Nonparametric Estimation of Quantile and Quantile Density Function

Yang X*, Hutson AD2,3 and Wang D2

1Department of Biostatistics, University at Buffalo, Buffalo, NY 14214, USA
2Roswell Park Cancer Institute, Buffalo, NY 14203, USA
3Department of Public Health and Preventive Medicine, SUNY Upstate Medical University, Syracuse, NY 13210, USA

Abstract

In this article, we derive a new and unique method of estimating quantile and quantile density function, which is based on moments of fractional order statistics. A comparison of the proposed estimators is made with existing popular nonparametric quantile and quantile density estimators, in terms of mean squared error (MSE) for censored and uncensored data. Recommendations for the choice of quantile and/or quantile density estimators are given.

Keywords: Quantile function estimators; Quantile density function estimators; Order statistics; Kernel function

Introduction

The quantile function

\[ Q(u) = \inf \{ x : F(x) \geq u \} , \]

where \( F(.) \) is the cumulative distribution function (CDF) of a continuous random variable \( X \) and \( 0 < u < 1 \), is an alternative to the probability density function (PDF), the CDF and the characteristic function for describing a probability distribution. The estimation of \( Q(u) \) is of great interest, especially when one is unwilling to assume the distribution is parametric or when the underlying distribution is skewed.

The use of quantile function estimation has been around for decades in exploratory data analysis, statistical analysis, reliability and medical studies [1-11]. A more recent application can be found in Jeong and Fine [12] and Sankaran et al. [13] for competing risk models.

Many nonparametric estimators of the quantile function have been proposed and studied extensively. For uncensored data, the simplest method is the empirical quantile (EQ) estimator based on a single order statistic. It is a piecewise constant function that does not provide a useful quantile density function estimation. Details about advantages of smoothed quantile estimators can be found in Cheng and Parzen [7]. Numerous smoothed quantile function estimators have been introduced. Here, only the most representative ones are outlined. The most commonly used estimator is the linear interpolation of successive order statistics, which is employed in applications, for example, Q-Q plots and popular software packages such as SAS, BMDP, and MINITAB. Parzen [1] developed kernel smoothing of the EQ estimator, which is well known as the kernel quantile estimator. It has been extensively studied and analyzed [3-5,8]. More complete literature reviews on kernel-based quantile estimators can be found in Sheather and Marron [8] and Cheng and Parzen [7]. However, kernel quantile estimators in general are complicated and analytically intractable. Their performance in the sense of MSEs is very sensitive to the choice of bandwidth. In addition, the approximations to kernel estimators may violate the monotonicity requirement as described by Yang [5] and this issue is also shown in our simulation study and the real data application in Section 5 and Section 6. Generalized order statistics were considered as alternatives to sample quantiles by Harrell and Davis [14] and Kaigh and Lachenbruch [15]. Huang [16] proposed a modification of the Harrell-Davis (HD) estimator based on developing a weighting scheme through the use of the level crossing empirical distribution function.

In the presence of right-censored data, the product-limit quantile (PLQ) estimator proposed by Sander [17] and the general kernel smoothing version of PLQ estimator by Padgett [18], which share similar problems as their parallels for uncensored data, have gained the most popularity in the literature. More recently, Wang et al. [19] extended the HD quantile function estimator for censored data and proposed an exact bootstrap procedure for optimization in terms of MSE related criteria.

In the same way that the CDF can be differentiated to give the PDF, Parzen [1] and Jones [20] defined the derivative of \( Q(u) \) as the quantile density function. That is, \( q(u)=Q'(u) \). Common applications of \( q(u) \) include but are not limited to constructing the asymptotic confidence interval of sample quantiles, inference procedures based on linear rank statistics in Hettmansperger [21], and quantile density based approach in the location scale problem, see Eubank [22].

To estimate the quantile density function \( q(u) \), given either censored or uncensored data, two main approaches can be applied. One is the mathematical derivative of quantile estimators (if differentiable), the other is the reciprocal of density quantile function \( f(Q(u)) \) obtained by differentiating on both sides of the equation \( F(Q(u))=u \). The former way is more advantageous over the latter in terms of efficiency, see Jones [20] for more information pertaining to the comparison of these two methods. In addition, kernel smoothing of the reciprocal of density quantile function has also been considered [23]. As mentioned before, PLQ and EQ estimators fail to provide useful quantile density function estimation. However, the linear interpolation of two successive order statistics is differentiable and the resulting quantile density function estimator is a histogram type estimator by Siddiqui [24]. The derivative of the kernel quantile estimator by Parzen [1] was introduced by Falk [25]. Xiang [26] proposed a natural derivative of the quantile estimator by Padgett [18]. More reviews on quantile density estimators can be found in Cheng and Parzen [7].

In this article we take a new and novel approach to quantile and quantile density function estimation based on estimating moments of

*Corresponding author: Yang X, Department of Biostatistics, University at Buffalo, Buffalo, NY 14214-3000, USA, Tel: (716) 845-1300; E-mail: xyang37@buffalo.edu

Received May 26, 2017; Accepted June 16, 2017; Published June 20, 2017

Citation: Yang X, Hutson AD, Wang D (2017) Nonparametric Estimation of Quantile and Quantile Density Function. J Biom Biostat 8: 356. doi: 10.4172/2155-6180.1000356

Copyright: © 2017 Yang X, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
fractional order statistics and solving a set of simultaneous equations pertaining to a series of moment expansions. We studied and compared the performance of our estimators with EQ estimator, PLQ estimator, the kernel smoothing of EQ and PLQ estimators, piecewise linear estimator and their corresponding quantile density estimators (if exist) for censored and uncensored data. The competing estimators were considered simply because they are commonly used for quantile and quantile density estimation. The advantages of our method are as follows: First, it does not require a selection of the optimal bandwidth and therefore can be more stable compared to the common kernel-based methods. Second, it at least outperforms the PL and the piecewise linear quantile estimators across all possible simulation parameters we considered in terms of MSEs and also appears to preserve the monotonicity of the quantile function curve. Third, the associated quantile density function estimator is shown to yield the smallest MSE among all quantile density estimators considered for both censored and uncensored data.

