



Research article

Statistical inference of an α -quantile past lifetime function with applications

Mohamed Kayid*

Department of Statistics and Operations Research, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia

* **Correspondence:** Email: drkayid@ksu.edu.sa; Tel: +9-661-467-6259; Fax: +9-661-467-6274.

Abstract: In reliability engineering and survival analysis, quantile functions are fundamental and often the most natural way to represent probability distributions and data samples. In this paper, the α -quantile function of past lifetime was estimated for right-censored data by applying the Kaplan-Meier survival estimator. The weak convergence of the proposed estimator to a Gaussian process was investigated. A confidence interval for the α -quantile of the past life function that does not depend on the density function was proposed. The strong convergence of the estimator to a Gaussian process was also discussed. The properties of the estimator and the confidence interval were investigated in a simulation study. Finally, two real datasets were analyzed.

Keywords: gaussian process; quantile past lifetime function; right censored data; survival analysis; Kaplan-Meier survival estimator

Mathematics Subject Classification: 62N01, 62N05

1. Introduction

Quantile-based reliability analysis is a novel approach in reliability theory that uses quantile functions as opposed to the traditional approach based on distribution functions. Quantile and distribution functions are mathematically equivalent methods for defining probability distributions. For a positive random variable T , the conditional past lifetime (also called inactivity time) at time t , $T_t = t - T|T < t$, is a dual for the residual lifetime and is the cornerstone of many studies in survival analysis and reliability theory. The α -quantile past lifetime (α -QPL) function is defined to be the α

quantile of the conditional past lifetime T_t and can be simplified to

$$q_\alpha(t) = t - Q(\bar{\alpha}F(t)), \quad t \geq 0, \quad (1.1)$$

where $\bar{\alpha} = 1 - \alpha$ and $Q(y) = \inf\{x: F(x) \geq y\}$ is called the quantile function of F . Unnikrishnan and Vineshkumar [1] examined some properties of this measure and its relationship to the reversed hazard rate function. Shafaei and Izadkhah [2] defined the α -QPL concept for parallel systems and investigated its properties. Shafaei [3] showed that the distribution can be characterized by two appropriate quantiles past life functions. Mahdy [4] investigated some ordering characteristics of the α -QPL function and discussed the problem of estimating this measure for uncensored data using an empirical distribution function. Balmert and Jeong [5] focused on the nonparametric estimation of the special case of 0.5-QPL, referred to as the median past life function. They also compared two or more groups of data based on the ratio of their median-quantile past life functions. In addition, Balmert et al. [6] investigated a log-linear quantile regression model for inactivity time when the data are right-censored.

The applicability and usefulness of α -QPL, discussed in two later references, motivated us to discuss the problem of its estimation for right-censored data and to investigate the statistical properties of the estimator. Another point that motivates me is that the median of the past lifetime may be preferred to the mean past lifetime, especially when the data are heavily censored or skewed or the moments do not exist. However, there is a very little literature on the topic of estimating the α -QPL function. In my approach, the Kaplan-Meier (KM) survival estimator is used to define the estimator as a continuous process in time. The rest of this paper is organized as follows. Section 2 reviews some preliminary remarks, including notations, the KM survival estimator and related processes. Section 3 defines the estimator of the α -QPL function. Its weak convergence to a Gaussian process, confidence intervals for the α -QPL function, and the strong convergence of the estimator are also topics that are investigated. Section 4 investigates the behavior of the estimator and the proposed confidence interval (CI) by simulating Gamma and Weibull models (see for example Teamah Abd-Elmonem et al. [7] and Elbatal et al. [8]). Section 5 presents applications to two datasets from the Mayo Clinic Primary Biliary Cirrhosis Study and North Central Cancer. Finally, in Section 6, I conclude the paper and suggest directions for future work.

2. The KM survival estimator and related processes

Let T_i , $i = 1, 2, \dots, n$ be n iid instances following the distribution function F which are censored by random censorship variable C with distribution function H . The random instance T_i is uncensored when $T_i \leq C_i$ and otherwise it is censored. Then the observations are n pairs (X_i, δ_i) where $X_i = T_i \wedge C_i$ and $\delta_i = I(T_i \leq C_i)$ is the censoring indicator. Let G show the distribution of X_i . Through the paper, let $b_F = \inf\{x: F(x) = 1\}$ shows the upper bound of the support of the distribution F . Similarly denote b_H and b_G for distributions H and G , respectively. Clearly, $b_G = b_F \wedge b_H$ where \wedge stands for the minimum operator. The reliability functions corresponding to F , H , and G , which are represented by \bar{F} , \bar{H} , and \bar{G} , respectively, are related by

$$\bar{G}(t) = \bar{F}(t)\bar{H}(t).$$

A concept that proves quite useful in our investigation is the probability that an event occurs before or at t and is uncensored, i.e.,

$$\tilde{F}(t) = P(X_i \leq t, \delta_i = 1) = \int_0^t \bar{H}(x) dF(x). \quad (2.1)$$

Let $N_i(t) = I(X_i \leq t, \delta_i = 1)$ and $Y_i(t) = I(X_i \geq t)$, then $\bar{N}(t) = \sum_{i=1}^n N_i(t)$ shows the number of items failed up to or at time t and $\bar{Y}(t) = \sum_{i=1}^n Y_i(t)$ represents the number of items at risk at time t . Also, let $\Delta\bar{N}(t) = \bar{N}(t) - \bar{N}(t^-)$. The KM estimator, also known as the product-limit (PL) estimator, of the survival function is given by

$$\bar{F}_n(t) = \prod_{X_i \leq t} \left(1 - \frac{\Delta\bar{N}(X_i)}{\bar{Y}(X_i)}\right), \quad 0 \leq t \leq X_{(n)}, \quad (2.2)$$

where $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$. It is widely used for estimating the survival function of right-censored data. The PL estimator of the distribution function is simply given by $F_n(t) = 1 - \bar{F}_n(t)$. A satisfactory property of this estimator is that it is asymptotically unbiased, more precisely,

