The Korean Communications
in Statistics Vol. 5, No. 3, 1998
pp. 887-894

Conditional Skewness and Kurtosis in Natural Exponential Models[1]

Chong Sun Hong[2] and Han Seung Lim[3]

Abstract

Let $\underset{\sim}{T}=(T_1,\ldots,T_k;k\geq 2)$ be a minimal sufficient and complete statistic for a k -parameter exponential model. Consider a partition of $\underset{\sim}{T}$ into $(\underset{\sim}{T}_1, \underset{\sim}{T}_2)$, where $\underset{\sim}{T}_1=(T_1,\ldots,T_r)$ and $\underset{\sim}{T}_2=(T_{r+1},\ldots,T_k;1\leq r\leq k-1)$. This article represents a way to obtain higher moments such as skewness and kurtosis for the distribution $\underset{\sim}{T}$ and the conditional distribution of $\underset{\sim}{T}_1$, given $\underset{\sim}{T}_2=\underset{\sim}{t}_2$. These results are illustrated by some examples.

1. Introduction

For a random variable X, the family of distribution $\{f(x;\underset{\sim}{\eta});\underset{\sim}{\eta}\in\underset{\sim}{H}\}$ is known as a k parameters exponential family if there exist real-valued functions $\{c_i(\underset{\sim}{\eta}), i=1,\ldots,k\}$, $d(\underset{\sim}{\eta})$ on a parameter space $\underset{\sim}{H}$, and real-valued functions $\{T_i(x), i=1,\ldots,k\}$, $S(x)$ on R, and a set $A\subset R$ such that the pdf may be written as

$$f(x;\underset{\sim}{\eta})=\exp[\sum_i^k c_i(\underset{\sim}{\eta})T_i(x)-d(\underset{\sim}{\eta})+S(x)]I_A(x), \tag{1.1}$$

where I_A is the indicator function of set A. Note that $\underset{\sim}{T}(X)=(T_1(X),\ldots,T_k(X))$ is well-known as a natural sufficient statistic of the family. These detail explanations are found in most of theoretical textbooks of statistics including Mood, Graybill and Boes (1974), Bickel and Doksum (1976), Barndorff-Nielsen (1978), Lehmann (1983, 1986), and Lindgren (1993), etc.

If a random sample $\underset{\sim}{X}=(X_1,X_2,\ldots,X_n)$ are taken from $f(x;\underset{\sim}{\eta})$ in (1.1), then its joint pdf becomes

---

1) This paper was supported by Sungkyunkwan University 1997. Authors thank to an anonymous referee for valuable comments.

2) Professor, Department of Statistics, Sungkyunkwan University, Seoul, 110-745, Korea.
3) Lecturer, Department of Statistics, Sungkyunkwan University, Seoul, 110-745, Korea.

$$f(\underline{x};\underline{\eta}) = \exp[\sum_i^k c_i(\underline{\eta})\sum_j^n T_i(x_j) - nd(\underline{\eta}) + \sum_j^n S(x_j)]I_{A^n}(x_1,\ldots,x_n), \tag{1.2}$$

where A^n is a subset of R^n. Note that the distribution of $\underline{X}$ forms a k parameters exponential family with the natural sufficient statistic

$$\underline{T}(\underline{X}) = (\sum_j^n T_1(X_j),\ldots,\sum_j^n T_k(X_j)).$$

A reparameterization of the exponential family in (1.2) is obtained by setting $\theta_i = c_i(\underline{\eta})$. Then for a parameter function $l(\underline{\theta}) = l(\theta_1,\ldots,\theta_k) = d(\eta_1,\ldots,\eta_k)$, the exponential family has the following form :

$$f(\underline{x};\underline{\theta}) = [\exp[\sum_i\theta_i\sum_j T_i(x_j) - nl(\underline{\theta}) + \sum_j S(x_j)]]I_{A^n}(x_1,\ldots,x_n). \tag{1.3}$$

Note that if a function $c(\cdot)$ is one-to-one, then $l(\theta_i) = d(c_i^{-1}(\theta_i)), i=1,\ldots,k.$

For a suitable measurable function $g(\cdot)$ satisfying $g(\underline{t}) = \exp(\sum_j^n S(x_j))$, the joint pdf (1.3) can be expressed as

$$f(\underline{t};\underline{\theta}) = g(\underline{t})\exp[\underline{\theta}\cdot\underline{t} - nl(\underline{\theta})]I_{A_T}(\underline{t}), \tag{1.4}$$

where $\underline{\theta}\cdot\underline{t} = \sum_i\theta_i\sum_j T_i(x_j)$.

