Statistics and Applications {ISSN 2454-7395 (online)}
Volume 23, No. 1, 2025 (New Series), pp 321–334
https://www.ssca.org.in/journal

Nonparametric Estimation of Extropy-Related Measures with Length-Biased Data

R. Dhanya Nair[1] and E. I. Abdul Sathar[2]
[1]*Department of Statistics, University College, Thiruvananthapuram - 695034, India*
[2]*Department of Statistics, University of Kerala, Thiruvananthapuram - 695581, India*

Received: 09 May 2024; Revised: 20 June 2024; Accepted: 30 June 2024

Abstract

Nonparametric estimators for extropy-related measures using length-biased data are proposed in this paper. The proposed estimators exhibit desirable properties, including consistency and asymptotic normality, which have been established. Furthermore, the precision of these estimators is assessed through the utilization of both simulated and real data sets, thereby validating their effectiveness in practical scenarios.

Key words: Extropy; Length-biased; Kernel estimator; Nonparametric estimation.

AMS Subject Classifications: 62N05, 62N02, 62G05.

1. Introduction

Length-biased sampling is a widely used technique for collecting lifetime data, primarily due to its cost-effectiveness and convenience. Unlike random sampling, length-biased sampling selects observations from the population of interest with probability proportional to their length. This approach finds significant applications in survival analysis, particularly when the onset time of diseases is unknown. In such scenarios, individuals who survive longer are more likely to be included in the sample, resulting in length-biased survival data. The phenomenon of length-bias was first noticed by Wicksell (1925) while investigating cell samples under a microscope. In his research, he noticed that only the cells that were larger than a particular size were visible in the microscope, leading to the study of a length-biased sample of cells. However, there are many other applications of length-biased data that make it crucial to understand the properties of this type of data. For instance, length-biased data arise in the study of diverse phenomena, such as ageing, epidemiology, and genetics. Therefore, exploring various aspects of length-biased data is essential for researchers and practitioners in fields such as medical research, public health, and social sciences.

Consider a random variable X with a probability density function (pdf), distribution function, and survival function denoted by f, F, and $\bar{F}$, respectively. Suppose a sample of size n is drawn from this population using a length-biased sampling technique, where the probability of including an observation in the sample is proportional to its size, volume,

Corresponding Author: E.I. Abdul Sathar
Email: sathare@gmail.com

length, or survival time. In other words, observations that are larger, longer, or have a longer survival time have a higher probability of being sampled. The resulting sample is length-biased, and the observed sample can be regarded as drawn from a distribution with pdf g given by

$$g(y) = \frac{y\,f(y)}{\mu},\ \ y \geq 0 \ and\ \mu \text{ is the mean of the population.} \tag{1}$$

One crucial problem is the nonparametric estimation of functionals of the distribution function F or pdf f based on a length-biased sample $Y_i, 1 \leq i \leq n$. This paper aims to propose nonparametric estimators for extropy-related measures of a population, using a length-biased sample drawn from it. Furthermore, the properties of these estimators are thoroughly investigated. The proposed estimators are useful in various fields, such as information theory, economics, and statistical physics, where the analysis of length-biased data is required. The study of these estimators' properties can aid in better understanding and utilizing length-biased data in practical applications.

Shannon's entropy, introduced by Shannon (1948), is one of the most widely used measures for assessing the uncertainty associated with a random variable. For a discrete random variable X taking values $\{x_1, x_2, x_3, ..., x_N\}$ with probability mass function (pmf) $\mathbf{p} = (p_1, p_2, p_3, ..., p_N)$, such that $N > 1$ is finite, Shannon's entropy is defined as

$$H(X) = -\sum_{i=1}^{N} p_i \log p_i. \tag{2}$$

Because equation (2) can be rewritten as $\mathcal{H}(\mathbf{p}) = E(-log\,\mathbf{p})$, the discrete entropy $\mathcal{H}(\mathbf{p})$ can be thought of as quantifying the average information content of X. That is, the entropy of a probability distribution is just the expected value of the information in the distribution. The entropy measure has far-reaching applications in many areas such as financial analysis, data compression, statistics and information theory. Lad *et al.* (2012) observed that the entropy measure on its own do not provide complete summary of the information in a distribution. This observation was substantiated in the context of its application in the logarithmic scoring rule, widely considered to be an eminent proper scoring rule used extensively for assessing and comparing sequential forecast distributions. The expected logarithmic score of a pmf $\mathbf{p}$ is in fact $-H(X)$, called negentropy. Lad pointed out that the logarithmic scoring function provides an incomplete assessment as it is a function only of the actual observation value of a quantity, ignoring other possible but unobserved values. To address this issue, a complementary scoring rule needs to be monitored concomitantly with the log score and this led to the expanded version of the logarithmic score, termed as the total log score. As a pair, the two complementary scores constitute the total logarithmic score and both components of the total log score are relevant to the assessment of forecasting distribution. Moreover, the expectation of the total log score equals the negentropy plus the negextropy of the distribution, where negextropy is the negative of a measure of a probability distribution suggested to be called as the extropy of the distribution by Lad *et al.* (2015) and is defined as follows.