$$E(\bar{F}_n(t) - \bar{F}(t)) \leq F(t)G^n(t). \quad (2.3)$$

This shows that the bias can be considerable for large values of t and/or heavily censored data. In the following, I discuss some necessary processes. More details can be found in Chung [9]. The PL process is defined by

$$\beta_n(t) = \sqrt{n}(F_n(t) - F(t)). \quad (2.4)$$

Burke et al. [10] showed that there is a sequence of Wiener processes $\{W_n(t), t \geq 0\}$, such that for any $t^* < b_G$

$$\sup_{0 < t < t^*} |\beta_n(t) - \bar{F}(t)W_n(d(t))| = O\left(n^{-\frac{1}{2}}(\log n)^2\right), \quad \text{almost surely}, \quad (2.5)$$

where

$$d(t) = \int_0^t \bar{G}^{-2}(x) d\tilde{F}(x),$$

and \tilde{F} is defined in (2.1). Moreover, (2.5) implies the following law of iterated logarithm for the PL-process, see Burke et al. [10] and Csorgo and Horvath [11].

$$\sup_{0 < t < t^*} |\beta_n(t)| = O\left((\log \log n)^{\frac{1}{2}}\right), \quad \text{almost surely}. \quad (2.6)$$

Also, Burke et al. [10] showed that

$$\sup_{0 < t < t^*} |d_n(t) - d(t)| = O\left(n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}\right), \quad \text{almost surely}, \quad (2.7)$$

in which

$$d_n(t) = \int_0^t \bar{G}_n^{-2}(x) d\tilde{F}_n(x),$$

and \bar{G}_n and \tilde{F}_n are their empirical functions, i.e.,

$$\bar{G}_n(t) = \frac{1}{n} \sum_{i=1}^n I(X_i > t),$$

and

$$\tilde{F}_n(t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t, \delta_i = 1).$$

Applying the PL estimator of the distribution function, the PL quantile process is defined by

$$Q_n(y) = \inf\{x: F_n(x) \geq y\}, \quad 0 \leq y \leq 1.$$

The PL-normed quantile process is defined by

$$\rho_n(t) = \sqrt{n}f(Q(y))(Q(y) - Q_n(y)), \quad 0 < y < 1, \quad (2.8)$$

where f is the density function of F . Aly et al. [12] showed that for $p^* \in (0,1)$ with $Q(p^*) < b_G$, there is a sequence of Wiener processes $\{W_n(t), t \geq 0\}$ such that

$$\sup_{0 < y < p^*} \left| \rho_n(y) - (1-y)W_n(d(Q(y))) \right| = O\left(n^{-\frac{1}{4}}(\log n)^{\frac{1}{2}}(\log \log n)^{\frac{1}{4}}\right). \quad (2.9)$$

The uniform PL-process is defined by

$$e_n(y) = \sqrt{n}(F_n^*(y) - y), \quad 0 \leq y \leq 1, \quad (2.10)$$

where $F_n^*(y) = F_n(Q(y))$ could be computed by PL estimator of the (W_i, δ_i) and $W_i = F(X_i)$. Then,

$$Q_n^*(p) = \inf\{y: F_n^*(y) \geq p\}, \quad 0 < p < 1,$$

and the uniform PL-quantile process is defined to be

$$u_n(x) = \sqrt{n}(x - Q_n^*(x)), \quad 0 \leq x \leq 1. \quad (2.11)$$

Let b_G^* be the upper bound of the distribution of W_i . Aly et al. [12] proved that for $y_0 < b_G^*$ such that $Q(y_0) < b_G$, there are sequences of Wiener processes $\{W_n(t), t \geq 0\}$ such that

$$\sup_{0 < y < y_0} \left| e_n(y) - (1-y)W_n(d(Q(y))) \right| = O\left(n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}\right), \quad \text{almost surely,} \quad (2.12)$$

and

$$\sup_{0 < y < y_0} \left| u_n(y) - (1-y)W_n(d(Q(y))) \right| = O\left(n^{-\frac{1}{4}}(\log n)^{\frac{1}{2}}(\log \log n)^{\frac{1}{4}}\right), \quad \text{almost surely.} \quad (2.13)$$

These equations imply the following law of iterated logarithm relations respectively.

$$\sup_{0 < y < y_0} |e_n(y)| = O\left((\log \log n)^{\frac{1}{2}}\right), \quad \text{almost surely,} \quad (2.14)$$

and

$$\sup_{0 < y < y_0} |u_n(y)| = O\left((\log \log n)^{\frac{1}{2}}\right), \quad \text{almost surely.} \quad (2.15)$$

For more details about these processes, refer to Chung [9].

3. The α -QPL estimator

The estimator of the α -QPL function is defined as in the following.

$$q_{\alpha,n}(t) = t - F_n^{-1}(\bar{\alpha}F_n(t)), \quad 0 < \alpha < 1, 0 \leq t \leq X_{(n)}. \quad (3.1)$$

Similar to the process adopted for the α -QRL function by Csorgo and Csorgo [13] and Chung [9], the scaled α -QPL process $r_n^\alpha(t)$ can be defined by

$$r_n^\alpha(t) = \sqrt{n}f(Q(\bar{\alpha}F(t)))(q_{\alpha,n}(t) - q_\alpha(t)), \quad 0 < t < X_{(n)}. \quad (3.2)$$

The following result shows that this process converges weakly to a Gaussian process.

Theorem 1. Assume that $0 < \alpha < 1$, $0 \leq t < b_G$, $Q(\bar{\alpha}F(t)) < b_G$ and the density-quantile function $f(Q(\cdot))$ is continuous at point $\bar{\alpha}F(t)$. Then, I have

$$r_n^\alpha(t) \rightarrow N(0, \sigma_{\alpha,t}^2), \quad \text{in distribution}, \quad (3.3)$$

where $\sigma_{\alpha,t}^2 = (1 - \bar{\alpha}F(t))^2 d(Q(\bar{\alpha}F(t))) + \bar{\alpha}^2 \bar{F}^2(t) d(t) - 2\bar{\alpha}(1 - \bar{\alpha}F(t))\bar{F}(t) d(Q(\bar{\alpha}F(t)))$.

