isid/ms/2012/11
August, 8, 2012
http://www.isid.ac.in/~statmath/eprints

Nonparametric estimation of quantile density function

Pooja Soni
Isha Dewan
Kanchan Jain

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

Nonparametric estimation of quantile density function

Pooja Soni[1], Isha Dewan[2] and Kanchan Jain[1]
[1]Department of Statistics, Panjab University
Chandigarh- 160014 (India)
soni06pooja@gmail.com and jaink14@gmail.com
[2] Stat-Math Unit
Indian Statistical Institute
New Delhi-110016 (India)
ishadewan@gmail.com

Abstract

In the present article, a new nonparametric estimator of quantile density function is defined and its asymptotic properties are studied. The comparison of the proposed estimator has been made with estimators given by Jones (1992), graphically and in terms of mean square errors for the uncensored and censored case.

Keywords: *Quantiles, quantile density function, kernel density estimators.*

1 Introduction

In classical statistics, most of distributions are defined in terms of their cumulative distribution function (cdf) or probability density function (pdf). There are some distributions which do not have the cdf/pdf in an explicit form but a closed form of the quantile function is available, for example Generalised Lambda distribution (GLD) and Skew logistic distribution (Gilchrist (2000)). Karian and Dudewicz (2000) showed the significance of different Lambda distributions for modelling failure time data. Quantile measures are less influenced by extreme observations. Hence the quantile function can also be looked upon as an alternative to the distribution function in lifetime data for heavy tailed distributions. Sometimes for those distributions whose reliability measures do not have a closed or explicit form, the

reliability characteristics can be represented through quantile function.
The quantile function approach is a useful tool in statistical analysis. It has been used in exploratory data analysis, applied statistics, reliability and survival analysis (See, for example, Reid (1981), Slud et al. (1984), Su and Wei (1993), Nair et al. (2008), Nair and Sankaran (2009) and Sankaran and Nair (2009)). For a unified study of this concept, one can refer to Parzen (1979), Jones (1992), Friemer et al. (1998), Gilchrist (2000) and Nair and Sankaran (2009). The concept of quantiles has been used by Peng and Fine (2007), Jeong and Fine (2009) and Sankaran et al. (2010) for modelling competing risk models.
Let X be a non-negative continuous random variable that represents the life time of a unit with cdf $F(x)$, survival function $S(x)$, the density function $f(x)$ and failure rate function $h(x) = \frac{f(x)}{S(x)}$. If X is censored by a non-negative random variable C, then we observe $T = min(X, C)$ and $\delta = I(X \leq C)$ where $I(.)$ is an indicator function.
A quantile is simply the value that corresponds to a specified proportion of sample or population (Gilchrist (2000)). Mathematically, it is defined as

$$\begin{aligned} &Q(u) = F^{-1}(u) = inf\{x : F(x) \geq u\},\ 0 \leq u \leq 1 \\ &\Longrightarrow\ F(Q(u)) = u. \end{aligned} \tag{1.1}$$

Parzen (1979) and Jones (1992) defined the quantile density function as the derivative of $Q(u)$, that is, $q(u) = Q'(u)$. Note that the sum of two quantile density functions is again a quantile density function.
Differentiating (1.1), we get

$$q(u) = \frac{1}{f(Q(u))}. \tag{1.2}$$

Nair and Sankaran (2009) defined the hazard quantile function as follows:
$H(u) = h(Q(u)) = \dfrac{f(Q(u))}{S(Q(u))} = ((1-u)q(u))^{-1}.$
Thus hazard rate of two populations would be equal if and only if their corresponding quantile density functions are equal. This has been used to construct tests for testing equality of failure rates of two independent samples. The results are being reported elsewhere.
We propose a kernel type quantile density estimator. The kernel $K(.)$ is a real valued function satisfying the following properties:

(i) $K(u) \geq 0$ for all u;

(ii) $\int_{-\infty}^{\infty} K(u)du = 1$;

(iii) $K(.)$ has finite support, that is $K(u) = 0$ for $|u| > c$ where $c > 0$ is some constant;

(iv) $K(.)$ is symmetric about zero;

(v) $K(.)$ satisfies Lipschitz condition, viz there exists a positive constant M such that $|K(u) - K(v)| \leq M|u - v|$.