For a discrete random variable X, the complementary dual of entropy, called extropy is defined as

$$J(X) = -\sum_{i=1}^{N} (1-p_i)\log(1-p_i).$$

The complementary of H and J arises from the fact that

$$J(\mathbf{p}) = (N-1)\left[H(\mathbf{q}) - \log(N-1)\right].$$

That is, the extropy of a pmf $\mathbf{p} = (p_1, p_2, p_3, ..., p_N)$ equals a location and scale transform of the entropy of another pmf $\mathbf{q} = \left(\frac{1-p_1}{N-1}, \frac{1-p_2}{N-1}, \frac{1-p_3}{N-1}, ..., \frac{1-p_N}{N-1}\right)$. The duality of entropy/extropy is a formal mathematical property of the pair of functions. For more details, one may refer Lad *et al.* (2015) and Lad *et al.* (2018).

As in entropy, extropy is interpreted as a measure of the amount of uncertainty represented by the distribution for X. Both entropy and extropy share many properties. They are invariant with respect to permutations of their mass functions and with respect to monotonic transformations. Moreover, the maximum extropy distribution is the uniform distribution and extropy satisfies Shannon's first and second axioms. As to differences in the two measures, the scale of the maximum entropy measure is unbounded as N increases while the scale of the maximum extropy is bounded by 1. It is evident that when $N = 2$, the entropy and extropy are identical. However, when $N > 2$, the measure bifurcates to yield distinct paired measurements $(H(X), J(X))$. As companions, these two measures relate as do the positive and negative images of a photographic film and they contribute together to characterizing the information in a distribution in much the same way. When the entropy is calculated for any assemblage such as the heat distribution for a galaxy of stars, a companion calculation of the extropy would allow us to complete our understanding of the variation inherent in its empirical distribution. An axiomatic characterization and several intriguing properties of this new measure was considered by Lad *et al.* (2015) and the results provided links to other notable information functions whose relation to entropy have not been recognized.

In the continuous context, a natural analog of discrete Shannon entropy for a probability density function f is called differential entropy and is defined as

$$H(X) = -\int_0^\infty f(x)\,\log\,f(x)\,dx.$$

The definition of differential entropy appears to be a natural extension of the Shannon entropy for discrete variables, defined in equation (2), to continuous variables. However, Shannon's differential entropy measure for a continuous density is actually derived from the limit of a linear translation of the discrete entropy measure. In order to define extropy for a continuous density, Lad *et al.* (2015) used the same procedure as the one followed by Shannon in defining differential entropy. Lad *et al.* (2015) noted that when the range of possibilities for X increases (as a result of larger N), the extropy measure $-\sum_{i=1}^{N}(1-p_i)\log(1-p_i)$ can be closely approximated by $1 - \frac{1}{2}\sum_{i=1}^{N} p_i^2$, which led to the definition of differential extropy.

Extropy of a non-negative absolutely continuous random variable X with pdf $f(x)$ is defined as

$$J(X) = -\frac{1}{2}\int_{0}^{\infty} f^2(x)\ dx = -\frac{1}{2}E(f(X)). \tag{3}$$

Here E denotes the expected value operator.

Differential entropy and extropy are obtained as the limit of a linear transformation of their corresponding discrete measures. The dual complementarity of extropy with entropy for continuous densities is derived in the context of relative entropy, also known as Kullback-Leibler divergence.

Through various illustrations Lad *et al.* (2012) showed that the extropies of the distributions do appear to provide interpretable complementary understandings of the character of distributions, already well-known to be summarised in a different dimension by their entropies.The total log score for densities is also better identified with the bivariate measure (negentropy, negextropy). Extropy can also be used to compare the uncertainties of two random variables. If the extropy of X is less than that of another random variable Y, that is, $J(X) \leq J(Y)$, then X is said to have more uncertainty than Y. By simultaneously considering entropy and extropy measures, researchers and practitioners can gain a more comprehensive understanding of the information and uncertainty within a given distribution. This broader perspective enables better-informed decision-making and more efficient utilization of statistical models in a range of applications. For further studies on extropy, one may also refer Noughabi and Jarrahiferiz (2019), Tahmasebi and Toomaj (2020), Buono *et al.* (2023) and Sathar and Nair (2024).

