axioms

Article

Bayesian Estimation of Variance-Based Information Measures and Their Application to Testing Uniformity

Luai Al-Labadi *, Mohammed Hamlili and Anna Ly

Department of Mathematical & Computational Sciences, University of Toronto Mississauga, Toronto, ON L5L 1C6, Canada; mohammed.hamlili@mail.utoronto.ca (M.H.); annahuynh.ly@mail.utoronto.ca (A.L.)
* Correspondence: luai.allabadi@utoronto.ca

Abstract: Entropy and extropy are emerging concepts in machine learning and computer science. Within the past decade, statisticians have created estimators for these measures. However, associated variability metrics, specifically varentropy and varextropy, have received comparably less attention. This paper presents a novel methodology for computing varentropy and varextropy, drawing inspiration from Bayesian nonparametric methods. We implement this approach using a computational algorithm in R and demonstrate its effectiveness across various examples. Furthermore, these new estimators are applied to test uniformity in data.

Keywords: Bayesian nonparametric inference; entropy; extropy information theory; goodness-of-fit tests

MSC: 94A17; 62F03

check for updates

Citation: Al-Labadi, L.; Hamlili, M.; Ly, A. Bayesian Estimation of Variance-Based Information Measures and Their Application to Testing Uniformity. *Axioms* **2023**, *12*, 887. https://doi.org/10.3390/axioms12090887

Academic Editors: Lechang Yang, Qingqing Zhai, Rui Peng and Aibo Zhang

Received: 16 August 2023
Revised: 13 September 2023
Accepted: 15 September 2023
Published: 17 September 2023

1. Introduction

In this section, we begin by reviewing the concepts of entropy and extropy, along with some existing estimators from the literature. Subsequently, we introduce estimators of varentropy and varextropy developed using a frequentist approach.

Entropy is a fundamental concept in information theory that was originally introduced in [1]. It has found numerous applications in various fields, such as thermodynamics, communication theory, reliability, computer science, biology, economics, and statistics [2,3]. Let X be a continuous random variable with support on $\mathcal{S}$, cumulative distribution function (CDF) F, and probability density function (PDF) f. The entropy $H(F)$ of X is defined as

$$H(F) = E_f(-\log f(x)) = -\int_{\mathcal{S}} f(x)\log f(x)dx, \tag{1}$$

where log is the natural logarithm.

Extropy, a concept introduced by [4], represents a relatively recent development in the field of statistics that serves as a dual counterpart to entropy. Its significance has been demonstrated in various studies, including those conducted by [5,6], where it has found applications in the context of goodness-of-fit tests.

For a continuous random variable X supported on $\mathcal{S}$ with CDF F and PDF f, the extropy of X, denoted by $J(F)$, is defined as

$$J(F) = -\frac{1}{2}E_f(f(x)) = -\frac{1}{2}\int_{\mathcal{S}} f^2(x)dx \tag{2}$$

In most practical scenarios, the true PDF f is unknown, and hence, we need to estimate the entropy (1) and extropy (2) from the available data, which can be a challenging task. Several frequentist methods are available for entropy estimation in the literature.

Among different approaches, Ref. [7] estimator has gained wide popularity due to its simplicity. Ref. [7] derived the expression for (1) in terms of the inverse of the distribution function F, given by

$$H(F) = \int_0^1 \log\left(\frac{d}{dt}F^{-1}(t)\right)dt.$$

Using the empirical distribution function F_n instead of the unknown F, Ref. [7] proposed an estimator for $H(F)$ based on the difference operator rather than the differential operator. The derivative of $F^{-1}(t)$ is estimated using a function of the order statistics. Specifically, if $X_1, X_2, \ldots, X_n$ is a sample from F, then [7] estimator is given by

$$H1_{m,n} = n^{-1}\sum_{i=1}^{n}\log\left(\frac{X_{(i+m)} - X_{(i-m)}}{2m/n}\right), \tag{3}$$

where m is a positive integer smaller than $n/2$, and $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are the order statistics of $X_1, X_2, \ldots, X_n$ with $X_{(i-m)} = X_{(1)}$ if $i \leq m$ and $X_{(i+m)} = X_{(n)}$ if $i \geq n-m$. [7] showed that $H1_{m,n} \xrightarrow{p} H(F)$ as $n \to \infty$, $m \to \infty$, and $\frac{m}{n} \to 0$, where $\xrightarrow{p}$ denotes convergence in probability. Note that the expression inside the log in (3) is the slope of the straight line that passes through the points $\left(\frac{i+m}{n}, X_{(i+m)}\right)$ and $\left(\frac{i-m}{n}, X_{(i-m)}\right)$, where $F_n(X_{(i+m)}) = \frac{i+m}{n}$ and $F_n(X_{(i-m)}) = \frac{i-m}{n}$. Ref. [8] proposed a modification to the estimator (3) as it does not provide the correct formula for the slope when $i \leq m$ or $i \geq n-m+1$. The proposed estimator, denoted by $H2_{m,n}$, is given by

$$H2_{m,n} = n^{-1}\sum_{i=1}^{n}\log\left(\frac{X_{(i+m)} - X_{(i-m)}}{c_i m/n}\right), \tag{4}$$

where

$$c_i = \begin{cases} \frac{m+i-1}{m} & \text{if } 1 \leq i \leq m \\ 2 & \text{if } m+1 \leq i \leq n-m \\ \frac{n+m-i}{m} & \text{if } n-m+1 \leq i \leq n \end{cases}. \tag{5}$$

