JOURNAL OF APPLIED STATISTICS, 2017
VOL. 44, NO. 10, 1856–1874
http://dx.doi.org/10.1080/02664763.2016.1238046

Nonparametric estimation of mean residual quantile function under right censoring

P.G. Sankaran[a] and N.N. Midhu [b]

[a]Department of Statistics, Cochin University of Science and Technology, Cochin, India; [b]Stat-Math Unit, Indian Statistical Institute, New Delhi, India

ABSTRACT

In this paper, we develop non-parametric estimation of the mean residual quantile function based on right-censored data. Two non-parametric estimators, one based on the empirical quantile function and the other using the kernel smoothing method, are proposed. Asymptotic properties of the estimators are discussed. Monte Carlo simulation studies are conducted to compare the two estimators. The method is illustrated with the aid of two real data sets.

ARTICLE HISTORY
Received 24 October 2015
Accepted 14 September 2016

KEYWORDS
Non-parametric estimator; mean residual quantile function; kernel smoothing

1. Introduction

Let X be a non-negative random variable with distribution function $F(x)$. Assume that the mean of X is finite. Then the mean residual life function (MRLF) of X is defined by

$$m(x) = E(X - x \,|\, X > x) = \frac{1}{\bar{F}(x)} \int_x^{\infty} \bar{F}(t)\, dt, \tag{1}$$

where $\bar{F}(x) = 1 - F(x)$ is the survival function of X. The $m(x)$ is interpreted as the expected remaining lifetime of a unit given that it has survival up to time x. The MRLF plays a vital role in the areas of engineering, medical science, survival studies, economics and many other fields. The mean residual life function has the property that it determines the distribution uniquely, see [11,12,17,22] . Ruiz and Navarro [32] have considered the problem of characterization of the distribution function by the relationship between the mean residual life function and the hazard rate function. For more properties and applications of Equation (1), one could refer to Lai and Xie [18], Marshall and Olkin [19] and Nair and Sankaran [24].

The functions $F(x)$ and $m(x)$ have no simple and closed-form expressions for many family of distributions. Accordingly, an alternative approach, using the quantile function is introduced for modelling and analysis of lifetime data. For any right continuous distribution function $F(x)$, the quantile function $Q(u)$ is defined by

$$Q(u) = F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad 0 \leq u \leq 1. \tag{2}$$

CONTACT N.N. Midhu midhu.nellikattu@gmail.com Stat-Math Unit, Indian Statistical Institute, New Delhi110 016, India

When $F(\cdot)$ is continuous, Equation (2) gives

$$F(Q(u)) = u. \tag{3}$$

The role of quantile function and other concepts derived from it are well established in exploratory data analysis and in various fields of applied statistics [9,29].

Recently, Nair and Sankaran [25] introduced basic concepts in reliability theory in terms of quantile functions. Nair and Sankaran [25] defined the mean residual quantile function $M(u)$ as

$$M(u) = \frac{1}{1-u}\int_u^1 (Q(p) - Q(u))\,\mathrm{d}p. \tag{4}$$

$M(u)$ can also be written as

$$M(u) = \frac{1}{1-u}\int_u^1 (1-p)q(p)\,\mathrm{d}p, \tag{5}$$

where $q(u) = \mathrm{d}Q(u)/\mathrm{d}u$ is the quantile density function of X. We can interpret $M(u)$ as the mean remaining life of a unit beyond the $100(1-u)\%$ of the distribution. $Q(u)$ is uniquely determined from $M(u)$ by the relation

$$Q(u) = \int_0^u \frac{M(p)}{(1-p)}\,\mathrm{d}p - M(u) + M(0).$$

The function $M(u)$ is related to the hazard quantile function $H(u)$ [25] by the relation

$$\frac{1}{H(u)} = M(u) - (1-u)M'(u), \tag{6}$$

where $H(u)$ is defined as

$$H(u) = \frac{1}{(1-u)q(u)},$$

and $M'(u)$ is the derivative of $M(u)$ with respect to u. As pointed out earlier, certain families of distributions have no closed-form expressions for $m(x)$, but have simple expressions for $M(u)$. For example, the linear mean residual quantile distribution (LMRQD) discussed in Midhu *et al.* [20] has $M(u)$ given by

$$M(u) = cu + \mu, \quad \mu > 0,\ -\mu \le c < \mu. \tag{7}$$

For the LMRQD, the quantile function is of the form

$$Q(u) = -(c+\mu)\log(1-u) - 2cu, \quad \mu > 0,\ -\mu \le c < \mu.$$

From Equation (7), it can be observed that $M(u)$ is increasing (decreasing) in u when $0 < c < \mu$ ($-\mu \le c < 0$) and is constant when $c = 0$. The LMRQD was employed for modelling failure data on electric carts. Another example is Govindarajulu distribution [10] with the quantile function given by

$$Q(u) = \sigma((\beta + 1)u^{\beta} - \beta u^{\beta+1}) \tag{8}$$

and the mean residual quantile function as

$$M(u) = \frac{\sigma(u^{\beta}(\beta(u-1)(\beta(u-1) + u - 3) + 2) - 2)}{(\beta + 2)(u - 1)}. \tag{9}$$

This quantile function was first used in Govindarajulu [10] for modelling failure time data on a set of refrigerators. For both LMRQD and Govindarajulu distribution, there are no closed-form expressions for functions $F(x)$, $f(x)$ and $m(x)$. In such contexts, the quantile function and measures derived from it are used for modelling and analysis of data. For more properties and applications on $M(u)$, one could refer to Nair *et al.* [26].

Right-censored data commonly arise in industrial life-testing and medical follow-up studies. For example, in life testing experiments, some test subjects are still surviving even at the time of termination of the study. The experimenter is unable to obtain complete information on lifetime of such subjects. The estimation of the quantile function using the kernel density approach under right censoring was first suggested in [28]. The estimation of the quantile function for left-truncated and right-censored data was discussed in Zhou *et al.* [41]. Buhamra *et al.* [4] addressed the problem of estimating the quantile function in a multiple sample setup when the data are left-truncated and right-censored. Sankaran and Nair [33] discussed the non-parametric estimation of the hazard quantile function under right censoring. The non-parametric estimation of the quantile density function for censored and uncensored data was discussed in Soni *et al.* [36]. For various estimation procedures on quantile function and related concepts, one can refer to Yang [39], Jones [13], Rojo [30] and Cheng [6].

As mentioned earlier both $m(x)$ and $M(u)$ characterises, the distribution of the lifetime variable X. Thus non-parametric estimators $m(x)$ and $M(u)$ enable us to identify an appropriate model for a given lifetime data. The non-parametric estimation of $m(x)$ was first studied by Yang [38]. A smooth estimator of the mean residual life was developed by Chaubey and Sen [5]. Kochar *et al.* [16] studied the estimation of monotone mean residual life. The non-parametric estimation of the mean residual life under right censoring was addressed by Abdous and Berred [2]. Recently, Zhao *et al.* [40] proposed estimation of the mean residual life function with left-truncated and right-censored data. For further reference on non-parametric estimation of the mean residual life under different setup, one could refer to Ghoral and Rejtö [8], Ruiz and Guillamón [31], Kochar [15] and Na and Kim [23]. However, non-parametric estimation of $M(u)$ has not yet studied in literature. Motivated by this, in the present paper, we discuss non-parametric estimation of $M(u)$ under right censoring.

The rest of the article is organised as follows. Section 2 presents two non-parametric estimators of $M(u)$. The first one is based on the empirical quantile function and the second one is developed using the kernel density approach. Asymptotic properties of the estimators are discussed in Section 3. We carry out a simulation study, to assess the performance of the estimators in Section 4. The method is applied to two real data sets in Section 5. Finally, Section 6 provides a brief conclusion of the study.

