

## A NUMERICAL STUDY OF ENTROPY AND RESIDUAL ENTROPY ESTIMATORS BASED ON SMOOTH DENSITY ESTIMATORS FOR NON-NEGATIVE RANDOM VARIABLES

YOGENDRA P. CHAUBEY\*

*Department of Mathematics and Statistics  
Concordia University, Montreal, Canada H3G 1M8  
Email: yogen.chaubey@concordia.ca*

NHAT LINH VU

*Department of Mathematics and Statistics  
Concordia University, Montreal, Canada H3G 1M8  
Email: nhatlinh.vu@concordia.ca*

### SUMMARY

In this paper, we are interested in estimating the entropy of a non-negative random variable. Since the underlying probability density function is unknown, we propose the use of the Poisson smoothed histogram density estimator to estimate the entropy. To study the performance of our estimator, we run simulations on a wide range of densities and compare our entropy estimators with the existing estimators based on different approaches such as spacing estimators. Furthermore, we extend our study to residual entropy estimators which is the entropy of a random variable given that it has been survived up to time  $t$ .

*Keywords and phrases:* Entropy estimator, Information theory, Residual entropy, Density estimator, Survival function

*AMS Classification:* 94A15; 94A17

## 1 Introduction

For a continuous random variable  $X$ , the differential entropy  $H(X)$  is defined as

$$H(X) = - \int_{-\infty}^{\infty} f(x) \log f(x) dx. \quad (1.1)$$

In the literature, it is also often referred as Shannon entropy as it generalizes the concept of entropy, for a discrete random variable proposed by Shannon (1948) in his landmark paper, that is given by

$$H(X) = - \sum_{x \in A} \mathbb{P}(x) \log \mathbb{P}(x), \quad (1.2)$$

---

\* Corresponding author

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

where  $A$  is the set of possible values of  $X$ , and  $\mathbb{P}(x)$  denotes its probability mass function (*pmf*). We will simply refer to  $H(X)$  as the entropy of the random variable  $X$ .

The concept of entropy takes a central place in statistical theory and applications (see Kapur (1993) and Kapur and Kesavan (1992)). One of the well-known applications of entropy in statistics is the test of normality for a random variable because of the characterizing property that the normal distribution attains the maximum entropy among all continuous distributions with a given variance (see Vasicek (1976) and an adaptation to testing exponentiality by Chaubey et al. (1993)).

Ebrahimi (1995) provided an interpretation of  $H(X)$  as a measure of uncertainty associated with pdf  $f$ . In the context of life testing,  $X$  being the lifetime of a unit, knowing that the unit has survived up to time  $t$ , a more appropriate measure for uncertainty in  $f$  is defined as

$$H(X; t) = - \int_t^\infty \left( \frac{f(x)}{R(t)} \right) \left( \log \frac{f(x)}{R(t)} \right) dx, \quad (1.3)$$

where  $R(t) = \mathbb{P}(X > t)$  is the reliability or the survival function.  $H(X; t)$  is called *residual entropy* of  $X$  given the event  $X > t$ . In this paper, we focus on estimating  $H(X)$  and more generally  $H(X; t)$  when  $X$  is continuous and more specifically, when  $X$  is a non-negative random variable. In recent decades, there exist a huge literature on entropy estimators, which can be classified as follows.

- (i) A naive approach is to discretize the support of the underlying density function into  $m$  bins, then for each bin we compute the empirical measures, corresponding to  $p_i = \int_{A_i} f(x) dx$ . That is if  $X_1, X_2, \dots, X_n$  are independently and identically distributed (i.i.d.) continuous random variables, then  $\hat{p}_i = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{A_i}(X_j)$ , where  $\{A_i\}_{i=1}^m$  are mutually disjoint subsets of  $\mathbb{R}$  such that  $\cup_{i=1}^m A_i = \text{support}(X)$ . Clearly,  $\hat{p}_i$  is the maximum likelihood estimator (MLE) for  $p_i$ . With this estimator  $\hat{p}_i$ , a “naive” - (Strong et al., 1998) or “plug-in” (Antos and Kontoyiannis, 2001) MLE estimator of entropy is given as

$$\hat{H}^{MLE}(X) = - \sum_{i=1}^m \hat{p}_i \log \hat{p}_i. \quad (1.4)$$

It has been shown that this MLE estimator of entropy results in heavy bias (see Panzeri and Treves (1996)). Consequently, the problem of bias correction and the choice of  $m$  have drawn the attention of many researchers; see Paninski (2003), §3 for a detailed account.

- (ii) One straightforward approach for entropy estimation is to estimate the underlying density function  $f(x)$  by some well-known density estimator  $\hat{f}_n(x)$ , then plug it into (1.1) to obtain the entropy estimator

$$\hat{H}^{PlugIn}(X) = - \int_0^\infty \hat{f}_n(x) \log \hat{f}_n(x) dx. \quad (1.5)$$

This approach was first proposed by Dmitriev and Tarasenko (1973) (with  $\hat{f}_n$  being the kernel density estimator) and later investigated by others, for example Györfi and van der Meulen (1987) and Joe (1989). The *fixed symmetric kernel density estimator*  $\hat{f}_n^{Fixed}(x)$ , which is

already a well-known and popular approach for estimating the *pdf* with an unbounded support [see Rosenblatt (1966), Silverman (1986)], is defined as

$$\hat{f}_n^{Fixed}(x) = \frac{1}{nb} \sum_{i=1}^n K\left(\frac{x - X_i}{b}\right), \tag{1.6}$$

where  $K$  is a symmetric probability density function with mean zero and variance one, and  $b$  is the smoothing parameter, called the bandwidth. It is obvious that the performance of this plugin entropy estimator totally depends on the density estimator  $\hat{f}_n$ . However, when dealing with non-negative random variables, this  $\hat{f}_n^{Fixed}(\cdot)$  is shown to produce a heavy bias near the boundary (Silverman, 1986). Consequently, a better estimator may result if we replace  $\hat{f}_n^{Fixed}$  by some density estimator that is free of boundary effect.

(iii) Motivated by the representation of entropy as an expected value

$$H(X) = - \int_0^\infty f(x) \log f(x) dx = -\mathbb{E}[\log f(X)],$$

by the strong law of large number we have  $-\frac{1}{n} \sum_{i=1}^n \log f(X_i) \xrightarrow{a.s.} H(X)$ . Thus we obtain a new entropy estimator if we replace  $f(\cdot)$  by an appropriate density estimator  $\hat{f}_n(\cdot)$ , as given by

$$\hat{H}^{Meanlog}(X) = -\frac{1}{n} \sum_{i=1}^n \log \hat{f}_n(X_i). \tag{1.7}$$

Ahmad and Lin (1976) suggested the use of (1.7) to estimate  $H(X)$ , taking  $\hat{f}_n$  as the kernel density estimator, whereas the asymptotic properties of (1.7) was studied by Hall and Morton (1993).

(iv) The entropy can be estimated by another approach, called “spacing”, which is initiated by Vasicek (1976). By the change of variable  $p = F(x)$ , the entropy (1.1) can be expressed in the form

$$H(X) = - \int_{-\infty}^\infty f(x) \log f(x) dx = \int_0^1 \log \left\{ \frac{d}{dp} F^{-1}(p) \right\} dp.$$

To estimate  $H(X)$ , the distribution  $F$  is replaced by the empirical distribution  $F_n$ , and the differential operator is replaced by a difference operator. As a result, the derivative  $F^{-1}(p)$  is estimated by  $\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})$  for  $(i-1)/n < p \leq i/n, i = m+1, m+2, \dots, n-m$ , where  $X_{(i)}$ 's are the order statistics and  $m$  is a positive integer smaller than  $n/2$ . When  $p \leq m/n$  or  $p > (n-m)/n$ , one-sided differences are used. That is,  $(X_{(i+m)} - X_{(1)})$  and  $(X_{(n)} - X_{(i-m)})$  are used for  $(X_{(i+m)} - X_{(i-m)})$  respectively. All together this leads to the following estimator of entropy

$$\hat{H}^{Vasicek}(X) = \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right\}. \tag{1.8}$$

Motivated by the idea of spacing, researchers have followed this direction and proposed other versions of entropy estimator, which are claimed to have a better performance. We will list some of them in the next section. One of the greatest weakness of spacing estimators is the choice of spacing parameter  $m$ , which does not have the optimal form.

- (v) Lastly, different from all above estimators, Bouzebda et al. (2013) presented a potentially useful estimator of entropy based on smooth estimator of quantile density function. Their idea again starts with the expression of the entropy

$$H(X) = \int_0^1 \log \left\{ \frac{d}{dp} F^{-1}(p) \right\} dp = \int_0^1 \log \left\{ \frac{d}{dp} Q(p) \right\} dp = \int_0^1 \log q(p) dp,$$

where  $Q(p) := \inf\{t : F(t) \geq p\}$  for  $0 < p < 1$  is the quantile function and  $q(p) := dQ(p)/dp = 1/f(Q(p))$  is the quantile density function. Then a new estimator of entropy can be obtained by replacing  $q(\cdot)$  with its appropriate estimator  $\hat{q}_n(\cdot)$ . That is

$$\hat{H}^{Quantile}(X) = \int_0^1 \log \hat{q}_n(p) dp. \quad (1.9)$$

Bouzebda et al. (2013) were motivated by the work of Cheng and Parzen (2010), which introduced a kernel type estimator  $\hat{q}_n(\cdot)$  for  $q(\cdot)$  that has good asymptotic properties.

For estimating the residual entropy  $H(X; t)$  we can write (1.3) as suggested in Belzunce et al. (2001),

$$H(X, t) = \log(R(t)) - \frac{1}{R(t)} \int_t^\infty f(x) \log f(x) dx. \quad (1.10)$$

Then the residual entropy can be estimated if we replace  $R(t)$  by an empirical or kernel estimator  $\hat{R}(t)$ , and an estimator  $\hat{g}_n(x) = \hat{f}_n(x) \log \hat{f}_n(x)$  in place of the functional  $g(x) = f(x) \log f(x)$ .

The organization of the paper is as follows. Section 2 introduces entropy and residual entropy estimators based on Poisson smooth density estimator (Chaubey and Sen, 2009), corresponding to non-negative random variables. The first part of Section 3 is dedicated to the entropy estimators comparison, based on a simulation for a wide range of densities. Finally, we perform the comparison of our residual entropy estimator with other estimators in the second part of Section 3.

## 2 Main results

### 2.1 Entropy estimators

In this sub-section, we will propose two entropy estimators based on the idea of (1.5) and (1.7). The asymptotic properties of these estimators, such as strong convergence and asymptotic normality in the context of densities with non-negative support can be established along the same lines as in Hall and Morton (1993) for estimators of type (1.7) and Eggermont and LaRiccia (1999) for estimators of type (1.5). We defer the discussion of these in another publication.

