Statistics and Probability Letters 104 (2015) 133–140

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Berry–Esseen bounds for the percentile residual life function estimators

Mu Zhao [a,b], Hongmei Jiang [a,*]

[a] *Department of Statistics, Northwestern University, Evanston, IL, 60208, USA*

[b] *The School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, 430073, China*

ARTICLE INFO

Article history:
Received 10 October 2014
Received in revised form 10 May 2015
Accepted 10 May 2015
Available online 21 May 2015

Keywords:
Berry–Esseen bound
Covering probability
Deficiency
Empirical estimator
Kernel estimator
Percentile residual life function

ABSTRACT

The Berry–Esseen bounds for two estimators of the percentile residual life function are established. The bound for the kernel estimator is shown sharper than in the previous work. The obtained bounds are applied to study the relative deficiency of the proposed estimators.

1. Introduction

The mean residual life function (MRLF) of a random variable T at time t, which is defined by $E(T - t|T > t)$, has been frequently used in the fields of biometry, actuarial studies and reliability. It is a very useful tool to describe the remaining life of a subject given that the subject has survived up to time t. However, Schmittlein and Morrison (1981) pointed out that the MRLF has a number of practical drawbacks. For example, the estimated mean residual life may be unstable due to the influence of outliers, and the MRLF even may not exist in some cases. Therefore, as an alternative, Haines and Singpurwalla (1974) originally introduced the percentile residual life function (PRLF).

Let F be the distribution function of T with support $(0, b_F)$, where $b_F = \sup\{t > 0 : F(t) < 1\} \leqslant \infty$. Let Q be the corresponding quantile function with $Q(p) = \inf\{t : F(t) \geq p\}$ for $0 < p < 1$ and $Q(0) = 0, Q(1) = b_F$. For $0 < p < 1$, the $(1 - p)$ PRLF at t is defined as

$$R^{(p)}(t) = Q(1 - p(1 - F(t))) - t. \tag{1.1}$$

Suppose that $X_1, X_2, \ldots, X_n$ are identically and independently distributed samples from F. Then, a natural estimator of $R^{(p)}(t)$ is its empirical analog

$$R_n^{(p)}(t) = Q_n(1 - p(1 - F_n(t))) - t, \tag{1.2}$$

where F_n and Q_n are the sample distribution function and the sample quantile function respectively, and $Q_n(1) = X_{(n)}$ with $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$. Csörgő and Csörgő (1987) showed that $R_n^{(p)}(t)$ is uniformly (in t and p) consistent and

* Corresponding author.
E-mail address: hongmei@northwestern.edu (H. Jiang).

http://dx.doi.org/10.1016/j.spl.2015.05.009
0167-7152/

$\sqrt{n}(R_n^{(p)}(t) - R^{(p)}(t)) \xrightarrow{\mathcal{D}} N(0, \sigma^2(p,t))$, where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, and

$$\sigma^2(p,t) = \frac{p(1-p)(1-F(t))}{f^2(Q(1-p(1-F(t))))}.$$

However, considering efficiency, numerous authors suggest kernel type estimator to substitute the empirical one. For example, Parzen (1979), Reiss (1981), Padgett (1986), Sheather and Marron (1990), and Xiang (1995b), among others, proposed the kernel smoothing distribution function and quantile function respectively. For the PRLF, the kernel type estimator is of the form

$$\widehat{R}_n^{(p)}(t) = h_n^{-1} \int_0^1 R_n^{(u)}(t) K\left(\frac{u-p}{h_n}\right) du, \tag{1.3}$$

where $K(\cdot)$ is a kernel function, and h_n is a sequence of bandwidths satisfying $h_n \to 0$ as $n \to \infty$. Under some general conditions, Zhao et al. (submitted for publication) showed that $\widehat{R}_n^{(p)}(t)$ is also asymptotically normally distributed with the same variance function $\sigma^2(p,t)$.