2. Non-parametric estimation

Let X be a non-negative random variable representing the lifetime of a unit (individual) with distribution function $F(x)$ and density function $f(x)$. Suppose that X is right censored by a non-negative random variable Z. In practice, one can observe (T, δ), where $T = Min(X, Z)$ and $\delta = I(X \leq Z)$ with $I(\cdot)$ being the usual indicator function. If all the $Z_i's$ are fixed constants, the observations are time censored. On the other hand, when all the $Z_i's$ are equal to the same constant, then we have type-I censoring situation. If $Z_i's$ constitute a random sample from a distribution $G(x)$ (usually unknown) and are independent of $X_1, \ldots, X_n$ then (T, δ) is called a randomly right-censored sample. Denote $L(x)$ as the distribution function of T. Then under independence of X and Z, we obtain

$$1 - L(x) = (1 - F(x))(1 - G(x)).$$

The aim of the present study is to develop non-parametric estimation of $M(u)$ under right censoring.

Based on the right-censored sample (T_i, δ_i), $i = 1, 2, \ldots, n$, a non-parametric estimator of $Q(u)$ is given by $Q_n(u) = \inf\{x : F_n(x) \geq u\}$, where $F_n(x) = 1 - S_n(x)$ with $S_n(x)$ as the product limit estimator of the survivor function $S(x) = 1 - F(x)$ ([14,28]). From Equation (4), we suggest a simple non-parametric estimator of $M(u)$ as

$$\hat{M}(u) = \frac{1}{1-u} \int_u^1 (Q_n(p) - Q_n(u))\, \mathrm{d}p. \tag{10}$$

By changing $u = F_n(T_{(k)})$ and approximating integral to summation, we can write Equation (10) as

$$\begin{aligned} \hat{M}(u) &= \frac{1}{1-u} \sum_{i=k}^{n} [Q_n(F_n(T_{(i)})) - Q_n(F_n(T_{(k)}))] \int_{S_{(i-1)}}^{S_{(i)}} \mathrm{d}p \\ &= \frac{1}{1-u} \sum_{i=k}^{n} [Q_n(F_n(T_{(i)})) - Q_n(F_n(T_{(k)}))](S_{(i)} - S_{(i-1)}). \end{aligned} \tag{11}$$

In Equation (11), k is the greatest integer value less than nu and

$$S_{(i)} = \begin{cases} 0 & i = 0, \\ F_n(T_{(i)}) & i = 1, 2 \ldots n-1, \\ 1 & i = n, \end{cases}$$

$T_{(i)}$ is the ith ordered observation of T_i, $i = 1, 2, \ldots, n$ and $T_{(0)} = 0$. When no censoring is present $S_{(i)} - S_{(i-1)} = 1/n$ for all i. The estimator $Q_n(u)$ is a step function with jumps at the observed values of uncensored times.

Now we present a smooth non-parametric estimator of $M(u)$ using the kernel density method. Define $K(\cdot)$ to be a real-valued function such that

(i) $K(x) \geq 0$ for all x
(ii) $\int K(x)\, \mathrm{d}x = 1$
(iii) $K(\cdot)$ has finite support, that is $K(x) = 0$ for $|x| > c$ for some constant $c > 0$

(iv) $K(\cdot)$ is symmetric about zero and
(v) $K(\cdot)$ satisfies Lipschitz condition,

viz. there exists a constant D such that

$$|K(x) - K(y)| \leq D|x - y|, \quad \text{for all } x, y.$$

Let $\{h_n\}$ be a bandwidth sequence of positive numbers such that $h(n) \to 0$ as $n \to \infty$. Parzen [29] proposed a smooth kernel estimator for $Q(u)$ in the uncensored case as

$$Q^*(u) = \int_0^1 K\left(\frac{u-p}{h}\right) \hat{Q}(p)\,\mathrm{d}p, \tag{12}$$

where $\hat{Q}(u)$ is the empirical quantile estimator. Sheather and Marron [35] have modified the estimator (12) by approximating the integral to a sum which is given by

$$Q_n^*(u) = \frac{\sum_{i=1}^n K\left(\frac{i-1/2}{n} - u\right) T_{(i)}}{\sum_{i=1}^n K\left(\frac{i-1/2}{n} - u\right)}. \tag{13}$$

The estimator (13) in the presence of censoring is obtained as

$$Q_n^*(u) = \frac{\sum_{i=1}^n K(F_n(T_{(i)}) - u) T_{(i)}}{\sum_{i=1}^n K(F_n(T_{(i)}) - u)}. \tag{14}$$

Now we propose a smooth non-parametric estimator of $M(u)$ based on Equation (14) as

$$M_n(u) = \frac{1}{1-u} \int_u^1 (Q_n^*(p) - Q_n^*(u))\,\mathrm{d}p. \tag{15}$$

In the above estimation, one problem is the optimal choice of the bandwidth $h(n)$. A possible approach is to find the $h(n)$ that minimises the mean-squared error of the estimator. However, as seen in Section 3, the exact mean-squared error of the proposed estimator (15) is not available. One can employ the bootstrap method for randomly right-censored data given in Efron [7] to establish the optimal bandwidth, which is discussed in Section 5.

3. Asymptotic properties

In this section we establish asymptotic properties of $\hat{M}(u)$ and $M_n(u)$. We first discuss the asymptotic properties of $\hat{M}(u)$.

Theorem 3.1: *Let $F(\cdot)$ be continuous. Then the estimator $\hat{M}(u)$ is uniformly strong consistent.*

Proof: From the definition of $M(u)$ and $\hat{M}(u)$, we have

$$\hat{M}(u) - M(u) = \frac{1}{1-u} \int_u^1 (Q_n(p) - Q(p))\,\mathrm{d}p - (Q_n(u) - Q(u)). \tag{16}$$

Since $\sup_u |Q_n(u) - Q(u)| \to 0$ almost surely [3], Equation (16) gives $\sup_u |\hat{M}(u) - M(u)| \to 0$ as $n \to \infty$. This completes the proof. ■

Theorem 3.2: *Let $F(\cdot)$ be continuous. Then for fixed $u \in (0,1)$ $\sqrt{n}(\hat{M}(u) - M(u))$ is asymptotically normal with mean zero and variance $\sigma^2(u)$ given by*

$$\sigma^2(u) = \frac{1}{(1-u)^2} E\left[\int_u^1 Q_n(p)\,\mathrm{d}p - Q_n(u)\right]^2. \tag{17}$$

Proof:

$$\sqrt{n}(\hat{M}(u) - M(u)) = \frac{\sqrt{n}}{1-u}\int_u^1 (Q_n(p) - Q(p))\,\mathrm{d}p - \sqrt{n}(Q_n(u) - Q(u)). \tag{18}$$

From Parzen [29], for fixed $u (0 < u < 1)$, $\sqrt{n}(Q_n(u) - Q(u))$ is asymptotically normal with mean zero and variance

$$\sigma_1^2(u) = (1-u)^2 \int_0^u \frac{\mathrm{d}p}{(1-p)^2(1-G(Q(p)))},$$

where $G(\cdot)$ is the distribution function of censoring time.

Further by the functional delta method [3], it follows that $\sqrt{n}\int_u^1 (Q_n(p) - Q(p))\,\mathrm{d}p$ follows asymptotic normal with mean zero and variance given by

$$\sigma_2(u)^2 = \int_u^1 (\mu - \mu(v))^2\,\mathrm{d}\sigma_1^2(v)$$

where $\mu = \int_0^1 Q(p)\,\mathrm{d}p$ and $\mu(v) = \int_v^1 Q(p)\,\mathrm{d}p$.

