



RESEARCH ARTICLE

ON THE MEAN AND VARIANCE OF A FUNCTION OF ORDER STATISTICS.

Ehtesham Hussain¹, Muhammad Ahsanuddin² and Masood Ul Haq³.

1. Department of Statistics, University of Karachi, Pakistan.
2. Department of Economics, University of Karachi, Pakistan.
3. Usman Institute of Technology, Karachi, Pakistan.

Manuscript Info

Manuscript History

Received: 20 January 2019

Final Accepted: 22 February 2019

Published: March 2019

Key words:-

Order statistics, Statistical Transformation, Distribution Theory, Taylor's expansion.

Abstract

Order statistics deals with the properties and applications of random variables and their functions. In this study the distribution of order statistics for samples is derived and Taylor's expansion technique is applied for the calculation of moments of the functions. The transformation of random variable X , i.e. $h(X)$ has been computed in several ways as well and it is found that the variance of $\ln(X_{(m+1)})$ is negligible as compared to the variance of $X_{(m+1)}$. In this paper the probability distribution of $\ln(X_{(m+1)})$ is completely evaluated for large samples.

Copy Right, IJAR, 2019,. All rights reserved.

Introduction:-

Order statistics is important to highlight significant properties of a sample, for instance its location, dispersion and skewness (Buitendag, 2014). Bairamov et al. (2001) investigated the finite sample and asymptotic distributions of the statistics based on order statistics as well as their use in statistical inference. The concept of order statistics is common in many areas as it makes use of data dealing with natural problems related to sciences and life sciences (David & Nagaraja, 2004; Menzel & Morganti, 2013; Balakrishnan, & Cohen, 2014; Parrado-Gallardo et al., 2014; Lück & Wolf, 2016).

Likewise, in empirical economic studies, the size of parameters across different populations is often compared to measure for instance relative growth rates between states, wage discriminations between firms and technical efficiencies amongst production units which involves handling of huge numbers of populations (Horrace, 2000; Kampelmann & Rycx, 2016; Grinza et al., 2018). In this process transformation of data is widely used to assist in the interpretation of data, to derive a pivot for the construction of confidence intervals, testing of hypotheses, for stabilizing the variance in an ANOVA set up, etc. (Gumbel, 2012).

Box and Cox (1964) developed a transformation for transforming data to a normal model. Transformations have been used in applied sciences with success (Carroll and Ruppert, 1988; Cohen, 2001; Lütkepohl & Xu, 2012). Chao et al. (2001) estimate the impact of data transformation on forecasting models by using Monte Carlo experiments.

The transformations on the sample mean, \bar{X} , have been frequently derived for a given probability distribution $f(x; \theta)$, using Taylor's expansion.

If $X_i \sim \text{Binomial}(1, p)$ with $E(X) = p$; $V(X) = p(1 - p)$

Corresponding Author:-Ehtesham Hussain.

Address:-Department of Statistics, University of Karachi, Pakistan.

Let $h(p) = p(1-p)$ and $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ is the sample mean then expected value and variance given by

$$E[\bar{X}(1-\bar{X})] = \frac{(n-1)}{n} p(1-p) \quad (1)$$

$$V[\bar{X}(1-\bar{X})] = \frac{p(1-p)}{n} \left\{ (1-2p)^2 + \frac{1}{n} [2p(1-p) - 2(1-p)^2] \right\} \quad (2)$$

with an error of order $\frac{1}{n^3}$.

Variance stabilizing transformation, i.e. $\sigma^2[h'(\mu)]^2 = C$, where σ^2 is the variance of the given distribution and $h'(\mu)$ is the derivative of $h(x)$ at $X=\mu$. This differential equation, with $h'(\mu) = h'(\sqrt{\bar{X}})$ when applied to Poisson probability function yields: $V(\sqrt{\bar{X}}) \cong \frac{1}{4n}$ such that $2\sqrt{n}(\sqrt{\bar{X}} - \sqrt{\lambda})$ has approximate normal distribution $N(0, 1/4)$.

Taylor's method has been used in the present paper for the approximate complex calculation of moments of the functions of order statistics, $h(X_{(i)})$, for a random sample of size n , and Taylor's expansion has been done around the median of the given distribution ($\tilde{\mu}$) instead of the mathematical expectation (μ).