Proof. It could be checked easily that $Q_n^*(p) = F(Q_n(p))$ for all $0 < p < 1$. Then, by substituting $F(Q_n(\bar{\alpha}F_n(t)))$ by $Q_n^*(\bar{\alpha}F(t))$, and applying Taylor expansion on the Q function, for $0 < t < X_{(n)}$, I have

$$\begin{aligned} r_n^\alpha(t) &= \sqrt{n}f(Q(\bar{\alpha}F(t)))(q_{\alpha,n}(t) - q_\alpha(t)) \\ &= \sqrt{n}f(Q(\bar{\alpha}F(t)))(Q(\bar{\alpha}F(t)) - Q(F(Q_n(\bar{\alpha}F_n(t)))) \\ &= \sqrt{n}f(Q(\bar{\alpha}F(t)))(Q(\bar{\alpha}F(t)) - Q(Q_n^*(\bar{\alpha}F_n(t)))) \\ &= \frac{f(Q(\bar{\alpha}F(t)))}{f(Q(\delta_{n,t,\alpha}))} \sqrt{n}(\bar{\alpha}F(t) - Q_n^*(\bar{\alpha}F_n(t))) \\ &= \frac{f(Q(\bar{\alpha}F(t)))}{f(Q(\delta_{n,t,\alpha}))} s_n^\alpha(t), \end{aligned} \quad (3.4)$$

where $\delta_{n,t,\alpha}$ lies between $\bar{\alpha}F(t)$ and $Q_n^*(\bar{\alpha}F_n(t))$ and $s_n^\alpha(t) = \sqrt{n}(\bar{\alpha}F(t) - Q_n^*(\bar{\alpha}F_n(t)))$. Clearly $\delta_{n,t,\alpha} \rightarrow \bar{\alpha}F(t)$ as $n \rightarrow \infty$ and by continuity of the density quantile function

$$\frac{f(Q(\bar{\alpha}F(t)))}{f(Q(\delta_{n,t,\alpha}))} \rightarrow 1, \quad \text{as } n \rightarrow \infty.$$

Then the well-known Slutsky's theorem implies that the asymptotic distribution of the $r_n^\alpha(t)$ and $s_n^\alpha(t)$ is the same. Now, by adding and subtracting the expression $\bar{\alpha}F_n(t)$, I have

$$\begin{aligned} s_n^\alpha(t) &= \sqrt{n}(\bar{\alpha}F(t) - \bar{\alpha}F_n(t) + \bar{\alpha}F_n(t) - Q_n^*(\bar{\alpha}F_n(t))) \\ &= u_n(\bar{\alpha}F_n(t)) - \bar{\alpha}e_n(F(t)). \end{aligned} \quad (3.5)$$

Then, applying (3.5), (2.12) and (2.13), I have

$$\left| s_n^\alpha(t) - \left((1 - \bar{\alpha}F(t))W_n \left(d(Q(\bar{\alpha}F(t))) \right) - \bar{\alpha}\bar{F}(t)W_n(d(t)) \right) \right| \rightarrow 0, \quad \text{almost surely}, \quad (3.6)$$

and in turn

$$s_n^\alpha(t) \rightarrow (1 - \bar{\alpha}F(t))W(d(Q(\bar{\alpha}F(t)))) - \bar{\alpha}\bar{F}(t)W(d(t)). \quad (3.7)$$

Note that $Q(\bar{\alpha}F(t)) \leq t$, so $d(Q(\bar{\alpha}F(t))) \leq d(t)$, and by properties of the Wiener processes, it results that

$$s_n^\alpha(t) \rightarrow N(0, \sigma_{\alpha,t}^2), \text{ in distribution,} \quad (3.8)$$

where $\sigma_{\alpha,t}^2$ is defined previously.

Let $z_{p/2}$ be the upper $\frac{p}{2}$ -quantile of the standard normal distribution, i.e., for the standard normal random variable Z , $P\left(Z > z_{\frac{p}{2}}\right) = p/2$. Using Theorem 1, I can construct a CI for the α -QPL function, which is described in the following theorem.

Theorem 2. Let $0 < \alpha < 1, 0 \leq t < b_G$, $Q(\bar{\alpha}F(t)) < b_G$ and the density-quantile function $f(Q())$ is continuous at point $\bar{\alpha}F(t)$. In addition, I assume that f_n is a consistent estimator of the density function f in the neighborhood of $Q(\bar{\alpha}F(t))$. Then, an asymptotic $(1 - p)$ percent CI for $q_\alpha(t)$ is

$$\left(q_{\alpha,n}(t) - c_n(t, p), q_{\alpha,n}(t) + c_n(t, p) \right), \quad (3.9)$$

where

$$c_n(t, p) = \frac{z_{p/2} \sqrt{\sigma_{\alpha,t,n}^2}}{\sqrt{n} f_n(Q_n(\bar{\alpha}F_n(t)))}$$

and

$$\begin{aligned} \sigma_{\alpha,t,n}^2 &= (1 - \bar{\alpha}F_n(t))^2 d_n(Q_n(\bar{\alpha}F_n(t))) + \bar{\alpha}^2 \bar{F}_n^2(t) d_n(t) \\ &\quad - 2\bar{\alpha}(1 - \bar{\alpha}F_n(t)) \bar{F}_n(t) d_n(Q_n(\bar{\alpha}F_n(t))). \end{aligned}$$

A consistent estimator of the density function f should be used to calculate the CI. A fundamental problem in statistics is the estimation of the density function, which many have already addressed. Specifically, for censored data, I can refer to Blum and Susarla [14], Burke and Horvath [15], Mielniczuk [16], Marron and Padgett [17], Lo et al. [18], and many other studies. The following theorem provides a different approach to constructing a CI for the α -QPL function that does not depend on the density function. This idea is similar to that of Csorgo and Csorgo [13] and Chung [9] for the α -quantile residual life function.

