Pak. J. Statist.
2021 Vol. 37(3), 303-313

NEUTROSOPHIC ENTROPY MEASURES FOR THE NORMAL DISTRIBUTION: THEORY AND APPLICATIONS

Rehan Ahmad Khan Sherwani[1], Muhammad Farooq[2], Nadia Saeed[1]
Zahoor Ahmad[3], Sana Saeed[1], Shumaila Abbas[1] and **Tooba Arshad[1]**

[1] College of Statistical and Actuarial Sciences,
University of the Punjab, Lahore, Pakistan.
Email: rehan.stat@pu.edu.pk; nadia.stat@pu.edu.pk
sana.stat@pu.edu.pk; shumaila.stat@pu.edu.pk
toobaarshad80@gmail.com

[2] Govt. College University, Lahore, Pakistan.
Email: muhammad.farooq@gcu.edu.pk

[3] University of Lahore, Sargodha Campus, Pakistan.
Email: zahoor.ahmad@math.uol.edu.pk

ABSTRACT

Entropy is a measure of uncertainty and often used in information theory to determine the precise testimonials about unclear situations. Different entropy measures available in the literature are based on the exact form of the observations and lacks in dealing with the interval-valued data. The interval-valued data often arises from the situations having ambiguity, imprecise, unclear, indefinite, or vague states of the experiment and is called neutrosophic data. In this research modified forms of different entropy measures for normal probability distribution have been proposed by considering the neutrosophic form data. The performance of the proposed neutrosophic entropies for normal distribution has been assessed via a simulation study. Moreover, the proposed measures are also applied to two real data sets for their wide applicability. The results of the study suggested the use of the proposed methods in the presence of fuzzy, interval-valued, or neutrosophic data.

KEYWORDS

Neutrosophic Statistics; Entropy; Shannon and Rényi entropy; Neutrosophic Entropy, Normal Distribution; fuzzy type data; interval type data.

1. INTRODUCTION

The normal distribution has gained immense importance due to its wide real-life applications. For a random variable $X \in \mathbb{R}$, the location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$, the probability density function (pdf) of a normal distribution is defined by

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

One of the reasons for the wide application of normal distribution is that the sum for a large number of observations from a random variable can be approximated by the normal distribution, irrespective of the prior original distribution of the random variable [1]. Furthermore, the random errors that usually appear in several measurements approximately follow a normal distribution, a reason to give the name of *error distribution* to the normal distribution [2]. Moreover, almost all the discrete and continuous distributions can be approximated by the normal distribution under some conditions [3-5].

The concept of entropy is proposed by Entropy Clausius in 1850 to measure of uncertainty and unpredictability of the state [6]. It is also introduced by [7, 8] as a measure of disorder. In information theory, entropy is defined as the average level of information or uncertainty inherent in a random variable's possible outcomes. Numerous studies have been conducted where different types of entropy measures have been proposed. Among those types, Shannon entropy and Rényi Entropy are well-known. Recall that the Shannon entropy is defined by [9]

$$H(x) = -\int f(x)\log\big(f(x)\big)\,dx \tag{2}$$

and Rényi Entropy [10] is defined by

$$H_\alpha(X) = \frac{1}{1-\alpha} log_2 \int f(x)^\alpha dx, \tag{3}$$

where α is the order of Rényi entropy. Note that the Rényi entropy for $\alpha = 0$ is Schur Concave.

Neutrosophic statistics is an extension of classical statistics. It has gain importance due to its capacity to deal with sets of values more specifically an interval. To be more precise, when the values or parameters have confusion attached with them, then that particular value or parameter is replaced with a set of values [11, 12]. To represent the neutrosophic version of parameter or statistic, a subscript "N" is used such as x_N [13, 14]. Analogously, any number x replaced by a set is denoted here as x_N i.e. neutrosophic x or uncertain x. It is important to note that fuzzy logic is a special case of neutrosophic logic [15]. This x_N is mostly the interval including x, generally it is a set which approximates x [16]. The neutrosophic statistical number is defined by

$$N = d + i$$

where "d" is the determining part of N and "i" is the indeterminate part (unsure) of N. For more details, we refer the readers to [5, 17-22].