The moment generating function of $\underline{T}$ in (1.4) could be easily found as follows:

$$\Psi(\underline{s}) = E[\exp(\sum_{i=1}^k s_iT_i)] = \exp[nl(\underline{\theta}+\underline{s}) - nl(\underline{\theta})]. \tag{1.5}$$

Moments of $\underline{T}$ can be obtained by using the moment generating property of $\Psi(\underline{s})$. And cumulants of $\underline{T}$ can also be derived by differentiating $nl(\theta)$ in Equation (1.5). For example, the mean vector and variance-covariance matrix of $\underline{T}$ are found in $n\partial l(\underline{\theta})/\partial\underline{\theta}$ and $n\partial^2 l(\underline{\theta})/\partial\underline{\theta}'\partial\underline{\theta}$, respectively (see Kendall and Stuart 1969; Barndorff-Nielsen, 1978; Lehmann 1986 for more detail).

Let us consider a partition of a complete minimum sufficient statistic vector $\underline{T}$ into $(\underline{T}_1, \underline{T}_2)$, where $\underline{T}_1 = (T_1,\ldots,T_r)$ and $\underline{T}_2 = (T_{r+1},\ldots,T_k; 1\le r\le k-1)$. And define parameters vector $\underline{\theta} = (\underline{\theta}_1,\underline{\theta}_2)$, $\underline{\theta}_1 = (\theta_1,\ldots,\theta_r)$ and $\underline{\theta}_2 = (\theta_{r+1},\ldots,\theta_k; 1\le r\le k-1)$. The conditional distribution of $\underline{T}_1$, given $\underline{T}_2 = \underline{t}_2$, is

$$f_{\underline{T}_1|\underline{T}_2}(\underline{t}_1;\underline{\theta}_1|\underline{t}_2) = g(\underline{t})\exp[\underline{\theta}_1\cdot\underline{t}_1 - \log b(\underline{\theta}_1;\underline{t}_2)], \tag{1.6}$$

where $\underline{\theta}_1\cdot\underline{t}_1 = \sum_{i=1}^r\theta_i\sum_{j=1}^r T_i(x_j)$ and $b(\underline{\theta}_1;\underline{t}_2) \equiv f_{\underline{T}_2}(\underline{t}_2;\underline{\theta})\exp[nl(\underline{\theta}) - \underline{\theta}_2\cdot\underline{t}_2]$ which can be defined as a partial parameter function (see Shaul, 1994). This conditional pdf belongs to a natural exponential family with $\underline{\theta}_1$ as the vector of natural parameters (see Lehmann 1986, lemma 8, p. 58). Accordingly, $\log b(\underline{\theta}_1;\underline{t}_2)$ in (1.6) plays the same role as $nl(\underline{\theta})$ in (1.4). The conditional moment generating function of $\underline{T}_1$, given $\underline{T}_2=\underline{t}_2$, is obtained as the following:

$$\begin{aligned}\Psi_{\underline{T}_1|\underline{T}_2}(\underline{s}) &= E[\exp(s_1\cdot T_1)] \\ &= \exp(\log b(\underline{\theta}_1+\underline{s}_1;\underline{t}_2)-\log b(\underline{\theta}_1;\underline{t}_2)).\end{aligned} \tag{1.7}$$

With similar arguments, the conditional cumulants of $\underline{T}_1$, given $\underline{T}_2=\underline{t}_2$, can be derived by differentiating $\log b(\underline{\theta}_1;\underline{t}_2)$ with respect to $\underline{\theta}_1$. That is, the conditional mean vector and conditional variance-covariance matrix of $\underline{T}_1$, given $\underline{T}_2=\underline{t}_2$, are given by the matrix of the first and second-order derivatives of $\log b(\underline{\theta}_1;\underline{t}_2)$ with respect to $\underline{\theta}_1$, respectively (see Shaul, 1994 for more detail).

In this paper, we are interested in deriving higher moments of $\underline{T}$ by further differentiating $l(\underline{\theta})$ with respect to $\underline{\theta}$. The skewness and kurtosis of $\underline{T}$ will be discussed. And the conditional skewness and kurtosis of $\underline{T}_1$, given $\underline{T}_2=\underline{t}_2$, will be also formulated by third and fourth differentiations of $\log b(\underline{\theta}_1;\underline{t}_2)$ with respect to $\underline{\theta}_1$, respectively.