Theorem 3. Let $0 < \alpha < 1$, t satisfy the conditions of Theorem 2 and the density-quantile function $f(Q())$ is continuous at point $\bar{\alpha}F(t)$. Then, an asymptotic $(1 - p)$ percent CI for $q_\alpha(t)$ is

$$\left(t - Q_n \left(\bar{\alpha}F_n(t) + n^{-\frac{1}{2}} z_{\frac{p}{2}} \sqrt{\sigma_{\alpha,t,n}^2} \right), t - Q_n \left(\bar{\alpha}F_n(t) - n^{-\frac{1}{2}} z_{\frac{p}{2}} \sqrt{\sigma_{\alpha,t,n}^2} \right) \right), \quad (3.10)$$

where $\sigma_{\alpha,t,n}^2$ is defined in Theorem 2.

Proof. For simplicity, take $u_n = \bar{\alpha}F_n(t) - n^{-\frac{1}{2}}z_{p/2}\sqrt{\sigma_{\alpha,t,n}^2}$ and $v_n = \bar{\alpha}F_n(t) + n^{-\frac{1}{2}}z_{p/2}\sqrt{\sigma_{\alpha,t,n}^2}$. I should have

$$\lim_{n \rightarrow \infty} P(t - Q_n(v_n) \leq q_\alpha(t) \leq t - Q_n(u_n)) = 1 - p. \quad (3.11)$$

This probability can be written as in the following.

$$\begin{aligned} P(Q_n(u_n) \leq Q(\bar{\alpha}F(t)) \leq Q_n(v_n)) \\ = P(U_n \cap V_n), \end{aligned} \quad (3.12)$$

where

$$\begin{aligned} U_n &= \{Q_n(u_n) \leq Q(\bar{\alpha}F(t))\} \\ &= \{Q_n(u_n) - Q(u_n) + Q(u_n) - Q(\bar{\alpha}F(t)) \leq 0\} \\ &= \{-\rho_n(u_n) + \sqrt{n}f(Q(u_n))(Q(u_n) - Q(\bar{\alpha}F(t))) \leq 0\} \\ &= \left\{-\rho_n(u_n) + \sqrt{n}\frac{f(Q(u_n))}{f(Q(\delta_n))}(u_n - \bar{\alpha}F(t)) \leq 0\right\}. \end{aligned} \quad (3.13)$$

In the last equality, I used Taylor expansion for Q and δ_n lies between u_n and $\bar{\alpha}F(t)$. Similarly

$$\begin{aligned} V_n &= \{Q_n(v_n) \geq Q(\bar{\alpha}F(t))\} \\ &= \{Q_n(v_n) - Q(v_n) + Q(v_n) - Q(\bar{\alpha}F(t)) \geq 0\} \\ &= \{-\rho_n(v_n) + \sqrt{n}f(Q(v_n))(Q(v_n) - Q(\bar{\alpha}F(t))) \geq 0\} \\ &= \left\{-\rho_n(v_n) + \sqrt{n}\frac{f(Q(v_n))}{f(Q(\delta'_n))}(v_n - \bar{\alpha}F(t)) \geq 0\right\}, \end{aligned} \quad (3.14)$$

where δ'_n lies between v_n and $\bar{\alpha}F(t)$. Note that u_n , v_n , δ_n and δ'_n all converge to $\bar{\alpha}F(t)$ as $n \rightarrow \infty$. Now, combining (3.13) and (3.14), I have

$$\begin{aligned} \lim_{n \rightarrow \infty} P(U_n \cap V_n) &= \lim_{n \rightarrow \infty} P\left(\left\{-\rho_n(u_n) + \sqrt{n}\frac{f(Q(u_n))}{f(Q(\delta_n))}(u_n - \bar{\alpha}F(t)) \leq 0\right\} \cap \right. \\ &\quad \left.\left\{-\rho_n(v_n) + \sqrt{n}\frac{f(Q(v_n))}{f(Q(\delta'_n))}(v_n - \bar{\alpha}F(t)) \geq 0\right\}\right) \\ &= \lim_{n \rightarrow \infty} P\left(\left\{-\rho_n(u_n) + \sqrt{n}(\bar{\alpha}F_n(t) - n^{-\frac{1}{2}}z_{p/2}\sqrt{\sigma_{\alpha,t,n}^2} - \bar{\alpha}F(t)) \leq 0\right\} \cap \right. \\ &\quad \left.\left\{-\rho_n(v_n) + \sqrt{n}(\bar{\alpha}F_n(t) + n^{-\frac{1}{2}}z_{p/2}\sqrt{\sigma_{\alpha,t,n}^2} - \bar{\alpha}F(t)) \geq 0\right\}\right) \\ &= \lim_{n \rightarrow \infty} P\left(-z_{\frac{p}{2}}\sqrt{\sigma_{\alpha,t}^2} \leq -\rho_n(\bar{\alpha}F(t)) + \bar{\alpha}e_n(F(t)) \leq z_{\frac{p}{2}}\sqrt{\sigma_{\alpha,t}^2}\right). \end{aligned} \quad (3.15)$$

The second equality is due to the fact that $\frac{f(Q(v_n))}{f(Q(\delta'_n))} \rightarrow 1$ and $\frac{f(Q(u_n))}{f(Q(\delta_n))} \rightarrow 1$ as $n \rightarrow \infty$. The third equality follows by the fact that u_n and v_n converge to $\bar{\alpha}F(t)$ and $\sigma_{\alpha,t,n}^2$ converges to $\sigma_{\alpha,t}^2$ as $n \rightarrow \infty$. On the other hand, the asymptotic distribution of $-\rho_n(\bar{\alpha}F(t)) + \bar{\alpha}e_n(F(t))$ is same as the

$$(1 - \bar{\alpha}F(t))W_n(d(Q(\bar{\alpha}F(t)))) + \bar{\alpha}\bar{F}(t)W_n(d(t)), \quad (3.16)$$

which by (3.6), is same with distribution of $s_n^\alpha(t)$ as $n \rightarrow \infty$. Thus, by Theorem 1, I have

$$\lim_{n \rightarrow \infty} P(U_n \cap V_n) = 1 - \alpha, \quad (3.17)$$

and this completes the proof.

