Comparing Shannon entropies and standard deviations of the order statistics for uncensored, semicensored, and censored distributions

Hayat Kılıç, Atıf Evren

Abstract— As well as being three maximum entropy distributions, uniform, exponential and normal distributions have different properties in terms of censorization. Uniform distribution is censored from two sides . Exponential distribution is censored from below, whereas the normal distribution is uncensored. In this study, Shannon entropies of order sta-tistics from uniform, exponential and normal distributions are considered.It has been found that entropies of order statistics are some functions of the entropies of parent distributions.They are also functions of sample size, and order of the statistic. Entropy estimates are then compared with standard deviations of order statistics. It has been detected that for order statistics of censored distributions, like uniform or exponential distribution; as the two measures of uncertainty; standard deviation, and entropy do not convey the same information, i.e., they are not positively correlated.This is probably because the lowest and/or highest order statistics do not show high variability due to censorization. For uncensored distributions like the normal distribution, entropy and standard deviation are positively correlated as expected a priori. Therefore in case of censorization using entropy statistics may not be appropriate to measure uncertainty or variability of order statistics.

Index Terms— Censored distributions, Order statistics, Shannon entropy, Uncertainty

I. INTRODUCTION

Order statistics are important in modelling meteorological, agricultural, or financial data. They are often used in analysing extreme events. The minimum, the first quartile, the median, the third quartile and the maximum are well-known examples of order statistics.

The entropy of a statistical experiment is a measure of uncertainty before sampling, so that entropy, and the amount of sample information are two closely related concepts. Various entropy measures are summarized well by Esteban&Morales(1995) and Pardo(2006). Among them, it should be emphasized that Shannon, Rényi and Tsallis entropies have been widely used in recent studies.

Maximum entropy principle; which was introduced by Jaynes(2005); is used in Bayesian modelling, mainly when there is little, or no information about prior distributions. Uniform, exponential and normal distributions are three distributions giving maximum entropy under some

Hayat Kılıç, Yildiz Technical University, Institute of of Natural Sciences , Department of Statistics, Davutpasa, Esenler, 34210, Istanbul, Turkey

Atıf Evren, Yildiz Technical University, Faculty of Sciences and Literature ,Department of Statistics, Davutpasa, Esenler,34210, Istanbul-Turkey

assumptions on probability distribution. Besides they have totally different characteristics in terms of censorization.

II. SHANNON ENTROPY

If the discrete random variable X takes on the values $x_1, x_2, \ldots, x_K$ with respective probabilities $p_1, p_2, \ldots, p_K$ on sample space S, then Shannon entropy is defined as

$$H_{SH} = -\sum_{i=1}^{K} p_i \ln p_i \qquad (1)$$

It can be shown that Shannon entropy is related to multinomial coefficient as below:

$$\lim_{n \to \infty} \ln \frac{n!}{n_1! n_2! \ldots n_K!} = -n \sum_{i=1}^{K} p_i \ln p_i \qquad (2)$$

where $n = \sum_{i=1}^{K} n_i$ and $p_i = \frac{n_i}{n}$ for $i = 1,2,3,\ldots,K$ (Kullback, p11)

This relation can be shown by taking $\log x! \approx x \log x - x$ for $0 < x < \infty$ approximately (Golan, p.25).

Rényi Entropy

Rényi(1961) entropy which is also called as α-class entropy (Ullah, 1996) is a generalization of Shannon entropy. For discrete case, it is defined as

$$H_R = \frac{\ln \sum_{i=1}^{K} p_i^\alpha}{1-\alpha} \text{ for } \alpha > 0 \text{ and } \alpha \neq 1 \qquad (3)$$

By L'Hospital's rule

$$\lim_{\alpha \to 1} \frac{\frac{d}{d\alpha}(\ln \sum_{i=1}^{K} p_i^\alpha)}{\frac{d}{d\alpha}(1-\alpha)} = \lim_{\alpha \to 1} \frac{\frac{\sum_{i=1}^{K} p_i^\alpha \ln p_i}{\sum_{i=1}^{K} p_i^\alpha}}{-1}$$

$$= -\sum_{i=1}^{K} p_i \ln p_i = H_s \qquad (4)$$

Therefore, Shannon entropy is a special case of Rényi entropy as $\alpha \to 1$.

