

Bootstrap Based Confidence Interval Estimation of Quantiles for Current Status Data

Wei Chen¹ & Fengling Ren²

¹ School of Zhangjiagang, Jiangsu University of Science and Technology, Zhangjiagang 215600, P.R. China

² School of Computer Science and Engineering, Xinjiang University of Finance and Economics, Urumqi 830012, P.R. China

Correspondence: Wei Chen, School of Zhangjiagang, Jiangsu University of Science and Technology, Zhangjiagang 215600, P.R. China

Received: May 25, 2021 Accepted: August 12, 2021 Online Published: August 20, 2021

doi:10.5539/ijsp.v10n5p38 URL: <https://doi.org/10.5539/ijsp.v10n5p38>

Abstract

In this paper, we proposed a bootstrap approach to construct the confidence interval of quantiles for current status data, which is computationally simple and efficient without estimating nuisance parameters. The reasonability of the proposed method is verified by the well performance presented in the extensive simulation study. We also analyzed a real data set as illustration.

Keywords: bootstrap, confidence interval, current status data, estimation of quantiles, nonparametric maximum likelihood estimator

1. Introduction

Current status data, also called the "case 1" interval censored data, arise extensively in epidemiological studies, clinical trials, and other areas, where the time of occurrence of some event is of interest, but one only know whether the event has occurred or not at the examination time. For example, Keiding et al. (1996) provided such a data set, which is also analyzed to illustrate our approach. The data set concerns 230 Austrian males, in which it is interested to acquire the time to infection by Rubella and each subject was tested once during the period 1-25 March 1988 for immunization against Rubella. Apparently, the exact time to infection is impossible to be known, and the only information available for each subject is whether the immunization is achieved or not at the examination time.

In the past three decades, there have been considerable researches on the analysis of current status data. Among of them, estimating the distribution of failure time and some other related topics are still attractive and many meaningful results have been obtained. Let X be the failure time having the distribution function F , and T the examination time having the distribution function G . Then the observed data consists of n independent and identically distributed copies of (T, Δ) , where $\Delta = I(X \leq T)$ and $I(\cdot)$ denotes the indicator function, namely, (T_i, Δ_i) , $i = 1, \dots, n$. When F and G are continuously differentiable at $t_0 \in (0, +\infty)$ with derivatives $f(t_0) > 0$ and $g(t_0) > 0$, Groeneboom and Wellner (1992) showed that for the nonparametric maximum likelihood estimator (NPMLE) F_n of F ,

$$n^{1/3}(F_n(t_0) - F(t_0)) \rightarrow \left(\frac{4f(t_0)F(t_0)(1 - F(t_0))}{g(t_0)} \right)^{1/3} \mathbb{C} \equiv \kappa \mathbb{C}, \quad (1)$$

in distribution, where $\mathbb{C} = \operatorname{argmin}_{t \in \mathbb{R}} W(t) + t^2$, and W is a standard two-sided Brownian motion process with $W(0) = 0$. Then it seems to be straightforward to construct the confidence interval for $F(t_0)$. But to do this, one need to estimate the nuisance parameter κ in the limiting distribution. As noted by Groeneboom and Wellner (2005), the difficulty in computation of κ is enormous, and the resulting performance is rather unstable. To avoid this trouble, they suggested the likelihood ratio approach and conducted a small simulation study.

Recently, several researchers renewed to discuss the problem of constructing the confidence interval of $F(t_0)$. Groeneboom et al. (2010) proposed the maximum smoothed likelihood estimation and compared it with the smoothed maximum likelihood estimation. See also Groeneboom (2012). Groeneboom and Jongbloed (2015) investigated the nonparametric confidence intervals for the general monotone functions. However, when applying these methods above, the resulting confidence intervals usually vary drastically and the tuning parameters therein had to be chosen cutely. In particular, Sen and Xu (2015) developed a model based bootstrap method for mixed case interval censored data, which is consistent and can be seen as a "bootstrap residual" approach combined with the smooth estimator of F , and their simulation study showed their procedure has superior performance. Rather, Groeneboom and Hendrickx (2017) proposed a nonparametric

bootstrap method, which can be seen as a "bootstrap pairs" approach and generates the data by sampling directly from the empirical distribution of the observations.

Although estimating the survival probability is very useful, quantile survival time has more flexible interpretation, and plays an important role in various statistical applications, especially in data modeling, reliability and medical studies. The existing researches about the quantile for survival data mainly focus on right censored data. However, the same problem for current status data just attracts little attention. When without covariates, Banerjee and Wellner (2001) briefly discussed the problem of constructing the confidence interval by using the likelihood-ratio test. When covariates exist, Ou et al. (2016) proposed a novel estimating method for linear quantile regression model, see also the reference therein. In this paper, we are interested to construct the confident interval of quantiles for current status data without covariates. Note that confidence region of quantile in Banerjee and Wellner (2001) is obtained by inverting the acceptance region of the likelihood ratio test, it is inevitable to tackle the same difficulty in the computation of confidence interval of $F(t_0)$, as argued by (Sen and Xu, 2015). In the light of the merits of bootstrap method and its superior performance having been achieved in constructing the the confidence interval of $F(t_0)$, we develop a current status model based bootstrap method to establish the confidence intervals of quantiles of F . It is worth noting that our bootstrap method is directly based on the NPMLE F_n , not relied on the smooth version of F_n , although the former is indeed adopted in the method of (Sen and Xu, 2015).

