پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان

15th Seminar On Probability and Stochastic Processes
August 30-31, 2025
University of Kurdistan

Some Estimators for the Residual Varextropy Function of Length-biased Data

R. Zamini*

Department of Mathematics, Faculty of Mathematical Sciences and Computer, Kharazmi University, Tehran, Iran.

Abstract

In reliability theory and survival analysis, the residual extropy is proposed to measure the residual uncertainty of lifetimes. In this paper under length-biased sampling, we introduce two estimators for the variance of the residual lifetimes, "residual varextropy" in short. To assess the performance of the proposed estimators, we conduct a finite-sample simulation study in terms of bias and root mean squared error (RMSE). We illustrate the application of proposed estimators through real data analysis.

Keywords: Residual varextropy, Length-biased data, Kernel density.
Mathematics Subject Classification (2020): 62G05, 62G07, 94A17.

1 Introduction

Consider the non-negative continuous random variable X with the distribution function $F(\cdot)$ and the density function $f(\cdot)$. Shannon (1948) introduced the entropy associated with X as

$$H(x) = -\int_0^{+\infty} f(x)\log\left(f(x)\right)\,dx. \tag{1}$$

$H(X)$ has applications in various fields, including physics, biology and computer sciences. Lad and *et al.* (2015) introduced the dual of entropy, extropy, as:

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

Among other authors who have applied extropy in statistical contexts, we can refer to the work of Qiu (2017), Raqab and Qiu (2019) and Gneiting and Raftery (2007). Qiu and Jia (2018) defined a dynamic metric of extropy, called residual extropy, as follows:

$$J_t(X) = J(X_t) = -\frac{1}{2(\overline{F}(t))^2}\int_t^{+\infty}\left(f(x)\right)^2\,dx, \tag{3}$$

where $\overline{F}(t) = 1-F(t)$ is the survival function of X. $J_t(X)$ measures uncertainty in the residual lifetime $X_t = (X-t|X>t)$ with density function $f_t(x) = \frac{f(x+t)}{\overline{F}(t)}$, $x>0$. Vaselabadi and *et*

*Corresponding author, zamini@khu.ac.ir

al. (2021) introduced the variance of $-\frac{1}{2}f(X)$ as another uncertainty measure:

$$VJ(X) = Var\left(-\frac{1}{2}f(X)\right) = \frac{1}{4}\int (f(x))^3\,dx - \left(\frac{1}{2}\int (f(x))^2\,dx\right)^2. \tag{4}$$

$VJ(X)$ is called varextropy measure. This measure indicates how the information content is scattered around the extropy. Goodarzi (2022) investigated the relationship between series and parallel systems and symmetric distributions with the help of varextropy. Alizadeh Noughabi and Shafaei Noughabi (2024) proposed some estimators for varextropy. Zamini and *et al.* (2025) proposed some non-parametric estimators for varextropy, residual and past varextropy under $\alpha-$mixing dependent condition. Vaselabadi and *et al.* (2021) introduced variance of $-\frac{1}{2}f(X_t)$ as the residual varextropy:

$$VJ_t(X) = VJ(X_t) = \frac{1}{4(\overline{F}(t))^3}\int_t^{+\infty} (f(x))^3\,dx - \left(\frac{1}{2(\overline{F}(t))^2}\int_t^{+\infty} (f(x))^2\,dx\right)^2. \tag{5}$$

This measure indicates how the information content is scattered around the residual extropy. In this paper, we study some estimators for $VJ_t(X)$ under length-biased data.

If observations are selected from the interest population with probability proportional to their length, rather than randomly, we will encounter length-biased data. Examples of such data can be found in studies related to cancer screening trails, as explained by Zelen and Feinleib (1969), biomedicine (Chakraborty and Rao (2000)), astronomy, renewal processes and econometrics.