Havrda-Charvat and Tsallis Entropies

Another generalized form of Shannon entropy is due to Constantino Tsallis(1988). Although a similar formulation by Havrda-Charvat(1967) is older, Tsallis entropy has become more popular. Indeed, there is a slight difference between the formulations of Havrda-Charvat and Tsallis entropies. Havrda-Charvat entropy is defined as

$$H_{HC} = \frac{2^{\alpha-1}}{2^{\alpha-1}-1}[1 - \sum_{i=1}^{K} p_i^\alpha] \text{ for } \alpha > 0 \text{ and } \alpha \neq 1 \qquad (5)$$

Comparing Shannon entropies and standard deviations of the order statistics for uncensored, semicensored, and censored distributions

Tsallis entropy also known as β-class entropy is defined as

$$H_T = \frac{1 - \sum_{i=1}^{K} p_i^\alpha}{\alpha - 1}, \; for \, \alpha > 0 \, and \, \alpha \neq 1 \tag{6}$$

By L'Hospital's rule

$$\lim_{\alpha \to 1} \frac{1 - \sum_{i=1}^{K} p_i^\alpha}{\alpha - 1} =$$

$$\lim_{\alpha \to 1} \frac{\frac{d}{d\alpha}(1 - \sum_{i=1}^{K} p_i^\alpha)}{\frac{d}{d\alpha}(\alpha - 1)} = \lim_{\alpha \to 1} \frac{-\sum_{i=1}^{K} p_i^\alpha \ln p_i}{1} =$$

$$-\sum_{i=1}^{K} p_i \ln p_i \tag{7}$$

In other words, Tsallis entropy approaches to Shannon entropy as $\alpha \to 1$ as well as Rényi entropy.

The asymptotic properties of several entropy measures are given in Pardo(2006), and Zhang (2013). For a comprehensive simulation study on asymptotic normality of the three entropy measures for discrete cases, one can also refer to Evren&Ustaoglu(2015).

Maximum Entropy Considerations

For discrete distributions, if all cases are equally probable, i.e.; $p_i = \frac{1}{K} for \, i = 1,2,\dots,K)$, then entropy is maximum. Maximum entropy formalism was first developed by Jaynes (1957), and it coincides with Laplace's principle of insufficient reasoning. It is generally used in Bayesian context. In determining posterior distributions, whenever there is little, or no information about the prior distributions, by this principle, some probability distributions are selected intentionally. They are mainly uniform, exponential and normal distributions. One may refer to Reza(1994), to check how these distributions are derived as a result of maximum entropy assumption, in case of certain limitations on probability distributions.

Differential Entropies

The entropy of a continuous distribution is called differential entropy. For continuous cases, the summation signs in the defining formulas (1), (3), (5) and, (6) are replaced by integration signs. Note that for continuous cases, differential entropies may be negative, since probability densities may be larger than one.

Differential Entropies of Uniform, Exponential and Normal Distributions

Uniform distribution has the probability density function and distribution function as below:

$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b \tag{8}$$

$$F(x) = \int_{-\infty}^{x} f(t)dt = \frac{x-a}{b-a} \quad a \leq x \leq b \tag{9}$$

Shannon entropy of uniform distribution is

$$H_{SH}(x) = -E(\ln(f(x)) = \ln(b-a) \tag{10}$$

For $a = 0 \, and \, b = 1$, Shannon entropy is not meaningful as a measure of variability.

Note that the variance of uniform distribution is $\sigma^2 = \frac{(b-a)^2}{12}$. Therefore,

$$(b-a) = 2\sqrt{3}\sigma$$
$$H_{SH}(x) = \ln(2\sqrt{3}\sigma) \tag{11}$$

from which, it is easy to see that for uniform distribution, entropy is a logarithmic function of standard deviation.