This paper is organised as follows. The current status model based bootstrap method is presented in Section 2, where a consistent estimator of standard error of NPMLE of a quantile is proposed. In Section 3, simulation studies are carried out to evaluate our proposal under different scenarios. In Section 4, we analyzed a real data set as illustration. Finally Section 5 provides some concluding remarks.

2. Methods

For the observed data (T_i, Δ_i) , $i = 1, \dots, n$, the NPMLE F_n of F can be obtained by maximizing the log-likelihood function

$$\sum_{i=1}^n \{\Delta_i \log(F(T_i)) + (1 - \Delta_i) \log(1 - F(T_i))\}$$

over all distributions. Then for a given level $\tau \in (0, 1)$, we can estimate the τ th quantile $F^{-1}(\tau)$, which is defined by $F^{-1}(\tau) = \inf\{s : F(s) \geq \tau\}$, by its empirical version $F_n^{-1}(\tau)$, namely, $F_n^{-1}(\tau) = \inf\{s : F_n(s) \geq \tau\}$. As F_n is a step function, $F_n^{-1}(\tau)$ is naturally also stepwise with respect to τ . Banerjee and Wellner (2001) have derived the asymptotic distribution of $F_n^{-1}(\tau)$ ($0 < \tau < 1$) as follows,

$$n^{1/3}(F_n^{-1}(\tau) - F^{-1}(\tau)) \rightarrow \left(\frac{4\tau(1-\tau)}{f^2(t)g(t)}\right)^{1/3} \mathbb{C} \equiv \gamma\mathbb{C} \tag{2}$$

in distribution, where $t = F^{-1}(\tau)$. Following the limiting distribution, a natural asymptotic confidence interval of $F^{-1}(\tau)$ with nominal level $1 - \alpha$, such as $\alpha = 0.05$, is

$$F_n^{-1}(\tau) \pm n^{-1/3} \gamma Q_{1-\alpha/2}(\mathbb{C}), \tag{3}$$

where $Q_{1-\alpha/2}(\mathbb{C})$ is the $(1 - \alpha/2)100\%$ th percentile of \mathbb{C} (Groeneboom and Wellner, 2001). As γ herein involves with both the density functions of F and G at t , one has to resort to some smooth technique like the kernel method to yield a consistent estimate of γ . It is well known that to ensure the kernel method works well, the sample size usually should be moderate at least. And the cumbersome and heavy computation cost should be paid. Therefore, a computationally simple and stable procedure for solving the problem would be desirable.

In many situations, when the limiting distribution of an estimator has been established and the variance of the estimator is very complicated, one usually seeks to find a consistent estimate of the variance as a counterpart. To do this, one popular strategy is to employ the resampling approach. Among of them, the bootstrap method is typically adopted. However, under the current status model, using the naive nonparametric bootstrap method to generate the limit distribution $\kappa\mathbb{C}$ has been shown to be inconsistent (Sen and Xu, 2015). As a result, other consistent bootstrap methods should be developed, more details see (Cattaneo et al.,2020) and the references therein.

In this paper, we do not follow them to approximate the limiting distribution, but develop a new bootstrap method aiming to find out a consistent estimation of the variance of the estimator of a quantile, not a consistent estimate of the limit distribution function. Since the variance of $F_n^{-1}(\tau)$ is approximately equal to $n^{-2/3}$ time by the variance of $\gamma\mathbb{C}$, and suppose that one can estimate the variance of $F_n^{-1}(\tau)$ using a bootstrap method, then by solving this equation, a consistent estimation of γ yields, denoted by γ_n . Through the simulation study in the next section, we find that the bootstrap method

outlined below performs well and indeed can provide a desirable estimator of variance of $F_n^{-1}(\tau)$. Furthermore, a bootstrap method to construct a asymptotic confidence interval of $F^{-1}(\tau)$ with nominal level $1 - \alpha$ is introduced as follows.

Step 1: For the original observations (T_i, Δ_i) , $i = 1, \dots, n$ and fixed τ , one compute the NPML F_n of F as mentioned above. Then $F_n^{-1}(\tau)$ is computed.

Step 2: For each T_i and $b = 1, \dots, B$, we generate the b-th bootstrap one of the corresponding indicator Δ_i through a bernoulli variable with success probability $F_n(T_i)$, denoted by Δ_i^b . Based on the b-th bootstrap sample (T_i, Δ_i^b) , $i = 1, \dots, n$, compute the corresponding NPML F_{nb} , then a bootstrap quantile $F_{nb}^{-1}(\tau)$ is obtained. Repeat the process above B times.

Step 3: Utilizing the bootstrap quantiles $F_{nb}^{-1}(\tau)$, $b = 1, \dots, B$, compute the variance of the limiting distribution through the sample variance of $F_{nb}^{-1}(\tau)$. Thus a equation about γ_n can be established by

$$\frac{1}{B-1} \sum_{b=1}^B \left\{ F_{nb}^{-1}(\tau) - \frac{1}{B} \sum_{b=1}^B F_{nb}^{-1}(\tau) \right\}^2 = \gamma_n^2 D, \tag{4}$$

where D is the variance of \mathbb{C} (Groeneboom and Wellner, 2001). The solution can be easily obtained.