Additionally, to capture the uncertainty of a random variable which has already survived for some time, Qiu and Jia (2018) suggested the measure residual extropy. The residual extropy, denoted as $J(X;t)$, is defined as

$$J(X;t) = -\frac{1}{2\ (1-F(t))^2}\int_t^{\infty} f^2(x)dx \tag{4}$$

Furthermore, Krishnan *et al.* (2020) introduced a measure called past extropy, which computes the uncertainty associated with the past lifetime of a component that failed before a specific time. Past extropy of a random life time X is of course the extropy of the random variable $[X|X \leq t]$ and is given by

$$\bar{J}(X;t) = -\frac{1}{2\ F(t)^2}\int_0^{t} f^2(x)dx. \tag{5}$$

For a non-negative rv X having a survival function $\bar{F}$, an alternative measure of extropy based on the survival function of a rv called survival extropy (SE) has been proposed by Sathar and Nair (2021) which is defined as

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

The survival extropy of the random variable $[X - t|X \geq t]$ called dynamic survival extropy (DSE), was also considered by Sathar and Nair (2021) and is defined as

$$J_s(X;t) = -\frac{1}{2}\int_t^\infty \frac{\bar{F}^2(x)}{\bar{F}^2(t)}dx = -\frac{1}{2}\int_t^\infty \frac{(1-F(x))^2}{(1-F(t))^2}dx. \tag{6}$$

It is worth noting that the SE and DSE have a close relationship with well-known economic measures such as the Gini index and statistical quantities including L-moments. These connections have been extensively studied by Nair and Sathar (2022) and Nair and Sathar (2023). These insights further contribute to the interpretation and application of the SE and DSE measures, offering valuable connections to economic analysis and statistical modeling.

These alternative measures of extropy, namely residual extropy, past extropy, and survival extropy, complement Shannon's entropy and offer additional perspectives on the uncertainty and information content of a random variable. These measures find applications in various fields, including reliability analysis, survival modeling, risk assessment, economics, finance, and actuarial science, where the analysis of time-dependent uncertainty is of paramount importance. By utilizing these measures, researchers and practitioners can gain deeper insights into the temporal aspects and survival behavior of random variables in practical scenarios. In this study, we introduce nonparametric estimators for extropy related measures of the population based on a length-biased data drawn from it. Length-biased sampling has proven to be valuable in various fields, and in Section 2, we present our proposed estimator for dynamic survival extropy (DSE). We also examine the asymptotic properties of the proposed estimator to ensure its reliability. Furthermore, in Section 3, we discuss the nonparametric estimation of residual and past extropy, and analyze their asymptotic properties. Finally, in Section 4, a simulated study and real-data analysis have been carried out to illustrate the precision of the estimators. By employing these empirical investigations, we showcase the accuracy and effectiveness of the estimators in practical settings. This empirical validation adds credibility to the proposed methodology and confirms its utility in real-world scenarios.

2. Nonparametric estimation of DSE using length-biased sample

This section proposes a nonparametric estimator for the DSE of a random variable X using a length-biased sample of size n drawn from X. Due to the use of a probability proportional to size (PPS) sampling scheme, the observed sample $Y_1, Y_2, Y_3, ..., Y_n$ cannot be treated as independent and identically distributed (iid) samples from X. Consequently, existing estimators of extropy measures based on a random sample from the population cannot be applied. Instead, a different estimator suitable for length-biased data needs to be considered. To this end, it is worth noting that the observed length-biased sample can be regarded as iid observations from the distribution of a random variable Y with a pdf $g(y)$ given by equation (1). Building upon this insight, Cox (1969) proposed an empirical estimator for the distribution function $F(x)$ in the length-biased setup. The estimator is given by

$$F_n(x) = \frac{\sum_{i=1}^{n} Y_i^{-1} I(Y_i \le x)}{\sum_{i=1}^{n} Y_i^{-1}}, \tag{7}$$

where $I(.)$ is the indicator random variable of the event specified in parentheses. It has been demonstrated by Chaubey *et al.* (2010) that as $n \to \infty$, the empirical estimator $F_n(x)$ converges almost surely to the true distribution function $F(x)$, as shown in equation (8). Furthermore, the estimator converges in distribution to a normal distribution, as expressed in equation (9).

$$\sup_{x \in R+} |F_n(x) - F(x)| \xrightarrow{a.s} 0, \text{ as} \, n \to \infty. \tag{8}$$

and

$$\sqrt{n}\,(F_n(x) - F(x)) \xrightarrow{D} N(0, \delta^2(x)), \tag{9}$$

where $\delta^2(x) = \mu \left\{ \int_0^x t^{-1} f(t)dt - 2F(x) \int_0^x t^{-1} f(t)dt + F^2(x) \int_0^\infty t^{-1} f(t)dt \right\}$.

