

# Entropy Estimation Using Numerical Methods

Hadi Alizadeh Noughabi<sup>1</sup>

Received: 18 June 2015 / Revised: 5 October 2015 / Accepted: 6 October 2015 /  
Published online: 16 October 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** Direct integration of the Riemann–Stieltjes integral has been used to computing convolution integrals. This approach has been established to be simple and accurate with good convergence property. In this paper, we used some numerical methods to estimation of entropy of a continuous random variable and then some estimators are introduced. Bounds on the error terms are derived for some direct Riemann–Stieltjes integration methods. Consistency of estimators is proved and by simulation, the proposed estimators are compared with some prominent estimators, namely Correa (Commun Stat Theory Methods 24:2439–2449, 1995), Ebrahimi et al. (Stat Probab Lett 20:225–234, 1994), van Es (Scand J Stat 19:61–72, 1992) and Vasicek (J R Stat Soc B 38:54–59, 1976). The results indicate that the proposed estimators have smaller mean squared error than other estimators.

**Keywords** Entropy estimator · Riemann–Stieltjes integration · Midpoint rectangle RS-integration · Mean-value rectangle RS-integration · Generalized Simpson rule

## 1 Introduction

The entropy  $H(f)$  of the random variable  $X$  with distribution function  $F(x)$  and continuous density function  $f(x)$  is defined by [23] to be

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

---

✉ Hadi Alizadeh Noughabi  
alizadehhadi@birjand.ac.ir

<sup>1</sup> Department of Statistics, University of Birjand, Birjand, Iran

Estimation of entropy from a random sample has been considered by many authors. For discrete random variables, References [8, 15, 29] have proposed estimators of entropy, while [5, 9, 11, 14, 16, 27, 28] have proposed solutions for the problem of estimating the entropy for continuous random variables.

Vasicek [28] expressed (1.1) as

$$H(f) = \int_0^1 \log \left\{ \frac{d}{dp} F^{-1}(p) \right\} dp, \tag{1.2}$$

and then by replacing the distribution function  $F$  by the empirical distribution function  $F_n$ , and using a difference operator instead of the differential operator proposed an entropy estimator. He also estimated the derivative of  $F^{-1}(p)$  by a function of the order statistics. Assuming that  $X_1, \dots, X_n$  is a random sample, Vasicek’s estimator is

$$HV_{mn} = \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right\}, \tag{1.3}$$

where the window size  $m$  is a positive integer smaller than  $n/2$ ,  $X_{(i)} = X_{(1)}$  if  $i < 1$ ,  $X_{(i)} = X_{(n)}$  if  $i > n$  and  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  are order statistics based on a random sample of size  $n$ . Vasicek proved that his estimator is consistent, i.e.,  $HV_{mn} \xrightarrow{\text{Pr.}} H(f)$  as  $n \rightarrow \infty, m \rightarrow \infty, \frac{m}{n} \rightarrow 0$ .

Van Es [27] using spacings introduced an estimator of entropy and proved the consistency and asymptotic normality of this estimator under some conditions. Van Es’ estimator is given by

$$HVE_{mn} = \frac{1}{n-m} \sum_{i=1}^{n-m} \left( \frac{n+1}{m} (X_{(i+m)} - X_{(i)}) \right) + \sum_{k=m}^n \frac{1}{k} + \log(m) - \log(n+1).$$

Ebrahimi et al. [9] modified Vasicek’s estimator and then proposed the following estimator.

$$HE_{mn} = \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{n}{c_i m} (X_{(i+m)} - X_{(i-m)}) \right\},$$

where

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

They showed that  $HE_{mn} \xrightarrow{\text{Pr.}} H(f)$  as  $n \rightarrow \infty, m \rightarrow \infty, m/n \rightarrow 0$ . Then they showed that their estimator has smaller bias and mean squared error than Vasicek’s estimator, by simulation.