The probability density, and distribution functions of exponential distribution are given as

$$f(x) = \lambda \exp(-\lambda x) \quad x \geq 0 \tag{12}$$

$$F(x) = \int_{-\infty}^{x} f(t)dt = 1 - \exp(-\lambda x) \quad x \geq 0 \tag{13}$$

Shannon entropy can be found as

$$H_{SH}(x) = -E(\ln(f(x)) = 1 - \ln(\lambda) \tag{14}$$

The variance of exponential distribution is $\sigma^2 = \frac{1}{\lambda^2}$. Thus

$$H_{SH}(x) = 1 + \ln(\sigma) \tag{15}$$

Finally probability density, distribution function, and Shannon entropy of normal distribution are as follows:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad -\infty < x < \infty$$

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\}dt \tag{16}$$

$$H_{SH}(x) = -E(\ln(f(x)) = \ln(\sigma\sqrt{2e\pi}) \tag{17}$$

Note that as uniform distribution, and exponential distribution (as in (11) and (15)), Shannon entropy of normal distribution is also a logarithmic function of standard deviation.

Distribution of Order Statistics

Theorem: Let $X_1, X_2, \dots X_n$ be a random sample from the probability density function f(x) with cumulative distribution function F(x). Let $Y_1 \leq Y_2 \leq \dots \leq Y_n$ denote the corresponding order statistics; then the probability density function of αth order statistic is given by

$$f_{Y_\alpha}(y) = \frac{n!}{(\alpha-1)!(n-\alpha)!}[F(y)]^{\alpha-1}[1-F(y)]^{n-\alpha}f(y) \tag{18}$$

Or identically;

$$f_{Y_\alpha}(y) = \alpha \binom{n}{\alpha}[F(y)]^{\alpha-1}[1-F(y)]^{n-\alpha}f(y) \tag{19}$$

(Mood et al., p254)

III. MOMENTS OF ORDER STATISTICS

Uniform Distribution

Let $Y \sim Uniform(0;1)$

$$f_{Y_\alpha}(y) = \frac{n!}{(\alpha-1)!(n-\alpha)!}[y]^{\alpha-1}[1-(y)]^{n-\alpha}$$

The Kth central moment of a continuous random variable Y_α is

$$E(Y_\alpha^K) =$$
$$\int_{-\infty}^{\infty} y^K f(y_\alpha)dy = \alpha \binom{n}{\alpha} \int_{-\infty}^{\infty} y^{K+\alpha-1}(1-y)^{n-\alpha}dy \tag{20}$$

$$E(Y_\alpha^K) = \alpha \binom{n}{\alpha} Beta(K+\alpha, n-\alpha+1)$$

$$E(Y_\alpha^K) = \frac{\alpha n!(K+\alpha-1)!(n-\alpha)!}{\alpha!(n-\alpha)!(K+n)!} = \frac{\alpha n!(K+\alpha-1)!}{\alpha!(K+n)!} \tag{21}$$

Thus

$$E(Y_\alpha) = \frac{\alpha}{(n+1)} \qquad (22)$$

$$E(Y_\alpha^2) = \frac{\alpha(1+\alpha)}{(n+1)(n+2)} \qquad (23)$$

$$Var(Y_\alpha) = E(Y_\alpha^2) - \left(E(Y_\alpha)\right)^2 = \frac{(n-\alpha+1)\alpha}{(n+2)(n+1)^2} \qquad (24)$$

Note that for uniform distribution, the variance of order statistic is not a monotonic function of α. By taking the first derivative of (24), with respect to α;

$$\frac{d}{d\alpha}\left[\frac{(n-\alpha+1)\alpha}{(n+2)(n+1)^2}\right] = constant * (n+1-2\alpha)$$

The algebraic sign of the derivative is positive for $\alpha < \frac{n+1}{2}$.
In other words;

$Var(Y_\alpha)$ is increasing for $\alpha < \frac{n+1}{2}$, and $Var(Y_\alpha)$ is decreasing for $\alpha > \frac{n+1}{2}$.