Step 4: Finally, one can construct the asymptotic confidence interval of $F^{-1}(\tau)$ with level $1 - \alpha$ by

$$F_n^{-1}(\tau) \pm n^{-1/3} \gamma_n Q_{1-\alpha/2}(\mathbb{C}). \tag{5}$$

Remark 1: Similar to the techniques used in Groeneboom and Hendrickx (2017), the consistence of the estimator in the left side of the equality in (4) can be shown.

Remark 2: In the typical bootstrap method, the left side of the equality in (4) has the form

$$A = \frac{1}{B} \sum_{b=1}^B \{F_{nb}^{-1}(\tau) - F_n^{-1}(\tau)\}^2,$$

which is centering the estimate $F_n^{-1}(\tau)$. As explained above, due to the inconsistency of the nonparametric bootstrap method (Sen and Xu, 2015), this expression should be adjusted. If one use A to estimate the unknown variance, our simulation study display that it exaggerate heavily the true error, and the confidence interval is longer and has more higher coverage probability than the nominal level except the level 0.99.

Remark 3: The computation of the NPML F_n of F can be implemented using the R packages "Icens" or "curstatCI", and so on.

3. Numerical Studies

To evaluate the finite performance of the proposed method, several simulation studies were conducted. Here we mainly considered three simulation settings. Explicitly, the failure time X was generated from the distribution $\exp(1)$, $|N(0, 1)|$, and $\exp(1)$ but truncated on the interval $[0,2]$, respectively. Corresponding to it, the examination time T was generated independently from the distribution $\exp(1)$, $\text{Unif}(0,2)$, and $\text{Unif}(0,2)$. These simulation settings are used in Groeneboom and Wellner (2005), Groeneboom et al. (2010), Groeneboom (2012), Groeneboom and Hendrickx (2017), and Sen and Xu (2015). The sample sizes $n = 50, 75, 100, 200, 400$ and 800 were chosen. For each value of n , we generated 10000 samples independently from the preceding distributions of the pair (X, T) . For each data set, we used the method described in Section 2 to bootstrap the Δ_i from the Bernoulli distribution with the success probability $F_n(T_i)$, fixing the values of T_i , by taking 1000 bootstrap samples and determining the $(1 - \alpha/2)100$ th percentile of the limiting distribution. All simulations are implemented using the software R. For each given setting and the NPML of the τ quantile, we record its empirical sample standard deviation (SD), standard error estimation (SEE), the observed average coverage probability (CP) and length (LEN) of confidence intervals with levels 80%, 90%, 95%, 99%, respectively, where the valves of τ vary from 0.1 to 0.9.

All simulation results are summarized in Tables 1-3. It can be found that the SD is close to the SSE. As the increase of the sample size, for any given τ th quantile, the average coverage probability of its confidence intervals with a fixed nominal level is more and more near to the true level, and the corresponding average length becomes shorter. It is noted that when $\tau \leq 0.2$ or $\tau \geq 0.9$, the resulting behavior is not satisfactory, and the coverage probability is slightly below the nominal level, even though the sample size reaches to 800. This maybe caused by that few data can be obtained below or above the true quantile. The bad behavior at the tails also occurs in other problems. When τ lies between 0.3 and 0.8, the proposed method indeed behaves well, especially when the sample size is greater than 100. Besides, we also found a interest result. If we use

$$\frac{1}{B} \sum_{b=1}^B \{F_{nb}^{-1}(\tau_0) - F_n^{-1}(\tau_0)\}^2$$

Table 1. Simulation results of confidence interval of quantiles under different levels, where $X \sim \exp(1)$ and $T \sim \exp(1)$