In Sections 2 and 3, we outline the existing methods of quantile and quantile density estimation considered in this investigation. In Section 4, our new quantile and quantile density estimators are introduced. The performance of our estimators is illustrated in terms of MSE by a Monte Carlo simulation study in Section 5. This is followed by an application of the switch life data reported by Nair [27] in Section 6. Recommendations of the choice of quantile, and/or quantile density estimators are summarized in Section 7.

Estimation of Q(u)

Let X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)} be the order statistics from a random sample with a continuous distribution F(t). And let T_1 \leq T_2 \leq \cdots \leq T_n be the order statistics corresponding to the i.i.d. sample of randomly right-censored times T_1, T_2, \ldots, T_n, and \delta_{(i)} \leq \delta_{(i)} \leq \cdots \leq \delta_{(0)} are censoring indicators corresponding to the ordered T_1, \ldots, T_n, respectively. A value of \delta_{(i)} = 0.1 indicates that T_1 is uncensored, while a value of \delta_{(i)} = 0 indicates that T_1 is censored. The methods described below can be readily applied for uncensored data by setting all \delta_{(i)} = 1.

Then the well-known PLQ estimator by Sander [17], is defined as

\hat{Q}_n(u) = \inf\{x : 1 - \hat{S}_n(x) \geq u\},

where \hat{S}_n is the common PL estimator of the survival function by Kaplan and Meier [28]:

\hat{S}_n(t) = \left\{ \begin{array}{ll}
\prod_{i=1}^{\min(n-t, n)} \left(1 - \frac{j}{n-j+1}\right) & \text{if } t < T_{(i)}; \\
0 & \text{if } t \geq T_{(i)}.
\end{array} \right.

Defining \hat{S}_n(t) = 0 for t \geq T_{(i)} has been studied by Xiang [26] with respect to convergence properties for a class of kernel quantile function estimators, and was shown to provide a more technically suitable definition in term of the large sample theory as compared to setting \hat{S}_n(t) as undefined.

In the absence of censoring, \hat{Q}_n(u) reduces to the EQ estimator.

\hat{Q}_0(u) = \inf\{x : \hat{F}_n(x) \geq u\},

where \hat{F}_n(x) is the empirical distribution, n^{-1} \sum_{i=1}^{n}I(X_{(i)} \leq x).

The linear interpolation estimator of the quantile function given uncensored data is denoted as

\hat{Q}_L(u) = (1 - \epsilon)X_{\lceil nu \rceil} + \epsilon X_{\lceil nu + \epsilon \rceil},

where \epsilon = n'u - \lceil nu \rceil and \epsilon = n' u - \lceil nu \rceil.

Padgett (1986) defined the kernel smoothing version of \hat{Q}_n(u) as

\hat{Q}_{kn}(u) = h^{-1} \int \hat{Q}_{kn}(u) K\left(\frac{t-u}{h}\right) dt

= h^{-1} \sum_{i=1}^{n} \int \hat{Q}_{kn}(u) K\left(\frac{t-u}{h}\right) dt

where S is the (3) at T_{(i)} and K(.) is a symmetric kernel function. If no censoring, the estimator in eqn. (6) reduces to the general kernel smoothing of the EQ estimator by Parzen [1],

\hat{Q}_{kn}(u) = h^{-1} \sum_{i=1}^{n} X_{(i)} K\left(\frac{t-u}{h}\right) dt.

Estimation of q(u)

Let, \frac{i}{n} \leq u < \frac{i+1}{n}, i = 1, 2, \ldots, n - 1 the first order derivative of \hat{Q}_n(u) is:

\hat{q}_n(u) = n' \left(X_{(i+1)} - X_{(i)}\right)

This is called the spacings of the sample, see Pyke [29,30], or a histogram type estimator by the fact that \hat{q}_n(u) = 1/\hat{f}(\hat{Q}(u)) , where \hat{f}(\hat{Q}(u)) is the density quantile function estimator based on finite differences introduced by Siddiqui [24].

Again, the PLQ estimator and the EQ estimator do not have the corresponding quantile density estimators. The natural derivative of (6), \hat{Q}_{kn}(u), was established by Xiang [26] as

\hat{q}_{kn}(u) = -h^{-2} \sum_{i=1}^{n} s_i K'\left(\frac{1 - \hat{S}_n(T_{(i)}) - t}{h}\right),

where s_i denotes the jump of \hat{S}_n(T_{(i)}) , that is,

s_i = \left\{ \begin{array}{ll}
1 - \hat{S}_n(T_{(i)}) & i = 1 \\
\hat{S}_n(T_{(i-1)}) - \hat{S}_n(T_{(i)}) & i = 2, 3, \ldots, n
\end{array} \right.

When no censoring is present, eqn. (9) reduces to the kernel quantile density estimator \hat{q}_{kn}(u), which is the complementary of Parzen [1] estimator in eqn. (7).

Fractional Order Statistic-Based Quantile and Quantile Density Estimator

In order to develop our new quantile and quantile density function estimators, we need to derive asymptotic expansions corresponding to the kth non-central moment of the fractional order statistic, E[X_{\nu,k}], where the fractional order statistic X_{\nu,\alpha} for an i.i.d. uniform sample jointly follow a particular Dirichlet process U_{\nu,\alpha}; see Stigler [31]. Even though fractional order statistics do not exist in the empirical sense, their respective expectations may be calculated.

