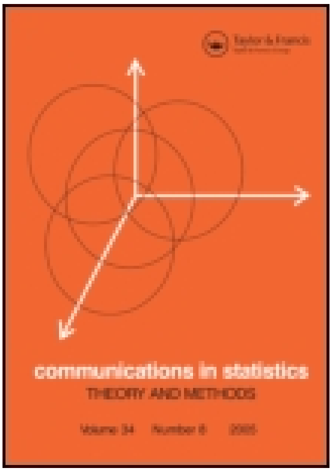




Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

# On the Nonparametric Mean Residual Life Estimator in Length-biased Sampling

V. Fakoor[a]

[a] Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

Accepted author version posted online: 14 Aug 2014.Published online: 21 Dec 2014.

To cite this article: V. Fakoor (2015) On the Nonparametric Mean Residual Life Estimator in Length-biased Sampling, Communications in Statistics - Theory and Methods, 44:3, 512-519, DOI: 10.1080/03610926.2012.748328

To link to this article: http://dx.doi.org/10.1080/03610926.2012.748328

*Communications in Statistics—Theory and Methods*, 44: 512–519, 2015

ISSN: 0361-0926 print / 1532-415X online
DOI: 10.1080/03610926.2012.748328



# On the Nonparametric Mean Residual Life Estimator in Length-biased Sampling

V. FAKOOR

Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

*In this article, we discuss nonparametric estimation of a mean residual life function from length-biased data. Precisely, we prove strong uniform consistency and weak converge of the nonparametric mean residual life estimator in length-biased setting.*

**Keywords** Length-biased; Mean residual life; Strong uniform consistency; Weak convergence.

**Mathematics Subject Classification** 62G05; 62G20.

## 1. Introduction

Let $X$ be a random variable (r.v.) with density function $f$ and distribution function (df) $F$, we say that random variable $Y$ has the length-biased distribution of $F$ if the df of $Y$ is given by

$$G(t) = \frac{1}{\mu} \int_0^t x f(x) dx, \quad t \geq 0, \tag{1.1}$$

where $\mu = \int_0^\infty x dF(x)$, and $\mu$ is assumed finite. Hence, the density of $Y$ is

$$g(t) = \frac{t f(t)}{\mu}, \quad t \geq 0. \tag{1.2}$$

The phenomenon of length-biased was first tackled in the context of anatomy by Wicksell (1925) as what he called the corpuscle problem. Length-biased was later systematically studied by McFadden (1962), Blumenthal (1967), then by Cox (1969), in the context of estimation of the distribution of fiber lengths in a fabric.

Length-biased data arise in many practical situations, including econometrics, survival analysis, renewal processes, biomedicine and physics. For instance, if $X$ represents the length of an item and the probability of this item being selected in the sample is proportional

Received July 10, 2012; Accepted October 29, 2012.
Address correspondence to V. Fakoor, Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran; E-mail: fakoor@math.um.ac.ir

to its length, then the distribution of the observed length is length-biased. In cross-sectional studies in survival analysis, for example, often the probability of being selected, for a particular subject, is proportional to his/her survival time. Interesting applications of length-biased data can be found in Cox (1969), Patil and Rao (1977, 1978), Colman (1979) and Vardi (1982b). The distribution function, $G$, is, from a slightly different perspective, the distribution of the randomly left-truncated r.v.'s $Y$, in the stationary assumption. If the incidence rate of the event has not changed over time, a stationary might reasonably describe the incidence of the event, this is equivalent to assuming that the randomly left truncation induced by the sampling is uniform (Wang, 1991). Throughout this article we assume that $G$ is continuous on $\mathcal{R}^+ = [0, \infty)$. An elementary calculation shows that $F$ is determined uniquely by $G$, namely

$$F(t) = \mu \int_0^t y^{-1} dG(y), \quad t \geq 0.$$

Cox (1969) and Vardi (1982a) considered the problem of finding a nonparametric maximum likelihood estimate (NPMLE) of $F$ on the basis of a sample $Y_1, Y_2, \ldots, Y_n$ from $G$. Let $G_n$ be empirical estimator of G is given by

