{\pi})(\frac{4}{\pi}-1)^2\frac{m_2^3}{m_3^2}\right)^{1/3}-1$$
$$=>\lambda^2=\frac{1}{\left((\frac{2}{\pi})(\frac{4}{\pi}-1)^2\frac{m_2^3}{m_3^2}\right)^{1/3}+\left(\frac{2}{\pi}-1\right)}$$

Since $\frac{1}{\left((\frac{2}{\pi})(\frac{4}{\pi}-1)^2\frac{m_2^3}{m_3^2}\right)^{1/3}+\left(\frac{2}{\pi}-1\right)}$ is a known quantity free from the two unknown parameters ,it can be suggested as MME estimator of shape parameter.

So, method of moment estimator (MME) of λ is suggested by $\widehat{\lambda}^2=\frac{1}{\left((\frac{2}{\pi})(\frac{4}{\pi}-1)^2\frac{m_2^3}{m_3^2}\right)^{1/3}+\left(\frac{2}{\pi}-1\right)}$...(2.1.2)

Now using the MME estimate of λ(2.1.2) ,we can get an estimate of μ,σ given by $\widehat{\mu},\widehat{\sigma}$

$$\sigma^2(1-\tfrac{2}{\pi}\tfrac{\lambda^2}{1+\lambda^2})=m_2$$
$$=>\widehat{\sigma}=\sqrt{\frac{m_2}{(1-\frac{2}{\pi}\frac{\lambda^2}{1+\lambda^2})}}\text{...(2.1.3)}$$

and

$$\mu+\sigma\sqrt{\tfrac{2}{\pi}}\tfrac{\lambda}{\sqrt{1+\lambda^2}}=m_1$$
$$=>\widehat{\mu}=m_1-\widehat{\sigma}\sqrt{\tfrac{2}{\pi}}\tfrac{\widehat{\lambda}}{\sqrt{1+\widehat{\lambda}^2}}\text{.....................................(2.1.4)}$$

The solution of (2.1.2) **exists** if $\frac{1}{\lambda^2} > 0$
$=> \frac{2}{\pi} + \left((\frac{2}{\pi})(\frac{4}{\pi} - 1)^2 \frac{m_2^3}{m_3^2}\right)^{1/3} - 1 > 0$
$=> \left((\frac{2}{\pi})(\frac{4}{\pi} - 1)^2 \frac{m_2^3}{m_3^2}\right)^{1/3} > 1 - \frac{2}{\pi}$
$=> (\frac{2}{\pi})(\frac{4}{\pi} - 1)^2 \frac{m_2^3}{m_3^2} > (1 - \frac{2}{\pi})^3$
$=> \frac{m_2^3}{m_3^2} > \frac{(1-\frac{2}{\pi})^3}{(\frac{2}{\pi})(\frac{4}{\pi}-1)^2}$..(2.1.5)

So,

- An MME estimator of λ is suggested as $\widehat{\lambda_{MME}} = \widehat{\lambda} \times sign(m_3)$ if (2.1.5) holds true.
- If (2.1.5) doesn't holds true in that situation MME estimator of λ doesn't exist.
- An MME estimator of σ is suggested as $\widehat{\sigma_{MME}} = \sqrt{\frac{m_2}{1-\frac{2}{\pi}\frac{\widehat{\lambda_{MME}}^2}{1+\widehat{\lambda_{MME}}^2}}}$ if (2.1.5) holds true.
- If (2.1.5) doesn't holds true in that case we take $\frac{\widehat{\lambda_{MME}}^2}{1+\widehat{\lambda_{MME}}^2} = 1$ then,$\widehat{\sigma_{MME}} = \sqrt{\frac{m_2}{(1-\frac{2}{\pi})}}$
- An MME estimator of μ is suggested as $\widehat{\mu_{MME}} = m_1 - \widehat{\sigma_{MME}}\sqrt{\frac{2}{\pi}}\frac{\widehat{\lambda_{MME}}}{\sqrt{1+\widehat{\lambda_{MME}}^2}}$ if (2.1.5) holds true.
- If (2.1.5) doesn't holds true in that case we take $\widehat{\mu_{MME}} = m_1 - \widehat{\sigma_{MME}}\sqrt{\frac{2}{\pi}}$

2.2 Maximum Likelihood estimator (MLE)-

Maximum likelihood estimator (MLE) is the most popular method used to obtain estimators of the 3 unknown parameters.

Now, we will derive the Maximum Likelihood estimator (MLE) of λ by writing the system of three equations. so,we will consider two cases:

Case 1 (Both μ, σ are known):- The likelihood function of μ, σ, λ given the n samples is given by:

$l(\mu, \sigma, \lambda | y_i : i = 1(1)n) = \prod_{i=1}^{n} \frac{2}{\sigma}\phi(\frac{(y_i-\mu)}{\sigma})\Phi(\frac{\lambda}{\sigma}(y_i - \mu))$

Then the Log likelihood function is given by-

$L(\mu, \sigma, \lambda | y_i : i = 1(1)n)$

$= log(l(\mu, \sigma, \lambda | y_i : i = 1(1)n))$

$= log \prod_{i=1}^{n} \frac{2}{\sigma}\phi(\frac{(y_i-\mu)}{\sigma})\Phi(\frac{\lambda}{\sigma}(y_i - \mu))$

$= \sum_{i=1}^{n} log(\frac{2}{\sigma}\phi(\frac{(y_i-\mu)}{\sigma})\Phi(\frac{\lambda}{\sigma}(y_i - \mu)))$

$= \sum_{i=1}^{n} log\frac{2}{\sigma} + \sum_{i=1}^{n} log\phi(\frac{(y_i-\mu)}{\sigma}) + \sum_{i=1}^{n} log\Phi(\frac{\lambda}{\sigma}(y_i - \mu))$