Also, as $n \to \infty$,

$$E(F_n(x)) = F(x) \text{ and } Var(F_n(x)) = \frac{\delta^2(x)}{n}. \tag{10}$$

Therefore, we can obtain a nonparametric estimator of DSE of X by substituting the estimator given in equation (7) into equation (6). The resulting estimator for DSE is given by

$$\hat{J}_s(X;t) \;= -\frac{1}{2} \int_t^\infty \frac{(1 - F_n(x))^2}{(1 - F_n(t))^2} dx. \tag{11}$$

Now let's examine the asymptotic properties of the proposed estimator. For simplifying the notation, we define the following terms:

$$a_n(t) \;=\; \int_t^\infty \bar{F}_n^2(x)\,dx,\;\; m_n(t) = \bar{F}_n^2(t),\;\; a(t) \;=\; \int_t^\infty \bar{F}^2(x)\,dx \text{ and } m(t) = \bar{F}^2(t).$$

Thus, the estimator $\hat{J}_s(X;t)$ can be expressed as

$$\hat{J}_s(X;t) = -\frac{1}{2}\frac{a_n(t)}{m_n(t)}, \text{ while the true DSE } J_s(X;t) \text{ is given by } J_s(X;t) = -\frac{1}{2}\frac{a(t)}{m(t)}.$$

Result 1:

$$\lim_{n \to \infty} |\hat{J}_s(X;t) - J_s(X;t)| = 0 \; a.s.$$

Moreover, mean square error (MSE) of $\hat{J}_s(X;t)$ tends to 0 as $n \to \infty$.

Proof: Using Taylor series expansion,

$$\bar{F}_n^2(t) = \bar{F}^2(t) + \left(\bar{F}_n(t) - \bar{F}(t)\right) 2\bar{F}(t) + o\left(\bar{F}_n(t) - \bar{F}(t)\right)^2.$$

It follows that

$$m_n(t) - m(t) = \left(\bar{F}_n(t) - \bar{F}(t)\right) 2\bar{F}(t) + o\left(\bar{F}_n(t) - \bar{F}(t)\right)^2.$$

Similarly, we obtain

$$a_n(t) - a(t) \simeq 2\int_t^{\infty} \bar{F}(x)\ \left(\bar{F}_n(x) - \bar{F}(x)\right) dx.$$

Now,

$$\frac{a_n(t)}{m_n(t)} - \frac{a(t)}{m(t)} \simeq \frac{m(t)\,[a_n(t) - a(t)] - a(t)\,[m_n(t) - m(t)]}{m^2(t)}.$$

Hence,
$\hat{J}_s(X;t) - J_s(X;t)$

$$\simeq -\frac{1}{m(t)}\int_t^{\infty} \bar{F}(x)\left(\bar{F}_n(x) - \bar{F}(x)\right)\, dx\, + \frac{a(t)}{m^2(t)}\left(\bar{F}_n(t) - \bar{F}(t)\right)\bar{F}(t). \tag{12}$$

By using the almost sure convergence of $F_n(x)$ given in equation (8), we obtain

$$\lim_{n\to\infty} |\hat{J}_s(X;t) - J_s(X;t)| = 0\ a.s.$$

Additionally, from equations (12) and (10), it can be easily seen that the bias and variance of $\hat{J}_s(X;t)$ tends to 0 as $n \to \infty$. Hence, as $n \to \infty$, MSE of $\hat{J}_s(X;t) \to 0$.

□

Next, we discuss the asymptotic normality of our estimator.

Result 2: $\hat{J}_s(X;t) - J_s(X;t)$ is asymptotically normal with mean 0 and variance

$$\frac{1}{n\,\bar{F}^4(t)}\left[\int_t^{\infty} \bar{F}^2(x)\,\delta^2(x)dx + \frac{a^2(t)\delta^2(t)}{\bar{F}^2(t)}\right].$$

Proof: Using equation (10), as $n \to \infty$,

$$E(\bar{F}_n(x) - \bar{F}(x)) = 0 \text{ and } Var(\bar{F}_n(x)) = \tfrac{\delta^2(x)}{n}.$$

Hence, from equation (12), we obtain the following.

As $n \to \infty$, $E(\hat{J}_s(X;t) - J_s(X;t)) = 0$ and

$$Var(\hat{J}_s(X;t) - J_s(X;t)) = \tfrac{1}{n\,\bar{F}^4(t)}\left[\int_t^{\infty} \bar{F}^2(x)\,\delta^2(x)dx + \tfrac{a^2(t)\delta^2(t)}{\bar{F}^2(t)}\right].$$

Again, using equation (9), we have,

$$\sqrt{n}\,(\bar{F}_n(x) - \bar{F}(x)) \xrightarrow{D} N(0, \delta^2(x))$$

Hence from equation (12), it follows that $\hat{J}_s(X;t) - J_s(X;t)$ is asymptotically normal. This completes the proof. □

In a similar manner, a nonparametric estimator for dynamic failure extropy (DFE) proposed by Nair and Sathar (2020) can be obtained.
The DFE of X is defined as

$$J_f(X;t) = -\frac{1}{2}\int_0^t \frac{F^2(x)}{F^2(t)}\,dx.$$

By plugging in the estimator given by equation (7) into the above equation, we can obtain the nonparametric estimator of DFE under length-biased setup, which is as follows.

$$\hat{J}_f(X;t) = -\frac{1}{2}\int_0^t \frac{F_n^2(x)}{F_n^2(t)}\,dx. \tag{13}$$

Consistency and asymptotic normality of this estimator can be proved by proceeding in a similar manner as in Result 1 and 2.