Let X denote a non-negative random variable with density function $f(\cdot)$ and distribution function $F(\cdot)$. The associated length-biased random variable Y, characterized by the distribution function $G(\cdot)$, is as follows:

$$G(y) = \frac{1}{\mu}\int_0^y xf(x)dx,\ y \geq 0, \tag{6}$$

where $\mu = \int_0^\tau xf(x)dx < \infty$ is its expected value and $\tau = \sup\{x; F(x) < 1\} < \infty$.

Under length-biased sampling, some kernel estimators for extropy and residual extropy obtained by Rajesh and *et al.* (2022) and Pavithradas and *et al.* (2025), respectively. Zamini and *et al.* (2023) presented non-parametric estimators for varextropy under length-biased sampling. They investigated the consistency and asymptotic normality of the proposed estimators under suitable regularity conditions. In this paper, we address the problem of kernel estimators of residual varextropy function under length-biased sampling.

The rest of the paper is organized as follows. In Section 2, we propose two kernel estimators of the residual varextropy for length-biased data and present the main theorems. We investigate the finite sample behaviour of the proposed estimators and a real dataset of length-biased lifetimes in Section 3. The proof of the main results are given in Section 4. We give briefly the conclusions of this study and suggestions for the future application in Section 5.

2 Main Results

In this section we propose two kernel estimators for the residual varextropy under length-biased scheme and investigate some of their asymptotic properties. In the following, we list some necessary assumptions in this paper.

Assumptions

A1 The symmetric density function K is bounded variation on $(-1, 1)$.

A2 $\int_{-1}^{1} tK(t)\,dt = 0$.

A3 $\int_{-1}^{1} t^2K(t)\,dt < \infty$.

A4 $\int_{-1}^{+1} |dK(t)| < \infty$.

A5 $\int_0^{+\infty} u^{-2}G^{1/r}(u)du < \infty$, for some $r > 2$.

A6 $f(t) = O(t) \quad as \quad t \to 0.$

Let X be the interest random variable with density and distribution functions $f(\cdot)$ and $F(\cdot)$ respectively. The associated length-biased random variable Y has the distribution function G. Cox (1969) proposed the following empirical estimator for F:

$$F_n(t) = \frac{\mu_n}{n} \sum_{i=1}^{n} \frac{1}{Y_i} I\left(Y_i \leq t\right), \tag{7}$$

where $Y_1, \ldots, Y_n$ is a sample of independent and identically distributed random variables from G and $\mu_n^{-1} = \frac{1}{n}\sum_{i=1}^{n} Y_i^{-1}$. Chaubey and *et al.* (2010) proved that:

$$\lim_{n\to\infty} \sup_{0<x<\infty} |F_n(x) - F(x)| = 0, \ a.s. \tag{8}$$

Jones (1991) proposed the following kernel estimator for f based on F_n as follows:

$$f_n(x) = \frac{\mu_n}{nh_n} \sum_{i=1}^{n} \frac{1}{Y_i} K\left(\frac{x - Y_i}{h_n}\right), \tag{9}$$

where $\{h_n\}_{n\geq 1}$ is a bandwidth sequence, satisfying the conditions $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} nh_n = \infty$. Under some conditions, Ajami and *et al.* (2013) showed that

$$\lim_{n\to\infty} \sup_{0<x<\tau} |f_n(x) - f(x)| = 0, \ a.s. \tag{10}$$

Based on f_n and F_n we propose an estimator for $VJ_t(X)$ as follows:

$$VJ_t^1(X) = \frac{1}{4(\overline{F_n}(t))^3} \int_t^{+\infty} \left(\tilde{f}_n(x)\right)^2 dF_n(x) - \left(J_n^1(f,t)\right)^2, \tag{11}$$

where

$$\tilde{f}_n(Y_i) = \frac{\mu_{n-1}}{(n-1)h_{n-1}} \sum_{j=1, j\neq i}^{n} \frac{1}{Y_j} K\left(\frac{Y_i - Y_j}{h_{n-1}}\right), \tag{12}$$

and

$$J_n^1(f,t) = -\frac{\mu_n}{2n(\overline{F_n}(t))^2} \sum_{i=1}^{n} {Y_i}^{-1} \tilde{f}_n(Y_i) I(Y_i \geq t), \tag{13}$$

is the first estimator of $J_t(X)$ proposed by Pavithradas and *et al.* (2025).