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables from $F(x)$. Jones (1992) proposed two smooth estimators of the quantile density function. The first estimator was given as

$$q_n^{j1}(u) = \frac{1}{f_n(Q_n(u))} \tag{1.3}$$

where $f_n(x)$ is a kernel density estimator of $f(x)$ and $Q_n(u)$ is the empirical estimator of quantile function $Q(u)$.
Note that $f_n(x) = \frac{1}{nh(n)} \Sigma_{i=1}^n K\Big(\frac{x - X_i}{h(n)}\Big)$, where $h(n)$ is the bandwidth and

$Q_n(u) = inf\{x : F_n(x) \geq u\}, 0 \leq u \leq 1$.
Another estimator of quantile density function given by Jones (1992) is

$$\begin{aligned} q_n^{j2}(u) &= \Sigma_{i=2}^n X_{(i)}\Big(K_{h(n)}(u - \frac{i-1}{n}) - K_{h(n)}(u - \frac{i}{n})\Big) \\ &= \Sigma_{i=2}^n (X_{(i)} - X_{(i-1)})K_{h(n)}(u - \frac{i-1}{n}) - X_{(n)}K_{h(n)}(u-1) + X_{(1)}K_{h(n)}(u) \end{aligned} \tag{1.4}$$

where $X_{(i)}$ is the i^{th}order statistic, i=1,2,...,n.
In Section 2, we propose a smooth estimator of the quantile density function and derive its asymptotic properties. In Section 3, three estimators of $q(u)$ are compared graphically. In Section 4, simulations have been carried out for comparing the two estimators given by Jones (1992) and the estimator proposed by us.

2 Estimation of quantile density function

Based on data $X_1, X_2, \ldots, X_n$, we propose a smooth estimator of the quantile density function as

$$q_n(u) = \frac{1}{h(n)} \int_0^1 \frac{K(\frac{t-u}{h(n)})}{f_n(Q_n(t))} dt \tag{2.1}$$

where $K(.)$ is a kernel and $h(n)$ is the bandwidth sequence.

(2.1) can also be written as

$$q_n(u) = \frac{1}{h(n)} \sum_{i=1}^{n} \frac{1}{f_n(X_{(i)})} \int_{S_{i-1}}^{S_i} K\Big(\frac{t-u}{h(n)}\Big) dt$$

where S_i is the proportion of observations less than or equal to $X_{(i)}$, the i^{th} order statistic.

In uncensored case, for small $S_i - S_{i-1}$, we use the mean value theorem to get $q_n(u) = \frac{1}{nh(n)} \Sigma_{i=1}^{n} \frac{K(\frac{S_i-u}{h(n)})}{f_n(X_{(i)})}$.

In case of censoring, we observe $T_i = min(X_i, C_i)$, where C_i is the censoring variable. For $\delta_i = I(X_i \leq C_i)$, when data is of the form $(T_i, \delta_i), i = 1, 2..., n$, the estimator of quantile density function is given by

$$q_n^c(u) = \frac{1}{h(n)} \Sigma_{i=1}^{n} \frac{1}{f_n(T_{(i)})} \int_{S_{i-1}}^{S_i} K\Big(\frac{t-u}{h(n)}\Big) dt \tag{2.2}$$

where

$$S_i = \begin{cases} 0 & i = 0, \\ F_n(T_{(i)}) & i = 1, 2, ..., n-1, \\ 1 & i = n, \end{cases}$$

and $T_{(i)}$ is the i^{th} order statistic. In the presence of censoring, the proposed quantile density estimator takes the form

$$q_n^c(u) = \frac{1}{h(n)} \Sigma_{i=1}^{n} \frac{(S_i - S_{i-1}) K(\frac{S_i-u}{h(n)})}{f_n(T_{(i)})}.$$

The following theorem proves a result that shall be used in the sequel.