Hodges and Lehmann (1970) firstly introduced the conception of *deficiency* for comparing two different estimating procedures. Suppose a less effective procedure requires k_n observations to give equally good performance as a statistical procedure based on n observations. The ratio $e = \lim_{n\to\infty} n/k_n$ which is known as *asymptotic relative efficiency* is the most common quantity used for this comparison. However, in many important statistical problems, e equals 1. In this situation, the limit value $d = \lim_{n\to\infty}(k_n - n)$ becomes a useful measure due to that d summarizes the comparison more revealingly than e. Hodges and Lehmann (1970) named d as *asymptotic deficiency*. The deficiencies of empirical distribution (quantile) estimator with respect to a type of kernel estimators under the criteria of equal mean squared error, mean absolute error and covering probability have been established by many authors. A short and incomplete list includes Falk (1984, 1985), Xiang (1995a,b), Lemdani and Ould-Saïd (2003) and Zhao et al. (2011, 2013).

In this paper, we first establish two Berry–Esseen theorems for $R_n^{(p)}(t)$ and $\widehat{R}_n^{(p)}(t)$ respectively. Based on these two theorems, we give the relative deficiency of $R_n^{(p)}(t)$ with respect to $\widehat{R}_n^{(p)}(t)$ under the criterion of equal covering probability. The paper is organized as follows. Section 2 gives the main results. In Section 3, we study the deficiency of the sample estimator with respect to its kernel smoothing counterpart and provide some simulation results. Section 4 gives some concluding remarks. The proofs of the main results are deferred to Appendix A.

2. Main results

We assume that the functions $F(\cdot)$, $K(\cdot)$ and $Q(\cdot)$ satisfy the following conditions:

A1. For fixed p and t, F has a positive continuous density f and a bounded second derivative $F^{(2)}$ in a neighborhood of $1-p(1-F(t))$;
A2. $K(x)$ is Lipschitz of order 1 and has compact support on $[-1, 1]$;
A3. $K(x)$ is a kernel function of $(r+1)$th order with $r \geq 1$. Hence, $\int_{-1}^1 K(x)dx = 1$, $\int_{-1}^1 x^j K(x)dx = 0$, for $j = 1, \ldots, r$, and $\int_{-1}^1 x^{r+1}K(x)dx = c_r$, where c_r is a constant;
A4. For fixed p and t, $Q(\cdot)$ has a bounded $(r+1)$th derivative in a neighborhood of $1-p(1-F(t))$;
A5. $nh_n^{2(r+1)} \to 0$ as $n \to \infty$.

The following two theorems give the Berry–Esseen bounds for the empirical estimator and the kernel estimator of the percentile residual life function respectively.

Theorem 1. *Assume that assumption* (A1) *holds. Then, for* $0 < p < 1$, $0 \leqslant t < b_F$, *as* $n \to \infty$,

$$\sup_{-\infty<x<\infty} \left| P\left(\frac{\sqrt{n}(R_n^{(p)}(t) - R^{(p)}(t))}{\sigma(p,t)} \leqslant x\right) - \Phi(x)\right| = O(n^{-\frac{1}{2}}), \tag{2.1}$$

where $\Phi(\cdot)$ *is the d.f. of standard normal distribution.*

To give the Berry–Esseen bound for $\widehat{R}_n^{(p)}(t)$, we denote

$$\Omega_{n1} = \frac{1}{\sqrt{n}h_n} \int_t^\infty K\left(\frac{\frac{1-F(x)}{1-F(t)} - p}{h_n}\right) \frac{(1-F(x))B(F(t))}{(1-F(t))^2} dx, \tag{2.2}$$

$$\Omega_{n2} = -\frac{1}{\sqrt{n}h_n} \int_t^\infty K\left(\frac{\frac{1-F(x)}{1-F(t)} - p}{h_n}\right) \frac{B(F(x))}{1-F(t)} dx, \tag{2.3}$$