2. Conditional Skewness and Kurtosis

It can be obtained easily the third and fourth cumulants of $\underline{T}$ by third and fourth-order derivatives of $nl(\theta)$ in (1.5) such that, for $i=1,\ldots,k$

$$\begin{aligned}\frac{n\partial^3 l(\underline{\theta})}{\partial\theta_i^3} &= E[(T_i-\mu_i)^3], \\ \frac{n\partial^4 l(\underline{\theta})}{\partial\theta_i^4} &= E[(T_i-\mu_i)^4]-3\sigma_i^4.\end{aligned} \tag{2.1}$$

Then the skewness s_i and kurtosis k_i of the statistic T_i are presented in the following:

<Proposition 1>

$$s_i=\frac{n\partial^3 l(\underline{\theta})/\partial\theta_i^3}{(n\partial^2 l(\underline{\theta})/\partial\theta_i^2)^{3/2}} \quad\text{and}\quad k_i=\frac{n\partial^4 l(\underline{\theta})/\partial\theta_i^4}{(n\partial^2 l(\underline{\theta})/\partial\theta_i^2)^2} \qquad \text{for } i=1,\ldots,k.$$

Since the fourth cumulant of T_i has the form of second equation in (2.1), we can say that the kurtosis for a normal density has value 0 without any adjustments. Positive value of kurtosis indicates that a corresponding density is more peaked around its center than the density of a normal curve, and negative value means that a density is more flat around its center than the normal curve.

Similarly we can present the third and fourth cumulants of $\underline{T}_1$, given $\underline{T}_2=\underline{t}_2$, by third and fourth-order derivatives of $\log b(\underline{\theta}_1;\underline{t}_2)$ in (1.7). Then the analogous results for $\underline{T}_1$, given $\underline{T}_2=\underline{t}_2$, could be found as the followings:

$$\frac{\partial^3 \log b(\underline{\theta}_1;\underline{t}_2)}{\partial\theta_i^3} = E[(T_i-\mu_i)^3|\underline{T}_2=\underline{t}_2]$$
$$\frac{\partial^4 \log b(\underline{\theta}_1;\underline{t}_2)}{\partial\theta_i^4} = E[(T_i-\mu_i)^4|\underline{T}_2=\underline{t}_2]-3\sigma_i^{*4},$$

where $\sigma_i^{*2}=V(T_i|T_2=t_2)$ for $i=1,\dots,r$.

Then the conditional skewness (s_i^*) and kurtosis (k_i^*) of the statistic T_i for $i=1,\dots,r$ $(r<k)$, given $\underline{T}_2=\underline{t}_2$, are expressed in <Proposition 2>.

<Proposition 2>

$$s_i^*=\frac{n\partial^3\log b(\underline{\theta}_1;\underline{t}_2)/\partial\theta_i^3}{(n\partial^2\log b(\underline{\theta}_1;\underline{t}_2)/\partial\theta_i^2)^{3/2}} \text{ and } k_i^*=\frac{n\partial^4\log b(\underline{\theta}_1;\underline{t}_2)/\partial\theta_i^4}{(n\partial^2\log b(\underline{\theta}_1;\underline{t}_2)/\partial\theta_i^2)^2} \text{ for } i=1,\dots,r.$$

Also we can recognize that for $i=1,\dots,r$ $(r<k)$, the conditional kurtosis of the sufficient statistic T_i , given $\underline{T}_2=\underline{t}_2$, is based on value 0 as the same as that of the marginal T_i for $i=1,\dots,k$.

3. Examples

1) Bivariate Normal Distribution

Consider a random vector (X, Y) has a bivariate normal distribution with mean vector $\underline{\mu}=(\mu_1, \mu_2)$ and covariance matrix $\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$.