We obtain another estimator for $VJ_t(X)$ as follows:

$$VJ_t^2(X) = \frac{1}{4(\overline{F_n}(t))^3} \int_t^{+\infty} \left(f_n(x)\right)^3 dx - \left(J_n^2(f,t)\right)^2, \tag{14}$$

where

$$J_n^2(f,t) = -\frac{1}{2(\overline{F}_n(t))^2} \int_t^{\infty} (f_n(x))^2 dx, \tag{15}$$

is the second estimator of $J_t(X)$ proposed by Pavithradas and *et al.* (2025).

In the following theorems, consistency of the $VJ_t^1(X)$ and $VJ_t^2(X)$ are investigated.

Theorem 2.1. *Let* $\lim_{n\to\infty} \frac{\log n}{n^{\frac{1}{2}+\lambda} h_n} = 0$ *for any* $0 < \lambda < \frac{1}{2} - \frac{1}{r}$ *and some* $r > 2$. *If* f *is bounded, then under Assumptions* $A_1 - A_4$ *we can write:*

$$\lim_{n\to\infty} VJ_t^1(X) = VJ_t(X), \ a.s. \tag{16}$$

Proof. See the Apendix □

Theorem 2.2. *Under conditions of Theorem 2.1, we can write:*

$$\lim_{n\to\infty} VJ_t^2(X) = VJ_t(X), \ a.s. \tag{17}$$

Proof. See the Apendix □

3 Simulation Results

To demonstrate the performance of the proposed estimators, we conduct simulation studies in two following scenarios.
Scenario 1. Let X be a random variable having a beta distribution with density function:

$$f(x) = 6x(1-x),\ 0 < x < 1.$$

Scenario 2. Let X be a random variable having a gamma distribution with density function:

$$f(x) = xe^{-x},\ 0 < x < \infty.$$

We use the notations $B(\alpha, \beta)$ and $G(\alpha, \beta)$ to indicate X has the beta and gamma distribution, respectively. We note that if X has $B(2,2)$, then the corresponding length-biased form is $B(3,2)$ and if X has $G(2,1)$, then the corresponding length-biased form is $G(3,1)$. We use 500 samples of various sample sizes $n = 50, 100, 500$ and the Epanechnikov density function $K(x) = \frac{3}{4}(1-x^2)I(|x| < 1)$ for these simulations. We select h_n by using the rule of thumb bandwidth selection technique for length-biased data. (See Borrajo and *et al.* (2017) for more details.)We compute the absolute values of bias (B_1 and B_2) and RMSE (RMSE1 and RMSE2) of the proposed estimators for various values of $u = F(t)$ and sample sizes. The results are given in Tables 1 and 2. From tables we can observe that as the sample sizes increase, B_1, B_2, RMSE1 and RMSE2 decrease. In beta distribution, VJ^1 generally has better performance than VJ^2 in terms of bias and RMSE. In gamma distribution, for $u = 0.2$, VJ^2 performs better in terms of bias and RMSE and for $u = 0.5, 0.8$, VJ^1 has better performance in terms of bias and RMSE. In the following, we assess the performance of the estimators using real data. Helu and *et al.* (2020) used 70 failure times of aircraft windshields with a unit of measurement of $1000h$. They showed that these data have gamma distribution with parameters $\alpha = 7.75$ and $\beta = 0.285$. Rajesh and *et al.* (2022) modified these data and obtained the following length-biased sample:

3.000	2.83	2.224	1.505	3.166	1.568	1.281	2.135	2.194	3.478
3.117	3.344	2.038	2.661	1.899	2.385	1.981	0.557	2.688	1.876
1.303	3.114	2.097	3.443	2.962	2.632	3.595	2.610	1.615	2.646
2.300	2.089	2.229	3.103	1.652	1.914	1.757	3.385	1.911	2.481
1.506	2.085	1.652	2.223	2.324	1.619	3.578	2.349	1.866	1.912

Table 3 shows values of the residual varextropy and its estimators for the real data. From the Table 3, we can observe that the residual values obtained using VJ_t^1 and VJ_t^2 are close to the residual varextropy VJ_t.