2. NEUTROSOPHIC ENTROPY OF NORMAL DISTRIBUTION

In this section, we first present the Shannon entropy and Rényi entropy for normal distribution. After that, the neutrosophic versions of these two entropies for the normal distribution are proposed by combining the logic of neutrosophic with the said entropies.

2.1 Shannon Entropy of Normal Distribution

The Shannon entropy for the normal probability distribution is defined as:

$$H(x) = -\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \log\left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right) dx \quad (4)$$

$$H(x) = \frac{1}{2}\ln(2\pi e\sigma^2) \quad (5)$$

Equation (5) shows the Shannon entropy of the normal distribution. For more details we referrer the reader to [23].

2.2 Neutrosophic Shannon Entropy of Normal Distribution

The statistic in equation (5) is ideal for the cases where we have crisp numbers and values for parameters. However, in real-life situations, we scarcely come across crisp values. In other words, there is always some uncertainty that makes it difficult for us to draw decisions with certainty. For example, the temperature does not have an exact value as it lies between the highest recorded temperatures and the lowest at a particular time point. The same is the situation with stock exchange records and numerous other cases in the real world. Neutrosophy being the extension of classical statistics could be helpful in these kinds of circumstances. To this end, using equation (5) we define neutrosophic Shannon entropy for normal distribution by

$$H_N(x) = \left\{\frac{1}{2}\ln(2\pi e{\sigma_L}^2), \frac{1}{2}\ln(2\pi e{\sigma_U}^2)\right\} \quad (6)$$

Here σ_L is the estimate of the parameter σ of the normal distribution that is obtained from the lower bound of the interval from the samples where the uncertainty lies and σ_U being the estimate of parameter σ of the normal distribution that is obtained from the upper bound of the interval from uncertain samples. In the situations where doubts appear on more than one values (an interval), the lowest among them is considered as “L” the largest value as “U”. Using equation (6) we can obtain two results, one for the lower and other for the upper bound of any interval and then we can state that the entropy in the particular case lies between these results.

2.3 Rényi Entropy of Normal Distribution

To develop a neutrosophic version of Rényi Entropy we here briefly discuss the derivation of Rényi entropy of normal distribution. For this purpose, we plug pdf of a normal distribution from equation (1) in the general formulation of Rényi entropy given in equation (3). Assuming that $\alpha = 2$ the Rényi entropy is

$$H_2(X) = \frac{1}{1-2} log_2 \int_{-\infty}^{+\infty} \left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right)^2 dx \quad (7)$$

Assume that $t = \frac{x-\mu}{2\sigma}$, then $\sigma dt = dx$. By plugging this result in equation (7) we obtain

$$H_2(X) = -log_2 \frac{1}{2\pi\sigma} \int_{-\infty}^{+\infty} \exp(-t^2) dt$$

$$H_2(X) = -log_2 \frac{1}{2\sqrt{\pi}\sigma} \tag{8}$$

Equation (8) shows the Rényi entropy of the Normal distribution.

2.4 Neutrosophic Rényi Entropy of Normal Distribution:

Analogous to the idea presented in the case of neutrosophic Shannon entropy in Section 2.2, neutrosophic Rényi entropy of normal distribution based on the statistic given in equation (7) is defined by

$$H_{2N}(X) = \left\{-log_2 \frac{1}{2\sqrt{\pi}\sigma_L}, -log_2 \frac{1}{2\sqrt{\pi}\sigma_U}\right\} \tag{9}$$

where σ_L be the lower and σ_U be the upper limit values of the parameter in the given interval. Furthermore, by equation (8) we can also obtain two results one for the lower and the other for the upper bound of an interval which will indicate that the entropy of that particular case lies between these bounds.

3. SIMULATION STUDY

In this Section, a simulation study is conducted to determine the efficiency of the proposed entropy measures for normal distribution. For the simulation study, the following steps are followed.