The joint pdf of n random sample $\{(X_1, Y_1),\dots,(X_n, Y_n)\}$ of (X, Y) has the form of Equation (1.4) which belongs to a five-parameter natural exponential family (see Freund and Walpole (1987), and Mood, Graybill and Boes (1974)). Let us partition the complete minimum sufficient statistic $\underline{T}=(\underline{T}_1,\underline{T}_2)$ and parameter vector $\underline{\theta}=(\underline{\theta}_1,\underline{\theta}_2)$ such that

$$\underline{T}_1=(T_1=\Sigma X_i^2,\ T_2=\Sigma Y_i^2,\ T_3=\Sigma X_iY_i),\qquad \underline{T}_2=(T_4=\Sigma X_i,\ T_5=\Sigma Y_i),$$

$$\underline{\theta}_1'=\begin{pmatrix}\theta_1 &=& -[2(1-\rho^2)\sigma_1^2]^{-1}\\ \theta_2 &=& -[2(1-\rho^2)\sigma_2^2]^{-1}\\ \theta_3 &=& \rho[(1-\rho^2)\sigma_1\sigma_2]^{-1}\end{pmatrix} \text{ and } \underline{\theta}_2'=\begin{pmatrix}\theta_4=(1-\rho^2)^{-1}\left(\frac{\mu_1}{\sigma_1^2}-\frac{\rho\mu_2}{\sigma_1\sigma_2}\right)\\ \theta_5=(1-\rho^2)^{-1}\left(\frac{\mu_2}{\sigma_2^2}-\frac{\rho\mu_1}{\sigma_1\sigma_2}\right)\end{pmatrix}.$$

From the pdf of $\underline{T}$, we obtain the parameter function

$$l(\underline{\theta}) = \frac{\theta_3\theta_4\theta_5-\theta_2\theta_4^2-\theta_1\theta_5^2}{(4\theta_1\theta_2-\theta_3^2)}-\frac{1}{2}\log(4\theta_1\theta_2-\theta_3^2).$$

Now suppose $\mu_1=\mu_2=0$, $\sigma_1=\sigma_2=1$ without any loss of generality. By differentiation of $nl(\underline{\theta})$, we can obtain the skewness vector and kurtosis vector of $\underline{T}$ as follows:

$$\begin{aligned} s_{\underline{T}} &=\left(2\sqrt{\frac{2}{n}},\ 2\sqrt{\frac{2}{n}},\ \frac{2\rho(3+\rho^2)}{(1+\rho^2)\sqrt{n(1+\rho^2)}},\ 0,\ 0\right)\\ k_{\underline{T}} &=\left(\frac{12}{n},\ \frac{12}{n},\ \frac{6(\rho^4+6\rho^2+1)}{n(1+\rho^2)^2},\ 0,\ 0\right). \end{aligned} \tag{3.1}$$

Also from the conditional pdf $f_{\underline{T}_1|\underline{T}_2}(\underline{t}_1;\underline{\theta}_1|\underline{t}_2)$, the logarithm partial parameter function can be obtained as

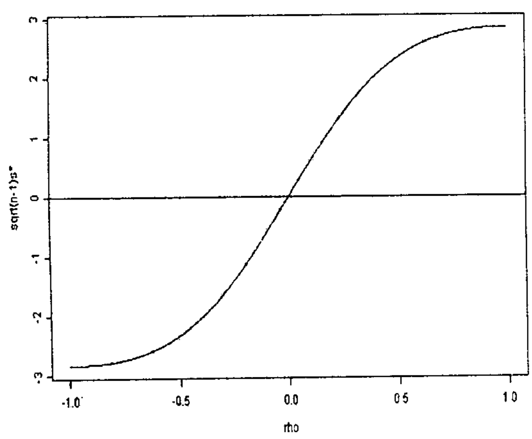
$$\log b(\underline{\theta}_1;\underline{t}_2)=-\log n+\frac{\theta_1t_4^2}{n}+\frac{\theta_2t_5^2}{n}+\frac{\theta_3t_4t_5}{n}-\frac{(n-1)}{2}\log(4\theta_1\theta_2-\theta_3^2).$$

Suppose $\mu_1=\mu_2=0$, $\sigma_1=\sigma_2=1$ without any loss of generality. By differentiation of $\log b(\underline{\theta}_1;\underline{t}_2)$, the conditional skewness vector and conditional kurtosis vector of $\underline{T}_1$, given $\underline{T}_2=\underline{t}_2$, are

