Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 28 (2025), No. 5, pp. 855–878
DOI : 10.47974/JSMS-1387

A new goodness of fit test for normal distribution based on the negative cumulative extropy

Hadi Alizadeh Noughabi
Department of Statistics
University of Birjand
Birjand
Iran

Abstract

In recent years, extropy has emerged as a compelling alternative measure of uncertainty in statistical analysis. This paper focuses on the negative cumulative residual extropy, as introduced by Tahmasebi and Toomaj [28], and leverages this concept to develop a novel goodness-of-fit test for normality. We rigorously examine the statistical properties of this new test statistic, including its mean and variance. Furthermore, we establish the percentiles of the test statistic's distribution, allowing for practical implementation. To assess its effectiveness, we evaluate the power of the proposed test against a range of alternative distributions. Our simulation results reveal that the new test exhibits competitive power compared to existing methods. The test is computationally straightforward, and we demonstrate its practical application through the analysis of a real-world dataset.

Subject Classification (2000): *62G10, 62G30.*

Keywords: *Empirical distribution function, Negative cumulative extropy, Testing normality, Power study.*

1. Introduction

Information theory is a branch of applied mathematics that quantifies information, providing a framework for understanding communication systems and data encoding. It is vital for analyzing the efficiency of data transmission and storage, as it introduces key concepts such as entropy, which measures uncertainty or information content. Applications of

ORCID-ID: `0000-0002-7515-1896`

E-mail: `alizadehhadi@birjand.ac.ir`

information theory are extensive, including telecommunications, where it optimizes signal processing and error correction, as well as machine learning, where it aids in feature selection and model evaluation through metrics like mutual information. Additionally, information theory underpins advancements in data compression techniques and cryptography, ensuring secure communication in an increasingly digital world. By facilitating a deeper understanding of information flow and processing, information theory plays a pivotal role in modern technology and data science.

Consider a non-negative random variable, X, with distribution function, $F(x)$, and reliability function, $\overline{F}(x) = P(X > x)$. The concept of Cumulative Residual Entropy (CRE) was discussed by Rao et al. [26] as a measure of uncertainty associated with this random variable:

$$CRE(F) = -\int_0^\infty \overline{F}(x) \log \overline{F}(x) dx.$$

The CRE offers several advantages over traditional Shannon entropy (1948) [27], including its non-negativity and ease of calculation from sample data. Subsequent research by numerous authors, such as Rao [25], Asadi and Zohrevand [5], Di Crescenzo and Longobardi [7-9], Drissi et al. [10], Navarro et al. [16], Kapodistria and Psarrakos [13], Kayal [14], Psarrakos and Toomaj [20], and Navarro and Psarrakos [17], has explored various properties and extensions of the CRE. Toomaj et al. [30] further discussed the CRE in mixed and coherent systems with identically distributed component lifetimes.

In recent years, an alternative uncertainty measure known as extropy was proposed by Lad et al. [15]. For a random variable X, with non-negative support and with density function $f(\mathrm{x})$, extropy is given as:

$$J(X) = -\frac{1}{2}\int_0^\infty f^2(x) dx. \qquad (1-1)$$

Some properties of the above measure are presented by Lad et al. [15]. Further contributions have been made by Qiu [21], Qiu and Jia [22-23], and Qiu et al. [24]. Alizadeh Noughabi and Jarrahiferiz [4] investigated extropy estimation techniques. Paralleling this development, Jahanshahi et al. [12] introduced a cumulative residual extropy (CREX):

$$J^*(X) = -\frac{1}{2}\int_0^\infty \overline{F}^2(x) dx. \qquad (1-2)$$

They noted that CREX is non-positive. Subsequently, a dynamic CRE proposed by Abdul Sathar and Dhanya Nair [1]. Tahmasebi and Toomaj

[28] introduced a new measure, Negative Cumulative Extropy (NCEX), similar to but distinct from (1-2). They defined it as:

$$NCEX(X) = \frac{1}{2}\int_0^\infty [1 - F^2(x)]dx\,.$$

They calculated the NCEX for various well-known distributions, observing that for a uniform distribution over 0,1, the NCEX value is 1/3. Their analysis extended to coherent systems with identically distributed components. More recently, Alizadeh Noughabi [3] utilized negative cumulative extropy to develop a test for uniformity. Building on this, Alizadeh Noughabi [3] presented a theorem establishing that for density f with support [0,1], $0 \le NCEX(f) \le 1/2,$, with $NCEX(f) = 1/3$ uniquely associated with the uniform density. Leveraging this characteristic, Alizadeh Noughabi [3] proposed a test statistic T_n for uniformity assessment:

$$T_n = \frac{1}{2}\sum_{i=1}^{n-1}\left[1 - \left(\frac{i}{n}\right)^2\right](X_{(i+1)} - X_{(i)}),$$

Through simulation studies, they determined the power values of this test, demonstrating its competitive performance compared to other existing tests.

The normal distribution holds a position of paramount importance in both statistics and engineering due to its pervasive presence in natural phenomena and its elegant mathematical properties.

The normal distribution holds a pivotal position within the realm of statistics, serving as a foundational concept underpinning diverse statistical analyses. Hypothesis testing, confidence interval construction, and regression analysis frequently rely on the assumption that data adheres to a normal distribution. This assumption empowers researchers to draw inferences about entire populations based on observations from smaller samples. The symmetry inherent in the normal distribution, coupled with its well-defined mean and standard deviation, facilitates precise quantifications of variability and probabilistic occurrences.

Within engineering disciplines, the normal distribution finds wide applications in quality control, reliability analysis, and signal processing. For instance, manufacturers utilize it to monitor production processes, ensuring that product dimensions fall within acceptable tolerances. Engineers employ it to predict the lifespan of components and design systems that withstand expected variations. In signal processing, the

assumption of Gaussian noise allows for the development of filtering techniques to extract meaningful information from noisy signals.

The widespread applicability of this distribution is attributed to the Central Limit Theorem (CLT). This theorem postulates that when aggregating a substantial number of independent variables, irrespective of their initial distributions, the resulting sum or average converges towards a normal distribution. This principle underscores the profound significance of the normal distribution in modeling intricate systems where numerous factors coalesce to influence an ultimate outcome.

In essence, the normal distribution provides a powerful and versatile framework for understanding and quantifying uncertainty in both statistical analysis and engineering applications. Its wide applicability and elegant mathematical properties continue to make it a cornerstone of modern scientific inquiry.