$$G_n(t) = \frac{1}{n} \sum_{i=1}^{n} I(Y_i \leq t),$$

where $I(A)$ denote the indicator of the event $A$. Empirical estimator of $F$ can be written in the form

$$F_n(t) = \mu_n \int_0^t y^{-1} dG_n(y),$$

where

$$\mu_n^{-1} = \int_0^\infty y^{-1} dG_n(y).$$

The mean residual life (MRL) function at age $x$ is defined to be the expected remaining life given survival to age $x$. It is a concept of obvious interest and, indeed, one of the most important notations in actuarial, reliability and survivorship studies. Yang (1978) and Hall and Wellner (1979) initiated investigations of the asymptotic behaviors of the empirical mean residual life process. They obtained results on the basis of a uniform on compact topology. Csörgő and Zitikis (1996) exposed the study of the mean residual life process over the whole positive half line. They establish the the strong uniform-over-$[0, \infty)$ consistency, and weak uniform-over-$[0, \infty)$ approximation of the empirical mean residual life process by employing weight functions. By representing the empirical mean residual life process as an integral form, Bae and Kim (2006) proved uniform asymptotic behaviours of the process over the whole positive half line. Under length-biased sampling and Type I censoring, pointwise consistency of MRL established by de Uña-Álvarez (2004), which un-censoring is its special case.

The main aim of this article is to derive asymptotic behaviors of the nonparametric estimator of a MRL function for a sample from the corresponding length-biased distribution. We prove uniform consistency and weak convergenc of the MLR estimator.

For any d.f. $L$ denotes the right endpoint of its support by a $\tau_L = \inf\{x : L(x) = 1\}$. Assuming that $\tau_F = \tau_G = \tau < \infty$. As mentioned above, the MRL function $M_F$ at $x \geq 0$

is defined by

$$M_F(x) := E(X - x|X > x) = \frac{I_{[0,\tau)}(x)}{1 - F(x)} \int_x^\infty (1 - F(t))dt. \tag{1.3}$$

$M_n$, is the empirical counterpart of $M_F$, defined by

$$M_n(x) := \frac{I_{[0,Y_{(n)})}(x)}{1 - F_n(x)} \int_x^\infty (1 - F_n(t))dt, \tag{1.4}$$

where $Y_{(n)} = \max_{1 \le i \le n} Y_i$. The layout of this article is as follows. In Sec. 2, we give some asymptotic results of MRL function. Proofs of the main results deffered to Sec. 3.

## 2. Asymptotic Study

### *2.1. Strong Uniform Consistency*

We have the following strong consistency result for $M_n - M_F$.

**Theorem 2.1.** *If* $0 < \mu < \infty$, *we have as* $n \to \infty$, *for any* $\epsilon > 0$

$$\sup_{0 \le x \le \tau - \epsilon} |M_n(x) - M_F(x)| \to 0 \quad a.s. \tag{2.1}$$

*Proof.* See Sec. 3. □

### *2.2. Weak Convergence*

It is the purpose of this subsection to study the weak convergence results for the normalized mean residual life process $U_n(x)$, which is defined by

$$U_n(x) = \sqrt{n}[M_n(x) - M_F(x)].$$

First, we introduce a class of functions $\{\varphi_x; \varphi_x(u) = (1 - \frac{x}{u})I_{(x,\infty)}(u), x \ge 0\}$. We notice that for each $x \ge 0$ the function $\varphi_x(.)$ is bounded by envelope $\varphi_0(u) = I_{(0,\infty)}(u)$ and