Step 1: Compute the theoretical neutrosophic Shannon and Rényi entropy for some chosen interval of σ in the interval form.

Step 2: Random number *pop 1 and pop 2* were generated from a normal distribution of sample size 10,000 Generate *pop 1* ~ Normal Distribution and *pop 2* ~ Normal Distribution

Step 3: Compute the estimates of standard deviation σ of normal distribution for the lower and upper bound from the obtained samples.

Step 4: Now $H_N(x)$ can be attained from the equation (6).

Step 5: Now $H_{2N}(x)$ can be attained from the equation (8).

Step 6: We repeat steps 1-4 for 1000 random samples and different parameters to check the efficacy of the novel measure.

Step 7: Compute the empirical neutrosophic Shannon and Rényi entropy for normal distribution by taking an average of the obtained matrix from the obtained entropies in the interval form.

Step 8: Compare the theoretical and empirical neutrosophic entropies. The empirical results obtained from simulations for the neutrosophic Shannon entropy of normal distribution are shown in Table 1 for different values of σ, keeping the mean as constant.

It can be observed from Table 1 that there are some patterns and relationships among the values of estimates of parameters and proposed entropy measure. Furthermore, by increasing the values of parameters the neutrosophic Shannon entropy also increases.

Table 1
Lower and Upper limits of Neutrosophic Shannon Entropy for Normal Distribution for Fixed Values of μ and Different Values of σ

Neutrosophic Shannon Entropy μ =(150,188)			
	σ	Lower Limit	Upper Limit
1.	(0.05,0.25)	-0.079	0.726
2.	(0.10,0.14)	0.267	0.436
3.	(0.25,0.37)	0.726	0.922
4.	(0.45,0.62)	1.019	1.179

On the contrary Table 2 shows the empirical results of neutrosophic Rényi entropy of normal distribution based on simulation studies for different values of σ and fixed mean. It can be seen from Table 2 that on increasing the value of the parameter the neutrosophic entropy decreases.

Table 2
Lower and Upper Limits of Neutrosophic Rényi Entropy for Normal Distribution for Fixed μ and Different Values of σ

Neutrosophic Rényi Entropy μ = (150,188)			
	σ	Lower Limit	Upper Limit
1.	(0.05,0.25)	2.175	4.497
2.	(0.10,0.14)	3.011	3.496
3.	(0.25,0.37)	1.609	2.174
4.	(0.45,0.62)	0.864	1.326

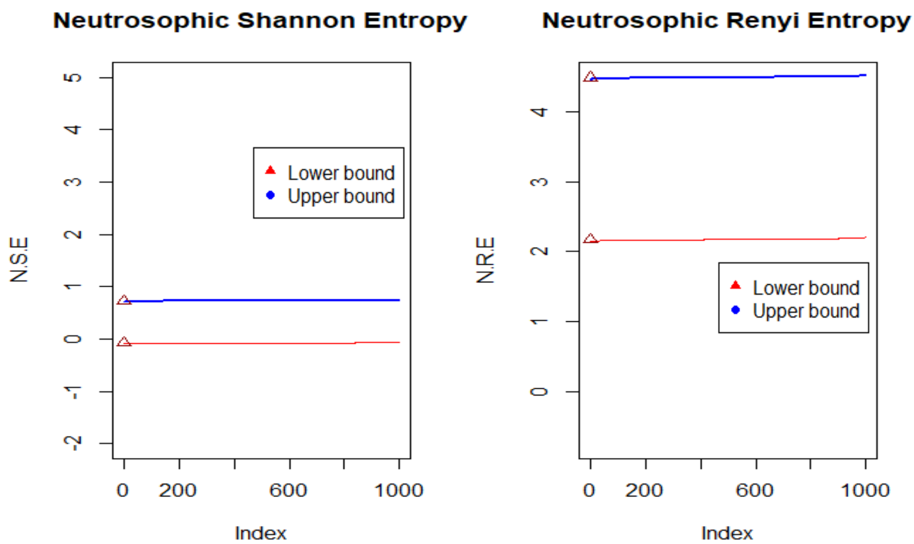

Figure 1: The lower and upper bounds of Neutrosophic Shannon and Rényi Entropy for Norma Distribution for σ = (0.05, 0.25) & μ = (150, 188)