3. Nonparametric estimation of residual and past extropies for length-biased sample

In this section, we focus on the nonparametric estimation of residual and past extropies defined by equations (4) and (5), respectively. To obtain the estimators of residual and past extropies using length-biased data, we utilize equation (7) and the kernel density estimator proposed by Jones (1991). By smoothing the estimator given in equation (7), Jones (1991) derived a new kernel density estimator given by

$$\hat{f}(x) = \frac{\sum_{i=1}^{n} \frac{1}{Y_i\,h}\,k\left(\frac{x-Y_i}{h}\right)}{\sum_{i=1}^{n} Y_i^{-1}}, \tag{14}$$

where k is the kernel function and $h = h_n$ is the band-width. The bias, variance and asymptotic properties of this estimator was obtained by Guillamon *et al.* (1998) as follows.

$$Bias(\hat{f}(x)) = \frac{1}{2}\,h^2\,\mu_2(k)\,f''(x) + o(h^2) \text{ and } Var(\hat{f}(x)) = \frac{1}{n\,h}\,\mu\,x^{-1}\,f(x)\,C_k + o\left(\frac{1}{nh}\right), \tag{15}$$

where $\mu_2(k) = \int_{-\infty}^{\infty} u^2 k(u)\,du$, $C_k = \int_{-\infty}^{\infty} k^2(u)\,du$ and $f''(x)$ is the 2^{nd} derivative of f with respect to x.

Also,

$$\sqrt{n\,h}\,(\hat{f}(x) - f(x)) \xrightarrow{D} N(0, \mu\, x^{-1}\, f(x)\, C_k). \tag{16}$$

Now, we propose a nonparametric estimator of residual extropy under length-biased set up.

Definition 1: A nonparametric kernel estimator for $J(X;t)$ shall be defined as

$$\hat{J}(X;t) = -\frac{1}{2}\left[\frac{\int_t^\infty \hat{f}^2(x)dx}{(1-F_n(t))^2}\right]. \tag{17}$$

In order to simplify the notations, define

$$p_n(t) = \int_t^\infty \hat{f}^2(x)dx, \quad p(t) = \int_t^\infty f^2(x)dx,$$

so that equation (17) can be written as $\hat{J}(X;t) = -\frac{1}{2}\left[\frac{p_n(t)}{m_n(t)}\right]$.

By using Taylor's series expansion, we get

$$p_n(t) - p(t) = 2\int_t^\infty f(x)\ \left(\hat{f}(x) - f(x)\right) dx + o\Big(\hat{f}(x) - f(x)\Big)^2.$$

Proceeding in a similar manner as in Section 2, we obtain

$$\begin{aligned}
&\hat{J}(X;t) - J(X;t)\\
&\qquad \simeq -\frac{1}{m(t)}\int_t^\infty f(x)\left(\hat{f}(x) - f(x)\right)\, dx\ + \frac{p(t)}{m^2(t)}\left(\bar{F}_n(t) - \bar{F}(t)\right)\bar{F}(t).
\end{aligned}$$

The asymptotic normality of $\hat{J}(X;t)$ can now be easily obtained on using equations (16) and (9). Furthermore, using equation (15), we observe that the MSE of $\hat{J}(X;t)$ tends to 0 as $n \to \infty$, and thus the estimator $\hat{J}(X;t)$ is strongly consistent.
Similarly, a consistent and asymptotically normal nonparametric estimator for $\bar{J}(X;t)$ under length-biased set up shall be defined as

$$\hat{\bar{J}}(X;t) = -\frac{1}{2\ F_n^2(t)}\int_0^t \hat{f}^2(x)dx.$$

4. Data analysis

To demonstrate the accuracy of the presented nonparametric estimators, we first apply the proposed methods to the simulated data sets. We generate length-biased samples from beta distribution of first kind with parameters $\alpha = 2$ and $\gamma = 4$. The bias and MSE of the suggested estimators of DSE and DFE given by equations (11) and (13) respectively, are computed for certain values of t, and the results obtained are presented in Tables 1 and 2. It can be observed from the tables that the bias and MSE are negligible. This indicates that the estimators perform well in accurately capturing the extropy measures. Figures 1 and 2 display plots of the actual and estimated values of DSE and DFE of the population for simulated data. Both graphs clearly show that the estimated values closely align with the actual values. Notably, even with a sample size of $n = 30$, the estimated values for DFE are very close to the actual values, highlighting the effectiveness of the proposed estimators. Furthermore, we computed the theoretical and estimated values of residual and past extropies using the Gaussian kernel function. These values, along with the bias and MSE, are presented in Tables 3 and 4. The results from these tables indicate that the estimators of residual and past extropies also perform well, further validating the reliability of the proposed nonparametric estimators. Overall, the results obtained from the simulations demonstrate the precision and accuracy of the nonparametric estimators proposed in this study.

