IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ORDER STATISTICS FROM A CLASS OF SYMMETRIC DISTRIBUTIONS WITH KURTOSIS THREE

Dr. James Kurian

Associate Professor, Department of Statistics, Maharaja's College, Ernakulam, Kerala, India, PIN- 682011

Abstract

In this paper, I consider the class of continuous distributions, belonging to the symmetric location scale family $\{F[(x-\theta)/\sigma], -\infty < \theta < \infty, \sigma > 0\}$, with Pearson's measure of kurtosis $\beta_2 = 3$. The study uses a sub class of symmetric mesokurtic distributions based on the Edgeworth expansions to the probability density function (p.d.f.) based on the first six moments. This paper derives the distributions of order statistics from such a family of distributions. Moreover, based on a sample of size n, the distribution of minimum, maximum and the form of their moments are also derived. The class of distributions considered here can be used as an alternative to the normal error in linear models and robustness studies.

Key words-*Edgeworth expansions, Hermite Polynomial, Kurtosis, Location Scale family, Order statistics, Robustness*.

Introduction

Assume that a continuous random variable X is distributed symmetric about θ and its distribution function (d.f.) belongs to the location scale family $\{F[(x-\theta)/\sigma], -\infty < \theta < \infty, \sigma > 0\}$. Note that the normal distribution $N(\theta, \sigma^2)$, $-\infty < \theta < \infty, \sigma > 0$ belongs to this family, which has a unique place as error distribution in robustness studies and linear models. Various alternatives to the normal model distributions in the measurement error model $X = \theta + \epsilon$ have been considered in this context by many authors, where ϵ is a random variable with mean zero is and variance σ^2. For example, Box and Tiao (1973) considers the exponential power family in which the normal is a particular case. Various types of contaminated mixture models are considered in Huber (1981). Among the Pearson's family of p.d.f.'s normal distribution is characterized by the property that $\beta_1 = 0$ and $\beta_2 = 3$, where where $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ and $\beta_2 = \frac{\mu_4}{\mu_2^2}$, the Pearson's measure of skewness and kurtosis and μ_k is the kth central moment of X. This property is often used in the literature in order to detect the departure of a distribution from normality. Various types of tests for normality have been developed based on sample skewness b_1 and sample kurtosis b_2 (See for example Pearson (1966) and D' Agostino and Tietjen (1973)). Kale and Sebastian (1995) introduced a wide class of non-normal symmetric distributions with $\beta_1 = 0$ and $\beta_2 = 3$ denoted by KS class. This class of p.d.fs is a class of mixture distributions of the form $F_\alpha = aH + (1-\alpha)G$, where H is symmetric around θ with $\beta_2 < 3$ and G is symmetric symmetric around θ with $\beta_2 > 3$. They showed that there exists a unique $a \epsilon (0,1)$ such that β_2 of F_α is equal to three. Further, if F_α is a symmetric distribution with kurtosis three, then $\gamma H + (1-\gamma)\Phi$ is symmetric with $\beta_1 = 0$ and $\beta_2 = 3$ for all $\gamma\epsilon[0,1]$, where Φ is the distribution function of a normal.

In the linear model $X = \theta + \epsilon$, where ϵ follows a symmetric distribution with $V(\epsilon) = \sigma^2$ $\beta_1 = 0$ and $\beta_2 = 3$ (ie, $X \sim KS(\theta, \sigma^2)$) has several optimal properties as in the case of $\epsilon \sim N(\theta, \sigma^2)$. Godambe and Thompson (1989) showed a semi-parametric set-up in which first four moments are known functions of θ and σ^2, the optimal estimating equation for (θ, σ^2) based on the samples $(X_1, X_2, \dots, X_n)$ coincides with that of the likelihood equations of θ and σ^2 in $N(\theta, \sigma^2)$ whenever, $\beta_1 = 0$ and $\beta_2 = 3$. We know that $\bar{X} = \frac{1}{n}\sum X_i$ and $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$ are UMVUEs of θ and σ^2 whenever, $X \sim N(\theta, \sigma^2)$. Further, $Var\,(\bar{X}) = \frac{\sigma^2}{n}$ and $Var\,(S^2) = \frac{2\sigma^4}{n-1}$. If $\epsilon \sim KS(\theta, \sigma^2)$, it can be shown that the variances of $\bar{X}$ and S^2 are the same as that of normal case even though, they are not UMVUE's of θ and σ^2 respectively (James Kurian and Sebastian, G. (2005)).