Recall that the entropy of a random variable can be estimated by the direct plugin approach:

$$\hat{H}^{Plugin}(X) = - \int_0^\infty \hat{f}_n(x) \log \hat{f}_n(x) dx, \quad (2.1)$$

or by the sample mean of  $\log f(X_i), i = 1, 2, \dots, n$ :

$$\hat{H}^{Meanlog}(X) = - \frac{1}{n} \sum_{i=1}^n \log \hat{f}_n(X_i). \quad (2.2)$$

We observe that both the entropy estimators presented above mainly depend on the density estimator, so we would expect to obtain a good entropy estimator if the chosen density estimator is well-behaved. In literature, regarding non-negative random variables, there are many candidates for the density estimator with nice asymptotic properties (Chaubey et al., 2012). Among these, we especially focus on the Poisson smoothed histogram density estimator, which has the following form (Chaubey and Sen, 2009),

$$\hat{f}_n^{Pois}(x) = k \sum_{i=0}^{\infty} \left[ F_n \left( \frac{i+1}{k} \right) - F_n \left( \frac{i}{k} \right) \right] e^{-kx} \frac{(kx)^i}{i!}, \quad (2.3)$$

where  $F_n(\cdot)$  is the empirical distribution function, and  $k := k(n)$  can be viewed as the smoothing parameter. The choice of this estimator is motivated by its computational and theoretical properties. A competitor of this estimator is the asymmetric kernel density estimator (Chaubey et al., 2012), however, it has a problem that it may not be able to accommodate non-zero values of the density at zero (such as exponential density), without further modification. This necessitates introduction of another constant that increases the computational complexity, without any clear advantage (see Chaubey et al. (2011)).

The estimator  $\hat{f}_n^{Pois}(x)$  can be interpreted as a random weighted sum of Poisson mass functions. The asymptotic properties of  $\hat{f}_n^{Pois}(x)$  have been studied and its first weak convergence was proved by Bouezmarni and Scaillet (2005) under the assumption  $\lim_{n \rightarrow \infty} k(n) = \infty$  and  $\lim_{n \rightarrow \infty} nk(n)^{-2} = \infty$ . They also obtained the weak convergence for the case of unbounded pdf  $f$  at  $x = 0$ . Later on, Chaubey et al. (2010) filled the gap in the asymptotic theory of  $\hat{f}_n^{Pois}(x)$  by studying the asymptotic bias, asymptotic variance, strong consistency and the asymptotic normality of the estimator. Particularly, under the assumptions that  $k(n) = cn^h$  for some constant  $c$  and  $0 < h < 1$ , and  $f'(x)$  satisfies the Lipschitz condition of order  $\alpha > 0$ , the asymptotic bias and variance of  $\hat{f}_n^{Pois}(\cdot)$  are given by

$$Bias[\hat{f}_n^{Pois}(x)] \approx \frac{f'(x)}{2cn^h}, \quad (2.4)$$

$$Var[\hat{f}_n^{Pois}(x)] \approx \frac{\mathbb{E}[X]}{2} \sqrt{\frac{c}{2\pi x^3}} f(x) n^{h/2-1}. \quad (2.5)$$

For the strong consistency of  $\hat{f}_n^{Pois}(x)$ , if  $\mathbb{E}[X^{-2}] < \infty$ ,  $f'(x)$  is bounded and  $k(n) = O(n^h)$ , then

$$\|\hat{f}_n^{Pois} - f\| \xrightarrow{a.s.} 0,$$

where  $\|\hat{f}_n^{Pois} - f\| = \sup_{x \in \mathbb{R}^+} |\hat{f}_n^{Pois}(x) - f(x)|$ .

For the asymptotic normality of  $\hat{f}_n^{Pois}(x)$ , if  $\mathbb{E}[X^{-2}] < \infty$ ,  $k(n) = O(n^{2/5})$ , and  $f'(x)$  satisfies the Lipschitz order  $\alpha$  condition, then for  $x$  in a compact set  $I \subset \mathbb{R}^+$ ,

$$n^{2/5}(\hat{f}_n^{Pois}(x) - f(x)) - \frac{1}{2\delta^2}f'(x) \xrightarrow{D} \mathcal{G},$$

where  $\mathcal{G}$  is the Gaussian process with covariance function  $\gamma_x^2 \delta_{xs}$ , where  $\gamma_x^2 = \frac{\mathbb{E}[X]}{2}(2\pi x^3)^{-1/2}f(x)\delta$ ,  $\delta_{xs} = 0$  for  $x \neq s$ ,  $\delta_{xs} = 1$  for  $x = s$ , and  $\delta = \lim_{n \rightarrow \infty} (n^{-1/5}k(n)^{1/2})$ .

Similar to the problem of bandwidth selection in kernel density estimators, the choice of the smoothing parameter  $k$  strongly affects the performance of the resulting density estimators. In general, large values of  $k$  correspond to over-smoothing, whereas small values of  $k$  correspond to under-smoothing.

With these nice asymptotic properties of  $\hat{f}_n^{Pois}(\cdot)$ , we propose here two entropy estimators of the form

$$\hat{H}^{Plugin-Pois}(X) = - \int_0^\infty \hat{f}_n^{Pois}(x) \log \hat{f}_n^{Pois}(x) dx, \quad (2.6)$$

$$\hat{H}^{Meanlog-Pois}(X) = - \frac{1}{n} \sum_{i=1}^n \log \hat{f}_n^{Pois}(X_i). \quad (2.7)$$

## 2.2 Residual entropy estimator

Motivated by the good properties of Poisson smooth density estimator  $\hat{f}_n^{Pois}(\cdot)$  (as mentioned earlier), we suggest a direct plugin residual entropy estimator where the  $\hat{f}_n^{Pois}(x)$  is substituted for  $f(x)$ . That is, our proposed residual entropy estimator is of the form

$$\hat{H}^{Plugin-Pois}(X, t) = \log(\hat{R}^{Pois}(t)) - \frac{1}{\hat{R}^{Pois}(t)} \int_t^\infty \hat{f}_n^{Pois}(x) \log \hat{f}_n^{Pois}(x) dx, \quad (2.8)$$

where  $\hat{R}^{Pois}(t) = \int_t^\infty \hat{f}_n^{Pois}(x) dx$ .

## 3 Simulation studies and discussions

To study the performance of entropy and residual entropy estimators, we run simulations for a wide range of densities which consists of fifteen non-negative densities below. They are categorized into three groups: monotone density, uni-modal density, and bimodal density.

### 1 Monotone density

- Standard exponential(1) with true  $H(X) = 1$ .
- Exponential(10) with true  $H(X) = -\log 10 - 1 \approx -1.3026$ .

- Pareto(2,1) with true  $H(X) \approx 0.8069$ .
- Log-Normal(0,2) with true  $H(X) \approx 2.1121$ .
- Weibull(0.5,0.5) with true  $H(X) \approx 0.4228$ .

## 2 Uni-modal density

- Gamma(2,2) with true  $H(X) \approx 0.8841$ .
- Gamma(7.5,1) with true  $H(X) \approx 2.3804$ .
- Log-Normal(0,0.5) with true  $H(X) \approx 0.7258$ .
- Maxwell(1) with true  $H(X) \approx 0.9962$ .
- Maxwell(20) with true  $H(X) \approx -0.5017$ .
- Weibull(2,2) with true  $H(X) \approx 1.2886$ .

## 3 Bimodal density

- Mix Gamma:  $(1/2)\text{Gamma}(0.5,0.5)+(1/2)\text{Gamma}(2,2)$  with true  $H(X) \approx 2.2757$ .
- Mix Lognormal:  $(1/2)\text{Lognormal}(0,0.5)+(1/2)\text{Lognormal}(0,2)$  with true  $H(X) \approx 1.6724$ .
- Mix Maxwell:  $(1/2)\text{Maxwell}(1)+(1/2)\text{Maxwell}(20)$  with true  $H(X) \approx 0.8014$ .
- Mix Weibull:  $(1/2)\text{Weibull}(0.5,0.5)+(1/2)\text{Weibull}(2,2)$  with true  $H(X) \approx 1.1330$ .

The densities in the monotone group are chosen with different rates of decay, from the slowest standard Exp(1) to the fastest Exp(10). The Pareto distribution is included here to show the effect of different support  $[1, \infty)$  on entropy estimation. On the contrary to the first group, the second group consists of well-behaved densities from highly concentrated Maxwell(20) to widely spread Gamma(7.5,1). Lastly, the simulations are extended to the bimodal densities where each density is a mixture of the same family density but with different parameters. Note that for some densities with a high rate of decay and small variance like Exp(10) or Maxwell(20), the integrand  $f(x) \log f(x)$  produces a non-finite value for a large value of  $x$  in  $\mathbb{R}$  software due to the round-up. However, since the contribution of the right tail of the integrand to the entropy is insignificant for sufficiently large  $x$ , we obtain the appropriate value of the entropy by cutting off the negligible right tail of the integration. That is, instead of integrating over the entire support of  $f(x) \log f(x)$ , we integrate up to a certain value at which the right tail is negligible.

### 3.1 Comparison of entropy estimators

In this sub-section, we run simulation experiments on different estimators which are classified into two groups, namely, ‘spacing estimators’ and ‘non-spacing estimators’.

### 3.1.1 Spacing estimators

1. Vasicek's estimator by Vasicek (1976):

$$\hat{H}_1 = \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right\}. \quad (3.1)$$

Vasicek showed that if the variance is finite, then as  $m \rightarrow \infty$  and  $m/n \rightarrow 0$ ,

$$\hat{H}_1 \xrightarrow{\mathbb{P}} H(X).$$

2. Ebrahimi's estimator by Ebrahimi et al. (1992):

$$\hat{H}_2 = \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{n}{c_i m} (X_{(i+m)} - X_{(i-m)}) \right\}, \quad (3.2)$$

where

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

It was shown in Ebrahimi et al. (1992) that  $\hat{H}_2 \xrightarrow{\mathbb{P}} H(X)$  as  $n, m \rightarrow \infty$ , and  $m/n \rightarrow 0$ . Also, its bias is smaller than that of Vasicek's estimator.

3. Van Es' estimator by Van Es (1992):

$$\hat{H}_3 = \frac{1}{n-m} \sum_{i=1}^{n-m} \log \left\{ \frac{n+1}{m} (X_{(i+m)} - X_{(i)}) \right\} + \sum_{k=m}^n \frac{1}{k} + \log \left( \frac{m}{n+1} \right). \quad (3.3)$$

Van Es (1992) established, under certain conditions, the strong consistency and asymptotic normality of  $\hat{H}_3$ .