	SD	SEE	CP				LEN			
			0.8	0.9	0.95	0.99	0.8	0.9	0.95	0.99
<i>n=50</i>										
$\tau = 0.1$	0.131	0.113	0.596	0.671	0.727	0.805	0.292	0.372	0.439	0.566
$\tau = 0.2$	0.165	0.151	0.716	0.806	0.859	0.925	0.392	0.499	0.589	0.759
$\tau = 0.3$	0.209	0.199	0.753	0.849	0.906	0.964	0.514	0.654	0.772	0.996
$\tau = 0.4$	0.245	0.237	0.778	0.881	0.933	0.976	0.614	0.781	0.923	1.189
$\tau = 0.5$	0.286	0.278	0.777	0.880	0.933	0.977	0.719	0.915	1.081	1.393
$\tau = 0.6$	0.390	0.342	0.749	0.860	0.920	0.974	0.885	1.126	1.330	1.714
$\tau = 0.7$	0.487	0.393	0.705	0.822	0.887	0.949	1.018	1.295	1.530	1.972
$\tau = 0.8$	0.608	0.423	0.609	0.726	0.793	0.871	1.094	1.392	1.644	2.120
$\tau = 0.9$	0.764	0.421	0.435	0.517	0.570	0.652	1.090	1.387	1.638	2.112
<i>n=75</i>										
$\tau = 0.1$	0.108	0.094	0.627	0.701	0.756	0.837	0.244	0.310	0.366	0.472
$\tau = 0.2$	0.138	0.130	0.740	0.831	0.884	0.946	0.337	0.429	0.507	0.653
$\tau = 0.3$	0.178	0.169	0.761	0.861	0.918	0.971	0.438	0.557	0.657	0.847
$\tau = 0.4$	0.209	0.203	0.786	0.889	0.938	0.980	0.526	0.669	0.790	1.018
$\tau = 0.5$	0.248	0.240	0.779	0.884	0.937	0.981	0.623	0.792	0.936	1.206
$\tau = 0.6$	0.313	0.294	0.771	0.878	0.931	0.977	0.762	0.970	1.146	1.477
$\tau = 0.7$	0.415	0.355	0.733	0.843	0.907	0.964	0.920	1.170	1.382	1.782
$\tau = 0.8$	0.525	0.398	0.635	0.756	0.837	0.912	1.031	1.311	1.549	1.997
$\tau = 0.9$	0.706	0.403	0.460	0.557	0.613	0.698	1.044	1.328	1.569	2.022
<i>n=100</i>										
$\tau = 0.1$	0.095	0.085	0.660	0.738	0.789	0.866	0.220	0.280	0.331	0.426
$\tau = 0.2$	0.125	0.118	0.754	0.847	0.904	0.961	0.305	0.388	0.459	0.591
$\tau = 0.3$	0.158	0.152	0.769	0.868	0.926	0.977	0.394	0.501	0.591	0.762
$\tau = 0.4$	0.189	0.184	0.788	0.890	0.943	0.985	0.477	0.607	0.717	0.925
$\tau = 0.5$	0.222	0.217	0.780	0.890	0.942	0.983	0.562	0.715	0.844	1.088
$\tau = 0.6$	0.284	0.268	0.774	0.880	0.932	0.978	0.695	0.884	1.044	1.346
$\tau = 0.7$	0.361	0.329	0.755	0.864	0.921	0.972	0.851	1.083	1.279	1.649
$\tau = 0.8$	0.479	0.377	0.662	0.782	0.851	0.928	0.977	1.243	1.468	1.893
$\tau = 0.9$	0.669	0.399	0.495	0.598	0.661	0.745	1.033	1.314	1.552	2.000
<i>n=200</i>										
$\tau = 0.1$	0.074	0.066	0.698	0.785	0.842	0.916	0.172	0.219	0.258	0.333
$\tau = 0.2$	0.096	0.093	0.771	0.871	0.924	0.978	0.242	0.308	0.363	0.468
$\tau = 0.3$	0.124	0.119	0.776	0.881	0.936	0.980	0.310	0.395	0.466	0.601
$\tau = 0.4$	0.148	0.144	0.792	0.896	0.948	0.985	0.375	0.477	0.563	0.726
$\tau = 0.5$	0.175	0.172	0.794	0.895	0.946	0.984	0.446	0.568	0.671	0.865
$\tau = 0.6$	0.218	0.212	0.785	0.890	0.941	0.983	0.550	0.699	0.826	1.065
$\tau = 0.7$	0.282	0.267	0.774	0.882	0.936	0.980	0.692	0.880	1.040	1.340
$\tau = 0.8$	0.380	0.328	0.713	0.827	0.894	0.957	0.851	1.082	1.278	1.648
$\tau = 0.9$	0.573	0.386	0.562	0.681	0.758	0.851	1.001	1.274	1.505	1.940
<i>n=400</i>										
$\tau = 0.1$	0.057	0.053	0.731	0.824	0.882	0.946	0.137	0.175	0.207	0.266
$\tau = 0.2$	0.076	0.074	0.779	0.881	0.935	0.980	0.192	0.244	0.288	0.372
$\tau = 0.3$	0.096	0.094	0.782	0.887	0.940	0.981	0.243	0.309	0.366	0.471
$\tau = 0.4$	0.115	0.114	0.794	0.900	0.950	0.990	0.296	0.376	0.444	0.573
$\tau = 0.5$	0.140	0.137	0.782	0.890	0.943	0.984	0.355	0.452	0.534	0.688
$\tau = 0.6$	0.171	0.169	0.791	0.896	0.949	0.985	0.438	0.558	0.659	0.849
$\tau = 0.7$	0.223	0.214	0.782	0.886	0.942	0.984	0.555	0.706	0.834	1.075
$\tau = 0.8$	0.298	0.274	0.743	0.861	0.915	0.969	0.711	0.905	1.069	1.377
$\tau = 0.9$	0.468	0.361	0.628	0.752	0.830	0.910	0.936	1.191	1.407	1.813
<i>n=800</i>										
$\tau = 0.1$	0.044	0.042	0.755	0.853	0.907	0.965	0.110	0.139	0.165	0.212
$\tau = 0.2$	0.059	0.059	0.791	0.896	0.946	0.985	0.153	0.195	0.230	0.296
$\tau = 0.3$	0.075	0.074	0.793	0.897	0.948	0.987	0.193	0.246	0.291	0.375
$\tau = 0.4$	0.091	0.090	0.791	0.900	0.951	0.988	0.233	0.297	0.351	0.452
$\tau = 0.5$	0.109	0.108	0.791	0.900	0.949	0.987	0.282	0.358	0.423	0.546
$\tau = 0.6$	0.137	0.134	0.789	0.894	0.947	0.985	0.347	0.442	0.522	0.672
$\tau = 0.7$	0.176	0.171	0.786	0.895	0.944	0.984	0.444	0.565	0.667	0.860
$\tau = 0.8$	0.236	0.225	0.771	0.875	0.929	0.977	0.582	0.741	0.875	1.128
$\tau = 0.9$	0.375	0.324	0.694	0.811	0.880	0.949	0.840	1.068	1.262	1.626

SD: sample standard deviation; SEE: standard error estimate; CP (LEN): empirical coverage probability (length) of confidence interval with level from 0.8 to 0.99; n: sample size.