Using the functional delta method and Slutsky's theorem ([34]), Equation (18) provides that $\sqrt{n}(\hat{M}(u) - M(u))$ is asymptotically normal with mean zero and variance $\sigma^2(u)$ given in Equation (17) which completes proof. ∎

Theorem 3.3: *Let $F(\cdot)$ be continuous. Assume that $K(\cdot)$ satisfies conditions* (i)–(v) *in Section 2. Then the estimator $M_n(u)$ defined by Equation* (15) *is uniformly strong consistent.*

Proof: We can write Equation (15) as

$$M_n(u) = \frac{1}{1-u}\int_u^1 (Q_n^*(p) - Q_n^*(u))\,\mathrm{d}p.$$

Thus

$$M_n(u) - M(u) = \frac{1}{1-u}\int_u^1 (Q_n^*(p) - Q(p))\,\mathrm{d}p - (Q_n^*(u) - Q(u)). \tag{19}$$

Since $\sup_u |Q_n^*(u) - Q(u)| \to 0$ almost surely ([35]), $\sup_u |M_n(u) - M(u)| \to 0$ as $n \to \infty$. This completes the proof. ∎

Theorem 3.4: *Let $F(\cdot)$ be continuous. Assume that $K(\cdot)$ satisfies conditions* (i)–(v) *in Section 2. Then for fixed $u \in (0,1)$ $\sqrt{n}(M_n(u) - M(u))$ is asymptotically normal with mean zero and variance $\sigma^2(u)$ given by*

$$\sigma_*^2(u) = \frac{1}{(1-u)^2} E\left[\int_u^1 Q_n^*(p)\,\mathrm{d}p - Q_n^*(u)\right]^2. \tag{20}$$

Proof:

$$\sqrt{n}(M_n(u) - M(u)) = \frac{\sqrt{n}}{1-u}\int_u^1 (Q_n^*(p) - Q(p))\,dp - \sqrt{n}(Q_n^*(u) - Q(u)). \tag{21}$$

From Sheather and Marron [35], it follows that for fixed $u(0 < u < 1)$, $\sqrt{n}(Q_n^*(u) - Q(u))$ is asymptotically normal with mean zero and variance

$$\sigma_3^2(u) = n^{-1}u(1-u)(q(u))^2 - n^{-1}h(n)q(u)^2\int_{-\infty}^{\infty} pK(p)K^{(-1)}(p)\,dp + o(n^{-1})h(n),$$

where $K^{(-1)}(\cdot)$ is the anti-derivative of $K(\cdot)$.

Using the functional delta method [3] and Slutsky's theorem [34], Equation (21) provides that $\sqrt{n}(M_n(u) - M(u))$ is asymptotically normal with mean zero and variance $\sigma_*^2(u)$ given in Equation (20) which completes proof. ■

Remark 3.1: As the estimation of $\sigma^2(u)$ and $\sigma_*^2(u)$ is difficult in practice, we use the naive bootstrap procedure of re sampling observations for finding there estimates.

4. Simulation study

We carried out simulation studies to assess the performance of the proposed estimators. We consider two quantile functions for the simulation. The first one is the linear hazard quantile distribution (LHQD) with

$$Q(u) = \frac{\log\frac{(a+bu)}{(a-au)}}{a+b}, \quad a > 0,\ a+b > 0,\ 0 \le u \le 1 \tag{22}$$

and

$$M(u) = \frac{\log\frac{(a+b)}{(a+bu)}}{b(1-u)}.$$

The quantile function (22) is extensively discussed in the context of reliability analysis by Midhu *et al.* [21]. Secondly, we consider Weibull distribution with quantile function given by

$$Q(u) = \beta(-\log(1-u))^{1/\alpha}, \quad \alpha, \beta > 0,\ 0 \le u \le 1. \tag{23}$$

The $M(u)$ for Equation (23) is obtained as

$$M(u) = \frac{\beta\Gamma\left(1+\frac{1}{\alpha}, -\log(1-u)\right)}{1-u} - \beta(-\log(1-u))^{1/\alpha}.$$

To check, how good the asymptotic normality, we generated the 1000 simulated samples of size 100 from from distributions (22) and (23). Histogram of resulting estimates $\hat{M}(u)$ and $M_n(u)$ are shown in Figures 1–4. We tested the normality in each case with the Kolmogorov–Smirnov test and passed in all cases. To asses the performance of the estimators, we generate random samples from Equations (22) and (23). We consider sample sizes

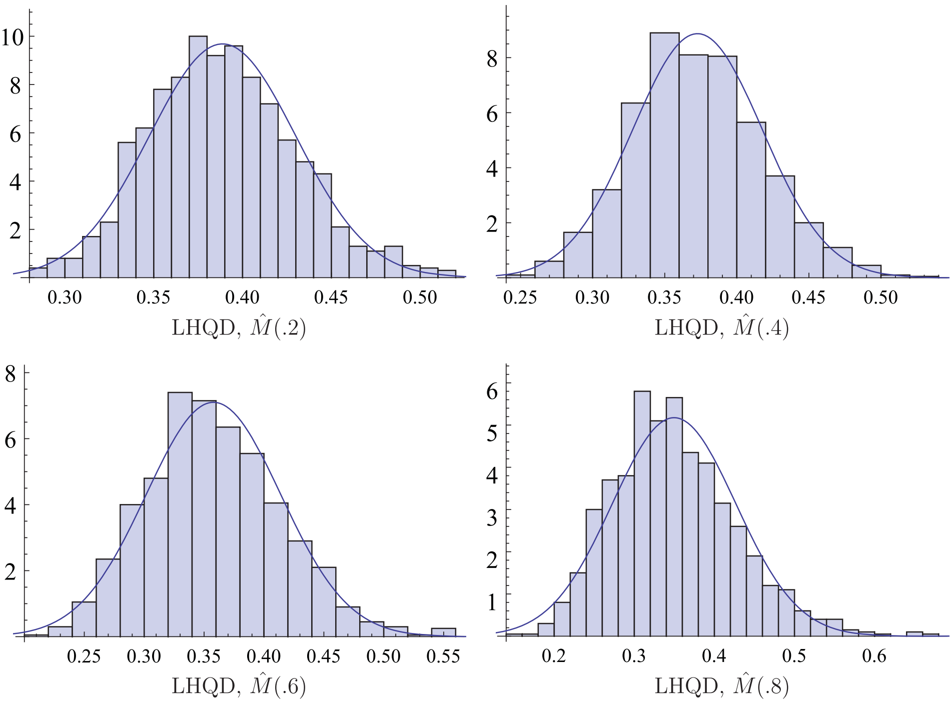

Figure 1. Histogram of $\hat{M}(u)$ with normal density curve for LHQD with $n = 100$.

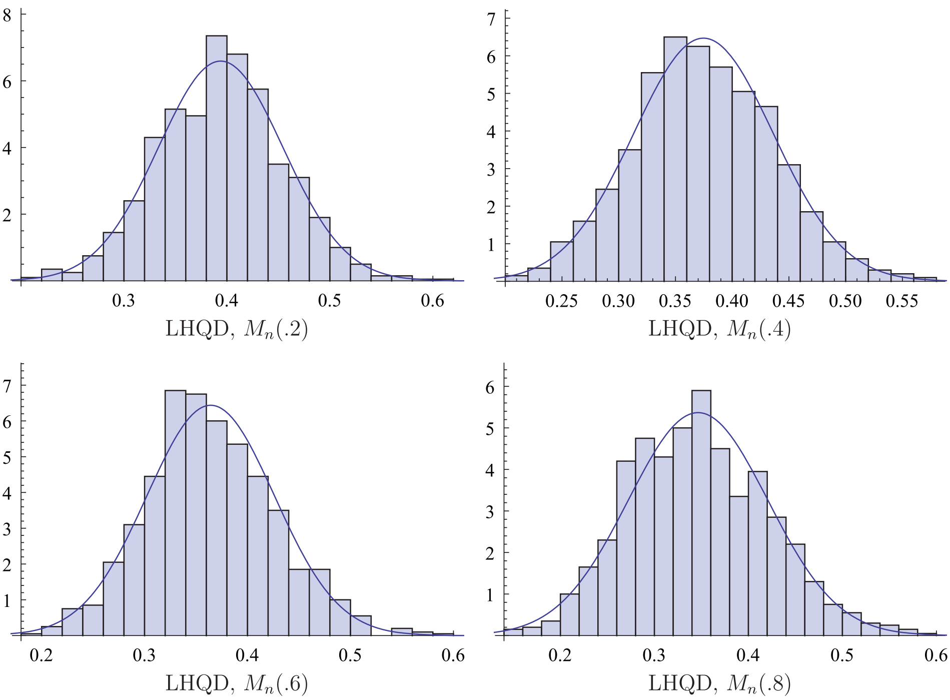

Figure 2. Histogram of $M_n(u)$ with normal density curve for LHQD with $n = 100$.