Distribution of order statistic

Consider the family of symmetric distributions introduced by Sebastian, G. and James Kurian (2002) of the form;

$$f(x) = \left\{1 + \frac{\varepsilon_6}{720} H_6(x)\right\} \phi(x)\, , -\infty < x < \infty \qquad (1)$$

where ε_6 is the sixth standardized cumulant which we assume to exist, $\phi(x)$ is the p.d.f. of a standard normal and $H_6(x)$ is the sixth degree Hermite polynomial given by $H_6(x) = x^6 - 15x^4 + 45x^2 - 15$.

The distribution function of the above p.d.f. is

$$F(x) = \Phi(x) + \frac{\varepsilon_6}{720} H_5(x)\, \phi(x)\, , \text{ where } \Phi(x) = \int_{-\infty}^{x} \phi(t)dt$$

$$= \Phi(x)\left(1 + \frac{\varepsilon_6}{720} H_5(x) T(x)\right)$$

Hence, $$1 - F(x) = (1 - \Phi(x))\left(1 - \frac{\varepsilon_6}{720} P(x) H_5(x) T(x)\right)$$

$$\text{where, } P(x) = \frac{\Phi(x)}{1-\Phi(x)} \text{ and } T(x) = \frac{\phi(x)}{\Phi(x)}$$

$$F(y) - F(x) = (\Phi(y) - \Phi(x))\left(1 + \frac{\varepsilon_6}{720}\sigma_5(x,y)\right), \text{ where } \sigma_5(x,y) = \frac{H_5(y)\phi(y) - H_5(x)\phi(x)}{\Phi(y) - \Phi(x)}$$

Let $x_{(i)} = y_i$ (i=1,2, … ,n), then the p.d.f. of $(y_1, y_2, \dots, y_n)$ is

$$g(y_1, y_2, \dots, y_n) = n!\,\left[\prod_{i=1}^{n} \phi(y_i)\right] \prod_{i=1}^{n}\left[1 + \frac{\varepsilon_6}{720} H_6(y_i)\right]$$

The distribution of the rth order statistic $g_r(r)$ is given by

$$g_r(r) = \frac{n!}{(r-1)!(n-r)!} f(y_r)[F(y_r)]^{r-1}[1 - F(y_r)]^{n-r} \qquad (2)$$

For the p.d.f. in equation (1)

$$[F(y_r)]^{r-1} = [\Phi(y_r)]^{r-1}\left[1 + \frac{\varepsilon_6}{720} H_5(y_r) T(y_r)\right]^{r-1}$$

$$\approx [\Phi(y_r)]^{r-1}\left[1 + (r-1)\frac{\varepsilon_6}{720} H_5(y_r) T(y_r)\right] \qquad \text{and}$$

$$[1-F(y_r)]^{n-r}=[1-\Phi(y_r)]^{n-r}\left[1-\frac{\varepsilon_6}{720}H_5(y_r)T(y_r)\right]^{n-r}$$

$$\approx [1-\Phi(y_r)]^{n-r}\left[1-\frac{(n-r)\varepsilon_6}{720}P(y_r)H_5(y_r)T(y_r)\right]$$

Substituting the above values in equation (2) we get

$$g_r(y_r)=\frac{n!}{(r-1)!(n-r)!}\,\phi(y_r)[\Phi(y_r)]^{r-1}[1-\Phi(y_r)]^{n-r}\left[1+(r-1)\frac{\varepsilon_6}{720}H_5(y_r)T(y_r)\right]$$

$$\left[1-\frac{(n-r)\varepsilon_6}{720}P(y_r)H_5(y_r)T(y_r)\right]$$

$$=\frac{\phi(y_r)\Phi^{r-1}(y_r)\left(1-\Phi(y_r)\right)^{n-r}}{B(r,n-r+1)}\left[1+\frac{\varepsilon_6}{720}H_6(y_r)\right]$$

$$\left[1-\frac{(r-1)\varepsilon_6}{720}\frac{\phi(y_r)}{\Phi(y_r)}H_5(y_r)\right]\left[1-\frac{(n-r)\varepsilon_6}{720}\frac{\phi(y_r)}{1-\Phi(y_r)}H_5(y_r)\right]$$