As for extropy estimation, Ref. [5] noticed that the extropy, first defined in (2) can be rewritten as

$$J(F) = -\frac{1}{2}\int_0^1 \left[\frac{d}{dt}F^{-1}(t)\right]^{-1} dt$$

and proposed the following estimator for $J(F)$:

$$J1_{m,n} = -\frac{1}{2n}\sum_{i=1}^{n}\frac{2m/n}{X_{(i+m)} - X_{(i-m)}},$$

As in $H1_{m,n}$, Ref. [5] found that $J1_{m,n}$ gives incorrect estimates for $i \leq m$ or $i \geq n-m+1$. Therefore, they proposed the revised estimator $J2_{m,n}$, where

$$J2_{m,n} = -\frac{1}{2n}\sum_{i=1}^{n}\frac{c_i m/n}{X_{(i+m)} - X_{(i-m)}}.$$

Here, c_i is defined as specified in Equation (5). Ref. [5] proved that $J1_{m,n}$ and $J2_{m,n}$ converges in probability to $J(F)$ under the same conditions as for $H1_{m,n}$ or $H2_{m,n}$.

Other frequentist nonparametric estimators of entropy include those proposed by [9–14]. A comprehensive review of nonparametric entropy estimators can be found in [15]. For extropy estimation, alternative approaches are provided by [16] as well as [17].

Bayesian estimation of entropy has not received as much attention as the frequentist approach. However, Ref. [18] developed a Bayes estimator of $H(F)$ based on the Dirichlet process [19]. Recently, Refs. [6,20,21] proposed an estimator of entropy and extropy based on an approximation of the Dirichlet process $DP(a, G)$ introduced by [22], where $a > 0$ and G is a known CDF. The approximation is defined as

$$P_N(\cdot) = \sum_{i=1}^{N} w_{i,N}\delta_{Y_i}(\cdot), \tag{6}$$

In this equation, the weights $(w_{1,N}, \ldots, w_{N,N})$ follow a Dirichlet distribution with parameters $(a/N, \ldots, a/N)$, while $Y_1, \ldots, Y_N$ are independent and identically distributed from the distribution G. The notation δ_{Y_i} represents the Dirac measure at the point Y_i. The sequences $(W_{i,N})1 \le i \le N$ and $(Y_i)1 \le i \le N$ are independent. We refer to $(Y_i)1 \le i \le N$ as the *data points* of P_N. Let

$$H_{m,N,a} = \frac{1}{N}\sum_{i=1}^{N} \log\left(\frac{Y_{(i+m)} - Y_{(i-m)}}{c_{i,a}}\right) \tag{7}$$

and

$$J_{m,N,a} = -\frac{1}{2N}\sum_{i=1}^{N} \frac{c_{i,a}}{Y_{(i+m)} - Y_{(i-m)}}, \tag{8}$$

where

$$c_{i,a} = \begin{cases} \sum_{k=2}^{i+m} w_{k,N} & 1 \le i \le m, \\ \sum_{k=i-m+1}^{i+m} w_{k,N} & m+1 \le i \le N-m, \\ \sum_{k=i-m+1}^{N} w_{k,N} & N-m+1 \le i \le N. \end{cases} \tag{9}$$

As $N \to \infty$, $m \to \infty$, $\frac{m}{N} \to 0$ and $a \to \infty$, Refs. [6,20] showed that

$$H_{m,N,a} \xrightarrow{p} H(G) = -\int_{\mathcal{S}} g(x)\log g(x)dx$$

and

$$J_{m,N,a} \xrightarrow{p} J(G) = -\frac{1}{2}\int_{\mathcal{S}} g^2(x)dx,$$

where $G'(x) = g(x)$. Observe that the slope of the straight line connecting the two points $\left(P_N(Y_{(i-m)}),\ Y_{(i-m)}\right)$ and $\left(P_N(Y_{(i+m)}),\ Y_{(i+m)}\right)$ is

$$\frac{Y_{(i+m)} - Y_{(i-m)}}{P_N(Y_{(i+m)}) - P_N(Y_{(i-m)})} = \frac{Y_{(i+m)} - Y_{(i-m)}}{c_{i,a}}.$$

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a sample from F and $DP(a, G)$ be a prior of F. Consider $H_{m,N,a}|\boldsymbol{X}$ to be the posterior version of $H_{m,N,a}$ as defined in (7) with P_N replaced by $P_N|\boldsymbol{X}$, an approximation of $DP(a+n, G_{\boldsymbol{X}})$, where

$$G_{\boldsymbol{X}} = a(a+n)^{-1}G + n(a+n)^{-1}F_n. \tag{10}$$

Then, as $N \to \infty$, $m \to \infty$, $n \to \infty$, $\frac{m}{N} \to 0$, and $\frac{a}{n} \to 0$, we have [6,20]

$$H_{m,N,a}|\boldsymbol{X} \xrightarrow{p} H(F) = -\int_{\mathcal{S}} f(x)\log f(x)dx,$$

and

$$J_{m,N,a}|\boldsymbol{X} \xrightarrow{p} J(F) = -\frac{1}{2}\int_{\mathcal{S}} f^2(x)dx,$$

where $F'(x) = f(x)$.