So, MLE estimator of λ is suggested as $\widehat{\lambda_{MLE}}$ which can be obtained by maximizing the log likelihood equation,i.e solving

$\frac{\delta log l(\mu,\sigma,\lambda|y_i:i=1(1)n)}{\delta\lambda} = 0$

$=> 0 + 0 + \sum_{i=1}^{n} \frac{\delta log\Phi(\frac{\lambda}{\sigma}(y_i-\mu)}{\delta\lambda} = 0$

$=> \sum_{i=1}^{n} \frac{\mathit{\Phi}(\frac{(y_i-\mu)}{\sigma})}{\Phi(\lambda\frac{(y_i-\mu)}{\sigma})}\frac{(y_i-\mu)}{\sigma^2} = 0$..(2.2.1)

Since μ, σ both are known,$\frac{\mathit{\Phi}(\frac{(y_i-\mu)}{\sigma})}{\Phi(\lambda\frac{(y_i-\mu)}{\sigma})}\frac{(y_i-\mu)}{\sigma^2}$ is free from unknown quantity

so,the MLE of λ is the solution of (2.2.1)

Case 2 (Both μ, σ are unknown):- Let us defne a mixture of parameters and sample observation:
$W(y_i) = \frac{\Phi(\frac{(y_i-\mu)}{\sigma})}{\Phi(\lambda\frac{(y_i-\mu)}{\sigma})}$

Since both μ, σ are unknown so we have to solve log likelihood function wrt the 3 unknown parameters μ, σ, λ:

$$\frac{\delta logl(\mu,\sigma,\lambda|y_i:i=1(1)n)}{\delta\lambda} = 0$$
$$=> \sum_{i=1}^{n} \frac{\Phi(\frac{(y_i-\mu)}{\sigma})}{\Phi(\lambda\frac{(y_i-\mu)}{\sigma})} \frac{(y_i-\mu)}{\sigma^2} = 0$$
$$=> \sum_{i=1}^{n} W(y_i)\frac{(y_i-\mu)}{\sigma^2} = 0$$
$$=> \sum_{i=1}^{n} W(y_i)(y_i-\mu) = 0 \text{...}(2.2.1)$$

$$\frac{\delta logl(\mu,\sigma,\lambda|y_i:i=1(1)n)}{\delta\mu} = 0$$
$$=> \frac{\delta}{\delta\mu}\{\sum_{i=1}^{n} log\frac{2}{\sigma} + \sum_{i=1}^{n} log\phi(\frac{(y_i-\mu)}{\sigma}) + \sum_{i=1}^{n} log\Phi(\frac{\lambda}{\sigma}(y_i-\mu))\} = 0$$
$$=> 0 + \sum_{i=1}^{n} \frac{(y_i-\mu)}{\sigma^2} - \frac{\lambda}{\sigma}\sum_{i=1}^{n} \frac{\Phi(\frac{(y_i-\mu)}{\sigma})}{\Phi(\lambda\frac{(y_i-\mu)}{\sigma})} = 0$$
$$=> \frac{(\bar{y}-\mu)}{\sigma^2} - \frac{\lambda}{n\sigma}\sum_{i=1}^{n} \frac{\Phi(\frac{(y_i-\mu)}{\sigma})}{\Phi(\lambda\frac{(y_i-\mu)}{\sigma})} = 0$$
$$=> n(\bar{y}-\mu) - \lambda\sigma\sum_{i=1}^{n} W(y_i) = 0 \text{...}(2.2.2)$$

$$\frac{\delta logl(\mu,\sigma,\lambda|y_i:i=1(1)n)}{\delta\sigma} = 0$$
$$=> \frac{\delta}{\delta\sigma}\{\sum_{i=1}^{n} log\frac{2}{\sigma} + \sum_{i=1}^{n} log\phi(\frac{(y_i-\mu)}{\sigma}) + \sum_{i=1}^{n} log\Phi(\frac{\lambda}{\sigma}(y_i-\mu))\} = 0$$
$$=> -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(y_i-\mu)^2}{\sigma^3} - \frac{\lambda}{\sigma^2}\sum_{i=1}^{n} \frac{\Phi(\frac{(y_i-\mu)}{\sigma})}{\Phi(\lambda\frac{(y_i-\mu)}{\sigma})}(y_i-\mu) = 0$$
$$=> \sum_{i=1}^{n}(y_i-\mu)^2 - \lambda\sigma\sum_{i=1}^{n} \frac{\Phi(\frac{(y_i-\mu)}{\sigma})}{\Phi(\lambda\frac{(y_i-\mu)}{\sigma})}(y_i-\mu) - n\sigma^2 = 0$$
$$=> \sum_{i=1}^{n}(y_i-\mu)^2 - \lambda\sigma\sum_{i=1}^{n} W(y_i)(y_i-\mu) - n\sigma^2 = 0 \text{....................}(2.2.3)$$

But Solving 3 equations to obtain the unknown parameters is non trivial due to several reasons:

First,An explicit analytical solution is not readily available due to the presence of the **nonlinear function** $\Phi(\cdot)$ and hence **numerical approximation is required**.

Second,The log-likelihood is **not convex** with respect to its parameters (μ, σ, λ) for the given data, which makes the optimization problem very difficult.Third ,there is always a stationary point at $\lambda = 0$

which causes some iterative optimization algorithms to stop at that point.

Liseo (1990) and **Azzalini** & **Arellano-Valle** (2013) noted that the maximum likelihood estimator of a finite λ can explode to $\pm\infty$ with non-negligible probability when the sample size is small.