3.2 Strong convergence

Here, I show that the defined process in (3.1) converges almost surely to $q_\alpha(t)$ under some mild conditions. The following result proves that the scaled process $r_n^\alpha(t)$ can be approximated by a Gaussian process with zero mean.

Let $G_n^\alpha(t)$ be defined by

$$G_n^\alpha(t) = (1 - \bar{\alpha}F(t))W_n(d(Q(\bar{\alpha}F(t)))) - \bar{\alpha}\bar{F}(t)W_n(d(t)). \quad (3.18)$$

where $\{W_n(t), t \geq 0\}$ is a sequence of Wiener processes converging to Wiener process $W(t)$. Also, let $G^\alpha(t)$ be defined by

$$G^\alpha(t) = (1 - \bar{\alpha}F(t))W(d(Q(\bar{\alpha}F(t)))) - \bar{\alpha}\bar{F}(t)W(d(t)). \quad (3.19)$$

Then, for each $n = 1, 2, \dots$, $\{G_n^\alpha(t), t \geq 0\}$ has the same distribution as $\{G^\alpha(t), t \geq 0\}$.

Theorem 4. Let $0 < \alpha < 1$ be fixed, $b \in (0, b_G)$ such that $Q(\bar{\alpha}F(b)) < b_G$, F be twice differentiable on $(0, b_F)$ and f be positive on the interval $(0, Q(\bar{\alpha}))$. Then, as $n \rightarrow \infty$, I have

$$\sup_{0 < t < b} |r_n^\alpha(t) - G_n^\alpha(t)| = O\left(n^{-\frac{1}{4}}(\log n)^{\frac{1}{2}}(\log \log n)^{\frac{1}{4}}\right).$$

Proof. By (3.4), I have

$$r_n^\alpha(t) = \frac{f(Q(\bar{\alpha}F(t)))}{f(Q(\delta_{n,t,\alpha}))} s_n^\alpha(t) = \left(\frac{f(Q(\bar{\alpha}F(t)))}{f(Q(\delta_{n,t,\alpha}))} - 1\right) s_n^\alpha(t) + s_n^\alpha(t), \quad (3.20)$$

where $\delta_{n,t,\alpha}$ lies between $\bar{\alpha}F(t)$ and $Q_n^*(\bar{\alpha}F_n(t))$ and $s_n^\alpha(t) = \bar{\alpha}F(t) - Q_n^*(\bar{\alpha}F_n(t))$. As demonstrated in proof of Theorem 1,

$$s_n^\alpha(t) = u_n(\bar{\alpha}F_n(t)) - \bar{\alpha}e_n(F(t)). \quad (3.21)$$

Now, applying (2.12) and (2.13), I find that

$$\sup_{0 \leq t \leq b} |s_n^\alpha(t) - G_n^\alpha(t)| = O\left(n^{-\frac{1}{4}}(\log n)^{\frac{1}{2}}(\log \log n)^{\frac{1}{4}}\right). \quad (3.22)$$

Recall the relations (2.6) and (2.15) that state, respectively,

$$\sup_{0 \leq t \leq b} \sqrt{n}|F_n(t) - F(t)| = O\left((\log \log n)^{\frac{1}{2}}\right), \quad (3.23)$$

and since $Q(\bar{\alpha}F(b)) < b_G$,

$$\sup_{0 \leq y \leq \bar{\alpha}F(b)} \sqrt{n}|y - Q_n^*(y)| = O\left((\log \log n)^{\frac{1}{2}}\right). \quad (3.24)$$

I have

$$\begin{aligned} \sup_{0 \leq t \leq b} |\delta_{n,t,\alpha} - \bar{\alpha}F(t)| &\leq \sup_{0 \leq t \leq b} n^{-\frac{1}{2}} |s_n^\alpha(t)| \\ &\leq \sup_{0 \leq t \leq b} |\bar{\alpha}F_n(t) - Q_n^*(\bar{\alpha}F_n(t))| + \bar{\alpha} \sup_{0 \leq t \leq b} |F_n(t) - F(t)| \\ &= O\left(n^{-\frac{1}{2}}(\log \log n)^{\frac{1}{2}}\right). \end{aligned} \quad (3.25)$$

The last equality follows from (3.23) and (3.24).

Since the density function f is differentiable and positive on $(0, Q(\bar{\alpha}))$, applying (3.25) I have

$$\sup_{0 \leq t \leq b} |s_n^\alpha(t)| = O\left(n^{-\frac{1}{2}}(\log \log n)^{\frac{1}{2}}\right), \quad (3.26)$$

and

$$\sup_{0 \leq t \leq b} \left| \frac{f(Q(\bar{\alpha}F(t)))}{f(Q(\delta_{n,t,\alpha}))} - 1 \right| = O\left(n^{-\frac{1}{2}}(\log \log n)^{\frac{1}{2}}\right). \quad (3.27)$$

Subsequently, the proof follows from (3.20.), (3.22), (3.26), and (3.27).

The following corollary follows from (2.14), (2.15), (3.2), (3.4) and (3.21).

Corollary 1. Under the conditions of Theorem 4, I have

$$\sup_{0 < t < b} |q_{\alpha,n}(t) - q_\alpha(t)| = O\left(n^{-\frac{1}{2}}(\log \log n)^{\frac{1}{2}}\right).$$

It implies that $q_{\alpha,n}(t)$ converges almost surely to $q_\alpha(t)$.

4. Simulation

The properties of the α -QPL function estimator and CI, defined by Theorem 3, were investigated through simulation studies. For this purpose, two important distributions, Weibull and Gamma, are considered, with the following distribution functions:

$$F(t) = 1 - \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\}, \quad t \geq 0, \quad (4.1)$$

and

$$F(t) = \frac{\gamma(\alpha, \beta t)}{\Gamma(\alpha)}, \quad t \geq 0, \quad (4.2)$$

where $\gamma(\alpha, x) = \int_0^x y^{\alpha-1} e^{-y} dy$ is the lower incomplete gamma function.