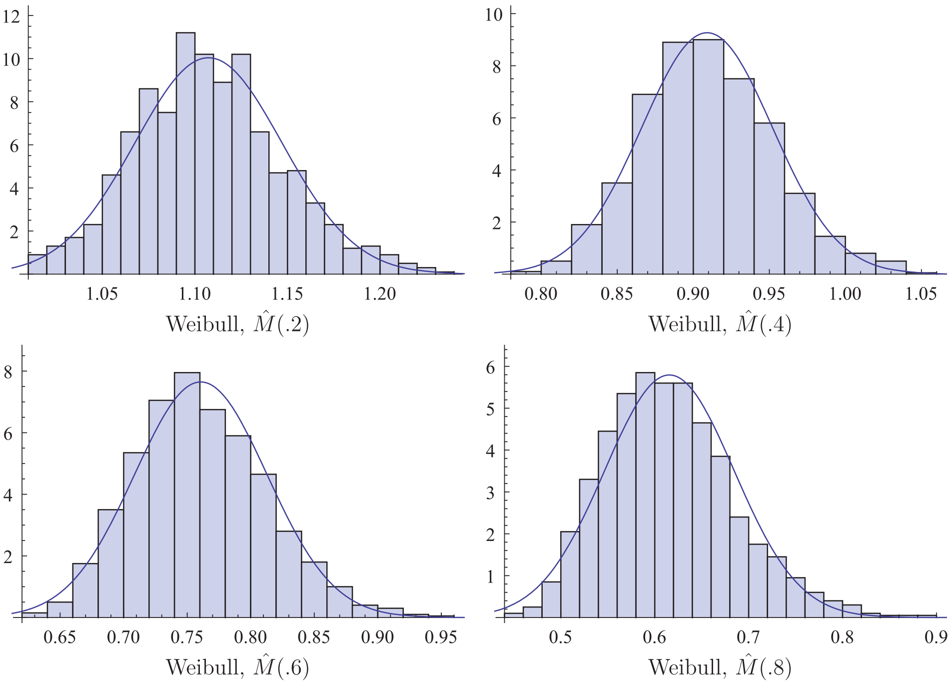

Figure 3. Histogram of $\hat{M}(u)$ with normal density curve for Weibull with $n = 100$.

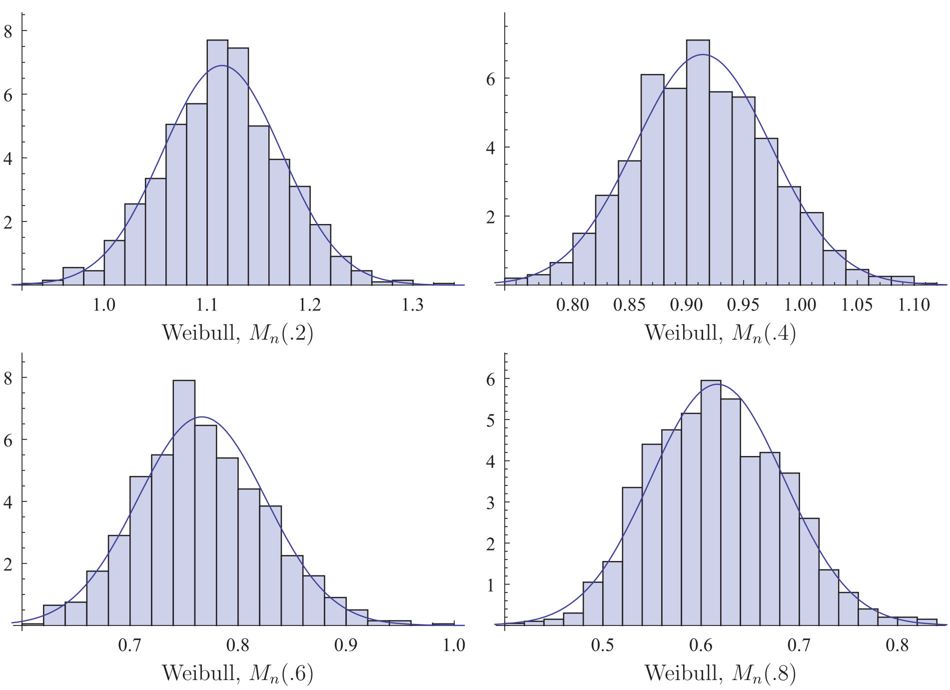

Figure 4. Histogram of $M_n(u)$ with normal density curve for Weibull with $n = 100$.

Table 1. Bias of the estimate: LHQD with $a = 2$ and $b = 1$.

		Uncensored			Censored		
u		$n = 100$	200	300	100	200	300
0.2	$\hat{M}(u)$	0.00612	0.00610	0.00311	0.00662	0.00649	0.00328
	$M_n(u)$	0.01812	0.01517	0.01447	0.01987	0.01847	0.01503
0.4	$\hat{M}(u)$	0.00664	0.00637	0.00354	0.00714	0.00702	0.00425
	$M_n(u)$	0.02241	0.0236	0.02002	0.02507	0.02432	0.02079
0.6	$\hat{M}(u)$	0.01131	0.00744	0.00595	0.0117	0.0082	0.00642
	$M_n(u)$	0.03691	0.03456	0.02887	0.04038	0.03737	0.03472
0.8	$\hat{M}(u)$	0.02594	0.01223	0.00945	0.02673	0.01371	0.01132
	$M_n(u)$	0.093	0.07934	0.0746	0.09969	0.09504	0.09319