This not only produces a huge bias in the estimation of λ, but also creates difficulty for likelihood inference as λ tends to $\pm\infty$ corresponds to a boundary point of the parameter space, which violates one of the regularity conditions for likelihood inference.

Due to these problems ,statistical inference related to SND becomes very difficult.So,getting an explicit solution of the 3 unknown parameters by solving $2.2.1, 2.2.2, 2.2.3$ is impossible.

2.3 Concept of Q Penalty:-

Let $y_1, y_2, ..., y_n$ be n random samples drawn from $SND(0, 1, \lambda)$.

Then the likelihood function of λ is given by

$l(\lambda|y_i : i = 1(1)n) = \prod_{i=1}^{n} 2\phi(y_i)\Phi(\lambda y_i)$

$= \prod_{i=1}^{n} \phi(y_i) \prod_{i=1}^{n} 2\Phi(\lambda y_i)$

$=> log(l(\lambda|y_i : i = 1(1)n)) = \sum log\phi(y_i) + \sum_{i=1}^{n} log(2\Phi(\lambda y_i))$

$\frac{\delta logl(\lambda|y_i:i=1(1)n)}{\delta\lambda} = 0$

$=> \sum_{i=1}^{n} \frac{\phi(\lambda y_i)}{\Phi(\lambda y_i)} y_i = 0$

This is known as **score function.**

Let,$S(\lambda) = \sum_{i=1}^{n} \frac{\phi(\lambda y_i)}{\Phi(\lambda y_i)} y_i$

$S'(\lambda)$ can be derived as follows,

$S'(\lambda) = -\sum_{i=1}^{n} \frac{y_i(\Phi(\lambda y_i)\phi'(\lambda y_i)y_i^2\lambda - (\phi(\lambda y_i))^2 y_i)}{(\Phi(\lambda y_i))^2}$

$= -\lambda \sum_{i=1}^{n} \frac{\phi'(\lambda y_i)y_i^3}{\Phi(\lambda y_i)} - \sum_{i=1}^{n} \frac{(\phi(\lambda y_i))^2 y_i^3}{(\Phi(\lambda y_i))^2}$

Sartori (2006) used a modified **score** function in an estimating equation for λ.The modified score function is

$S(\lambda)+Q(\lambda) = 0$ where $Q(\lambda) = E(S'(\lambda)b(\lambda)), b(\lambda) = Bias$ of λ..........................$(*)$

Sartori's estimator $\widehat{\lambda}$ is the solution of $(*)$ after replacing $Q(\lambda)$ by $Q_1(\lambda)$ as follows,

$Q_1(\lambda) = -\frac{\lambda}{2} \frac{a_{42}(\lambda)}{a_{22}(\lambda)}$

where $a_{ij}(\lambda) = E(Y^i \left(\frac{\phi(\lambda Y)}{\Phi(\lambda Y)}\right)^j)$

The expected values need to be numerically computed.

Bayes and Branco's (2007) $\widehat{\lambda}$ modified estimator is the solution of Equation $(*)$ after replacing $Q(\lambda)$ by $Q_2(\lambda)$ as follows,

$Q_2(\lambda) = -\frac{3\lambda}{2} \frac{1}{(1+\frac{8\lambda^2}{\pi^2})}$

Azzalini & **Arellano-Valle** (2013) added a **penalty function** of the shape parameter (λ) to control the explosion.They called it **Q-penalty.**

They replace $Q(\lambda)$ in Equation ($*$) by $Q_3(\lambda)$ as follows,

$Q_3(\lambda) = \frac{2c_1c_2\lambda}{1+c_2\lambda^2}$ where $c_1 = 0.8759$ and $c_2 = 0.2854$

Clearly, as $\lambda \to \infty$, $Q_3(\lambda) \to \infty$. Thus, by minimizing the **penalized loss function** created by adding the Q-penalty to the negative log likelihood.

The 3 estimator used in minimizing the modified score function , $Q_1(\lambda), Q_2(\lambda), Q_3(\lambda)$ all are $O(1/n)$ .So the finite MLE of λ exists for all of the three methods.

2.4 Iterative method:

The penalized log likelihood function is given by:

$log(l^*(\mu, \sigma, \lambda | y_i : i = 1(1)n)) = L^* = L - Q_3(\lambda)$

If we partially differentiate the penalized log likelihood function w.r.t the 3 unknown parameters we'll get the two equations (2.2.2), (2.2.3) along with a new equation:

$\frac{\delta L^*}{\delta\lambda} = 0$

$=> \frac{\delta L}{\delta\lambda} - Q_3(\lambda) = 0$

$=> \sum_{i=1}^{n} W_i \frac{(y_i-\mu)}{\sigma^2} - \frac{2c_1c_2\lambda}{1+c_2\lambda^2} = 0$..............................(2.4.1)

The **Maximum Likelihood estimators** (**MLE**) of μ, σ, λ given by $\widehat{\mu_{MLE}}, \widehat{\sigma_{MLE}}, \widehat{\lambda_{MLE}}$ are obtained numerically.