Therefore comparing the variances of order statistics with entropies is not meaningful for censored distributions like uniform distribution.

Exponential and Normal Distributions
The Kth central moment of a continuous random variable Y_α is

$$E(Y_\alpha^K) = \int_{-\infty}^{\infty} y^K f(y_\alpha)dy = \int_{-\infty}^{\infty} y^K \alpha \binom{n}{\alpha}(F(y))^{\alpha-1}(1-F(y))^{n-\alpha}f(y)dy \qquad (25)$$

The moments of the order statistics of uniform destribution can be easily found by (20), and by simply exploiting the properties of beta function. But for exponential and normal distributions, deriving the moments of order statistics is not easy. For exponential distribution evaluating the integral in (25) is complicated . For normal distribution, even worse, the distribution function does not have an explicit form. For these reasons, numeric methods (series expansion or integrating by simulation) may be appropriate.
By the law of large numbers ,

$$\frac{\sum_{j=1}^{m} g(y_j)}{m} \to E(g(Y)) \quad \text{for large m.} \qquad (26)$$

Let

$$g(y_j) = y_j^K \alpha \binom{n}{\alpha}\left(F(y_j)\right)^{\alpha-1}\left(1-F(y_j)\right)^{n-\alpha} \qquad (27)$$

Then

$$\frac{\sum_{j=1}^{m} \alpha y_j^K \binom{n}{\alpha}\left(F(y_j)\right)^{\alpha-1}\left(1-F(y_j)\right)^{n-\alpha}}{m} \to E(Y_\alpha^K) \qquad (28)$$

for large m.

In other words, by simulating large number of random observations from predetermined distributions, one can estimate $E(Y_\alpha^K)$ (Ross).

Shannon-Entropy of Order Statistics
Shannon entropies of order statistics can be found by (1) and (17):

$$\ln\left(f_{Y_\alpha}(y)\right) = \ln\left(\alpha\binom{n}{\alpha}\right) + (\alpha-1)\ln(F(y) + (n-\alpha)\ln(1-F(y)) + \ln(f(y)) \qquad (29)$$

$$-E\left(\ln\left(f_{Y_\alpha}(y)\right)\right) = -\ln\left(\alpha\binom{n}{\alpha}\right) - (\alpha-1)E(\ln(F(y)) - (n-\alpha)E(\ln(1-F(y))) - E(\ln(f(y))) \qquad (30)$$

$$H_{SH}(y_\alpha) = -\ln\left(\alpha\binom{n}{\alpha}\right) - (\alpha-1)E(\ln(F(y)) - (n-\alpha)E(\ln(1-F(y))) + H_{SH}(y) \qquad (31)$$

To simplify (31), one may find the probability distributions of $lnF(Y)$, and $\ln(1-F(Y))$ by the following theorem:

Theorem
If X is a continuous random variable with probability density function f, and $Y = g(X)$ is a monotone function in the range space of X, then the probability density function of Y is

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right| \qquad (32)$$

(Mood et al., p200)

Probability Integral Transform
Specifically if $Y = F(X)$ then $\frac{dy}{dx} = f_x(x)$ and

$$f_Y(y) = f_X(x)\left|\frac{1}{\frac{dy}{dx}}\right| = 1$$

Therefore $U = F(Y) \sim Uniform(0;1)$. Similarly $V = (1-F(Y)) \sim Uniform(0;1)$