$n = 100, 200$ and 300, under censored and uncensored cases. Observations are censored using uniform distribution $U(0, b)$ where b is chosen in such a way that 20% observations are censored. The triangular density function $K(x) = (1 - |x|)I(|x| \leq 1)$ is used as the kernel function for the estimation. The triangular kernel is one of the simple kernel with finite support, which is commonly employed in the estimation of functions of lifetime data. We consider various values for the bandwidth $h(n)$ and choose the value of $h(n)$ that minimises the bootstrap mean-squared error (BMSE). In both cases, it is observed that the value of $h(n) = 0.15$ that minimises BMSE. We then calculate the proposed estimators at different values of u for various parameter combinations. Tables 1 and 2 represent the empirical bias and MSE of the estimates based on 1000 simulations for the LHQD model with parameters a and b. In Table 2, BMSE is also computed, which is given in brackets. Tables 3 and 4 represent the empirical bias, MSE and BMSE (in brackets), of the estimates based on 1000 simulations for Weibull distribution. Both bias and MSE decrease as sample size increases. The bias and MSE of estimates for censored case is more than that of uncensored case. From Table 1, we see that the bias of $\hat{M}(u)$ is smaller than $M_n(u)$. When u is large, the bias of $M_n(u)$ is considerably large. For example, when $u = 0.8$, the bias of $M_n(u)$ is 0.093; but that of $\hat{M}(u)$ is 0.02594. The MSE of the $\hat{M}(u)$ is smaller than that of $M_n(u)$ in both censored and uncensored cases, when $n = 100$ and $n = 200$. When sample size is large, $M_n(u)$ has almost same MSE in both censored and uncensored setup. For Weibull distribution $\hat{M}(u)$ outperforms $M_n(u)$ in respect of bias (see Table 3). From Table 4, it follows that MSE values of $\hat{M}(u)$ is significantly lower than MSE of $M_n(u)$ in both censored and uncensored cases when $u = 0.2, 0.4, 0.8$. But when $u = 0.6$, $M_n(u)$ gives slightly better MSE compared to $\hat{M}(u)$. Figure 5 shows plot of $M(u)$, $\hat{M}(u)$ and $M_n(u)$ using a simulated sample from LHQD of size n = 100 for $h(n) = 0.18$, $a = 2$ and $b = 1$. Figure 6 shows plot of $M(u)$, $\hat{M}(u)$ and $M_n(u)$ using a simulated sample from Weibull distribution of size $n = 100$ for $h(n) = 0.18$, $\alpha = 2$ and $\beta = 2$. In Figures 5 and 6, thick line represents $M(u)$, thin line represents $\hat{M}(u)$ and dotted line represents $M_n(u)$. From Figures 5 and 6, we see that $M_n(u)$ is a smooth estimator compared to $\hat{M}(u)$ and that $\hat{M}(u)$ is close to $M(u)$ than $M_n(u)$.

Empirical coverage % with 95% asymptotic confidence interval for LHQD and Weibull are given in Tables 5 and 6. In Tables 7–10, we compared the asymptotic confidence interval and bootstrap confidence interval for LHQD and Weibull. Bootstrap confidence interval is wider than the other in most of the cases.

Table 2. MSE of the estimate : LHQD with $a = 2$ and $b = 1$.

		Censored			Uncensored		
u		100	200	300	100	200	300
0.2	$\hat{M}(u)$	0.00307 (0.00121)	0.00164 (0.00196)	0.00106 (0.00065)	0.00302 (0.00287)	0.00162 (0.00127)	0.00088 (0.00087)
	$M_n(u)$	0.00306 (0.00124)	0.00172 (0.00146)	0.00117 (0.00064)	0.0027 (0.0024)	0.00146 (0.00111)	0.00113 (0.00107)
0.4	$\hat{M}(u)$	0.00386 (0.00129)	0.00209 (0.00298)	0.00131 (0.00079)	0.00374 (0.00342)	0.00171 (0.00139)	0.00109 (0.00107)
	$M_n(u)$	0.00381 (0.00133)	0.0023 (0.00228)	0.00151 (0.00077)	0.00312 (0.00263)	0.00191 (0.0014)	0.00125 (0.00116)
0.6	$\hat{M}(u)$	0.00571 (0.00255)	0.00299 (0.00437)	0.00191 (0.00115)	0.00571 (0.00423)	0.00251 (0.00177)	0.0019 (0.00186)
	$M_n(u)$	0.00573 (0.00276)	0.00354 (0.00325)	0.00258 (0.00122)	0.00481 (0.00323)	0.00304 (0.00196)	0.00212 (0.00197)
0.8	$\hat{M}(u)$	0.01012 (0.00338)	0.00569 (0.01172)	0.00365 (0.0023)	0.00858 (0.00607)	0.00533 (0.00335)	0.00365 (0.00322)
	$M_n(u)$	0.01458 (0.00641)	0.01155 (0.01393)	0.01028 (0.00465)	0.01277 (0.00783)	0.01105 (0.00584)	0.00961 (0.00802)

Table 3. Bias of the estimate : Weibull with $\beta = 2$ and $\alpha = 2$.

		Uncensored			Censored		
u		$n = 100$	200	300	100	200	300
0.2	$\hat{M}(u)$	0.01299	0.01049	0.00561	0.01403	0.01078	0.00613
	$M_n(u)$	−0.07835	−0.07936	−0.05828	−0.08586	−0.0799	−0.07094
0.4	$\hat{M}(u)$	0.01639	0.01085	0.00786	0.01924	0.01276	0.00801
	$M_n(u)$	−0.02928	−0.02689	−0.01121	−0.0338	−0.02755	−0.01176
0.6	$\hat{M}(u)$	0.02523	0.01348	0.01114	0.0262	0.01626	0.01299
	$M_n(u)$	0.04623	0.03132	0.02546	0.05193	0.03485	0.02791
0.8	$\hat{M}(u)$	0.04628	0.02593	0.01557	0.05628	0.02882	0.01907
	$M_n(u)$	0.18905	0.16728	0.13609	0.19127	0.16751	0.16271

5. Data analysis

For the illustration of estimation procedure, we first consider a failure time data given in Aarset [1]. The data represent the failure times of 50 devices. We compute estimators (11) and (15) for the given data. To find the optimum value of $h(n)$, we use the bootstrap procedure given in Efron [7]. The value of $h(n)$ that minimises the bootstrap mean-squared error of the estimate can be chosen as optimal $h(n)$. We use 1000 bootstrap samples of size 50. The value of $h(n) = 0.15$ is chosen to calculate $M_n(u)$ for $0 < u \leq 0.25$, the value of $h(n) = 0.25$ is used for $0.25 < u < 0.50$ and the value of $h(n) = 0.35$ is used for $0.5 < u < 1$.

Nair *et al.* [27] have modelled the Aarset data using the Govindarajulu distribution (8). The estimates of the parameters using L-moments given in Nair *et al.* [27] are $\hat{\beta} = 2.0915$ and $\hat{\sigma} = 93.463$. The parametric estimate of $M(u)$, using $\hat{\beta}$ and $\hat{\sigma}$, is

$$M^*(u) = \frac{-147.701u^{4.0915} + 390.956u^{3.0915} - 288.941u^{2.0915} + 45.6864}{1-u}.$$

The plots of $\hat{M}(u)$, $M_n(u)$ and $M^*(u)$ are shown in Figure 7. In the figure thick line represents $M^*(u)$, thin line represents $\hat{M}(u)$ and dotted line represents $M_n(u)$. It can be seen from Figure 7 that $\hat{M}(u)$ is a step function and $M_n(u)$ is a smooth estimate close to parametric estimate $M^*(u)$ for most of the values of u.

The second dataset is taken from Wingo [37] which represent time to failure of various suspensions. The data consist of 33 failure times. We compute estimators (11) and (15) for these data. We employ bootstrap procedure mentioned earlier to find the optimum bandwidth $h(n)$ to calculate $M_n(u)$. But in this case, the optimum bandwidths are different from the earlier case. The optimum value of $h(n)$ is 0.35 for $0 < u \leq 0.25$ and 0.15 for $0.25 < u < 1$.

A three-parameter Weibull distribution is employed to fit the data by Wingo [37]. The quantile function of the three-parameter Weibull is

$$Q(u) = \mu + \beta(-\log(1-u))^{1/\alpha} \tag{24}$$

with $M(u)$ as

$$M(u) = \frac{\beta((u-1)(-\log(1-u))^{1/\alpha} + \Gamma(1+\frac{1}{\alpha}, -\log(1-u)))}{1-u}. \tag{25}$$

Table 4. MSE of the estimate: Weibull with $\beta = 2$ and $\alpha = 2$.