The computation of $\widehat{\mu_{MLE}}, \widehat{\sigma_{MLE}}, \widehat{\lambda_{MLE}}$ are done by following iterative method:

- We Start with an initial value of λ, call it λ_0 and let $\lambda_0 = \widehat{\lambda}_{MME}$ (We have already obtained the value of $\widehat{\lambda}_{MME}$)
- At m^{th} iteration ,we'll get the value of $\widehat{\mu^{(m)}}, \widehat{\sigma^{(m)}}$ by solving the equation $\frac{\delta L^*}{\delta\mu} = 0$, $\frac{\delta L^*}{\delta\sigma} = 0$
- So we'll get $L^*(\widehat{\mu^{(m)}}, \widehat{\sigma^{(m)}}, \lambda_0)$.
- Now we'll get the value of $\widehat{\lambda^{(m)}}$ by maximizing $L^*(\widehat{\mu^{(m)}}, \widehat{\sigma^{(m)}}, \lambda)$ wrt λ i.e.,solving $\frac{\delta L^*}{\delta\lambda} = 0$.
- So we'll get $L^*(\widehat{\mu^{(m)}}, \widehat{\sigma^{(m)}}, \widehat{\lambda^{(m)}})$.
- After getting $\widehat{\lambda^{(m)}}$ we'll get $\widehat{\mu^{(m+1)}}, \widehat{\sigma^{(m+1)}}$ by maximizing $L^*(\mu, \sigma, \widehat{\lambda^{(m)}})$ wrt μ, σ
- We repeat the above step untill $|L^*(\widehat{\mu^{(m+1)}}, \widehat{\sigma^{(m+1)}}, \widehat{\lambda^{(m+1)}}) - L^*(\widehat{\mu^{(m)}}, \widehat{\sigma^{(m)}}, \widehat{\lambda^{(m)}})|$ is very less.

2.5 Limitation and Bias Corrected MLE Estimator-

Let $\widehat{\lambda_1}, \widehat{\lambda_2}, \widehat{\lambda_3}$ be the maximum likelihood estimator (MLE) of λ respectively suggested by **Sartori,Bayes and Branco** and **Azzalini** & **Arellano-Valle.**

Then the Bias of the 3 estimator is given by $E(\widehat{\lambda_1}) - \lambda \approx E(\widehat{\lambda_2}) - \lambda \approx E(\widehat{\lambda_3}) - \lambda = O(1/n^2)$.

So this estimators gives **large bias** when sample size is moderate which is not at all desirable.

For large sample size n,

$\sqrt{n}(\widehat{\lambda_{MLE}} - \lambda) \rightarrow N(0, 1/i(\lambda))$ as $\widehat{\lambda_{MLE}}$ is the Best asymptotically normal (**BAN**) estimator.

where $i(\lambda)$ = Fisher information based on the single observation drawn from Skew Normal distribution.

So asymptotically the ML estimator of shape parameter goes close to the original λ.

Let us

$L^*(\widehat{\lambda_3}) = L(\lambda) + Q_3(\lambda)$ where $Q_3(\lambda) = \frac{2c_1c_2\lambda}{1+c_2\lambda^2}$

Now we take the derivative of the above equation w.r.t λ,

$L^{'*}(\widehat{\lambda_3}) = L^{'}(\lambda) + Q_3^{'}(\lambda)$

where,$Q_3^{'}(\lambda) = \frac{(1+c_2\lambda^2)2c_1c_2 - 2c_1c_2\lambda \times 2c_2\lambda}{(1+c_2\lambda^2)^2}$

$= \frac{2c_1c_2(1-c_2\lambda^2)}{(1+c_2\lambda^2)^2}$

Now we apply **taylor's approximation** for $L^{'*}(\widehat{\lambda_3})$ at the the neighborhood of λ,

$L^*(\widehat{\lambda_3}) = L^*(\lambda) + L^{'*}(\widehat{\lambda_3})(\widehat{\lambda_3} - \lambda)$....................................(2.5.1)

$=> 0 = L^*(\lambda) + L^{'*}(\widehat{\lambda_3})(\widehat{\lambda_3} - \lambda)$

$=> E(L^*(\lambda) + L^{'*}(\widehat{\lambda_3})(\widehat{\lambda_3} - \lambda)) = 0$

$=> L^*(\lambda) + E(L^{'*}(\widehat{\lambda_3}))(\widehat{\lambda_3} - \lambda) = 0$

$=> (\widehat{\lambda_3} - \lambda) = -\frac{L^*(\lambda)}{E(L^{'*}(\widehat{\lambda_3}))}$

$=> (\widehat{\lambda_3} - \lambda) = -\frac{L(\lambda)+Q_3(\lambda)}{E(L^{'}(\lambda)+Q_3^{'}(\lambda))}$

$-E(L^{'}(\lambda)) = n.i(\lambda)$

So,

$(\widehat{\lambda_3} - \lambda) = -\frac{L(\lambda)+Q_3(\lambda)}{-n.i(\lambda)+Q_3^{'}(\lambda))} = \frac{L(\lambda)+Q_3(\lambda)}{n.i(\lambda)-Q_3^{'}(\lambda))}$....................................(2.5.2)

Take the expectations in both side of 2.5.1 and we get,
$E(L^*(\widehat{\lambda}_3)) = E(L^*(\lambda)) + E(L^{'*}(\widehat{\lambda}_3)(\widehat{\lambda}_3 - \lambda)))$
$=> 0 = E(L^*(\lambda)) + E(L^{'*}(\lambda))E(\widehat{\lambda}_3 - \lambda) + Cov(L^{'*}(\lambda), \widehat{\lambda}_3 - \lambda)$
$=> 0 = E(L^*(\lambda)) + E(L^{'*}(\lambda))E(\widehat{\lambda}_3 - \lambda) + \{E(L^{'*}(\lambda)(\widehat{\lambda}_3 - \lambda)) - E(L^{'*}(\lambda))E(\widehat{\lambda}_3 - \lambda)\}$
$=> 0 = Q_3(\lambda) + E(\widehat{\lambda}_3 - \lambda) \times Q_3'(\lambda) - E(\widehat{\lambda}_3 - \lambda) \times na_{22}(\lambda) + \frac{1}{na_{22}(\lambda) - Q_3'(\lambda)}(-n(a_{42}(\lambda)\lambda + a_{33}(\lambda)))$
$=> 0 = Q_3(\lambda) + E(\widehat{\lambda}_3 - \lambda) \times (Q_3'(\lambda) - na_{22}(\lambda)) + \frac{1}{na_{22}(\lambda) - Q_3'(\lambda)}(-n(a_{42}(\lambda)\lambda + a_{33}(\lambda)))$
$=> E(\widehat{\lambda}_3 - \lambda) = \frac{-Q_3(\lambda) - \frac{1}{na_{22}(\lambda) - Q_3'(\lambda)}(-n(a_{42}(\lambda)\lambda + a_{33}(\lambda)))}{Q_3'(\lambda) - na_{22}(\lambda)}$
$=> E(\widehat{\lambda}_3 - \lambda) = \frac{-Q_3(\lambda)\left(Q_3'(\lambda) - na_{22}(\lambda)\right) - n(a_{42}(\lambda)\lambda + a_{33}(\lambda))}{\left(Q_3'(\lambda) - na_{22}(\lambda)\right)^2}$..(2.5.3)