Correa [5] proposed a modification of Vasicek estimator as follows:

He considered the sample information as

$$(F_n(X_{(1)}), X_{(1)}), (F_n(X_{(2)}), X_{(2)}), \dots, (F_n(X_{(n)}), X_{(n)}),$$

and written Eq. (1.2) as

$$HV_{mn} = -\frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{(i+m)/n - (i-m)/n}{X_{(i+m)} - X_{(i-m)}} \right\}.$$

Then he noted that the argument of log is the equation of the slope of the straight line that joins the points  $(F_n(X_{(i+m)}), X_{(i+m)})$  and  $(F_n(X_{(i-m)}), X_{(i-m)})$ , and therefore applied a local linear model based on  $2m + 1$  points to estimate the density of  $F(x)$  in the interval  $(X_{(i+m)}, X_{(i-m)})$ ,

$$F(x_{(j)}) = \alpha + \beta x_{(j)} + \varepsilon \quad j = m - i, \dots, m + i.$$

Via the least square method, he proposed an estimator of entropy as

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

where

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

He compared his estimator with Vasicek’s and van Es’s estimators and concluded the mean square error (MSE) of his estimator is smaller than the MSE of Vasicek’s estimator. Also for some  $m$  his estimator behaves better than van Es estimator.

Entropy estimators are used to developing entropy-based statistical procedures. See for example, [1, 2, 10, 13, 20, 21]. Therefore, new entropy estimators can be useful in practice.

For many computational problems in applied probability and statistics, we have to compute the Riemann–Stieltjes integral of the following form

$$\int_a^b f(x)dg(x), \tag{1.4}$$

where the function  $g(x)$  is usually a distribution function.

Direct integration of the Riemann–Stieltjes integral can be used to computing convolution integrals. This approach has been established to be simple and accurate with good convergence property.

The above integral can be approximated directly using the definition of the Riemann–Stieltjes integral. The function  $f(x)$  may approximate by a piecewise constant function  $\tilde{f}(x)$  and by transforming the integration to a summation, we have

$$\int_a^b f(x)dg(x) \approx \sum_{i=1}^n \tilde{f}(x_i) \cdot (g(x_i) - g(x_{i-1})).$$

Therefore, a numerical algorithm can be obtained based on the above formula.

This procedure is used by [30] in solving renewal equation. Independently, a similar idea is used by [26]. Numerical results in these papers show that direct RS-integration is very simple and accurate compared with the existing algorithms ([6, 19], for example) that are more complicated. Some applications of this approach can be found in [4, 17, 25]. Xie et al. [31] developed further methodologies of the Riemann–Stieltjes integration under general conditions.

In Sect. 2, some direct integration methods and the bounds on the errors are given. In Sect. 3, numerical methods are used to estimation of entropy of continuous random variables. We show that the proposed estimators are consistent. Scale invariance of variance and mean squared error of the estimators are established. In the last section we present results of a comparison of the proposed estimators with the competing estimators by a simulation study.

## 2 Some Direct Integration Methods

Let the interval  $I = [a, b]$  be partitioned into  $n$  equal length intervals denoted by  $I_i = [x_{i-1}, x_i]$ ,  $i = 1, \dots, n$  where  $x_0 = a$  and  $x_i = x_{i-1} + h$ , where  $h = (b - a)/n$  is the length of each interval. By using the direct Riemann–Stieltjes integration method ([30]), we approximate the function  $f(x)$  with a function  $\tilde{f}(x)$  and then use the definition of the Riemann–Stieltjes integral. Similar to the case of rectangle and trapezoidal rules for the Riemann integral, Xie et al. [31] distinguished between the midpoint and mean value RS-integration method, and described these methods as follows.

### 2.1 The Midpoint Rectangle RS-Integration

By the rectangle RS-integration method the function  $f(x)$  is approximated by its value at the midpoint

$$x_{i-.5} = \frac{x_{i-1} + x_i}{2},$$

of each interval. Then

$$\sum_{i=1}^n \int_{x_{i-1}}^{x_i} f(x)dg(x) \approx \sum_{i=1}^n f(x_{i-.5}) (g(x_i) - g(x_{i-1})).$$

Numerical accuracy and convergence are demonstrated in [30].

