**JIRSS (2025)**
**Vol. 24, No. 01, pp 11-34**
**DOI: 10.22034/jirss.2025.2049691.1092**

# New Estimations for Varextropy under Complete Data and Uniformity Testing

**Faranak Goodarzi** [1] **and Raheleh Zamini** [2]
[1] Faculty of Mathematics, Department of Statistics, University of Kashan, Kashan, Iran.
[2] Department of Mathematics, Faculty of Mathematical Sciences and Computer, Kharazmi University, Tehran, Iran.

Received: 03/01/2025, Accepted: 09/06/2025, Published online: 26/06/2025

**Abstract.** Recently Alizadeh Noughabi and Shafaei Noughabi (2024) introduced some estimators for the varextropy of an absolutely continuous random variable. In this paper, we propose other nonparametric estimators for the varextropy function. Additionally, we prove asymptotic properties of two estimators given in Alizadeh Noughabi and Shafaei Noughabi (2024). Asymptotic properties of the proposed estimators are established under suitable regularity conditions. Moreover, a simulation study is performed to compare the performance of the proposed estimators based on mean squared error (MSE) and bias. Furthermore, by using the proposed estimators some tests are constructed for uniformity. It is shown that the varextropy-based test proposed in this paper performs well in terms of power when compared to other uniformity hypothesis tests. Real datasets are utilized to evaluate the performance of the varextropy estimators.
**Keywords.** Varextropy Estimator, Consistency, Goodness-of-fit Test, Testing Uniformity, Monte Carlo Simulation.

**MSC:** 94A15, 62G05.

## 1 Introduction

Suppose that $X$ be a non-negative and an absolutely continuous random variable with the probability density function (pdf) $f$, distribution function $F$ and survival function $\overline{F}$. A prevalent uncertainty measure is the expectation of $-\log f(X)$, as defined by

Corresponding Author: Faranak Goodarzi (f-goodarzi@kashanu.ac.ir)
Raheleh Zamini (zamini@khu.ac.ir).

Shannon (1948), and given by

$$H(X) = -\int_0^{+\infty} f(x) \log f(x) dx. \tag{1.1}$$

Many researchers have estimated the entropy of continuous random variables in their studies. Some of these researchers are Vasicek (1976), van Es (1992), Ebrahimi et al. (1994), Correa (1995) and Zamanzade (2015). The varentropy of a random variable $X$ is defined as

$$\begin{aligned} VE(X) &= Var[-\log f(X)] \\ &= \int_0^{+\infty} f(x)[\log f(x)]^2 dx - \left[\int_0^{+\infty} f(x) \log f(x) dx\right]^2. \end{aligned} \tag{1.2}$$

Hence the variability in the information content of $X$ is measured by the varentropy. The relevance of this measure to physics, computer and mathematics sciences has been pointed out in various studies, including Madiman and Barron (2007), Jiang et al. (2011), Fradelizi et al. (2016) and Li et al. (2016). Recently Alizadeh Noughabi and Shafaei Noughabi (2023) proposed some estimators for varentropy of a continuous random variable.

The notion of entropy is recently intertwined with a complementary dual measure, designated as extropy, by Lad et al. (2015). The extropy of the random variable $X$ is defined as:

$$J(X) = -\frac{1}{2}\int_0^{+\infty} f^2(x) dx. \tag{1.3}$$

It is clear that $J(X) \leq 0$. Scoring the forecasting distributions using the total log-scoring rule is one of the statistical applications of extropy (see Gneiting and Raftery (2007)). Furthermore, the extropy has been universally investigated in commercial or scientific areas such as astronomical measurements of heat distributions in galaxies, (see Furuichi and Mitroi (2012) and Vontobel (2013)). The concept of extropy is useful in automatic speech recognition (Refer to Becerra et al. (2018)). Moreover, extropy is a measure better than entropy in some scenarios in statistical mechanics and thermodynamics. For more studies on extropy, see Qiu and Jia (2018b), Yang et al. (2019). Recently, the problem of estimating $J(X)$ has been noted by Qiu and Jia (2018a), Alizadeh Noughabi and Jarrahiferiz (2019) and Al-Labadi and Berry (2022).

Vaselabadi et al. (2021) defined the varextropy for the absolutely continuous random variable $X$ as:

$$VJ(X) = Var\left[-\frac{1}{2}f(X)\right] = \frac{1}{4}E[f^2(X)] - J^2(X). \tag{1.4}$$

Moreover, Goodarzi (2024) obtained lower bounds for varextropy. The varextropy measure can be used as an alternative measure to Shannon entropy and a measure of the amount of uncertainty.

The rest of the work is structured as follows. In Section 2, we propose some new nonparametric estimators for the varextropy function and investigate the consistency of these estimators under suitable conditions. Moreover, the convergence in probability of the fourth estimator and the almost sure convergence of the fifth estimator, which are the second and third estimators proposed by Alizadeh Noughabi and Shafaei Noughabi (2024), respectively, are rigorously proven in Section 2. We present the results of the Monte Carlo studies on biases and mean squared errors (MSEs) of the proposed estimators in Section 3. In Section 4, we introduce some goodness-of–fit tests for uniformity by using our proposed estimators and compare the powers of our proposed test statistics with some known test statistics and the two estimators Alizadeh Noughabi and Shafaei Noughabi (2024). We apply the newly introduced test procedures to real data sets for illustration in Section 5.

## 2 Varextropy Estimation for Complete Data

In this section, we propose three new estimators for varextropy function under complete data and investigate some of their asymptotic properties. We also prove asymptotic properties of two estimators given in Alizadeh Noughabi and Shafaei Noughabi (2024).

### 2.1 The First Estimator

Before constructing the first estimator for $VJ(X)$, we first review some required concepts and notations. Let $X$ be a continuous random variable with density function $f$ and distribution function $F$. Consider the class of functions based on $f(x)$ as

$$T(f) = \int_{-\infty}^{+\infty} f(x)\Phi(f(x))w(x)dx, \tag{2.1}$$

where $\Phi$ is a real valued differentiable function on $[0, \infty)$ and $w$ is a real valued function on $[0, \infty)$.

Assuming that $f$ is bounded, van Es (1992) proposed the following estimator for $T(f)$:

$$T_{m,n} = \frac{1}{2(n-m)} \sum_{j=1}^{n-m} \Phi\left(\frac{m}{n+1}(X_{(j+m)} - X_{(j)})^{-1}\right)\left(w(X_{(j)}) + w(X_{(j+m)})\right), \tag{2.2}$$

where $X_{(1)} \leq \cdots \leq X_{(n)}$ are the order statistics corresponding to the sample $X_1, \ldots, X_n$ and $m$ is an integer such that $1 \leq m \leq n$. The estimator given in (2.2) is constructed according to the fact that $f(X_{(j)})$, the density at the point $X_{(j)}$, is replaced by $m\left((n+1)(X_{(j+m)} - X_{(j)})\right)^{-1}$, where is the local histogram estimate of $f(X_{(j)})$.

Under some conditions, van Es (1992) showed that $T_{m,n}$ is a.s. consistent, i.e.,

$$T_{m,n} \to T(f), \quad a.s. \quad as \quad m, n \to \infty. \tag{2.3}$$

(See Theorem 1 of van Es (1992) for some details). Two special cases of $T(f)$ are

$$T'(f) = \int_0^{+\infty} f^3(x)dx, \tag{2.4}$$

and

$$T''(f) = \int_0^{+\infty} f^2(x)dx, \tag{2.5}$$

obtained by taking $\Phi(x)$ equal to $x^2$ and $x$, respectively, and $w$ is identically equal to $I_{[0,\infty)}(x)$. According to $T_{m,n}$ defined in (2.2), estimators for $T'(f)$ and $T''(f)$ can be proposed as

$$T'_{m,n} = \frac{1}{n-m}\sum_{j=1}^{n-m}\Big(\frac{m}{n+1}(X_{(j+m)} - X_{(j)})^{-1}\Big)^2, \tag{2.6}$$

and

$$T''_{m,n} = \frac{1}{n-m}\sum_{j=1}^{n-m}\Big(\frac{m}{n+1}(X_{(j+m)} - X_{(j)})^{-1}\Big), \tag{2.7}$$

respectively.

Now, the expression given in (1.4) for $VJ(X)$, motivates us to construct the first estimator of varextropy based on $T'_{m,n}$ and $T''_{m,n}$ as follows:

$$VJV_{m,n} = \frac{1}{4}\{T'_{m,n} - (T''_{m,n})^2\}. \tag{2.8}$$

In the following theorem, we establish the a.s. consistency of $VJV_{m,n}$.