The above expression in (2.5.2) is bias of the Maximum likelihood estimator of λ
.

So the Bias corrected estimator is given as $\widehat{\lambda}_3 - E(\widehat{\lambda}_3 - \lambda)$
$= \widehat{\lambda}_3 + \frac{Q_3(\lambda)\left(Q_3'(\lambda) - na_{22}(\lambda)\right) + n(a_{42}(\lambda)\lambda + a_{33}(\lambda))}{\left(Q_3'(\lambda) - na_{22}(\lambda)\right)^2}$

3.Regression -

So far we have derived the Method of moments estimator(**MME**) and Maximum Likelihood Estimator (**MLE**) of the shape parameter (λ) of a Skew-Normal Distribution.

Now our target is to fit the distribution based on a regression model.

So our aim is to fit an appropriate linear regression model for Y on vector of covariates $(\underset{\sim}{x_1}, \underset{\sim}{x_2}, ..., \underset{\sim}{x_p})$ and then to estimate the model parameters by least square method for a dataset of size say n,

$\{Y_i, \underset{\sim}{x_i} = (x_{i1}, x_{i2}, ..., x_{ip}) : i = 1(1)n\}$ corresponding to a random sample $(n >> p)$ drawn from the population under investigation.r

Here we exploit the property of assumed Skew-Normal distribution-

Y has same distribution as $\frac{\lambda|U|}{\sqrt{1+\lambda^2}} + \frac{V}{\sqrt{1+\lambda^2}}$ where $U \sim N(0,1)$ independently of $V \sim N(0,1)$

Let us define ,

$Y(\mu,\sigma,\lambda) = \mu + (\sqrt{1+\lambda^2}Y - \lambda|U|) \times \sigma$

where,$Y \sim SN(\mu,\sigma,\lambda)$

and $U \sim N(0,1)$

By the above property of Skew-Normal Distribution we get,$Y(\mu,\sigma,\lambda) \sim N(\mu,\sigma^2)$

So,

$E(Y(\mu,\sigma,\lambda)|\underset{\sim}{x}) = \mu(\underset{\sim}{x})$

So,we regress $Y(\mu,\sigma,\lambda)$ on $(\underset{\sim}{x_1}, \underset{\sim}{x_2}, ..., \underset{\sim}{x_p})$.

Our Proposed regression model is given by:$\underset{\sim}{Y}(\mu,\sigma,\lambda) = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$ where ,

$\underset{\sim}{Y}(\mu,\sigma,\lambda) = \begin{pmatrix} Y_1(\mu,\sigma,\lambda) \\ Y_2(\mu,\sigma,\lambda) \\ \vdots \\ Y_n(\mu,\sigma,\lambda) \end{pmatrix}$ be the response vector with each $Y_i(\mu,\sigma,\lambda) \sim N(\mu,\sigma^2), i = 1(1)n$,

the design matrix $X^{n\times p} = \begin{pmatrix} \underset{\sim}{x_1} \\ \underset{\sim}{x_2} \\ \vdots \\ \underset{\sim}{x_n} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$,

$\underset{\sim}{\beta}^{p\times 1} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$ be the regreesion coefficient and

$\underset{\sim}{\epsilon}^{n\times 1} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$ =Random Error vector we assume:

- Rank$(X^{n \times p}) = p(<< n)$ i.e X is a full column rank matrix.
- $E(\underset{\sim}{\epsilon}) = \underset{\sim}{0}$
- Dispersion matrix $disp(\underset{\sim}{\epsilon}) = \sigma^2 I_n$

We can see that the regressed variable $Y(\mu, \sigma, \lambda)$ is a function of unknown quantity μ, σ, λ and $|U|$.

we have already get an estimate (MME,MLE) of μ, σ, λ.We pluged those estimator into $Y(\mu, \sigma, \lambda)$.

For $|U|$ we simmulate n random observations $z_1, z_2, ..., z_n$ on $|U|$ as the distribution of $|U|$ is completely known.

Hence the plug in estimator $Y^*(\widehat{\mu_{MME}}, \widehat{\sigma_{MME}}, \widehat{\lambda_{MME}})$ and $Y^*(\widehat{\mu_{MLE}}, \widehat{\sigma_{MLE}}, \widehat{\lambda_{MLE}})$ is completely known to us and and we can use this as regressed variable to estimate β.

Then the Least Square estimate of β is given by $\widehat{\beta} = (X'X)^{-1}X'Y^*$which depends on the estimate of μ, σ, λ.

$\widehat{\beta}$ will be unique unbiased estimator of β and $\widehat{\beta}$ always exist because $R(X'X^{p \times p}) = R(X) = p$ i.e $(X'X)^{-1}$ always exist.

