Moment Generating Functions of Observed Gaps between Hypopnea Using Saddlepoint Approximations

Nur Zakiah Mohd Saat and Abdul Aziz Jemain

Abstract—Saddlepoint approximations is one of the tools to obtain an expressions for densities and distribution functions. We approximate the densities of the observed gaps between the hypopnea events using the Huzurbazar saddlepoint approximation. We demonstrate the density of a maximum likelihood estimator in exponential families.

Keywords—Exponential, maximum likehood estimators, observed gap, Saddlepoint approximations.

I. INTRODUCTION

PEOPLE with sleep apnea stop breathing for 10 to 30 seconds at a time while they are sleeping.There are two kinds of sleep apnea which is obstructive sleep apnea and central sleep apnea. Obstructive sleep apnea (OSA) is the most common type. The person that have the OSA feel like something is blocking the trachea that brings air into the body. While central sleep apnea is a rare cases and it is related to the function of the central nervous system[1].

Doctors estimated that about 12 million Americans people have sleep apnea.Men and people who are over 40 years old are more likely to have sleep apnea. Apnea is defined as a complete cessastion of airflow for at least 10 seconds. Whilst hypopnea is defined as decreased in airflow of at least 50% with a concomitant fall of at least 4% in the arterial oxygen saturation,followed by an arousal response. The symptom is heavy snoring or long pauses in breathing during sleep[2].

The duration of the hypopnea is recorded in every 30 seconds. Between the 30 seconds there might be a hypopnea that is not recorded because it is less than 10 seconds. Let Y be the duration of the gap between hypopnea's, which is between the real hypopnea and the hidden hypopnea or vice-versa. While X denoted the duration of hypopnea's which is the duration of the real hypopnea and the duration of the hidden hypopnea. Next, Z is denoted as the true gap or observed gap, which is sum of the duration of gap between the real hypopnea and the hidden hypopnea denotes as Y_1, the duration of the hidden hypopnea denotes as X_2 and the duration of the gap between the hidden hypopnea and the real hypopnea Y_2. .Y's and X's are in alternating sequence.The summation of the X's and Y's are geometric

Nur Zakiah is with the Department of Biomedical Science, FSKB, UKM, Kuala Lumpur,MALAYSIA email: nur14504@gmail.com
Abdul Aziz Jemain is with the School of Mathematical Sciences,FST,UKM, Bangi,MALAYSIA email: azizj@pkrisc.cc.ukm.my

distributed. X's and Y's are independent and identically distributed (iid). $X_1, X_2, X_3, ... X_n$ and $Y_1, Y_2, Y_3, ..., Y_n$ are also iid. This study are interested to determine the effect of hidden hypopnea and the process of hidden hypopnea is not homogeneous in time because the subjects sleep stages is changing through time. The observed gap is a sum of n th hidden hypopnea and sum of (n+1) true gaps, n is the number of events.
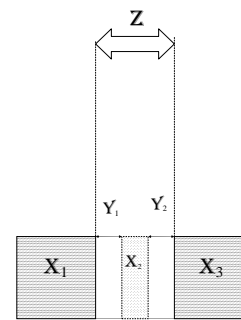

Fig. 1.   Diagram of the observed gap, Z. Y is the duration of gap. X_1 and X_3 is the duration of hypopnea's, X_2 is the hidden hypopnea

II. MOMENT GENERATING FUNCTIONS

The distribution of X and Y were determined in order to find the distribution of the observed gaps. Moreover, the observed distribution of gaps and the distribution of hypopnea durations were used to make inferences about the distributions of the gaps(Y).Suppose the moment generating functions of X and Y are $M_X(t)$ and $M_Y(t)$ respectively

Let p denotes the probability that a hypopnea is not observed

$$p = \int_0^{10} f(x)dx = Pr(\text{a hypopnea is hidden})$$

The distribution of the number of hidden hypopneas that occur between two observed hypopneas whilst n is the number of false events during the gap, has the mgf,

$$M_z(t) = \prod_{i=1}^{n} M_{x_i}(t) \prod_{i=1}^{n} M_{y_i}(t) \qquad (1)$$

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:2, No:2, 2008

and the mgf of X given that X < 10

$$
\begin{aligned}
M_{X_1+X_2+...+X_n} &= E\left[e^t(X_1 + X_2 + ... + X_n)\right] \\
&= E(e^t X_1)...E(e^t X_n) \\
&= \prod_{i=1}^{n} M_{X_i}(t)
\end{aligned}
$$