**Theorem 2.1.** *Let the density function $f$ be bounded, then $VJV_{m,n} \to VJ(X)$ a.s., provided $m, n \to \infty$, $\frac{m}{\log n} \to \infty$ and $\frac{m}{n} \to 0$.*

*Proof.*

$$\begin{aligned} VJV_{m,n} - VJ(X) &= \frac{1}{4}\{T'_{m,n} - (T''_{m,n})^2\} - \frac{1}{4}\{T'(f) - (T''(f))^2\} \\ &= \frac{1}{4}\{T'_{m,n} - T'(f)\} - \frac{1}{4}\{(T''_{m,n})^2 - (T''(f))^2\}. \end{aligned} \tag{2.9}$$

Theorem 1 of van Es (1992) ensures that

$$T'_{m,n} \to T'(f) \ \ a.s. \ \ as \ \ m, n \to \infty, \tag{2.10}$$

and

$$(T''_{m,n})^2 \to (T''(f))^2 \ \ a.s. \ \ as \ \ m, n \to \infty. \tag{2.11}$$

(2.9), (2.10) and (2.11) complete the proof. □

## 2.2 The Second Estimator

The second estimator is proposed by:

$$\frac{1}{4}\left(\frac{1}{n}\sum_{i=1}^{n} f^2(X_i) - \Big(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\Big)^2\right), \tag{2.12}$$

using the empirical distribution function $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)$.

Since $f$ is unknown, a natural estimator for $VJ(X)$ can be obtained by substituting $f$ by $\hat{f}$ defined in (2.36), and is given by

$$\begin{aligned} VJB_n &= \frac{1}{4}\left(\int_0^{+\infty} \hat{f}^2(x)dF_n(x) - \Big(\int_0^{+\infty} \hat{f}(x)dF_n(x)\Big)^2\right) \\ &= \frac{1}{4}\left(\frac{1}{n}\sum_{i=1}^{n}(\tilde{f}_n(X_i))^2 - \Big(\frac{1}{n}\sum_{i=1}^{n}\tilde{f}_n(X_i)\Big)^2\right), \end{aligned} \tag{2.13}$$

where

$$\tilde{f}_n(X_i) = \frac{1}{(n-1)h_{n-1}} \sum_{j=1, j\neq i}^{n} K\left(\frac{X_i - X_j}{h_{n-1}}\right). \tag{2.14}$$

In the following theorem, we investigate a.s. consistency of $VJ_n(X)$.

**Theorem 2.2.** *Suppose that* $\lim_{x\to\infty} f(x)$ *exists and is finite. If* $f$ *is bounded, then under the conditions of Theorem 2.4, we can write:*

$$\lim_{n\to\infty} |VJB_n - VJ(X)| = 0, \quad a.s.$$

*Proof.* By usng (1.4) and (2.13) we have

$$|VJB_n - VJ(X)| \leq |I'_n| + |I''_n|, \tag{2.15}$$

where

$$I'_n = \frac{1}{4n}\sum_{i=1}^{n}(\tilde{f}_n(X_i))^2 - \frac{1}{4}\int_0^{+\infty} f^2(x)dF(x), \tag{2.16}$$

and

$$I''_n = \left(-\frac{1}{2}\int_0^{+\infty} f^2(x)dx\right)^2 - \left(-\frac{1}{2n}\sum_{i=1}^{n}(\tilde{f}_n(X_i))\right)^2. \tag{2.17}$$

Using the definition of step function $F_n(x)$ and adding and subtracting $\frac{1}{4}\int_0^{+\infty} f^2(x)dF_n(x)$, $I'_n$ can be rewritten as

$$\begin{aligned} I'_n &= \frac{1}{4}\int_0^{+\infty} \hat{f}^2(x)dF_n(x) - \frac{1}{4}\int_0^{+\infty} f^2(x)dF(x) \\ &= \frac{1}{4}\int_0^{+\infty} (\hat{f}^2(x) - f^2(x))dF_n(x) + \frac{1}{4}\int_0^{+\infty} f^2(x)d(F_n(x) - F(x)). \end{aligned} \tag{2.18}$$

On the other hand, since $f$ is bounded and $\lim_{x\to\infty} f(x)$ exists and is finite, by using the integration by parts we have:

$$\frac{1}{4}\int_0^{+\infty} f^2(x)d(F_n(x)-F(x)) = -\frac{1}{2}\int_0^{+\infty}(F_n(x)-F(x))f(x)f'(x)dx. \tag{2.19}$$

From (2.18) and (2.19), we can write:

$$I'_n = \frac{1}{4}\int_0^{+\infty}(\hat{f^2}(x)-f^2(x))dF_n(x) - \frac{1}{2}\int_0^{+\infty}(F_n(x)-F(x))f(x)f'(x)dx. \tag{2.20}$$

Hence

$$|I'_n| \le \frac{1}{4}\sup_x|\hat{f^2}(x)-f^2(x)| + \frac{1}{4}\sup_x|F_n(x)-F(x)|(\lim_{x\to\infty} f^2(x) + (\sup_x f(x))^2). \tag{2.21}$$

The a.s. convergence of $\hat{f^2}(x)$ given in (2.44) and a.s. convergence of $F_n(x)$ given in the Glivenko-Cantelli theorem ensure that

$$\lim_{n\to\infty}|I'_n| = 0. \tag{2.22}$$

In a similar way, using the triangle inequality and adding and subtracting $\int_0^{+\infty} f(x)dF_n(x)$, we can see that

$$\left|\int_0^{+\infty}\hat{f}(x)dF_n(x) - \int_0^{+\infty} f(x)dF(x)\right| \le \left|\int_0^{+\infty}(\hat{f}(x)-f(x))dF_n(x)\right| + \left|\int_0^{+\infty} f(x)d(F_n(x)-F(x))\right|. \tag{2.23}$$

Since $f$ is bounded and $\lim_{x\to\infty} f(x)$ exists and is finite, using the integration by parts we can write:

$$\int_0^{+\infty} f(x)d(F_n(x)-F(x)) = -\int_0^{+\infty}(F_n(x)-F(x))f'(x)dx. \tag{2.24}$$

Therefore, (2.23) and (2.24) imply that

$$\left|\int_0^{+\infty}\hat{f}(x)dF_n(x) - \int_0^{+\infty} f(x)dF(x)\right| \le \sup_x|\hat{f}(x)-f(x)| + \sup_x|F_n(x)-F(x)|(\lim_{x\to\infty} f(x) + \sup_x f(x)). \tag{2.25}$$

It can be concluded from the a.s. convergence of $\hat{f}$ given in (2.42), the Glivenko-Cantelli theorem and the finitness of $\lim_{x\to\infty} f(x)$ and $\sup_x f(x)$ that

$$\lim_{n\to\infty}|I''_n| = 0. \tag{2.26}$$

(2.15), (2.22) and (2.26) complete the proof. □

### 2.3 The Third Estimator

Let $Q(u) = \inf\{x : F(x) \geq u\}, 0 \leq u \leq 1$, be the quantile function corresponding to the distribution function $F(x)$ and $q(u) = Q'(u) = \frac{1}{f(Q(u))}$ be its quantile density function. Soni et al. (2012) proposed a smooth estimator for $q(u)$ given by

$$\begin{aligned} \tilde{q}_n(u) &= \frac{1}{h_n} \int_0^1 \frac{K(\frac{t-u}{h_n})}{f_n(Q_n(t))} dt \\ &= \frac{1}{nh_n} \sum_{i=1}^{n} \frac{K(\frac{S_i-u}{h_n})}{f_n(X_{(i)})}, \end{aligned} \tag{2.27}$$

where $S_i$ is the proportion of observations less than or equal to $X_{(i)}$, the *ith* order statistic, $f_n(u)$ is the same $\hat{f}(u)$ given in (2.36), and $Q_n(u) = \inf\{x : F_n(x) \geq u\}, 0 \leq u \leq 1$, is the empirical estimator of $Q(u)$.

The third estimator for varextropy function $VJ(X)$ is obtained by considering the fact that $VJ(X)$ can be rewritten in terms of the quantile function as follows:

$$\begin{aligned} VJ(X) &= \frac{1}{4}\left[\int_0^1 f^2(Q(u))du - \left(\int_0^1 f(Q(u))du\right)^2\right] \\ &= \frac{1}{4}\left[\int_0^1 \frac{du}{(q(u))^2} - \left(\int_0^1 \frac{du}{q(u)}\right)^2\right]. \end{aligned} \tag{2.28}$$

We consider the following plug-in estimator for quantile-based varextropy, given by

$$VJS_n = \frac{1}{4}\left[\int_0^1 \frac{du}{(\tilde{q}_n(u))^2} - \left(\int_0^1 \frac{du}{\tilde{q}_n(u)}\right)^2\right]. \tag{2.29}$$

In the following theorem, we prove the a.s. consistency of $VJS_n$.