Now, to get initial estimate of λ which is to be pluged into Y^* can be obtained by regressing Y^* on $|U|$ as $y_i^* = \alpha + \lambda z_i + \epsilon_i, \forall i = 1(1)n$.

where,$\beta = \begin{pmatrix} \alpha \\ \lambda \end{pmatrix}$

So we can get an estimate $\widehat{\lambda}_0$ but it depends on a particular sample $\{z_1, z_2, ..., z_n\}$

Finally we repeat the process a large number of times and will take the average to get the initial estimate of the shape parameter $\widehat{\lambda}_0$.

4.Simulation study-

In this part ,we'll work with the simmulated data and see the result of our experiment:

First We'll see the method of moment (MOM) estimator of shape parameter for a skew normal distribution :

Table - 1:Method of moment estimator

Skew Normal	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\lambda}$
$\mu = 0, \sigma = 1, \lambda = -10$	-0.095	0.7713	-3.385
$\mu = 0, \sigma = 1, \lambda = -5$	-0.102	0.768	-2.592
$\mu = 0, \sigma = 1, \lambda = 1$	-0.019	1.009	1.055
$\mu = 0, \sigma = 1, \lambda = 5$	0.0367	0.999	4.999
$\mu = 0, \sigma = 1, \lambda = 10$	0.0348	0.999	9.999

The above table shows the method of moment estimator of $\widehat{\mu}, \widehat{\sigma}, \widehat{\lambda}$ and can see that as the value of the shape parameter increases the estimated $\widehat{\lambda}$ goes very close to the original shape parameter .So the method of moment estimator works very well for the Skew-normal distribution.

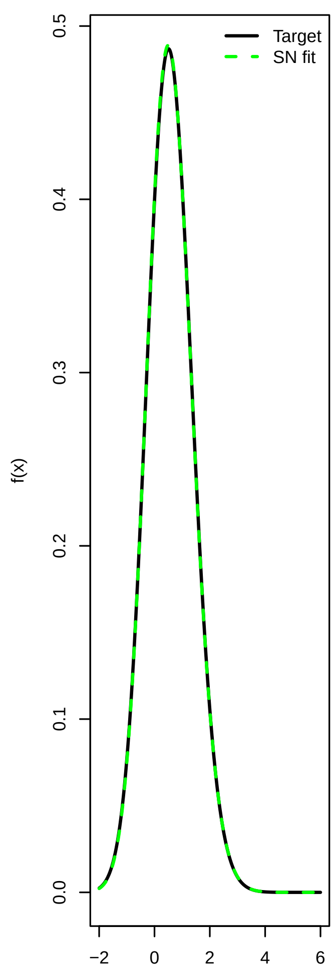


Fig 2.1 SKN(0,1,1)

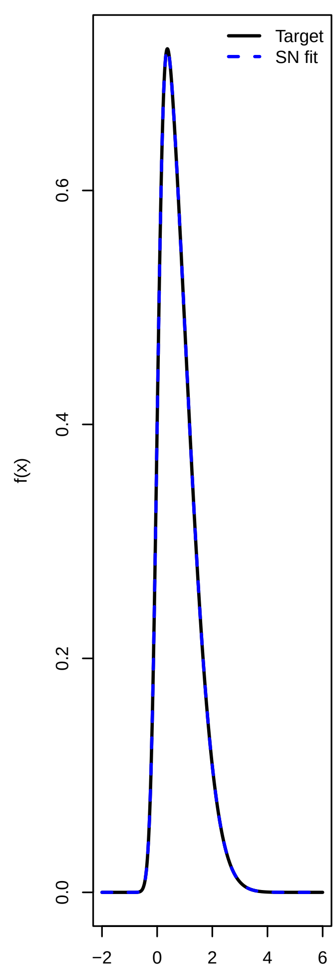


Fig 2.1 SKN(0,1,5)

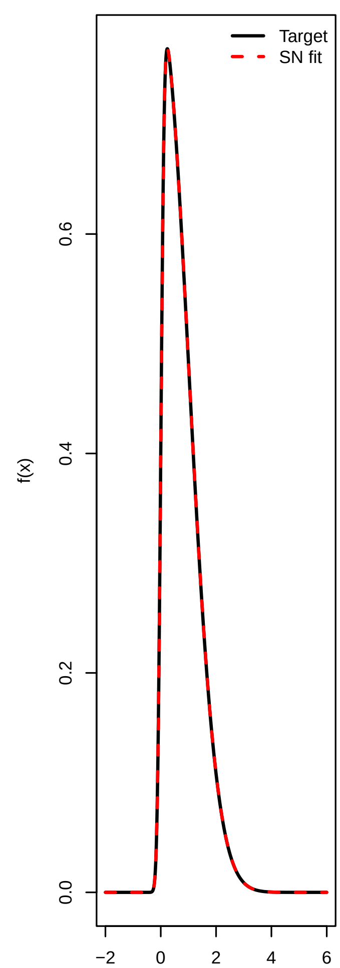


Fig 2.1 SKN(0,1,10)

So the above Fig $2.1, 2.2, 2.3$ shows that if we plot the fitted skew normal distribution (**MOM**) to the actuall skew normal distribution,thay are almost coincide.