If we denote $g_r^{(N)}(y)=\frac{\phi(y)\Phi^{r-1}(y)\left(1-\Phi(y)\right)^{n-r}}{B(r,n-r+1)}$ then we have

$$g_r(y)=g_r^{(N)}(y)\left[1+\frac{\varepsilon_6}{720}H_6(y)\right]$$

$$\left[1-\frac{(r-1)\varepsilon_6}{720}\frac{\phi(y)}{\Phi(y)}H_5(y)\right]\left[1-\frac{(n-r)\varepsilon_6}{720}\frac{\phi(y)}{1-\Phi(y)}H_5(y)\right]$$

Therefore,

$$g_r(y)\approx g_r^{(N)}(y)\left[1+\frac{\varepsilon_6}{720}H_6(y)+\frac{(r-1)\varepsilon_6}{720}\frac{\phi(y)}{\Phi(y)}H_5(y)-\frac{(n-r)\varepsilon_6}{720}\frac{\phi(y)}{1-\Phi(y)}H_5(y)\right]$$

Moments

Define $\Psi(s,\alpha,\beta,\gamma)=\int_{-\infty}^{\infty}x^s\ \phi^\alpha(y)\,\Phi^\beta(y)\left(1-\Phi(y)\right)^\gamma dy$

Then the sth moment is $E_{g_r^{(N)}}(x^s)=\Psi(s,1,r-1,n-r)$

Let $E_{g_r}(x^t)=\mu'_{t|r}$, then

$$\mu'_{t|r}=\frac{1}{B(r,n-r+1)}\Big[\Psi(t,1,r-1,n-r)+\frac{\varepsilon_6}{720}\{\Psi(t+6,1,r-1,n-r)-15\Psi(t+4,1,r-1,n-r)+45\Psi(t+2,1,r-1,n-r)-15\Psi(t,1,r-1,n-r)\}-\frac{(n-r)\varepsilon_6}{720}\{\Psi(t+5,2,r-2,n-r-1)\}+\frac{(r-1)\varepsilon_6}{720}\{\Psi(t+5,2,r-2,n-r)-10\Psi(t+3,2,r-2,n-r)+15\Psi(t+1,2,r-2,n-r)-10\Psi(t+3,2,r-1,n-r-11)+15\Psi(t+1,2,r-1,n-r-1)\}\Big] \qquad (3)$$

Distribution of the order statistic and its moments

The distributions of the maximum and the minimum, based on a random sample of size n are derived below:

Distribution of the maximum

The distribution of the nth order statistic is given by $g_n(y_n)=nf(y_n)F^{n-1}(y_n)$

Therefore $g_n(y_n)=n\left[1+\frac{\varepsilon_6}{720}H_6(y_n)\right]\ \phi(y_n)\Phi^{n-1}(y_n)\left[1-\frac{\varepsilon_6}{720}H_5(y_n)T(y_n)\right]^{n-1}$

$$= n\Phi^{n-1}\,\phi(y_n)\left[1-\frac{\varepsilon_6}{720}H_5(y_n)T(y_n)\right]^{n-1}\left[1+\frac{\varepsilon_6}{720}H_6(y_n)\right]$$

$$\approx\; n\Phi^{n-1}\,\phi(y_n)\left[1+\frac{\varepsilon_6}{720}H_6(y_n)+\frac{(n-1)\varepsilon_6}{720}H_5(y_n)T(y_n)\right]$$

$$= g_n^{(N)}(y_n)\left[1+\frac{\varepsilon_6}{720}H_6(y_n)+\frac{(n-1)\varepsilon_6}{720}\frac{\phi(y_n)}{\Phi(y_n)}H_5(y_n)\right]$$

Therefore, using equation (3), the tth moment is

$$\mu'_{t|n} = \frac{1}{B(r,n-r+1)}\Big[\Psi(t,1,n-1,0)+\frac{\varepsilon_6}{720}\{\Psi(t+6,1,n-1,0)-15\Psi(t+4,1,n-1,0)+45\Psi(t+2,1,n-1,0)-15\Psi(t,1,n-1,0)\}+\frac{(n-1)\varepsilon_6}{720}\{\Psi(t+5,2,n-2,0)-10\Psi(t+3,2,n-2,0)+15\Psi(t+1,2,n-2,0)\Big]$$