Theorem 2.1. *Let $q(u)$ be the quantile density function corresponding to a density function $f(x)$ and $q_n^{j1}(u)$ denote the estimator of $q(u)$ given by Jones (1992). Then $\sup_u |q_n^{j1}(u) - q(u)| \to 0$ as $n \to \infty$.*

Proof. We consider

$$\begin{aligned} q_n^{j1}(u) &= \frac{1}{f_n(Q_n(u))} \\ &= \frac{1}{f_n(Q_n(u)) - f(Q(u)) + f(Q(u))} \\ &= \frac{1}{f(Q(u))} \Big[\frac{1}{1 + \frac{f_n(Q_n(u)) - f(Q(u))}{f(Q(u))}} \Big]. \end{aligned}$$

Using Binomial theorem,

$$q_n^{j1}(u) = \frac{1}{f(Q(u))}\Big[1 - \frac{f_n(Q_n(u)) - f(Q(u))}{f(Q(u))} + \Big(\frac{(f_n(Q_n(u)) - f(Q(u)))^2}{f^2(Q(u))}\Big) - \dots\Big] \tag{2.3}$$

Hence

$$q_n^{j1}(u) - q(u) = \frac{-f_n(Q_n(u)) + f(Q(u))}{f^2(Q(u))} + \Big(\frac{f_n(Q_n(u)) - f(Q(u)))^2}{f^3(Q(u)}\Big) - \dots$$

Writing Taylor series expansion of $f_n(Q_n(u))$ about $Q(u)$, we have

$$f_n(Q_n(u)) = f_n(Q(u)) + (Q_n(u) - Q(u))f_n^{'}(Q(u)) + \frac{(Q_n(u) - Q(u))^2 f_n^{''}(Q(u))}{2!} + \dots,$$

assuming higher derivatives of f_n exist.
Hence $f_n(Q_n(u)) - f(Q(u)) =$

$$f_n(Q(u)) - f(Q(u)) + (Q_n(u) - Q(u))f_n^{'}(Q(u)) + \frac{(Q_n(u) - Q(u))^2 f_n^{''}(Q(u))}{2!} + \dots$$

For $n \to \infty$, $sup_t|Q_n(t) - Q(t)| \to 0$ (Serfling (1980)) and $sup_t|f_n(t) - f(t)| \to 0$, (Prakasa Rao (1983)) which implies that $f_n(Q_n(u)) - f(Q(u)) \to 0$.
Hence $sup_u|q_n^{j1}(u) - q(u)| \to 0$ as $n \to \infty$. □

The next thoerem proves consistency of the proposed estimator of the quantile density function.

Theorem 2.2. *For large n, $sup_u|q_n(u) - q(u)| \to 0$ where $q_n(u)$ given by (2.1) is the proposed estimator of $q(u)$, the quantile density function.*

Proof. (2.1) gives the estimator of the quantile density function $q(u)$ as

$$q_n(u) = \frac{1}{h(n)} \int_0^1 \frac{K(\frac{t-u}{h(n)})}{f_n(Q_n(t))} dt.$$

Hence

$$q_n(u) - q(u) = \frac{1}{h(n)} \int_0^1 \frac{K(\frac{t-u}{h(n)})}{f_n(Q_n(t))} dt - q(u)$$

$$= \frac{1}{h(n)} \int_0^1 \frac{K(\frac{t-u}{h(n)})}{f_n(Q_n(t))} dt - \frac{1}{h(n)} \int_0^1 \frac{K(\frac{t-u}{h(n)})}{f(Q(t))} dt + \frac{1}{h(n)} \int_0^1 \frac{K(\frac{t-u}{h(n)})}{f(Q(t))} dt - \frac{1}{f(Q(u))}$$

$$= \frac{1}{h(n)} \int_0^1 K\Big(\frac{t-u}{h(n)}\Big)\Big[\frac{1}{f_n(Q_n(t))} - \frac{1}{f(Q(t))}\Big] dt + \frac{1}{h(n)} \int_0^1 \frac{K(\frac{t-u}{h(n)})}{f(Q(t))} dt - \frac{1}{f(Q(u))}$$

$$=\frac{1}{h(n)}\int_0^1 K\Big(\frac{t-u}{h(n)}\Big)\Big[\frac{f(Q(t))-f_n(Q_n(t))}{f_n(Q_n(t))f(Q(t))}\Big]dt+\frac{1}{h(n)}\int_0^1\frac{K(\frac{t-u}{h(n)})}{f(Q(t))}dt-\frac{1}{f(Q(u))}\ .$$