and the mgf of time of gap which is the distribution of y

$$
\begin{aligned}
M_{Y_1+Y_2+...+Y_n} &= E\left(e^t(X_1 + X_2 + ... + X_n)\right) \\
&= E(e^t Y_1)...E(e^t Y_n) \\
&= \prod_{i=1}^{n} M_{y_i}(t)
\end{aligned}
$$

thus the mgf of observed gap distribution

$$
M_z(t) = M(X_1 + X_2 + ... + X_n)(t)M(Y_1 + Y_2 + ... + Y_n)(t)
$$
$$
= \prod_{i=1}^{n} M_{x_i}(t) \prod_{i=1}^{n+1} M_{y_i}(t)
$$

Now for conditional on N=n, we have mgf of z

$$
\begin{aligned}
M_z(t) &= \sum_{n=0}^{\infty} M_{z|n}(t)(1-p)^{n-1}p \\
&= \sum_{n=0}^{\infty} M_x^n(t)M_y^{n+1}(t)(1-p)^{n-1}p \\
&= pM_y(t) \sum_{n=0}^{\infty} [M_x(t)M_y(t)(1-p)]^n \\
&= \frac{pM_y(t)}{1 - M_x(t)M_y(t)(1-p)} \quad (2)
\end{aligned}
$$

A. *Mean and variance of z*

Cumulant generating function(CGF)

$$
\begin{aligned}
R(t) = \ln M_z(t) &= ln\left[\frac{pM_y(t)}{1 - [M_y(t)M_x(t)(1-p)]}\right] \\
&= \ln[pM_y(t)] - ln[1 - M_x(t)M_y(t)(1-p)]
\end{aligned}
$$

First derivatives

$$
R'(t) = \frac{M_y'(t)}{M_y(t)} + \frac{(1-p)\left[M_x(t)M_y'(t) + M_x'(t)M_y(t)\right]}{[1 - M_x(t)M_y(t)(1-p)]}
$$

and the mean of z is
$E(z) = R'(0)$

$$
E(z) = \frac{M_y'(0)}{M_y(0)} + \frac{(1-p)\left[M_x(0)M_y'(0) + M_x'(0)M_y(0)\right]}{1 - M_x(0)M_y(0)(1-p)}
$$

and

$M_y(0) = 1 , M_x(0) = 1[3]$
thus

$$
E(z) = \frac{M_y'(0) + M_x'(0) - pM_x'(0)}{p}
$$
$$
M_y'(0) = E(Y), M'x(0) = E(X)
$$

$$
E(z) = \frac{E(Y) + (1-p)E(X)}{p} \quad (3)
$$

To find the variance of z,

$$
\begin{aligned}
Var(z) &= \frac{M_y'(t)}{M_y(t)} + \frac{(1-p)M_x(t)M_y'(t)}{1 - M_x(t)M_y(t)(1-p)} + \\
& \quad \frac{(1-p)M_y(t)M_x'(t)}{1 - M_x(t)M_y(t)(1-p)}
\end{aligned}
$$

Let $Var(Z) = R_a'(t) + R_b'(t) + R_c'(t)$
and

$$
\begin{aligned}
R_a'(t) &= \frac{M_y'(t)}{M_y(t)} \\
R_a''(t) &= \frac{(M_y(t)M_y''(t) - M_y'(t)M_y'(t))}{M_y(t)^2} \\
R_a''(0) &= \frac{M_y''(0) - [M_y(0)]^2}{M_y(0)^2} \\
&= Var(Y)
\end{aligned}
$$

and

$$
R_b'(t) = \frac{(1-p)M_x(t)M_y'(t)}{1 - M_x(t)M_y(t)(1-p)}
$$

$$
R_b''(t) = [[1 - M_x(t)M_y(t)][M_x(t)M_x''(t) + M_x'(t)M_y'(t)]
$$

$$
-[M_x(t)M_y'(t)] - \left[M_x(t)M_y'(t) + M_y(t)M_x'(t)\right](1-p)] \Big/
$$

$$
[1 - M_x(t)M_y(t)(1-p)]^2
$$

$$
R_b''(0) = [p[M_y''(0) + [M_y'(0)]^2] +
$$

$$
[E(Y)(1-p)[E(Y) + E(X)]](1-p) \Big/ p^2
$$

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:2, No:2, 2008

$$= \left[\frac{\mathrm{Var}(Y)}{p} + \frac{[E(Y)^2] + E(Y)E(X)}{p^2} \right] (1-p)$$