Table - 2: Showing sample of size $n = 5$

Skew Normal	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\lambda}$	Iterations
$\mu = 0, \sigma = 1, \lambda = -10$	-0.367	1.273	0.036	150
$\mu = 0, \sigma = 1, \lambda = -5$	-0.361	0.653	0.002	126
$\mu = 0, \sigma = 1, \lambda = 1$	-0.304	0.699	0.003	118
$\mu = 0, \sigma = 1, \lambda = 5$	0.053	0.633	0.003	149
$\mu = 0, \sigma = 1, \lambda = 10$	-0.1305	0.319	0.004	121

The above table shows the maximum likelihood estimate of the 3 unknown parameters with a small or moderate sample size $n = 5$

Table - 3: Showing sample of size $n = 1000$

Skew Normal	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\lambda}$	Iterations
$\mu = 0, \sigma = 1, \lambda = -10$	0.049	0.947	0.1433	40
$\mu = 0, \sigma = 1, \lambda = -5$	0.03	0.981	0.2	26
$\mu = 0, \sigma = 1, \lambda = 1$	0.022	1.034	1.486	12
$\mu = 0, \sigma = 1, \lambda = 5$	0.041	1.029	4.460	17
$\mu = 0, \sigma = 1, \lambda = 10$	0.002	1.003	10.966	29

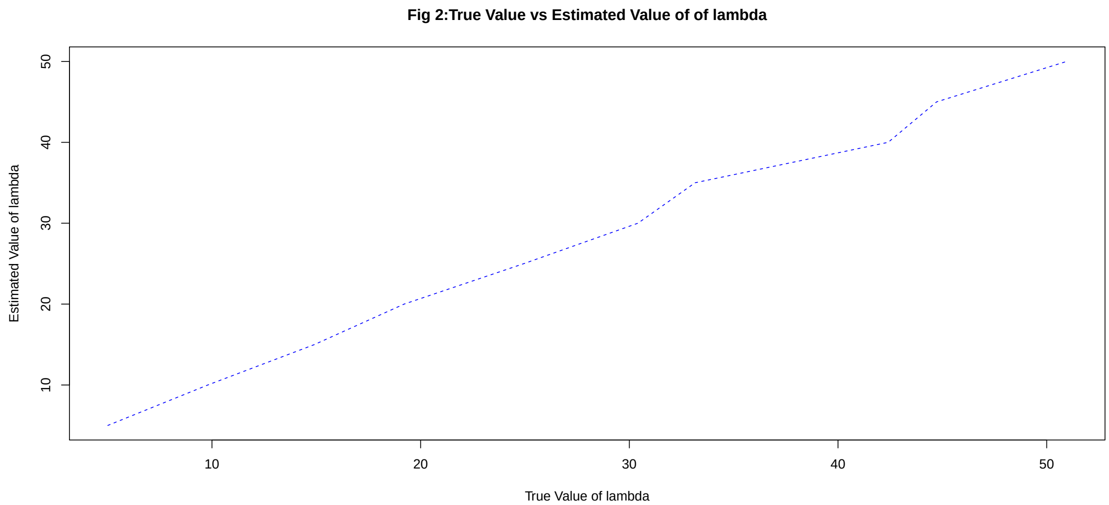
The above table shows the maximum likelihood estimate of the 3 unknown parameters with a large sample size $n = 1000$

Hence from the above two table $(2, 3)$ we can clearly conclude that for small or moderate sample size fitted skew-Normal distribution doesn't perform well as the bias increases as a result the **Mean Square Error** (MSE) increases and also number of iteration is very large for small sample size data.

On the other side as the sample size increases, the fitted $\widehat{\mu}, \widehat{\sigma}, \widehat{\lambda}$ goes very close to the actual parameter value.So the bias of λ reduces for large sample size

and $|L^*(\widehat{\mu^{(m+1)}}, \widehat{\sigma^{(m+1)}}, \widehat{\lambda^{(m+1)}}) - L^*(\widehat{\mu^{(m)}}, \widehat{\sigma^{(m)}}, \widehat{\lambda^{(m)}})|$ is approximately very small with small no of iterations.So,Maximum likelihood estimator (MLE) performs very well for large sample size

We already saw that the Maximum likelihood estimator of the shape parameter goes asymptotically to the original value of shape parameter and we show it for simmulated data also.



The above plot (Fig 2) shows how the maximum likelihood estimator of λ changes for different value of λ.We take the range of λ from 5 to 50 for very large sample size = 10000.