**Theorem 2.3.** *Let $f$ be bounded, then*

$$VJS_n \to VJ(X), \ \ a.s. \ \ n \to \infty.$$

*Proof.* We have

$$\begin{aligned} VJS_n - VJ(X) &= \frac{1}{4}\left[\int_0^1 \frac{du}{(\tilde{q}_n(u))^2} - \int_0^1 \frac{du}{(q(u))^2}\right] \\ &\quad - \frac{1}{4}\left[\left(\int_0^1 \frac{du}{\tilde{q}_n(u)}\right)^2 - \left(\int_0^1 \frac{du}{q(u)}\right)^2\right] \\ &:= J_1 + J_2. \end{aligned} \tag{2.30}$$

Now

$$\int_0^1 \Big(\frac{1}{(\tilde{q}_n(u))^2} - \frac{1}{(q(u))^2}\Big)du = \int_0^1 (q(u) - \tilde{q}_n(u)).\frac{q(u) + \tilde{q}_n(u)}{(\tilde{q}_n(u))^2 q^2(u)}du. \tag{2.31}$$

Using proof of Theorem 3.4 of Subhash et al. (2023), $\sup_t |\tilde{q}_n(t) - q(t)| \to 0 \ \ as \ \ n \to \infty$, hence the above right expression asymptotically reduces to

$$\int_0^1 (q(u) - \tilde{q}_n(u))\frac{2}{q^3(u)}du. \tag{2.32}$$

Since $f$ is bounded, there exist $M > 0$, such that $|f| \le M$ and hence

$$\begin{aligned}\int_0^1 (q(u) - \tilde{q}_n(u))\frac{2}{q^3(u)}du &\le 2\sup_u |q(u) - \tilde{q}_n(u)| \int_0^1 f^3(Q(u))du \\ &\le 2\mathrm{M}^3 \sup\nolimits_u |q(u) - \tilde{q}_n(u)| \to 0 \ \ as \ \ n \to \infty.\end{aligned} \tag{2.33}$$

As a result

$$J_1 \to 0 \ \ as \ \ n \to \infty. \tag{2.34}$$

In a similar way, we can show that

$$J_2 \to 0 \ \ as \ \ n \to \infty. \tag{2.35}$$

(2.30), (2.34) and (2.35), ensure that

$$VJS_n \to VJ(X) \ \ \ as \ \ n \to \infty.$$

□

## 2.4 The Fourth Estimator

Let $\hat{f}(x)$ be the kernel density function of $f$ defined by

$$\hat{f}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\Big(\frac{x - X_i}{h_n}\Big), \tag{2.36}$$

where $K$ is the kernel density function and $h_n$ is the bandwidth parameter. The fourth estimator for $VJ(X)$ had been proposed by Alizadeh Noughabi and Shafaei Noughabi (2024) based on $\hat{f}(x)$ as:

$$VJD = \frac{1}{4}\Big[\int \hat{f}^3(x)dx - \Big(\int \hat{f}^2(x)dx\Big)^2\Big]. \tag{2.37}$$

Here, we aim to prove the almost sure convergence of this estimator.

The following assumptions on $K$ are used in the Theorem 2.4 of this section.
$(A_1)$ : $K$ is uniformly continuous and of bounded variation $V(K)$.
$(A_2)$ : $K(x) \to 0 \ \ as \ \ |x| \to \infty$.
$(A_3)$ : $\int |x \log |x||^{\frac{1}{2}} |dK(x)| < \infty$.

**Theorem 2.4.** *Suppose $K$ satisfies Assumptions $A_1 - A_3$ and $f$ is uniformly continuous. If $h_n \to 0$ and $(nh_n)^{-1}\log n \to 0$ as $n \to \infty$, then*

$$VJD \to VJ(X) \ \ a.s. \ \ as \ \ n \to \infty.$$

*Proof.* From (1.4) and (2.37) we have

$$VJD - VJ(X) = I_n + II_n, \tag{2.38}$$

where

$$I_n = \frac{1}{4}\left(\int_0^{+\infty} \hat{f}^3(x)dx - \int_0^{+\infty} f^3(x)dx\right), \tag{2.39}$$

and

$$II_n = \left(-\frac{1}{2}\int_0^{+\infty} f^2(x)dx\right)^2 - \left(-\frac{1}{2}\int_0^{+\infty} \hat{f}^2(x)dx\right)^2. \tag{2.40}$$

Observe that

$$\begin{aligned} |I_n| &= \frac{1}{4}\left|\int_0^{+\infty} \hat{f}^3(x)dx - \int_0^{+\infty} f^3(x)dx\right| \\ &\leq \frac{1}{4}\left|\int_0^{+\infty} \hat{f}^2(x)\Big(\hat{f}(x) - f(x)\Big)dx\right| \\ &+ \frac{1}{4}\left|\int_0^{+\infty} \Big(\hat{f}^2(x) - f^2(x)\Big)f(x)dx\right|. \end{aligned} \tag{2.41}$$

Under Assumptions $A_1 - A_3$, from Theorem A of Silverman (1978), one can see that

$$\lim_{n\to\infty} \sup_x |\hat{f}(x) - f(x)| = 0, \ \ a.s. \tag{2.42}$$

On the other hand, a simple algebra calculation shows that

$$\hat{f}^2(x) - f^2(x) = \sum_{i=1}^{2} (f(x))^{i-1}(\hat{f}(x))^{2-i}(\hat{f}(x) - f(x)). \tag{2.43}$$

(2.42) and (2.43) ensure that

$$\lim_{n\to\infty} \sup_x |\hat{f}^2(x) - f^2(x)| = 0, \ \ a.s. \tag{2.44}$$

From (2.41), (2.42) and (2.44), one can observe that

$$\lim_{n\to\infty} I_n = 0, \qquad a.s. \tag{2.45}$$

Next, to deal with $II_n$, observe that

$$\begin{aligned}\left|\int_0^{+\infty}\hat{f}^2(x)dx-\int_0^{+\infty}f^2(x)dx\right| &\leq \left|\int_0^{+\infty}\hat{f}^2(x)dx-\int_0^{+\infty}\hat{f}(x)dF(x)\right| \\ &+\left|\int_0^{+\infty}\hat{f}(x)dF(x)-\int_0^{+\infty}f^2(x)dx\right| \\ &=\left|\int_0^{+\infty}\hat{f}(x)(\hat{f}(x)-f(x))dx\right| \\ &+\left|\int_0^{+\infty}(\hat{f}(x)-f(x))dF(x)\right|. \end{aligned} \tag{2.46}$$

Since from (2.42), $\hat{f}(x)$ is a.s. consistent, we can write

$$\lim_{n\to\infty}\int_0^{+\infty}\hat{f}^2(x)dx=\int_0^{+\infty}f^2(x)dx, \quad a.s. \tag{2.47}$$

As a result

$$\lim_{n\to\infty} II_n=0, \quad a.s. \tag{2.48}$$

(2.38), (2.45) and (2.48) complete the proof. □

## 2.5 The Fifth Estimator

By analogy to Qiu and Jia (2018a) and using the (2.28), Alizadeh Noughabi and Shafaei Noughabi (2024) proposed the following estimator for varextropy function:

$$VJQ_{m,n}=\frac{1}{4}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{c_i\frac{m}{n}}{X_{(i+m)}-X_{(i-m)}}\right)^2-\left(\frac{1}{n}\sum_{i=1}^{n}\frac{c_i\frac{m}{n}}{X_{(i+m)}-X_{(i-m)}}\right)^2\right], \tag{2.49}$$

where the window size $m$ is a positive integer smaller than $n/2$, $X_{(i)}=X_{(1)}$ if $i<1$, $X_{(i)}=X_{(n)}$ if $i>n$ and

$$c_i=\begin{cases}1+\frac{i-1}{m}, & 1\leq i\leq m,\\ 2, & m+1\leq i\leq n-m,\\ 1+\frac{n-i}{m}, & n-m+1\leq i\leq n.\end{cases}$$

The consistency of the proposed estimator is shown in the following theorem.

**Theorem 2.5.** *Let $X_1,\ldots,X_n$ be a random sample from distribution function $F$ with bounded probability density function $f$ and finite variance. Then we can write:*

$$VJQ_{m,n}\xrightarrow{\text{p}} VJ(X), \quad as\ n\to\infty,\ m\to\infty,\ \frac{m}{n}\to 0. \tag{2.50}$$

*Proof.* A simple algebra shows that:

$$VJQ_{m,n} - VJ(X) := I_1 + I_2, \tag{2.51}$$

where

$$I_1 = \frac{1}{4}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{c_i\frac{m}{n}}{X_{(i+m)} - X_{(i-m)}}\right)^2 - \int_0^{+\infty} f^3(x)dx\right], \tag{2.52}$$

and

$$I_2 = \frac{1}{4}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\frac{c_i\frac{m}{n}}{X_{(i+m)} - X_{(i-m)}}\right)^2 - \left(\int_0^{+\infty} f^2(x)dx\right)^2\right]. \tag{2.53}$$

Theorem 2.1 of Qiu and Jia (2018a), clearly shows that $I_2 \xrightarrow{p} 0$. For $I_1$, first we write $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{c_i\frac{m}{n}}{X_{(i+m)}-X_{(i-m)}}\right)^2$ as follows

$$\frac{1}{n}\sum_{i=1}^{n}(c_i\frac{m}{n})^2\Big(\frac{1}{X_{(i+m)} - X_{(i-m)}}\Big)^2 = \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\frac{2m}{n}}{X_{(i+m)} - X_{(i-m)}}\Big)^2$$
$$-\frac{2}{n}\sum_{i=1}^{n}\xi_{im}\frac{2m}{n}\frac{1}{X_{(i+m)} - X_{(i-m)}}(1 - \frac{c_i}{2}) := J_1 - J_2. \tag{2.54}$$

where $\xi_{im}$ is a random point between the points $\frac{c_i\frac{m}{n}}{X_{(i+m)}-X_{(i-m)}}$ and $\frac{\frac{2m}{n}}{X_{(i+m)}-X_{(i-m)}}$.