$$\int_x^\infty (1 - F(u))du = \mu \int \varphi_x(u)dG(u).$$

Then, for $x \ge 0$, we get the integral representations of $M$ and $M_n$ as

$$M_F(x) = \mu \frac{I_{[0,\tau)}(x)}{1 - F(x)} H(x) \tag{2.2}$$

and

$$M_n(x) = \mu_n \frac{I_{[0,Y_{(n)})}(x)}{1 - F_n(x)} H_n(x). \tag{2.3}$$

where

$$H(x) = \int \varphi_x(u)dG(u)$$

and

$$H_n(x) = \int \varphi_x(u)dG_n(u)$$

Notice further that, for each $x \geq 0$

$$M_n(x) - M_F(x) = \mu_n \frac{I_{[0,Y_{(n)})}(x)}{1 - F_n(x)}[H_n(x) - H(x)] + R_n(x)H(x), \tag{2.4}$$

where

$$R_n(x) = \mu_n \frac{I_{[0,Y_{(n)})}(x)}{1 - F_n(x)} - \mu \frac{I_{[0,\tau)}(x)}{1 - F(x)}. \tag{2.5}$$

We will use the representation (2.4) in studing the weak convergence of $U_n(x)$. In the following we assume that $Q_G(t) = Q(t)$ is the quntile function of $G$ for $0 < t < 1$. Let $W = \{W(t); 0 < t < 1\}$ be a mean zero Gaussian process with covariance function

$$Cov(W(x), W(y)) = \sigma_{x\wedge y} - A(y)\sigma_x - A(x)\sigma_y + A(x)A(y)\sigma, \tag{2.6}$$

for each $x, y \in (0, 1)$, where $A(t) = \mu^2(1 - 2F(Q(t)))$,

$$\sigma_t = \int_0^{Q(t)} \frac{1}{y^2} dG(y), 0 < t < 1,$$

and

$$\sigma = \lim_{t\to\infty} \sigma_t = \int_0^{\infty} \frac{1}{y^2} dG(y).$$

Also, let $Z = \{Z(x); x \geq 0\}$ be a mean zero Gaussian process with covariance function

$$Cov(Z(x), Z(y)) = \int_{x\vee y}^{\infty} \left(1 - \frac{x}{u}\right)\left(1 - \frac{y}{u}\right) dF(u).$$

**Theorem 2.2.** *Suppose that $E(X^{-r}) < \infty$ for some $r > 1$. Then we have*

$$U_n(\cdot) \xrightarrow{\mathcal{D}} U(\cdot)$$

*over $D[0, \infty)$, where*

$$U(x) = \mu \frac{I_{[0,\tau)}(x)}{1 - F(x)} Z(x) + \frac{I_{[0,\tau)}(x) W(G(x))}{(1 - F(x))^2} \left( \int_x^{\infty} \left(1 - \frac{x}{u}\right) dG(u) \right) \tag{2.7}$$

*and $D[0, \infty)$ is the space of cadlag functions on $[0, \infty)$ endowed with Skorokhod metric* d *on $[0, \infty)$ [cf. Billingsley (1968)].*

*Proof.* See Sec. 3. □

## 3. Proofs

In order to prove Theorem 2.1 we need the following lemma.

**Lemma 3.1.** *(Horváth, 1985) Let* $0 < \mu < \infty$, *then we have*

$$\sup_{0 \le t < \infty} |F_n(t) - F(t)| \to 0 \quad a.s.$$

*Proof of Theorem 2.1.* First we have

$$M_n(x) - M_F(x) = (1 - F_n(x))^{-1}\left( -\int_x^{\infty} (F_n(t) - F(t))dt + M_F(x)(F_n(x) - F(x))\right). \tag{3.1}$$

Hence, (2.1) follows by Lemma 3.1, if we can show that

$$I_n = \int_0^{\infty} |F_n(t) - F(t)|dt \to 0 \quad a.s.$$

Since $\mu = \int_0^{\infty}(1-F(t))dt < \infty$, therefore, for $\epsilon > 0$ arbitrary small, we can choose $\alpha > 0$ so large that

$$I(\alpha) = \int_{\alpha}^{\infty} (1 - F(t))dt < \epsilon/2.$$

Thus,

$$I_n \le I(\alpha) + I_{n1} + I_{n2},$$

where

$$I_{n1} = \mu_n^{-1} n^{-1} \sum_{i=1}^{n} \int_{\alpha}^{\infty} \frac{1}{Y_i} I(Y_i > t)dt$$

and

$$I_{n2} = \int_0^{\alpha} |F_n(t) - F(t)|dt \le \alpha \sup_{0 \le x < \infty} |F_n(t) - F(t)|.$$