		Censored			Uncensored		
u		100	200	300	100	200	300
0.2	$\hat{M}(u)$	0.02166 (0.02736)	0.01094 (0.00962)	0.00695 (0.01132)	0.01801 (0.01951)	0.00987 (0.01279)	0.00619 (0.00766)
	$M_n(u)$	0.02751 (0.02791)	0.01815 (0.01615)	0.01444 (0.01774)	0.0241 (0.02654)	0.01644 (0.0181)	0.01218 (0.01334)
0.4	$\hat{M}(u)$	0.01977 (0.01879)	0.01022 (0.00767)	0.00623 (0.01077)	0.01634 (0.02219)	0.00921 (0.00795)	0.00564 (0.00397)
	$M_n(u)$	0.02144 (0.01913)	0.01173 (0.00922)	0.00766 (0.00989)	0.01995 (0.02515)	0.01089 (0.00947)	0.00709 (0.00457)
0.6	$\hat{M}(u)$	0.02296 (0.01691)	0.013 (0.01056)	0.0087 (0.01905)	0.01879 (0.01795)	0.01098 (0.00985)	0.00801 (0.0072)
	$M_n(u)$	0.02144 (0.01728)	0.01173 (0.00967)	0.00766 (0.01136)	0.01969 (0.01874)	0.00983 (0.00861)	0.00699 (0.00546)
0.8	$\hat{M}(u)$	0.02912 (0.01319)	0.01599 (0.00873)	0.01047 (0.03587)	0.02679 (0.02738)	0.01476 (0.01713)	0.01015 (0.00896)
	$M_n(u)$	0.05293 (0.02651)	0.03756 (0.02497)	0.03292 (0.06646)	0.05168 (0.05015)	0.03246 (0.03498)	0.02735 (0.02066)

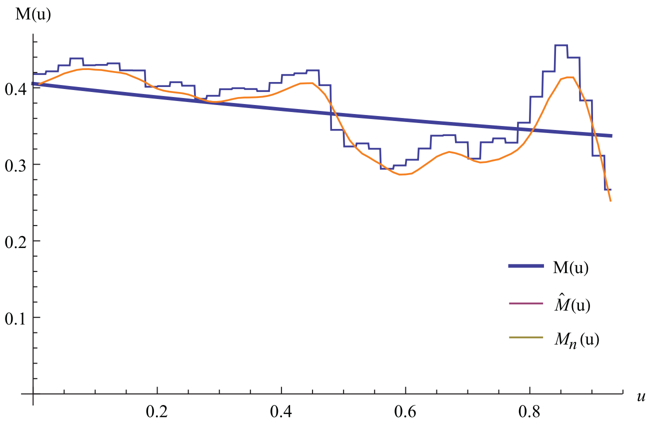

Figure 5. Plot of $M(u)$, $\hat{M}(u)$ and $M_n(u)$ using a simulated sample from LHQD of size $n = 100$ for $h(n) = 0.15$, $a = 2$ and $b = 1$ under censored setup.

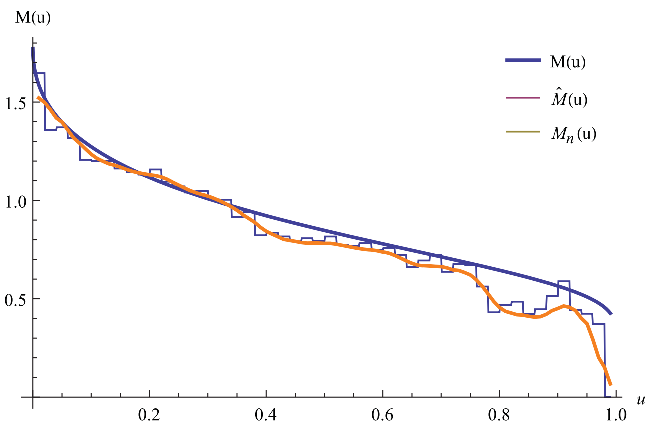

Figure 6. Plot of $M(u)$, $\hat{M}(u)$ and $M_n(u)$ using a simulated sample from Weibull of size $n = 100$ for $h(n) = 0.15$, $\alpha = 2$ and $\beta = 2$ under censored setup.

Table 5. Empirical coverage % with 95% asymptotic confidence interval for LHQD with $a = 2$ and $b = 1$.

		Uncensored			Censored		
u		$n = 100$	200	300	100	200	300
0.2	$\hat{M}(u)$	94.9	95.7	94.5	92.2	93.4	93.6
	$M_n(u)$	93.1	95.6	94.3	75.3	54.8	34.9
0.4	$\hat{M}(u)$	95.1	95.3	95.2	92.6	93.3	94.7
	$M_n(u)$	93.9	95.1	94.1	84.6	69.1	55.2
0.6	$\hat{M}(u)$	94.8	95.8	95.1	91.1	94.1	94.7
	$M_n(u)$	92.9	94.8	94.2	90.4	82.9	76.8
0.8	$\hat{M}(u)$	95.2	95.7	94.7	93.1	93.8	93.9
	$M_n(u)$	93.3	92.7	93.2	94.4	93.4	90.9

Table 6. Empirical coverage % with 95% asymptotic confidence interval for Weibull with $\beta = 2$ and $\alpha = 2$.

		Uncensored			Censored		
u		$n = 100$	200	300	100	200	300
0.2	$\hat{M}(u)$	94.7	95.0	94.7	93.7	93.3	94.2
	$M_n(u)$	95.2	94.4	94.7	85.2	76.2	60.8
0.4	$\hat{M}(u)$	94.7	95.7	95.1	94.6	95.1	94.6
	$M_n(u)$	95.0	95.2	94.5	90.7	83.1	75.6
0.6	$\hat{M}(u)$	94.3	94.8	94.4	89.7	91.8	92.4
	$M_n(u)$	94.1	94.5	94.5	94.6	90.4	87.7
0.8	$\hat{M}(u)$	95.1	95.4	95.2	90.8	93.3	93.5
	$M_n(u)$	93.4	93.9	94.5	92.2	94.5	94.3

Table 7. 95% Confidence intervals based on asymptotic normal and bootstrap for LHQD-uncensored sample.

		100		200		300	
u		Asym CI	Boot CI	Asym CI	Boot CI	Asym CI	Boot CI
0.2	$\hat{M}(u)$	(0.286, 0.437)	(0.112, 1.111)	(0.121, 0.59)	(0.145, 0.494)	(0.24, 0.503)	(0.147, 0.817)
	$M_n(u)$	(0.276, 0.418)	(0.112, 1.03)	(0.212, 0.549)	(0.181, 0.645)	(0.193, 0.487)	(0.105, 0.894)
0.4	$\hat{M}(u)$	(0.252, 0.412)	(0.084, 1.233)	(0.369, 0.59)	(0.414, 0.525)	(0.352, 0.476)	(0.213, 0.787)
	$M_n(u)$	(0.245, 0.395)	(0.086, 1.129)	(0.362, 0.531)	(0.358, 0.537)	(0.344, 0.45)	(0.175, 0.884)
0.6	$\hat{M}(u)$	(0.162, 0.392)	(0.073, 0.877)	(0.335, 0.618)	(0.422, 0.489)	(0.299, 0.453)	(0.179, 0.755)
	$M_n(u)$	(0.158, 0.369)	(0.076, 0.765)	(0.323, 0.529)	(0.296, 0.577)	(0.288, 0.413)	(0.136, 0.875)
0.8	$\hat{M}(u)$	(0.156, 0.452)	(0.052, 1.353)	(0.183, 0.706)	(0.343, 0.378)	(0.242, 0.491)	(0.153, 0.781)
	$M_n(u)$	(0.129, 0.401)	(0.057, 0.913)	(0.196, 0.539)	(0.237, 0.447)	(0.234, 0.41)	(0.106, 0.907)

Table 8. 95% Confidence intervals based on asymptotic normal and bootstrap for LHQD-censored sample.