4. Correa's estimator by Correa (1995):

$$\hat{H}_4 = -\frac{1}{n} \sum_{i=1}^{n-m} \log \left\{ \frac{\sum_{j=i-m}^{i+m} (X_{(j)} - \bar{X}_{(i)})(j-i)}{n \sum_{j=i-m}^{i+m} (X_{(j)} - \bar{X}_{(i)})^2} \right\}, \quad (3.4)$$

where  $\bar{X}_{(i)} := \frac{1}{2m+1} \sum_{j=i-m}^{i+m} X_{(j)}$ . The resulting estimator of entropy is shown to attain a smaller mean squared error (MSE) compared to that of Vasicek's estimator.

5. WG's estimator by Wiczorkowski and Grzegorzewski (1999):

$$\hat{H}_5 = \hat{H}_1 + \log \left( \frac{2m}{n} \right) - \frac{n-2m}{n} \Psi(2m) + \Psi(n+1) - \frac{2}{n} \sum_{i=1}^m \Psi(i+m-1), \quad (3.5)$$

where  $\Psi(k) = \sum_{i=1}^{k-1} \frac{1}{i} - \gamma$  is the di-Gamma function defined on integer set, and  $\gamma = 0.57721566\dots$  is the Euler's constant.

6. Noughabi's estimator (kernel-spacing) by Noughabi (2010):

$$\hat{H}_6 = -\frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{\hat{f}_n^{Fixed}(X_{(i+m)}) + \hat{f}_n^{Fixed}(X_{(i-m)})}{2} \right\}, \quad (3.6)$$

where the bandwidth in  $\hat{f}_n^{Fixed}(x)$  is fixed to  $b = 1.06sn^{-1/5}$  and  $s$  is the sample standard deviation. Similar to  $\hat{H}_1$ , Noughabi (2010) showed that  $\hat{H}_6$  weakly converges to  $H(X)$ .

### 3.1.2 Non-spacing estimators

To see how the Poisson smoothed histogram density estimator performs in the estimation of entropy, we also run simulation on other direct plug-in and mean of log entropy estimators that use the fixed symmetric kernel density estimator  $\hat{f}_n^{Fixed}$  (with the Gaussian kernel) and the Gamma asymmetric kernel density estimator (Chen, 2010)  $\hat{f}_n^{Gam}$  which has the form

$$\hat{f}_n^{Gam}(x) = \frac{1}{n} \sum_{i=1}^n K_{x/b+1,b}^{Gam}(X_i) = \sum_{i=1}^n \frac{X_i^{x/b} \exp(-X_i/b)}{b^{x/b+1} \Gamma(x/b+1)}. \quad (3.7)$$

Thus, we obtain another four entropy estimators to compare with our proposed estimators. They are

$$\text{Plug-in Fixed} : \hat{H}_7 = - \int_0^\infty \hat{f}_n^{Fixed}(x) \log \hat{f}_n^{Fixed}(x) dx, \quad (3.8)$$

$$\text{Meanlog Fixed} : \hat{H}_8 = -\frac{1}{n} \sum_{i=1}^n \log \hat{f}_n^{Fixed}(X_i), \quad (3.9)$$

$$\text{Plugin Gam} : \hat{H}_9 = - \int_0^\infty \hat{f}_n^{Gam}(x) \log \hat{f}_n^{Gam}(x) dx, \quad (3.10)$$

$$\text{Meanlog Gam} : \hat{H}_{10} = -\frac{1}{n} \sum_{i=1}^n \log \hat{f}_n^{Gam}(X_i). \quad (3.11)$$

The last entropy estimator we want to compare with is based on the quantile density estimator. Here the quantile function is first estimated by the method of Bernstein polynomial of degree  $m$ , which is of the form

$$\tilde{Q}_n(p) = \sum_{i=0}^m \hat{Q}_n\left(\frac{i}{m}\right) b(i, m, p) \quad p \in (0, 1),$$

where  $\hat{Q}_n(\cdot)$  is the empirical quantile function,  $b(i, m, p) = \mathbb{P}[Y = i]$  where  $Y$  follows the binomial( $m, p$ ), and  $m$  is a function of  $n$  such that  $m \rightarrow \infty$  as  $n \rightarrow \infty$ . Then the estimator of  $q(\cdot)$  can be obtained by differentiating  $\tilde{Q}_n(\cdot)$

$$\tilde{q}_n(p) = \frac{d\tilde{Q}_n(p)}{dp} = \sum_{i=0}^m \hat{Q}_n\left(\frac{i}{m}\right) b(i, m, p) \left[ \frac{i - mp}{p(1-p)} \right]. \quad (3.12)$$

Indeed,  $\tilde{q}_n(\cdot)$  is the special case of the quantile kernel density estimator, proposed by Cheng and Parzen (2010), which has the general form

$$\tilde{q}_n^{CP}(p) := \frac{d}{dp} \tilde{Q}_n^{CP}(p) = \frac{d}{dp} \int_0^1 \hat{Q}_n(t) K_n(p, t) d\mu_n(t).$$

As a result, the last entropy estimator is of the form:

$$\hat{H}_{11} = \int_0^1 \log \tilde{q}_n(p) dp. \quad (3.13)$$

And the new Poisson smoothed entropy estimators are labelled as

$$\hat{H}_{12} = \hat{H}^{Plugin-Pois}(X), \quad (3.14)$$

$$\text{and } \hat{H}_{13} = \hat{H}^{Meanlog-Pois}(X). \quad (3.15)$$

### 3.1.3 The Simulation Study

The simulation study is organized as follows. For each density, 500 replicated data are generated for three sample sizes:  $n = 10$ ,  $n = 50$ , and  $n = 100$ . To obtain the entropy estimator, we need to assign a value to the smoothing parameter. Particularly, since the optimal choice for  $m$  in spacing estimators is still an opening problem, for each spacing estimator, we compute the sample entropies for all values of  $m$  from 2 to  $n/2 - 1$ , then we use the optimal entropy with the smallest MSE to compare with other entropy estimators. On the other hand, the bandwidth selection in  $\hat{f}_n^{Fixed}(\cdot)$  is set to  $b = 1.06sn^{-1/5}$  while  $b = sn^{-2/5}$  is used in  $\hat{f}_n^{Gam}(\cdot)$ . The choice of the polynomial degree  $m$  in the quantile density estimator  $\tilde{q}_n(\cdot)$  is fixed to  $m = \lfloor n/\log n \rfloor$ , where  $\lfloor x \rfloor$  denotes the greatest integer less than or equal to  $x$ . Lastly, for Poisson smoothed density estimators, we applied the smoothing parameter  $k = \lfloor n^{2/5} + 1 \rfloor$ . To compare the performance between estimators, for each density and each estimator, we compute the point estimate and its MSE shown in the parentheses, given in Table A1 to Table A9 in Appendix A, for various densities and sample sizes  $n = 10, 50$ , and  $n = 100$ . Each column in the table corresponds to one density with the associated true entropy right below, and the bold value in that column indicates the best estimator with the smallest MSE for that density. By observing the simulation results in the Appendix A, we have the following general remarks.

*Remark 1.* There does not exist the uniquely best entropy estimator in all cases.

It is clear from the simulation results that depending on the density and the sample size, the best entropy estimator switches from one to another. When comparing entropy estimators, not only the smallest MSE is considered, but the computation expense is also an important key to determine the best estimator. Indeed, in terms of speed, the entropy estimators in the spacing group seem to dominate estimators in the other group. For instance, for the sample size  $n = 50$ , it takes at most two seconds to obtain the spacing entropy estimators while it is up to thirty seconds for the direct plugin entropy estimator to be computed. This makes sense because the amount of computations in the latter estimators is much heavier than those in former estimators due to the presence of integration

in their form. This could be the main reason why most papers in literature only focus on spacing estimators but a few on the direct plugin estimators.

*Remark 2.* The direct plugin entropy estimators by means of  $\hat{f}_n^{Fixed}$ ,  $\hat{f}_n^{Gam}$ , and  $\hat{q}_n^{Quantile}$  encounter the convergence problem when computing the integral.

Another issue when we run the direct plugin entropy estimators is the convergence of the integral. Perhaps, this is the greatest weakness of the direct plugin estimators. Indeed, in all cases, we could not evaluate the integral in  $\hat{H}_7 = \hat{H}^{Plugin-Fixed}(X)$  and  $\hat{H}_9 = \hat{H}^{Plugin-Gam}(X)$  over the same interval that we used to compute the true entropy because they produce non-finite values on the right tail. Thus, we have to cut off the larger tail of the integral to a shorter finite interval up to the point that the value of the integrand is finite. This issue, however, does not appear in our entropy estimator  $\hat{H}_{12} = \hat{H}^{Plugin-Pois}(X)$ , so it makes our entropy estimator more favourable than other entropy estimators based on the direct plugin approach.

*Remark 3.* Our entropy estimator by the mean of log of Poisson density point

$$\hat{H}_{13} = \hat{H}^{Meanlog-Pois}(X)$$

seems to be a potential candidate for entropy estimator when the sample size is small.

We see that most of the times, the MSE of  $\hat{H}_{13}$  is the smallest compared to that of other estimators for small sample size in the same density. Note that we have to take into account that while the spacing entropy estimators are computed with the best optimal spacing parameter  $m$ , the smoothing parameter in our estimators is set to  $k = \lfloor n^{2/5} + 1 \rfloor$ , which may be far away from the optimal value. Still, our entropy estimator  $\hat{H}_{13}$  outperforms other entropy estimators in most cases.

*Remark 4.* As the sample size increases, the spacing entropy estimators,  $\hat{H}_2$ ,  $\hat{H}_3$ ,  $\hat{H}_4$ , and  $\hat{H}_5$ , perform much better and converge to the true value.

Although for a small sample size like  $n = 10$  (Table A1-A3), the performance of the entropy estimators from the spacing group is quite poor comparing to the estimators in the other group, their rate of convergence, however, pick up very fast as the sample size increases as in  $n = 100$  (Tables A7–A9). For instance, the number of the best entropy estimator that falls into the spacing group increase as the sample size increases from 10 to 100. Therefore, together with the fast computation, the spacing entropy estimators would be a better choice in the case of a large sample size.

**Conclusion:** There is a trade-off between the speed and the performance for the two groups of entropy estimators. Any spacing estimator (Ebrahimi's estimator, Van Es' estimator, Correa's estimator, or WG's estimator) would do a good job when dealing with large sample sizes due to its fast rate of convergence and fast computation; though the answer to the question of what value of  $m$  would be considered for large sample size is still unknown. However, when the sample size is small, the choice of an estimator for entropy becomes more difficult and requires careful study. This is why in the small sample size, the entropy estimators utilizing density estimator come in handy because they give a more precise result. It may be highlighted that the Poisson smoothed entropy estimator of the mean-log form (namely  $\hat{H}_{13}$ ) consistently performs better in comparison to the other spacing estimators, under small sample sizes.