4 Appendix

The proofs of theorems are given in this section.

Proof of Theorem 2.1.

$$VJ_t^1(X) - VJ_t(X) = I_n' + II_n', \tag{18}$$

where

$$I_n' = \left(\frac{1}{4(\overrightarrow{F}_n(t))^3} \int_t^\tau (f_n(x))^2 dF_n(x) - \frac{1}{4(\overrightarrow{F}(t))^3} \int_t^\tau (f(x))^2 dF(x) \right),$$

and

$$II_n' = \left((J_t(X))^2 - (J_n^1(f,t))^2 \right).$$

Table 1: Bias and RMSE of B(2,2).

n	u	VJ^1	B1	RMSE1	VJ^2	B2	RMSE2
50	0.2	0.0384	0.0081	0.0205	0.0551	0.0249	0.0364
50	0.5	0.1044	0.0016	0.0573	0.1371	0.0342	0.0819
50	0.8	0.4464	0.0278	0.3316	0.2404	0.2337	0.7234
100	0.2	0.0346	0.0044	0.0149	0.0497	0.0195	0.0262
100	0.5	0.1018	0.0010	0.0450	0.1328	0.0300	0.0632
100	0.8	0.4407	0.0334	0.2467	0.4139	0.0603	0.3035
500	0.2	0.0317	0.0015	0.0069	0.0407	0.0104	0.0131
500	0.5	0.1034	0.0006	0.0230	0.1226	0.0197	0.0332
500	0.8	0.4419	0.0322	0.1168	0.4953	0.0211	0.1207

.

Table 2: Bias and RMSE of G(2,1).

n	u	VJ^1	B1	RMSE1	VJ^2	B2	RMSE2
50	0.2	0.0024	0.0024	0.0027	0.0031	0.0017	0.0020
50	0.5	0.0063	0.0019	0.0035	0.0048	0.0034	0.0041
50	0.8	0.0146	0.0030	0.0091	0.0022	0.0093	0.0202
100	0.2	0.0027	0.0021	0.0023	0.0034	0.0014	0.0016
100	0.5	0.0068	0.0014	0.0025	0.0057	0.0024	0.0029
100	0.8	0.0138	0.0021	0.0056	0.0075	0.0040	0.0059
500	0.2	0.0036	0.0013	0.0014	0.0040	0.0008	0.0010
500	0.5	0.0076	0.0006	0.0012	0.0070	0.0012	0.0015
500	0.8	0.0127	0.0011	0.0026	0.0104	0.0011	0.0021

From Theorem 1 of Pavithradas and *et al.* (2025), we can observe that

$$\lim_{n\to\infty} II'_n = 0, \ a.s. \tag{19}$$

On the other hand

$$I'_n = \frac{An}{4}\int_t^\tau (f_n(x))^2 dF_n(x) + \frac{1}{4(\overline{F}(t))^3}\left[\int_t^\tau (f_n(x))^2 dF_n(x) - \int_t^\tau (f(x))^2 dF(x)\right], \tag{20}$$

where

$$A_n = \frac{1}{(\overline{F}_n(t))^3} - \frac{1}{(\overline{F}(t))^3}. \tag{21}$$