In practical applications, ensuring that data meets the normality assumption can significantly influence decision-making processes. For example, in quality control settings, manufacturers may assume that product measurements follow a normal distribution to apply control charts effectively. In finance, risk assessment models often assume that asset returns are normally distributed to estimate potential losses accurately. Therefore, understanding and testing for normality not only enhances the robustness of statistical analyses but also serves as a foundation for making informed decisions based on empirical data.

The assumption of normality is fundamental to many statistical methods, some of which are highly sensitive to deviations from this assumption. Consequently, evaluating normality has become a cornerstone of goodness-of-fit testing. Let $X_1, ..., X_n$ be a random sample drawn from a population with a density function $f(x)$, the goal is to test the null hypothesis H_0 that these X_i values follow a normal distribution.

Numerous tests for H_0 exist in statistical literature, including those developed by D'Agostino and Stephens [6]. Recent contributions include Eidous and Abu-Hawwas [11], and Nwezza and Ugwuowo [18], who introduced an extended normal distribution tailored for reliability data analysis. Additionally, Abdul-Nabi et al. [2] explored shrinkage techniques for estimating the mean.

Goodness-of-fit (GOF) tests are fundamental tools in statistics used to assess how well a statistical model or distribution describes a set of observed data. They provide a quantitative measure of agreement between the empirical distribution of the data and the theoretical distribution assumed by the model. The core principle behind GOF tests is to compare

the observed frequencies of events with the expected frequencies predicted by the chosen model. Discrepancies between these observed and expected values are quantified using statistical metrics, such as chi-squared or Kolmogorov-Smirnov statistics. A small p-value resulting from these tests indicates that the observed data deviate significantly from the theoretical distribution, suggesting that the chosen model may not be an adequate representation of the underlying data generating process. In practice, goodness-of-fit tests find applications across diverse fields. In quality control, they help determine if a manufacturing process consistently produces products within specified tolerances. For instance, analyzing the distribution of product dimensions can reveal if a production line needs adjustment to meet quality standards. In healthcare, researchers utilize GOF tests to evaluate the effectiveness of treatments by comparing observed patient outcomes with expected outcomes based on historical data or theoretical models. In finance, these tests are applied to assess the fit of statistical models used for predicting asset prices or portfolio returns. By comparing observed market behavior with model predictions, investors can gain insights into the accuracy and limitations of their analytical frameworks. Ultimately, goodness-of-fit tests provide a crucial mechanism for evaluating the validity and reliability of statistical models, guiding informed decision-making across various domains.

This paper introduces a novel test based on NCEX to evaluate the hypothesis that a sample originates from a normal distribution. We delve into its theoretical properties and explore its performance through extensive simulations. Specifically, we determine percent points for our test statistic across diffeerent sample sizes using a Monte Carlo experiment with 100,000 sample values generated from different distributions. Finally, power of the NCEX-based test is computed and compared by competitors, revealing that it significantly outperforms some competitors for specific alternative hypotheses.

2. The proposed test

The normal distribution, with its characteristic bell shape, holds a central position in statistics due to its prevalence in natural phenomena and elegant mathematical properties. It underpins numerous statistical tests and provides a framework for understanding variability and probability. However, before applying these tools confidently, it's crucial to ensure that the data actually follows a normal distribution. This is GOF into play, specifically those tailored for the normal distribution.

When assessing whether a dataset aligns with a normal distribution, we employ GOF tests. These tests hinge on comparing the actual frequencies of data points against the expected frequencies if the data were indeed normally distributed, assuming specific mean and standard deviation parameters. Two widely used tests for this purpose are Kolmogorov-Smirnov test and chi-squared test. Both calculate a statistic that quantifies the divergence between observed and anticipated frequencies. This calculated statistic is then juxtaposed against critical values obtained from established probability distributions. If the resulting p-value is statistically significant (indicating a low probability of observing such discrepancies under normality), it suggests that the data deviates considerably from a normal distribution. In such cases, researchers may consider alternative statistical models or explore data transformations to better represent the underlying pattern.

In practice, verifying normality through goodness-of-fit tests is essential in various fields. For instance, in quality control, manufacturers rely on normal distribution assumptions for process monitoring and control charts. A failed normality test might signal a need to investigate process variability or adjust control limits. Similarly, in medical research, analyzing patient outcomes often assumes normality. If this assumption is violated, researchers might need to use non-parametric tests or explore alternative statistical models that accommodate skewed data.

Financial analysts frequently utilize normal distribution assumptions for risk assessment and portfolio optimization. A goodness-of-fit test can reveal if asset returns deviate from expected normality, potentially influencing investment strategies and risk management practices. Ultimately, the importance of normality testing lies in ensuring the validity of statistical inferences and guiding data analysis towards meaningful and accurate conclusions.

Let $X_1,...,X_n$ be a random sample drawn from a population with a probability distribution *F* with a density function $f(x)$, we are interested in testing the hypothesis:

$$H_0 : f(x) = f_0(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad \text{for some } (\mu,\sigma)\in\Theta,$$

where μ and σ are unspecified parameters and $\Theta = \mathbb{R}\times\mathbb{R}^+$. The alternative hypothesis is:

$$H_1 : f(x) \neq f_0(x;\mu,\sigma) \qquad \text{for any } (\mu,\sigma)\in\Theta.$$

Without loss of generality, we can simplify this GOF problem to testing for uniformity on the interval [0,1] using the probability integral transformation $U = F_0(X)$. Thus, if we define $U_i = F_0(X_i)$ for $i = 1, 2, ..., n$, the problem transforms into testing for uniformity:

$$H_0 : f(u) = 1, \quad 0 < u < 1$$

against

$$H_1 : f(u) \neq 1, \quad 0 < u < 1$$

.

We will utilize the test of uniformity proposed by Alizadeh Noughabi [3]. The test statistic is constructed as follows: Suppose that $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ are the order statistics of the sample. The NCEX is defined as:

$$NCEX(F) = NCEX(f) = \frac{1}{2}\int_0^{\infty}[1 - F^2(x)]dx = \int_0^1 \frac{\phi(\mathrm{u})}{f(\mathrm{F}^{-1}(\mathrm{u}))}du\,,$$

where $\phi(u) = \frac{1-u^2}{2}$, $0 < u < 1$.