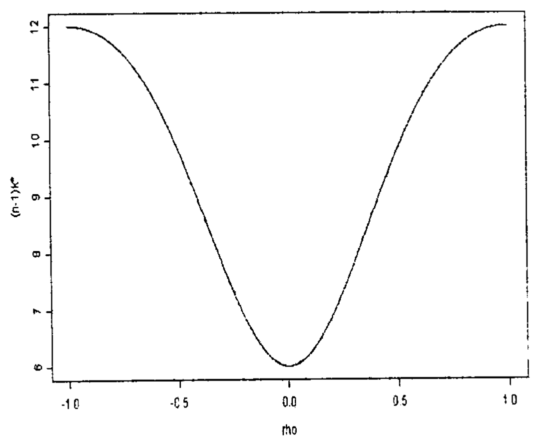
$$\begin{aligned} \underline{s}^*_{\underline{T}_1|\underline{t}_2} &=\left(2\sqrt{\frac{2}{(n-1)}},\ 2\sqrt{\frac{2}{(n-1)}},\ \frac{2\rho(3+\rho^2)}{(1+\rho^2)\sqrt{(n-1)(1+\rho^2)}}\right)\\ \underline{k}^*_{\underline{T}_1|\underline{t}_2} &=\left(\frac{12}{(n-1)},\ \frac{12}{(n-1)},\ \frac{6(1+6\rho^2+\rho^4)}{(n-1)(1+\rho^2)^2}\right). \end{aligned} \tag{3.2}$$

One can easily find that the difference between (3.1) and (3.2) is due to the degree of freedom, because the degree of freedoms of the conditional skewness and conditional kurtosis of $\underline{T}_1$, given $\underline{t}_2$, are all $n-1$. Especially, we are interested in the conditional skewness and the conditional kurtosis of $T_3=\sum_{i=1}^{n}X_iY_i$ given $T_4=\Sigma x_i$ $T_5=\Sigma y_i$. As sample size is increasing, the conditional kurtosis value decreases more rapidly than that of the conditional skewness. The relationships between the conditional skewness ($\sqrt{n-1}\, s^*_{T_3|\underline{t}_2}$) and the correlation coefficient (ρ), and the relationships between the conditional kurtosis

($(n-1)\ k^*_{T_3|t_2}$) and the correlation coefficient are represented in <Figure 1> and <Figure 2>, respectively. From <Figure 1> when ρ has a positive (negative) value, the value of the conditional distribution of $T_3=\sum_{i=1}^{n} X_iY_i$ given $T_4=\Sigma x_i$ $T_5=\Sigma y_i$ skews to the right (left). And as the value of ρ goes from -1 to +1, the value of $\sqrt{n-1}\, s^*_{T_3|t_2}$ increases from $-2\sqrt{2}$ to $2\sqrt{2}$, whose absolute value is that of the conditional skewness of $T_1=\sum X_i^2$ (or $T_2=\sum Y_i^2$). And <Figure 2> shows that the value of $(n-1)\ k^*_{T_3|t_2}$ converges to 12, which is the value of the conditional kurtosis of $\sum X_i^2$ (or $\sum Y_i^2$) when ρ goes to ± 1. $(n-1)\ k^*_{T_3|t_2}$ has the minimum value 6 at $\rho=0$.



<Figure 1>



<Figure 2>

2) Multinomial Distribution

Let $\underline{X}=(X_1, \dots, X_k)$ be a random vector that has a multinomial distribution with m trial and cell probabilities $p_1, \dots, p_k$, where $\sum_{i=1}^{k} X_i=m$, $p_i>0$ $(i=1, \dots, k)$ and $\sum_{i=1}^{k} p_i=1$. The corresponding pdf constitutes a $k-1$ parameter natural exponential family with $T_i=X_i$ and $\theta_i=\log(p_i/p_k)$, for $i=1,\dots,k-1$. Then the parameter function of $\underline{\theta}$ and the logarithm partial parameter function of θ_i are, respectively, $l(\underline{\theta})=m\log(\sum_{i=1}^{k-1} e^{\theta_i}+1)$, and

$$\log b(\theta_i;x_j)=\log\binom{m}{x_j}+(m-x_j)\log(1+e^{\theta_1}+\cdots+e^{\theta_{j-1}}+e^{\theta_{j+1}}+\cdots+e^{\theta_{k-1}}),$$

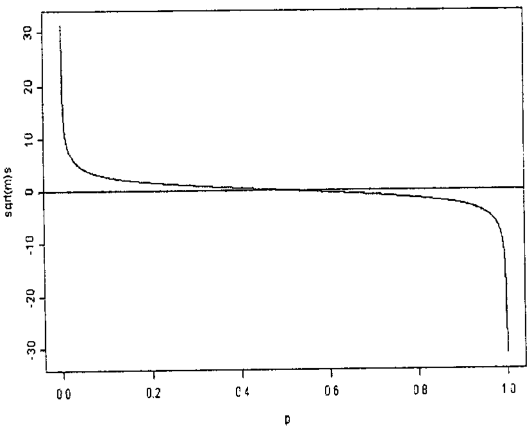
for $i\neq j$.