### 2.2 The Mean-Value Rectangle RS-Integration

Another method to approximate  $f(x)$  is to use the mean value of  $f(x)$  at each interval. We approximate the function  $f(x)$  at the  $i$ th interval by the mean value of the endpoints  $(f(x_i) + f(x_{i-1}))/2$ . In this case, the Eq. (1.4) is approximated by

$$\begin{aligned}
 Q_n &= \sum_{i=1}^n \int_{x_{i-1}}^{x_i} \left( \frac{f(x_{i-1}) + f(x_i)}{2} \right) dg(x) \\
 &= \sum_{i=1}^n \left( \frac{f(x_{i-1}) + f(x_i)}{2} \right) (g(x_i) - g(x_{i-1})). \tag{2.1}
 \end{aligned}$$

Since the mean value of  $f(x)$  at points  $x_i$  and  $x_{i-1}$  is used in approximating  $f(x)$ , we call the mean value rectangle RS-integration method or the generalized trapezoidal rule. For the Riemann integral, this method is usually called the trapezoidal rule since it is equivalent to the approximating  $f(x)$  by a piecewise linear function. This is not the case for Riemann–Stieltjes integral unless  $g(x)$  is a linear function.

This method is a simple method that gives satisfactory results and it has been used by [3, 26].

### 2.3 The Generalized Simpson Rule

Xie et al. [31] generalized other integration method to Riemann–Stieltjes equation (1.4) which is called the generalized Simpson rule.

Let the interval  $I = [a, b]$  be divided into  $2n$  equal parts of length and let  $h = (b - a)/2n$ . Now the Eq. (1.4) approximated by

$$\int_a^b f(x)dg(x) \approx \frac{4Q_{2n}}{3} - \frac{Q_n}{3},$$

where  $Q_n$  is given by the mean-value rectangle equation (2.1) with respect to  $n$  subintervals of length  $(b - a)/n$  and  $Q_{2n}$  is the rule (2.1) for  $2n$  subintervals of length  $(b - a)/2n$ . Then

$$\begin{aligned}
 \int_a^b f(x)dg(x) &\approx \frac{4}{3} \sum_{i=1}^{2n} \int_{x_{i-1}}^{x_i} \left( \frac{f(x_{i-1}) + f(x_i)}{2} \right) (g(x_i) - g(x_{i-1})) \\
 &\quad - \frac{1}{3} \sum_{i=1}^n \left( \frac{f(x_{2i-2}) + f(x_{2i})}{2} \right) (g(x_{2i}) - g(x_{2i-2})) \\
 &= \frac{1}{6} \sum \{ (-3g(x_{2i-2}) + 4g(x_{2i-1}) - g(x_{2i})) f(x_{2i-2}) - 4(g(x_{2i-2}) \\
 &\quad - g(x_{2i})) f(x_{2i-1}) + (g(x_{2i-2}) - 4g(x_{2i-1}) + 3g(x_{2i})) f(x_{2i}) \},
 \end{aligned}$$

is considered as generalized Simpson rule.

## 2.4 General Bounds on the Truncation Errors

Numerical results show that all the mentioned methods can be used successfully, see, e.g., [4, 7, 26, 30]. In [3] some convergence results are presented and then [31] derived some further bounds on the truncation errors of the method under very general assumptions. They only assumed that  $g(x)$  is an increasing function which is the usual assumption used for defining the Riemann–Stieltjes integral ([22], p. 122). The following theorems have proved by [31].

**Theorem 1** *For the midpoint method, the global truncation error  $|\varepsilon|$  may be bounded by*

$$(h/2) \sup_{x \in I} |f'(x)|(g(b) - g(a)),$$

*assuming  $f(x)$  is continuously differentiable and  $g(x)$  is an increasing function.*

*Remark* The convergence of this method is thus of order 1. As the number of intervals  $n$  increases, the error will decrease at least as fast as  $1/n$ .