It is assumed that the censoring random variable C follows a uniform model in interval $(0, M)$. Based on the censorship percentage, p , I compute M by $M = E(T)/p$ where T is a determined

Weibull or gamma random variable. Some parameter values are selected and $r = 5000$ replicates of samples with sizes $n = 25$ or 50 are simulated. For each sample, the median past lifetime function is computed at four decile points $q_{0.1}$, $q_{0.2}$, $q_{0.4}$, and $q_{0.6}$. Table 1 shows the results for the Weibull model which includes the bias (B) and the mean squared error (MSE). Table 2 presents the simulation results for the gamma model.

From these tables I observe the following:

- When n increases, the MSE decreases significantly, which indicates that the α -QPL estimator is consistent.

- As expected by (2.3), for large deciles, the MSE is affected by the censoring percentage.

In a further simulation study, which is summarized in Table 3, the coverage probability (CP) of the CI defined by Theorem 3 is investigated. Weibull and Gamma distributions, each with a set of parameters, were assumed as the underlying models. Three censoring rates of 0.05, 0.20, and 0.30 and three points $q_{0.1}$, $q_{0.2}$, and $q_{0.4}$. In each run, $r = 5000$ repetitions of samples of size $n=50$ or 100 were simulated and the CI of the 0.5-QPL at the points considered were calculated. The CP, which shows the ratio of CIs containing the true 0.5-QPL to the mean length (ML) of the CIs, was reproduced. The true values of the 0.5-QPL are also shown in the table for comparison.

The results show that the CI coverage is high and acceptable, but ML coverage increases significantly with time. Moreover, the ML coverage increases slightly with the censoring percentage.

Table 1. Simulation results for the Weibull distribution.

Parameters	q	p (censorship)			
		0.05		0.25	
		B	MSE	B	MSE
$\alpha = 10, \beta = 1.2, n = 25$	0.1	0.0411	0.1692	0.0308	0.1577
	0.2	0.0105	0.3414	-0.0426	0.3402
	0.4	-0.1100	0.6936	-0.0536	0.7035
	0.6	-0.1152	1.1710	-0.0389	1.2517
$\alpha = 10, \beta = 0.8, n = 25$	0.1	-0.0033	0.0295	-0.0072	0.0292
	0.2	-0.0159	0.1036	-0.0442	0.1134
	0.4	-0.1429	0.4540	-0.1129	0.4640
	0.6	-0.1700	1.1377	-0.1241	1.1837
$\alpha = 10, \beta = 1.2, n = 50$	0.1	0.0465	0.0975	-0.0117	0.0951
	0.2	0.0282	0.1781	-0.0348	0.1722
	0.4	-0.0078	0.3480	-0.0148	0.3584
	0.6	-0.0195	0.5768	-0.0229	0.6639
$\alpha = 10, \beta = 0.8, n = 50$	0.1	0.0130	0.0162	0.0001	0.0169
	0.2	0.0094	0.0571	-0.0291	0.0614
	0.4	-0.0055	0.2217	-0.0346	0.2365
	0.6	-0.0432	0.5645	-0.0433	0.6000

Table 2. Simulation results for the gamma distribution.

Parameters	q	p (censorship)			
		0.05		0.25	
		B	MSE	B	MSE
$\alpha = 1.5, \beta = 2, n = 25$	0.1	0.00487	0.0133	0.00270	0.00130
	0.2	0.00268	0.00231	-0.00231	0.00223
	0.4	-0.00628	0.00412	-0.00367	0.00436
	0.6	-0.00765	0.00647	-0.00340	0.00735
$\alpha = 0.9, \beta = 2, n = 25$	0.1	0.00026	0.00012	0.000051	0.000113
	0.2	-0.00017	0.00034	-0.00172	0.000363
	0.4	-0.000601	0.00112	-0.00371	0.001102
	0.6	-0.00700	0.00240	-0.00406	0.002585
$\alpha = 1.5, \beta = 2, n = 50$	0.1	0.00473	0.000782	-0.000771	0.000744
	0.2	0.00283	0.001191	-0.002069	0.00119
	0.4	-0.00070	0.002114	-0.000270	0.00227
	0.6	-0.00056	0.003517	-0.000260	0.00377
$\alpha = 0.9, \beta = 2, n = 50$	0.1	0.00114	0.000064	0.000055	0.000066
	0.2	0.00093	0.000184	-0.001362	0.000192
	0.4	-0.00086	0.000567	-0.001156	0.00058
	0.6	-0.00101	0.00119	-0.002002	0.00129

Table 3. Coverage probabilities and the mean lengths of the CIs.

Model	0.5-QPL	n	q	p (censorship)					
				0.05		0.20		0.30	
				CP	ML	CP	ML	CP	ML
Weibull: $\alpha = 10, \beta = 1.2$	0.6916	50	0.1	0.9360	1.1936	0.9678	1.2595	0.9742	1.2855
	1.3321		0.2	0.9960	2.2770	0.9982	2.3648	0.9990	2.3928
	2.8482	100	0.4	0.9804	2.8177	0.9824	2.9058	0.9796	2.9457
	0.6916		0.1	0.9948	1.2768	0.9976	1.3126	0.9998	1.3272
	1.3321		0.2	0.9964	1.7928	0.9984	1.8027	0.9978	1.8104
	2.8482		0.4	0.9820	1.9813	0.9804	2.0212	0.9760	2.0642
Gamma: $\alpha = 0.9, \beta = 0.1$	0.4223	50	0.1	0.9324	0.6470	0.9510	0.6585	0.9538	0.6651
	0.9819		0.2	0.9946	1.4623	0.9972	1.4934	0.9986	1.5016
	2.5465	100	0.4	0.9802	2.2176	0.9822	2.2400	0.9792	2.2784
	0.4223		0.1	0.9954	0.6751	0.9982	0.6813	0.9972	0.6861
	0.9819		0.2	0.9980	1.1499	0.9982	1.1579	0.9982	1.1637
	2.5465		0.4	0.9808	1.5640	0.9776	1.5899	0.9802	1.6183

5. Applications

5.1 Mayo Clinic study on primary biliary cirrhosis data

The dataset of the Mayo Clinic study on primary biliary cirrhosis (PBC) presented by Fleming and Harrington [19] was used. The dataset is also available in the ‘survival’ library of the famous statistical programming language R and has been analyzed by many authors. I focused on time, i.e. the number of days between registration and the earlier date of death, liver transplantation or analysis of the study in July 1986. More than 60% of the items were censored, resulting in a minimum KM survival function of approximately 0.3534. The KM survival function is plotted in Figure 1 (left). Each censored item was marked with a positive sign in the survival curve.