5.R-Codes

```
options(warn=-1)

library(fGarch)

par(mfrow=c(3,1))

#SKN(0,1,-10)
r1=rsnorm(1000, mean = 0, sd = 1, xi =-10)
x1= seq(min(r1), max(r1), length =201)
hist(r1, n=25, probability = TRUE,main="SKN(0,1,-10)",xlab="Fig.1.1:SKN(0,1,-10)", c
box()
lines(x1, dsnorm(x1), lwd = 2,lty=2)

#SKN(0,1,-5)
r2=rsnorm(1000, mean = 0, sd = 1, xi =-5)
x2= seq(min(r2), max(r2), length =201)
hist(r2, n =25, probability = TRUE,main="SKN(0,1,-5)",xlab="Fig.1.2:SKN(0,1,-5)", co
box()
lines(x2, dsnorm(x2), lwd = 2,lty=2)


#SKN(0,1,0)
r3 = rsnorm(n =1000)
hist(r3, n = 25, probability = TRUE,main="SKN(0,1,0)",xlab="Fig.1.3:SKN(0,1,0)", col
box()
x3= seq(min(r3), max(r3), length = 201)
```

```
lines(x3, dsnorm(x3), lwd = 2,lty=2)

#SKN(0,1,5)
r4=rsnorm(1000, mean = 0, sd = 1, xi =5)
x4= seq(min(r4), max(r4), length =201)
hist(r4, n =25, probability = TRUE,main="SKN(0,1,5)",xlab="Fig.1.4:SKN(0,1,5)", col
box()
lines(x4, dsnorm(x3), lwd = 2,lty=2)


#SKN(0,1,10)
r5=rsnorm(1000, mean = 0, sd = 1, xi =10)
x5= seq(min(r5), max(r5), length =201)
hist(r5, n =25, probability = TRUE,main="SKN(0,1,10)",xlab="Fig.1.5:SKN(0,1,10)", co
box()
lines(x5, dsnorm(x5), lwd = 3,lty=2)

#SKN(0,1,20)
r6=rsnorm(1000, mean = 0, sd = 1, xi =20)
x6= seq(min(r6), max(r6), length =201)
hist(r6, n =25, probability = TRUE,main="SKN(0,1,20)",xlab="Fig.1.6:SKN(0,1,20)", co
box()
lines(x6, dsnorm(x6), lwd = 3,lty=2)

#Method of Moment Estimator
library(blapsr)
par(mfrow=c(3,1))

sn.target <- function(x, location, scale, shape)
{
val <- 2 * stats::dnorm(x, mean = location, sd = scale) * pnorm(shape * (x - locatio
return(val)
}
x.grid <- seq(-2, 6, length = 200)

par(mfrow=c(1,3))

domx <- seq(-2, 6, length = 1000)
y.grid1 <- sapply(x.grid, sn.target, location = 0, scale = 1, shape = 1)
plot(domx, sapply(domx, sn.target, location = 0, scale =1, shape = 1),type = "l", yl
fit1 <- snmatch(x.grid, y.grid1)
lines(fit1$xgrid, fit1$snfit, type="l", col = "green", lwd = 2, lty = 2)
legend("topright", lty = c(1,2), col = c("black", "green"), lwd = c(2, 2),c("Target'

y.grid2 <- sapply(x.grid, sn.target, location = 0, scale = 1, shape = 5)
fit2 <- snmatch(x.grid, y.grid2)
plot(domx, sapply(domx, sn.target, location = 0, scale =1, shape = 5),type = "l", yl
lines(fit2$xgrid, fit2$snfit, type="l", col = "blue", lwd = 2, lty = 2)
```

```
legend("topright", lty = c(1,2), col = c("black", "blue"), lwd = c(2, 2),c("Target",

y.grid3 <- sapply(x.grid, sn.target, location = 0, scale = 1, shape = 10)
fit3 <- snmatch(x.grid, y.grid3)
plot(domx, sapply(domx, sn.target, location = 0, scale =1, shape = 10),type = "l", y
lines(fit3$xgrid, fit3$snfit, type="l", col = "red", lwd = 2, lty = 2)
legend("topright", lty = c(1,2), col = c("black", "red"), lwd = c(2, 2),c("Target","




#True lambda vs Estimated lambda
r1= rsnorm(n=10000,mean = 0, sd = 1, xi =5)
lmbda1=snormFit(r1)$par[3]
r2= rsnorm(n=10000,mean = 0, sd = 1, xi =10)
lmbda2=snormFit(r2)$par[3]
r3= rsnorm(n=10000,mean = 0, sd = 1, xi =15)
lmbda3=snormFit(r3)$par[3]
r4= rsnorm(n=10000,mean = 0, sd = 1, xi =20)
lmbda4=snormFit(r4)$par[3]
r5= rsnorm(n=10000,mean = 0, sd = 1, xi =25)
lmbda5=snormFit(r5)$par[3]
r6= rsnorm(n=10000,mean = 0, sd = 1, xi =30)
lmbda6=snormFit(r6)$par[3]
r7= rsnorm(n=10000,mean = 0, sd = 1, xi =35)
lmbda7=snormFit(r7)$par[3]
r8= rsnorm(n=10000,mean = 0, sd = 1, xi =40)
lmbda8=snormFit(r8)$par[3]
r9= rsnorm(n=10000,mean = 0, sd = 1, xi =45)
lmbda9=snormFit(r9)$par[3]
lmbda9
r10= rsnorm(n=10000,mean = 0, sd = 1, xi =50)
lmbda10=snormFit(r10)$par[3]
lmbda=c(lmbda1,lmbda2,lmbda3,lmbda4,lmbda5,lmbda6,lmbda7,lmbda8,lmbda9,lmbda10)
plot(lmbda,seq(5,50,by=5),type="l",xlab="True Value",ylab="Estimated Value",main="Tr
```

Conclusion:

This detailed work on estimation of an Skew-Normal distribution shape parameter (λ) gives a closer look how both the Method of moment estimators and Maximum likelihood estimators are obtained in a classical way, and how do they perform for small to large sample sizes.We conclude that the method of moment estimator fitted the skew-normal data very well irrespective of the sample size and the maximum likelihood estimator method does not perform well for small sample size but the bias reduction method perform very well for large sample size data and using the estimator as an initial choice if we perform a linear regression model and then estimate the model parameter and finally we'll get the fitted shape parameter.

Future prospects:

The above discussed results may be applied for any skewed data and anyone can estimate the shape parameter of the skewed data.The estimator will be very good for large sample size data and can approximate its original value by the estimated value.Then one can perform testing of hypothesis.In our project we assumed that the dispersion matrix of the random error is independent and homoscedastic.If anyone want to work with heteroscedastic model then it will involve lot of parameters and there is a chance of overfitting.So the resulting shape parameter may overfit the data.

References

[1] Azzalini, A. A class of distributions which includes the normal ones[J]. Scandinavian Journal of Statistics, 1985

[2] Gomez, H.W., Venegas, O., Bolfarine, H. Skew-symmetric distributions generated by the distribution function of the normal distribution[J]. Environmetrics, 2007

[3] Azzalini, A., The skew-normal and related families. Cambridge University Press. 2013

[4] Azzalini, A. & Arellano-Valle, R, Maximum penalized likelihood estimation for skew-normal and skew-t distributions. Journal of Statistical Planning and Inference,2013

[5] Raqab et. al.(2019): Compound Zero-Truncated Poisson Normal Distribution and its Applications, Communication in Statistics- Theory and Methods

[6] Regression Diagnostics by Sen & Srivastava