Table 1: Bias and MSE of the estimator of DSE for simulated data

	$n = 50$		$n = 100$	
t	Bias	MSE	Bias	MSE
0.4	0.00147	0.00009	-0.00180	0.00004
0.5	-0.00269	0.00007	0.00029	0.00002
0.6	0.00009	0.00005	-0.00363	0.00003
0.7	-0.00961	0.00031	-0.00626	0.00018

Table 2: Bias and MSE of the estimator of DFE for simulated data

	$n = 50$		$n = 100$	
t	Bias	MSE	Bias	MSE
0.4	-0.00501	0.00038	-0.00029	0.00004
0.5	0.00239	0.00021	-0.00113	0.00005
0.6	0.00388	0.00011	-0.00173	0.00011
0.7	0.00044	0.00047	-0.00379	0.00014

Table 3: Theoretical and estimated values of residual extropy together with its bias and MSE for simulated data

		$n = 50$			$n = 100$		
t	Theory	Estimate	Bias	MSE	Estimate	Bias	MSE
0.4	-1.62017	-1.71256	-0.09239	0.08763	-1.69013	-0.06996	0.01232
0.5	-2.02822	-2.11318	-0.08496	0.09544	-2.16395	-0.13573	0.00642
0.6	-2.62262	-2.72248	-0.09986	0.05238	-2.68369	-0.06107	0.01003
0.7	-3.59451	-3.65942	-0.06491	0.03416	-3.63571	-0.04120	0.00237

Next, we consider the empirical estimator of DSE and DFE, which were proposed by Sathar and Nair (2021) and Nair and Sathar (2020), respectively. These empirical estimators

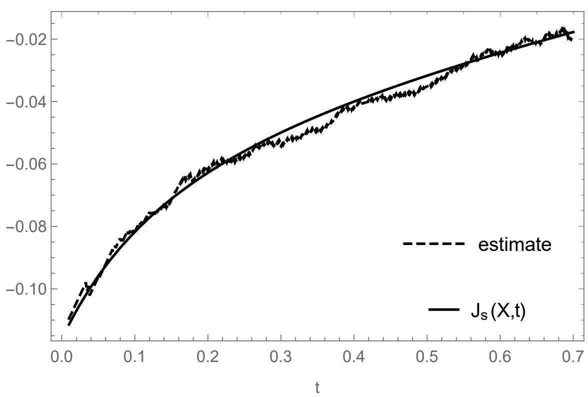

Figure 1: Plots of actual and estimated values of DSE using a simulated sample of size $n = 100$

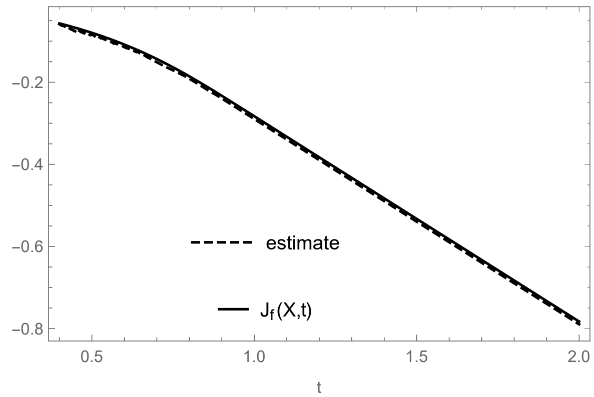

Figure 2: Plots of actual and estimated values of DFE using a simulated sample of size $n = 30$

Table 4: Theoretical and estimated values of past extropy together with its bias and MSE for simulated data

		$n = 50$			$n = 100$		
t	Theory	Estimate	Bias	MSE	Estimate	Bias	MSE
0.4	-1.38686	-1.35128	0.03558	0.00483	-1.38890	-0.00204	0.00311
0.5	-1.09420	-1.15662	-0.06242	0.00468	-1.10053	-0.00633	0.00265
0.6	-0.92836	-0.98423	-0.05587	0.00957	-0.94362	-0.01526	0.00403
0.7	-0.84123	-0.88562	-0.04439	0.00348	-0.84563	-0.00439	0.00114

are based on an iid sample from the population. The empirical dynamic survival extropy and dynamic failure extropy estimators are respectively as follows.

$$J_s(\hat{\bar{F}}_n;t) = -\frac{1}{2}\int_t^{\infty}\left[\frac{\hat{\bar{F}}_n(x)}{\hat{\bar{F}}_n(t)}\right]^2 dx \tag{18}$$

and

$$J_f(\hat{F}_n;t) = -\frac{1}{2}\int_0^{t}\left[\frac{\hat{F}_n(x)}{\hat{F}_n(t)}\right]^2 dx, \tag{19}$$

where $\hat{\bar{F}}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i > x)$, $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)$, with I being the indicator function.