Using integration by parts we can write:

$$\begin{aligned}
\left|\int_t^\tau (f_n(x))^2 dF_n(x) - \int_t^\tau (f(x))^2 dF(x)\right| &= \Big|\int_t^\tau ((f_n(x))^2 - (f(x))^2) dF_n(x) \\
&- \int_t^\tau (f(x))^2 d(F_n(x) - F(x))\Big| \\
&= \Big|\int_t^\tau \left((f_n(x))^2 - (f(x))^2\right) dF_n(x) \\
&- f^2(t)\left(F_n(t) - F(t)\right) - \int_t^\tau \left(F_n(x) - F(x)\right) d(f^2(x))\Big| \\
&\leq \sup_{0<x<\tau} \left|(f_n(x))^2 - (f(x))^2\right| \\
&+ 3M^2 \sup_{0<x<\infty} |F_n(x) - F(x)|,
\end{aligned} \tag{22}$$

where $M = sup_x f(x) < \infty$. (22), Equation (23) of Zamini and *et al.* (2023),(8) and (20) ensure that

$$\lim_{n\to\infty} I'_n = 0, \ a.s. \tag{23}$$

Table 3: Values of estimators for real data

u	VJ_t	VJ^1	VJ^2
0.2	0.00005790913	0.00008251021	0.00006029210
0.5	0.0001446675	0.0001892130	0.0001952106
0.8	0.0002971429	0.0003312519	0.0004151892

.

(18), (19) and (23) complete the proof. □

Proof of Theorem 2.2.

$$VJ_t^2(X) - VJ_t(X) = I_n + II_n, \tag{24}$$

where

$$I_n = \left(\frac{1}{4(\overrightarrow{F}_n(t))^3} \int_t^\tau (f_n(x))^3 d(x) - \frac{1}{4(\overrightarrow{F}(t))^3} \int_t^\tau (f(x))^3 d(x) \right),$$

and

$$II_n = \left((J_t(X))^2 - (J_n^2(f,t))^2 \right).$$

From Theorem 3 of Pavithradas and *et al.* (2025), we can write:

$$\lim_{n\to\infty} II_n = 0, \ a.s. \tag{25}$$

Next, to deal with In, observe that

$$I_n = \frac{An}{4} \int_t^\tau (f_n(x))^3 d(x) + \frac{1}{4(\overline{F}(t))^3} \left[\int_t^\tau (f_n(x))^3 d(x) - \int_t^\tau (f(x))^3 d(x) \right]. \tag{26}$$

It is easy to see that

$$\begin{aligned}
\left| \int_t^\tau (f_n(x))^3 d(x) - \int_t^\tau (f(x))^3 d(x) \right| &\leq \left| \int_t^\tau ((f_n(x)) - (f(x)))(f_n(x))^2 d(x) \right| \\
&+ \left| \int_t^\tau ((f_n(x))^2 - (f(x))^2) d(F(x)) \right| \\
&\leq \sup_{0<x<\tau} |(f_n(x)) - (f(x))| \int_t^\tau (f_n(x))^2 dx \\
&+ \sup_{0<x<\infty} \left| (f_n(x))^2 - (f(x))^2 \right| \\
&\leq \sup_{0<x<\tau} |(f_n(x)) - (f(x))| \\
&\times \left[\int_t^\tau \left| (f_n(x))^2 - (f(x))^2 \right| dx + \int_t^\tau (f(x))^2 dx \right] \\
&+ \sup_{0<x<\tau} \left| (f_n(x))^2 - (f(x))^2 \right| \\
&< \sup_{0<x<\tau} |(f_n(x)) - (f(x))| \\
&\times \left[\tau \sup_{0<x<\tau} \left| (f_n(x))^2 - (f(x))^2 \right| + \tau M^2 \right] \\
&+ \sup_{0<x<\tau} \left| (f_n(x))^2 - (f(x))^2 \right|,
\end{aligned} \tag{27}$$

where $M = sup_x f(x) < \infty$. Equation (23) of Zamini and *et al.* (2023),(10),(27),(26) and (8) ensure that

$$\lim_{n\to\infty} I_n = 0, \ a.s. \tag{28}$$

(24), (25) and (28) complete the proof. □

Discussion and Results

We proposed some non-parametric kernel estimator for the residual varextropy under length-biased sampling and proved some asymptotic properties for the proposed estimators. Monte Carlo simulation studies were carried out to evaluate the performance of the estimators using the bias and RMSE. A real dataset was used to evaluate the performance of the residual varextropy estimators. It can be easily shown that if there is a t_0 for which $VJ_{t_0}(X) = 0$, then this is equivalent to the fact that $f_{t_0}(x) = \frac{f(x)}{\overline{F}(t_0)} = \frac{1}{1-t_0},\ x > t_0$. This is equivalent to the hazard rate function $\lambda(x) = \frac{f(x)}{\overline{F}(x)}$ is increasing for $x > t_0$. In other words, t_0 is the starting time that the unite in the system is becoming more and more prone to failure. Therefore, the residual varextropy can be used as a way to to detect when systems start to fail.