In deriving the expansion we assume that the first three derivatives of Q are bounded in a neighborhood of u and denote them as Q' = Q(u), Q'' = Q'(u) and Q''' = Q''(u). We also assume, similar to the results pertaining to the kth order statistics as in Section 3.1 and Section
is given by
\[ \mu = \frac{1}{2(n+2)} \left( u - 2Q(u)^3 + 2Q(u)^2 Q'(u) \right) + O(n^{-2}), \]
(11)
\[ \hat{\mu}_{\text{KPL}}^2 = Q(u)^2 + \frac{1}{2(n+2)} \left( 2Q(u)^3 + 2Q(u)^2 Q'(u) \right), \]
(12)
\[ \hat{\mu}_{\text{M.i}}^2 = Q(u)^2 + \frac{1}{2(n+2)} \left( 6Q(u)^3 + 3Q(u)^2 Q'(u) \right), \]
(13)

Proof: First note that \( X_{u:n} - Q(U_{u:n}) \). Expanding \( Q(U_{u:n}) \) about \( U_{u:n} = u \) and taking expectations of each term in the expansion based upon a Beta distribution \( B(nu,n(1-u)) \) proves the result. Note that the bounds in terms of \( Q'' \) is with respect to the remainder term of the expansion.

We may utilize equations 11-13 to find an approximation for \( Q(u) \) as a function of \( E(X_{u:n}) \), \( k = 1, 2, 3 \). Hence, we propose:

Definition: The moment quantile estimator, \( \hat{\mu}_{\text{m}}(u) \), is given by solving the set of simultaneous nonlinear equations
\[ \hat{\mu}_{\text{m}} = Q(u) + \frac{1}{2(n+2)} \left( u - 2Q(u)^3 + 2Q(u)^2 Q'(u) \right), \]
(14)
\[ \hat{\mu}_{\text{KPL}}^2 = Q(u)^2 + \frac{1}{2(n+2)} \left( 2Q(u)^3 + 2Q(u)^2 Q'(u) \right), \]
(15)
\[ \hat{\mu}_{\text{M.i}}^2 = Q(u)^2 + \frac{1}{2(n+2)} \left( 6Q(u)^3 + 3Q(u)^2 Q'(u) \right), \]
(16)

at a fixed \( u \) in terms of \( Q(u) \), \( Q(u) \) and \( Q'(u) \) in order to estimate \( Q(u) \), where
\[ \hat{\mu}_{\text{m}}^2, k = 1, 2, 3, \] is given by the exact bootstrap \( k \)th moment estimator of the \( u \)th fraction order statistic by Hutsun and Ernst [32],
\[ \hat{\mu}_{\text{m}}^2 = \sum_{i=1}^{n} X_i^k \]
(17)

The weight \( w_i \) is given as
\[ w_i = B_{n,u,n-1} \left( \frac{i}{n} - B_{n,u,n-1} \left( \frac{i-1}{n} \right) \right) \]
(18)

for uncensored data and
\[ w_i = B_{n,u,n-1} \left( 1 - S_{PL} (T_{i-1}) \right) \]
(19)

for censored data in Wang et al. [19], where \( B_{n}(x) = \frac{1}{n} \binom{n-1}{x-1} \) is the incomplete beta function. As an aside, estimates of \( Q(u) \) and \( Q'(u) \) are also available as part of this process. In this investigation, we only interested in the performance of the first-order derivative of \( Q(u) \), i.e., the quantile density function \( q(u) \), and let's denote the quantile density estimator as \( \hat{q}_u(u) \). In the later simulation, we will show that \( \hat{q}_u(u) \) is at least more reliable than \( \hat{q}_{\text{KPL}}(u) \), \( \hat{q}_{\text{M.i}}(u) \), and \( \hat{q}_u(u) \) in terms of MSE.

If we are only interested in an estimate for \( Q(u) \) then the numerical solution with respect to \( Q(u) \) in terms of our system of equations 17-19 is relatively straightforward and reduces to solving the simple cubic equation
\[ E'\left(X_{u:n} - Q(u)\right)^3 = 0. \]
(20)

with respect to \( Q(u) \), \( u \) fixed, where \( E' \) denotes the exact bootstrap moment estimator of the quantity at (4.10). Alternatively, we can reformulate (20) and define it as follows:

Definition: The cubic quantile estimator, \( \hat{Q}_c(u) \), is given by minimizing
\[ \sum_{i=1}^{n} |X(i) - Q(u)|^3 w_i, \]
(21)

with respect to \( Q(u) \), where the weights \( w_i \) are defined at (19) for censored data or at (18) for uncensored data. The associated quantile density function \( q_u(u) \) can still be estimated by taking the numerical derivative of an interpolated function from cubic quantile estimates at a proper number of \( u \) values. As motivated by this, an alternative way of estimating the moment quantile and quantile density functions can be interpolating a quantile function curve from the moment quantile estimates at a proper number of \( u \) values and taking the numerical first-order derivative of the interpolated function at \( u \) points, respectively. To distinguish with \( \hat{Q}_u(u) \) and \( \hat{q}_u(u) \), which are obtained by solving the equation system (14-16) simultaneously, we denote the alternatives as \( \hat{Q}_{\text{m}}(u) \) and \( \hat{q}_{\text{m}}(u) \). The simulation study in Section 5 shows that \( \hat{Q}_{\text{m}}(u) \) and \( \hat{Q}_{\text{m}}(u) \) yield basically the same quantile estimations. Even though \( \hat{Q}_{\text{m}}(u) \) and \( \hat{Q}_{\text{m}}(u) \) are different, they are both significantly better than \( \hat{q}_{\text{KPL}}(u) \), \( \hat{q}_{\text{M.i}}(u) \), and \( \hat{q}_u(u) \) in terms of MSE criteria.

Simulation Results

For the purpose of illustrating the behavior of our estimators, a straight forward simulation study was carried out for samples of size \( n=30, 50, 100 \) for Weibull distribution with the quantile function, \( Q(u) = (-\log(1-u))^\theta \), where \( \theta=0.5,1,1.5 \) and across the standardized normal, exponential, extreme value, and logistic distributions. The censoring distribution was given as a uniform distribution uniform \((0,T) \) with \( T=2.5 \) and uncensored case was also considered. We utilized fixed quantiles of \( u=0.25, 0.5, \) and 0.75. For each combination, 2,000 Monte Carlo simulations were utilized.