Note that

$$E\int_{\alpha}^{\infty} \frac{1}{Y} I(Y > t)dt = \int_{\alpha}^{\infty} (1 - F(t))dt = \mu^{-1} I(\alpha).$$

Strong law of large numbers implies that

$$I_{n1} \to I(\alpha) \quad a.s. \tag{3.2}$$

Also, Lemma 3.1, implies that

$$I_{n2} \to 0 \quad a.s. \tag{3.3}$$

Therefore,

$$\limsup_{n\to\infty} I_n \le \epsilon \quad a.s.$$

for all small $\epsilon$ and this completes the proof of (2.1).

**Lemma 3.2.** *(Bae and Kim, 2006) As $n \to \infty$,*

$$\sqrt{n} I_{[Y_{(n)},\tau)}(\cdot) \xrightarrow{\mathcal{D}} 0$$

*as random elements of $D[0, \infty)$.*

Let $W = \{W(x); 0 < x < 1\}$ be the Gaussian process with covariance function mentioned in (2.6).

**Lemma 3.3.** *Under assumption of Theorem 2.2, we have*

$$\sqrt{n} R_n(\cdot) \xrightarrow{\mathcal{D}} \frac{W(G(\cdot))}{(1-F(\cdot))^2} I_{[0,\tau)}(\cdot) in D[0,\infty)$$

*where $R_n(\cdot)$ is Eq. (2.5).*

*Proof.* It is easy to see that

$$\begin{aligned}
R_n(x) &= \mu_n \frac{I_{[0,Y_{(n)})}(x)}{1-F_n(x)} - \mu \frac{I_{[0,\tau)}(x)}{1-F(x)} \\
&= \mu_n I_{[0,Y_{(n)})}(x) \left[ \frac{1}{1-F_n(x)} - \frac{1}{1-F(x)} \right] \\
&\quad + (\mu_n - \mu) \frac{I_{[0,Y_{(n)})}(x)}{1-F(x)} - \mu \frac{I_{[Y_{(n)},\tau)}(x)}{1-F(x)} \\
&= \mu_n I_{[0,Y_{(n)})} \frac{(F_n(x) - F(x))}{(1-F(x))(1-F_n(x))} + (\mu_n - \mu) \frac{I_{[0,Y_{(n)})}(x)}{1-F(x)} \\
&\quad - \mu \frac{I_{[Y_{(n)},\tau)}(x)}{1-F(x)} \\
&=: R_{n1}(x) + R_{n2}(x)
\end{aligned}$$

To deal with $R_{n2}(x)$, it follows from Lemma 3.2 that

$$\sqrt{n} R_{n2}(x) = -\sqrt{n} \mu \frac{I_{[Y_{(n)},\tau)}}{1-F(x)} \xrightarrow{\mathcal{D}} 0$$

Next, following the notations of Sen (1984, p. 65),

$$\begin{aligned}
\sqrt{n} R_{n1}(x) &= \sqrt{n} \frac{I_{[0,Y_{(n)})}}{(1-F(x))(1-F_n(x))} [\mu_n (F_n(x) - F(x)) + (\mu_n - \mu)(1 - F_n(x))] \\
&= \frac{I_{[0,Y_{(n)})}}{(1-F(x))(1-F_n(x))} [\mu_n W_n^*(G(x)) + \sqrt{n}(\mu_n - \mu)(1 - F_n(x))]
\end{aligned}$$

$$= \frac{I_{[0,Y_{(n)})}}{(1-F(x))(1-F_n(x))}[W_n^{**}(G(x))],$$

where

$$W_n^{**}(G(x)) = \mu_n W_n^*(G(x)) + (1-F_n(x))\mu^2 \int_0^\infty \frac{1}{y} d[\sqrt{n}(G_n(y) - G(y))].$$