Next to deal with $J_1$, observe that

$$J_1 = \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\frac{2m}{n}}{X_{(i+m)} - X_{(i-m)}}\Big)^2 = \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{F(X_{(i+m)}) - F(X_{(i-m)})}{X_{(i+m)} - X_{(i-m)}}\Big)^2\left[\Big(\frac{\frac{2m}{n}}{F(X_{(i+m)}) - F(X_{(i-m)})}\Big)^2 - 1\right]$$
$$+ \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{F(X_{(i+m)}) - F(X_{(i-m)})}{X_{(i+m)} - X_{(i-m)})}\Big)^2$$
$$:= J_{11} + J_{12}. \tag{2.55}$$

Now, since $f(x)$ is positive and continuous, there exists a value $X_i \in (X_{(i-m)}, X_{(i+m)})$ such that

$$f(X_i) = \frac{F(X_{(i+m)}) - F(X_{(i-m)})}{X_{(i+m)} - X_{(i-m)}}, \tag{2.56}$$

hence

$$J_{11} = \frac{1}{n}\sum_{i=1}^{n}(f(X_i))^2\left[\Big(\frac{\frac{2m}{n}}{F(X_{(i+m)}) - F(X_{(i-m)})}\Big)^2 - 1\right]. \tag{2.57}$$

On the other hand, since $F_n(X_{(i+m)}) - F_n(X_{(i-m)}) = \frac{2m}{n}$, Chung's law of the iterated logarithm (LIL) implies that:

$$\begin{aligned}\frac{F(X_{(i+m)}) - F(X_{(i-m)})}{F_n(X_{(i+m)}) - F_n(X_{(i-m)})} &\leq \frac{|F_n(X_{(i+m)}) - F(X_{(i+m)})|}{F_n(X_{(i+m)}) - F_n(X_{(i-m)})} + \frac{|F_n(X_{(i-m)}) - F(X_{(i-m)})|}{F_n(X_{(i+m)}) - F_n(X_{(i-m)})} \\ &+ \frac{F_n(X_{(i+m)}) - F_n(X_{(i-m)})}{F_n(X_{(i+m)}) - F_n(X_{(i-m)})} \leq \frac{n}{2m}\sup_x |F_n(x) - F(x)| \\ &+ \frac{n}{2m}\sup_x |F_n(x) - F(x)| + 1 \\ &\leq \frac{n}{m} O\left(\sqrt{\frac{\log\log n}{n}}\right) + 1. \end{aligned} \tag{2.58}$$

Now, since $\lim_{n\to\infty} \frac{\sqrt{n \log\log n}}{m} = 0$, we can see that

$$\lim_{n\to\infty} \frac{\frac{2m}{n}}{F(X_{(i+m)}) - F(X_{(i-m)})} = \lim_{n\to\infty} \frac{F_n(X_{(i+m)}) - F_n(X_{(i-m)})}{F(X_{(i+m)}) - F(X_{(i-m)})} \geq 1. \tag{2.59}$$

Therefore, (2.57), (2.59) and condition of $\sup_x f(x) < M < \infty$ result that for enough large $n$ we can write

$$J_{11} \leq \frac{M^2}{n} \sum_{i=1}^{n} \left[ \frac{(\frac{2m}{n})^2}{(F(X_{(i+m)}) - F(X_{(i-m)}))^2} - 1 \right]. \tag{2.60}$$

Now, we prove that when $m \to \infty$,

$$\lim_{n\to\infty} E\left( \frac{M^2}{n} \sum_{i=1}^{n} \left[ \frac{(\frac{2m}{n})^2}{(F(X_{(i+m)}) - F(X_{(i-m)}))^2} - 1 \right] \right) = 0. \tag{2.61}$$

Note that

$$\begin{aligned}\lim_{n\to\infty} E\left( \frac{M^2}{n} \sum_{i=1}^{n} \left[ \frac{(\frac{2m}{n})^2}{(F(X_{(i+m)}) - F(X_{(i-m)}))^2} - 1 \right] \right) &= \lim_{n\to\infty} E\left( \frac{M^2}{n} \sum_{i=1}^{m} \left[ \frac{(\frac{2m}{n})^2}{(F(X_{(i+m)}) - F(X_{(1)}))^2} - 1 \right] \right) \\ + \lim_{n\to\infty} E\left( \frac{M^2}{n} \sum_{i=m+1}^{n-m} \left[ \frac{(\frac{2m}{n})^2}{(F(X_{(i+m)}) - F(X_{(i-m)}))^2} - 1 \right] \right) \\ + \lim_{n\to\infty} E\left( \frac{M^2}{n} \sum_{i=n-m+1}^{n} \left[ \frac{(\frac{2m}{n})^2}{(F(X_{(n)}) - F(X_{(i-m)}))^2} - 1 \right] \right). \end{aligned} \tag{2.62}$$

Next notice that $F(X_{(1)}), \ldots, F(X_{(n)})$ is an orderd sample from uniform distribution and the random variable $F(X_{(i+j)}) - F(X_{(i)})$ has the beta distribution with parameters $j$ and $n - j + 1$, so we have

$$\begin{aligned}&\lim_{n\to\infty} E\left( \frac{M^2}{n} \sum_{i=1}^{n} \left[ \frac{(\frac{2m}{n})^2}{(F(X_{(i+m)}) - F(X_{(i-m)}))^2} - 1 \right] \right) \\ &= \lim_{n\to\infty} M^2 \Big( \tfrac{2m^2(n-1)(2m^2+mn+2m-2n}{n^2(m-1)(m-2)(2m-1)} - 1 \Big) = 0. \end{aligned} \tag{2.63}$$

Since convergence in expectation to 0 implies convergence in probability to 0 for non-negative random variables, thus

$$J_{11} \xrightarrow{p} 0. \tag{2.64}$$

Note that by (2.56) and the law of large numbers (LLN), we can write:

$$J_{12} = \frac{1}{n}\sum_{i=1}^{n} f^2(X_i) = \int f^2(x)dF_n(x) \xrightarrow{p} \int f^3(x)dx. \tag{2.65}$$

So, (2.55), (2.64) and (2.65) conclude that

$$J_1 \xrightarrow{p} \int f^3(x)dx. \tag{2.66}$$

To deal with $J_2$, we notice that since $\xi_{im} \leq \frac{\frac{2m}{n}}{X_{(i+m)}-X_{(i-m)}}$, we can write

$$\begin{aligned}
|J_2| &\leq \frac{2}{n}\sum_{i=1}^{n} \frac{(\frac{2m}{n})^2}{(X_{(i+m)} - X_{(i-m)})^2}(1 - \frac{c_i}{2}) \\
&= \frac{2}{n}\sum_{i=1}^{n} \left(\frac{2m}{n}\right)^2 \left(\frac{F(X_{(i+m)}) - F(X_{(i-m)}}{X_{(i+m)} - X_{(i-m)}}\right)^2 \frac{1 - \frac{c_i}{2}}{(F(X_{(i+m)}) - F(X_{(i-m)})^2}.
\end{aligned}$$

From (2.56) and boundedness of $f(X_i)$, we have

$$\begin{aligned}
|J_2| &\leq \frac{2M^2}{n}\sum_{i=1}^{n} \frac{4m^2}{n^2} \frac{1 - \frac{c_i}{2}}{(F(X_{(i+m)}) - F(X_{(i-m)})^2} \\
&= \frac{4M^2 m}{n^3}\left[\sum_{i=1}^{m} \frac{m - i + 1}{(F(X_{(i+m)}) - F(X_{(1)})^2} + \sum_{i=n-m+1}^{n} \frac{m - n + i}{(F(X_{(n)}) - F(X_{(i-m)})^2}\right],
\end{aligned} \tag{2.67}$$

thus

$$\begin{aligned}
E(J_2) &\leq \frac{8M^2 m}{n^2}\sum_{i=1}^{m} \frac{(n-1)(m-i+1)}{(i+m-2)(i+m-3)} \\
&= \frac{8M^2 m(n-1)}{n^2}\left\{\frac{m(2m-1)}{2(m-1)(m-2)} + \psi(m-2) - \psi(2m-2)\right\},
\end{aligned} \tag{2.68}$$

where $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ and converges to zero with $n \to \infty$, $m \to \infty$ and $m/n \to 0$. Thus, $J_2$ form a series of non-positive variables with expectations approaching zero, and consequently

$$J_2 \xrightarrow{p} 0. \tag{2.69}$$

Thus, (2.51), (2.66) and (2.69) complete the proof. □

*Remark* 1. In Theorem 2.5, many m's can be found that satisfy the conditions of theorem, for example $m = O((\log\log n)^{\alpha} n^{\beta})$ for every $\alpha \in \mathbb{R}$ and $\frac{1}{2} < \beta < 1$.