**Theorem 2** *Under the same conditions as in Theorem 1, the global truncation error may be bounded by*

$$h \sup_{x \in I} |f'(x)|(g(b) - g(a)),$$

*for the mean value method.*

**Theorem 3** *Under the same conditions as in Theorem 1, the global truncation error  $|\varepsilon|$  for the generalized Simpson rule may be bounded by*

$$2h \sup_{x \in I} |f'(x)|(g(b) - g(a)).$$

## 3 Estimation of Entropy

Suppose  $X_1, \dots, X_n$  are order statistics of a random sample of size  $n$  from an unknown continuous distribution  $F$  with a probability density function  $f(x)$ . We use the methods discussed in previous section to estimate the entropy  $H(f)$  of an unknown continuous probability density function  $f$ . Using these methods the following approximations for entropy can be derived.

$$(1) \quad H(f) = - \int \log f(x) dF(x) \approx - \sum_{i=1}^n \log f(x_i) \{F(x_i) - F(x_{i-1})\} = H_1,$$

$$(2) \quad H(f) \approx - \sum_{i=1}^n \log f\left(\frac{x_i + x_{i-1}}{2}\right) \{F(x_i) - F(x_{i-1})\} = H_2,$$

$$(3) H(f) \approx - \sum_{i=1}^n \frac{\log f(x_{i-1}) + \log f(x_i)}{2} \{F(x_i) - F(x_{i-1})\} = H_3,$$

$$(4) H(f) \approx \frac{4Q_{2n}}{3} - \frac{Q_n}{3} = H_4.$$

Now, we replace  $F(x_i)$  and  $f(x_i)$  with  $i/n$  (the empirical distribution function) and  $\hat{f}(x_i)$  (the kernel density estimate), respectively. Therefore, we obtain

$$H_1 = -\frac{1}{n} \sum_{i=1}^n \log \hat{f}(x_i),$$

$$H_2 = -\frac{1}{n} \sum_{i=1}^n \log \hat{f}\left(\frac{x_i + x_{i-1}}{2}\right),$$

$$H_3 = -\frac{1}{n} \sum_{i=1}^n \frac{\log \hat{f}(x_{i-1}) + \log \hat{f}(x_i)}{2},$$

$$H_4 = -\frac{1}{6n} \sum_{i=1}^n \left\{ \log \hat{f}(x_{2i-2}) + 4 \log \hat{f}(x_{2i-1}) + \log \hat{f}(x_{2i}) \right\}.$$

The kernel density function estimator is well-known and is defined by

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

where  $h$  is a bandwidth and  $k$  is a kernel function which satisfies

$$\int_{-\infty}^{\infty} k(x)dx = 1.$$

Usually,  $k$  will be a symmetric probability density function (see, [24]).

Here, the kernel function is chosen the standard normal density function and the bandwidth  $h$  is chosen the normal optimal smoothing formula,  $h = 1.06sn^{-\frac{1}{5}}$ , where  $s$  is the sample standard deviation.

Using Theorems 1–3, the global truncation error  $|\varepsilon|$  can be obtained as follows.

We have

$$f(x) = -\log(\hat{f}(x)) \quad \text{and} \quad g(x) = F_n(x),$$

therefore,

$$M = \sup_{x \in I} |f'(x)| = \sup_{x \in [x_1, x_n]} \left| -\frac{\hat{f}'(x)}{\hat{f}(x)} \right| = \sup_{x \in [x_1, x_n]} \left| -\frac{1}{h} \frac{\sum_{j=1}^n k'\left(\frac{x-x_j}{h}\right)}{\sum_{j=1}^n k\left(\frac{x-x_j}{h}\right)} \right|.$$

So, for the midpoint, mean value and generalized Simpson methods, the global truncation error  $|\varepsilon|$  may be bounded by

$$\frac{x_n - x_1}{2n} M \left(1 - \frac{1}{n}\right), \quad \frac{x_n - x_1}{n} M \left(1 - \frac{1}{n}\right), \quad \frac{x_n - x_1}{2n} M \left(1 - \frac{1}{n}\right),$$

respectively.

The scale of the random variable  $X$  has no effect on the accuracy of  $H_1, H_2, H_3$  or  $H_4$  in estimating  $H(f)$ . The following theorem shows this fact. Similar results have been obtained for  $HV_{mn}$  and  $HE_{mn}$  by [18] and [9], respectively.