## 3.2 Comparison of residual entropy estimators

### 3.2.1 The residual entropy estimators

In this sub-section, we want to study the performance of our proposed estimator for residual entropy function as well as compare it with the existing ones. Particularly, we use the same set of testing distributions as used in the previous section, whereas for the residual entropy estimators, we consider the two estimators given by Belzunce et al. (2001) and the estimators based on quantile smoothing and Poisson smooth density estimator. The first estimator considered by Belzunce et al. (2001) is the direct plugin estimator,

$$\hat{H}_1^{Belzunce}(X, t) = \log \hat{R}(t) - \frac{1}{\hat{R}(t)} \int_t^\infty \hat{f}_n^{Fixed}(x) \log \hat{f}_n^{Fixed}(x) dx \quad (3.16)$$

where  $\hat{f}_n^{Fixed}(\cdot)$  is based on the Gaussian kernel with the bandwidth  $b = 1.06sn^{-1/5}$  and  $\hat{R}(t)$ , the estimated survival function, is computed based on the integration of  $\hat{f}_n^{Fixed}(\cdot)$ . The second estimator of Belzunce et al. (2001), based on convolution arguments, is given by

$$\hat{H}_2^{Belzunce}(X, t) = \log \hat{R}(t) - \frac{1}{\hat{R}(t)} \sum_{i=1}^n R_K\left(\frac{t - X_i}{b}\right) \log \hat{f}_{i,n}^{Fixed}(X_i), \quad (3.17)$$

where  $R_K(\cdot)$  is the survival function associated with kernel  $K$  and

$$\hat{f}_{i,n}^{Fixed}(X_i) = \frac{1}{(n-1)b} \sum_{j \neq i} K\left(\frac{X_i - X_j}{b}\right).$$

The quantile smoothing and Poisson smooth density estimators are given by

$$\hat{H}^{Quantile}(X, t) = \log(\hat{R}^{Pois}(t)) - \frac{1}{\hat{R}^{Pois}(t)} \int_{\hat{F}_n^{Pois}(t)}^1 \log \tilde{q}_n(p) dp, \quad (3.18)$$

$$\hat{H}^{Pois}(X, t) = \log \hat{R}^{Pois}(t) - \frac{1}{\hat{R}^{Pois}(t)} \int_t^\infty \hat{f}_n^{Pois}(x) \log \hat{f}_n^{Pois}(x) dx, \quad (3.19)$$

where  $\hat{R}^{Pois}(t)$ , is computed based on the integration of  $\hat{f}_n^{Pois}(\cdot)$ ,  $\hat{F}_n^{Pois}(t) = 1 - \hat{R}^{Pois}(t)$ , the smoothing parameter  $k(n)$  being  $\lfloor n^{2/5} + 1 \rfloor$ , and the parameter  $m$  in  $\tilde{q}_n(\cdot)$  is set to  $\lfloor n / \log n \rfloor$ .

### 3.2.2 The simulation study

The setup of the simulation study is as follows. We present here both performance comparison by graphics (for  $n = 50$ ) and by mean integrated squared error (MISE) (for  $n = 50, 100$  and  $500$ ). To produce the plots of residual entropy estimators, for each distribution, we run 500 replications, then we compute the sample mean residual entropy estimator functions of these 500 replications. As mentioned in Belzunce et al. (2001), the behaviour of residual entropy function (see Figures B1-B3) is smooth up to a certain bound because  $R(t) \rightarrow 0$  very fast when  $t \rightarrow \infty$ , it is recommended to

stop the estimation at the time  $t^* = \inf\{t : R(t) \leq 0.01\}$ . Therefore, the MISE's are just computed on the interval  $[0, t^*]$  (for Pareto(2,1) the interval is  $[1, t^*]$ ). Similarly, the MISE's comparison between estimators are done based on 500 replications for sample size  $n = 50$  and  $n = 100$  for each distribution, but only 100 replications for sample size  $n = 500$ . The plots and tables, presented in Appendix B, show the performance of our estimators and the ones in Belzunce et al. (2001). To illustrate the rate of convergence of each estimator as the sample size increases, we also show the graphs of MISEs as a function of sample size (see Figures B4-B6) for each distribution. We draw the following conclusions based on these computations.

*Remark 5.*  $\hat{H}_2^{Belzunce}(X, t)$  performs very poorly in most cases.

By the plot of residual entropy estimators from Figures B1-B3, we see that  $\hat{H}_2^{Belzunce}(X, t)$  is not as good as the others among all estimators. It is even out of the range of MISE in some distributions. Therefore, we decide to remove it from our consideration.

*Remark 6.* All estimators tend to diverge away from the true residual entropy over time in most cases.

We notice that most estimators have a quite good startup as  $t$  is small. However, as the time  $t$  increases, in which the true survival function tends to zero rapidly, the estimators begin to diverge away from the true residual entropy. How large of the time  $t$  this phenomenon happens depends on the true distribution.

*Remark 7.* None of the estimators outperforms the others in all cases. However, the  $\hat{H}^{Pois}(X, t)$  estimator seems to achieve a better estimation in most cases.

If we just focus on the performance comparison between estimators, we see that in most cases, our estimator  $\hat{H}^{Pois}(X, t)$  captures the trend of the true function throughout the plots and achieves better precision in terms of MISE shown in the Table B1–B3. It shows inferior performance to  $\hat{H}_1^{Belzunce}(X, t)$  only for Maxwell distributions which have the bell-normal shape. Therefore, the  $\hat{H}^{Pois}(X, t)$  is recommended for small sample sizes.

*Remark 8.* The rate of convergence of the estimator  $\hat{H}^{Quantile}(X, t)$  seems to be faster than other competitors.

Similar to the entropy estimation, the residual entropy estimator,  $\hat{H}^{Quantile}(X, t)$  still has the fastest rate of convergence in most cases. We observe that for small sample size,  $\hat{H}^{Quantile}(X, t)$  does not show up to be a good estimator, and it is not recommended. However, if the sample size is sufficiently large, it becomes one of the best potential choices for residual entropy estimators.

**Conclusion:** The Poisson smoothed residual entropy estimator  $\hat{H}^{Pois}(X, t)$  is recommended for small sample sizes, whereas  $\hat{H}^{Quantile}(X, t)$ , based on quantile smoothing may be preferred for larger sample sizes.

## A Comparison of Entropy Estimators

		Monotone densities				
		Exp(1)	Exp(10)	Pareto(2,1)	Lnorm(0,2)	Weibull(0.5,0.5)
		H=1	H=-1.3026	H=0.8069	H=2.1121	H=0.4228
Spacing estimators	$\hat{H}_1$	0.5694(0.3167)	-1.7332(0.3167)	0.4857(0.3907)	1.8084(0.6484)	0.4196(0.4067)
	$\hat{H}_2$	0.8694(0.1505)	-1.4332(0.1505)	0.6350(0.2904)	2.0045(0.5677)	0.4785(0.3998)
	$\hat{H}_3$	0.8869(0.1503)	-1.4157(0.1503)	0.7083(0.2581)	<b>1.9921(0.5311)</b>	0.5426(0.4297)
	$\hat{H}_4$	0.7625(0.1896)	-1.5401(0.1896)	0.6311(0.2988)	2.0206(0.5815)	0.4936(0.4015)
	$\hat{H}_5$	0.9908(0.1313)	-1.3117(0.1313)	0.8515(0.2628)	2.2210(0.5680)	0.6950(0.4707)
	$\hat{H}_6$	1.0627(0.1428)	-1.2399(0.1428)	1.1180(0.5314)	2.9613(1.8878)	1.2738(1.2629)
Non-spacing estimators	$\hat{H}_7$	1.0142(0.0704)	divergent	<b>0.7446(0.0659)</b>	1.2540(0.9033)	divergent
	$\hat{H}_8$	1.0934(0.1525)	-1.2091(0.1525)	1.1658(0.5780)	3.0210(2.0037)	1.3304(1.3736)
	$\hat{H}_9$	1.0390(0.0842)	divergent	1.1095(0.2202)	1.7242(0.3123)	0.9027(0.3854)
	$\hat{H}_{10}$	1.0102(0.1178)	<b>-1.2923(0.1178)</b>	1.2326(0.4528)	2.5934(1.0828)	0.9867(0.7638)
	$\hat{H}_{11}$	0.7272(0.2140)	1.5754(0.2140)	0.6924(0.3503)	2.2866(0.8415)	0.6751(0.5048)
	$\hat{H}_{12}$	<b>1.0610(0.0676)</b>	-0.2004(1.2198)	1.1612(0.1958)	1.7180(0.3854)	0.8643(0.3612)
	$\hat{H}_{13}$	0.8677(0.0983)	-0.8690(0.2064)	1.0988(0.1748)	1.6043(0.5424)	<b>0.4957(0.2262)</b>

Table A1: Simulation results for  $n = 10$  of the monotone density group. The point estimate and MSE (in parentheses) are computed for each estimator. The bold value is the best estimator with the smallest MSE.

		Unimodal densities					
		Gam(2,2)	Gam(7.5,1)	Lnorm(0,0.5)	Maxwell(1)	Maxwell(20)	Weibull(2,2)
		H=0.8841	H=2.3804	H=0.7258	H=0.9962	H=-0.5017	H=1.2886
Spacing estimators	$\hat{H}_1$	0.3676(0.3560)	1.8444(0.3569)	0.2076(0.2723)	0.4790(0.3319)	-1.0189(0.3319)	0.7885(0.3228)
	$\hat{H}_2$	0.6342(0.1489)	2.0746(0.1576)	0.4608(0.1746)	0.7084(0.1407)	-0.7895(0.1407)	1.0229(0.1371)
	$\hat{H}_3$	0.7103(0.1300)	2.1784(0.1191)	0.5316(0.1482)	0.8290(0.1045)	-0.6688(0.1045)	1.1352(0.1096)
	$\hat{H}_4$	0.5401(0.2063)	2.0139(0.2043)	0.3786(0.2259)	0.6481(0.1860)	-0.8498(0.1860)	0.9577(0.1822)
	$\hat{H}_5$	0.7745(0.0973)	2.2571(0.0848)	0.6203(0.1149)	0.8917(0.0754)	-0.6062(0.0754)	1.1787(0.0801)
	$\hat{H}_6$	0.9023(0.0874)	2.3403(0.0646)	0.7244(0.1042)	0.9756(0.0588)	-0.5223(0.0588)	1.2893(0.0668)
Non-spacing estimators	$\hat{H}_7$	0.8891(0.0613)	2.4273(0.0548)	0.8021(0.0895)	1.0301(0.0435)	divergent	1.2912(0.0442)
	$\hat{H}_8$	0.8332(0.0951)	2.2495(0.0802)	0.6503(0.1156)	0.8840(0.0711)	-0.6139(0.0711)	1.2016(0.0756)
	$\hat{H}_9$	0.9395(0.0571)	2.6388(0.0968)	0.9217(0.0959)	1.1949(0.0648)	-0.2888(0.0731)	1.4138(0.0467)
	$\hat{H}_{10}$	0.8406(0.0742)	2.3963(0.0444)	0.7333(0.0801)	0.9933(0.0395)	<b>-0.5046(0.0395)</b>	1.2688(0.0489)
	$\hat{H}_{11}$	0.5040(0.2332)	1.9697(0.2387)	0.3450(0.2533)	0.5998(0.2212)	0.8980(0.2212)	0.9114(0.2112)
	$\hat{H}_{12}$	1.0245(0.0532)	<b>2.3802(0.0257)</b>	1.0476(0.1271)	1.2644(0.0832)	0.2563(0.5828)	1.4288(0.0387)
	$\hat{H}_{13}$	<b>0.8120(0.0504)</b>	2.1908(0.0707)	<b>0.7665(0.0400)</b>	<b>1.0046(0.0215)</b>	-0.1038(0.1655)	<b>1.2183(0.0353)</b>

Table A2: Simulation results for  $n = 10$  of the uni-modal density group. The point estimate and MSE (in parentheses) are computed for each estimator. The bold value is the best estimator with the smallest MSE.