Let $W = g(U) = ln(U)$. Hence

$$f_W(w) = f_u(g^{-1}(w)) \quad \left|\frac{du}{dw}\right| \quad ;$$
$$f_u(u) = 1 \, for \, 0 < u < 1$$
$$f_u(g^{-1}(w))=1; \quad U = \exp(W) \quad \frac{du}{dw} = \exp(W)$$
$$f_W(w) = \exp(w) \quad -\infty < w < 0$$
$$E(W) = \int_{-\infty}^{0} w\exp(w)\,dw = \int_{-\infty}^{0} x\exp(x)\,dx$$

By the technique of integration by parts;
$$u = x \Rightarrow du = dx; dv = \exp(x)\,dx \Rightarrow v = \exp(x)$$
$$\int_{a}^{b} uv\,dx = uv|_a^b - \int_{a}^{b} v\,du$$
$$\int_{-\infty}^{0} x\exp(x)\,dx =$$
$$x\exp(x)|_{-\infty}^{0} - \int_{-\infty}^{0}\exp(x)\,dx = -1. \quad \text{Therefore}$$
$$E(\ln(F(Y)) = E(Ln(1-(F(Y))) = -1 \quad \text{and}$$

hence,

$$H_{SH}(y_\alpha) = -\ln(\alpha) - \ln\binom{n}{\alpha} + (n-1) + H_{SH}(y) \qquad (33)$$

For uniform distribution;

$$H_{SH}(y_\alpha) = -\ln(\alpha) - \ln\binom{n}{\alpha} + (n-1) + \ln(b-a) \quad (34)$$

For exponential distribution;

$$H_{SH}(y_\alpha) = -\ln(\alpha) - ln\binom{n}{\alpha} + n - \ln(\lambda) \quad (35)$$

For normal distribution;

$$H_{SH}(y_\alpha) = -\ln(\alpha) - \ln\binom{n}{\alpha} + (n-1) + \ln(\sigma\sqrt{2\pi e}) \quad (36)$$

IV. SIMULATION STUDY

From uniform, exponential and normal distributions; random samples, consisting of 30, 50 and 100 observations have been selected. Simulations have been realized by the help of Microsoft Excel.Then by using the equations (34), (35) and (36) Shannon entropies of order statistics have been calculated. These quantities have been then compared with the sample standard deviations of order statistics. The results of 9 simulation studies are summarized by the following graphs.

1)Distribution: uniform, sample size=30, number of runs=1000.
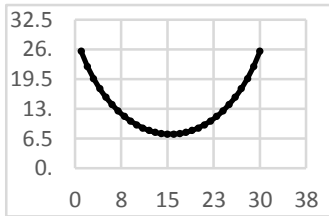
Fig 1-a)Entropy versus the order of the statistic
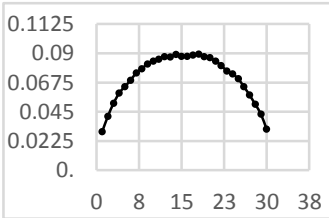
Fig-1-b) Sample standard deviation of order statistics versus order of the statistic
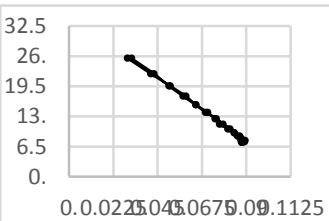
Fig 1-c)Entropy versus sample versus standard deviation of order statistic

2)Distribution: uniform, sample size=50, number of runs=1000.
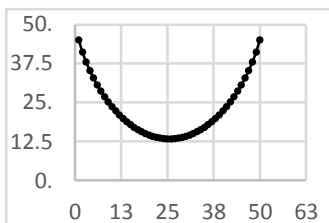
Fig 2-a)Entropy versus the order of the statistic
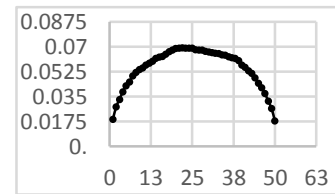
Fig-2-b) Sample standard deviation of order statistic versus order of the statistic
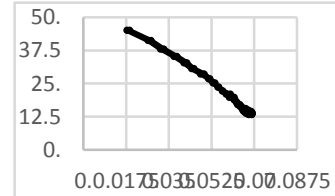
Fig 2-c)Entropy versus sample standard deviation of the order statistic