For censored data, a comparison of \( \hat{Q}_{\text{m}}, \hat{q}_{\text{KPL}}, \hat{q}_{\text{M.i}}, \hat{Q}_{\text{c}} \) and \( \hat{Q}_{\text{c}} \) was made in terms of MSEs in Table 1-4. For uncensored data, quantile estimators we considered are \( \hat{Q}_{\text{KPL}}, \hat{Q}_{\text{KPL}}, \hat{Q}_{\text{M.i}}, \hat{Q}_{\text{c}}, \) and \( \hat{Q}_{\text{c}} \), as shown in Tables 5 and 6. MSEs of the corresponding quantile density estimators, if exist, were also summarized in Tables 7-12. Epanechnikov kernel function, 0.75\((1-u^2)^2([u] \leq 1)\), was utilized here for the considered kernel estimators. Note that even though the triangular kernel is more commonly used for kernel quantile estimators in literature, see Padgett [18], Nair and Sankaran [33], and Soni [23], it fails to provide a useful derivative when calculating the \( K'() \) in the kernel quantile density estimators. The Epanechnikov kernel, which gives the optimal kernel, see Prakasa Rao [34], was studied by Soni [23] for comparing non-parametric quantile density estimators and our simulation showed that \( Q(u) \) estimation behaviors under Epanechnikov kernel and triangular kernel were quite close, which was also confirmed in the study of Soni [23]. For simplicity, only the result of the Epanechnikov kernel was presented here. Bandwidth
\( \theta = 0.5 \) \( \theta = 1 \) \( \theta = 1.5 \)

| n   | method | \( u = 0.25 \)  | \( u = 0.50 \)  | \( u = 0.75 \) | \( u = 0.25 \)  | \( u = 0.50 \)  | \( u = 0.75 \) | \( u = 0.25 \)  | \( u = 0.50 \)  | \( u = 0.75 \) |
|-----|--------|-----------------|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 30  | \( \hat{Q}_n \) | 13.247          | 16.832          | 36.021         | 14.756         | 45.677         | 173.423        | 12.858         | 70.408         | 571.736        |
| 30  | \( \hat{Q}_{k,n} \) | 5.413           | 13.528          | 64.776         | 16.357         | 32.939         | 651.067        | 10.903         | 36.465         | 1200.693       |
| 30  | \( \hat{Q}_w \) | 11.117          | 16.527          | 72.228         | 14.665         | 44.066         | 527.885        | 14.576         | 66.225         | 1309.573       |
| 30  | \( \hat{Q}_{v,n} \) | 11.117          | 16.527          | 72.228         | 14.665         | 44.066         | 527.905        | 14.576         | 66.225         | 1309.574       |
| 30  | \( \hat{Q}_c \) | 9.492           | 14.319          | 22.232         | 13.03          | 33.949         | 173.423        | 13.362         | 47.344         | 389.307        |
| 50  | \( \hat{Q}_n \) | 7.456           | 11.262          | 23.24          | 8.372          | 30.323         | 129.169        | 5.771          | 43.967         | 497.303        |
| 50  | \( \hat{Q}_{k,n} \) | 3.247           | 8.085           | 28.04          | 9.741          | 17.729         | 360.658        | 5.642          | 87.146         | 1218.629       |
| 50  | \( \hat{Q}_w \) | 6.433           | 10.034          | 24.181         | 8.021          | 28.461         | 318.258        | 6.907          | 45.937         | 1162.492       |
| 50  | \( \hat{Q}_{v,n} \) | 6.433           | 10.034          | 24.181         | 8.021          | 28.461         | 318.258        | 6.907          | 45.937         | 1162.487       |
| 50  | \( \hat{Q}_c \) | 5.697           | 8.746           | 16.202         | 7.177          | 24.194         | 54.206         | 6.263          | 36.878         | 240.124        |
| 100 | \( \hat{Q}_n \) | 3.681           | 5.323           | 10.703         | 3.935          | 13.744         | 75.536         | 2.591          | 21.057         | 378.403        |
| 100 | \( \hat{Q}_{k,n} \) | 2.584           | 5.629           | 10.762         | 12.72          | 11.341         | 157.495        | 16.762         | 16.339         | 794.296        |
| 100 | \( \hat{Q}_w \) | 3.198           | 4.763           | 9.513          | 3.61           | 13.204         | 138.789        | 2.739          | 22.763         | 787.276        |
| 100 | \( \hat{Q}_{v,n} \) | 3.198           | 4.763           | 9.513          | 3.61           | 13.204         | 138.779        | 2.739          | 22.763         | 787.276        |
| 100 | \( \hat{Q}_c \) | 2.929           | 4.4             | 8.781          | 3.317          | 11.965         | 27.973         | 2.468          | 20.106         | 111.144        |

Table 1: Mean squared error of \( Q(x \times 1000), T=2 \).
\[ \theta = 0.5 \]