Using Theorem 2.1, $sup_t|q_n^{j1}(t)-q(t)|\to 0$ as $n\to\infty$. Hence the above expression asymptotically reduces to

$$\frac{1}{h(n)}\int_0^1 K\Big(\frac{t-u}{h(n)}\Big)[q(t)]^2[f(Q(t))-f_n(Q_n(t))]dt+\frac{1}{h(n)}\int_0^1\frac{K(\frac{t-u}{h(n)})}{f(Q(t))}dt-\frac{1}{f(Q(u))}$$

$$=\frac{1}{h(n)}\int_0^1 K\Big(\frac{t-u}{h(n)}\Big)q(t)[f(Q(t))q(t)-f_n(Q_n(t))q_n(t)]dt+\frac{1}{h(n)}\int_0^1\frac{K(\frac{t-u}{h(n)})}{f(Q(t))}dt-\frac{1}{f(Q(u))}.$$

Since $dF(Q(t))=f(Q(t))q(t)dt$, hence
$q_n(u)-q(u)$

$$=\frac{1}{h(n)}\int_0^1 K\Big(\frac{t-u}{h(n)}\Big)q(t)[dF(Q(t))-dF_n(Q_n(t))]+\frac{1}{h(n)}\int_0^1\frac{K(\frac{t-u}{h(n)})}{f(Q(t))}dt-\frac{1}{f(Q(u))}.$$

Writing $K^*(u,t)=K(\frac{t-u}{h(n)})q(t)$ and integrating by parts in the first integral, we get
$q_n(u)-q(u)$
$=\frac{1}{h(n)}(K^*(u,t))[F(Q(t))-F_n(Q_n(t))]|_0^1-\frac{1}{h(n)}\int_0^1 dK^*(u,t)[F(Q(t))-F_n(Q_n(t))]$
$$+\frac{1}{h(n)}\int_0^1\frac{K(\frac{t-u}{h(n)})}{f(Q(t))}dt-\frac{1}{f(Q(u))}.$$
Since $F(Q(0))=F_n(Q(0))$ and $F(Q(1))=F_n(Q_n(1))$, hence the above expression transforms to
$\frac{1}{h(n)}\int_0^1 dK^*(u,t)[F_n(Q_n(t))-F(Q(t))]+\dfrac{1}{h(n)}\displaystyle\int_0^1\frac{K(\frac{t-u}{h(n)})}{f(Q(t))}dt-\frac{1}{f(Q(u))}.$

Putting $\dfrac{t-u}{h(n)}=v$ and using (1.2),

$$\frac{1}{h(n)}\int_0^1\frac{K(\frac{t-u}{h(n)})}{f(Q(t))}dt-\frac{1}{f(Q(u))}=\int_{\frac{-u}{h(n)}}^{\frac{1-u}{h(n)}}K(v)q(u+vh(n))dv-q(u)\qquad(2.4)$$

Using Taylor series expansion, we can write $q(u+vh(n))=q(u)+vh(n)q'(u)+...$, assuming higher derivatives of $q(u)$ exist and are bounded.
Hence (2.4) can be written as

$$\int_{\frac{-u}{h(n)}}^{\frac{1-u}{h(n)}}K(v)(q(u)+vh(n)q'(u)+...)dv-q(u).\qquad(2.5)$$

For $n \to \infty$, $h(n) \to 0$ and hence (2.5) converges to
$\int_{-\infty}^{\infty} K(v)q(u)dv - q(u)$ which equals zero as $\int_{-\infty}^{\infty} K(v)dv$=1.
This gives

$$q_n(u) - q(u) = \frac{1}{h(n)} \int_0^1 dK^*(u,t)[F_n(Q_n(t)) - F(Q(t))]dt. \tag{2.6}$$

Since $\sup_t |F_n(Q_n(t))F(Q(t))| \to 0$ as $n \to \infty$, hence $sup_u|q_n(u) - q(u)| \to 0$.