To investigate the performance of the empirical estimators defined by equations (18) and (19) when applied to a length-biased sample, we compare the actual values of DSE and DFE of the population with the estimated values obtained using these estimators. The results are displayed in Figures 3 and 4. Analyzing Figures 1 to 4, we observe that the deviation between actual and estimated values is more when the empirical estimators are used instead of the proposed estimators. This suggests that the estimators defined by equations (18) and (19) are suitable when an iid sample is available from the population whereas for the length-biased sample, the estimators defined by equations (11) and (13) should be employed.

In summary, the comparison of the estimators highlights the importance of choosing the appropriate estimator based on the characteristics of the sample. The proposed nonparametric estimators are specifically tailored for length-biased data and demonstrate superior accuracy in estimating extropy measures when applied to length-biased samples, as evidenced by the smaller deviations between the actual and estimated values.

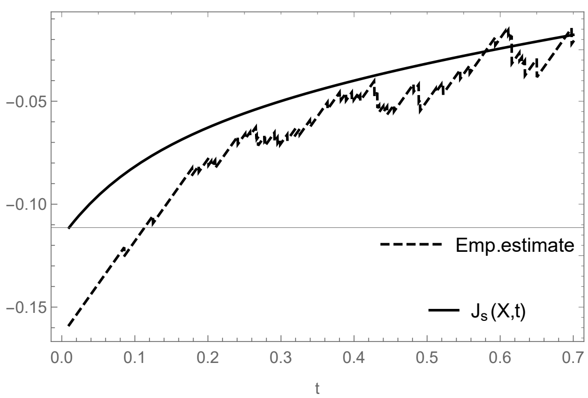

Figure 3: Plots of actual and estimated values of DSE using the empirical estimator for a simulated sample

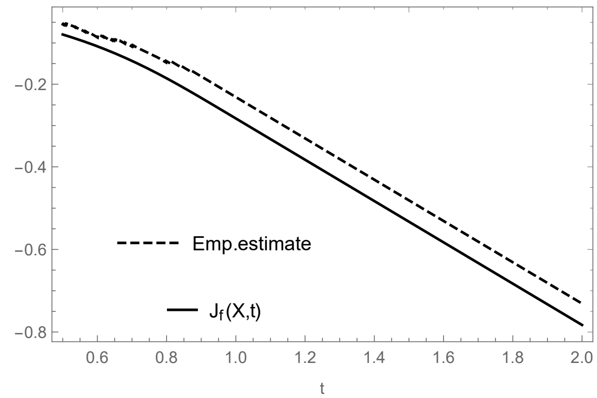

Figure 4: Plots of actual and estimated values of DFE using the empirical estimator for a simulated sample

To further assess the performance of the proposed estimators defined by equations (11) and (13), we apply them to a real-world scenario using a data that was previously investigated by Helu *et al.* (2020). The data set consists of 70 failure times of aircraft windshields, from which a sample of size 50 is drawn with probability proportional to size. The best-fitted distribution to the original data set is the Gamma distribution with parameters $\alpha = 7.75$ and $\beta = 0.285$. We plot the theoretical and estimated values of DSE and DFE for the real data in Figures 5 and 6, respectively. Upon analysis of the plots, we observe that the estimated values are remarkably close to the actual values. This indicates that the proposed estimators perform well in real-world circumstances. The accuracy of the estimators in estimating the extropy measures for the length-biased sample demonstrates their reliability and applicability in practical scenarios.