Recently, there has been significant interest in studying the variability of information measures in the literature. In certain situations, it is possible to encounter two random variables with identical entropy or extropy. As a result of these circumstances, researchers ponder whether entropy or extropy would be the most suitable criterion for measuring uncertainty. One way of determining which probability distribution is more suitable is to check its variance. This scenario serves as a motivation for exploring two variance measures associated with entropy and extropy, known as varentropy and varextropy, respectively.

For a random variable X, the varentropy, denoted by $VH(F)$, is defined as follows:

$$\begin{aligned} VH(F) = \mathrm{Var}_f(\log(f(X))) &= \int_S f(x)[\log f(x)]^2 dx - \left[\int_S f(x)\log f(x)dx\right]^2 \\ &= E_f\left(\log((f(x))^2)\right) - [H(F)]^2. \end{aligned}$$

Ref. [23] introduced varentropy as a compelling alternative to the kurtosis measure, particularly for comparing heavy-tailed distributions. In fact, varentropy has proven to be a valuable tool for assessing heavy-tailed distributions instead of relying solely on kurtosis. Subsequently, varentropy has found diverse applications across various fields. In computer science, varentropy plays an instrumental role in data compression, studying the variability of uncertainty measures [24], testing uniformity [25], and analyzing statistical issues [26]. Moreover, researchers have effectively employed varentropy to explore the variability of interval entropy measures [27]. Additionally, varentropy has proven valuable in applications related to proportional hazard rate models [24] and residual lifetime distributions [26].

Ref. [25] presented six estimators for computing varentropy. In this context, we will focus on two specific estimators based on the estimators (3) and (4). Specifically, $VH(F)$ can be expressed as

$$VH(F) = \int_0^1 \log^2\left(\frac{d}{dt}F^{-1}(t)\right)dt - \left[\int_0^1 \log\left(\frac{d}{dt}F^{-1}(t)\right)dt\right]^2, \tag{11}$$

the two estimators of [25] are

$$\begin{aligned} VH1_{mn} &= \frac{1}{n}\sum_{i=1}^{n}\log^2\left(\frac{X_{(i+m)} - X_{(i-m)}}{2m/n}\right) - [H1_{m,n}]^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\log^2\left(X_{(i+m)} - X_{(i-m)}\right) - \left[\frac{1}{n}\sum_{i=1}^{n}\log(X_{(i+m)} - X_{(i-m)})\right]^2 \end{aligned} \tag{12}$$

and

$$VH2_{mn} = \frac{1}{n}\sum_{i=1}^{n}\log^2\left(\frac{X_{(i+m)} - X_{(i-m)}}{c_i m/n}\right) - [H2_{m,n}]^2,$$

where $H1_{m,n}$, $H2_{m,n}$, and c_i are defined in (3), (4), and (5), respectively. Noughabi and Noughabi (2023) showed that $VH1_{mn}$ and $VH2_{mn}$ converge in probability to $VH(F)$ under the same conditions as for $H1_{m,n}$ or $H2_{m,n}$.

Another measure of information variability is the varextropy. Let X be an absolutely continuous random variable, then its varextropy, denoted as $VJ(X)$, is defined as follows [28]:

$$VJ(X) = \mathrm{Var}_f(-\frac{1}{2}f(X)) = \frac{1}{4}E_f((f(X))^2) - (J(X))^2 \tag{13}$$

$$= \frac{1}{4}\int_S f^3(x)dx - \frac{1}{4}\left[\int_S f^2(x)dx\right]^2. \tag{14}$$

Unlike the varentropy $VH(F)$, the estimation of $VJ(F)$ has not been intensively discussed in the literature. Notice that

$$VJ(F) = \frac{1}{4}\int_0^1 \left[\frac{d}{dt}F^{-1}(t)\right]^{-2} dt - \frac{1}{4}\left[\int_0^1 \left[\frac{d}{dt}F^{-1}(t)\right]^{-1} dt\right]^2. \tag{15}$$

Accordingly, our proposed estimators of $VJ(X)$ based on the two estimator of [5] are:

$$VJ1_{mn} = \frac{1}{4n}\sum_{i=1}^{n}\left[\frac{2m/n}{x_{(i+m)} - x_{(i-m)}}\right]^2 - [J1_{m,n}]^2 \tag{16}$$

and

$$VJ2_{mn} = \frac{1}{4n}\sum_{i=1}^{n}\left[\frac{c_i m/n}{x_{(i+m)} - x_{(i-m)}}\right]^2 - [J2_{m,n}]^2, \tag{17}$$

where c_i is defined in (5). The convergence in probability of $VJ1_{mn}$ and $VJ2_{mn}$ to $VJ(F)$ straightforwardly follows from the consistency of $J1_{m,n}$ and $J2_{m,n}$.

The rest of this paper is structured as follows. Section 2 presents the proposed Bayesian estimator based on the Dirichlet process. Section 3 details the proposed approach, including a computational algorithm. In Section 4, a test for uniformity is developed. Section 5 presents several examples to illustrate the approach. Finally, Section 6 contains concluding remarks and discussions.

2. Bayesian Estimation of Varentropy and Varextropy

In this section, we derive Bayesian nonparametric estimators for varentropy and varextropy. Define the following two quantities:

$$VH_{m,N,a} = \frac{1}{N}\sum_{i=1}^{N}\log^2\left(\frac{Y_{(i+m)} - Y_{(i-m)}}{c_{i,a}}\right) - [H_{m,N,a}]^2 \tag{18}$$

and

$$VJ_{m,N,a} = \frac{1}{4N}\sum_{i=1}^{N}\left[\frac{c_{i,a}}{Y_{(i+m)} - Y_{(i-m)}}\right]^2 - [J_{m,N,a}]^2, \tag{19}$$

where $c_{i,a}$, $H_{m,N,a}$, and $J_{m,N,a}$ are defined in (9), (7), and (8), respectively. The following proposition presents the prior formulation of varentropy and varextropy. The proof follows from the consistency of $H_{m,N,a}$ and $J_{m,N,a}$.