Theory Of Transformation

Suppose a probability distribution is given as $f(x; \theta)$ and assume $h(X_{(i)})$ is considered where $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ are the ordered variables. Then

$$h(X_{(i)}) \cong h(\tilde{\mu}) + (X_{(i)} - \tilde{\mu})h'(\tilde{\mu}) + (X_{(i)} - \tilde{\mu})^2 \frac{h''(\tilde{\mu})}{2} \quad (3)$$

The higher powers of $(X_{(i)} - \tilde{\mu})$ have been ignored.

Thus, applying expectation on both the sides:

$$E[h(X_{(i)})] \cong h(\tilde{\mu}) + E[(X_{(i)} - \tilde{\mu})]h'(\tilde{\mu}) + E[(X_{(i)} - \tilde{\mu})^2] \frac{h''(\tilde{\mu})}{2} \quad (4)$$

For the mathematical expectation of $h(X_{(i)})$, we require $E(X_{(i)})$, which would be calculated by the probability distribution of $X_{(i)}$ in a random sample size n , i.e.

$$f(X_{(i)}) = K(F(X_{(i)})^{i-1})(1-F(X_{(i)}))^{n-i+1}f(X_{(i)}) \quad (5)$$

For the variance of $h(X_{(i)})$, we have

$$V[h(X_{(i)})] = E[h^2(X_{(i)})] - (E[h(X_{(i)})])^2.$$

To obtain $E[h^2(X_{(i)})]$, squaring both sides in the expansion of $h(X_{(i)})$ using (3) and applying expectation we have

$$E[h^2(X_{(i)})] \cong h^2(\tilde{\mu}) + 2E[(X_{(i)} - \tilde{\mu})]h'(\tilde{\mu})h(\tilde{\mu}) + E[(X_{(i)} - \tilde{\mu})^2]\{[h'(\tilde{\mu})]^2 + h(\tilde{\mu})h''(\tilde{\mu})\}$$

$$V[h(X_{(i)})] = CVar(\text{median})$$

The model of $h(X_{(i)})$ would be different for each $X_{(i)}$, in a given probability distribution.

Example:

Consider the exponential distribution:

$$f(x) = \lambda \text{Exp}(-\lambda) \quad X > 0 \quad \lambda > 0$$

defining $h(X_{(i)})$ as function of $\ln(X_{(i)})$ where $X_{(i)}$ for $i=m$ is the median in a random sample of size $n = 2m+1$. Thus

$$X_{(m+1)} \text{ is the median. The median of given exponential model is } \tilde{\mu} = \frac{\ln(2)}{\lambda} \text{ and } h(\tilde{\mu}) = \ln\left(\frac{\ln(2)}{\lambda}\right); h'(\tilde{\mu}) = \frac{\lambda}{\ln(2)};$$

$$h''(\tilde{\mu}) = \frac{-\lambda^2}{(\ln(2))^2}.$$

The probability density function (pdf) of $X_{(m+1)}$ by expression (5) provides $E(X_{(m+1)})$:

$$E(X_{(m+1)}^r) = \frac{(2m+1)!}{(m!)^2} \frac{(-1)^r}{\lambda^r} \int_0^1 (\ln(y))^r y^m (1-y)^m dy; \quad r = 1, 2, 3, \dots$$

The following tables give the computed values of $E(X_{(m+1)})$ and $V(X_{(m+1)})$

The computed results of (median) of exponential distribution:

Table 1:-The expected values of median $X_{(m+1)}$ of exponential distribution: Median values of exponential distribution for different values of parameter λ .

λ	0.5	1.0	1.5	2.0	2.5
$\tilde{\mu}$	1.3862	0.6931	0.4621	0.3466	0.2773

Table 2:-The expected values of median of exponential distribution for different values of sample size n and parameter λ .

$E(X_{(m+1)})$						
n	m	$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 1.5$	$\lambda = 2.0$	$\lambda = 2.5$
5	2	1.5666	0.7833	0.5222	0.3917	0.3133
7	3	1.5190	0.7595	0.5063	0.3796	0.3038
9	4	1.4913	0.7456	0.4971	0.3728	0.2983
11	5	1.4731	0.7365	0.4910	0.3682	0.2946
13	6	1.4602	0.7301	0.4867	0.3650	0.2920

Table 3:-The variance values of median of exponential distribution for different values of sample size n and parameter λ .