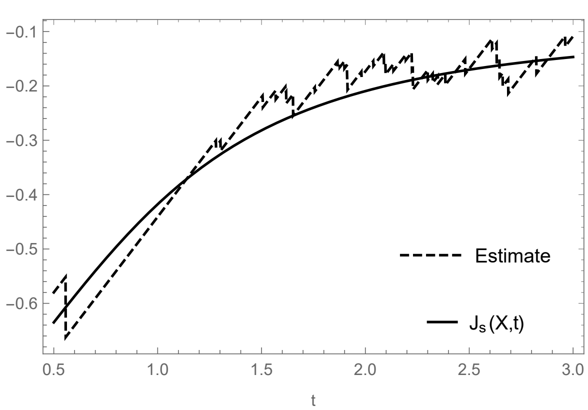

Figure 5: Plots of actual and estimated values of DSE for real data

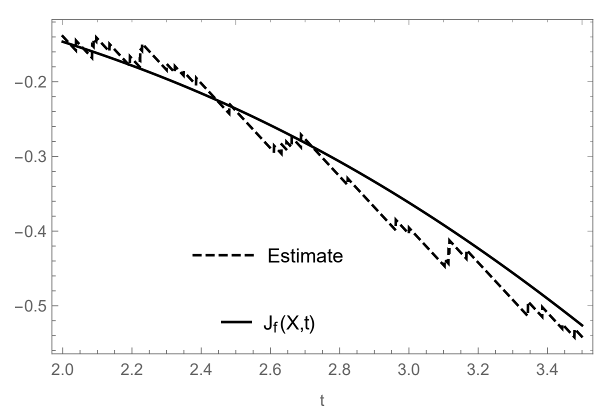

Figure 6: Plots of actual and estimated values of DFE for real data

5. Conclusions

This work proposes nonparametric estimators for extropy-related measures under length-biased sampling. The consistency and asymptotic normality of the proposed estimators are established, demonstrating their reliability in estimating these measures. The performance of the estimators is evaluated using both simulated and real data sets. The simulation results provide strong evidence of the accuracy and precision of the proposed estimators. The negligible bias and mean squared error observed in the estimators confirm their ability to closely approximate the true values of the extropy-related measures. Furthermore, the analysis of a real data set reinforces the practical utility of the proposed estimators. By evaluating the extropy-related measures using the real data, it is evident that the estimators perform well in real-life scenarios. This highlights the applicability of the estimators in various domains, such as reliability analysis, survival analysis, and engineering, where accurate estimation of extropy-related measures is crucial for making informed decisions and understanding complex systems.

In summary, this work contributes valuable nonparametric estimators for extropy-related measures under length-biased sampling. The established properties of consistency and asymptotic normality, coupled with the demonstrated accuracy in both simulated and real data settings, make these estimators highly reliable tools for researchers and practitioners. The availability of such estimators facilitates the estimation of extropy-related measures, enabling deeper insights into the dynamics of failure and survival processes in diverse fields of study.

Acknowledgements

The authors would like to express their gratitude to the editor-in-chief and the reviewers for their valuable feedback and constructive comments. Their insights and suggestions have played a significant role in enhancing the quality of this article.

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

Buono, F., Kamari, O., and Longobardi, M. (2023). Interval extropy and weighted interval extropy. *Ricerche di Matematica*, **72**, 283-298.

Chaubey, Y. P., Sen, P. K., and Li, J. (2010). Smooth density estimation for length-biased data. *Journal of the Indian Society of Agricultural Statistics*, **64**, 145–155.

Cox, D.R. (1969). Some sampling problems in technology. In *New Developments in Survey Sampling*, Wiley-Interscience, New-York. 506–527.

Guillamon, A., Navarro, J., and Ruiz, J. M. (1998). Kernel density estimation using weighted data. *Communications in Statistics - Theory and Methods*, **27**, 2123–2135.

Helu, A., Samawi, H., and Rochani, H. (2020). Kernel density estimation based on progressive type-II censoring. *Journal of the Korean Statistical Society*, **49**, 475–498.

Jones, M. C. (1991). Kernel density estimation for length-biased data. *Biometrika*, **78**, 511–519.

Krishnan, A. S., Sunoj, S. M., and Nair, N. U. (2020). Some reliability properties of extropy for residual and past lifetime random variables. *Journal of the Korean Statistical Society*, **49**, 457–474 .

Lad, F., Sanfilippo, G., and Agro, G. (2012). Completing the logarithmic scoring rule for assessing probability distributions. *AIP Conference Proceedings*, **1490**, 13-30.

Lad, F., Sanfilippo, G., and Agro, G. (2015). Extropy: complementary dual of entropy. *Statistical Science*, **30**, 40-58.

Lad, F., Sanfilippo, G., and Agro, G. (2018). The duality of entropy/extropy, and completion of the Kullback information complex. *Entropy*, **20**, 593.

Nair, R. D. and Sathar, E. I. A. (2020). On dynamic failure extropy. *Journal of the Indian Society for Probability and Statistics*, **21**, 287-313.

Nair, R. D. and Sathar, E. I. A. (2022). A study on some properties of dynamic survival extropy and its relation to economic measures. *Stochastic and Quality Control*, **37**, 65-74. doi:10.1515/eqc-2021-0050.

Nair, R. D. and Sathar, E. I. A. (2023). Some useful results related to various measures of extropy and their interrelationship. *Statistics and Probability Letters*, **193**, doi: 10.1016/j.spl.2022.109729.

Noughabi, H. A., and Jarrahiferiz, J. (2019). On the estimation of extropy. *Journal of Non-parametric Statistics*, **31**, 88-99.

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

Sathar, E. I. A. and Nair, R.D. (2021). On dynamic survival extropy. *Communications in Statistics-Theory and Methods*, **50**, 1295-1313.

Sathar, E. I. A. and Nair, R.D. (2024). Properties of extropy and its weighted version for doubly truncated random variables. *Ricerche di Matematica*, https://doi.org/10.1007/s11587-024-00863-8.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical*, **27**, 379-423 and 623-656.

Tahmasebi, S. and Toomaj, A. (2020). On negative cumulative extropy with applications. *Communications in Statistics - Theory and Methods*, **51**, 5025–5047.

Wicksell, S. D. (1925). The corpuscle problem: A mathematical study of a biometric problem. *Biometrika*, **17**, 84–99.