\[
\begin{array}{lccccccccccc}
\text{n} & \text{method} & u = 0.25 & u = 0.50 & u = 0.75 & u = 0.25 & u = 0.50 & u = 0.75 & u = 0.25 & u = 0.50 & u = 0.75 \\
30 & \hat{Q}_n & 10.859 & 13.27 & 22.34 & 14.146 & 37.672 & 142.768 & 11.943 & 64.733 & 457.705 \\
30 & \hat{Q}_{\text{w}} & 3.597 & 8.835 & 17.584 & 13.142 & 23.019 & 120.039 & 15.396 & 105.723 & 324.724 \\
30 & \hat{Q}_{\text{i}} & 9.309 & 12.372 & 20.408 & 13.291 & 42.257 & 164.929 & 13.767 & 105.495 & 414.619 \\
30 & \hat{Q}_c & 8.127 & 10.686 & 17.804 & 14.146 & 37.672 & 142.768 & 11.943 & 64.733 & 457.705 \\
50 & \hat{Q}_n & 6.414 & 8.686 & 17.804 & 6.414 & 8.686 & 17.804 & 6.414 & 8.686 & 17.804 \\
50 & \hat{Q}_{\text{w}} & 5.674 & 7.417 & 12.188 & 5.674 & 7.417 & 12.188 & 5.674 & 7.417 & 12.188 \\
50 & \hat{Q}_{\text{i}} & 1.329 & 4.891 & 8.733 & 1.329 & 4.891 & 8.733 & 1.329 & 4.891 & 8.733 \\
50 & \hat{Q}_c & 2.837 & 3.85 & 6.242 & 2.837 & 3.85 & 6.242 & 2.837 & 3.85 & 6.242 \\
100 & \hat{Q}_n & 3.245 & 4.175 & 6.825 & 3.245 & 4.175 & 6.825 & 3.245 & 4.175 & 6.825 \\
100 & \hat{Q}_{\text{w}} & 1.329 & 4.891 & 8.733 & 1.329 & 4.891 & 8.733 & 1.329 & 4.891 & 8.733 \\
100 & \hat{Q}_{\text{i}} & 2.837 & 3.85 & 6.242 & 2.837 & 3.85 & 6.242 & 2.837 & 3.85 & 6.242 \\
100 & \hat{Q}_c & 2.62 & 3.617 & 5.779 & 2.62 & 3.617 & 5.779 & 2.62 & 3.617 & 5.779 \\
\end{array}
\]

Table 3: Mean squared error of \( Q(\times 1000) \), \( T = 5 \).

\[ \theta = 1 \]

\[
\begin{array}{lccccccccccc}
\text{n} & \text{method} & u = 0.25 & u = 0.50 & u = 0.75 & u = 0.25 & u = 0.50 & u = 0.75 & u = 0.25 & u = 0.50 & u = 0.75 \\
30 & \hat{Q}_n & 10.859 & 13.27 & 22.34 & 14.146 & 37.672 & 142.768 & 11.943 & 64.733 & 457.705 \\
30 & \hat{Q}_{\text{w}} & 3.597 & 8.835 & 17.584 & 13.142 & 23.019 & 120.039 & 15.396 & 105.723 & 324.724 \\
30 & \hat{Q}_{\text{i}} & 9.309 & 12.372 & 20.408 & 13.291 & 42.257 & 164.929 & 13.767 & 105.495 & 414.619 \\
30 & \hat{Q}_c & 8.127 & 10.686 & 17.804 & 14.146 & 37.672 & 142.768 & 11.943 & 64.733 & 457.705 \\
50 & \hat{Q}_n & 6.414 & 8.686 & 17.804 & 6.414 & 8.686 & 17.804 & 6.414 & 8.686 & 17.804 \\
50 & \hat{Q}_{\text{w}} & 5.674 & 7.417 & 12.188 & 5.674 & 7.417 & 12.188 & 5.674 & 7.417 & 12.188 \\
50 & \hat{Q}_{\text{i}} & 1.329 & 4.891 & 8.733 & 1.329 & 4.891 & 8.733 & 1.329 & 4.891 & 8.733 \\
50 & \hat{Q}_c & 2.837 & 3.85 & 6.242 & 2.837 & 3.85 & 6.242 & 2.837 & 3.85 & 6.242 \\
100 & \hat{Q}_n & 3.245 & 4.175 & 6.825 & 3.245 & 4.175 & 6.825 & 3.245 & 4.175 & 6.825 \\
100 & \hat{Q}_{\text{w}} & 1.329 & 4.891 & 8.733 & 1.329 & 4.891 & 8.733 & 1.329 & 4.891 & 8.733 \\
100 & \hat{Q}_{\text{i}} & 2.837 & 3.85 & 6.242 & 2.837 & 3.85 & 6.242 & 2.837 & 3.85 & 6.242 \\
100 & \hat{Q}_c & 2.62 & 3.617 & 5.779 & 2.62 & 3.617 & 5.779 & 2.62 & 3.617 & 5.779 \\
\end{array}
\]

Table 4: Mean squared error of \( Q(\times 1000) \), Weibull, uncensored.
| n  | method | U = 0.25 | U = 0.50 | U = 0.75 | U = 0.25 | U = 0.50 | U = 0.75 | U = 0.25 | U = 0.50 | U = 0.75 |
|----|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 30 | \(\hat{\theta}_c\) | 9.62     | 11.277   | 18.28    | 11.899   | 33.147   | 109.119  | 9.352    | 57.895   | 380.881  |
| 30 | \(\hat{\theta}_{EQ}\) | 10.7     | 12.471   | 20.634   | 14.211   | 36.217   | 128.025  | 11.423   | 61.913   | 437.653  |
| 30 | \(\hat{\theta}_{EQ}\) | 3.091    | 22.452   | 33.251   | 11.977   | 16.033   | 91.378   | 9.72     | 58.631   | 516.648  |
| 30 | \(\hat{\theta}_h\) | 8.589    | 10.724   | 16.607   | 12.523   | 37.139   | 124.11   | 12.983   | 87.72    | 605.788  |
| 30 | \(\hat{\theta}_{Ea}\) | 8.589    | 10.724   | 16.611   | 12.523   | 37.139   | 124.11   | 12.983   | 87.72    | 605.788  |
| 30 | \(\hat{\theta}_c\) | 7.489    | 9.382    | 14.54    | 11.086   | 31.1     | 97.904   | 11.767   | 67.257   | 418.041  |
| 50 | \(\hat{\theta}_c\) | 5.707    | 6.859    | 10.929   | 6.851    | 19.865   | 63.43    | 4.967    | 33.457   | 212.619  |
| 50 | \(\hat{\theta}_{EQ}\) | 6.126    | 7.566    | 12.106   | 7.656    | 22.171   | 69.311   | 5.712    | 36.968   | 231.98   |
| 50 | \(\hat{\theta}_{EQ}\) | 1.947    | 4.382    | 8.281    | 6.367    | 12.328   | 49.203   | 5.31     | 40.178   | 231.919  |
| 50 | \(\hat{\theta}_h\) | 5.169    | 6.47     | 9.996    | 6.963    | 20.852   | 66.823   | 6.426    | 42.851   | 280.134  |
| 50 | \(\hat{\theta}_{Ea}\) | 5.169    | 6.47     | 9.993    | 6.963    | 20.852   | 66.823   | 6.426    | 42.851   | 280.134  |
| 50 | \(\hat{\theta}_c\) | 4.633    | 5.869    | 9.074    | 6.245    | 18.353   | 56.985   | 5.653    | 35.693   | 217.612  |
| 100 | \(\hat{\theta}_c\) | 2.955    | 3.553    | 5.385    | 3.464    | 9.998    | 30.3     | 2.363    | 16.108   | 97.287   |
| 100 | \(\hat{\theta}_{EQ}\) | 3.092    | 3.874    | 5.97     | 3.617    | 10.569   | 33.835   | 2.414    | 17.005   | 111.129  |
| 100 | \(\hat{\theta}_{EQ}\) | 0.939    | 2.171    | 4.534    | 7.287    | 8.746    | 21.288   | 4.971    | 38.55    | 64.046   |
| 100 | \(\hat{\theta}_h\) | 2.649    | 3.354    | 4.859    | 3.307    | 9.873    | 29.37    | 2.649    | 17.553   | 106.727  |
| 100 | \(\hat{\theta}_{Ea}\) | 2.649    | 3.354    | 4.859    | 3.307    | 9.873    | 29.37    | 2.649    | 17.553   | 106.727  |
| 100 | \(\hat{\theta}_c\) | 2.458    | 3.16     | 4.536    | 3.046    | 9.188    | 26.55    | 2.351    | 15.903   | 91.397   |