		Bimodal densities			
		MixGam	MixLnorm	MixMaxwell	MixWeibull
		H=2.2757	H=1.6724	H=0.8014	H=1.1330
Spacing estimators	$\hat{H}_1$	2.0233(0.0945)	1.3541(0.5505)	0.4218(0.1888)	0.9726(0.1635)
	$\hat{H}_2$	2.3402(0.0350)	1.5502(0.4641)	0.7387(0.0486)	1.2228(0.1480)
	$\hat{H}_3$	2.4074(0.0504)	1.4707(0.3565)	0.8033(0.0595)	1.3422(0.1414)
	$\hat{H}_4$	2.2469(0.0333)	1.5604(0.4861)	0.6319(0.0734)	1.1713(0.1532)
	$\hat{H}_5$	2.3214(0.0432)	1.7667(0.4581)	0.8247(0.0444)	1.3253(0.1918)
	$\hat{H}_6$	2.6307(0.1614)	2.2060(1.5798)	0.9993(0.0913)	1.5081(0.2682)
Non-spacing estimators	$\hat{H}_7$	2.4427(0.0539)	1.2077(0.3345)	1.0057(0.0703)	<b>1.3375(0.0623)</b>
	$\hat{H}_8$	2.4700(0.0946)	2.2567(1.6634)	1.0161(0.0967)	1.5286(0.1927)
	$\hat{H}_9$	2.5721(0.1097)	<b>1.4677(0.1810)</b>	1.0041(0.0675)	1.3452(0.0695)
	$\hat{H}_{10}$	2.5106(0.0952)	1.9468(0.8679)	0.9495(0.0585)	1.3924(0.1293)
	$\hat{H}_{11}$	2.1290(0.0563)	1.7298(0.8283)	0.5777(0.0961)	1.0935(0.1563)
	$\hat{H}_{12}$	<b>2.2509(0.0190)</b>	1.5151(0.1883)	1.0488(0.0793)	1.3312(0.0664)
	$\hat{H}_{13}$	2.0440(0.0813)	1.3316(0.2847)	<b>0.8673(0.0264)</b>	1.1107(0.1077)

Table A3: Simulation results for  $n = 10$  of the bimodal density group. The point estimate and MSE (in parentheses) are computed for each estimator. The bold value is the best estimator with the smallest MSE.

		Monotone densities				
		Exp(1) H=1	Exp(10) H=-1.3026	Pareto(2,1) H=0.8069	Lnorm(0,2) H=2.1121	Weibull(0.5,0.5) H=0.4228
Spacing estimators	$\hat{H}_1$	0.8818(0.0368)	-1.4207(0.0368)	0.7405(0.0663)	2.0568(0.1184)	0.4555(0.1011)
	$\hat{H}_2$	0.9941(0.0229)	<b>-1.3085(0.0229)</b>	0.7653(0.0587)	2.0646(0.1143)	0.4475(0.1029)
	$\hat{H}_3$	0.9516(0.0270)	-1.3509(0.0270)	0.7454(0.0547)	2.0646(0.1123)	0.4677(0.1002)
	$\hat{H}_4$	0.9758(0.0236)	-1.3268(0.0236)	0.7875(0.0570)	2.0793(0.1120)	0.4595(0.1054)
	$\hat{H}_5$	1.0174(0.0230)	-1.2851(0.0230)	<b>0.8142(0.0535)</b>	2.1403(0.1111)	0.4647(0.1065)
	$\hat{H}_6$	1.1124(0.0364)	-1.1902(0.0364)	1.2279(0.3808)	3.4125(2.2625)	1.3432(0.9828)
Non-spacing estimators	$\hat{H}_7$	1.3702(0.1554)	divergent	1.0947(0.1223)	<b>2.1172(0.0657)</b>	1.2002(0.6431)
	$\hat{H}_8$	1.5531(0.3478)	-1.1404(0.0506)	1.3116(0.4532)	3.5042(2.4923)	1.4557(1.1993)
	$\hat{H}_9$	1.0648(0.0219)	-1.0389(0.0915)	1.1494(0.1534)	1.7962(0.1410)	1.0498(0.4362)
	$\hat{H}_{10}$	1.0577(0.0250)	-1.2449(0.0250)	1.2538(0.2882)	2.8416(0.8425)	1.0126(0.4482)
	$\hat{H}_{11}$	1.0215(0.0250)	1.2811(0.0250)	0.9867(0.1284)	2.5532(0.3923)	0.7510(0.2102)
	$\hat{H}_{12}$	1.0665(0.0215)	-0.5605(0.5543)	1.1189(0.1188)	1.8701(0.1160)	0.7985(0.1898)
	$\hat{H}_{13}$	<b>0.9650(0.0211)</b>	-1.0823(0.0581)	1.0811(0.1031)	1.9167(0.1194)	<b>0.5376(0.0785)</b>

**Table A4:** Simulation results for  $n = 50$  of the monotone density group. The point estimate and MSE (in parentheses) are computed for each estimator. The bold value is the best estimator with the smallest MSE.

		Unimodal densities					
		Gam(2,2) H=0.8841	Gam(7.5,1) H=2.3804	Lnorm(0,0.5) H=0.7258	Maxwell(1) H=0.9962	Maxwell(20) H=-0.5017	Weibull(2,2) H=1.2886
Spacing estimators	$\hat{H}_1$	0.7377(0.0363)	2.2204(0.0379)	0.5805(0.0419)	0.8436(0.0320)	-0.6543(0.0320)	1.1291(0.0361)
	$\hat{H}_2$	0.8597(0.0165)	2.3436(0.0151)	0.6915(0.0238)	0.9655(0.0103)	-0.5324(0.0103)	1.2480(0.0135)
	$\hat{H}_3$	0.8264(0.0214)	2.3025(0.0216)	0.6508(0.0279)	0.9360(0.0157)	-0.5619(0.0157)	1.2244(0.0172)
	$\hat{H}_4$	0.8327(0.0178)	2.3166(0.0166)	0.6756(0.0234)	0.9398(0.0123)	-0.5581(0.0123)	1.2238(0.0153)
	$\hat{H}_5$	0.8792(0.0149)	2.3619(0.0126)	0.7183(0.0206)	0.9869(0.0087)	-0.5109(0.0087)	1.2707(0.0109)
	$\hat{H}_6$	0.8962(0.0164)	2.3811(0.0124)	0.7300(0.0194)	0.9848(0.0102)	-0.5131(0.0102)	1.2865(0.0110)
Non-spacing estimators	$\hat{H}_7$	0.9466(0.0164)	2.469(0.0188)	0.5744(0.0347)	1.0689(0.0126)	divergent	1.3212(0.0088)
	$\hat{H}_8$	0.9191(0.0175)	2.3518(0.0125)	0.7390(0.0221)	0.9785(0.0093)	-0.5194(0.0093)	1.2748(0.0107)
	$\hat{H}_9$	0.9593(0.0171)	2.5833(0.0489)	0.9091(0.0492)	1.1594(0.0323)	-0.3360(0.0332)	1.3908(0.0172)
	$\hat{H}_{10}$	0.8970(0.0139)	2.4078(0.0105)	0.7630(0.0195)	1.0153(0.0079)	<b>-0.4825(0.0079)</b>	1.2899(0.0089)
	$\hat{H}_{11}$	0.8705(0.0169)	2.3480(0.0151)	0.7267(0.0259)	0.9621(0.0105)	-0.5358(0.0105)	1.2434(0.0138)
	$\hat{H}_{12}$	1.0109(0.0247)	<b>2.4258(0.0096)</b>	0.9796(0.0726)	1.2045(0.0468)	0.1018(0.3659)	1.3998(0.0176)
	$\hat{H}_{13}$	<b>0.8797(0.0113)</b>	2.3276(0.0122)	<b>0.7750(0.0150)</b>	<b>1.0245(0.0067)</b>	-0.2296(0.0765)	<b>1.2728(0.0078)</b>

**Table A5:** Simulation results for  $n = 50$  of the uni-modal density group. The point estimate and MSE (in parentheses) are computed for each estimator. The bold value is the best estimator with the smallest MSE.

		Bimodal densities			
		MixGam H=2.2757	MixLnorm H=1.6724	MixMaxwell H=0.8014	MixWeibull H=1.1330
Spacing estimators	$\hat{H}_1$	2.2335(0.0117)	1.6370(0.0742)	0.6712(0.0237)	1.1196(0.0440)
	$\hat{H}_2$	2.2732(0.0086)	1.6386(0.0707)	0.7997(0.0069)	1.1524(0.0431)
	$\hat{H}_3$	2.2847(0.0080)	1.5939(0.0704)	0.7801(0.0100)	1.1969(0.0306)
	$\hat{H}_4$	2.2792(0.0088)	1.6486(0.0684)	0.7649(0.0082)	1.1160(0.0438)
	$\hat{H}_5$	2.2955(0.0099)	1.7084(0.0690)	<b>0.8109(0.0067)</b>	1.1781(0.0444)
	$\hat{H}_6$	2.5927(0.1088)	2.6941(1.7221)	0.9969(0.0462)	1.5212(0.1882)
Non-spacing estimators	$\hat{H}_7$	2.5123(0.0619)	1.5856(0.0733)	1.0167(0.0515)	1.3733(0.0263)
	$\hat{H}_8$	2.5520(0.0864)	2.7858(1.9013)	1.0072(0.0503)	1.5531(0.1118)
	$\hat{H}_9$	2.6064(0.1150)	<b>1.6203(0.0302)</b>	0.9806(0.0369)	1.3559(0.0239)
	$\hat{H}_{10}$	2.5520(0.0845)	2.1751(0.5432)	0.9149(0.0185)	1.3840(0.0428)
	$\hat{H}_{11}$	2.3299(0.0121)	2.1617(0.4259)	0.8168(0.0079)	1.3226(0.0996)
	$\hat{H}_{12}$	<b>2.3168(0.0077)</b>	1.5664(0.0495)	1.0213(0.0522)	<b>1.3105(0.0195)</b>
	$\hat{H}_{13}$	2.2166(0.0118)	1.5316(0.0676)	0.8761(0.0097)	1.1669(0.0378)

**Table A6:** Simulation results for  $n = 50$  of the bimodal density group. The point estimate and MSE (in parentheses) are computed for each estimator. The bold value is the best estimator with the smallest MSE.