Alizadeh Noughabi [3] demonstrated that for a density *f* concentrated on the unit interval, it holds that $0 \leq NCEX(f) \leq 1/2$, with the specific value of $1/3$ uniquely corresponding to the uniform distribution. Utilizing this property, we can formulate our test for H_0. For the hypothesis of uniformity, a consistent test can be expressed as:

$$T_n = NCEX(F_n),$$

where $NCEX(F_n)$ is the sample estimate of $NCEX(F)$ as suggested by Tahmasebi and Toomaj [28]:

$$NCEX(F_n) = \frac{1}{2}\int [1 - F_n^2(x)]dx,$$

and F_n being the empirical CDF given by

$$F_n(x) = \sum_{i=1}^{n-1} \frac{i}{n} I_{(X_{(i)}, X_{(i+1)}]}(x), \quad x \in \mathbb{R}$$

where I_A represents the indicator function for the set *A*. Consequently, we can express the our test statistic as:

$$T_n = NCEX(F_n) = \frac{1}{2}\sum_{i=1}^{n-1}\int_{X_{(i)}}^{X_{(i+1)}}\left[1 - \left(\frac{i}{n}\right)^2\right]dx = \frac{1}{2}\sum_{i=1}^{n-1}\left[1 - \left(\frac{i}{n}\right)^2\right](U_{(i+1)} - U_{(i)})$$

$$= \frac{1}{2}\sum_{i=1}^{n-1} W_i (U_{(i+1)} - U_{(i)}) = \frac{1}{2}\sum_{i=1}^{n-1} W_i Z_i ,$$

where $W_i = 1 - \left(\frac{i}{n}\right)^2$, $U_i = F_0(X_i;\, \hat{\mu},\, \hat{\sigma})$, $i = 1, 2, ..., n$ and $Z_i = U_{(i+1)} - U_{(i)}$, $i = 1, ..., n-1$, represent the transformed sample spacings. Here, $\hat{\mu}$ and $\hat{\sigma}$ are the MLEs of the parameters μ and σ, respectively.

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad ; \qquad \hat{\sigma} = s = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2} .$$

Under the null hypothesis H_0, as $n \to \infty$, T_n converges in probability to $1/3$; while under an alternative distribution, the test statistic T_n converges in probability to a value either less than or greater than $1/3$. For the significance level α and a sample size n, the critical region is as:

$$T_n \le T^*_{\alpha/2} \quad \text{or} \quad T_n \ge T^*_{1-\alpha/2} \qquad (2-1)$$

where $T^*_{\alpha/2}$ and $T^*_{1-\alpha/2}$ are determined to ensure that the test maintains the desired significance level α for a given n. For different values of α and n, the thresholds T^*_{α} can be derived using simulation.

Theorem 1: *Let $X_1, ..., X_n$ represent a sample drawn from a papulation with probability density function $f(x)$. Then, it follows that*

$$0 \le T_n \le 1/2.$$

Proof: Define the function $g(p) = \frac{1}{2}(1 - p^2)$, $0 < p < 1$. It can be easily demonstrated that the maximum value of $g(p)$ is $1/2$. Consequently

$$T_n = NCEX(F_n) = \frac{1}{2}\sum_{i=1}^{n-1}\left[1 - \left(\frac{i}{n}\right)^2\right](U_{(i+1)} - U_{(i)})$$

$$\le \frac{1}{2}\sum_{i=1}^{n-1}(U_{(i+1)} - U_{(i)}) = \frac{1}{2}(U_{(n)} - U_{(1)}) \le \frac{1}{2}.$$

It is also evident that $NCEX(F_n) \ge 0$. Hence, we conclude that $0 \le T_n \le 1/2$.

Theorem 2: *The suggested test statistic T_n is consistent.*

Proof: According to the Glivenko-Cantelli theorem, we have

$$\sup_x \left|F_n(x) - F(x)\right| \xrightarrow{a.s.} 0 \qquad \text{as } n \to \infty,$$

Tahmasebi and Toomaj [28] demonstrated that

$$NCEX(F_n) \to NCEX(F),$$

almost surely. Additionally, as $n \to \infty$, the MLEs $(\hat{\mu}, \hat{\sigma})$ approach (μ, σ). Thus, the proof is complete.

Theorem 3: *Under* H_0, *the expected value of the proposed test statistic* T_n *is given by*

$$E(T_n) = \frac{\sum_{i=1}^{n-1} W_i}{2(n+1)},$$

where $W_i = 1 - \left(\frac{i}{n}\right)^2$ *for* $i = 1, ..., n-1$.

Proof: It is obvious that the sample spacing $Z_i = U_{(i+1)} - U_{(i)}$, for a uniform distribution, follows a beta distribution, i.e., $Z_i \sim Beta(1, n)$, which implies that $E(Z_i) = (n+1)^{-1}$. Hence,

$$E(T_n) = E\left[\frac{1}{2}\sum_{i=1}^{n-1} W_i Z_i\right] = \frac{1}{2}\sum_{i=1}^{n-1} W_i E(Z_i) = \frac{1}{2(n+1)}\sum_{i=1}^{n-1} W_i.$$

From this theorem, we immediately find that

$$\lim_{n\to\infty} E(T_n) = \frac{1}{3} = NCEX(U).$$

Theorem 4: *Under* H_0, *the variance of* T_n *is*

$$Var(T_n) = \frac{n}{4(n+1)^2(n+2)}\sum_{i=1}^{n-1} W_i^2,$$

and the second moment of T_n *is*

$$E(T_n^2) = \frac{1}{2(n+1)^2}\left[\left(1 - \frac{1}{n+2}\right)\sum_{i=1}^{n-1} W_i^2 + \sum\sum_{i<j} W_i W_j\right].$$

Proof: Since $Z_i \sim Beta(1, n)$, we have $Var(Z_i) = \frac{n}{(n+1)^2(n+2)}$. This leads us to derive

$$Var(T_n) = Var\left[\frac{1}{2}\sum_{i=1}^{n-1} W_i Z_i\right] = \frac{1}{4}\sum_{i=1}^{n-1} W_i^2 Var(Z_i) = \frac{n}{4(n+1)^2(n+2)}\sum_{i=1}^{n-1} W_i^2.$$

$$E(T_n^2) = Var(T_n) + E^2(T_n) = \frac{n}{4(n+1)^2(n+2)} \sum_{i=1}^{n-1} W_i^2 + \frac{\left(\sum_{i=1}^{n-1} W_i\right)^2}{4(n+1)^2}$$

$$= \frac{\sum_{i=1}^{n-1} W_i^2 + \left(\sum_{i=1}^{n-1} W_i\right)^2}{4(n+1)^2} - \frac{\sum_{i=1}^{n-1} W_i^2}{2(n+1)^2(n+2)}$$

$$= \frac{\left[\sum_{i=1}^{n-1} W_i^2 + \sum\sum_{i<j} W_i W_j\right]}{2(n+1)^2} - \frac{\sum_{i=1}^{n-1} W_i^2}{2(n+1)^2(n+2)}$$

$$= \frac{1}{2(n+1)^2}\left[\left(1 - \frac{1}{n+2}\right)\sum_{i=1}^{n-1} W_i^2 + \sum\sum_{i<j} W_i W_j\right].$$

Remark 1: It is evident that $Var(T_n)$ approaches zero as $n \to \infty$.