Lemma 1. *Let $VH_{m,N,a}$ and $VJ_{m,N,a}$ be defined as in (18) and (19), respectively. Let P_N be an approximation of the $DP(a, G)$ as defined (6). As $N \to \infty$, $m \to \infty$, $\frac{m}{N} \to 0$ and $a \to \infty$, we have*

$$VH_{m,N,a} \xrightarrow{p} VH(G)$$

and

$$VJ_{m,N,a} \xrightarrow{p} VJ(G),$$

where $G'(x) = g(x)$.

The following proposition demonstrates that as the sample size increases (with the concentration parameter a being relatively small compared to the sample size n), the posterior distributions of $VH_{m,N,a}$ and $VJ_{m,N,a}$ converge in probability to $VH(F)$ and $VJ(F)$, respectively. The proof follows from consistency of $H_{m,N,a}|\boldsymbol{X}$ and $J_{m,N,a}|\boldsymbol{X}$.

Lemma 2. *Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from F. Let the prior on F be $DP(a, H)$. Let $VH_{m,N,a}$ and $VJ_{m,N,a}$ be as defined in (18) and (19), respectively. As $N \to \infty$, $m \to \infty$, $n \to \infty$, $\frac{m}{N} \to 0$, and $\frac{a}{n} \to 0$, we have*

$$VH_{m,N,a}|\mathbf{X} \xrightarrow{p} VH(F)$$

and

$$VJ_{m,N,a}|\mathbf{X} \xrightarrow{p} VJ(F),$$

where $VH(F)$ and $VJ(F)$ are defined in (11) and (14), respectively, with $F'(x) = f(x)$.

3. Computational Algorithms

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from a continuous distribution F. The objective is to approximate $VH(F)$ and $JH(F)$ using the approximation discussed in Section 2. To proceed with this approximation, it is important to determine the values of m, a, and G. We begin by considering the choice of m. A commonly used formula, proposed by [29], is given by

$$m = \lfloor \sqrt{N} + 0.5 \rfloor, \tag{20}$$

where $\lfloor y \rfloor$ denotes the largest integer less than or equal to y. Note that the value of m in (20) is used for the prior. For the posterior, the value of N should be replaced with the number of distinct data points in $P_N|\mathbf{X}$, an approximation of $F|\mathbf{X}$. It is worth noting that, from (10), if a/n is close to zero, the number of distinct data points in $P_N|\mathbf{X}$ will be approximately n.

Regarding the hyperparameters a and G of the Dirichlet process, their selection depends on the specific application of interest. For varentropy and varextropy estimation, any choice of a such that a/n is close to zero should be suitable, regardless of the choice of G. This property is evident from (10), as when a/n approaches 0, the sample will dominate the prior guess G. Consequently, the approach becomes invariant to the choice of G. As an illustrative example, by setting $a = 0.01$ and $n = 20$ in (10), we obtain

$$G_{\mathbf{X}} = 0.0005G + 0.9995F_n.$$

This suggests a 99.95% likelihood of selecting a sample from the gathered data rather than drawing a new sample from G. To facilitate estimation, we will consider G as the uniform distribution over $(0, 1)$ and set $a = 0.01$, although alternative choices are certainly possible. Within Section 4, we will include an example that investigates the sensitivity of the approach to the choices of a and G.

For a given observed data set $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, we employ the following computational algorithm to estimate $VH(F)$ and $VJ(F)$ based on Equations (18) and (19).

Algorithm 1. *(Nonparametric Estimation of Varentropy and Varextropy):*

(i) *Generate a sample from P_N, where P_N is an approximation of $DP(a = 0.01, G = U(0, 1))$.*
(ii) *Generate a sample from $P_N|\mathbf{X}$, where $P_N|\mathbf{X}$ is an approximation of $\sim DP(a + n, G_{\mathbf{X}})$.*
(iii) *Compute $VH_{m,N,a}|\mathbf{X}$ and $VJ_{m,N,a}|\mathbf{X}$ as specified in Lemma 2.*
(iv) *Repeat steps (i) and (iii) to obtain a sample of r values from $VH_{m,N,a}|\mathbf{X}$ and $VJ_{m,N,a}|\mathbf{X}$. As r increases, the average of the generated r values becomes the estimator for varentropy and varextropy.*

4. Testing for Uniformity

Suppose that $\mathbf{X} = (X_1, \ldots, X_n)$ is a sample from an unknown continuous distribution F. The objective is to test the hypothesis $\mathcal{H}_0 : F(x) = F_0(x)$, for all $x \in \mathbb{R}$, where F_0 represents a fully specified distribution. By utilizing the probability integral transform property, we can deduce that $F(X_1), \ldots, F(X_n)$ follows the uniform distribution on the interval $(0, 1)$. Therefore, testing the null hypothesis is equivalent to testing $\mathcal{H}_0 : U(x) = x$

for all $x \in (0,1)$, where $U(x)$ represents the CDF of a uniform $(0,1)$ random variable. For further details on testing uniformity, please consult the work of [30].