□

The following theorem proves asymptotic normality of the proposed estimator.

Theorem 2.3. $\sqrt{n}(q_n(u) - q(u))$ *is asymptotically normal with mean zero and variance* $\sigma^2(u) = \frac{n}{(h(n))^2} E(\int_0^1 dK^*(u,t)F_n(Q_n(t)))^2$.

Proof. Using (2.6), we have
$\sqrt{n}(q_n(u) - q(u)) = \frac{\sqrt{n}}{h(n)} \int_0^1 dK^*(u,t)[F_n(Q_n(t)) - F(Q(t))]dt.$
Using the result of Anderson et al. (1993), for $0 < u < 1$, $\sqrt{n}[Q_n(u) - Q(u)]$ is asymptotically normal with mean 0 and variance $\sigma_1^2(u) = (S(u))^2 \displaystyle\int_0^u \frac{-dS(t)}{S(t)S^*(u)}$
where $S^*(u)$ is the probability that a unit is alive and uncensored at time t.

Since $\frac{d}{du}F(Q(u)) = 1$, $\sqrt{n}[F_n(Q_n(u)) - F(Q(u))]$ is asymptotically normal with mean zero and variance $\sigma_1^2(u)$.
Using Delta method and Slustky's theorem (Serfling (1980)), we get that
$\sqrt{n}(q_n(u) - q(u))$ is asymptotically normal with mean zero and variance

$\sigma^2(u) = \frac{n}{(h(n))^2} E(\int_0^1 dK^*(u,t)F_n(Q_n(t)))^2.$ □

The expression of $\sigma^2(u)$ in the above theorem can't be simplified analytically and one can estimate it using bootstrapping.

3 Examples of Quantile Density Estimators

In this section, we consider two distributions Exp(1) and Generalised Lambda $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. The corresponding quantile functions are $Q_E(u) = -log(1-u)$ and $Q_{GL}(u) = \lambda_1 + \frac{(u^{\lambda_3} - (1-u)^{\lambda_4})}{\lambda_2}$, where λ_1 and λ_2 are location and inverse scale parameters, respectively and λ_3 and λ_4 jointly determine the shape (with λ_3 mostly affecting the left tail and λ_4 mostly affecting the right tail). Though the GLD is defined on the entire real line, we consider those choices of parameters that give support as $(0, \infty)$. For Exp(1), the quantile density function is

$q_E(u) = \frac{1}{1-u}, 0 < u < 1$ and for GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, the quantile density function

is $q_{GL}(u) = \dfrac{(\lambda_3 u^{\lambda_3 - 1} + \lambda_4(1-u)^{\lambda_4 - 1})}{\lambda_2}, 0 < u < 1.$

For finding the estimators, we consider two different types of kernels

(i) Triangular: $K(u) = (1 - |u|)I(|u| \leq 1)$ and

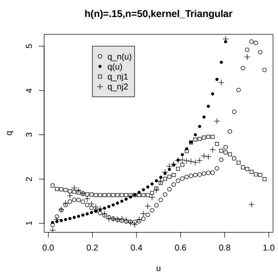
(ii) Epanechnikov: $K(u) = .75(1 - u^2)I(|u| \leq 1)$.

Triangular kernel was used by Nair and Sankaran (2009) for studying non-parametric estimators of the hazard quantile function and Epanechnikov kernel gives the optimal kernel Prakasa Rao (1983). We are reporting the results for $h(n) = .15$. Similar results were found for $h(n)$=.19 and .25 but are not being reported for brevity.

The figures (1.1)-(1.4) show the original quantile density function $q(u)$ and three estimators $q_n(u)$, $q_n^{j1}(u)$, $q_n^{j2}(u)$ in uncensored case when the observations are from Exp(1) and Triangular and Epanechikov kernels are used.