3. Percentage points and power study

Critical values stand as fundamental elements in hypothesis testing, including GOF tests, acting as benchmarks to gauge the strength of evidence against a null hypothesis. In a GOF test, H_0 typically asserts that the data adheres to a specific distribution. The critical value functions as a threshold: if the test statistic surpasses this value, we reject H_0, indicating a poor fit between the assumed distribution and the observed data.

In practice, these critical values guide decision-making regarding model fit. For example, a financial analyst analyzing asset returns might use a chi-squared GOF test to assess whether the data follows a normal distribution. When the calculated statistic exceeds the critical point, the analyst would conclude that the normality assumption is inadequate and explore alternative models for risk assessment.

Computation of critical points for test statistics often relies on theoretical distributions (like chi-squared or Kolmogorov-Smirnov) and statistical tables. These tables provide pre-calculated critical values for various degrees of freedom and significance levels. Alternatively, statistical software packages can efficiently compute these critical values based on user-specified parameters.

Determining the crucial thresholds, or critical values, for a test statistic often necessitates a reliance on simulations, especially when analytical

solutions prove elusive or computationally demanding. The process commences by generating numerous random samples, each mimicking the presumed population distribution. For every generated sample, we calculate the test statistic using our proposed method. By aggregating these calculated values across all simulations, we effectively build under H_0 a simulated distribution of the statistic. From this simulated distribution, we can pinpoint critical values corresponding to specific significance levels. For instance, if we aim for a 5% significance level, we would identify the values within the simulated distribution that demarcate the upper 5% of the simulated test statistic's range. These determined critical values then function as benchmarks for making decisions about whether to reject or retain the null hypothesis when analyzing real-world datasets. Simulation-based critical values provide a versatile and robust strategy, particularly advantageous when confronting intricate test statistics or non-standard distributions where analytical solutions are difficult to obtain.

Through a simulation study, we derive the percentage points. For given sample size n, we generate 100,000 samples from the standard normal distribution. Then we compute the test statistic. For a significance level α, the lower and upper tail percentage points $T^*_{\alpha/2}$ and $T^*_{1-\alpha/2}$ of the distribution of T_n are determined using the $\alpha/2$ and $1-\alpha/2$ percentiles from the empirical distribution function of T_n using the 100,000 samples collected. The critical points are summarized in Table 1. Moreover, in Figure 1, we present the empirical density functions of the test statistic created using 100,000 samples with sizes n=10, 20, 30, 50, and 100 drawn from the standard normal distribution.

Table 1

Critical points of the new test statistic for $\alpha = 0.05$.

n	lower	upper
10	0.2336	0.3323
20	0.2705	0.3370
30	0.2865	0.3378
40	0.2961	0.3379
50	0.3022	0.3379
75	0.3105	0.3377
100	0.3152	0.3374

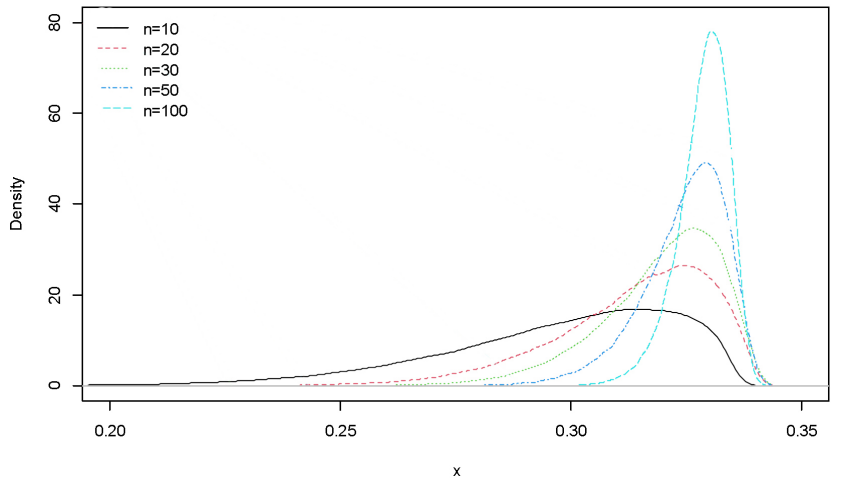

Figure 1

Empirical density functions of the test statistic created using 100,000 samples with sizes n=10, 20, 30, 50, and 100 drawn from the standard normal distribution.

While critical values inform our decision-making based on observed data, power studies provide crucial insights into the sensitivity of our tests. In essence, it quantifies the ability of a test to detect a genuine deviation from the assumed distribution. Conducting a power study for a GOF test involves simulating data under various scenarios where the alternative hypothesis (the data doesn't follow the assumed distribution) holds true. By varying factors like sample size, effect size (magnitude of deviation), and significance level, we can estimate the power of the test under different conditions.

The importance of power studies in practice stems from their ability to guide methodological choices and ensure reliable results. For instance, a researcher might discover that their chosen test lacks sufficient power to detect small deviations from normality. This insight could lead to increasing sample size, adopting a more sensitive test statistic, or adjusting the significance level to achieve an acceptable level of power.

In finance, analysts utilize power studies to assess the ability of their statistical models to detect market anomalies or predict future trends. A low power value might indicate that the model is insufficiently robust to capture subtle shifts in market behavior, necessitating model refinements or alternative approaches.

Calculating power values often involves numerical methods and statistical software packages. Popular approaches include using formulas based on the chosen test statistic's distribution or employing simulation techniques to generate data under the alternative hypothesis and calculate the rejection probability. Ultimately, incorporating power considerations into goodness-of-fit tests ensures that our analyses are not only statistically sound but also practically meaningful. By understanding the limitations of our tests and striving for adequate power, we can draw more confident conclusions and also make informed decisions based on reliable evidence.