		Monotone densities				
		Exp(1)	Exp(10)	Pareto(2,1)	Lnorm(0,2)	Weibull(0.5,0.5)
		H=1	H=-1.3026	H=0.8069	H=2.1121	H=0.4228
Spacing estimators	$\hat{H}_1$	0.9246(0.0166)	-1.3780(0.0166)	0.7705(0.0290)	2.0973(0.0518)	0.4337(0.0466)
	$\hat{H}_2$	0.9975(0.0107)	-1.3051(0.0107)	0.7820(0.0265)	2.1137(0.0509)	0.4323(0.0469)
	$\hat{H}_3$	0.9595(0.0129)	-1.3431(0.0129)	0.7715(0.0256)	2.0754(0.0521)	<b>0.4448(0.0455)</b>
	$\hat{H}_4$	0.9854(0.0107)	-1.3172(0.0107)	0.8010(0.0250)	2.1238(0.0507)	0.4409(0.0479)
	$\hat{H}_5$	1.0030(0.0104)	<b>-1.2996(0.0104)</b>	<b>0.8152(0.0244)</b>	2.1296(0.0504)	0.4469(0.0485)
	$\hat{H}_6$	1.0916(0.0196)	-1.2110(0.0196)	1.2405(0.3582)	3.4532(2.1463)	1.3183(0.8771)
Non-spacing estimators	$\hat{H}_7$	1.0980(0.0175)	divergent	1.1411(0.1391)	<b>2.1739(0.0389)</b>	1.2275(0.6685)
	$\hat{H}_8$	1.1501(0.0335)	-1.1525(0.0335)	1.3429(0.4410)	3.5585(2.4107)	1.4594(1.1432)
	$\hat{H}_9$	1.0501(0.0111)	-1.0950(0.0534)	1.1478(0.1347)	1.8295(0.1033)	1.0531(0.4198)
	$\hat{H}_{10}$	1.0451(0.0121)	-1.2575(0.0121)	1.2469(0.2491)	2.8244(0.6626)	0.9872(0.3663)
	$\hat{H}_{11}$	1.0375(0.0128)	-1.2651(0.0128)	0.6637(0.0502)	1.9686(0.0903)	divergent
	$\hat{H}_{12}$	<b>1.0041(0.0066)</b>	-0.6973(0.3688)	1.2950(0.2498)	1.8994(0.0776)	0.7686(0.1453)
	$\hat{H}_{13}$	0.9733(0.0103)	-1.1530(0.0280)	1.0713(0.0851)	1.9845(0.0585)	0.5433(0.0476)

**Table A7:** Simulation results for  $n = 100$  of the monotone density group. The point estimate and MSE (in parentheses) are computed for each estimator. The bold value is the best estimator with the smallest MSE.

		Unimodal densities					
		Gam(2,2)	Gam(7.5,1)	Lnorm(0,5)	Maxwell(1)	Maxwell(20)	Weibull(2,2)
		H=0.8841	H=2.3804	H=0.7258	H=0.9962	H=-0.5017	H=1.2886
Spacing estimators	$\hat{H}_1$	0.7951(0.0150)	2.2896(0.0147)	0.6469(0.0151)	0.9033(0.0132)	-0.5946(0.0132)	1.193(0.0140)
	$\hat{H}_2$	0.8742(0.0075)	<b>2.3668(0.0060)</b>	0.7153(0.0090)	0.9882(0.0049)	-0.5096(0.0049)	1.2763(0.0055)
	$\hat{H}_3$	0.8446(0.0102)	2.3333(0.0097)	0.6770(0.0125)	0.9569(0.0077)	-0.5410(0.0077)	1.2508(0.0080)
	$\hat{H}_4$	0.8629(0.0076)	2.3587(0.0071)	0.7106(0.0088)	0.9732(0.0053)	-0.5247(0.0053)	1.2635(0.0056)
	$\hat{H}_5$	0.8843(0.0070)	2.3777(0.0064)	0.7280(0.0085)	0.9983(0.0046)	<b>-0.4995(0.0046)</b>	1.2851(0.0049)
	$\hat{H}_6$	0.8939(0.0079)	2.3560(0.0068)	0.7296(0.0087)	0.9779(0.0049)	-0.5200(0.0049)	1.2794(0.0048)
Non-spacing estimators	$\hat{H}_7$	0.9397(0.0093)	2.4614(0.0124)	0.8477(0.0235)	1.0597(0.0079)	divergent	1.3233(0.0049)
	$\hat{H}_8$	0.9180(0.0089)	2.3663(0.0063)	0.7400(0.0093)	0.9866(0.0046)	<b>-0.5113(0.0046)</b>	1.2873(0.0047)
	$\hat{H}_9$	0.9464(0.0097)	2.5499(0.0331)	0.8808(0.0310)	1.1301(0.0210)	-0.3666(0.0214)	1.3782(0.0114)
	$\hat{H}_{10}$	0.8930(0.0068)	2.4025(0.0059)	0.7537(0.0086)	1.0094(0.0041)	-0.4885(0.0041)	1.2935(0.0042)
	$\hat{H}_{11}$	0.8628(0.0100)	2.3522(0.0103)	0.6953(0.0137)	0.9700(0.0073)	0.5278(0.0073)	1.2546(0.0079)
	$\hat{H}_{12}$	0.9891(0.0156)	2.4262(0.0065)	0.9407(0.0503)	1.1704(0.0324)	0.0272(0.2806)	1.3856(0.0121)
	$\hat{H}_{13}$	<b>0.8816(0.0058)</b>	2.3514(0.0063)	<b>0.7639(0.0076)</b>	<b>1.0180(0.0038)</b>	-0.2849(0.0485)	<b>1.2834(0.0038)</b>

**Table A8:** Simulation results for  $n = 100$  of the uni-modal density group. The point estimate and MSE (in parentheses) are computed for each estimator. The bold value is the best estimator with the smallest MSE.

		Bimodal densities			
		MixGam	MixLnorm	MixMaxwell	MixWeibull
		H=2.2757	H=1.6724	H=0.8014	H=1.1330
Spacing estimators	$\hat{H}_1$	2.2615(0.0054)	1.6434(0.0346)	0.7194(0.0103)	1.1248(0.0216)
	$\hat{H}_2$	2.2818(0.0037)	1.6476(0.0335)	0.7998(0.0037)	1.1424(0.0213)
	$\hat{H}_3$	<b>2.2834(0.0033)</b>	1.6211(0.0348)	0.7926(0.0053)	1.1735(0.0154)
	$\hat{H}_4$	2.2817(0.0037)	1.6850(0.0327)	0.7854(0.0040)	1.1513(0.0215)
	$\hat{H}_5$	2.2859(0.0041)	1.6869(0.0326)	<b>0.8035(0.0036)</b>	1.1524(0.0214)
	$\hat{H}_6$	2.5557(0.0824)	2.8079(1.8217)	0.9702(0.0328)	1.5147(0.1646)
Non-spacing estimators	$\hat{H}_7$	2.5170(0.0611)	1.8478(0.0672)	1.0020(0.0432)	1.3790(0.0175)
	$\hat{H}_8$	2.5496(0.0795)	2.9098(2.0306)	0.9793(0.0359)	1.5534(0.0919)
	$\hat{H}_9$	2.5966(0.1058)	<b>1.6365(0.0158)</b>	0.9584(0.0273)	1.3569(0.0144)
	$\hat{H}_{10}$	2.537(0.0755)	2.1910(0.4629)	0.8873(0.0104)	1.3706(0.0239)
	$\hat{H}_{11}$	2.4534(0.0393)	1.5709(0.0510)	0.7508(0.0076)	divergent
	$\hat{H}_{12}$	2.3167(0.0044)	1.5792(0.0287)	0.9955(0.0398)	<b>1.2996(0.0100)</b>
	$\hat{H}_{13}$	2.2402(0.0050)	1.5759(0.0345)	0.8617(0.0059)	1.1753(0.0250)

**Table A9:** Simulation results for  $n = 100$  of the bimodal density group. The point estimate and MSE (in parentheses) are computed for each estimator. The bold value is the best estimator with the smallest MSE.

## B Comparison of Residual Entropy Estimators

	Intergrated Mean Squared Error (IMSE)			
	Interval	$\hat{H}^{Pois}(X,t)$	$\hat{H}_1^{Belzunce}(X,t)$	$\hat{H}^{Quantile}(X,t)$
Exp(1)	(0,4.605)	<b>0.608</b>	2.5816	9.8387
Exp(10)	(0,0.461)	0.2535	1.3814	<b>0.0719</b>
Gamma(7.5,1)	(0,15.289)	<b>0.9315</b>	1.5832	14.4769
Gamma(2,2)	(0,3.319)	<b>0.1993</b>	0.9969	3.1009
LogNormal(0,2)	(0,3.200)	2.2236	<b>0.9033</b>	2.9778
LogNormal(0,0.5)	(0,3.200)	<b>0.2631</b>	2.188	3.1081
Weibull(0.5,0.5)	(0,10.604)	<b>15.6378</b>	>10000	>10000
Weibull(2,2)	(0,4.292)	<b>0.2945</b>	0.3276	1.797
Maxwell(1)	(0,3.368)	0.4408	<b>0.1853</b>	0.6615
Maxwell(20)	(0,0.753)	0.7173	0.0379	<b>0.0367</b>
Pareto(2)	(1,10.024)	<b>19.6957</b>	6509.037	181.6876
MixGamma	(0,14.130)	1.1988	<b>0.8582</b>	17.3737
MixLogNormal	(0,3.200)	<b>2.5622</b>	3.2082	3.6732
MixWeibull	(0,4.292)	2.3112	<b>0.7388</b>	3.9595
MixMaxwell	(0,3.136)	0.365	<b>0.1949</b>	0.9866