Then by the differentiation of $l(\underline{\theta}) = m\log(\sum_{i=1}^{k-1} e^{\theta_i} + 1)$, the skewness and kurtosis of X_i could be obtained as the followings:

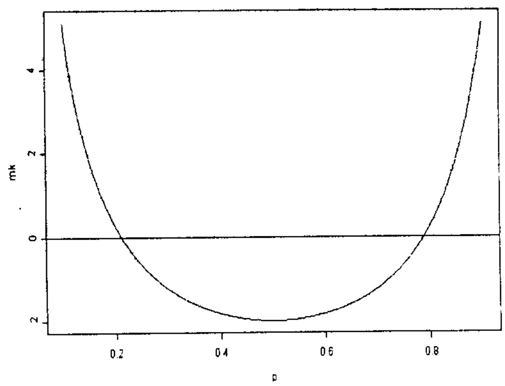
$$\begin{aligned} s_{X_i} &= \frac{(1-2p_i)}{\sqrt{mp_i(1-p_i)}}, \\ k_{X_i} &= \frac{(1-6p_i+6p_i^2)}{mp_i(1-p_i)}. \end{aligned} \tag{3.3}$$

And from $\log b(\theta_i; x_j)$, we can get the conditional skewness and kurtosis of X_i, given $X_j = x_j$ ($i \neq j$), such that

$$\begin{aligned} s^*_{X_i|x_j} &= \frac{(1-p_j-2p_i)}{\sqrt{(m-x_j)p_i(1-p_i-p_j)}}, \\ k^*_{X_i|x_j} &= \frac{(1-p_j)^2 - 6p_i(1-p_j) + 6p_i^2}{(m-x_j)p_i(1-p_i-p_j)}. \end{aligned} \tag{3.4}$$



<Figure 3>



<Figure 4>

From the conditional skewness and kurtosis of X_i given x_j ($i \neq j$) in (3.4), one could recognize the fact that the total trial number m and total probability 1 in (3.3) are replaced with $m - x_j$ and $1 - p_j$, respectively. So we would like to explore the skewness and kurtosis for the multinomial distribution defined in (3.3). The relationships between the skewness and kurtosis ($\sqrt{m}\ s_{X_i}$ and $m\ k_{X_i}$), and p_i are represented in <Figure 3> and <Figure 4>, respectively. When p_i is less (greater) than 0.5 ($0.5 - p_j/2$ for the conditional skewness), the skewness has positive (negative) value, so that one can say that the distribution of the multinomial random variable X_i skews to the right (left). From <Figure 4>, the value of the kurtosis get larger when p_i approaches near 0 or 1. And the kurtosis of the multinomial variable X_i has negative value when p_i belongs to $(\frac{1}{2}(1-\frac{1}{\sqrt{3}}),\ \frac{1}{2}(1+\frac{1}{\sqrt{3}}))$ = (0.21, 0.79). For the conditional kurtosis of X_i given x_j ($i \neq j$), this interval has changed into $(\frac{1-p_j}{2}(1-\frac{1}{\sqrt{3}}),\ \frac{1-p_j}{2}(1+\frac{1}{\sqrt{3}}))$.

References

[1] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*, Chichester, England, John Wiley & Sons, New York.

[2] Bickel, P. J. and Doksum, K. (1976). *Mathematical Statistics*, John Wiley, New York.

[3] Freund, J, E. and Walpole, R. E. (1987). *Mathematical Statistics* (4th ed.), Englewood Cliffs, NJ: Prentice-Hall.

[4] Kendall, M. G. and Stuart, A. (1969). *The Advanced Theory of Statistics*, Volume 1 : Distribution Theory (3rd ed), Hafner.

[5] Lehmann, E . L. (1983). *Testing of point estimation*, John Wiley & Sons, New York.

[6] Lehmann, E . L. (1986). *Testing Statistical Hypothesis* (2nd ed.), John Wiley & Sons, New York.

[7] Lindgren, B. W. (1993). *Statistical Theory*, Fourth Edition, Chapman & Hall.

[8] Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). *Introduction to the Theory of Statistics*, Third Edition, McGraw-Hill.

[9] Shaul K. B. (1994). A Derivation of conditional cumulants in Exponential Model, *The American statistician*, May, Vol. 48, No. 2, 126-129.