For $\sigma = (0.05, 0.25)$ and $\mu = (150, 188)$, the neutrosophic Shannon and Rényi entropy for the normal distribution is shown in Figure 1. The red line represents the lower bound of the obtained interval of neutrosophic entropy while the blue line represents the upper bound of the obtained interval of neutrosophic entropy of the normal distribution. There exists an indeterminacy among an interval of values of parameters indicating the need of neutrosophic entropy in this situation.

4. REAL-LIFE APPLICATIONS

4.1 Application 1

In this example, we consider the monthly temperature of Oslon, France is collected for ten years (2010 – 2020) from World Weather Online in the form of average low and average high temperature (see Appendix for data). The normality of these two levels of temperature data is checked using a probability-probability plot using Minitab Software version 18.0. From Figure 2 and Figure 3, we see that the low average temperature, as well as high average temperature, follows approximately normal distributions at 5% level of significance.

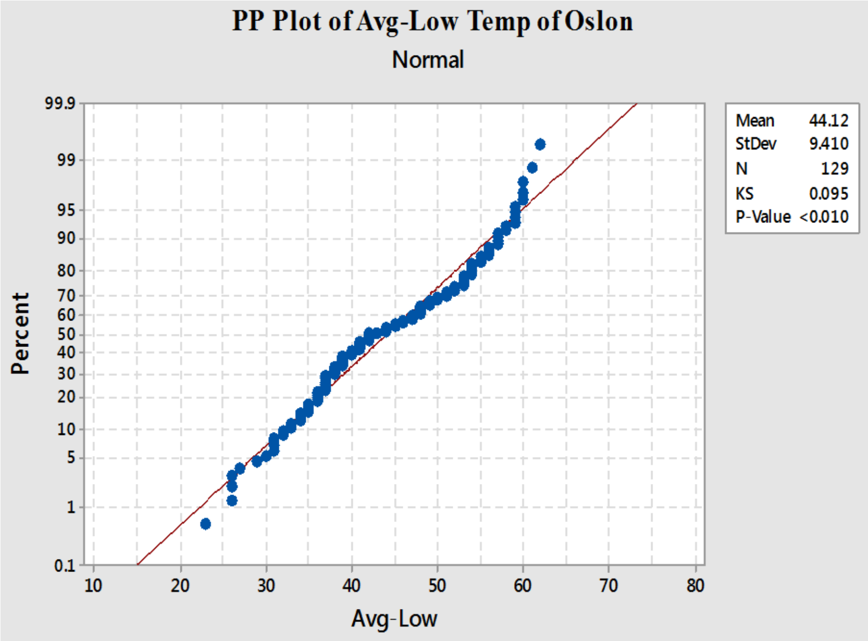

Figure 2: PP-Plots of Average Low Temperatures of Oslon, France during the period of 2010-2020