## 3 Simulation Study

In this section, the results of the Monte Carlo studies on mean squared errors (MSEs) and biases of our introduced estimators are presented. We consider the gamma, uniform and exponential distributions, which are considered in many references. For each sample size, 10,000 samples were generated and MSEs and biases of estimators were computed. We used the $m = [\sqrt{n} + 0.5]$ formula to estimate the varextropy suggested by Grzegorzewski and Wieczorkowski (1999). They used this heuristic formula for entropy estimation. In addition, we choose the standard normal density as the kernel and its corresponding optimal value of $h_n = 1.06sn^{-\frac{1}{5}}$, where $s$ is the sample standard deviation. Tables 1-3 give simulated biases and MSEs of $VJV_{m,n}$, $VJB_n$ and $VJS_n$ for gamma, uniform and exponential distributed samples, respectively. It is apparent from the tables that both the absolute bias and MSE decrease as the sample size increases. The bold type in these tables indicates the varextropy estimator achieves the minimal MSE (absolute bias). In the case of gamma distribution, Table 1, it is observed that, the

Table 1: MSE and Bias of estimators to estimate the varextropy $VJ(X)$ of the gamma distribution (G(2, 1)).

| | MSE(Bias) | | |
|---|---|---|---|
| $n$ | $VJV_{m,n}$ | $VJB_n$ | $VJS_n$ |
| 10 | 0.0120008(0.0198195) | **0.0000048(-0.0011958)** | 0.0000072(-0.0023191) |
| 20 | 0.0016044(0.0116757) | **0.0000026(-0.0013097)** | 0.0000064(-0.0024858) |
| 30 | 0.0002339(0.0074458) | **0.0000022(-0.0012534)** | 0.0000062(-0.0024651) |
| 40 | 0.0000904(0.0052193) | **0.0000019(-0.0012115)** | 0.0000060(-0.0024254) |
| 50 | 0.0000483(0.0039833) | **0.0000017(-0.0011682)** | 0.00000577(-0.0023838) |
| 100 | 0.0000125(0.0024089) | **0.0000013(-0.0010154)** | 0.0000048(-0.0021808) |

Table 2: MSE and Bias of estimators to estimate the varextropy $VJ(X)$ of the uniform distribution(U(0, 1))

| | MSE(Bias) | | |
|---|---|---|---|
| $n$ | $VJV_{m,n}$ | $VJB_n$ | $VJS_n$ |
| 10 | 5.634314(0.3022407) | **0.0007618(0.0191408)** | 0.0007793(0.0197080) |
| 20 | 0.2218065(0.1563245) | 0.0003397(0.0141829) | **0.0001894(0.0109698)** |
| 30 | 0.0295767(0.0993167) | 0.0002256(0.0119855) | **0.0001103(0.0086761)** |
| 40 | 0.0113412(0.0710308) | 0.0001722(0.0107598) | **0.0000784(0.0074447)** |
| 50 | 0.0058115(0.0547022) | 0.0001409(0.0099317) | **0.0000610(0.0066770)** |
| 100 | 0.0015858(0.0336733) | 0.0000811(0.0079505) | **0.0000300(0.0048685)** |

estimator $VJB_n$ performs best in terms of bias and MSE. Under the uniform distribution, Table 2, it can be seen that, the MSEs and biases of $VJS_n$ are always smaller than other proposed estimators except in $n = 10$. According to Table 3, under exponential distribution, estimator $VJB_n$ performs better than other proposed estimators in terms of MSE, and in terms of absolute bias.

Finally, based on the results of our simulation study, we can conclude that the estimators $VJB_n$ and $VJS_n$ have better performance.

Table 3: MSE and Bias of estimators to estimate the varextropy $VJ(X)$ of the exponential distribution with mean equal to one.

| | MSE(Bias) | | |
|---|---|---|---|
| $n$ | $VJV_{m,n}$ | $VJB_n$ | $VJS_n$ |
| 10 | 0.4596275(0.0899478) | **0.0003007(-0.0159281)** | 0.0003360(-0.0167951) |
| 20 | 0.0252052(0.0485333) | **0.0002770(-0.0163934)** | 0.0003233(-0.0175661) |
| 30 | 0.0089156(0.0322723) | **0.0002678(-0.0162028)** | 0.0003118(-0.0173596) |
| 40 | 0.0034209(0.0234392) | **0.0002598(-0.0159925)** | 0.0002999(-0.0170549) |
| 50 | 0.0015958(0.0177894) | **0.0002522(-0.0157660)** | 0.0002883(-0.0167369) |
| 100 | 0.0004385(0.0106603) | **0.0002277(-0.0150125)** | 0.0002375(-0.0151743) |

# 4 Some Tests of Uniformity

In this section, we introduce a few goodness-of–fit tests for uniformity. These tests are based on our proposed varextropy estimators and the two estimators Alizadeh Noughabi and Shafaei Noughabi (2024) and their percentage points and also power values are obtained by Monte Carlo simulation.

## 4.1 Test Statistic

Consider the class of continuous distribution functions $F$ with density function $f(x)$ defined on interval $[0, 1]$. Due to the fact that the variance is always nonnegative, we have $VJ(X) = Var[-\frac{1}{2}f(X)] \geq 0$. If $f$ is a standard uniform density function, it is easy to see that $VJ(X) = 0$. On the other hand, if $VJ(X) = Var[-\frac{1}{2}f(X)] = 0$, then we have $f(x) = c$ on $[0, 1]$, where $c$ is a constant and from the property of density function, it is obvious that $f(x) = 1$. Therefore, we can claim that $VJ(X)$ is always nonnegative and its minimum value is uniquely obtained by the standard uniform distribution. The above discussion gives us the idea to use the proposed varextropy estimators as the test statistics of goodness-of-fit tests of uniformity.

Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution function $F(x)$ on $[0, 1]$. The null hypothesis is $H_0 : F(x) = x$ and the alternative hypothesis (denoted $H_1$) is the opposite of $H_0$. Given any significance level $\alpha$, our hypothesis-testing procedure can be defined by the critical region:

$$G_n = \hat{V}J(X) \geq C_{1-\alpha}, \tag{4.1}$$

where $\hat{V}J(X)$ is one of the proposed estimators and $C_{1-\alpha}$ is the critical value for the test with level $\alpha$. Notice when $\hat{V}J(X) \xrightarrow{p} VJ(X)$, under $H_0$, $G_n \xrightarrow{p} 0$ and $H_1$, $G_n$ converges to a number larger than zero in probability.

Due to the complexity of the calculations, it is not easy to determine the distribution of the test statistics under $H_0$, so the critical values are calculated using the Monte Carlo method. By utilizing five different varextropy estimators, we introduce the following test statistics for testing the uniformity:

$$\begin{aligned} GV_n &= VJV_{m,n}, \\ GB_n &= VJB_n, \\ GS_n &= VJS_n, \\ GD_n &= VJD, \\ GQ_n &= VJQ_{m,n}. \end{aligned}$$

It is necessary to mention here that, also Alizadeh Noughabi and Shafaei Noughabi (2023) used varentropy for testing uniformity. They showed that $VE(X) = 0$ if and only if $f(x)$ is the standard uniform density function and used their proposed varentropy estimators as the test statistics for goodness-of-fit tests of uniformity.