Table 5: Mean squared error of Q(-1000), uncensored.
\[ \theta = 0.5 \]
\[ \theta = 1 \]
\[ \theta = 1.5 \]

| n | method | \( u = 0.25 \) | \( u = 0.50 \) | \( u = 0.75 \) | \( u = 0.25 \) | \( u = 0.50 \) | \( u = 0.75 \) | \( u = 0.25 \) | \( u = 0.50 \) | \( u = 0.75 \) |
|---|---|---|---|---|---|---|---|---|---|---|
| 30 | \( \hat{q}_{L/C} \) | 8149.28 | 21114.3 | 302430 | 5279.13 | 81367 | 10753.5 | 184243 | 30426.8 | 437681 |
| 30 | \( \hat{q}_M \) | 190.06 | 138.632 | 1773.02 | 357.404 | 416.272 | 19065.2 | 554.12 | 1016.54 | 60623.1 |
| 30 | \( \hat{q}_L \) | 346.405 | 454.448 | 7033.58 | 375.337 | 1521.34 | 58892.9 | 472.144 | 2854.32 | 131391 |
| 30 | \( \hat{q}_C \) | 140.399 | 164.01 | 588.519 | 233.121 | 461.381 | 38923.3 | 334.335 | 808.027 | 35986.6 |
| 50 | \( \hat{q}_{L/C} \) | 5365.27 | 9166.75 | 79146.1 | 2799.42 | 31653.6 | 3.2E+08 | 8938081 | 6779.78 | 33467.5 |
| 50 | \( \hat{q}_M \) | 171.934 | 115.401 | 546.322 | 263.807 | 186.036 | 16876.7 | 334.348 | 589.967 | 71765.9 |
| 50 | \( \hat{q}_L \) | 218.037 | 234.301 | 3712.65 | 246.717 | 560.172 | 46448.3 | 282.841 | 561.617 | 32809.8 |
| 50 | \( \hat{q}_C \) | 3457.75 | 5540.73 | 39824.2 | 2876.18 | 350462 | 4700535 | 1754.15 | 96355.8 | 125608.3 |
| 100 | \( \hat{q}_{L/C} \) | 1.8E+09 | 8264.46 | 21800.3 | 2806.18 | 93756.5 | 141038 | 4700535 | 1754.15 | 96355.8 |
| 100 | \( \hat{q}_M \) | 1055.82 | 550.683 | 1276.25 | 273.726 | 176.808 | 16926.5 | 605.194 | 508.001 | 18993.3 |
| 100 | \( \hat{q}_L \) | 960.359 | 1667.24 | 2538.14 | 249.059 | 742.928 | 69260 | 1087.41 | 1370.22 | 56219.3 |
| 100 | \( \hat{q}_C \) | 375.799 | 215.821 | 659.105 | 149.814 | 426.399 | 68805.3 | 30426.8 | 71765.9 | 131391 |

Table 8: Mean squared error of \( q(\times 1000) \), T = 5.
\[ \theta = 0.5 \]  
\[ \theta = 1 \]  
\[ \theta = 1.5 \]