and

$$R'_c(t) = \frac{(1-p)M_y(t)M'_x(t)}{1 - M_x(t)M_y(t)(1-p)}$$

$$R''_c(0) = \left[\frac{\mathrm{Var}(X)}{p} + \frac{[E(X)]^2 + E(Y)E(X)}{p^2} \right] (1-p)$$

thus

$$Var(Z) = Var(Y) + \frac{1-p}{p} \left[Var(X) + Var(Y) \right] +$$

$$\left[E(X)^2 + E(Y)^2 + 2E(Y)E(X) \right] \frac{1-p}{p^2}$$

III. SADDLEPOINT APPROXIMATION

$$\hat{f}_T(t) = [2\pi K''(s)]^{-1/2} \exp[K(s) - t\hat{s}] \qquad (4)$$

$$K'_z(s) = t \qquad (5)$$

where $\hat{s}$ is the saddlepoint is solved by computing $K'_z(s) = t$ numerically

A. Exponential distribution

Let p be the probability of the hidden hypopnea ($< 10s$) with exponential distribution

$$\begin{aligned} p &= \int_0^c \lambda \exp(-\lambda k)dk \\ &= \left| -\exp(-\lambda k) \right|_0^c \\ &= \exp(-\lambda c) \end{aligned}$$

Let X denotes the duration of the hypopnea has an exponent distribution. The mean of X,

$$\begin{aligned} E(X) &= \frac{1}{1-p} \int_c^\infty \lambda x \exp(-\lambda x)dx \\ &= \frac{\exp(-\lambda c) + \exp(-\lambda c)\lambda c) - 1}{\lambda(p-1)} \\ &= \frac{p + p\lambda c - 1}{\lambda(p-1)} \\ &= \frac{1 - p - p\lambda c}{\lambda(1-p)} \end{aligned}$$

To find the variance of X, we need to find

$$\begin{aligned} E(X^2) &= \int_c^\infty \frac{\lambda x^2 \exp(-\lambda x)}{1-p} dx \\ &= \frac{-2\exp(-\lambda c) + 2\exp(-\lambda c)\lambda c + \exp(-\lambda c)c^2\lambda^2 - 2}{\lambda^2(1-p)} \\ &= \frac{-2p + 2p\lambda c + pc^2\lambda^2 - 2}{\lambda^2(1-p)} \end{aligned}$$

thus

$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{-2p + 2p\lambda c + pc^2\lambda^2 - 2}{\lambda^2(1-p)} - \left[\frac{1-p-p\lambda c}{\lambda(1-p)} \right]^2 \end{aligned}$$

Moment generating function of X

$$\begin{aligned} M_x(t) &= \frac{\lambda}{1-p} \int_c^\infty \exp(tx)\lambda \exp(-\lambda x)\mathrm{dx} \\ &= \frac{\lambda}{1-p} \left[\frac{\exp(tc - \lambda c)}{-t + \lambda} + \frac{1}{-t + \lambda} \right] \\ &= \frac{(1 - \exp(tc - \lambda c))\lambda}{(1-p)(\lambda - t)} \end{aligned}$$

Y is the duration of the gap between the hypopnea, which has the standard exponetial mean and variance.

$$\mu = R'(0) = \frac{1}{\mu}$$

$$\sigma^2 = R''(0) = \frac{1}{\mu^2}$$

Moment generating function of Y

$$M_y(t) = \int_0^\infty \exp(ty)\mu \exp(-\mu y)\mathrm{dy} = \frac{\mu}{\mu - t} \qquad (6)$$

Mean of z by substituting formula (3)

$$E(z) = \frac{\lambda + 1 - p - p\lambda c}{\mu\lambda p} \qquad (7)$$

Moment generating function of z

$$M_z(t) = \frac{p\mu(\lambda - t)}{(\lambda - t)(\mu - t) - \lambda\mu(1 - \exp(tc - \lambda c))} \qquad (8)$$