**Theorem 3.1** *Let  $X_1, \dots, X_n$  be a sequence of i.i.d. random variables with entropy  $H^X(f)$  and let  $Y_i = cX_i, i = 1, \dots, n$ , where  $c > 0$ . Let  $H_1^X$  and  $H_1^Y$  be entropy estimators for  $H^X(f)$  and  $H^Y(g)$ , respectively (here  $g$  is pdf of  $Y = cX$ ). Then the following properties hold.*

- (i)  $E(H_1^Y) = E(H_1^X) + \log c$ ,
- (ii)  $Var(H_1^X) = Var(H_1^Y)$ ,
- (iii)  $MSE(H_1^X) = MSE(H_1^Y)$ .

*Proof* We have

$$\hat{f}(cX_i) = \frac{1}{nh_y} \sum_{j=1}^n k \left( \frac{cX_i - cX_j}{h_y} \right) = \frac{1}{nch_x} \sum_{j=1}^n k \left( \frac{cX_i - cX_j}{ch_x} \right) = \frac{\hat{f}(X_i)}{c},$$

where  $h_y = 1.06s_y n^{-1/5} = 1.06cs_x n^{-1/5} = ch_x$ .

Therefore,  $H_1^Y = H_1^X + \log c$ , and the theorem is established.

The above theorem is hold for the other proposed estimators.

The following theorem establishes the consistency of estimators.

**Theorem 3.2** *Let  $C$  be the class of continuous densities with finite entropies and let  $X_1, \dots, X_n$  be a random sample from  $f \in C$ . If  $n \rightarrow \infty$ , then*

$$H_i \xrightarrow{\text{Pr.}} H(f) \quad i = 1, \dots, 4.$$

*Proof* It is obvious by consistency of  $\hat{f}(x)$  and continuity of  $f(x)$ .

### 4 Simulation Study

A simulation study is carried out to analyze the behavior of the proposed entropy estimators,  $H_i, i = 1, 2, 3, 4$ . The proposed estimators are compared with some prominent estimators, namely Vasicek’s estimator [28], van Es’s estimator [27], Ebrahimi et al.’s estimator [9] and Correa’s estimator [5]. For comparisons, we consider normal, exponential and uniform distributions which are the same three distributions



**Table 1** Root of mean square error (and standard deviation) of estimators in estimate of entropy  $H(f)$  for standard normal distribution

$n$	$m$	RMSE (SD)							
		$HV_{mn}$	$HVE_{mn}$	$HC_{mn}$	$HE_{mn}$	$H_1$	$H_2$	$H_3$	$H_4$
5	2	0.994	0.509	0.793	0.665	0.457	0.511	0.501	0.553
		(0.413)	(0.452)	(0.418)	(0.424)	(0.397)	(0.405)	(0.408)	(0.413)
10	3	0.618	0.366	0.470	0.402	0.287	0.317	0.318	0.327
		(0.264)	(0.283)	(0.271)	(0.265)	(0.256)	(0.254)	(0.254)	(0.259)
15	4	0.474	0.318	0.348	0.301	0.221	0.250	0.249	0.278
		(0.211)	(0.220)	(0.213)	(0.211)	(0.186)	(0.200)	(0.201)	(0.207)
20	4	0.373	0.276	0.265	0.247	0.186	0.208	0.215	0.205
		(0.178)	(0.185)	(0.182)	(0.178)	(0.170)	(0.168)	(0.174)	(0.171)
30	5	0.282	0.243	0.194	0.186	0.149	0.169	0.168	0.163
		(0.144)	(0.148)	(0.146)	(0.142)	(0.136)	(0.138)	(0.136)	(0.138)
50	7	0.198	0.212	0.135	0.128	0.110	0.125	0.126	0.120
		(0.109)	(0.110)	(0.112)	(0.109)	(0.102)	(0.103)	(0.104)	(0.105)

**Table 2** Root of mean square error (and standard deviation) of estimators in estimate of entropy  $H(f)$  for exponential distribution with mean one