| n | method | \( u = 0.25 \) | \( u = 0.50 \) | \( u = 0.75 \) | \( u = 0.25 \) | \( u = 0.50 \) | \( u = 0.75 \) | \( u = 0.25 \) | \( u = 0.50 \) | \( u = 0.75 \) |
|---|---|---|---|---|---|---|---|---|---|---|
| 30 | \( \hat{q}_{L} \) | 38.45 | 85.5 | 5.1 | 54.3 | 7.2 | 54.1 | 7.2 | 719 | 6.9 | 596.6 | 4.5 |
| 30 | \( \hat{q}_{M} \) | 125.3 | 654 | 1 | 127.2 | 2.8 | 280.4 | 1.7 | 173.8 | 1.7 | 744.3 | 1.2 |
| 30 | \( \hat{q}_{M} \) | 1495.8 | 2643 | 93 | 1774 | 43 | 270.2 | 54 | 773.5 | 99 | 4735.6 | 3.6 |
| 30 | \( \hat{q}_{C} \) | 812.46 | 417.749 | 964.279 | 185.737 | 500.469 | 2634.03 | 576.176 | 743.603 | 3127.67 | 1038.3 | 5797.39 |
| 50 | \( \hat{q}_{L} \) | 2.4E+07 | 8598.16 | 24053.7 | 39256.3 | 5429.6 | 17970.9 | 154250 | 274176.7 | 31780.1 | 109331 | 22086.3 |
| 50 | \( \hat{q}_{M} \) | 960.4 | 1641.38 | 1224 | 6582 | 496.723 | 3758.39 | 422.942 | 478.437 | 2876.6 | 3477.37 |
| 50 | \( \hat{q}_{C} \) | 594.953 | 311.538 | 721 | 009 | 140.66 | 315.854 | 2450.21 | 999.108 | 453.36 | 10658 | 3477.37 |
| 100 | \( \hat{q}_{L} \) | 131996 | 8640.7 | 29730 | 1009626 | 5460.92 | 20570.2 | 73271.3 | 11335.3 | 74396.3 | 666840 | 22098.9 |
| 100 | \( \hat{q}_{M} \) | 917.3 | 350.549 | 941.464 | 164.87 | 109.83 | 7394.43 | 427.406 | 331.716 | 10603.5 | 850.49 | 1450.13 |
| 100 | \( \hat{q}_{C} \) | 618.658 | 818.937 | 771.579 | 130.775 | 281.672 | 1825.94 | 528.506 | 577.48 | 2267.99 | 1658.73 | 1721.42 |
| 100 | \( \hat{q}_{L} \) | 375.79 | 205.667 | 445.164 | 88.231 | 197.073 | 1307.94 | 309.365 | 324.425 | 1596.78 | 1199.19 | 551.284 |

Table 10: Mean squared error of \( q(\times 1000) \), Weibull, uncensored.

h for kernel estimators was chosen based on minimizing the bootstrap MSE, in which 300 bootstrap samples with replacement at each value of \( u \) for each sample size, distribution and censoring combination were used [35-41].

From Tables 1-6, we conclude for nonparametric quantile estimators that:

\( \hat{Q}_{L} \) has the smallest MSE in the majority of cases except when \( u \) is large. But again, its performance is unstable or sensitive to the choice of \( h \) (e.g., A large MSE occurs in Table 1, \( n=100, \theta=1 \), and \( u=0.25 \)). In addition, the process of computing the optimal bandwidth based on minimizing MSE can be time-consuming.

When data are heavily censored and skewed, it seems that only \( \hat{Q}_{C} \) and \( \hat{Q}_{L} \) are good at tails. For example, in Tables 1 and 2, when \( u=0.75 \), MSEs of other estimators under Weibull, exponential, extreme value, and logistic distributions are at least twice as large as the MSE of \( \hat{Q}_{C} \). A further discussion on this is mentioned in Section 6.
Among estimators without the bandwidth selection, behaves the best for both censored and uncensored cases. And $\hat{Q}$ and $M_{i\hat{Q}}$ are almost the same given fixed $u$.

1. And from Tables 7-12, we conclude for nonparametric quantile density estimators that: $C_{\hat{q}}$ produces the smallest MSE in almost all cases. $KPL_{\hat{q}}$ or $KEQ_{\hat{q}}$ yields the largest MSE and is substantially larger than other quantile density estimators $M_{\hat{q}}$ is the second best quantile density estimator in terms of MSE, which also implies that it is better than $M_{i\hat{q}}$.

Application

A real life test data set with $n=40$ mechanical switches by Nair [27] was used for the purpose of illustration. The observed data is:

$$T=(1.151^*, 1.667, 2.119^*, 2.547, 1.170^*, 1.695, 2.135, 2.548, 1.248^*, 1.710, 2.197, 2.738^*, 1.331^*, 1.955^*, 2.199^*, 2.794, 1.381^*, 1.965, 2.227, 2.883^*, 2.261^*, 3.015, 1.577^*, 2.109, 2.349^*, 3.017, 1.584^*, 2.116^*, 2.369, 3.793^*)$$

where $^*$ denotes a censored.

Observation: There is 57.5% censoring or 24 censored observations in this dataset. The choice of bandwidth $h$ was borrowed from Padgett (1986) based on minimizing the bootstrap MSE: $h=0.28$ if $0 < u \leq 0.25$, $h=0.34$ if $0.25 < u \leq 0.90$, and $h=0.40$ if $0.90 < u \leq 1$.

$\hat{Q}_{m}$ and $\hat{Q}_{i m}$ were not included here since they do not show any additional advantages compared with the moment quantile method and the cubic quantile method in Definition 4.1-2 as shown in the simulation study.

Figures 1 and 2 showed the estimates of quantile and quantile density for this example data. We see from Figure 1 that $Q_{KPL}$, $Q_{M}$, and $Q_{c}$ do not differ much except at tails, and furthermore, only $Q_{m}$ and $Q_{c}$ preserve monotonicity of quantile function curves for this data. These matches with our finding in simulation study that $Q_{m}$ performs the best among all quantile density estimators. The fluctuated curve of $KPL_{\hat{q}}$ in Figure 2 may also give a little hint about how bad the KPL method may perform in estimating quantile density functions.

Conclusion

In this article, we proposed three types of smooth quantile and
quantile density function estimators. \( \hat{Q}_M \) and \( \hat{q}_M \) are better than \( \hat{Q}_{KL} \) and \( \hat{q}_{KL} \) in terms of MSEs and computational efficiency. But the cubic method, \( \hat{Q}_C \) and \( \hat{q}_C \), performs better than the two moment quantile methods mentioned above for both \( Q(u) \) and \( q(u) \) estimation, and also shows an obvious advantage in MSE to all the other nonparametric estimators, especially when it comes to quantile density function estimation.

In summary, if one is only interested in quantile estimation, \( \hat{Q}_{KLQ} \) or \( \hat{Q}_{KLQ} \), with modification at tails if in need, is a good choice. But if one prefers a more stable estimator, \( \hat{Q}_L \) or \( \hat{Q}_C \) may also be considered. For the estimation of quantile density function, \( \hat{Q}_C \) is clearly an optimal choice based on considerations of MSE, smoothness and simplicity. In addition, use \( \hat{Q}_L \) and \( \hat{q}_{KL} \) or \( \hat{q}_{KLQ} \) with care for quantile density estimation since the bias can be extremely large compared to other alternatives.