Calculating the power of a test often involves simulations. We can simulate data under different scenarios (e.g., with varying effect sizes) and see how often our test correctly rejects the null hypothesis when it's false. This gives us an estimate of the power for different situations. We compute the power of our test against different alternatives by simulations. We compare these power values with those of existing tests, specifically focusing on well-established tests commonly applied in practice and available in statistical software. A brief description of the test statistics for these comparisons is provided below. For more comprehensive information about these tests, refer to D'Agostino and Stephens [6].

Goodness-of-fit tests play a crucial role in statistical analysis by assessing how well a theoretical distribution aligns with observed data. Several powerful tests are based on the empirical distribution function (EDF), which is a step function that estimates the cumulative distribution function (CDF) of a random variable. These EDF-based tests quantify the discrepancy between the empirical CDF of the sample and the CDF of the hypothesized distribution. Several such tests are widely used, including the Cramer-von Mises, Kolmogorov-Smirnov, Kuiper, and Anderson-Darling statistics. Each of these tests employs a different approach to measure the discrepancies between the empirical and hypothesized CDFs, making them sensitive to different types of deviations.

Suppose that $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ denote the order statistics derived from the sample $X_1, ..., X_n$.

1. **The Cramer-von Mises statistic:** The Cramer-von Mises statistic (W^2) is a quadratic EDF statistic that measures the integrated squared difference between the empirical and hypothesized CDFs. It calculates the sum of the squared deviations between the empirical CDF and the hypothetical CDF at all observed data points. The W^2 statistic is calculated by:

$$W^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left(\frac{2i-1}{2n} - F_0(X_{(i)};\hat{\mu},\hat{\sigma})\right)^2.$$

The W^2 test is generally considered to be a robust measure of the overall discrepancy between the empirical and hypothesized distributions and it is sensitive to deviations in the center of the distribution.

2. **The Kolmogorov-Smirnov statistic:** The KS statistic (D), in contrast to the W^2 statistic, is a supremum-type statistic. It quantifies the maximum absolute difference between the empirical and hypothesized CDFs. The D statistic is given by:

$$D = \max(D^+, D^-),$$

where

$$D^+ = \max_{1\le i\le n}\left\{\frac{i}{n} - F_0(X_{(i)};\hat{\mu},\hat{\sigma})\right\};\ D^- = \max_{1\le i\le n}\left\{F_0(X_{(i)};\hat{\mu},\hat{\sigma}) - \frac{i-1}{n}\right\}.$$

The KS test is more sensitive to deviations in the tail regions of the distribution and is well known for being simple to implement but considered less powerful than the other tests for many alternatives. It's a non-parametric test, making it versatile across various distributions.

3. **The Kuiper statistic:** The Kuiper statistic, similar to the Kolmogorov-Smirnov statistic, is also a supremum-type statistic but considers the circular nature of the distribution. It calculates the sum of the maximum positive and maximum negative deviations between the empirical and hypothesized CDFs, thus making the test more sensitive to deviations in the tail regions and robust to changes in the scale of distribution. The statistic is calculated as follows:

$$V = D^+ + D^-.$$

4. **The Anderson-Darling statistic:** The Anderson-Darling (A^2) statistic is another quadratic EDF statistic but, unlike the statistic W^2, gives more weight to the tails of the distribution. This makes the test A^2 particularly sensitive to deviations in the tails of the distribution. The statistic is calculated as follows:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\{\log F_0(X_{(i)};\hat{\mu},\hat{\sigma}) + \log[1 - F_0(X_{(n-i+1)};\hat{\mu},\hat{\sigma})]\}.$$

The Anderson-Darling test is widely recognized as a powerful test for goodness-of-fit, especially when detecting departures from normality in the tails.

Furthermore, we include the well-known Shapiro-Wilk test (SW) in our power comparisons:

$$SW = \frac{\left(\sum_{i=1}^{\lceil n/2 \rceil} a_{(n-i+1)}(X_{(n-i+1)} - X_{(i)})\right)^2}{\sum_{i=1}^{n}(X_{(i)} - \overline{X})^2},$$

where the coefficients a_i can be found in Pearson and Hartley [19]. The test SW is a powerful and widely used statistical test specifically designed to assess the normality of a dataset. Unlike the EDF-based tests, such as the D or W^2, the SW test is based on the correlation between the sample order statistics and the corresponding expected order statistics from a normal distribution. This approach makes it particularly sensitive to departures from normality, especially in the tails of the distribution, and it is generally considered one of the most powerful normality tests when dealing with small to moderate sample sizes. The Shapiro-Wilk test, SW, is computed by comparing the sample order statistics to the expected order statistics under the assumption of normality. A SW value close to 1 indicates that the data are consistent with the normality assumption, while values significantly less than 1 suggest a departure from normality.

The Shapiro-Wilk test is known for its high power against various non-normal alternatives, particularly for skewed distributions and departures in the tails. It is generally considered more powerful than the Kolmogorov-Smirnov test and some other EDF-based tests for detecting non-normality. However, the Shapiro-Wilk test is not without limitations. Its calculation becomes more computationally demanding as sample sizes increase, and the test can become less accurate with very large datasets. Additionally, for distributions with heavy tails, the Shapiro-Wilk test might have reduced sensitivity.

In order to evaluate the power values of our test, we examine its behavior against several alternative distributions. These alternatives are chosen to represent a variety of shapes and characteristics, providing a comprehensive assessment of the test's sensitivity to departures from

normality. The selected distributions include those with exponential, gamma, lognormal, extreme value, and heavy-tailed properties, among others, and they are common in modeling various real-world phenomena. We examine the following alternative distributions for our power study of the tests:

- the exponential distribution, $Exp(1)$;
- the Gamma distribution, $\Gamma(0.5,1)$ and $\Gamma(2,1)$;
- the lognormal distribution, $LN(0,1)$;
- the Gumbel distribution, $Gu(0,1)$;
- the Weibull distribution, $W(0.5,1)$ and $W(2,1)$;
- the inverse Gaussian distribution, $IG(1,0.5)$, $IG(1,1)$ and $IG(1,2)$;
- the student's t distribution, $t(3)$;
- the Cauchy distribution, $C(0,1)$;
- the skew normal distribution, $SN(0,1,2)$ and $SN(0,1,3)$.

Table 2

Empirical power of the normality tests at level of 5%.