$n$	$m$	RMSE (SD)							
		$HV_{mn}$	$HVE_{mn}$	$HC_{mn}$	$HE_{mn}$	$H_1$	$H_2$	$H_3$	$H_4$
5	2	0.930	0.596	0.743	0.652	0.569	0.583	0.577	0.592
		(0.559)	(0.586)	(0.554)	(0.552)	(0.569)	(0.578)	(0.572)	(0.563)
10	3	0.570	0.392	0.435	0.401	0.390	0.386	0.378	0.376
		(0.360)	(0.373)	(0.361)	(0.358)	(0.378)	(0.381)	(0.375)	(0.375)
15	4	0.421	0.310	0.328	0.310	0.328	0.325	0.313	0.297
		(0.282)	(0.290)	(0.290)	(0.290)	(0.301)	(0.306)	(0.302)	(0.297)
20	4	0.356	0.274	0.272	0.261	0.294	0.287	0.276	0.271
		(0.242)	(0.250)	(0.247)	(0.242)	(0.259)	(0.261)	(0.258)	(0.256)
30	5	0.276	0.227	0.208	0.203	0.256	0.253	0.238	0.232
		(0.198)	(0.201)	(0.197)	(0.193)	(0.204)	(0.211)	(0.205)	(0.203)
50	7	0.194	0.179	0.155	0.151	0.223	0.218	0.204	0.202
		(0.148)	(0.149)	(0.152)	(0.148)	(0.159)	(0.162)	(0.157)	(0.155)

considered in [5]. For each sample size 10,000 samples were generated and the RMSEs of the estimators are computed.

For competitor estimators, we chose value of  $m$  using the following heuristic formula (see [12]):

$$m = [\sqrt{n} + 0.5].$$

**Table 3** Root of mean square error (and standard deviation) of estimators in estimate of entropy  $H(f)$  for uniform distribution on  $(0, 1)$ 

$n$	$m$	RMSE (SD)							
		$HV_{mn}$	$HVE_{mn}$	$HC_{mn}$	$HE_{mn}$	$H_1$	$H_2$	$H_3$	$H_4$
5	2	0.774	0.407	0.566	0.456	0.326	0.343	0.343	0.382
		(0.346)	(0.407)	(0.336)	(0.343)	(0.326)	(0.336)	(0.338)	(0.359)
10	3	0.455	0.216	0.295	0.234	0.186	0.185	0.187	0.191
		(0.166)	(0.216)	(0.169)	(0.167)	(0.172)	(0.183)	(0.184)	(0.188)
15	4	0.343	0.155	0.208	0.160	0.153	0.146	0.146	0.148
		(0.110)	(0.155)	(0.112)	(0.111)	(0.125)	(0.134)	(0.133)	(0.142)
20	4	0.274	0.121	0.157	0.135	0.139	0.129	0.130	0.134
		(0.087)	(0.121)	(0.088)	(0.087)	(0.101)	(0.108)	(0.108)	(0.106)
30	5	0.210	0.086	0.110	0.097	0.126	0.115	0.116	0.122
		(0.059)	(0.086)	(0.061)	(0.060)	(0.075)	(0.079)	(0.078)	(0.079)
50	7	0.156	0.058	0.076	0.063	0.115	0.107	0.107	0.112
		(0.037)	(0.058)	(0.039)	(0.037)	(0.052)	(0.055)	(0.055)	(0.054)

Tables 1, 2 and 3 present the RMSE values (and standard deviation) of the eight estimators at different sample size.

We can see from Tables 1, 2 and 3 that the proposed estimators compare favorably with their competitors. Also, note that the proposed estimators do not depend on  $m$ . In general, the proposed estimators have a good performance for small sample sizes.

## 5 Conclusion

In this paper, we first have described some direct integration methods and the bounds on the errors have been given. We next have introduced some estimators of entropy of a continuous random variable using numerical methods and the bounds on the errors of estimators are obtained. We then have compared the proposed estimators with some prominent existing estimators and observed that for small sample sizes the new estimators behave better than the competitors. The advantage of the proposed estimators were the fact that they do not depend on the window size  $m$ , unlike the competitor estimators.

**Acknowledgments** The author thanks the Associate Editor and anonymous referees for making some valuable suggestions which led to a considerable improvement in the presentation of this manuscript.