Cumulant generating function(CGF) of z

$$K_z(s) = \ln(M_z(s)) \qquad (9)$$

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:2, No:2, 2008

First derivatives

$$K'_z(s) = -s^2 - sc\exp(-c(\lambda-s))\lambda\mu + 2\lambda s - \lambda\mu - \lambda^2 + \lambda\mu$$
$$\exp(-c(\lambda-s)) + \lambda^2\mu c\exp(-c(\lambda-s)) \Big/ (\lambda-s)(\lambda s + s\mu - s^2$$
$$-\lambda\mu\exp(-c(\lambda-s))))$$

Second derivatives
$$K''_z(s) = (\exp(-c(\lambda-s))(-2\lambda^3\mu - 2\mu^2 c\lambda s^2 + 4\mu c\lambda s^3$$
$$+ \mu\lambda^4 c^2 t - 3\mu\lambda^3 c^2 s^2 + \mu^2\lambda^3 c^2 s$$
$$-2\mu^2\lambda^2 c^2 s^2 + 3\mu\lambda^2 c^2 s^3 + \mu^2 c^2\lambda s^3$$
$$-\mu c^2\lambda s^4 + 8\mu\lambda^3 cs + 6\lambda^2\mu s - 2\mu\lambda^4 c - 2\lambda^3\mu^2 c + 2\lambda\mu^2 s$$
$$-4\lambda\mu s^2 + 4\mu^2\lambda^2 cs - 10\mu\lambda^2 cs^2) - \lambda^2\mu^2\exp(-2c(\lambda-s))$$
$$-4\lambda^3 s + 6\lambda^2 s^2 - 4\lambda s^3 - 6\lambda^2 s\mu + 4\lambda s^2\mu - 2\lambda\mu^2 + \lambda^4$$
$$+2\lambda^3\mu + \lambda^2\mu^2 + s^4) \Big/ (\lambda-t)^2(-\lambda s - s\mu + s^2 + \lambda\mu\exp(-c(s-$$
$$\lambda)))^2$$

The mgf of the gap distribution which is defined as y, is assumed exponential with parameter μ with the standard mgf exponent. The mgf of the hypopnea distribution which is defined as x, assumed to be exponential with parameter λ but left-truncated at c=10. In order to be sure that the mgf exists around zero it must exist in an interval around the origin (c_1, c_2) where $c_1 < 0$ and $c_2 > 0$ [4]

M_y is the MGF of an exponential distribution. It exists when $s < \mu$, s is the saddlepoint approximation. This means that the MGF of the truncated distribution X is exists for all values of s. MGF of the z is obtained from the observed gap by summing a geometric series,under conditions,which is within the range(-1,1). Thus, $pm_x m_y$ has to lie in the range(-1,1).Recalling that p is the duration of hidden hypopnea. Hence, $pm_x m_y$ will always be positive. It is certain when $pm_x m_y < 1$. This puts a limit on the maximum value of s that is allowed.

A smaller value than this is require to ensure that m_z exists. Because the mgf exists for any negative value of s, we can take c_1 to be large negative value such as -5000. Next, to solve the saddlepoint equation $k_1 - t = 0$ for the different values of t , we can look for solutions in the range (c_1, c_2)[4].

B. Gamma distribution

Let p be the probability of the hidden hypopnea $(< 10s)$with gamma distribution. The probability of the duration of hidden hypopnea is the integration of the Incomplete gamma function $[0, c]$ [5]

$$p_x = \int_0^c \frac{x^{\alpha-1}\exp(-x/\beta)}{\Gamma(\alpha)\beta^\alpha}dx$$
$$= \frac{1}{\Gamma(\alpha)\beta}\int_0^c \left[\frac{x}{\beta}\right]^{\alpha-1}\exp(-x/\beta)dx$$

Let
$$u = \frac{x}{\beta}$$

and
$$du = \frac{dx}{\beta}$$

$$p_u = \int_0^{\frac{c}{\beta}} u^{\alpha-1}\exp(-u)du = \Gamma(\alpha, c/\beta)\Big/\Gamma(\alpha) \quad (10)$$

and the MGF of the duration of the gap between the hypopnea is given by the MGF of the standard gamma function

$$M_y(s) = \int_0^\infty \frac{y^{k-1}\exp(sy - y/\lambda)}{\Gamma(k)}dy = \frac{1}{(1-\lambda s)^k} \quad (11)$$

and the MGF of the duration of hypopnea is given by
$$M_x(s) = \int_c^\infty \frac{x^{\alpha-1}}{\beta^\alpha\Gamma(\alpha)p}\exp\left[-\frac{x}{\beta} + sx\right]dx$$