$V[X_{(m+1)}]$						
n	m	$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 1.5$	$\lambda = 2.0$	$\lambda = 2.5$
5	2	0.8544	0.2136	0.0949	0.0534	0.0342
7	3	0.6027	0.1507	0.0670	0.0377	0.0241
9	4	0.4646	0.1162	0.0516	0.0290	0.0186
11	5	0.3777	0.0944	0.0496	0.0236	0.0151
13	6	0.3180	0.0795	0.0318	0.0199	0.0127

Now substituting these values in the expressions (4) and (6) to obtain the mean and variance of function of $\ln(x)$
The computed results of Log(median) of exponential distribution:

Table 4:-The Log (median) values of exponential distribution for different values of parameter λ .

λ	0.5	1.0	1.5	2.0	2.5
$\ln(\tilde{\mu})$	0.3266	-0.3665	-0.7720	-1.0520	-1.2828

The expected values of Log (median) of exponential distribution:

Table 5:-The expected values of log (median) of exponential distribution for different values of sample size n and parameter λ .

$E[\ln(X_{(m+1)})]$						
n	m	$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 1.5$	$\lambda = 2.0$	$\lambda = 2.5$
5	2	0.2676	-0.4255	-0.8310	-1.1187	-1.3418
7	3	0.2835	-0.4059	-0.8152	-1.1028	-1.3260
9	4	0.2926	-0.4005	-0.8060	-1.0937	-1.3168
11	5	0.2986	-0.3946	-0.8000	-1.0877	-1.3109
13	6	0.30276	-0.3938	-0.7958	-1.0835	-1.3067

The variance values of Log (median) of exponential distribution:

Table 6: The variance values of log (median) of exponential distribution for different values of sample size n and parameter λ .

$V[\ln(X_{(m+1)})]$						
n	m	$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 1.5$	$\lambda = 2.0$	$\lambda = 2.5$
5	2	0.4060	0.4060	0.4060	0.4056	0.4060
7	3	0.2927	0.2927	0.2927	0.2927	0.2927
9	4	0.2287	0.2287	0.2287	0.2287	0.2287
11	5	0.1876	0.1876	0.1876	0.1876	0.1876
13	6	0.1590	0.1590	0.1590	0.1590	0.1590

Simulation: To confirm the results, we performed a simulation study (100 replications) for samples of size $n=5, 7, 9, 11$ and 13 . The samples were generated from the exponential distribution $F(x) = 1 - \exp(-x/\lambda)$, $x > 0$ with $\lambda = 0.5, 1.0, 1.5, 2.0$ and 2.5 . The results are given in following tables. The simulations are produced by using software package Minitab (Ryan & Joiner, 1994).

Table 7:-Simulated values of $E[x_{(m+1)}]$ and $V[x_{(m+1)}]$ for different values of parameter λ .

$E[x_{(m+1)}]$ and $V[x_{(m+1)}]$						
n	m	$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 1.5$	$\lambda = 2.0$	$\lambda = 2.5$
5	2	1.4491 (0.7646)	0.7289 (0.2191)	0.5679 (0.1141)	0.4250 (0.0754)	0.3089 (0.0339)
7	3	1.4727 (0.5477)	0.7524 (0.1407)	0.5485 (0.1266)	0.3592 (0.0347)	0.2946 (0.0281)
9	4	1.4972 (0.5257)	0.7644 (0.1035)	0.5245 (0.0723)	0.4084 (0.0278)	0.3141 (0.0281)
11	5	1.4832 (0.4413)	0.7253 (0.1143)	0.4728 (0.0384)	0.3686 (0.0352)	0.2921 (0.0199)
13	6	1.4891 (0.4382)	0.7587 (0.0958)	0.5132 (0.0542)	0.3572 (0.0158)	0.2889 (0.0103)

Table 8:-Simulated values of $E[\text{Log}(x_{(m+1)})]$ and $V[\text{Log}(x_{(m+1)})]$ for different values of parameter λ .