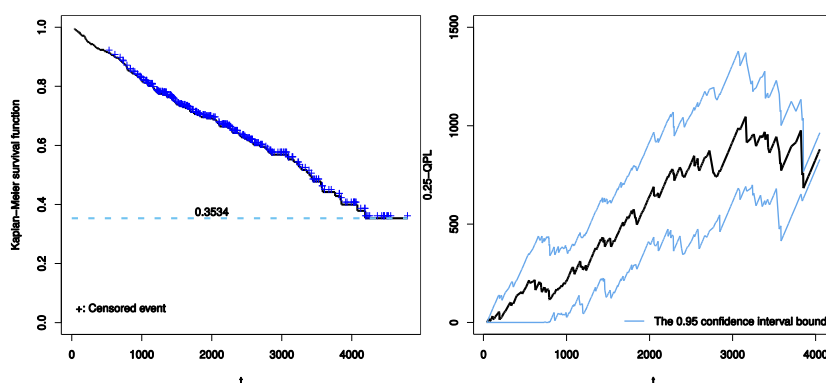


Figure 1. Left: The KM survival function of the PBC dataset. Right: The 0.25-QPL function of this dataset along with its 95% CIs.

Three functions 0.25-QPL, 0.5-QPL, and 0.75-QPL, namely the first, second, and third quartiles of the past lifetime, were estimated within the range of the data. The 0.5-QPL function is also referred to the median past lifetime function.

Figures 1 (right) and 2 plot these functions and show that they have increasing form. For the median past lifetime and specially, at two selected times 1077 and 4079, I have $q_{0.5}(1077) = 480$ and $q_{0.5}(4079) = 2068$. This means that I expect half of the patients that experienced the event before 1077 days, experienced it after (before) 1077-480=597 days. Similarly, I expect that half of the patients experienced the event before 4079 days, experienced it after (before) 2011 days. In addition, the 95% CIs defined in Theorem 3 are calculated and presented in graphs.

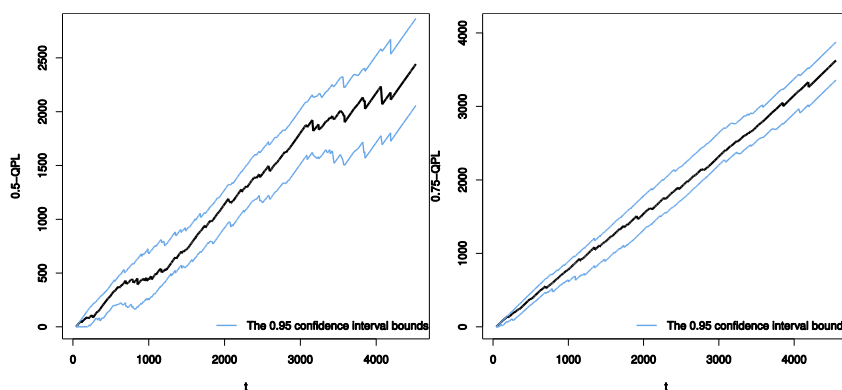


Figure 2. The 0.5-QPL and 0.75-QPL functions of the PBC dataset along with their 95% CIs.

5.2 North central cancer treatment group lung cancer data

This data set contains the survival times of 288 advanced lung cancer patients from the North Central Cancer Treatment Group (NCCTG) and is available in the ‘survival’ library of the statistical programming language R. The column ‘time’ of this dataset gives the time to death or censor of the patients and ‘status’ shows the censoring indicator. About 27% of the death times are censored. Figure 3 shows the KM survival function. Also, the first to third quartile functions 0.25-QPL, 0.5-QPL, and 0.75-QPL are plotted in Figure 3 (right) and Figure 4 and exhibit increasing form.

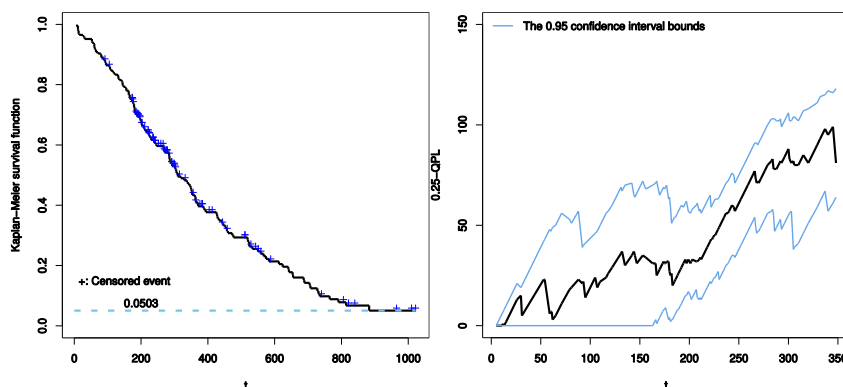


Figure 3. The KM survival function for NCCTG lung cancer data and the 0.25-QPL function of this data along with its 95% CIs.