Let
$$u = sx - \frac{x}{\beta}$$

and
$$du = (s - \frac{1}{\beta})\mathrm{dx}$$

$$M_u(s) = \frac{1}{(1-\beta s)^\alpha(1-p)\Gamma(\alpha)}\int_{\frac{c}{\beta}-cs}^\infty u^{\alpha-1}\exp(-u)du$$
$$= \frac{\Gamma(\alpha, c/\beta - cs)}{(1-p)\Gamma(\alpha)(1-\beta s)^\alpha}$$

thus the MGF of the duration of observed gap can be written as

$$M_z(s) = \Gamma(\alpha, \frac{c}{\beta})\Gamma(\alpha)(1-p)(1-\beta s)^\alpha \Big/ \Gamma(\alpha) - \Gamma(\alpha, \frac{c}{\beta} - cs)\Gamma(\alpha, \frac{c}{\beta}) \quad (12)$$

Cumulant generating function(CGF) of z
$$K_z(s) = ln(M_z(s))$$

First derivatives
$$K'_z(s) = gradient(kz, s) \quad (13)$$

Second derivatives
Let $K_z(s) = L(\alpha, \beta)$ [6]

$$K''_z(s) = hessian(kz, s) = \begin{bmatrix} \delta L/\delta\alpha^2 & \delta^2 L/\delta\alpha\beta \\ \delta^2 L/\delta\alpha\beta & \delta L/\delta\beta^2 \end{bmatrix} \quad (14)$$

IV. RESULTS

Saddlepoint approximation is a method to approximate density function from a MGF[7]. Using the saddlepoint approximation, we want to approximate the PDF of the M_z, which is the observed gap. Firstly we use the exponent distribution as it has an explicit form of MGF. This is generally to get the idea on the working of saddlepoint aprroximation in the R program. Furthermore, other positive skewed distribution has a special case of exponent distribution. Exponent distribution is a special case of gamma distribution when the shape parameter is equal to 1 or usually named as

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:2, No:2, 2008

Incomplete Gamma Function. So, it is essential to check the model of gamma distributions with exponent distribution in order to clarify the adequacy of the MGF that is generated.

Secondly, the saddlepoint approximation that produce the MGF of M_z will be compared with the simulation of the distribution that has been used for example the MGF of gamma distribution will be compared with the simulation of random numbers generated from gamma distribution.

A. Exponential distribution

A exponent random variable T$\sim$ Exp(λ) with mean $(1/\lambda)$ has a CGF given in section 2. Solution of the saddlepoint approximation density given in section 2, yields a saddlepoint that can be solved numerically which is by giving a value of t, see section 2. Calculation of the density requires approximations the first derivatives $K'(s)$ and second derivatives $K''(s)$ of the exponent function. Figure 2 is the density function for a Exp(5) variable with $\lambda = 1/5$ with mean 5 and variance 25.The saddlepoint normalizing constant ,is the area under the graph which is 1 for all nonnegative function [5],and the normalizing constant for exponent saddlepoint approximation is 1.06 which slightly deviates from 1 for t from 1 until 500.
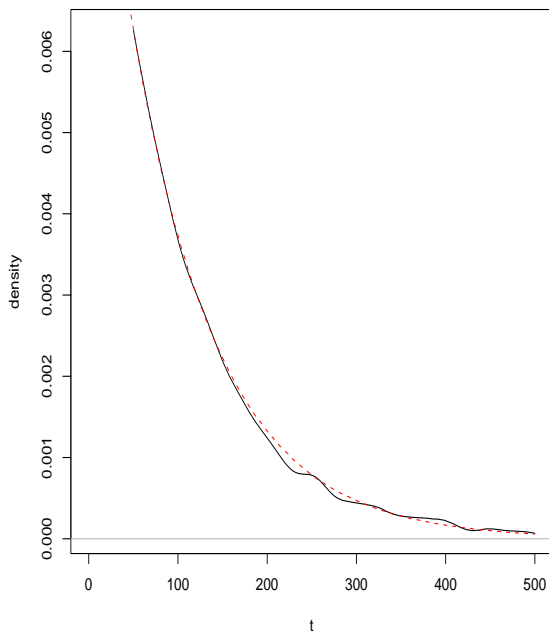