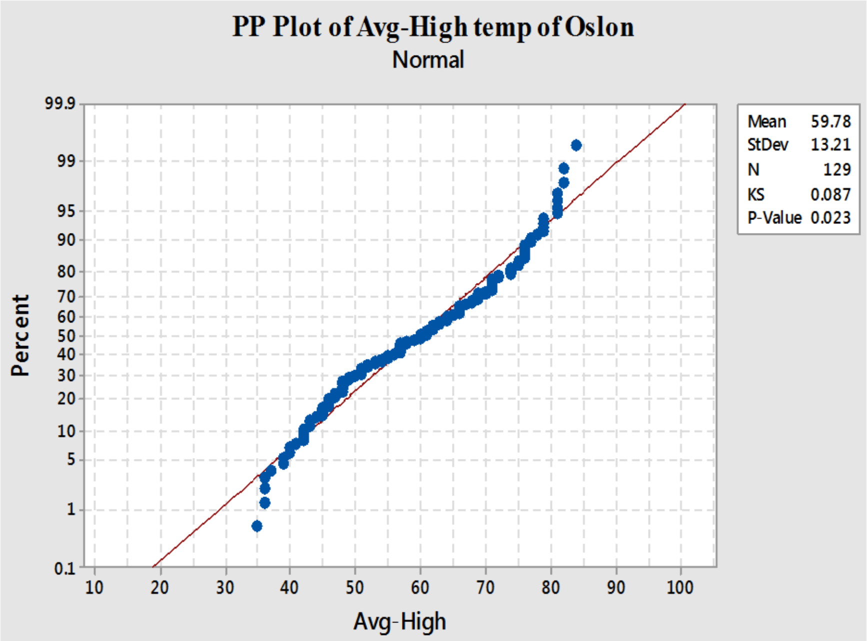

Figure 3: PP-Plots of Average High Temperatures of Oslon, France during the period of 2010-2020

Furthermore, from Figure 2 and Figure 3, we obtain the following estimates of lower and upper bounds for two parameters of normal distribution.

$$(\mu_L, \mu_U) = (44.12, 59.78) \quad (10)$$

$$(\sigma_L, \sigma_U) = (9.410, 13.21) \;\&\; N = 129 \quad (11)$$

By using equation (6), the neutrosophic Shannon entropy of normal distribution is

$$H_N(\mathrm{X}) = \{1.590\ ,1.737\} \quad (12)$$

We see that the entropy values for both the average high temperatures and average low temperatures are close to 1. This indicates that there lies maximum uncertainty in the weather condition of Oslon, France between 2010 and 2020. In other words, a close to neutrosophic value means that this kind of data has a huge amount of information indicating that weather conditions keep changing even in an intense cold situation in France.

4.2 Application 2

A juice company situated in Lahore, Pakistan performed a quality check on their frozen orange juice concentrate cardboard cans weighing 6-0z to determine the leakage. These cans are manufactured in a machine by spinning them from cardboard stock and attaching the metal bottom panel. By inspection of a can, we may determine whether a can could leak either on the side seam or around the bottom joint during the process of filling. This creates nonconformity in the data. However, the industrial experimenter is uncertain about the classification of some items either conforming or non-conforming during the inspection. Due to the uncertainty, the industrial engineers can expect the percent nonconforming product from 0.028 to 0.0379. The data is shown in Table 3.

Table 3
Number of Nonconforming Units in the Quality Check of Juice Company

Sr. No.	Number of Nonconforming Units D_i		Sr. No.	Number of Nonconforming Units D_i	
1.	12	13	2.	8	8
3.	15	15	4.	10	10
5.	8	10	6.	5	8
7.	10	10	8.	13	13
9.	4	4	10.	11	13
11.	7	7	12.	20	20
13.	16	16	14.	18	20
15.	9	11	16.	24	24
17.	14	14	18.	15	15
19.	10	10	20.	9	12
21.	5	8	22.	12	12
23.	6	8	24.	7	10
25.	17	17	26.	13	15
27.	12	15	28.	9	9
29.	22	22	30.	6	9

In this situation, when some observations are unclear and uncertain, the classical entropy measures are not suitable and the proposed entropy measures fit well. By probability-probability plot of upper and lower limits, we see that it follows approximately normal distribution at 5% level of significance.