For any random variable X, it holds that $VH(F) \geq 0$ and $VJ(F) \geq 0$. The following proposition demonstrates that equality is achieved when F corresponds to the CDF of a uniform distribution on the interval $(0,1)$. This property plays a crucial role in testing the hypothesis H_0.

Proposition 1. *Let f be a probability density function with support in $[0,1]$, we have*

(i) $VH(F) = 0$ *if and only if* $f(x) = 1$ *for all* $x \in (0,1)$ *(i.e., f is the PDF of the uniform random variable on $(0,1)$).*

(ii) $VJ(F) = 0$ *if and only if* $f(x) = 1$ *for all* $x \in (0,1)$ *(i.e., f is the PDF of the uniform random variable on $(0,1)$).*

Proof. For the proof of (i), see Theorem 4.1 of [25]. For (ii), if $f(x) = 1$ for all $x \in (0,1)$, then $VJ(F) = \mathrm{Var}(-0.5f(X)) = 0$. Also, if $VJ(F) = 0$, then $f(x) = c$, for all $x \in (0,1)$. Since $\int_0^1 f(x)dx = 1$, we have $c = 1$. Hence, $f(x) = 1$ for all $x \in (0,1)$. $\square$

The proposed test for uniformity involves comparing $VH_{m,N,a}|\boldsymbol{X}$ and $VJ_{m,N,a}|\boldsymbol{X}$. When the null hypothesis $\mathcal{H}_0$ is true, it is expected that $VH_{m,N,a}|\boldsymbol{X} \approx VJ_{m,N,a}|\boldsymbol{X} \approx 0$. Conversely, if there is evidence that $VH_{m,N,a}|\boldsymbol{X}$ or $VJ_{m,N,a}|\boldsymbol{X}$ deviates significantly from zero, $\mathcal{H}_0$ is rejected.

5. Examples

5.1. Simulation Study

In this subsection, we focus on investigating the efficiency and robustness of the proposed estimator for varentropy and varextropy. Additionally, we demonstrate the implementation of the uniformity test using these estimations. To evaluate the performance of our proposed Bayesian estimator, we compare it with the non-Bayesian counterparts obtained from (12), (13), (16), and (17). The comparison between Bayesian and non-Bayesian methods holds particular significance in this context, particularly when we consider the scenario where $a = 0.01$. In this case, the estimator remains unaffected by the selection of the prior guess G, as demonstrated in Section 4.

To carry out the computations, we implemented the required program codes in the programming language `R`, and these codes are made available by the authors. For demonstration purposes, we constructed a demo for the algorithms and presented them using R Shiny, as shown here: https://annaly.shinyapps.io/BayesianVarentropyVarextropy/ (accessed on 15 July 2023). In Algorithm 1, we set the parameters $r = 1000$ and $N = 500$ to ensure accurate and sufficient evaluations. Additionally, to ensure reproducibility, the `set.seed(100)` function in `R` was utilized for all examples.

Throughout this section, we use the following notation: $N(\mu, \sigma^2)$ denotes the normal distribution with mean μ and standard deviation σ, t_r represents the t distribution with r degrees of freedom, $\mathrm{Exp}(\lambda)$ corresponds to the exponential distribution with mean $1/\lambda$, $U(a,b)$ signifies the uniform distribution over the interval (a,b), and beta(α, β) denotes the beta distribution with parameters α and β.

In Tables 1 and 2, for each sample size ($n = 20, 50, 100$), 1000 samples were generated. We have considered three distributions: uniform on $(0,1)$ (exact varentropy and varextopy are both 0), exponential with mean 1 (exact varentropy and varextropy are 1 and $1/48 \approx 0.0208$, respectively), $N(0,1)$ (exact varentropy and exact varextropy are 0.5 and $(2-\sqrt{3})/16\pi\sqrt{3} \approx 0.0031$, respectively). The estimators and their root mean squared errors are computed and reported in Tables 1 and 2. The reported value of the estimator (Est) is the average of the 1000 estimates. On the other hand, the root mean squared error (RMSE) is computed as follows: $\sqrt{\sum_{i=1}^{1000}(\mathrm{Est}_i - \text{true value})^2/1000}$, where Est_i is the estimated value based on the ith sample.

Table 1. Varentropy Measure Estimates.

			$VH_{m,n,a}\|X$	$VH1_{m,n}$	$VH2_{m,n}$
Distribution	n	m		**Est(RMSE)**	
U(0,1)	20	4	0.2055(0.2248)	0.1749(0.218)	0.1099(0.1465)
	50	7	0.1266(0.1334)	0.1059(0.1214)	0.0641(0.0769)
	100	10	0.0919(0.0954)	0.0783(0.0873)	0.0469(0.0536)
Exp(1)	20	4	0.7446(0.4699)	0.6663(0.5105)	0.6586(0.5258)
	50	7	0.9025(0.3148)	0.7929(0.3454)	0.8344(0.3396)
	100	10	0.9671(0.2315)	0.8641(0.2533)	0.9203(0.2417)
$N(0,1)$	20	4	0.2646(0.2621)	0.1221(0.3885)	0.1620(0.3574)
	50	7	0.3051(0.2245)	0.1449(0.3633)	0.2351(0.2864)
	100	10	0.3666(0.1677)	0.2136(0.2971)	0.319(0.2076)

Table 2. Varextropy Measure Estimates.