Alternative	n	W^2	D	V	A^2	SW	T_n
$Exp(1)$	10	0.3819	0.2954	0.3552	0.4078	0.4424	0.4984
	20	0.7243	0.5814	0.6946	0.7759	0.8360	**0.8962**
	30	0.8972	0.7854	0.8870	0.9347	0.9678	**0.9877**
	50	0.9906	0.9629	0.9904	0.9964	0.9994	**0.9999**
$\Gamma(0.5,1)$	10	0.6625	0.5311	0.6557	0.6937	0.7318	0.7816
	20	0.9523	0.8830	0.9549	0.9698	0.9840	**0.9929**
	30	0.9954	0.9834	0.9968	0.9983	0.9996	**0.9999**
	50	1.0000	0.9999	1.0000	1.0000	1.0000	**1.0000**
$\Gamma(2,1)$	10	0.2037	0.1667	0.1771	0.2194	0.2385	0.2709
	20	0.4164	0.3254	0.3504	0.4629	0.5296	**0.6029**
	30	0.5970	0.4693	0.5159	0.6616	0.7474	**0.8274**
	50	0.8377	0.7008	0.7683	0.8921	0.9482	**0.9804**
$LN(0,0.5)$	10	0.2175	0.1808	0.1865	0.2318	0.2493	0.2667
	20	0.4270	0.3418	0.3539	0.4662	0.5210	**0.5681**
	30	0.6028	0.4902	0.5092	0.6557	0.7242	**0.7714**
	50	0.8278	0.7131	0.7398	0.8719	0.9234	**0.9515**

Contd...

$LN(0,1)$	10	0.5524	0.4566	0.5222	0.5756	0.6045	**0.6504**
	20	0.8820	0.7941	0.8596	0.9057	0.9320	**0.9562**
	30	0.9739	0.9335	0.9665	0.9839	0.9914	**0.9963**
	50	0.9992	0.9958	0.9988	0.9997	1.0000	**1.0000**
$LN(0,2)$	10	0.8940	0.8198	0.8912	0.9073	0.9228	**0.9374**
	20	0.9975	0.9913	0.9976	0.9985	0.9992	**0.9996**
	30	1.0000	0.9997	1.0000	1.0000	1.0000	**1.0000**
	50	1.0000	1.0000	1.0000	1.0000	1.0000	**1.0000**
$Gu(0,1)$	10	0.1342	0.1169	0.1154	0.1430	0.1511	**0.1610**
	20	0.2479	0.2042	0.1947	0.2725	0.3135	**0.3314**
	30	0.3504	0.2811	0.2670	0.3889	0.4576	**0.4844**
	50	0.5490	0.4421	0.4205	0.6036	0.6888	**0.7161**
$W(0.5,1)$	10	0.8570	0.7534	0.8557	0.8757	0.8973	**0.9176**
	20	0.9954	0.9838	0.9961	0.9973	0.9989	**0.9996**
	30	0.9999	0.9996	1.0000	1.0000	1.0000	**1.0000**
	50	1.0000	1.0000	1.0000	1.0000	1.0000	**1.0000**
$W(2,1)$	10	0.0757	0.0694	0.0688	0.0796	0.0850	**0.0884**
	20	0.1195	0.1038	0.0950	0.1309	0.1537	**0.1783**
	30	0.1643	0.1382	0.1239	0.1877	0.2346	**0.2854**
	50	0.2626	0.2101	0.1815	0.3110	0.4134	**0.5407**
$IG(1,0.5)$	10	0.6737	0.5657	0.6502	0.6975	0.7254	**0.7653**
	20	0.9524	0.8943	0.9447	0.9650	0.9775	**0.9874**
	30	0.9947	0.9813	0.9943	0.9973	0.9990	**0.9996**
	50	1.0000	0.9997	0.9999	1.0000	1.0000	**1.0000**
$IG(1,1)$	10	0.4983	0.4023	0.4657	0.5232	0.5531	**0.6006**
	20	0.8433	0.7325	0.8093	0.8731	0.9061	**0.9401**
	30	0.9606	0.9012	0.9480	0.9751	0.9874	**0.9942**
	50	0.9983	0.9905	0.9971	0.9993	0.9999	**1.0000**
$IG(1,2)$	10	0.3260	0.2628	0.2924	0.3469	0.3724	**0.4118**
	20	0.6341	0.5140	0.5685	0.6789	0.7354	**0.7911**
	30	0.8239	0.7061	0.7688	0.8657	0.9092	**0.9430**
	50	0.9682	0.9082	0.9461	0.9824	0.9932	**0.9976**
$t(3)$	10	0.1791	0.1592	0.1595	0.1851	**0.1863**	0.1396
	20	0.3039	0.2588	0.2760	0.3242	**0.3378**	0.1548
	30	0.4085	0.3440	0.3767	0.4378	**0.4594**	0.1543
	50	0.5748	0.4864	0.5413	0.6071	**0.6368**	0.1686

Contd...

$C(0,1)$	10	**0.6135**	0.5767	0.5876	0.6127	0.5928	0.2956
	20	0.8779	0.8439	0.8632	**0.8802**	0.8656	0.3328
	30	0.9634	0.9450	0.9579	**0.9649**	0.9585	0.3814
	50	0.9971	0.9941	0.9962	**0.9974**	0.9965	0.6756
$SN(0,1,2)$	10	0.0677	0.0639	0.0628	0.0699	0.0716	**0.0737**
	20	0.0890	0.0829	0.0754	0.0947	0.1040	**0.1051**
	30	0.1110	0.0992	0.0886	0.1205	0.1368	**0.1383**
	50	0.1556	0.1343	0.1125	0.1702	0.2013	**0.2062**
$SN(0,1,3)$	10	0.0857	0.0780	0.0758	0.0891	0.0934	**0.1005**
	20	0.1396	0.1208	0.1077	0.1513	0.1697	**0.1834**
	30	0.1990	0.1669	0.1494	0.2197	0.2518	**0.2639**
	50	0.3139	0.2529	0.2188	0.3469	0.4030	**0.4271**

By simulation we evaluate the power of considered tests across these alternatives. For each alternative, we generate 100,000 samples of sizes 10, 20, 30, and 50, calculating the respective test statistics. The power of each test is then determined by the frequency with which the statistic falls within the critical region. The resulting power are presented in Table 2 for $\alpha = 0.05$ and sample sizes of 10, 20, 30, and 50.

In these tables, bold type represents the tests that achieved the highest power for each alternative.