Distribution of the minimum

$$g_1(y_1) = nf(y_1)\big(1-F(y_1)\big)^{n-1}$$

$$\approx n\,\phi(y_1)(1-\Phi(y_1))^{n-1}\left[1+\frac{\varepsilon_6}{720}H_6(y_1)+\frac{(n-1)\varepsilon_6}{720}H_5(y_1)\right]$$

$$= g_1^{(N)}(y_1)\left[1+\frac{\varepsilon_6}{720}H_6(y_1)-\frac{(n-1)\varepsilon_6}{720}\frac{\phi(y_1)}{1-\Phi(y_1)}H_5(y_1)\right]$$

Similarly tth moment

$$\mu'_{t|1} = \frac{1}{B(n,n-r+1)}\Big[\Psi(t,1,0,n-1)+\frac{\varepsilon_6}{720}\{\Psi(t+6,1,0,n-1)-15\Psi(t+4,1,0,n-1)+45\Psi(t+2,1,0,n-1)-15\Psi(t,1,0,n-1)\}-\frac{(n-1)\varepsilon_6}{720}\{\Psi(t+5,2,0,n-2)-10\Psi(t+3,2,0,n-2)+15\Psi(t+1,2,0,n-2)\Big]$$

Characteristic function

Let $x_1, x_2, \dots, x_n$ be n random samples from the distribution defined by the density function $(x) = \left\{1+\frac{\varepsilon_6}{720}H_6(x)\right\}\phi(x)$, $-\infty < x < \infty$, then the characteristic function $\phi_x(t)$ is

$$\phi_x(t) = E\left(e^{itx}\right) = \int_{-\infty}^{\infty} e^{itx}\left\{1+\frac{\varepsilon_6}{720}H_6(x)\right\}\phi(x)\;dx$$

$$= \left(1+\frac{\varepsilon_6}{720}\,(it)^6\right)e^{-\frac{t^2}{2}}$$

Result- If $= \sum_{j=1}^{n} l_j x_j$, where $l_1, l_2, \dots, l_n$ are some real constant then the characteristic function of y is

$$\phi_y(t) = \prod_{j=1}^{n}\phi_{x_j}(l_j t) = \prod_{j=1}^{n}\left(1+\frac{\varepsilon_6}{720}{l_j}^6\,(it)^6\right)e^{\sum_{j=1}^{n} l_j^2 (it)^6}$$

Conclusion

Study derived the distribution and expression for mean of order statistics from a class of symmetric mesokurtic distributions with kurtosis three. Functional form of the characteristic function from the above family distributions also derived in this study.

Reference

Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Don Mills, Anderson-Wesley.

Cramer, H. (1962). *Mathematical Methods of Statistics*. Princeton University Press.

Godambe, V.P. and Thompson, M.E.(1989). An extension of quasi likelihood estimation. *J. Statist. Plann. Infer*, 22,137-172.

Huber, P.J. (1981). *Robust Statistics*, New York Wiley.

James Kurian and Sebastian, G. (2005). A comparative study between a class of Mesokurtic Distributions and Normal model. *STARS Int. Journal*. Vol. 6 17-32.

James Kurian and Sebastian, G. (2007). On a Class of Symmetric Mesokurtic Distributions. *International Journal of Mathematical Sciences* (Special Issue “Recent Trends in Computational Mathematics"), Vol. 6. (Serials Publications, ISSN: 0972-754X.

Johnson, N.L. and Kots, S. (1970). *Continuous univeriate distributions-1*. John Wiley and Sons.

Kale, B.K. and Sebastian, G. (1996). On a class of Non-normal Symmeric Distibutions with Kurtosis three. *Statistical Theory and Applications, Paper in honour of H.A. David Ed. H.N. Nagaraja, P.K. Sen and P.R. Morrison, Springer-Verlag*, New York 55-63.

Kale, B.K. (1999). *A first course on Parametric Inference*. Narosa Publishing House, New Delhi.

Sebastian, G. and James Kurian (2002). On robustness of mean for non-normal symmetric distributions with kurtosis three. *Journal of Indian Statistical Association*. Vol. 40, 1,71-79