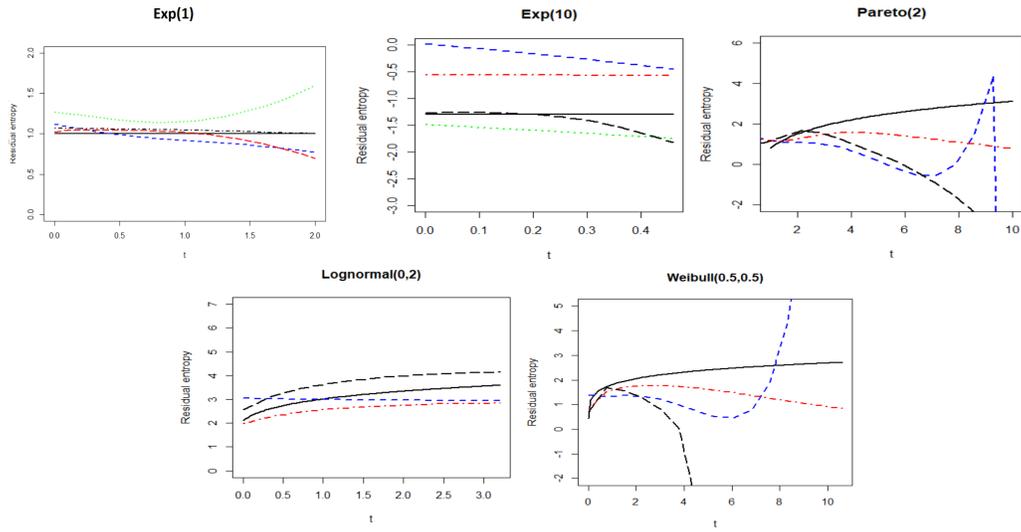
**Table B1:** MISEs of three estimators on 500 replications with sample size  $n = 50$ . The MISEs are computed on the indicated interval for each distribution.

	Intergrated Mean Squared Error (IMSE)			
	Interval	$\hat{H}^{Pois}(X,t)$	$\hat{H}_1^{Belzunce}(X,t)$	$\hat{H}^{Quantile}(X,t)$
Exp(1)	(0,4.605)	<b>0.4124</b>	2.4588	4.5337
Exp(10)	(0,0.461)	0.1674	0.2396	<b>0.0362</b>
Gamma(7.5,1)	(0,15.289)	<b>0.6</b>	0.944	5.9302
Gamma(2,2)	(0,3.319)	<b>0.1505</b>	0.6284	1.5077
LogNormal(0,2)	(0,3.200)	<b>0.7871</b>	0.9182	1.7078
LogNormal(0,0.5)	(0,3.200)	<b>0.1826</b>	0.8065	1.4114
Weibull(0.5,0.5)	(0,10.604)	<b>9.5602</b>	40.3328	62.3692
Weibull(2,2)	(0,4.292)	0.2113	<b>0.2012</b>	0.7826
Maxwell(1)	(0,3.368)	0.3158	<b>0.1346</b>	0.3129
Maxwell(20)	(0,0.753)	0.5532	<b>0.03</b>	0.031
Pareto(2)	(1,10.024)	<b>13.5616</b>	26773.74	66.4542
MixGamma	(0,14.130)	<b>0.7175</b>	0.8547	6.9898
MixLogNormal	(0,3.200)	<b>1.0251</b>	2.8335	2.8198
MixWeibull	(0,4.292)	2.5913	<b>0.9241</b>	4.6132
MixMaxwell	(0,3.136)	<b>0.258</b>	0.2852	0.4339

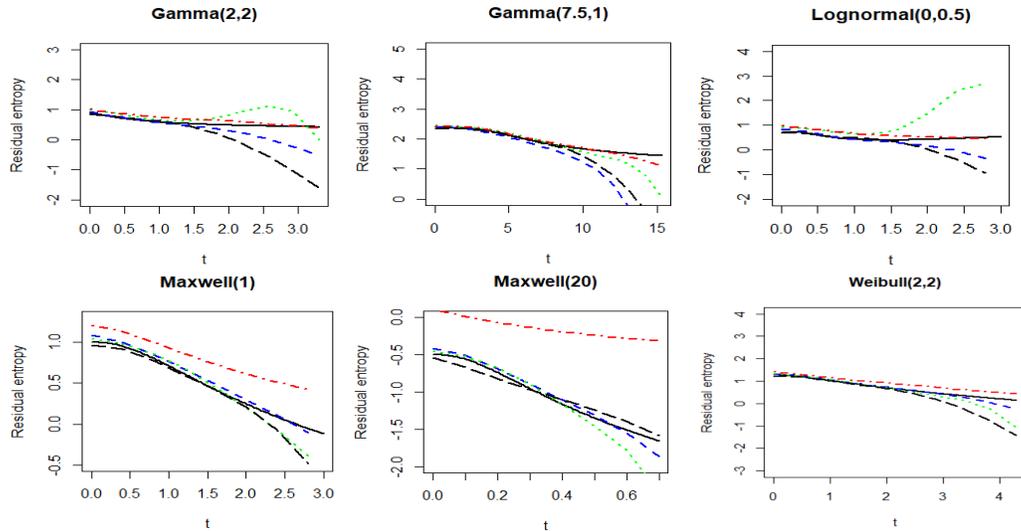
**Table B2:** MISEs of three estimators on 500 replications with sample size  $n = 100$ . The MISEs are computed on the indicated interval for each distribution.

	Intergrated Mean Squared Error (IMSE)			
	Interval	$\hat{H}^{Pois}(X,t)$	$\hat{H}_1^{Belzunce}(X,t)$	$\hat{H}^{Quantile}(X,t)$
Exp(1)	(0,4.605)	<b>0.1013</b>	0.7366	0.4834
Exp(10)	(0,0.461)	0.0583	0.06	<b>0.0099</b>
Gamma(7.5,1)	(0,15.289)	<b>0.1677</b>	0.3434	0.5875
Gamma(2,2)	(0,3.319)	<b>0.0537</b>	0.1211	0.1674
LogNormal(0,2)	(0,3.200)	0.3641	0.7894	<b>0.2377</b>
LogNormal(0,0.5)	(0,3.200)	<b>0.0722</b>	0.1486	0.2194
Weibull(0.5,0.5)	(0,10.604)	<b>1.8837</b>	28.6434	3.6078
Weibull(2,2)	(0,4.292)	0.0843	<b>0.0597</b>	0.124
Maxwell(1)	(0,3.368)	0.1273	<b>0.0356</b>	0.0548
Maxwell(20)	(0,0.753)	0.2719	<b>0.0066</b>	0.0181
Pareto(2)	(1,10.024)	<b>3.3839</b>	>1000	3.4142
MixGamma	(0,14.130)	<b>0.1736</b>	0.8307	0.5792
MixLogNormal	(0,3.200)	<b>0.4343</b>	1.9141	0.5238
MixWeibull	(0,4.292)	3.02	<b>0.8817</b>	6.2564
MixMaxwell	(0,3.136)	0.103	0.1164	<b>0.0786</b>

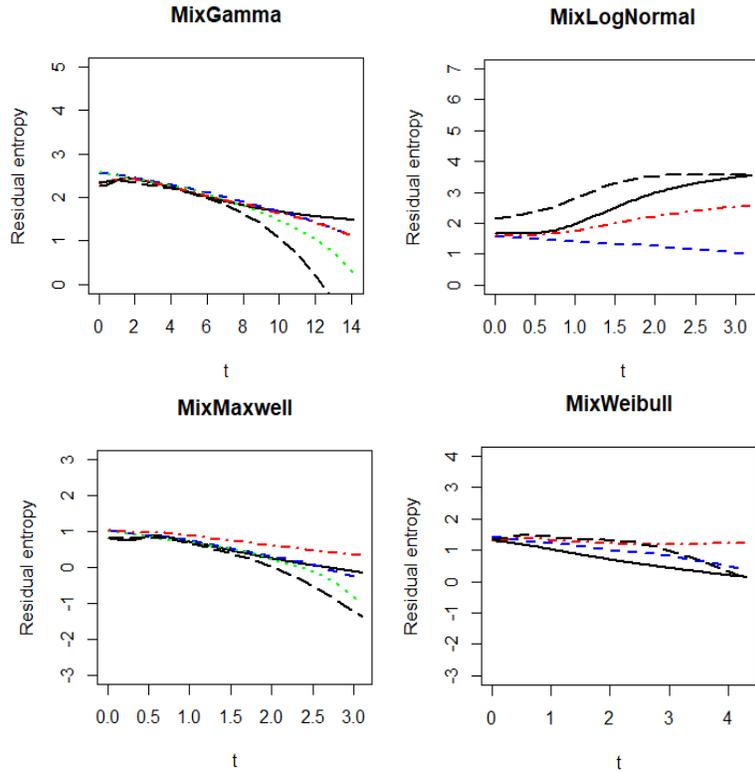
**Table B3:** MISEs of three estimators on 100 replications with sample size  $n = 500$ . The MISEs are computed on the indicated interval for each distribution.



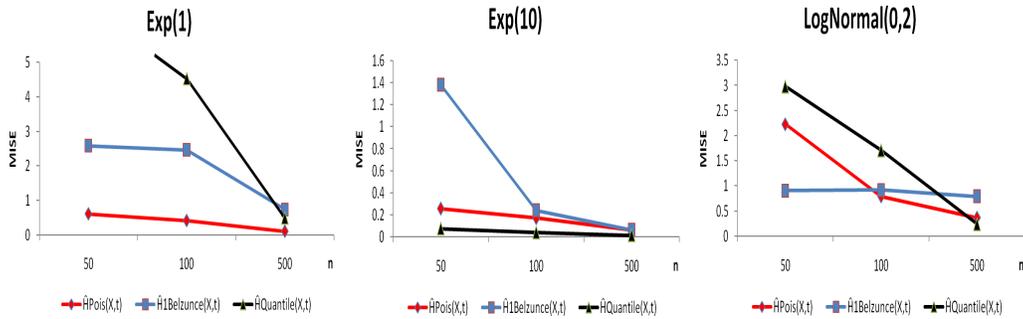
**Figure B1:** Plots of residual entropy estimators for  $n = 50$  for monotone densities. The true function is in black solid line,  $\hat{H}^{Pois}(X, t)$  in red dotted-dashed line,  $\hat{H}^{Quantile}(X, t)$  in black long dashed line,  $\hat{H}_1^{Belzunce}(X, t)$  in blue dashed line, and  $\hat{H}_2^{Belzunce}(X, t)$  in green dotted line.