Acknowledgement

This work was supported by Roswell Park Cancer Institute and National Cancer Institute (NCI) grant P30CA016056 and NRG Oncology Statistical and Data Management Center grant U10CA180822.

References

1. Parzen E (1979) Nonparametric statistical data modeling. Journal of the American Statistical Association 74: 105-121.
2. Parzen E (1993) Change PP plot and continuous sample quantile function. Communications in Statistics Theory and Methods 22: 3287-3304.
3. Reiss RD (1980) Estimation of quantiles in certain nonparametric models. Annals of Mathematical Statistics 8: 87-105.
4. Reiss RD (1989) Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics. Springer-Verlag, New York.
5. Yang S (1985) A smooth nonparametric estimator of a quantile function. Journal of the American Statistical Association 80: 1004-1011.
6. Cheng C (1995) The Bernstein Polynomial Estimator of a Smooth Quantile Function. Statistics & Probability Letters 24: 321-330.
7. Cheng C, Parzen E (1997) Unified Estimators of Quantile and Quantile Density Functions, Journal of Statistical Planning and Inference 59: 291-230.
8. Sheather SJ, Marron JS (1990) Kernel quantile estimators. Journal of the American Statistical Association 85: 410-416.
9. Hutson AD (1999) Calculating nonparametric confidence intervals for quantiles using fractional order statistics. Journal of Applied Statistics 26: 343-353.
10. Hutson AD (2001) Rational Spline Estimators of the Quantile Function. Communications and Statistics-Simulation and Computation 30: 377-390.
11. Hutson AD (2002) A Semiparametric Quantile Function Estimator for use in Bootstrap Estimation Procedures. Statistics and Computing 12: 331-338.
12. Jeong JH, Fine JP (2009) Parametric regression on cumulative incidence function. Biostatistics 8: 184-196.
13. Sankaran PG, Nair NU, Sreedevi EP (2010) A quantile based test for comparing cumulative incidence functions of competing risks models. Statistics and Probability Letters 80: 866-891.
14. Harrell FE, Davis CE (1982) A new distribution-free quantile estimator. Biometrika 69: 635-640.
15. Kaigh WD, Lachenbruch PA (1982) A generalized quantile estimator. Comm. Statist A 11: 2217-2238.
16. Huang ML (2001) On a distribution-free quantile estimator, Computational Statistics Data Analysis 37: 477-486.
17. Sander J (1975) The Weak Convergence of Quantiles of the Product Limit Estimator, Technical, Report 5, Stanford University. Dept of Statistics.
18. Padgett (1986) A kernel-Type Estimator of a Quantile Function from right-censored data, Journal of the American Statistical Association 81: 215-222.
19. Wan D, Hutson AD, Gaile DP (2010) An exact bootstrap approach towards modification of the Harrell-Davis quantile function estimator for censored data. Journal of Nonparametric Statistics 22: 1039-1051.
20. Jones MC (1992) Estimating densities, quantiles, quantile densities and density quantiles. Annals of the Institute of Mathematical Statistics 44: 721-727.
21. Hettmansperger TP (1984) Statistical Inference Based on Ranks. Wiley, New York.
22. Eubank RL (1981) A density-quantile function approach to optimal spacings selection. Ann Statist 9: 494-500.
23. Soni, Dewamb I, Jain K (2012) Nonparametric estimation of quantile density function. Computational Statistics and Data Analysis.
24. Siddiqui MM (1960) Distribution of quantiles in samples from a bivariate population. J Res Nat Bureau Standards Section B 64: 145-150.
25. Falk M (1986) On the estimation of the quantile density function. Statist Probab Lett 4: 697-93.
26. Xiang X (1994) A Law of the Logarithm for Kernel Quantile Density Estimators. The Annals of Probability 22: 1078-1091.
27. Nair VN (1984) Confidence Bands for Survival Functions With Censored Data: A Comparative Study. Technometrics 26: 265-275.
28. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 52: 457-481.
29. Pyke R (1965) Spacings. Journal of the Royal Statistical Society Ser B 27: 395-449.
30. Pyke R (1972) Spacings Revisited. Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, 1, 417-427.
31. Sliger SM (1977) Fractional Order Statistics with Applications. American Statistical Society 72: 544-550.
32. Hutson AD, Ernst MD (2000) The Exact Bootstrap Mean and Variance of an Lestimator, Journal of the Royal Statistical Society–Series B 62: 89-94.
33. Nair NU, Sankaran PG (2009) Quantile based reliability analysis. Communications in Statistics Theory and Methods 38: 222-232.
34. Prakasa Rao BLS (1983) Nonparametric Functional Estimation. Academic Press, New York.
35. Cheng (2002) Almost-sure uniform error bounds of general smooth estimator’s of quantile density functions, Statistics and Probability Letters 59: 183-194.
36. Hutson AD (1999) Calculating nonparametric confidence intervals for quantiles using fractional order statistics, Journal of Applied Statistics 26: 343-353.
37. Hutson AD (2001) Rational Spline Estimators of the Quantile Function. Communications and Statistics-Simulation and Computation 30: 377-390.
38. Hutson AD (2002) A Semiparametric Quantile Function Estimator for use in Bootstrap Estimation Procedures. Statistics and Computing 12: 331-338.

39. Srivastava DK, Mudholkar GS, Mudholkar A (1992) Assessing the significance of difference between two quick estimates of location. Journal of Applied Statistics 19: 405-416.

40. Vanzwet WR (1964) Convex Transformation of Random Variables. Mathematical Centre Tracts, Amsterdam.

41. Wang D, Hutson AD (2011) A fractional order statistic towards defining a smooth quantile function for discrete data. J Statist Plann Inference 141: 3142-3150.