Table 2. Simulation results of confidence interval of quantiles under different levels, where $X \sim \exp(1)$ but truncated on the interval $[0,2]$ and $T \sim Unif(0, 2)$

	SD	SEE	CP				LEN			
			0.8	0.9	0.95	0.99	0.8	0.9	0.95	0.99
<i>n=50</i>										
$\tau = 0.1$	0.135	0.101	0.457	0.533	0.583	0.666	0.260	0.331	0.391	0.504
$\tau = 0.2$	0.156	0.126	0.613	0.690	0.743	0.825	0.326	0.414	0.489	0.631
$\tau = 0.3$	0.181	0.154	0.660	0.759	0.822	0.903	0.397	0.506	0.597	0.770
$\tau = 0.4$	0.206	0.177	0.704	0.815	0.878	0.944	0.458	0.582	0.688	0.886
$\tau = 0.5$	0.224	0.194	0.714	0.831	0.895	0.958	0.503	0.640	0.755	0.974
$\tau = 0.6$	0.248	0.215	0.720	0.836	0.904	0.964	0.556	0.707	0.836	1.077
$\tau = 0.7$	0.267	0.229	0.709	0.832	0.899	0.962	0.593	0.754	0.891	1.148
$\tau = 0.8$	0.284	0.232	0.663	0.790	0.858	0.931	0.600	0.763	0.902	1.162
$\tau = 0.9$	0.284	0.209	0.593	0.684	0.743	0.826	0.541	0.688	0.812	1.047
<i>n=75</i>										
$\tau = 0.1$	0.112	0.088	0.520	0.593	0.647	0.732	0.230	0.293	0.346	0.445
$\tau = 0.2$	0.134	0.113	0.671	0.749	0.806	0.880	0.294	0.375	0.443	0.570
$\tau = 0.3$	0.159	0.139	0.699	0.804	0.864	0.936	0.362	0.460	0.543	0.700
$\tau = 0.4$	0.179	0.159	0.728	0.844	0.907	0.964	0.414	0.526	0.622	0.801
$\tau = 0.5$	0.198	0.178	0.739	0.859	0.917	0.972	0.462	0.587	0.694	0.894
$\tau = 0.6$	0.218	0.197	0.745	0.867	0.927	0.977	0.511	0.650	0.768	0.990
$\tau = 0.7$	0.236	0.212	0.740	0.854	0.919	0.972	0.549	0.698	0.825	1.063
$\tau = 0.8$	0.250	0.216	0.699	0.823	0.888	0.952	0.561	0.713	0.842	1.086
$\tau = 0.9$	0.251	0.197	0.619	0.719	0.786	0.869	0.511	0.651	0.768	0.991
<i>n=100</i>										
$\tau = 0.1$	0.100	0.080	0.545	0.621	0.677	0.761	0.208	0.265	0.313	0.404
$\tau = 0.2$	0.123	0.105	0.678	0.766	0.823	0.896	0.272	0.346	0.409	0.527
$\tau = 0.3$	0.148	0.130	0.699	0.806	0.871	0.944	0.338	0.430	0.508	0.655
$\tau = 0.4$	0.162	0.148	0.748	0.856	0.916	0.971	0.385	0.490	0.579	0.746
$\tau = 0.5$	0.181	0.166	0.749	0.863	0.927	0.979	0.431	0.548	0.648	0.835
$\tau = 0.6$	0.198	0.183	0.759	0.874	0.933	0.979	0.476	0.605	0.715	0.921
$\tau = 0.7$	0.215	0.198	0.753	0.864	0.927	0.977	0.513	0.652	0.770	0.993
$\tau = 0.8$	0.229	0.203	0.716	0.836	0.899	0.958	0.528	0.671	0.793	1.022
$\tau = 0.9$	0.230	0.188	0.649	0.756	0.822	0.902	0.488	0.621	0.734	0.946
<i>n=200</i>										
$\tau = 0.1$	0.077	0.064	0.618	0.692	0.749	0.831	0.168	0.214	0.253	0.325
$\tau = 0.2$	0.098	0.087	0.716	0.816	0.875	0.942	0.226	0.287	0.339	0.438
$\tau = 0.3$	0.114	0.107	0.743	0.849	0.909	0.968	0.278	0.353	0.417	0.538
$\tau = 0.4$	0.131	0.123	0.767	0.881	0.934	0.981	0.320	0.407	0.481	0.620
$\tau = 0.5$	0.143	0.137	0.775	0.885	0.941	0.983	0.356	0.453	0.535	0.690
$\tau = 0.6$	0.158	0.151	0.776	0.890	0.943	0.983	0.393	0.499	0.590	0.760
$\tau = 0.7$	0.172	0.164	0.773	0.882	0.937	0.982	0.425	0.541	0.639	0.824
$\tau = 0.8$	0.184	0.172	0.745	0.864	0.920	0.975	0.446	0.568	0.671	0.865
$\tau = 0.9$	0.186	0.165	0.701	0.813	0.877	0.947	0.427	0.544	0.642	0.828
<i>n=400</i>										
$\tau = 0.1$	0.060	0.053	0.676	0.767	0.828	0.903	0.138	0.176	0.208	0.268
$\tau = 0.2$	0.078	0.072	0.746	0.855	0.914	0.970	0.188	0.239	0.283	0.364
$\tau = 0.3$	0.091	0.087	0.761	0.872	0.928	0.978	0.226	0.288	0.340	0.438
$\tau = 0.4$	0.104	0.100	0.774	0.883	0.942	0.984	0.260	0.331	0.391	0.504
$\tau = 0.5$	0.114	0.111	0.780	0.888	0.941	0.985	0.289	0.367	0.434	0.559
$\tau = 0.6$	0.127	0.123	0.779	0.889	0.942	0.985	0.319	0.406	0.480	0.618
$\tau = 0.7$	0.138	0.133	0.776	0.890	0.942	0.984	0.346	0.440	0.520	0.670
$\tau = 0.8$	0.148	0.141	0.762	0.874	0.936	0.981	0.365	0.465	0.549	0.707
$\tau = 0.9$	0.151	0.140	0.732	0.846	0.905	0.965	0.362	0.461	0.545	0.702
<i>n=800</i>										
$\tau = 0.1$	0.048	0.043	0.712	0.809	0.869	0.940	0.113	0.144	0.170	0.219
$\tau = 0.2$	0.062	0.059	0.767	0.872	0.929	0.975	0.154	0.195	0.231	0.298
$\tau = 0.3$	0.072	0.070	0.777	0.887	0.942	0.983	0.183	0.233	0.276	0.355
$\tau = 0.4$	0.082	0.080	0.783	0.890	0.942	0.986	0.208	0.264	0.312	0.403
$\tau = 0.5$	0.091	0.089	0.785	0.895	0.948	0.985	0.232	0.296	0.349	0.450
$\tau = 0.6$	0.100	0.098	0.788	0.897	0.951	0.986	0.256	0.326	0.385	0.496
$\tau = 0.7$	0.110	0.107	0.781	0.891	0.946	0.986	0.278	0.354	0.418	0.539
$\tau = 0.8$	0.117	0.115	0.782	0.892	0.949	0.986	0.299	0.380	0.449	0.579
$\tau = 0.9$	0.121	0.115	0.757	0.870	0.926	0.975	0.299	0.380	0.449	0.579