Comparing the power values of GOF tests designed for assessing normality is crucial for selecting the most appropriate test for a given situation. Since each test statistic and its underlying distribution have unique properties, their sensitivity to deviations from normality can vary significantly. Consider a scenario where we need to determine if a dataset come from a normal distribution or not. The chi-squared test, often favored for categorical data, might not be as sensitive as the Kolmogorov-Smirnov test when dealing with continuous data. The latter directly compares the theoretical normal distribution with the empirical CDF of the observed data, potentially capturing subtle deviations missed by the chi-squared test.

Power comparison helps us identify which test is more likely to correctly reject the null hypothesis (data follows a normal distribution) when it's actually false. This becomes particularly important when dealing with small sample sizes or weak deviations from normality. A less powerful test might fail to detect a genuine departure from normality, leading to potentially misleading conclusions.

In practice, researchers and analysts often consult power tables or simulation studies to compare the power values of various GOF tests under different conditions. These resources provide insights into how factors like sample size, effect size (magnitude of deviation), and chosen significance level influence the test's ability to detect deviations from normality.

By carefully considering the power comparison, researchers can select the most appropriate test for their specific needs, ensuring that they have a high probability of detecting true departures from normality and making accurate inferences about their data.

From Table 2, it is clear that the new test T_n demonstrates strong performance and also it is a powerful test overall (except when Cauchy and $t(3)$ were the alternatives). The differences in power between our test and other tests are significant. Thus, we confidently recommend the proposed test T_n for practical use.

4. Application to a real data set

The application of goodness-of-fit (GOF) tests to real-world datasets is paramount for verifying the normality assumption, which is fundamental to the validity and reliability of many statistical analyses. Numerous statistical methods, including t-tests, ANOVA, and regression analysis, are predicated on the assumption that the underlying data is normally distributed. When this assumption is violated, the results of these methods can be unreliable and misleading. For example, consider a dataset of exam scores from a class. If we incorrectly assume these scores are normally distributed, we might misinterpret the average score, incorrectly identify significant differences between groups, or produce inaccurate individual performance predictions. A skewed distribution of scores, for instance, could invalidate conclusions drawn from analyses based on the normality assumption.

A GOF test provides a formal and objective evaluation of the agreement between the distribution of observed data and a theoretical normal distribution. When a GOF test reveals a significant departure from normality, it necessitates a careful reconsideration of the chosen analytical approach. This might involve utilizing alternative statistical methods that are robust to departures from normality, such as non-parametric tests, or applying transformations to the data to achieve a closer approximation to a normal distribution.

In practical applications, GOF tests for normality function as essential gatekeepers before proceeding with more complex statistical analyses. These tests ensure that our statistical inferences are grounded in a valid understanding of the underlying data distribution, thereby preventing us from drawing inaccurate conclusions based on flawed assumptions. This next section demonstrates the application of our new test to real-world data.

Example 1: We utilize the leghorn chicks data set analyzed by Tavakoli et al. [29], which includes the weights (in grams) of $n = 20$ twenty-one-day-old leghorn chicks. The authors found that the normal distribution adequately fits this data at the 5% significance level. The data is presented in Table 3, and Figure 2 illustrates the histogram of the chicks' weights in kilograms.

Table 3

The weights of the chicks data

156	162	168	182	186	190	190	196	202	210
214	220	226	230	230	236	236	242	246	270

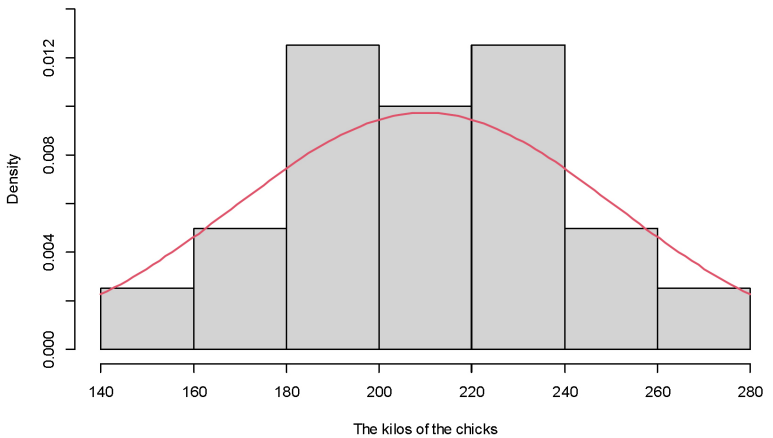

Figure 2

Histogram representing the data set from Example 1.

Using the new test, we evaluate whether this data originates from a normal distribution. The calculated value of the new test statistic is 0.3167,

while the lower and upper tail percentage points at the 5% level, as indicated in Table 1, are 0.2705 and 0.3370, respectively.

Since the computed value of the new test statistic does not fall within the critical region, we do not reject the normality assumption, which is consistent with previous findings.

5. Conclusions

This study embarked on an exploration of negative cumulative extropy for continuous random variables, ultimately leading to the development of a new GOF test for normality. This new test hinges on an estimator of negative cumulative extropy and boasts a simple computational structure. We rigorously demonstrated its consistency and elucidated its mean, variance, and other essential statistical properties.

To assess its efficacy, we provided critical values and conducted power simulations against various alternative distributions for diverse sample sizes. These simulations, comparing our NCEX test to established methods, reveal its robustness in evaluating normality. Notably, the NCEX test demonstrates significantly superior power compared to several competing tests.

Despite the promising results obtained in this study, it is important to acknowledge certain limitations. First, the power analysis was conducted using a specific set of alternative distributions. Future studies should explore the performance of our test against a wider variety of non-normal alternatives, including those with more complex characteristics. Additionally, the analysis was primarily conducted using Monte Carlo simulations. While this provides a solid basis for evaluation, the test's practical performance in real-world applications, particularly with complex or high-dimensional datasets, should be investigated further. Another limitation is that our proposed test relies on the assumption of continuous random variables, and this assumption may not always hold true in practical scenarios. Hence, further investigation is required for discrete distributions, as well as data sets with missing values.

Recognizing both the strengths and limitations of this study, several promising avenues for future research have been identified. One important direction involves exploring the test's performance with different sample sizes, particularly focusing on very small and very large sample sizes. Additionally, investigating the test's robustness to outliers and its behavior under various types of data contamination is a worthwhile pursuit. Given that the proposed test is based on extropy, further investigation into this

measure, specifically the residual version, as a method for detecting different aspects of departures from normality, is of particular importance. Comparing the test against new methods that have recently shown good power for testing normality should be investigated. A crucial extension would be to examine the applicability of the new test in different statistical contexts such as in regression analysis, where checking normality of residuals is essential.