		100		200		300	
u		Asym CI	Boot CI	Asym CI	Boot CI	Asym CI	Boot CI
0.2	$\hat{M}(u)$	(0.269, 0.422)	(0.256, 0.444)	(0.35, 0.449)	(0.275, 0.572)	(0.343, 0.433)	(0.34, 0.437)
	$M_n(u)$	(0.263, 0.401)	(0.233, 0.452)	(0.347, 0.441)	(0.263, 0.582)	(0.34, 0.427)	(0.32, 0.452)
0.4	$\hat{M}(u)$	(0.267, 0.43)	(0.244, 0.47)	(0.334, 0.451)	(0.272, 0.554)	(0.323, 0.431)	(0.316, 0.441)
	$M_n(u)$	(0.253, 0.406)	(0.213, 0.481)	(0.325, 0.435)	(0.239, 0.593)	(0.318, 0.417)	(0.296, 0.449)
0.6	$\hat{M}(u)$	(0.234, 0.424)	(0.174, 0.572)	(0.269, 0.399)	(0.19, 0.565)	(0.314, 0.441)	(0.308, 0.45)
	$M_n(u)$	(0.219, 0.392)	(0.147, 0.584)	(0.262, 0.38)	(0.169, 0.59)	(0.307, 0.423)	(0.285, 0.455)
0.8	$\hat{M}(u)$	(0.174, 0.415)	(0.123, 0.587)	(0.192, 0.356)	(0.121, 0.567)	(0.267, 0.432)	(0.235, 0.49)
	$M_n(u)$	(0.152, 0.362)	(0.093, 0.59)	(0.184, 0.432)	(0.097, 0.818)	(0.252, 0.403)	(0.211, 0.483)

For the dataset, the maximum likelihood estimates of the parameters, proposed by Wingo [37], are

$$\hat{\mu} = 14.5, \quad \hat{\beta} = 106.5 \quad \text{and} \quad \hat{\alpha} = 3.8.$$

The parametric estimate of $M(u)$ is given by

$$M^*(u) = \frac{\Gamma(1.26316, -(\log(1-u)))}{1-u} - 106.5(-((1-u)\log))^{0.263158}.$$

Table 9. 95% Confidence intervals based on asymptotic normal and bootstrap for Weibull-uncensored sample.

		100		200		300	
u		Asym CI	Boot CI	Asym CI	Boot CI	Asym CI	Boot CI
0.2	$\hat{M}(u)$	(0.964, 1.746)	(1.218, 1.382)	(1.035, 1.375)	(0.91, 1.564)	(1.039, 1.387)	(0.851, 1.693)
	$M_n(u)$	(0.923, 1.69)	(0.937, 1.665)	(1.026, 1.359)	(0.913, 1.527)	(1.025, 1.321)	(1.075, 1.26)
0.4	$\hat{M}(u)$	(0.817, 1.335)	(0.776, 1.405)	(0.856, 1.172)	(0.642, 1.563)	(0.812, 1.185)	(0.686, 1.404)
	$M_n(u)$	(0.814, 1.274)	(0.726, 1.427)	(0.844, 1.157)	(0.663, 1.472)	(0.79, 1.102)	(0.854, 1.02)
0.6	$\hat{M}(u)$	(0.486, 0.969)	(0.358, 1.316)	(0.677, 1.035)	(0.55, 1.274)	(0.643, 1.125)	(0.514, 1.408)
	$M_n(u)$	(0.456, 0.939)	(0.367, 1.165)	(0.673, 0.998)	(0.555, 1.211)	(0.614, 0.99)	(0.668, 0.911)
0.8	$\hat{M}(u)$	(0.303, 0.823)	(0.137, 1.817)	(0.484, 0.869)	(0.264, 1.592)	(0.561, 1.349)	(0.394, 1.923)
	$M_n(u)$	(0.287, 0.73)	(0.144, 1.458)	(0.452, 0.813)	(0.3, 1.223)	(0.519, 1.072)	(0.531, 1.048)

Table 10. 95% Confidence intervals based on asymptotic normal and bootstrap for Weibull-censored sample.

		100		200		300	
u		Asym CI	Boot CI	Asym CI	Boot CI	Asym CI	Boot CI
0.2	$\hat{M}(u)$	(0.909, 1.319)	(0.984, 1.217)	(0.94, 1.266)	(0.977, 1.218)	(0.897, 1.17)	(0.946, 1.11)
	$M_n(u)$	(0.893, 1.296)	(0.983, 1.177)	(0.946, 1.237)	(1.041, 1.124)	(0.905, 1.148)	(0.991, 1.049)
0.4	$\hat{M}(u)$	(0.677, 1.117)	(0.823, 0.92)	(0.754, 1.002)	(0.651, 1.16)	(0.719, 1.115)	(0.506, 1.584)
	$M_n(u)$	(0.66, 1.073)	(0.832, 0.851)	(0.733, 0.97)	(0.638, 1.116)	(0.71, 1.189)	(0.458, 1.843)
0.6	$\hat{M}(u)$	(0.531, 0.941)	(0.507, 0.985)	(0.603, 0.883)	(0.541, 0.984)	(0.658, 0.885)	(0.592, 0.984)
	$M_n(u)$	(0.497, 0.885)	(0.473, 0.93)	(0.585, 0.843)	(0.512, 0.962)	(0.642, 0.845)	(0.502, 1.081)
0.8	$\hat{M}(u)$	(0.404, 0.912)	(0.413, 0.893)	(0.426, 0.808)	(0.495, 0.696)	(0.431, 0.705)	(0.381, 0.798)
	$M_n(u)$	(0.375, 0.82)	(0.364, 0.845)	(0.398, 0.738)	(0.429, 0.685)	(0.413, 0.645)	(0.312, 0.854)

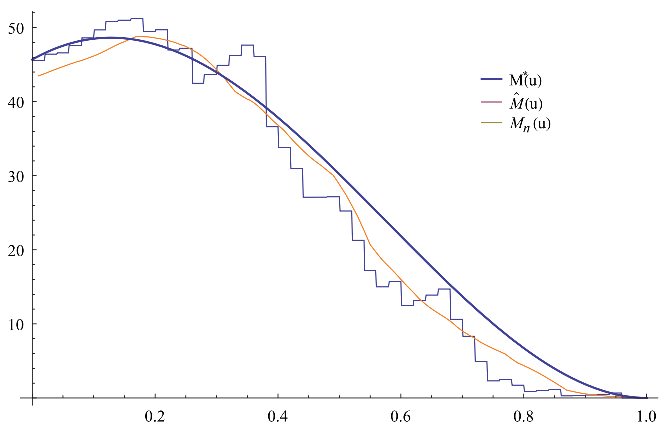

Figure 7. Non-parametric estimates of mean residual quantile function and its parametric estimate for Aarset data.

Figure 8 shows $\hat{M}(u)$, $M_n(u)$ and $M^*(u)$. In the figure thick line represents $M^*(u)$, thin line represents $\hat{M}(u)$ and dotted line represents $M_n(u)$. It follows from Figure 8 that $M_n(u)$ is a smooth curve close to the parametric estimate of the $M(u)$ for most of the values of u.

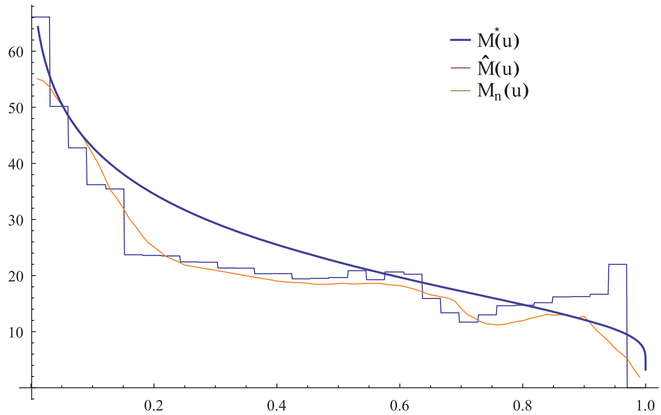

Figure 8. Non-parametric estimates of mean residual quantile function and its parametric estimate for suspension data.