Fig.1.1

Fig.1.2



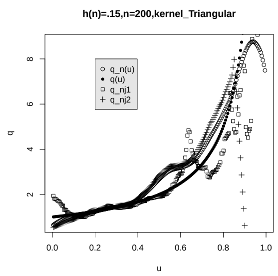
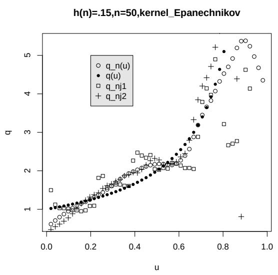
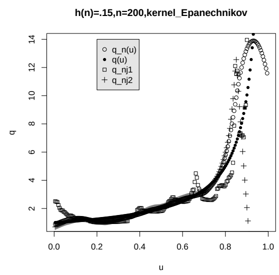


Fig.1.3

Fig.1.4





The figures (2.1)-(2.4) show the original quantile density function $q(u)$ and three estimators $q_n^c(u)$, $q_n^{j1}(u)$, $q_n^{j2}(u)$ in censored case when observations follow Exp(1) and censoring distribution is assumed to be Exp(.25) to ensure 20 percent censoring.

Fig.2.1

Fig.2.2

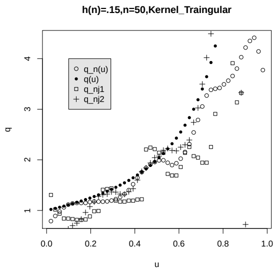


Fig.2.3

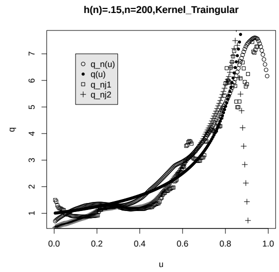
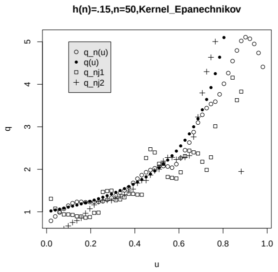
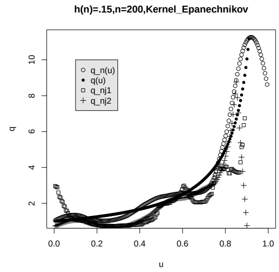


Fig.2.4





The figures (3.1)-(3.4) show the original quantile density function $q(u)$ and three estimators $q_n(u)$, $q_n^{j1}(u)$, $q_n^{j2}(u)$ in uncensored case when observations are from Generalised Lambda distribution with parameters $\lambda_1 = 1, \lambda_2 = -1, \lambda_3 = -1/8, \lambda_4 = -1/8$ for both kernels.

Fig.3.1

Fig.3.2

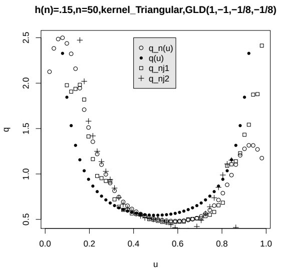

h(n)=.15,n=50,kernel_Triangular,GLD(1,−1,−1/8,−1/8)
q_n(u)
q(u)
q_nj1
q_nj2
q
u

Fig.3.3

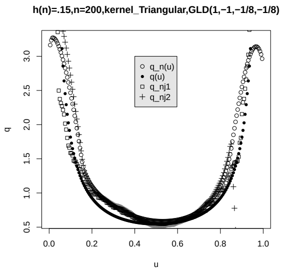
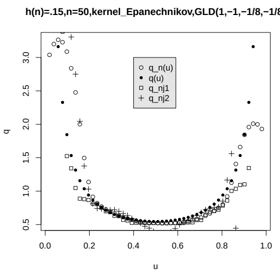
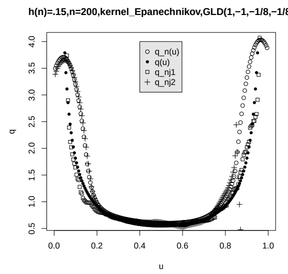

h(n)=.15,n=200,kernel_Triangular,GLD(1,−1,−1/8,−1/8)
q_n(u)
q(u)
q_nj1
q_nj2
q
u

Fig.3.4





The figures (4.1)-(4.4) display $q(u)$ and $q_n^c(u)$, $q_n^{j1}(u)$, $q_n^{j2}(u)$ in censored case, when the observations follow GLD(1,-1,-1/8,-1/8) and censoring distribution is assumed to be Uniform(0,4.1) so that 20 percent censoring is ensured.