			$VJ_{m,n,a}\|X$	$JQ1_{m,n}$	$JQ2_{m,n}$
Distribution	n	m		**Est(RMSE)**	
U(0,1)	20	4	0.0828(0.1159)	0.1698(0.3129)	0.0502(0.0875)
	50	7	0.0404(0.0460)	0.0642(0.0943)	0.0217(0.0291)
	100	10	0.0262(0.0283)	0.0352(0.0460)	0.0135(0.0166)
Exp(1)	20	4	0.0589(0.1451)	0.1419(0.5516)	0.0505(0.1466)
	50	7	0.0370(0.0291)	0.0621(0.0696)	0.0311(0.0238)
	100	10	0.0337(0.0197)	0.0514(0.0435)	0.0294(0.0161)
$N(0,1)$	20	4	0.0063(0.0065)	0.0062(0.0136)	0.0043(0.0047)
	50	7	0.0053(0.0034)	0.0033(0.0022)	0.0041(0.0026)
	100	10	0.0069(0.0044)	0.0037(0.0018)	0.0043(0.0021)

Based on the findings presented in Tables 1 and 2, it is evident that the estimators for varentropy and varextropy demonstrate overall good performance.

It is also of interest to examine the impact of utilizing different base measures G and concentration parameters a on the methodology. To explore this, we consider two distinct values for a, namely 0.01 and 5, and examine various choices for G. In our analysis, we utilize a dataset generated from the exponential distribution with a mean of 1. Based on the findings presented in Table 3, it can be concluded that the estimators exhibit robustness to the choice of G when $a = 0.01$.

In this last example, we generated samples of sizes $n = 20$, 50, and 100 from the uniform distribution on the interval $(0, 1)$. The goal is to test the hypothesis $H_0 : F(x) = F_0(x)$ using the proposed test of uniformity. To achieve this, we considered a range of candidate distribution functions $F_0(x)$ as outlined in Table 4, where

$$A_k : F_0(x) = 1 - (1 - x)^k, 0 \leq x \leq 1 \ \text{(for } k = 1.5, 2);$$

$$B_k : F_0(x) = \begin{cases} 2^{k-1}x^k & 0 \leq x \leq 0.5, \\ 1 - 2^{k-1}(1 - x)^k & 0.5 \leq x \leq 1 \end{cases} \ \text{(for } k = 1.5, 2, 3);$$

$$C_k : F_0(x) = \begin{cases} 0.5 - 2^{k-1}(0.5 - x)^k & 0 \leq x \leq 0.5, \\ 0.5 + 2^{k-1}(0.5 - x)^k & 0.5 \leq x \leq 1 \end{cases} \ \text{(for } k = 1.5, 2).$$

Table 3. Analysis of the impact of different values of a and G on the proposed estimators.

G	a	$VH_{m,n,a}\|X$	$VJ_{m,n,a}\|X$
$N(0,1)$	0.01	0.8758	0.0663
	5	2.4917	0.0255
$N(3,9)$	0.01	0.8877	0.0667
	5	3.2328	0.0073
t_1	0.01	0.8932	0.0681
	5	7.7335	0.0215
Exp(1)	0.001	0.8587	0.0657
	5	1.6113	0.0725
$U(0,1)$	0.01	0.8699	0.0668
	5	0.7525	0.0776

These candidate distribution functions A_k, B_k, and C_k of $F_0(x)$ have been previously studied by various authors, including [25,31]. The distributions of Exp(2) and $N(0,1)$ are included here to explore cases with support different from $[0,1]$. The results are presented in Table 4, where we also included the p-values obtained from the Kolmogorov–Smirnov test.

Table 4. Goodness-of-Fit Test.

n	F_0	Est. of VH	Est. of VJ	Threshold of VH	Threshold of VJ	p-Value
20	$A_{1.5}$	0.1573	0.0423	0.2461	0.0955	0.2953
	A_2	0.2494	0.0839			0.0442
	$B_{1.5}$	0.1322	0.0337			0.7472
	B_2	0.1553	0.0421			0.4646
	B_3	0.2934	0.1163			0.1469
	$C_{1.5}$	0.3082	0.1687			0.1958
	C_2	0.5543	0.6276			0.0745
	$U(0,1)$	0.1529	0.0484			0.6307
	Exp(2)	0.2604	0.1169			0.1300
	beta(3,1)	1.1438	2.6605			0.0000
	$N(0,1)$	0.1944	0.3892			0.0000
50	$A_{1.5}$	0.2378	0.1044	0.1469	0.0475	0.0073
	A_2	0.6105	0.7342			0.0000
	$B_{1.5}$	0.1176	0.0391			0.4119
	B_2	0.2564	0.1421			0.0857
	B_3	0.7699	1.6625			0.0048
	$C_{1.5}$	0.2449	0.1046			0.0430
	C_2	0.54717	0.4817			0.0045
	$U(0,1)$	0.0876	0.0232			0.3349
	Exp(2)	0.3819	0.2553			0.0013
	beta(3,1)	1.0189	1.6964			0.0000
	$N(0,1)$	0.1361	0.2849			0.0000
100	$A_{1.5}$	0.2073	0.0873	0.1042	0.03107	0.0019
	A_2	0.5906	0.8231			0.0000
	$B_{1.5}$	0.0958	0.0329			0.6042
	B_2	0.2602	0.1665			0.0764
	B_3	0.9135	3.0551			0.0011
	$C_{1.5}$	0.2877	0.1323			0.0080
	C_2	0.7074	0.8427			0.0001
	$U(0,1)$	0.0755	0.0193			0.2672
	Exp(2)	0.3507	0.1926			0.0000
	beta(3,1)	1.3669	4.1632			0.0000
	$N(0,1)$	0.1006	0.1951			0.0000