Similar to the proof of Relation (4.9) in the proof of weak convergence of $W_n^*$ in Sen (1984), it can be shown by employing weak convergence of $\sqrt{n}[G_n(\cdot) - G(\cdot)]$ to a tied-down Wiener process $B = \{B(t); 0 < t < 1\}$ that

$$\sqrt{n}R_{n1}(x) \xrightarrow{\mathcal{D}} \frac{I_{[0,\tau)}(x)}{(1-F(x))^2} W(G(x)),$$

where

$$W(t) = \frac{\mu^2}{Q_G(t)} B(t) + \mu^2 \int_0^{Q_G(t)} \frac{1}{Q_G(s)^2} B(s) dQ_G(s)$$
$$+ (1 - 2F(Q_G(t)))\mu^3 \int_0^1 \frac{1}{Q_G(s)^2} B(s) dQ_G(s), \tag{3.4}$$

and $Q_G(t)$ is the quntile function of $G$ for $0 < t < 1$. Note that covariance function of $W(t)$ is given by (2.6). The proof of Lemma 3.3 is now complete. □

*Proof of Theorem 2.2.* The result follows from representation (2.4), Lemma 3.3 and Proposition 2 of Bae and Kim (2006).

## Acknowledgments

The author would like to sincerely thank an anonymous referee for the careful reading of the manuscript.

## Funding

This research was supported by a grant from Ferdowsi University of Mashhad; (No. MS90229FAK).

## References

Bae, J., Kim, S. (2006). Uniform asymptotics in the empirical mean residual life process. *J. Kor. Math. Soc.* 43(2):225–239.

Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley.

Blumenthal, S. (1967). Proportional sampling in life length studies. *Technometrics* 9:205–218.

Colman, R. (1979). An introduction to mathematical sterology. *Memoirs No. 3*, Dept. of Theoretical Statistics, Univ. of Aarhus, Denmark.

Cox, D. R. (1969). Some sampling problems in technology. In: Johnson, N. L., Smith, H., Eds. , *New Developments in Survey Sampling*. New York: Wiley.

Csörgő, M., Zitikis, R. (1996). Mean residual life processes. *Ann. Statist.* 24(4):1717–1739.

de Uña-Álvarez, J. (2004). Nonparametric estimation under length-biased sampling and Type I censoring: a moment based approach. *Ann. Instit. Statist. Math.* 56:667–681.

Hall, W. J., Wellner, J. A. (1979). Estimation of mean residual life. Unpublished Manuscript. University of Rochester.

Horváth, L. (1985). Estimation from a length-biased distribution. *Statist. Decis.* 3:91–113.

McFadden, J. A. (1962). On the lengths of intervals in a stationary point process. *J. Roy. Statist. Soc. Ser. B* 24:364–382.

Patil, G. P., Rao, C. R. (1977). The weighted distributions: A survey their applications. In: Krishnaiah, P. R., Eds. , *Applications of Statistics*. Amesterdam: North Holland, pp. 383–405.

Patil, G. P., Rao, C. R. (1978). Weighted distributions and size-biased of sampling with applications to wildlife populations and human families. *Biometrics* 34:179–189.

Sen, P. K. (1984). On asymptotic representations for reduced quantiles in sampling from a length-biased distribution. *Calcutta Statist. Assoc. Bull.* 33:59–67.

Vardi, Y. (1982a). Nonparametric estimation in the presence of length-bias. *Ann. Statist.* 10:616–620.

Vardi, Y. (1982b). Nonparametric estimation in renewal processes. *Ann. Statist.* 10:772–785.

Wang, M. C. (1991). Nonparametric estimation from cross-sectional survival data. *J. Amer. Statist. Assoc.* 86:130–143.

Wicksell, S. D. (1925). The corpuscle problem. A mathematical study of a biometric problem. *Biometrika* 17(1/2):84–99.

Yang, G. L. (1978). Estimation of a biometric function. *Ann. Statist.* 6(1):112–116.