### 4.2 Power Comparisons

In this subsection, we compare the powers of our proposed test statistics with Kolmogorov-Smirnov statistic (Kolmogorov (1933), Smirnov (1939)), statistics of Zhang (2002) and the test statistics proposed by Alizadeh Noughabi and Shafaei Noughabi (2023). The test statistic of Kolmogorov-Smirnov is given as follows:

$$KS = max\left(max_{1\leq i\leq n}\left\{\frac{i}{n} - X_{(i)}\right\}, max_{1\leq i\leq n}\left\{X_{(i)} - \frac{i-1}{n}\right\}\right), \tag{4.2}$$

where $X_{(1)}, \ldots, X_{(n)}$ are order statistics. Moreover, based on the tests introduced by Zhang (2002), we derive the following statistics to test the goodness-of-fit for the uniform distribution.

$$Z_K = \max_{1\leq i\leq n}\left(\left(i-\frac{1}{2}\right)\log\left\{\frac{i-\frac{1}{2}}{nX_{(i)}}\right\} + \left(n-i+\frac{1}{2}\right)\log\left\{\frac{n-i+\frac{1}{2}}{n\{1-X_{(i)}\}}\right\}\right), \tag{4.3}$$

$$Z_A = -\sum_{i=1}^{n}\left[\frac{\log\{X_{(i)}\}}{n-i+\frac{1}{2}} + \frac{\log\{1-X_{(i)}\}}{i-\frac{1}{2}}\right], \tag{4.4}$$

and

$$Z_C = \sum_{i=1}^{n}\left[\log\left\{\frac{(X_{(i)})^{-1}-1}{(n-\frac{1}{2})/(i-\frac{3}{4})-1}\right\}\right]^2. \tag{4.5}$$

Furthermore, the Alizadeh Noughabi-Shafaei Noughabi Statistics are given as:

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

$$TE_n = \frac{1}{n}\sum_{i=1}^{n}\log^2\left(\frac{c_i m/n}{X_{(i+m)} - X_{(i-m)}}\right) - \left[\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{c_i m/n}{X_{(i+m)} - X_{(i-m)}}\right)\right]^2,$$

where $X_{(i)} = X_{(1)}$ if $i < 1$ and $X_{(i)} = X_{(n)}$ if $i > n$,

$$TD_n = \int_{-\infty}^{+\infty} \hat{f}(x)[\log \hat{f}(x)]^2 dx - \left[\int_{-\infty}^{+\infty} \hat{f}(x) \log \hat{f}(x) dx\right]^2,$$

where $\hat{f}$ is the kernel density function estimation of $f$ and is defined by

$$\hat{f}(x) = \frac{1}{nh} \sum_{j=1}^{n} k\left(\frac{x - X_j}{h}\right),$$

$$TB_n = \frac{1}{n} \sum_{i=1}^{n} (\log \hat{f}(X_i))^2 - \left[\frac{1}{n} \sum_{i=1}^{n} \log \hat{f}(X_i)\right]^2,$$

where $\hat{f}(X_i) = \frac{1}{n} \sum_{j=1}^{n} k(\frac{X_i - X_j}{h})$,

$$TC_n = \frac{1}{n} \sum_{i=1}^{n} \log^2\left(\frac{\sum_{j=i-m}^{i+m}(X_{(j)} - \bar{X}_{(i)})(j-i)}{n \sum_{j=i-m}^{i+m}(X_{(j)} - \bar{X}_{(i)})^2}\right) - \left[\frac{1}{n} \sum_{i=1}^{n} \log\left(\frac{\sum_{j=i-m}^{i+m}(X_{(j)} - \bar{X}_{(i)})(j-i)}{n \sum_{j=i-m}^{i+m}(X_{(j)} - \bar{X}_{(i)})^2}\right)\right]^2,$$

where $\bar{X}_{(i)} = \frac{1}{2m+1} \sum_{j=i-m}^{i+m} X_{(j)}$, and

$$TA_n = \frac{1}{n} \sum_{i=1}^{n} \log^2\{\hat{f}(X_{(i+m)}) + \hat{f}(X_{(i-m)})\} - \left[\frac{1}{n} \sum_{i=1}^{n} \log\{\hat{f}(X_{(i+m)}) + \hat{f}(X_{(i-m)})\}\right]^2. \tag{4.6}$$

Table 4: Percentage points of the proposed test statistics at the level $\alpha = 0.05$.

| n | $GV_n$ | $GB_n$ | $GS_n$ | $GD_n$ | $GQ_n$ |
|---|---|---|---|---|---|
| 10 | 0.9102 | 0.0574 | 0.0558 | 0.0665 | 0.4024 |
| 20 | 0.4937 | 0.0374 | 0.0277 | 0.0485 | 0.1813 |
| 30 | 0.2845 | 0.0297 | 0.0205 | 0.0413 | 0.1365 |
| 40 | 0.1933 | 0.0256 | 0.0169 | 0.0371 | 0.0767 |
| 50 | 0.1389 | 0.0226 | 0.0148 | 0.0343 | 0.0583 |
| 75 | 0.0872 | 0.0183 | 0.0116 | 0.0296 | 0.0385 |
| 100 | 0.0715 | 0.0158 | 0.0098 | 0.0270 | 0.0320 |

In the above statistics, $m$ is the window size as defined in Subsection 2.5.

Table 5: Power comparisons of the proposed tests at a significance level of 0.05.

| n | *Alternative* | $GV_n$ | $GB_n$ | $GS_n$ | $GD_n$ | $GQ_n$ |
|---|---|---|---|---|---|---|
| 10 | $A_{1.5}$ | 0.0739 | 0.1072 | 0.1663 | **0.1767** | 0.0764 |
| | $A_2$ | 0.1150 | 0.2033 | 0.3423 | **0.3959** | 0.1327 |
| | $B_{1.5}$ | 0.0719 | **0.1629** | 0.1109 | 0.1589 | 0.0456 |
| | $B_2$ | 0.1126 | 0.3292 | 0.2167 | **0.3494** | 0.0624 |
| | $B_3$ | 0.2156 | 0.6421 | 0.5038 | **0.7326** | 0.1294 |
| | $C_{1.5}$ | 0.0808 | 0.0296 | 0.0573 | 0.0415 | **0.1182** |
| | $C_2$ | 0.1295 | 0.0300 | 0.0660 | 0.0406 | **0.2160** |
| 20 | $A_{1.5}$ | 0.0884 | 0.1575 | 0.2497 | **0.2889** | 0.1088 |
| | $A_2$ | 0.1684 | 0.3648 | 0.5377 | **0.6680** | 0.2489 |
| | $B1.5$ | 0.0879 | **0.2543** | 0.0977 | 0.2424 | 0.0444 |
| | $B_2$ | 0.1640 | 0.5493 | 0.2611 | **0.5707** | 0.1061 |
| | $B_3$ | 0.3689 | 0.8948 | 0.7059 | **0.9459** | 0.3549 |
| | $C_{1.5}$ | 0.09628 | 0.0251 | 0.1010 | 0.0373 | **0.1812** |
| | $C_2$ | 0.1900 | 0.0281 | 0.1339 | 0.0440 | **0.3924** |
| 30 | $A_{1.5}$ | 0.1114 | 0.2092 | 0.3206 | **0.3933** | 0.1580 |
| | $A_2$ | 0.2425 | 0.5078 | 0.6832 | **0.8319** | 0.3994 |
| | $B1.5$ | 0.1100 | **0.3416** | 0.1080 | 0.3160 | 0.0690 |
| | $B_2$ | 0.2333 | 0.7132 | 0.3674 | **0.7251** | 0.2224 |
| | $B_3$ | 0.5500 | 0.9736 | 0.8706 | **0.9907** | 0.6455 |
| | $C_{1.5}$ | 0.1236 | 0.0228 | 0.1409 | 0.0385 | **0.2450** |
| | $C_2$ | 0.2827 | 0.0263 | 0.2118 | 0.0498 | **0.5757** |

For power comparisons, we compute the powers of our proposed tests and the powers of tests mentioned above under the following alternative distributions:

$$A_k : F(x) = 1 - (1-x)^k, \quad 0 \le x \le 1 \ \text{(for } k = 1.5, 2);$$

$$B_k : F(x) = \begin{cases} 2^{k-1}x^k, & 0 \le x \le 0.5 \\ 1 - 2^{k-1}(1-x)^k, & 0.5 \le x \le 1 \end{cases} \quad \text{(for } k = 1.5, 2, 3);$$

$$C_k : F(x) = \begin{cases} 0.5 - 2^{k-1}(0.5-x)^k, & 0 \le x \le 0.5 \\ 0.5 + 2^{k-1}(x-0.5)^k, & 0.5 \le x \le 1 \end{cases} \quad \text{(for } k = 1.5, 2).$$

Stephens (1974) uses these alternative distributions in his study on power comparisons of some uniformity tests.

Table 6: Power comparisons of the Zhang test at a significance level of 0.05.

| n | *Alternative* | $Z_K$ | $Z_A$ | $Z_C$ |
|---|---|---|---|---|
| 10 | $A_{1.5}$ | 0.1381 | **0.1712** | 0.1390 |
| | $A_2$ | 0.3471 | **0.4387** | 0.3678 |
| | $B_{1.5}$ | 0.0308 | **0.06336** | 0.02736 |
| | $B_2$ | 0.04738 | **0.1429** | 0.0654 |
| | $B_3$ | 0.1065 | **0.4442** | 0.2787 |
| | $C_{1.5}$ | **0.1220** | 0.06846 | 0.09516 |
| | $C_2$ | **0.2302** | 0.1179 | 0.1617 |
| 20 | $A_{1.5}$ | 0.2579 | **0.3284** | 0.2738 |
| | $A_2$ | 0.6726 | **0.7786** | 0.7194 |
| | $B1.5$ | 0.08053 | **0.1891** | 0.1091 |
| | $B_2$ | 0.2168 | **0.5399** | 0.4073 |
| | $B_3$ | 0.6289 | **0.9606** | 0.9268 |
| | $C_{1.5}$ | **0.1538** | 0.07005 | 0.1046 |
| | $C_2$ | **0.3478** | 0.1683 | 0.2330 |
| 30 | $A_{1.5}$ | 0.3857 | **0.4744** | 0.4210 |
| | $A_2$ | 0.8668 | **0.9292** | 0.9066 |
| | $B1.5$ | 0.1403 | **0.3221** | 0.2191 |
| | $B_2$ | 0.4393 | **0.8020** | 0.7197 |
| | $B_3$ | 0.9258 | **0.9986** | 0.9978 |
| | $C_{1.5}$ | **0.1926** | 0.080 | 0.1242 |
| | $C_2$ | **0.4757** | 0.2529 | 0.33086 |