n	m	$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 1.5$	$\lambda = 2.0$	$\lambda = 2.5$
5	2	0.1733 (.4420)	-.5022 (0.3951)	-0.7484 (0.4060)	-0.0315 (0.3957)	-1.3555 (0.4040)
7	3	0.2600	-0.4051	-0.7905	-1.1555	-1.3902

		(.2870)	(0.2747)	(0.3009)	(0.2996)	(0.3043)
9	4	0.2829 (0.2565)	-0.3649 (0.2436)	-0.7913 (0.2544)	-0.9821 (0.2994)	-1.3127 (0.2434)
11	5	-1.2149 (0.1931)	-0.4187 (0.1982)	-0.8308 (0.1900)	-1.1169 (0.2060)	-1.3402 (0.2068)
13	6	0.3097 (0.1772)	-0.3615 (0.1807)	-0.7680 (0.2090)	-1.0938 (0.1873)	-1.2999 (0.1797)

Conclusion:-

The transformations of random variable X , i.e. $h(X)$ have been used in several ways. Employing order statistics, it is seen that the variance of $\ln(X_{(m+1)})$ is very small compared to the variance of $X_{(m+1)}$. The reduction of variance increases the precision of the observations. In the present case $\ln(X_{(m+1)})$ would be asymptotically normally distributed, as it is known that $X_{(m+1)}$ is asymptotically normal. Thus, the probability distribution of $\ln(X_{(m+1)})$ is completely evaluated for large samples.

References:-

1. Bairamov, I. G., Gebizlioglu, O. L., & Kaya, M. F. (2001). Asymptotic distributions of statistics based on order statistics and record values and invariant confidence intervals. In *Asymptotic methods in probability and statistics with applications* (pp. 309-320). Birkhäuser, Boston, MA.
2. Balakrishnan, N., & Cohen, A. C. (2014). *Order statistics & inference: estimation methods*. Elsevier.
3. Box, G.E.P., and Cox, D.R. (1964). An Analysis of Transformations. *Jour. of the Royal Stat. Soc., Ser. B.*, 26, 211-252.
4. Buitendag, S. (2014). *An Investigation into Order Statistics: Asymptotic Distributions and Expansions*. Thesis. Stellenbosch University.
5. Carroll, R.J. and Ruppert, D. (1988). Transformation in regression: A robust analysis. *Technometric*, 27, 1-27.
6. Chao, J. C., Corradi, V., & Swanson, N. R. (2001). Data transformation and forecasting in models with unit roots and cointegration. *Annals of Economics and Finance*, 2(1), 59-76.
7. Cohen, D. S. (2001). Linear data transformations used in economics. Federal Reserve Board, FEDS Working Paper, (2001-59).
8. David, H. A., & Nagaraja, H. N. (2004). Order statistics. *Encyclopedia of Statistical Sciences*, 9.
9. Grinza, E., Kampelmann, S., & Rycx, F. (2018). L'union fait la force? Evidence for Wage Discrimination in Firms with High Diversity. *IZA Discussion Papers* 11520
10. Gumbel, E. J. (2012). *Statistics of extremes*. Courier Corporation.
11. Horrace, W. C. (2000). Selection Procedures for Order Statistics in Empirical Economic Studies.
12. Kampelmann, S., & Rycx, F. (2016). Wage discrimination against immigrants: measurement with firm-level productivity data. *IZA Journal of Migration*, 5(1), 15.
13. Lück, A., & Wolf, V. (2016). Generalized method of moments for estimating parameters of stochastic reaction networks. *BMC systems biology*, 10(1), 98.
14. Lütkepohl, H., & Xu, F. (2012). The role of the log transformation in forecasting economic variables. *Empirical Economics*, 42(3), 619-638.
15. Menzel, K., & Morganti, P. (2013). Large sample properties for estimators based on the order statistics approach in auctions. *Quantitative Economics*, 4(2), 329-375.
16. Parrado-Gallardo, E. M., Bárcena-Martín, E., & Imedio-Olmedo, L. J. (2014, September). Inequality, Welfare, and Order Statistics. In *Economic Well-Being and Inequality: Papers from the Fifth ECINEQ Meeting* (pp. 383-399). Emerald Group Publishing Limited.
17. Ryan, B.F., and Joiner, B.L. (1994). *Minitab Handbook Third Edition*. Blement, C A: Duxbing Press