SD: sample standard deviation; SEE: standard error estimate; CP (LEN): empirical coverage probability (length) of confidence interval with level from 0.8 to 0.99; n: sample size.

Table 3. Simulation results of confidence interval of quantiles under different levels, where $X \sim |N(0, 1)|$ and $T \sim Unif(0, 2)$

	SD	SEE	CP				LEN			
			0.8	0.9	0.95	0.99	0.8	0.9	0.95	0.99
<i>n=50</i>										
$\tau = 0.1$	0.154	0.114	0.483	0.553	0.605	0.686	0.294	0.374	0.442	0.569
$\tau = 0.2$	0.175	0.140	0.634	0.716	0.769	0.848	0.361	0.459	0.543	0.699
$\tau = 0.3$	0.202	0.168	0.663	0.768	0.836	0.915	0.434	0.552	0.652	0.840
$\tau = 0.4$	0.217	0.187	0.707	0.819	0.884	0.953	0.484	0.616	0.727	0.937
$\tau = 0.5$	0.230	0.204	0.725	0.845	0.909	0.963	0.527	0.671	0.792	1.021
$\tau = 0.6$	0.251	0.223	0.727	0.846	0.910	0.967	0.577	0.734	0.867	1.117
$\tau = 0.7$	0.270	0.237	0.727	0.842	0.907	0.967	0.613	0.779	0.920	1.186
$\tau = 0.8$	0.291	0.239	0.674	0.801	0.871	0.940	0.620	0.788	0.931	1.200
$\tau = 0.9$	0.294	0.219	0.594	0.680	0.740	0.826	0.567	0.721	0.852	1.098
<i>n=75</i>										
$\tau = 0.1$	0.129	0.100	0.534	0.607	0.660	0.744	0.259	0.330	0.390	0.502
$\tau = 0.2$	0.153	0.127	0.662	0.755	0.816	0.891	0.329	0.419	0.495	0.638
$\tau = 0.3$	0.173	0.153	0.703	0.815	0.873	0.943	0.396	0.504	0.596	0.768
$\tau = 0.4$	0.189	0.171	0.737	0.853	0.916	0.972	0.442	0.562	0.664	0.856
$\tau = 0.5$	0.204	0.186	0.749	0.864	0.922	0.973	0.483	0.614	0.725	0.935
$\tau = 0.6$	0.219	0.201	0.754	0.867	0.923	0.976	0.521	0.663	0.784	1.010
$\tau = 0.7$	0.240	0.218	0.745	0.860	0.922	0.975	0.563	0.716	0.846	1.090
$\tau = 0.8$	0.259	0.224	0.706	0.826	0.894	0.956	0.581	0.739	0.873	1.125
$\tau = 0.9$	0.264	0.207	0.617	0.722	0.782	0.866	0.535	0.681	0.805	1.037
<i>n=100</i>										
$\tau = 0.1$	0.117	0.093	0.575	0.650	0.703	0.787	0.242	0.308	0.364	0.469
$\tau = 0.2$	0.140	0.118	0.683	0.779	0.839	0.915	0.308	0.392	0.463	0.596
$\tau = 0.3$	0.162	0.143	0.712	0.823	0.887	0.955	0.371	0.472	0.557	0.719
$\tau = 0.4$	0.173	0.158	0.749	0.861	0.917	0.974	0.410	0.521	0.616	0.794
$\tau = 0.5$	0.185	0.173	0.763	0.876	0.932	0.980	0.450	0.572	0.676	0.871
$\tau = 0.6$	0.201	0.187	0.758	0.880	0.935	0.981	0.485	0.617	0.729	0.940
$\tau = 0.7$	0.217	0.201	0.760	0.870	0.928	0.977	0.522	0.664	0.784	1.011
$\tau = 0.8$	0.239	0.211	0.722	0.833	0.900	0.960	0.547	0.695	0.821	1.059
$\tau = 0.9$	0.243	0.198	0.650	0.759	0.820	0.896	0.513	0.652	0.771	0.993
<i>n=200</i>										
$\tau = 0.1$	0.091	0.076	0.642	0.721	0.779	0.863	0.198	0.252	0.298	0.384