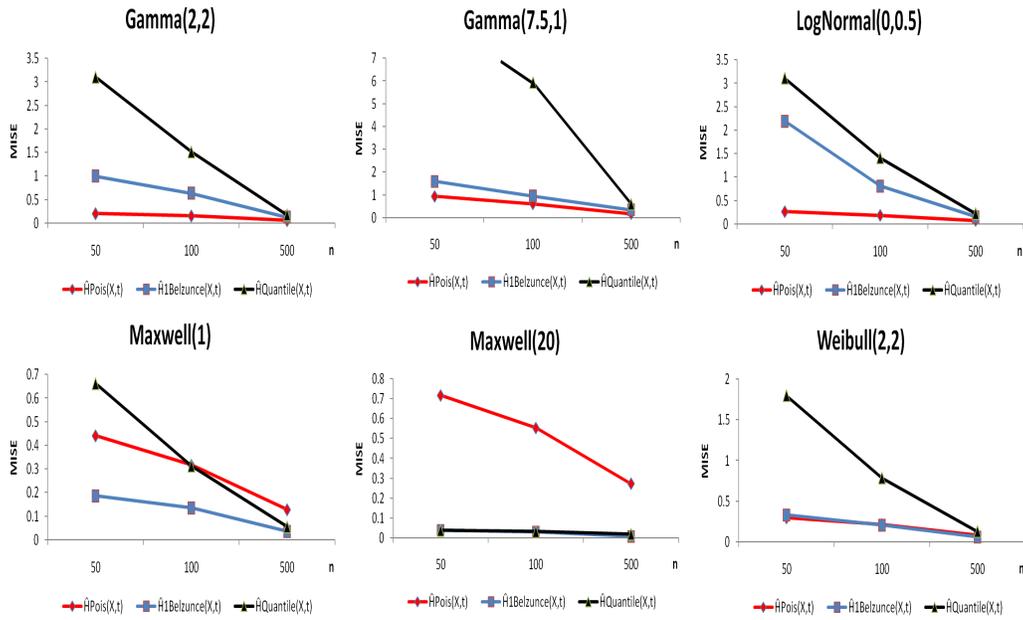
**Figure B2:** Plots of residual entropy estimators for  $n = 50$  for uni-modal densities. The true function is in black solid line,  $\hat{H}^{Pois}(X, t)$  in red dotted-dashed line,  $\hat{H}^{Quantile}(X, t)$  in black long dashed line,  $\hat{H}_1^{Belzunce}(X, t)$  in blue dashed line, and  $\hat{H}_2^{Belzunce}(X, t)$  in green dotted line.



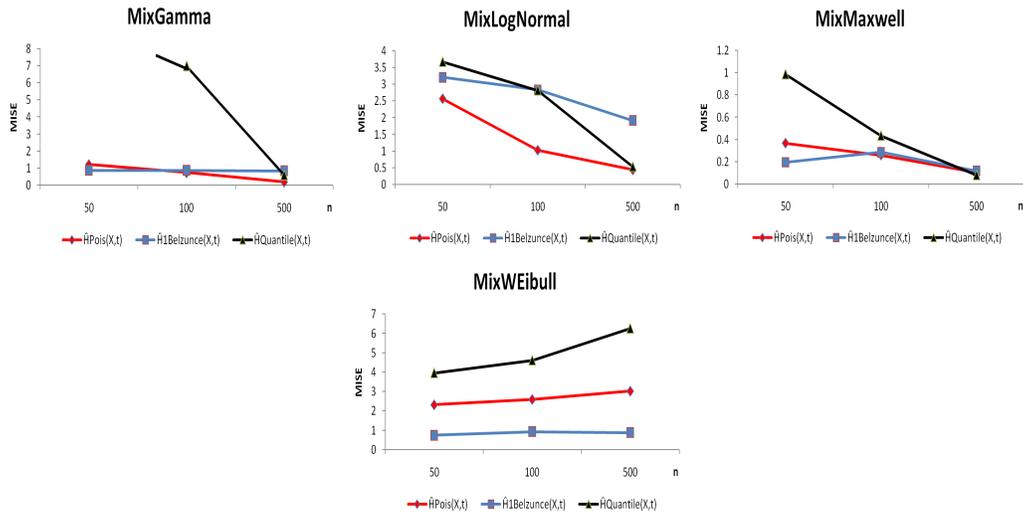
**Figure B3:** Plots of residual entropy estimators for  $n = 50$  for mixture densities. The true function is in black solid line,  $\hat{H}^{Pois}(X, t)$  in red dotted-dashed line,  $\hat{H}^{Quantile}(X, t)$  in black long dashed line,  $\hat{H}_1^{Belzunce}(X, t)$  in blue dashed line, and  $\hat{H}_2^{Belzunce}(X, t)$  in green dotted line.



**Figure B4:** Residual entropy estimation MISE comparison with different sample size  $n = 50, 100$  and  $500$ , for monotone densities, where diamond-red is  $\hat{H}^{Pois}(X, t)$ , square-blue is  $\hat{H}_1^{Belzunce}(X, t)$ , and triangle-black is  $\hat{H}^{Quantile}(X, t)$ .



**Figure B5:** Residual entropy estimation MISE comparison with different sample size  $n = 50, 100$  and  $500$ , for uni-modal densities, where diamond-red is  $\hat{H}^{Pois}(X, t)$ , square-blue is  $\hat{H}^{Belzunce}(X, t)$ , and triangle-black is  $\hat{H}^{Quantile}(X, t)$ .



**Figure B6:** Residual entropy estimation MISE comparison with different sample size  $n = 50, 100$  and  $500$  for mixture densities, where diamond-red is  $\hat{H}^{Pois}(X, t)$ , square-blue is  $\hat{H}^{Belzunce}(X, t)$ , and triangle-black is  $\hat{H}^{Quantile}(X, t)$ .

## Acknowledgments

The authors thank an anonymous referee for constructive comments which have led to an improved version of the paper. The first author would like to acknowledge the partial support for this research from NSERC of Canada through a Discovery Grant.

## References

- Ahmad, I. A. and Lin, P. E. (1976), "A non-parametric estimation of the entropy for absolutely continuous distribution," *IEEE Transactions on Information Theory*, 22, 372–375.
- Antos, A. and Kontoyiannis, I. (2001), "Convergence properties of functional estimates for discrete distributions," *Random Structures and Algorithms*, 19, 163–193.
- Belzunce, F., Guillamon, A., Navarro, J., and Ruiz, J. M. (2001), "Kernel estimation of residual entropy," *Communications in Statistics-Theory and Methods*, 30, 1243–1255.
- Bouezmarni, T. and Scaillet, O. (2005), "Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data," *Econometric Theory*, 21, 390–412.
- Bouzebda, S., Elhattab, I., Keziou, A., and Lounis, T. (2013), "New entropy estimator with an application to test of normality," *Communications in Statistics-Theory and Methods*, 42, 2245–2270.
- Chaubey, Y. P., Dewan, I., and Li, J. (2011), "Smooth estimation of survival and density functions for a stationary associated process using Poisson weights," *Statistics and Probability Letters*, 81, 267–276.
- Chaubey, Y. P., Li, J., Sen, A., and Sen, P. K. (2012), "A new smooth density estimator for non-negative random variables," *Journal of the Indian Statistical Association*, 50, 83–104.
- Chaubey, Y. P., Mudholkar, G. S., and Smethurst, P. A. (1993), "On entropy-based goodness-of-fit tests: a practical strategy," in *Probability and Statistics*, eds. Basu, S. K. and Sinha, B. K., New Delhi, India: Narosa Publishing House, pp. 116–120.
- Chaubey, Y. P. and Sen, P. K. (2009), "On the selection of the smoothing parameter in Poisson smoothing of histogram estimator: Computational aspects," *Pakistan Journal of Statistics*, 25, 385–401.
- 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.
- Chen, S. X. (2010), "Probability density function estimation using gamma kernels," *Annals of the Institute of Statistical Mathematics*, 52, 471–480.

- Cheng, C. and Parzen, E. (2010), "Unified estimators of smooth quantile and quantile density functions," *Journal of Statistical Planning and Inference*, 59, 291–307.
- Correa, J. C. (1995), "A new estimator of entropy," *Communications in Statistics: Theory and Methods*, 24, 2439–2449.
- Dmitriev, Y. G. and Tarasenko, F. P. (1973), "On the estimation functions of the probability density and its derivatives," *Theory of Probability and its Applications*, 18, 628–633.
- Ebrahimi, N. (1995), "How to measure uncertainty in the residual life time distribution," *Sankhya, The Indian Journal of Statistics*, A58, 48–56.
- Ebrahimi, N., Habibullah, M., and Soofi, E. (1992), "Testing exponentiality based on Kullback-Leibler information," *Journal of the Royal Statistical Society*, B54, 739–748.
- Eggermont, P. B. and LaRiccia, V. N. (1999), "Best asymptotic normality of the kernel density entropy estimator for smooth densities," *IEEE Transactions On Information Theory*, 45, 1321–1326.
- Györfi, L. and van der Meulen, E. C. (1987), "Density-free convergence properties of various estimators of entropy," *Computational Statistics and Data Analysis*, 5, 425–436.
- Hall, P. and Morton, S. C. (1993), "On the estimation of entropy," *Annals of Institute of Statistical Mathematics*, 45, 69–88.
- Joe, H. (1989), "On the estimation of entropy and other functionals of a multivariate density," *Annals of Institute of Statistical Mathematics*, 41, 683–697.
- Kapur, J. N. (1993), *Maximum Entropy Models in Science and Engineering*, New York, USA: John Wiley and Sons, Inc.
- Kapur, J. N. and Kesavan, H. K. (1992), *Entropy optimization principles with applications*, New York, USA: Academic Press Inc.
- Noughabi, H. A. (2010), "A new estimator of entropy and its application in testing normality," *Journal of Statistical Computation And Simulation*, 80, 1151–1162.
- Paninski, L. (2003), "Estimation of entropy and mutual information," *Neural Computation*, 15, 1191–1253.
- Panzeri, S. and Treves, A. (1996), "Analytical estimates of limited sampling biases in different information measures," *Network: Computation in Neural Systems*, 7, 87–107.
- Rosenblatt, M. (1966), "Remarks on some non-parametric estimates of a density function," *Annals of Mathematics Statistics*, 27, 832–837.
- Shannon, C. (1948), "A mathematical theory of communication," *Bell System Technical Journal*, 27, 379–423 & 623–656.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London, UK: Chapman and Hall.

Strong, S., Koberle, R., de Ruyter van Steveninck, R., and Bialek, W. (1998), "Entropy and information in neural spike trains," *Physical Review Letters*, 80, 197–202.

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

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

Wieczorkowski, R. and Grzegorzewski, P. (1999), "Entropy estimators improvements and comparisons," *Communications in Statistics-Simulation and Computation*, 28, 541–567.

*Received: July 28, 2020*

*Accepted: December 20, 2020*