## References

1. Alizadeh Noughabi H, Arghami NR (2013) General treatment of goodness-of-fit tests based on Kullback–Leibler information. *J Stat Comput Simul* 83:1556–1569
2. Balakrishnan N, Habibi Rad A, Arghami NR (2007) Testing exponentiality based on Kullback–Leibler information with progressively type-II censored data. *IEEE Trans Reliab* 56:301–307

3. Boehme TK, Preuss W, van der Wall V (1991) On a simple numerical method for computing Stieltjes integrals in reliability theory. *Probab Eng Inf Sci* 5:113–128
4. Chaudhry ML (1995) On computations of the mean and variance of the number of renewals—a unified approach. *J Oper Res Soc* 46:1352–1364
5. Correa JC (1995) A new estimator of entropy. *Commun Stat Theory Methods* 24:2439–2449
6. Deligonul ZS, Bilgen S (1984) Solution of the Volterra equation of renewal theory with the Galerkin technique using cubic splines. *J Stat Comput Simul* 20:37–45
7. Den Iseger PW, Smith MAJ, Dekker R (1997) Computing compound distribution faster!. *Insur Math Econ* 20:23–34
8. Dobrushin RL (1958) Simplified method of experimental estimate of entropy of stationary sequence. *Theory Probab Appl* 3:462–464
9. Ebrahimi N, Pflughoeft K, Soofi E (1994) Two measures of sample entropy. *Stat Probab Lett* 20:225–234
10. Esteban MD, Castellanos ME, Morales D, Vajda I (2001) Monte Carlo comparison of four normality tests using different entropy estimates. *Commun Stat Simul Comput* 30:761–785
11. Gorla MN, Leonenko NN, Mergel VV, Novi Inverardi PL (2005) A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *Nonparametr Stat* 17:277–297
12. Grzegorzewski P, Wiczorkowski R (1999) Entropy-based goodness-of-fit test for exponentiality. *Commun Stat Theory Methods* 28:1183–1202
13. Habibi Rad A, Yousefzadeh F, Balakrishnan N (2011) Goodness-of-fit test based on Kullback–Leibler information for progressively Type-II censored data. *IEEE Trans Reliab* 60:570–579
14. Hall P, Morton SC (1993) On the estimation of entropy. *Ann Inst Stat Math* 45:69–88
15. Hutcheson K, Shenton LR (1974) Some moments of an estimate of Shannon’s measure of information. *Commun Stat* 3:89–94
16. Joe H (1989) Estimation of entropy and other functionals of a multivariate density. *Ann Inst Stat Math* 41:683–697
17. Kao EPC (1997) An introduction to stochastic processes. Duxbury Press, Belmont
18. Mack SP (1988) A comparative study of entropy estimators and entropy based goodness-of-fit tests. PhD Dissertation, University of California, Riverside
19. McConalogue DJ (1981) Numerical treatment of convolution integrals involving distributions with densities having singularities at the origin. *Commun Stat Simul Comput* 10:265–280
20. Pakyari R, Balakrishnan N (2012) A general purpose approximate goodness-of-fit test for progressively type-II censored data. *IEEE Trans Reliab* 61:238–244
21. Pakyari R, Balakrishnan N (2013) Goodness-of-fit tests for progressively type-II censored data from location-scale distributions. *J Stat Comput Simul* 83:167–178
22. Rudin W (1976) Principles of mathematical analysis, 3rd edn. McGraw-Hill, New York
23. Shannon CE (1948) A mathematical theory of communications. *Bell Syst Tech J* 27:379–423; 623–656
24. Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London
25. Tijms HC (1994) Stochastic models: an algorithmic approach. Wiley, New York
26. van der Wall V, Preuss W (1989) A simple numerical procedure for solving Volterra–Stieltjes integral equations. In: Proceedings of the conference on complex analysis, Sofia, pp 424–430
27. van Es B (1992) Estimating functionals related to a density by class of statistics based on spacings. *Scand J Stat* 19:61–72
28. Vasicek O (1976) A test for normality based on sample entropy. *J R Stat Soc B* 38:54–59
29. Vatutin VA, Michailov VG (1995) Statistical estimation of entropy of discrete random variables with large numbers of results. *Russ Math Surv* 50:963–976
30. Xie M (1989) On the solution of renewal-type integral equations. *Commun Stat Simul Comput* 18:281–293
31. Xie M, Preuss W, Cui L (2003) Error analysis of some integration procedures for renewal equation and convolution integrals. *J Stat Comput Simul* 73:59–70