$\tau = 0.2$	0.111	0.100	0.734	0.839	0.897	0.955	0.260	0.331	0.391	0.504
$\tau = 0.3$	0.126	0.118	0.750	0.861	0.917	0.971	0.307	0.391	0.462	0.595
$\tau = 0.4$	0.138	0.131	0.762	0.881	0.936	0.981	0.340	0.432	0.510	0.658
$\tau = 0.5$	0.150	0.141	0.767	0.881	0.936	0.984	0.367	0.467	0.551	0.710
$\tau = 0.6$	0.160	0.152	0.765	0.882	0.940	0.982	0.393	0.501	0.591	0.762
$\tau = 0.7$	0.172	0.164	0.773	0.886	0.939	0.984	0.426	0.543	0.641	0.826
$\tau = 0.8$	0.189	0.178	0.758	0.868	0.928	0.977	0.462	0.588	0.694	0.895
$\tau = 0.9$	0.206	0.176	0.683	0.803	0.875	0.946	0.457	0.581	0.686	0.885
<i>n=400</i>										
$\tau = 0.1$	0.071	0.063	0.690	0.784	0.843	0.916	0.165	0.210	0.248	0.320
$\tau = 0.2$	0.089	0.083	0.750	0.860	0.920	0.972	0.216	0.275	0.325	0.419
$\tau = 0.3$	0.100	0.096	0.767	0.876	0.932	0.978	0.249	0.316	0.374	0.482
$\tau = 0.4$	0.110	0.106	0.780	0.892	0.943	0.985	0.276	0.351	0.414	0.534
$\tau = 0.5$	0.117	0.115	0.779	0.895	0.945	0.986	0.298	0.379	0.448	0.577
$\tau = 0.6$	0.125	0.123	0.783	0.892	0.946	0.985	0.320	0.407	0.481	0.619
$\tau = 0.7$	0.138	0.133	0.783	0.888	0.947	0.985	0.346	0.441	0.520	0.671
$\tau = 0.8$	0.151	0.145	0.774	0.877	0.933	0.979	0.377	0.480	0.567	0.730
$\tau = 0.9$	0.170	0.152	0.718	0.840	0.907	0.968	0.393	0.500	0.591	0.762
<i>n=800</i>										
$\tau = 0.1$	0.057	0.052	0.719	0.827	0.885	0.953	0.137	0.174	0.206	0.265
$\tau = 0.2$	0.070	0.067	0.768	0.875	0.929	0.978	0.176	0.223	0.264	0.340
$\tau = 0.3$	0.079	0.077	0.779	0.882	0.940	0.983	0.201	0.256	0.303	0.390
$\tau = 0.4$	0.087	0.085	0.785	0.895	0.949	0.988	0.220	0.280	0.331	0.427
$\tau = 0.5$	0.093	0.092	0.787	0.897	0.949	0.987	0.239	0.304	0.359	0.462
$\tau = 0.6$	0.102	0.099	0.779	0.893	0.947	0.987	0.257	0.327	0.386	0.498
$\tau = 0.7$	0.109	0.106	0.786	0.893	0.946	0.986	0.276	0.351	0.415	0.534
$\tau = 0.8$	0.120	0.117	0.772	0.889	0.943	0.985	0.303	0.386	0.455	0.587
$\tau = 0.9$	0.141	0.129	0.745	0.861	0.924	0.978	0.336	0.427	0.505	0.651

SD: sample standard deviation; SEE: standard error estimate; CP (LEN): empirical coverage probability (length) of confidence interval with level from 0.8 to 0.99; n: sample size.

to estimate the unknown variance, the resulting SSE tends to heavily exaggerate the true error, and the confidence interval has more higher coverage probability except the level 0.99. This seems to be in line with the fact that the bootstrap method is inconsistent when approximating the limiting distribution, see Theorem 2 in Sen and Xu (2015).