Using Monte Carlo simulation, we can determine an appropriate cut-off for both $VH_{m,n,a}|\boldsymbol{X}$ and $VJ_{m,n,a}|\boldsymbol{X}$ in the test of uniformity under $\mathcal{H}0$. We recommend using Q_3, the third quartile, as a suitable threshold. In Table 4, we present the thresholds for $n = 20, 50$, and 100, based on 5000 values of $VHm, n, a|\boldsymbol{X}$ and $VJ_{m,n,a}|\boldsymbol{X}$. If both estimates are less than their respective thresholds, it is advisable not to reject the null hypothesis $\mathcal{H}_0$. However, if one of the estimates is greater than its threshold, it is recommended to reject H_0.

For example, when $n = 50$ and $F_0 = A_{1.5}$, as $VH_{m,n,a}|\boldsymbol{X} > 0.1469$ (or $VJ_{m,n,a}|\boldsymbol{X} > 0.0475$), $\mathcal{H}0$ is rejected. Conversely, when $F_0 = B_{1.5}$, as both $VH_{m,n,a}|\boldsymbol{X} < 0.1469$ and $VJ_{m,n,a}|\boldsymbol{X} < 0.0475$, $\mathcal{H}_0$ is not rejected.

5.2. Real Data Examples

Military Personnel Carriers Dataset [32]: The following data represents mileages for 19 military personnel carriers that failed in service. The mileages are as follows:

162, 200, 271, 320, 393, 508, 539, 629, 706, 778, 884, 1003, 1101, 1182, 1463, 1603, 1984, 2355, 2880.

The aim of the study is to test whether the data follows an exponential distribution with a mean of 998. Employing Algorithm 1 with parameters $n = 19$, $N = 500$, and $r = 1000$, we obtained $VH_{m,n,a}|\boldsymbol{X} = 0.1124$ and $VJ_{m,n,a}|\boldsymbol{X} = 0.0331$. Since both of these values are significantly lower than their respective thresholds (0.2509 for $VH_{m,n,a}|\boldsymbol{X}$ and 0.0969 for $VJ_{m,n,a}|\boldsymbol{X}$), we cannot reject the hypothesis that the failure time is exponentially distributed with a mean of 998. This conclusion aligns with the findings of [33].

Chick Dataset [30]: The dataset below represents the weights of 20 chicks in grams:

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

The goal of this study is to test whether the data follow a normal distribution with a mean of 200 and a variance of 1225. Using Algorithm 1 with parameters $n = 20$, $N = 500$, and $r = 1000$, we obtained $VH_{m,n,a}|\boldsymbol{X} = 0.1396$ and $VJ_{m,n,a}|\boldsymbol{X} = 0.0000$. Both of these values are significantly lower than their respective thresholds (0.2461 for $VH_{m,n,a}|\boldsymbol{X}$ and 0.0955 for $VJ_{m,n,a}|\boldsymbol{X}$). Consequently, we cannot reject the hypothesis that the data follow a normal distribution with a mean of 200 and a variance of 1225. This conclusion is consistent with the findings of [30].

6. Conclusions

In this paper, we introduced a novel estimator for varentropy and varextropy, drawing inspiration from Bayesian nonparametric statistical methods. This method exhibits flexibility as it does not rely on any specific assumptions about the underlying distribution. Furthermore, we also presented a goodness-of-fit test. Extensive testing and validation of our estimator were conducted using multiple simulated examples and a real-life application. The results clearly demonstrate that our estimator displays favorable and accurate performance.

Moreover, the applicability of our approach is not limited to varentropy and varextropy alone. It is possible to extend the results presented in this paper to study other dispersion indices. For instance, dispersion indices based on Kerridge inaccuracy measure and Kullback–Leibler divergence, as studied by [34], can be explored using a similar Bayesian nonparametric framework.

Author Contributions: Methodology, L.A.-L. and A.L.; Software, L.A.-L., A.L. and M.H.; Resources L.A.-L.; Writing—review & editing, L.A.-L., A.L. and M.H.; Supervision, L.A.-L. All authors have read and agreed to the published version of the manuscript.

Funding: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Data Availability Statement: There are no data associated with this paper.

Acknowledgments: We would like to express our sincere gratitude to the Editor and the anonymous referees for their valuable and constructive comments. Their insightful feedback and suggestions have greatly contributed to the improvement of this paper.

Conflicts of Interest: On behalf of all authors, there are no conflict of interest.