Acknowledgment

The author thanks the editor and referees for their comments and helpful suggestions which helped to improve the presentation.

References

Ajami M., Fakoor V. and Jomhoori S. (2013), Some asymptotic results of kernel density estimator in length-biased sampling, *J.Sci. Islamic Republic of Iran,* **24**(1), 55-62.

Alizadeh Noughabi H. and Shafaei Noughabi M. (2024), On the estimation of varextropy, *Statistics: A journal of Theoretical and Applied Statistics,* **58**(4),1-15.

Borrajo M. I., Gonzalez-Manteiga W. and Martinez-Miranda M. (2017), Bandwidth selection for kernel density estimation with length-biased data, *J. Nonparametric Stat.,* **29(3), 636-668.**

Chakraborty R. and Rao C.R. (2000), 23 selection biases of samples and their resolutions, *Handbook of Statistics,* **18, 675-712.**

Cox D. (1969), Some sampling problems in technology, In: *New developments in survey sampling. Wiley Interscience, New York,* **506-527.**

Chaubey Y., Sen P. and Li J. (2010), Smooth density estimation for length-biased data, *J. Indian Soc. Agri. Stat.,* **64(2), 145-155.**

Gneiting T. and Raftery A.E. (2007), Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.,* **102(477), 359-378.**

Goodarzi F. (2022), Characterizations of some discrete distributions and upper bounds on discrete residual varentropy, *JIRSS,* **21(2), 231-248.**

Helu A., Samawi H., Rochani H., Yin J. and Vogel R. (2020), Kernel density estimation based on progressive type-II censoring, *J. Korean Stat. Soc.,* **49(2), 475-498.**

Jones M. C. (1991), Kernel density estimation for length-biased data, *Biometrika*, **78(3), 511-519.**

Lad F., Sanfilippo G. and Agro G. (2015), Extropy: complementary dual of entropy, *Stat. Sci.,* **30(1), 40-58.**

Pavithradas V., Rajesh G. and Rajesh R. (2025), Nonparametric estimation of residual extropy function under length-biased sampling, *Ricerche di Matematica,* **1-17.**

Qiu G. (2017), The extropy of order statistics and record values, *Statistics and Probability Letters,* **120(C), 52-60.**

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

Raqab M. Z. and Qiu G. (2019), On entropy properties of ranked set sampling, *Statistics: A Journal of Theoretical and Applied Statistics,* **53(1), 210-226.**

Rajesh R., Rajesh G. and Sunoj S. (2022), Kernel estimation of extropy function under length-biased sampling, *Stat. Prob. Lett.,* **181, 8684-8693.**

Shannon C. (1998), A mathematical theory of communication, *Bell Syst. Tech. J.,* **27(3), 379-423.**

Vaselabadi N. M., Tahmasebi S., Kazemi M. R. and Buono F. (2021), Results on varextropy measure of random variables, *Entropy,* **23(3), 356.**

Zamini R., Goodarzi F. and Salimi M. (2025), Nonparametric estimators for varextropy under $\alpha-$mixing condition with application in exponential $AR(1)$ model, *J. Mahani Math. Res.,* **14(1), 45-61.**

Zamini R., Goodarzi F. and Hashemi F. (2023), Some kernel estimators for varextropy function under length-biased sampling, *Communications in Statistics-Simulation and Computation,* **1-21.**

Zelen M. and Feinleib M. (1969), On the theory of screening for chronic diseases, *Biometrika,* **56(3), 601-614.**