4. Application

In this section, we applied the proposed bootstrap method to analyze the Rubella data set described in the Section 1 or Keiding et al. (1996). The data set contains 230 observations on the prevalence of Rubella in Austrian males. The goal here is to estimate selected quantiles of the distribution of the time to immunization. Several procedures to construct the confidence interval for this data set have been proposed. Explicitly speaking, MLE-based (p) denotes the MLE-based confidence intervals estimated using the Weibull parametric fits for density functions as obtained in Keiding et al. (1996), MLE-based (np) denotes the MLE-based confidence intervals where the density functions involved therein are estimated non-parametrically using the kernel estimates, LR-based denotes the confidence intervals constructed by the likelihood ratio-based method. We compared them with our method for quantiles for τ varying across the sequence 0.4, 0.5, 0.7, 0.9, and the result is exhibited in column 6 of Table 4, where data in the columns 3-5 origins from Table 5 of Banerjee and Wellner (2005). The confidence level is chosen to be 0.95, and the bootstrap time here is 5000. We found that the bootstrap based confidence interval has some similar behaviors with the existing methods and they overlap heavily. It is remarkable that for $\tau < 0.9$, the difference between the results using different methods is slight, but for $\tau = 0.9$, the difference is heavy. Maybe it is led by the sparse of data around it.

Table 4. Confidence intervals of quantiles by Rubella data using different methods

τ	$F_n^{-1}(\tau)$	MLE-based (p)	MLE-based (np)	LR-based	Bootstrap-based
0.4	1.2301	0-3.9648	0-3.2914	0.5288-9.3425	0-5.5632
0.5	5.6411	0.8887-10.3935	3.5516-7.7306	0.5342-12.4877	0-12.8494
0.7	12.4548	4.2224-20.6872	10.4833-14.4263	7.2630-15.8082	7.6097-17.2998
0.9	23.2219	6.1027-40.3411	21.7355-24.7083	12.4274-34.7890	5.1344-41.3093

5. Discussions

In this paper, we suggested a bootstrap based method to construct the confidence interval of quantiles for current status data, which is computationally simple and efficient without estimating nuisance parameters. The reasonability of the proposed method is verified by the well performance presented in the simulation study.

Note that we only use the NPMLE of the distribution function of the failure time in the proposed method. As the convergence rate of the NPMLE is cubic root of n , the behavior may be not well when the sample size is small, such as $n < 200$ in our simulation study. To overcome this problem, one strategy is to replace the NPMLE by its smooth version. The smoothed MLE has a faster convergence rate as discussed in Groeneboom et al. (2010) and Sen and Xu (2015). However, when applying this method, one has to select a proper bandwidth involved therein, which is critical and has drastic influence in making statistical inference. This is interesting and will be investigated in future.

Another interesting extension to the current research is the construction of confidence bands for the quantiles varying with τ . Besides, it can be conjectured that results similar to those presented in the current paper will follow for the case 2 interval censored data. These topics will be studied further but beyond the scope of this paper.

References

Banerjee, M., Wellner, J. A. (2001). Likelihood ratio tests for monotone functions. *The Annals of Statistics*, 29(6), 1699-1731. <https://doi.org/10.1214/aos/1015345959>

Banerjee, M., & Wellner, J. A. (2005). Confidence intervals for current status data. *Scandinavian Journal of Statistics*, 32(3), 405-424. <https://doi.org/10.1111/j.1467-9469.2005.00454.x>

Cattaneo, M. D., Jansson, M., & Nagasawa, K. (2020). Bootstrap based inference for cube root asymptotics. *Econometrica*, 88(5), 2203-2219. <https://doi.org/10.3982/ECTA17950>

Groeneboom, P., & Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. Dmv Seminar. Birkhäuser Basel.

Groeneboom, P., & Wellner, J. A. (2001). Computing Chernoff’s distribution. *Journal of Computational and Graphical Statistics*, 10(2), 388-400. <https://doi.org/10.1198/10618600152627997>

Groeneboom, P., Jongbloed, G., & Witte, B. I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *The Annals of Statistics*, 38(1), 352-387. <https://doi.org/10.1214/09->

AOS721

- Groeneboom, P. (2012). Likelihood ratio type two sample tests for current status data. *Scandinavian Journal of Statistics*, 39(4), 645-662. <https://doi.org/10.1111/j.1467-9469.2012.00789.x>
- Groeneboom, P., & Jongbloed, G. (2015). Nonparametric confidence intervals for monotone functions. *The Annals of Statistics*, 43(5), 2019-2054. <https://doi.org/10.1214/15-AOS1335>
- Groeneboom, P., & Hendrickx, K. (2017). The nonparametric bootstrap for the current status model. *Electronic Journal of Statistics*, 11(2), 3446-3484. <https://doi.org/10.1214/17-EJS1345>
- Keiding, N., Begtrup, K., Scheike, T. H., & Günther, H. (1996). Estimation from current-status data in continuous time. *Lifetime Data Analysis*, 2(2), 119-129. <https://doi.org/10.1007/BF00128570>
- Ou, F. S., Zeng, D., & Cai, J. (2016). Quantile regression models for current status data. *Journal of Statistical Planning & Inference*, 178, 112-127. <https://doi.org/10.1016/j.jspi.2016.06.001>
- Sen, B., & Xu, G. (2015). Model based bootstrap methods for interval censored data. *Computational Statistics & Data Analysis*, 81, 121-129. <https://doi.org/10.1016/j.csda.2014.07.007>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).