3)Distribution: uniform, sample size=100, number of runs=1000.
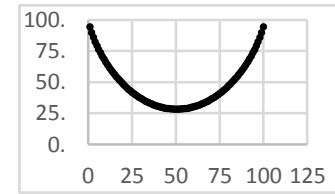
Fig 3-a)Entropy versus order of the statistic
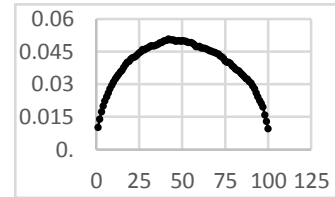
Fig 3-b) Sample standard deviation of order of the statistic
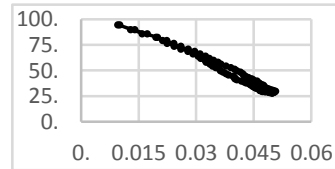
Fig 3-c)Entropy versus sample standard deviation of order statistic

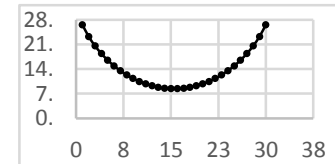4)Distribution: exponential, sample size=30, number of runs=1000.
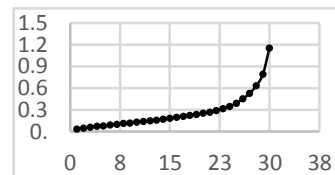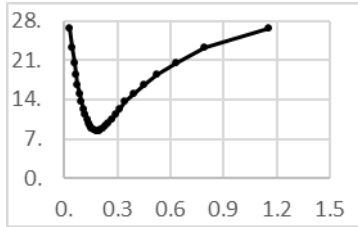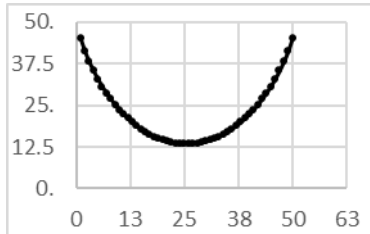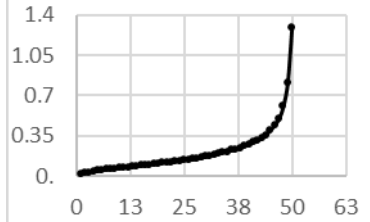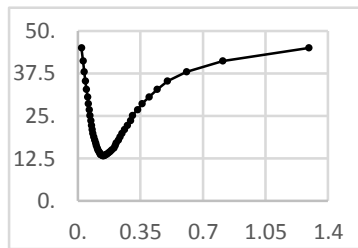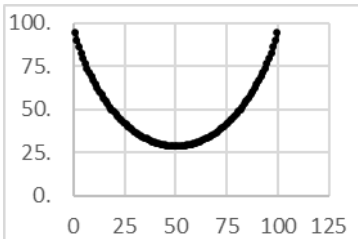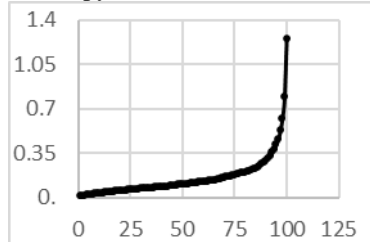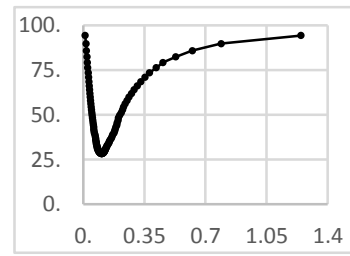

Fig 4-a)Entropy versus the order of the statistic

Fig 4-b) Sample standard deviation of order statistic versus order of the statistic

Fig 4-c)Entropy versus sample standard deviation of order statistic

5)Distribution: exponential, sample size=50, number of runs=1000.