6. Conclusion

The present study provided two non-parametric estimators, one based on the empirical quantile function and another using the kernel smoothing method, for the mean residual quantile function under right censoring which will be useful for modelling and analysis of lifetime data. Simulation studies proved that the estimators have small bias and less MSE. The non-parametric estimator $\hat{M}(u)$ is better than the kernel based estimator $M_n(u)$, in particular for large values of u. The method has been illustrated using two real data sets. The optimal choice of the bandwidth $h(n)$ depends on the distribution of the lifetime random variable X. More simulation studies are required to investigate this aspect, which is a topic of future research.

Acknowledgements

We thank the editor and the anonymous reviewer for their constructive comments, which helped us to improve the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCiD

N.N. Midhu http://orcid.org/0000-0001-7570-3446

References

[1] M.V. Aarset, *How to identify a bathtub hazard rate*, IEEE Trans. Reliab. 36 (1987), pp. 106–108.

[2] B. Abdous and A. Berred, *Mean residual life estimation*, J. Statist. Plan. Inference 132 (2005), pp. 3–19.

[3] P.K. Andersen, R.D. Gill, and N. Keiding, *Statistical Models Based on Counting Processes*, Springer, New York, 1993.

[4] S.S. Buhamra, N.M. Al-Kandari, and S.E. Ahmed, *Nonparametric inference strategies for the quantile functions under left truncation and right censoring*, J. Nonparametr. Stat. 19 (2007), pp. 189–198.
[5] Y.P. Chaubey and P.K. Sen, *On smooth estimation of mean residual life*, J. Statist. Plan. Inference 75 (1999), pp. 223–236.
[6] C. Cheng, *Almost-sure uniform error bounds of general smooth estimators of quantile density functions*, Stat. Probab. Lett. 59 (2002), pp. 183–194.
[7] B. Efron, *Censored data and the bootstrap*, J. Amer. Statist. Assoc. 76 (1981), pp. 312–319.
[8] J.K. Ghoral and L. Rejtö, *Estimation of mean residual life with censorded data under the proportional hazard model*, Commun. Stat – Theory Methods 16 (1987), pp. 2097–2114.
[9] W. Gilchrist, *Statistical Modelling with Quantile Functions*, CRC Press, Boca Raton, 2000.
[10] Z. Govindarajulu, *A class of distributions useful in life testing and reliability with applications to nonparametric testing*, Theory Appl. Reliab, 1 (1977), pp. 109–130.
[11] F. Guess and F. Proschan, *Mean residual life: Theory and applications*, Handbook of Statistics 7 (1988), pp. 215–224.
[12] W.J. Hall and J.A. Wellner, *Mean residual life*, Stat. Relat. Topics 169 (1981), pp. 184.
[13] M.C. Jones, *Estimating densities, quantiles, quantile densities and density quantiles*, Ann. Instit. Statist. Math. 44 (1992), pp. 721–727.
[14] E.L. Kaplan and P. Meier, *Nonparametric estimation from incomplete observations*, J. Amer. Statist. Assoc. 53 (1958), pp. 457–481.
[15] S.C. Kochar, *Stochastic comparisons of spacings and order*, Front. Reliab., Qual. Reliab. 4 (1998), pp. 201.
[16] S.C. Kochar, H. Mukerjee, and F.J. Samaniego, *Estimation of a monotone mean residual life*, Ann. Stat. 28 (2000), pp. 905–921.
[17] S. Kotz and D.N. Shanbhag, *Some new approaches to probability distributions*, Adv. Appl. Probab. 12 (1980), pp. 903–921.
[18] C.D. Lai and M. Xie, *Stochastic Ageing and Dependence for Reliability*, Springer Science Business Media, Inc., New York, 2006.
[19] A.W. Marshall and I. Olkin, *Life Distributions: Structure on Nonparametric, Semiparametric and Parametric Familes*, Springer, New York, 2007.
[20] N.N. Midhu, P.G. Sankaran, and N.U. Nair, *A class of distributions with the linear mean residual quantile function and it's generalizations*, Statist. Methodol. 15 (2013), pp. 1–24.
[21] N.N. Midhu, P.G. Sankaran, and N.U. Nair, *A class of distributions with linear hazard quantile function*, Commun. Stat – Theory Methods 43 (2014), pp. 3674–3689.
[22] E.J. Muth, *Reliability models with positive memory derived from the mean residual life function*, Theory Appl. Reliab. 2 (1977), pp. 401–434.
[23] M.H. Na and J.J. Kim, *On inference for mean residual life*, Commun. Stat – Theory Methods 28 (1999), pp. 2917–2933.
[24] N.U. Nair and P.G. Sankaran, *Characterizations of multivariate life distributions*, J. Multivariate Anal. 99 (2008), pp. 2096–2107.
[25] N.U. Nair and P.G. Sankaran, *Quantile-based reliability analysis*, Commun. Stat – Theory Methods 38 (2009), pp. 222–232.
[26] N.U. Nair, P.G. Sankaran, and N. Balakrishnan, *Quantile-Based Reliability Analysis*, Springer, New York, 2013.
[27] N.U. Nair, P.G. Sankaran, and B. Vineshkumar, *The Govindarajulu distribution: Some properties and applications*, Commun. Stat – Theory Methods 41 (2012), pp. 4391–4406.
[28] W.J. Padgett, *A kernel-type estimator of a quantile function from right-censored data*, J. Amer. Statist. Assoc. 81 (1986), pp. 215–222.
[29] E. Parzen, *Nonparametric statistical data modeling*, J. Amer. Statist. Assoc. 74 (1979), pp. 105–121.
[30] J. Rojo, *Estimation of the quantile function of an ifra distribution*, Scand. J. Stat. 25 (1998), pp. 293–310.

[31] J.M. Ruiz and A. Guillamón, *Nonparametrie recursive estimator for mean residual life and vitality function under dependence conditions*, Commun. Stat. – Theory Methods 25 (1996), pp. 1997–2011.
[32] J.M. Ruiz and J. Navarro, *Characterizations based on conditional expectations of the doubled truncated distribution*, Ann. Instit. Statist. Math. 48 (1996), pp. 563–572.
[33] P.G. Sankaran and N.U. Nair, *Nonparametric estimation of hazard quantile function*, J. Nonparametr. Stat. 21 (2009), pp. 757–767.
[34] R.J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.
[35] S.J. Sheather and J.S. Marron, *Kernel quantile estimators*, J. Amer. Statist. Assoc. 85 (1990), pp. 410–416.
[36] P. Soni, I. Dewan, and K. Jain, *Nonparametric estimation of quantile density function*, Comput. Stat. Data Anal. 56 (2012), pp. 3876–3886.
[37] D.R. Wingo, *Solution of the three-parameter weibull equations by constrained modified quasilinearization (progressively censored samples)*, IEEE Trans. Reliab. 22 (1973), pp. 96–102.
[38] G.L. Yang, *Estimation of a biometric function*, Ann. Stat. 6 (1978), pp. 112–116.
[39] S.S. Yang, *A smooth nonparametric estimator of a quantile function*, J. Amer. Statist. Assoc. 80 (1985), pp. 1004–1011.
[40] M. Zhao, H. Jiang, and X. Liu, *A note on estimation of the mean residual life function with left-truncated and right-censored data*, Stat. Probab. Lett. 83 (2013), pp. 2332–2336.
[41] X. Zhou, L. Sun, and H. Ren, *Quantile estimation for left truncated and right censored data*, Statist. Sin. 10 (2000), pp. 1217–1230.