Fig. 2. Exponential density function saddlepoint approximation(dashed curve) and simulation(solid curve) of the observed gap M_z sample size 200,for 5000 simulation

B. Gamma

A gamma random variable X$\sim$ Gamma(α,β)with mean $\alpha\beta$ has a CGF given in section 2. Solution of the saddlepoint ap-

proximation density given in section 2 and for gamma density function requires solving the Incomplete gamma function, produce a saddlepoint that can be solved numerically which is by giving a value of t, see section 2. The computation of the density requires the first derivatives $K'(s)$ and second derivatives $K''(s)$ of the gamma function,using the gradient and hessian matrix[5]. The density function for X$\sim$ Gamma(11,60.5) with $\alpha = 2$,and $\beta = 5.5$ and X is the duration of the hypopnea .While Z is the observed gap, that was approximated using the saddlepoint approximation. Y$\sim$(45,675) with $\kappa = 3$,$\lambda = 15$, Y is the gap between hypoponea.The saddlepoint normalizing constant ,is 1.04 which deviates from 1 for t from 1 until 500.The M_z was compared with the simulation of the observed gap , (see Figure 3).
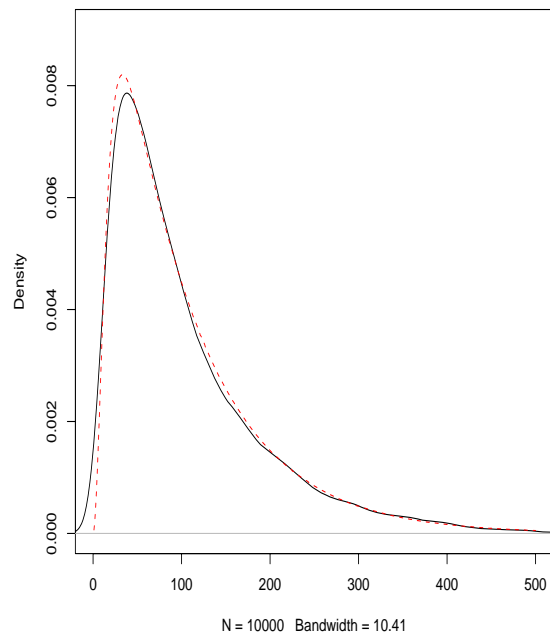

N = 10000   Bandwidth = 10.41

Fig. 3. Gamma density function from saddlepoint approximation(dashed curve) and simulation(solid curve) of the observed gap M_z sample size 500 for 5000 simulation

V. DISCUSSION AND CONCLUSION

We applied the saddlepoint to the MLEs in exponential families, but more general classes can be handled. For example for discrete distributions such as Geometric and Poisson distributions. In this paper we illustrate the use of saddlepoint to approximate a skewed distribution which is gamma distribution. From (5) where we take the transformation $K'_z(s) = t$. This transformation can be found in [7] and this transformation allows the evaluation of the integral with only the saddlepoint approximation using a numerical method to solve the expression (5) The normalizing constant deviates slightly from 1 which indicates that it is

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:2, No:2, 2008

close to the accurate approximation. However, the saddlepoint approximations have not yet received much attention in statistical applications. It is due to the computation involved difficulties to solve the $K'z(s) = t$ numerically.

REFERENCES

[1] Malhotra, A. & White, DP. (2002). Obstructive sleep apnea, *Lancet* **360**, 237-45
[2] Black J.(2003) Sleepiness and residual sleeponess in adults with obstructive sleep apnea. *Respiratory Physiology & Neurobiology*, **136**, 211-220
[3] Hogg R.V. and Tanis E.A. ,*Probability and Statistical Inference* New York: Macmillan Publishing,1977
[4] Huzurbazar A.,*Flowgraph Models for Multistate Time-to-Event Data* ,New York:John Wiley & Sons,2004
[5] Cox D.R., *Renewal Theory* . London:Methuen and Co., 1962
[6] Balakrishnan,N. & Mi, J.(2003) Existence and uniqueness of the MLEs for normal distribution based on general progressively Type-II censored samples *Statistics & Probability Letters*, **64**(4), 407-414
[7] Huzurbazar, S. (1999) Practical Saddlepoint Approximations *The American Statistician*, **53**(3),pp 225-232