Based on 100,000 repetitions, the critical values are estimated as shown in Table 4. The power performance of the proposed and competing test statistics under various alternatives and for different sample sizes ($n = 10, 20, 30$), at a significance level of $\alpha = 0.05$, is summarized in Tables 5, 6 and 7. In these tables, the highest power observed in each setting is highlighted in bold. The simulation results indicate that the power of the tests depends substantially on the type of alternative distribution. For type $A$ alternatives, the proposed $GD_n$ and KS tests show competitive performance. Specifically, under alternative $A_{1.5}$ and for ($n = 10$), $GD_n$ achieves the highest power, whereas under $A_2$, the $Z_A$ test outperforms the others. For larger samples ($n = 20$) and ($n = 30$), when the alternative distribution is of type $A$, the $Z_A$ test consistently yields superior power. In contrast, the Alizadeh Noughabi-Shafaei Noughabi test statistics have relatively low power in these cases. For the $B_{1,5}$ alternative, our proposed $GB_n$ test shows the best performance across all sample sizes. Competing tests such as $TV$, $TE$, $TC$, $TA$, $KS$ and Zhang statistics perform poorly in this situation. Furthermore, under the $B_2$ alternative, the $GD_n$ statistics provides the highest power for ($n = 10$) and ($n = 20$), while for ($n = 30$), the $Z_A$ statistic has the most powerful. For the $B_3$ alternative, $GD_n$ is optimal at ($n = 10$), whereas $Z_A$ dominates at larger sample sizes. For alternatives of type $C$, the test statistic $TC$ provides the highest power. Moreover, the $GQ_n$ statistic shows better performance than the KS test under these settings.

Table 7: Power comparisons of the Alizadeh Noughabi-Shafaei Noughabi tests and KS test at a significance level of 0.05.

| n | *Alternative* | $TV$ | $TE$ | $TD$ | $TB$ | $TC$ | $TA$ | $KS$ |
|---|---|---|---|---|---|---|---|---|
| 10 | $A_{1.5}$ | 0.0526 | 0.0625 | 0.0811 | 0.0814 | 0.0642 | 0.0832 | **0.1542** |
| | $A_2$ | 0.0828 | 0.1014 | 0.1311 | 0.1244 | 0.0883 | 0.1427 | **0.3871** |
| | $B_{1.5}$ | 0.0295 | 0.0213 | **0.0826** | 0.0810 | 0.0298 | 0.0414 | 0.0367 |
| | $B_2$ | 0.0195 | 0.0213 | 0.1110 | **0.1226** | 0.0169 | 0.0422 | 0.0447 |
| | $B_3$ | 0.0185 | 0.0319 | **0.1873** | 0.1810 | 0.0264 | 0.0609 | 0.0873 |
| | $C_{1.5}$ | 0.1627 | 0.1409 | 0.0530 | 0.0519 | **0.1779** | 0.0736 | 0.1148 |
| | $C_2$ | 0.3329 | 0.2909 | 0.0733 | 0.0620 | **0.3530** | 0.0924 | 0.2001 |
| 20 | $A_{1.5}$ | 0.0731 | 0.1185 | 0.1421 | 0.1429 | 0.0804 | 0.1502 | **0.2796** |
| | $A_2$ | 0.1523 | 0.2684 | 0.2924 | 0.2992 | 0.1772 | 0.3322 | **0.6979** |
| | $B1.5$ | 0.0082 | 0.0335 | 0.1526 | **0.1582** | 0.0196 | 0.0585 | 0.0579 |
| | $B_2$ | 0.0053 | 0.0593 | 0.3020 | **0.3067** | 0.0126 | 0.0848 | 0.1246 |
| | $B_3$ | 0.0093 | 0.1317 | 0.5122 | 0.**5166** | 0.0385 | 0.1746 | 0.4133 |
| | $C_{1.5}$ | 0.2930 | 0.2316 | 0.0324 | 0.0334 | **0.2933** | 0.1027 | 0.1501 |
| | $C_2$ | 0.6430 | 0.5415 | 0.0425 | 0.0384 | **0.6673** | 0.1326 | 0.3159 |
| 30 | $A_{1.5}$ | 0.0942 | 0.1835 | 0.2001 | 0.2039 | 0.1155 | 0.2326 | **0.4066** |
| | $A_2$ | 0.2342 | 0.4228 | 0.4429 | 0.4532 | 0.2913 | 0.5028 | **0.8751** |
| | $B1.5$ | 0.0053 | 0.0582 | 0.2433 | **0.2441** | 0.0192 | 0.1215 | 0.0837 |
| | $B_2$ | 0.0083 | 0.1622 | 0.4851 | **0.4946** | 0.0339 | 0.2716 | 0.2411 |
| | $B_3$ | 0.0299 | 0.3916 | 0.7583 | **0.7629** | 0.1242 | 0.5309 | 0.7524 |
| | $C_{1.5}$ | 0.4023 | 0.3224 | 0.0259 | 0.0232 | **0.4029** | 0.0832 | 0.1892 |
| | $C_2$ | 0.8224 | 0.7425 | 0.0353 | 0.0244 | **0.8450** | 0.0942 | 0.4383 |

# 5 Real Data

In this section, we present a few examples to demonstrate the behavior of the estimators. For each example, we utilize the uniformity tests introduced in the previous section. Initially, the data are transformed into a uniform distribution via the probability integral transform, followed by an evaluation of the uniformity hypothesis.

**Example 5.1.** Illowsky and Dean (2018) used 55 data set of babies' smiling time measured in seconds. The data follows a $U(0, 23)$ distribution. We transformed this data to $U(0, 1)$ and obtained test statistics as:

$$GV_n = 0.01735047, GB_n = 0.01267519, GS_n = 0.002416893, GD_n = 0.02292441, GQ_n = 0.008132604.$$

The values belong to the acceptance region. Therefore, Assumption $H_0$ is accepted, that is, the data follows the standard uniform distribution.

**Example 5.2.** We consider the data of 20 wild lizards collected by the zoologists in the southwestern United States. The total length (mm) of each given as follows:

179, 157, 169, 146, 143, 131, 159, 142, 141, 130, 142, 116, 130, 140, 138, 137, 134, 114, 90, 114.

Using common tests such as Kolmogorov Smirnov, Shapiro-Wilk and Anderson-Darling, it can be seen that the data follows a normal distribution. Now we want to transform the data into standard uniform distribution using the probability integral transformation. Therefore if $U_i = F_0(X_i)$ for $i = 1, \ldots, n$, where $F_0(X_i)$ is distribution function of the normal distribution with estaimated mean and variance of the data. The transform data is obtained as follows:

0.9804, 0.8326, 0.9408, 0.6620, 0.6056, 0.3715, 0.8562, 0.5864, 0.5670, 0.3530, 0.5864, 0.1419, 0.3530, 0.5475, 0.5081, 0.4884, 0.4289, 0.1205, 0.0091, 0.1205.

Next, the values of the test statistics based on transformed sample are computed as: $GV_n = 0.1453019$, $GB_n = 0.02535139$, $GS_n = 0.002548617$, $GD_n = 0.03378436$, $GQ_n = 0.03390633$. By comparing with the percentage points of the test statistics at the 0.05 level given in Table 4, which are equal to 0.4937, 0.0374, 0.0277, 0.0485, and 0.1813, respectively, since the values of the test statistics are smaller than the corresponding critical values it can be conclude that the assumption of normality of the data cannot be rejected at the 0.05 significance level.

**Example 5.3.** In this example, the data set symbolizes the simulated strengths of glass fibers presented by Mahmoud and Mandouh (2000)

1.014, 1.081, 1.082, 1.185, 1.223, 1.248, 1.267, 1.271, 1.272, 1.275, 1.276, 1.278, 1.286, 1.288, 1.292, 1.304, 1.306, 1.355, 1.361, 1.364, 1.379, 1.409, 1.426, 1.459, 1.460, 1.476, 1.481, 1.484, 1.501, 1.506, 1.524, 1.526, 1.535, 1.541, 1.568, 1.579, 1.581, 1.591, 1.593, 1.602, 1.666, 1.670, 1.684, 1.691, 1.704, 1.731, 1.735, 1.747, 1.748, 1.757, 1.800, 1.806, 1.867, 1.876, 1.878, 1.910, 1.916, 1.972, 2.012, 2.456, 2.592, 3.197, 4.121.