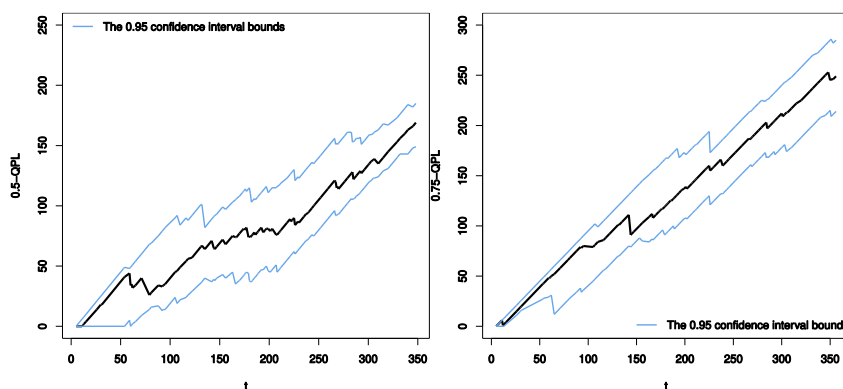


Figure 4. The estimated 0.50-QPL and 0.75-QPL functions for NCCTG lung cancer data along with their 95% CIs.

6 Conclusions

The α -QPL function is a useful rival of the mean inactivity time function in reliability theory and survival analysis. There are situations where the α -QPL function is preferable to the mean inactivity time function, e.g., when the data are heavily censored or the underlying distribution is skewed or have heavy right tail. The Kaplan-Meier product limit estimator of the survival function was applied to define an estimator for the α -QPL function. The proposed estimator converges weakly to a Gaussian

process, and this result is used to construct a confidence interval that depends on the density estimator. Another confidence interval is proposed that does not depend on the density estimator. Simulation results show that the proposed estimator of the α -QPL function is consistent. The coverage probability and mean length of the confidence intervals were also investigated. The applicability of the estimator and the proposed confidence interval is demonstrated using a real dataset. Further properties and applications of the α -QPL function can be considered in the future of this research. In particular, extending the concepts of the α -QPL function and its estimation problem in a multivariate context are interesting and remain as open problems.

Use of AI tools declaration

The author declares he has not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

Researchers Supporting Project number (RSP2024R392), King Saud University, Riyadh, Saudi Arabia.

The author thanks the two anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions. This work is supported by Researchers Supporting Project number (RSP2024R392), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

There are no conflicts of interest.

References

1. N. Unnikrishnan, B. Vineshkumar, Reversed percentile residual life and related concepts, *J. Korean Stat. Soc.*, **40** (2011), 85–92. <https://doi.org/10.1016/j.jkss.2010.06.001>
2. M. Shafaei Noughabi, S. Izadkhah, On the quantile past lifetime of the components of the parallel systems, *Commun. Stat. Theory Meth.*, **45** (2016), 2130–2136. <https://doi.org/10.1080/03610926.2013.875573>
3. M. Shafaei Noughabi, Solving a functional equation and characterizing distributions by quantile past lifetime functions, *Econ. Qual. Control*, **31** (2016), 55–58. <https://doi.org/10.1515/eqc-2015-0017>
4. M. Mahdy, M. Further results involving percentile inactivity time order and its inference, *Metron*, **72** (2014), 269–282. <https://doi.org/10.1007/s40300-013-0017-9>
5. L. Balmert, J. H. Jeong, Nonparametric inference on quantile lost lifespan, *Biometrics*, **73** (2017), 252–259. <https://doi.org/10.1111/biom.12555>
6. L. C. Balmert, R. Li, L. Peng, J. H. Jeong, Quantile regression on inactivity time, *Stat. Meth. Med. Res.*, **30** (2021), 1332–1346. <https://doi.org/10.1177/0962280221995977>
7. A. M. Teamah Abd-Elmonem, A. Elbanna, A. M. Gemeay, Frechet-Weibull mixture distribution: properties and applications, *Appl. Math. Sci.*, **14** (2020), 75–86. <https://doi.org/10.12988/ams.2020.912165>

8. I. Elbatal, S. M. Alghamdi, F. Jamal, S. Khan, E. M. Almetwally, M. Elgarhy, Kavya-manoharan weibull-g family of distributions: statistical inference under progressive type-ii censoring scheme, *Adv. Appl. Stat.*, **87** (2023), 191–223. <https://doi.org/10.17654/0972361723034>
9. C. F. Chung, Confidence bands for percentile residual lifetime under random censorship model, *J. Multivariate Anal.*, **29** (1989), 94–126. [https://doi.org/10.1016/0047-259X\(89\)90079-1](https://doi.org/10.1016/0047-259X(89)90079-1)
10. M. D. Burke, S. Csorgo, L. Horvath, Strong approximations of some biometric estimates under random censorship, *Z. Wahrsch. Verw. Gebiete*, **56** (1981), 87–112.
11. S. Csorgo, L. Horvath, The rate of strong uniform consistency for the product-limit estimator, *Z. Wahrsch. Verw. Gebiete*, **62** (1983), 411–426.
12. E. E. Aly, M. Csorgo, L. Horvath, Strong approximations of the quantile process of the product-limit estimator, *J. Multivariate Anal.*, **16** (1985), 185–210. [https://doi.org/10.1016/0047-259X\(85\)90033-8](https://doi.org/10.1016/0047-259X(85)90033-8)
13. M. Csorgo, S. Scorgo, Estimation of percentile residual life, *Oper. Res.*, **35** (1987), 598–606. <https://doi.org/10.1287/opre.35.4.598>
14. J. R. Blum, V. Susarla, Maximal deviation theory of density and failure rate function estimates based on censored data, *Multivariate Anal.*, **5** (1980), 213–222.
15. M. D. Burke, L. Horvath, Density and failure rate estimation in competing risks model, *Sankhya A.*, **46** (1982), 135–154. <https://doi.org/10.1016/j.res.2010.04.006>
16. J. Mielniczuk, Some asymptotic properties of kernel estimators of a density function in case of censored data, *Ann. Statist.*, **14** (1986), 766–773. <https://doi.org/10.1214/aos/1176349954>
17. J. S. Marron, W. J. Padgett, Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples, *Ann. Statist.*, **15** (1987), 1520–1535. <https://doi.org/10.1214/aos/1176350607>
18. S. H. Lo, Y. P. Mack, J. L. Wang, Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator, *Probab. Theory Related Fields*, **80** (1989), 461–473.
19. T. R. Fleming, D. P. Harrington, Counting processes and survival analysis, New York: Wiley, 1991. <https://doi.org/10.1002/9781118150672>



AIMS Press

© 2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)