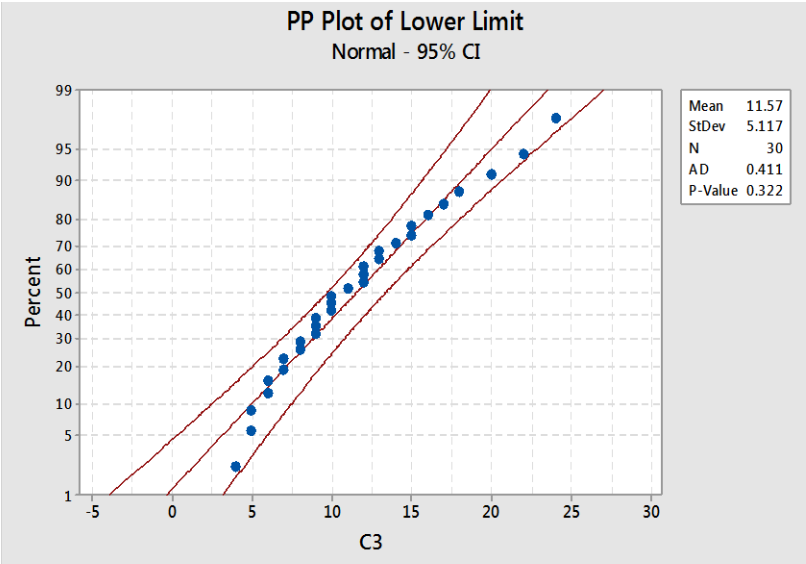

Figure 4: PP Plots of Lower Limit of Non-Conforming Units

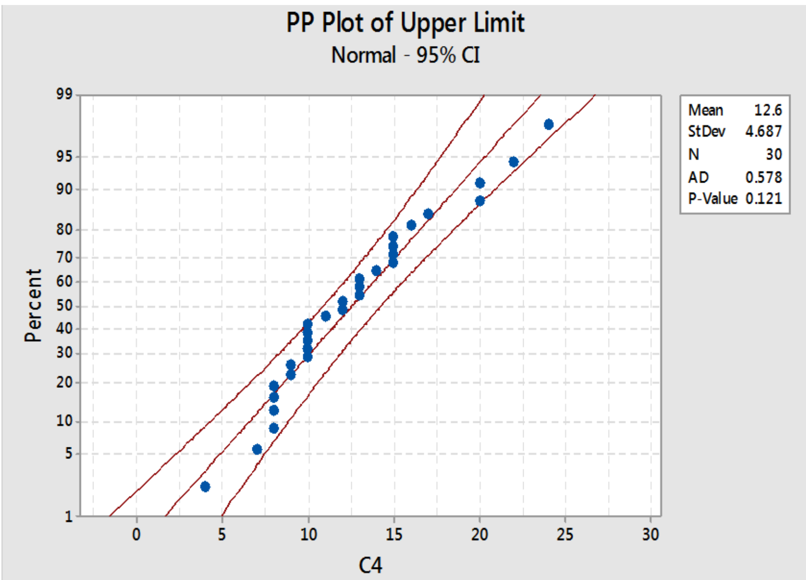

Figure 5: PP Plots of Upper Limit of Non-Conforming Units

$(\sigma_L, \sigma_U) = (4.687, 5.117)$ & $N = 30$.

We then calculate the Neutrosophic Shannon entropy of normal distribution using equation (20) that is by

$$H_N(X) = \left\{\frac{1}{2}\ln(2\pi e{\sigma_L}^2), \frac{1}{2}\ln(2\pi e{\sigma_U}^2)\right\}$$

which by considering $e = 2.7182818$ becomes

$$H_N(X) = \left\{\frac{1}{2}(2.574268), \frac{1}{2}(2.65051)\right\}$$

$$H_N(X) = \{1.287, 1.325\}.$$

The obtained neutrosophic entropy $\{1.287,\ 1.325\}$ is close to 1 indicating that the amount of uncertainty in the given situation is very high. This means that the leakage can be caused from both the side and bottom of the can. The manufacturer needs to take good care of both these parts to prevent this kind of error.

5. CONCLUSION

In this study neutrosophic versions of Shannon entropy and R*é*nyi entropy for the normal distribution are proposed by the combination of neutrosophic statistics and information theory, which has not been discussed in the past, especially with the combination of entropy. The proposed neutrosophic versions of entropies are found suitable for the situations where the data or the parameters occur in form of intervals (having some vagueness and indeterminacy). To check the performance and efficiency of the proposed measure of entropy, a simulation study was performed considering normal distribution and two real-life datasets. In both real-life examples, both upper and lower limits of neutrosophic entropies are found close to 1 which ensure the existence of a huge amount of uncertainty in the data and confirm the need of applying neutrosophic measures instead of classical entropy measures. Following the same idea, the neutrosophic entropy measures can be derived for some other distribution and implemented in real-life situations where uncertainty exists.