The proposed test, based on negative cumulative extropy, exhibits a strong potential for various real-world applications. Future studies should explore its applicability in different statistical contexts such as in regression analysis, where checking the normality of residuals is essential. Further modifications of the test can also lead to more practical tests for different distributions. In conclusion, the novel approach introduced in this study presents a valuable new tool for assessing normality, with further development holding significant promise for future applications and extensions in the field of statistical inference.

Acknowledgements

The author extends sincere gratitude to the anonymous reviewers and the associate editor for their insightful feedback on a preceding manuscript draft. Their contributions were instrumental in refining this work.

References

[1] E.I. Abdul Sathar and R. Dhanya Nair, "On dynamic survival extropy," *Communications in Statistics - Theory and Methods,* vol. 50, pp. 1295-1313 (2021).

[2] A.I. Abdul-Nabi, A.N. Salman, and M.M. Ameen, "Employ shrinkage technique during estimate normal distribution mean," *Journal of Interdisciplinary Mathematics,* vol. 24, no. 6, pp. 1685–1692 (2021).

[3] N. Alizadeh Noughabi, "Testing uniformity based on negative cumulative extropy," *Communications in Statistics - Theory and Methods,* vol. 52, pp. 4998-5009 (2023).

[4] N. Alizadeh Noughabi and H. Jarrahiferiz, "On the estimation of extropy," *Journal of Nonparametric Statistics,* vol. 31, pp. 88-99 (2019).

[5] M. Asadi and Y. Zohrevand, "On the dynamic cumulative residual entropy," *Journal of Statistical Planning and Inference,* vol. 137, pp. 1931–1941 (2007).

[6] R.B. D'Agostino and M.A. Stephens, Eds., Goodness-of-fit Techniques, New York: Marcel Dekker (1986).

[7] A. Di Crescenzo and M. Longobardi, "On cumulative entropies," *Journal of Statistical Planning and Inference*, vol. 139, pp. 4072-4087 (2009).

[8] A. Di Crescenzo and M. Longobardi, "On weighted residual and past entropies," *Scientiae Mathematicae Japonicae*, vol. 64, pp. 255–266 (2006).

[9] A. Di Crescenzo and M. Longobardi, "On cumulative entropies and lifetime estimations," in IWINAC 2009, Part I, J. Mira et al., Eds., LNCS 5601, Springer, Berlin, pp. 132–141 (2009).

[10] N. Drissi, T. Chonavel, and J.M. Boucher, "Generalized cumulative residual entropy for distributions with unrestricted supports," Research Letters in Signal Processing, article ID 790607, 5 p. (2008).

[11] O. Eidous and J. Abu-Hawwas, "An accurate approximation for the standard normal distribution function," *Journal of Information and Optimization Sciences*, vol. 42, no. 1, pp. 17–27 (2021).

[12] S.M.A. Jahanshahi, H. Zarei, and A.H. Khammar, "On cumulative residual extropy," *Probability in the Engineering and Informational Sciences*, vol. 34, pp. 605-625 (2020).

[13] S. Kapodistria and G. Psarrakos, "Some extensions of the residual lifetime and its connection to the cumulative residual entropy," *Probability in the Engineering and Informational Sciences,* vol. 26, pp. 129-146 (2012).

[14] S. Kayal, "On generalized cumulative entropies," Probability in the Engineering and Informational Sciences, vol. 30, pp. 640-662 (2016).

[15] F. Lad, G. Sanfilippo, and G. Agro, "Extropy: Complementary dual of entropy," *Statistical Science,* vol. 30, pp. 40–58 (2015).

[16] J. Navarro, Y. del Aguila, and M. Asadi, "Some new results on the cumulative residual entropy," *Journal of Statistical Planning and Inference,* vol. 140, pp. 310-322 (2010).

[17] J. Navarro and G. Psarrakos, "Characterizations based on generalized cumulative residual entropy functions," *Communication in Statistics - Theory and Methods*, vol. 46, pp. 1247-1260 (2017).

[18] E.E. Nwezza and F.I. Ugwuowo, "An extended normal distribution for reliability data analysis," *Journal of Statistics and Management Systems*, vol. 25, no. 2, pp. 369–392 (2022).

[19] E.S. Pearson and H.O. Hartley, Biometrika Tables for Statisticians, Cambridge, U.K.: Cambridge University Press (1972).

[20] G. Psarrakos and A. Toomaj, "On the generalized cumulative residual entropy with applications in actuarial science," *Journal of Computational and Applied Mathematics*, vol. 309, pp. 186-199 (2017).

[21] G. Qiu, "The extropy of order statistics and record values," *Statistics & Probability Letters*, vol. 120, pp. 52-60 (2017).

[22] G. Qiu and K. Jia, "Extropy estimators with applications in testing uniformity," *Journal of Nonparametric Statistics*, vol. 30, pp. 182-196 (2018).

[23] G. Qiu and K. Jia, "The residual extropy of order statistics," *Statistics & Probability Letters,* vol. 133, pp. 15-22 (2018).

[24] G. Qiu, L. Wang, and X. Wang, "On extropy properties of mixed systems," *Probability in the Engineering and Informational Sciences,* vol. 33, pp. 471–486 (2019).

[25] M. Rao, "More on a new concept of entropy and information," *Journal of Theoretical Probability*, vol. 18, pp. 967–981 (2005).

[26] M. Rao, Y. Chen, B.C. Vemuri, and F. Wang, "Cumulative residual entropy: a new measure of information," *IEEE Transactions on Information Theory*, vol. 50, pp. 1220–1228 (2004).

[27] C.E. Shannon, "A mathematical theory of communications," *Bell System Technical Journal,* vol. 27, pp. 379–423; 623–656 (1948).

[28] S. Tahmasebi and A. Toomaj, "On negative cumulative extropy with applications," *Communications in Statistics - Theory and Methods,* vol. 51, pp. 5025-5047 (2022).

[29] M. Tavakoli, N. Alizadeh Noughabi, and G.R. Mohtashami Borzadaran, "An estimation of Phi divergence and its application in testing normality," *Hacettepe Journal of Mathematics and Statistics*, vol. 49, pp. 2104-2118 (2020).

[30] A. Toomaj, S.M. Sunoj, and J. Navarro, "Some properties of the cumulative residual entropy of coherent and mixed systems," *Journal of Applied Probability*, vol. 54, pp. 379–393 (2017).

Received December, 2023