The values of the proposed test statistics after obtaining transformed sample are $GV_n = 0.4381156$, $GB_n = 0.03980463$, $GS_n = 0.03055442$, $GD_n = 0.05200372$, $GQ_n = 0.05855942$. By looking at the Table 4, we can see that because the test statistics are larger than the percentage points obtained for n=63, therefore, the assumption of normality will be rejected at the significance level of $\alpha = 0.05$.

**Example 5.4.** The following data had presented in Fuller et al. (1994) and represents the strength of glass for aircraft window.

18.83, 20.8, 21.657, 23.03, 23.23, 24.05, 24.321, 25.5, 25.52, 25.8, 26.69, 26.77, 26.78, 27.05, 27.67, 29.9, 31.11, 33.2, 33.73, 33.76, 33.89, 34.76, 35.75, 35.91, 36.98, 37.08, 37.09, 39.58, 44.045, 45.29, 45.381.

By using Kolmogorov-Smirnov (K-S) distance test statistic, Alshenawy (2020) showed that, among the distributions fitted to the data, including the exponential and A distributions, A distribution have the best fitted for data. Now by calculating the proposed test statistics after obtaining transformed sample, we show that A distribution, unlike the exponential distribution, is a suitable distribution to fit the data.

Table 8: The CDF, MLE and the proposed test statistics.

| Model | CDF | MLE | $GV_n$ | $GB_n$ | $GS_n$ | $GD_n$ | $GQ_n$ |
|---|---|---|---|---|---|---|---|
| Exponential | $1-\exp(-\beta x)$ | 0.032 | 0.7052102 | 0.1541674 | 0.08482313 | 0.2151081 | 0.07347814 |
| A | $\exp(\frac{1}{\beta}(1-\exp(\frac{\beta}{x})))$ | 125.662 | 0.1404678 | 0.0144915 | 0.01595046 | 0.02569515 | 0.07580088 |

Since, for A distribution, test statistics are lower than critical values 0.1875, 0.028615, 0.020166, 0.04088, and 0.07852028, thus it appears to fit the data very well.

## 6 Conclusion

In this paper, we proposed new estimators for varextropy. The consistency of the proposed estimators is proved. We employed Monte-Carlo simulation to derive the mean squared error (MSE) and bias of the estimators across various distributions. In addition, we assessed the suggested estimators based on their bias and MSE characteristics for the gamma distribution with paramters $(2,1)$, the standard uniform distribution and the exponential distribution with mean 1. Based on the bias and MSE results, we concluded that $VJB_n$ and $VJS_n$ had better performance. We introduced some goodness-of-fit tests of uniformity using the proposed estimators. Then, we obtained the critical and power values of the proposed tests by Monte Carlo simulation and compared the powers with the competitive powers in Alizadeh Noughabi and Shafaei Noughabi (2023), Kolmogorov-Smirnov and EDF-based tests proposed by Zhang (2002) . Our simulation findings demonstrated that the suggested tests performed well in comparison with the other tests of uniformity. In particular, for the $B_{1,5}$ alternative distribution, our proposed $GB_n$ test showed the best performance across all sample sizes. In our future work, we aim to propose new estimators for varextropy under length-biased sampling and to investigate their properties.

## References

Alizadeh Noughabi, H., and Jarrahiferiz, J. (2019), On the estimation of extropy. *Journal of Nonparametric Statistics*, **31**(1), 88-99.

Alizadeh Noughabi, H., and Shafaei Noughabi, M. (2023). Varentropy estimators with applications in testing uniformity. *Journal of Statistical Computation and Simulation*, **93**, 2582-2599.

Alizadeh Noughabi, H., and Shafaei Noughabi, M. (2024), On the estimation of varextropy. *Statistics*, **58**, 827-841.

Al-Labadi, L., and Berry, S. (2022), Bayesian estimation of extropy and goodness of fit tests. *Journal of Applied Statistics*, **49**, 357-370.

Alshenawy, R. (2020), A new one parameter distribution: properties and estimation with applications to complete and type II censored data. *Journal of Taibah University for Science*, **14**(1), 11-18.

Becerra, A., de la Rosa, J. I., González, E., Pedroza, A. D., and Escalante, N. I. (2018), Training deep neural networks with non-uniform frame-level cost function for automatic speech recognition. *Multimedia Tools and Applications*, 77, 27231-27267.

Correa, J.C. (1995), A new estimator of entropy. *Communication in Statistics-Theory and Methods*, **24**(10), 2439-2449.

Ebrahimi, N., Pflughoeft, K., and Soofi, E. (1994), Two measures of sample entropy. *Statistics and Probability Letters*, **20**(3), 225-234.

Fradelizi, M., Madiman, M., and Wang, L. (2016), Optimal concentration of information content for logconcave densities. In C. Houdré, D. Mason, P. Reynaud-Bouret & J. Rosiński (eds.), High Dimensional Probability VII. Progress in Probability, vol. 71, Cham, Springer, pp. 45-60.

Fuller, E.J., Frieman, S., Quinn, J., Quinn, G., Carter, W. (1994), Fracture mechanics approach to the design of glass aircraft windows. *A case study. International Society for Optics and Photonics*, 419-430.

Furuichi, S., and Mitroi, F.C. (2012), Mathematical inequalities for some divergences. *Physica A*, **391**, 388-400.

Goodarzi, F. (2024), Results on conditional variance in parallel system and lower bounds for varextropy.*Communication in Statistics-Theory and Methods*, **53**, 1590-1610.

Gneiting, T., and Raftery, A.E. (2007), Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, **102**(477), 359-378.

Grzegorzewski, P., and Wieczorkowski, R. (1999), Entropy-based goodness-of-fit test for exponentiality. *Communications in Statistics–Theory and Methods*, **28**, 1183-1202.

Illowsky, B., and Dean, S. (2018), Introductory statistics. Houston: OpenStax.

Jiang, J., Wang, R., Pezeril, M., and Wang, Q.A. (2011), Application of varentropy as a measure of probabilistic uncertainty for complex networks. *Science Bulletin*, **56**, 3677-3682.

Kolmogorov, A.N. (1933), Sulla determinazione empirica di una legge didistibuziane. *Giornale Z. dell'IstitutaItaliano degli Attuari*, **4**, 83-91.

Lad, F., Sanfilippo, G., and Agrò, G. (2015), Extropy: Complementary Dual of Entropy. *Statistical Science*, **30**, 40-58.

Li, J., Fradelizi, M., and Madiman, M. (2016), Information concentration for convex measures. *IEEE International Symposium on Information Theory*, Barcelona, 1128-1132.

Madiman, M., and Barron, A. (2007), Generalized entropy power inequalities and monotonicity properties of information. *IEEE Transactions on Information Theory*, **53**, 2317-2329.

Mahmoud, M.R., and Mandouh, R.M. (2013), On the transmuted frechet distribution. *Journal of Applied Sciences Research*, **9**(10), 5553-5561.

Shannon, C.E. (1948), A mathematical theory of communication. *The Bell System Technical Journal*, **27**, 379-432.

Silverman, B.W. (1978), Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The annals of statistics*, **6**, 177-184.

Smirnov, N. (1939), Estimate of derivation between empirical distribution functions in two independent samples (in Russian). *Bulletin Moscow University*, **2**, 3–16.

Soni, P., Dewan, I., and Jain, K. (2012), Nonparametric estimation of quantile density function. *Computational Statistics and Data Analysis*, **56**(12), 3876-3886.

Stephens, M.A. (1974), EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, **69**(347), 730-737.

Subhash, S., Sunoj, S.M., Sankaran, P.G., and Rajesh, G. (2023), Nonparametric estimation of quantile-based entropy function. *Communications in Statistics - Simulation and Computation*, **52**, 1805-1821.

Qiu, G., and Jia, K. (2018a), Extropy estimators with applications in testing uniformity. *Journal of Nonparametric Statistics*, **30**(1), 182-196.

Qiu, G., and Jia, K. (2018b), The residual extropy of order statistics. *Statistics and Probability Letters*, **133**, 15-22.

van Es, B. (1992), Estimating functionals related to a density by a class of statistics based on spacings. *Scandinavian Journal of Statistics*, **19**(1), 61-72.

Vaselabadi, N.M., Tahmasebi, S., Kazemi, M.R., and Buono, F. (2021), Results on varextropy measure of random variables. *Entropy*, **23**, 356.

Vasicek, O. (1976), A test for normality based on sample entropy. *Journal of the Royal Statistical Society: Series B*, **38**, 54-59.

Vontobel, P.O. (2013), The Bethe permanent of a nonnegative matrix. *IEEE Transactions on Information Theory*, **59**, 1866-1901.

Yang, J., Xia, W., and Hu, T. (2019), Bounds on extropy with variational distance constraint. *Probability in the Engineering and Informational Sciences*, **33**(2), 186-204.

Zamanzade, E. (2015), Testing uniformity based on new entropy estimators. *Journal of Statistical Computation and Simulation*, **85**(16), 3191-3205.

Zhang, J. (2002), Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society Series B*, **64**(2), 281-294.