Fig 5-a)Entropy versus order of the statistic

Fig 5-b)Sample standard deviation of order statistic versus order of the statistic

Fig 5-c)Entropy versus sample standard deviation of order statistic

6)Distribution: exponential, sample size=100, number of runs=1000.

Fig 6-a)Entropy versus the order of the statistic

Fig 6-b)Sample standard deviation versus order of the statistic

Fig 6-c)Entropy versus sample standard deviation of order statistic

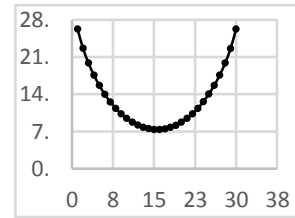7)Distribution:normal, sample size=30, number of runs=1000.

Fig 7-a)Entropy versus the order of the statistic
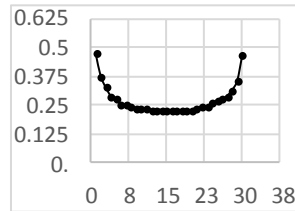

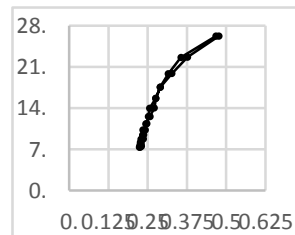Fig 7-b)Sample standard deviation versus order of the statistic

Fig 7-c)Entropy versus sample standard deviation of order statistic

8)Distribution: normal ,sample size=50, number of runs=1000.

Fig 8-a)Entropy versus the order of the statistic

Fig 8-b)Sample standard deviation versus order of the statistic

Fig 8-c)Entropy versus sample standard deviation of order statistic

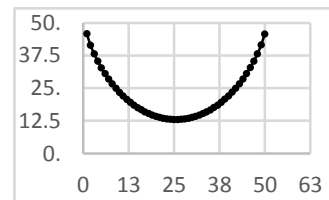9)Distribution: normal, sample size=100, number of runs=1000.
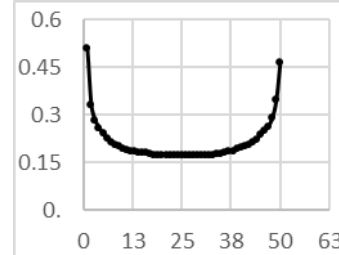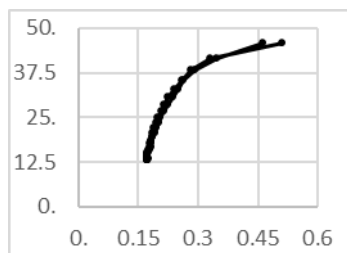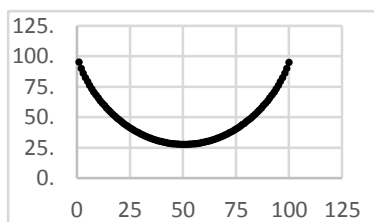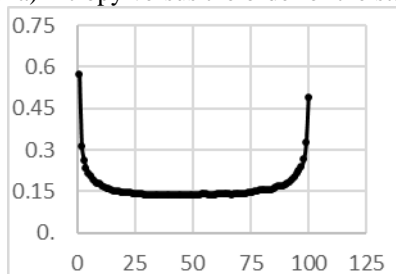

Fig 9-a)Entropy versus the order of the statistic

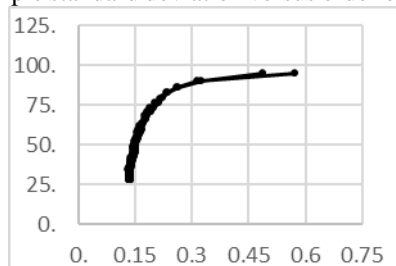Fig 9-b)Sample standard deviation versus order of the statistic

Fig 9-c)Entropy versus sample standard deviation of order statistic

V. DISCUSSION

For nominal and ordinal distributions, one cannot compute standard deviation. In such situations, the only way to measure variability is to use tools based on frequencies only. By this reasoning, various formulations of entropy should have been put forward to measure the dispersion in nominal and ordinal distributions.