References

1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
2. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1991.
3. Kamavaram, S.; Goseva-Popstojanova, K. Entropy as a measure of uncertainty in software reliability. In Proceedings of the 13th International Symposium on Software Reliability Engineering, Annapolis, MD, USA, 12–15 November 2002; pp. 209–210.
4. Lad, F.; Sanfilippo, G.; Agro, G. Extropy: Complementary dual of entropy. *Stat. Sci.* **2015**, *30*, 40–58. [CrossRef]
5. Qiu, G.; Jia, K. Extropy estimators with applications in testing uniformity. *J. Nonparametr. Stat.* **2018**, *30*, 182–196. [CrossRef]
6. Al-Labadi, L.; Berry, S. Bayesian estimation of extropy and goodness of fit tests. *J. Appl. Stat.* **2020**, *49*, 357–370. [CrossRef]
7. Vasicek, O. A test for normality based on sample entropy. *J. R. Stat. Soc. B* **1976**, *38*, 54–59. [CrossRef]
8. Ebrahimi, N.; Habibullah, M.; Soofi, E.S. Testing exponentiality based on Kullback-Leibler information. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1992**, *54*, 739–748. [CrossRef]
9. Van Es, B. Estimating functionals related to a density by a class of statistics based on spacings. *Scand. J. Stat.* **1992**, *19*, 61–72.
10. Correa, J.C. A new estimator of entropy. *Commun. Stat.—Theory Methods* **1995**, *24*, 2439–2449. [CrossRef]
11. Wieczorkowski, R.; Grzegorzewski, P. Entropy estimators-improvements and comparisons. *Commun. Stat.—Simul. Comput.* **1999**, *28*, 541–567. [CrossRef]
12. Alizadeh Noughabi, H. A new estimator of entropy and its application in testing normality. *J. Stat. Comput. Simul.* **2010**, *80*, 1151–1162. [CrossRef]
13. Alizadeh Noughabi, H.; Arghami, N.R. A new estimator of entropy. *J. Iran. Stat. Soc.* **2010**, *9*, 53–64.
14. Al-Omari, A.I. A new measure of entropy of continuous random variable. *J. Stat. Theory Pract.* **2016**, *10*, 721–735. [CrossRef]
15. Beirlant, J.; Dudewicz, E.J.; Györia, L.; van der Meulen, E.C. Nonparametric entropy estimation: An overview. *Int. J. Math. Stat.* **1997**, *6*, 17–39.
16. Noughabia, H.A.; Jarrahiferizb, J. Extropy of order statistics applied to testing symmetry. *Commun. Stat.—Simul. Comput.* **2020**, *51*, 3389–3399. [CrossRef]
17. Källberg, D.; Seleznjev, O. Estimation of entropy-type integral functionals. *Commun. Stat.—Theory Methods* **2016**, *45*, 887–905. [CrossRef]
18. Mazzuchi, T.A.; Soofi, E.S.; Soyer, R. Bayes estimate and inference for entropy and information index of fit. *Econom. Rev.* **2008**, *27*, 428–456. [CrossRef]
19. Ferguson, T.S. A Bayesian analysis of some nonparametric problems. *Ann. Stat.* **1973**, *1*, 209–230. [CrossRef]
20. Al-Labadi, L.; Patel, V.; Vakiloroayaei, K.; Wan, C. A Bayesian nonparametric estimation to entropy. *Braz. J. Probab. Stat.* **2021**, *35*, 421–434. [CrossRef]
21. Al-Labadi, L.; Patel, V.; Vakiloroayaei, K.; Wan, C. Kullback-Leibler divergence for Bayesian nonparametric model checking. *J. Korean Stat. Soc.* **2021**, *50*, 272–289. [CrossRef]
22. Ishwaran, H.; Zarepour, M. Exact and approximate sum representations for the Dirichlet process. *Can. J. Stat.* **2002**, *30*, 269–283. [CrossRef]
23. Song, K.S. Rényi information, log likelihood and an intrinsic distribution measure. *J. Stat. Plan. Inference* **2001**, *93*, 51–69. [CrossRef]
24. Saha, S.; Kayal, S. Weighted (residual) varentropy with properties and applications. *arXiv* **2023**, arXiv:2305.00852.
25. Noughabi, H.A.; Noughabi, M.S. Varentropy estimators with applications in testing uniformity. *J. Stat. Comput. Simul.* **2023**, *93*, 2582–2599. [CrossRef]
26. Maadani, S.; Mohtashami Borzadaran, G.R.; Rezaei Roknabadi, A.H. A new generalized varentropy and its properties. *Ural. Math. J.* **2020**, *6*, 114–129. [CrossRef]
27. Sharma, A.; Kundu, C. Varentropy of doubly truncated random variable. *Probab. Eng. Inf. Sci.* **2022**, *7*, 852–871. [CrossRef]
28. Vaselabadi, N.M.; Tahmasebi, S.; Kazemi, M.R.; Buono, F. Results on varextropy measure of random variables. *Entropy* **2021**, *23*, 356. [CrossRef]
29. Grzegorzewski, P.; Wieczorkowski, R. Entropy-based goodness-of-fit test for exponentiality. *Commun. Stat.—Theory Methods* **1999**, *28*, 1183–1202. [CrossRef]
30. D'Agostino, R.B.; Stephens, M.A. *Goodness-of-Fit Techniques*; Marcel Dekker: New York, NY, USA, 1986.
31. Stephens, M.A. EDF statistics for goodness of fit and some comparisons. *J. Am. Stat. Assoc.* **1974**, *69*, 730–737. [CrossRef]
32. Grubbs, F.E. Fiducial bounds on reliability for the two-parameter negative exponential distribution. *Technometrics* **1971**, *13*, 873–876. [CrossRef]

33. Ebrahimi, N.; Pflughoeft, K.; Soofi, E.S. Two measures of sample entropy. *Stat. Probab. Lett.* **1994**, *20*, 225–234. [CrossRef]
34. Balakrishnan, N.; Buono, F.; Calì, C.; Longobardi, M. Dispersion indices based on Kerridge inaccuracy measure and Kullback-Leibler divergence. *Commun. Stat.—Theory Methods* **2023**. [CrossRef]