REFERENCES

1. Altman, D.G. and Bland, J.M. (1995). Statistics notes: the normal distribution. *Bmj*, 310(6975), 298.
2. Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and Methods*, 31(4), 497-512.
3. Bell, J. (2020). *A simple and pragmatic approximation to the normal cumulative probability distribution.* Available at SSRN 2579686.
4. Dombi, J. and Jónás, T. (2018). Approximations to the normal probability distribution function using operators of continuous-valued logic. *Acta Cybernetica*, 23(3), 829-852.
5. Hernandez, H. (2020). Probability Distributions in Groups of Random Elements. *ForsChem Research Reports*, 5, 2020-06.

6. Kapur, J.N. and Kesavan, H.K. (1992). Entropy optimization principles and their applications. In *Entropy and energy dissipation in water resources* (pp. 3-20). Springer, Dordrecht.
7. Jaynes, E.T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, 106(4), 620-630.
8. Amigó, J.M., Balogh, S.G. and Hernández, S. (2018). A brief review of generalized entropies. *Entropy*, 20(11), 813.
9. Shannon, C.E. and Weaver, W. (1949). *The Mathematical Theory of Communication*, The University of Illinois Press, United States.
10. Rényi, A. (1961). *On measures of entropy and information*. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California.
11. Smarandache, F. (2003). *Definitions derived from neutrosophics*, Infinite Study.
12. Zeina, M.B. and Hatip, A. (2021). Neutrosophic Random Variables. *Neutrosophic Sets and Systems*, 39(1), p. 4.
13. Smarandache, F. (2010). Neutrosophic logic-a generalization of the intuitionistic fuzzy logic. *Multispace & multistructure. Neutrosophic transdisciplinarity (100 collected papers of science)*, 4, 396.
14. Smarandache, F., Vladutescu, S., Colhon, M., Al-Omeri, W., Jafari, S., Khan, M.Z., Khan, M.F., Aslam, M. and Mughal, A.R. (2020). New Challenges in Neutrosophic Theory and Applications. *Mathematics and Statistics Faculty and Staff Publications*, University of New Mexico.
15. Yager, R.R. and Zadeh, L.A. (2012). *An introduction to fuzzy logic applications in intelligent systems*. Vol. 165, Springer Science & Business Media.
16. Smarandache, F. (2014). *Introduction to Neutrosophic Statistics*. Sytech Publishing House, Craiava.
17. Alhabib, R., Ranna, M.M., Farah, H. and Salama, A.A. (2018). Some neutrosophic probability distributions. *Neutrosophic Sets and Systems*, 22, 30-38.
18. Alhasan, K.F.H. and Smarandache, F. (2019). Neutrosophic Weibull distribution and Neutrosophic Family Weibull Distribution. *Neutrosophic Sets and Systems*, 28, 192-194.
19. Biswas, P., Pramanik, S. and Giri, B.C. (2014). Entropy based grey relational analysis method for multi-attribute decision making under single valued neutrosophic assessments. *Neutrosophic Sets and Systems*, 2, 102-110.
20. Cincotta, P.M., Giordano, C.M., Silva, R.A. and Beaugé, C. (2021). The Shannon entropy: an efficient indicator of dynamical stability. *Physica D: Nonlinear Phenomena*, 417, 132816.
21. Pramanik, S., Dalapati, S., Alam, S., Smarandache, F. and Roy, T.K. (2018). NS-cross entropy-based MAGDM under single-valued neutrosophic set environment. *Information*, 9(2), 37.
22. Sethna, J. (2021). *Statistical mechanics: entropy, order parameters, and complexity*. Vol. 14. Oxford University Press, USA.
23. Misra, N., Singh, H. and Demchuk, E. (2005). Estimation of the entropy of a multivariate normal distribution. *Journal of Multivariate Analysis*, 92(2), 324-342.