Various measures of entropy are also taken into consideration in analysing continuous data. But apart from discrete case, there are some unexpected results. As in the case of uniform distribution, differential entropy may be zero. For some other distributions it may even possess negative values. In such cases, using entropy measures in statistical analysis may be useless, or misleading.

In this study, we have considered Shannon entropies of order statistics of three continuous distributions; namely, uniform, exponential and the normal distributions. In general, Shannon entropies are some logarithmic functions of standard deviations. In this respect, one may expect to see positive correlation coefficients between entropy and standard deviation. From the findings of 9 simulations, which are summarized by the figures from "Fig 1-a" to "Fig 9-c", we have not observed this phenomenon, because of censoring.

For uniform distribution, since the distribution is censored from below and above, the standard deviations and entropies of order statistics failed to be correlated with each other. For exponential distribution, the situation is recovered moderately, since the distribution is censored from below. Therefore, for exponential distribution, one may expect high correlations between standard deviations and entropies of order statistics as the order increases considerably. One remedy may be found in increasing the sample size, if one checks figures 4-c, 5-c, and 6-c. Especially in 6-c,   order statistics   whose standard deviations are not positively correlated with entropies are lower order statistics. The situation may be improved, by increasing the sample size if it is possible. Finally, we have observed that entropies and standard deviations of order statistics   from normal distribution are positively correlated, since the normal distribution is uncensored. Therefore our emphasis is on the existence of   positive correlations between standard deviations and entropies of order statistics   whenever there is no censorization.

REFERENCES

[1]Esteban, M.D., Morales, D. (1995), A summary on entropy statistics, Kybernetika, Vol. 31(1995), No.4, 337-346

[2]Evren, A, Ustaoğlu, E.,(2015), On Asymptotic Normality of Entropy Measures, International Journal of Applied Science and Technology, Vol 5., No.5, 31-38

[3]Golan, A. (2006), Information and Entropy Econometrics-A Review and Synthesis, Foundations and Trends in Econometrics 2:1-2 (2006), now Publishers Inc.

[4]Havrda, J and F.,Charvat (1967), Quantification Method in Classification Processes: Concept of Structural α-entropy, Kibernetika 3, 30-35

[5] Jaynes E.T. (1957), Information Theory and Statistical Mechanics, Phys. Rev. 106, 620-630

[6]Jaynes, E.T.(2005), Probability Theory. The Logic of Chance, Cambridge University Press, UK.

[7] Kullback, S.(1996), Information Theory and Statistics, Dover Publications, NY ,p6

[8] Mood, A.M., Graybill F.A., Boes, D. (1974), Introduction to the Theory of Statistics, McGraw-Hill International Editions, third edition.

[9]Pardo, L. (2006), Statistical Inference Measures Based on Divergence Measures, Chapman&Hall/CRC, 99.

[10] Rényi, A.(1961), On Measures of Entropy and Information, In Proc. 4th Berkeley Symp. Statist; Vol I, Probability 1, University of California Press, California, 547-561

[11] Reza, F.M. (1994), An Introduction to Information Theory, Dover Publications, NY.

[12] Ross,S.M. (2002), Simulation, Academic Press, An Elsevier Science Imprint, Third edition, pp38-40.

[13] Tsallis, C. (1988), Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics 52: 479–487.

[14] Ullah, A.(1996), Entropy, Divergence and Distance Measures with Econometric Applications, Journal of Statistical Planning and Inference, 49(1996), 137-162

[15] Zhang, Xing (2013), Asymptotic Normality of Entropy Estimators, proposal submitted to the faculty of The University of North Carolina at Charlotte in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Applied Mathematics, Charlotte) http://libres.uncg.edu/ir/uncc/f/Zhang_uncc_0694D_10428.pdf