Fig.4.1

Fig.4.2

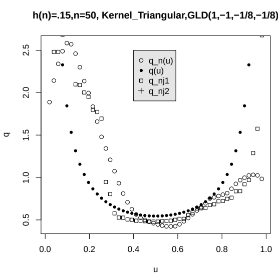


Fig.4.3

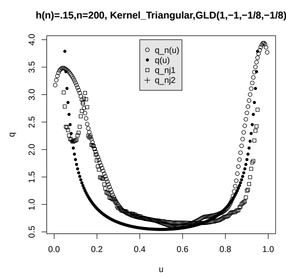
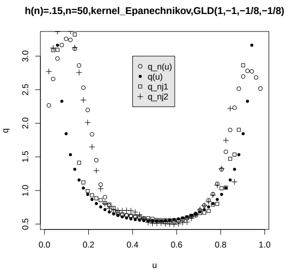
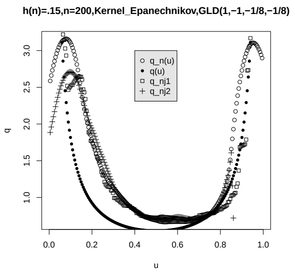


Fig.4.4


h(n)=.15,n=50,kernel_Epanechnikov,GLD(1,−1,−1/8,−1/8)
q_n(u)
q(u)
q_nj1
q_nj2
q
u


h(n)=.15,n=200,Kernel_Epanechnikov,GLD(1,−1,−1/8,−1/8)
q_n(u)
q(u)
q_nj1
q_nj2
q
u

From all the figures, we conclude that

(i) our newly proposed estimator is closer to the unknown quantile density function as compared to both estimators $q_n^{j1}(u)$ and $q_n^{j2}(u)$ proposed by Jones (1992),

(ii) the choice of kernel is immaterial. Both the kernels give similar looking graphs of estimators for both exponential and GLD distributions,

(iii) Jones (1992) gave only theoretical study of the estimators. These graphs show that $q_n^{j1}(u)$ is closer to the unknown $q(u)$ as compared to $q_n^{j2}(u)$,

(iv) our estimator and those given by Jones are away from the true quantile density function for large values of u. This means that the estimators need to be adjusted at the tails. $q_n^{j2}(u)$ is observed to be the worst performer. Some techniques used for correction in tails are mentioned in the concluding section.

(v) the estimators get closer to the true quantile density function as sample size increases,

(vi) for Generalised Lambda distribution, the three estimators are relatively close. There are problems for small values of u,

(vii) even for the censored case, the estimators are closer to the unknown quantile density function for n=200.

4 Simulation Results

In this section, we carry out simulations for comparing mean square error (MSEs) of our proposed estimator with those of $q_n^{j1}(u)$ and $q_n^{j2}(u)$ proposed by Jones (1992). We calculate the MSEs of these estimators by using bootstrap sampling technique with 5000 bootstrap samples. The chosen bandwidths are .15, .19 and .25 and sample size is 200. The Triangular and Epanechnikov kernels are used. The results are found using R package. For uncensored case, the underlying distribution is assumed to be Exp(1). For GLD, the choice of parameteric values is as given in Section 3. This has been done in order to get positive values of the generated random variable. The validity of the chosen parameters has been checked using the package fbasics(gld) in R.

Table 3.1 gives the values of MSEs for Exp(1) distribution for n= 200. The values in the brackets give MSEs for Epanechnikov kernel and the other values are for Triangular kernel.

Table 3.1 MSEs for Triangular (Epanechnikov) kernel, n=200, Exp(1)

		h(n)		
u	estimate	.15	.19	.25
.2	$q_n(u)$	0.0753(0.0329)	0.0222(0.1828)	0.0226(0.0882)
	$q_n^{j1}(u)$	0.1438(0.0341)	0.0453(0.1792)	0.0199(0.0254)
	$q_n^{j2}(u)$	0.1221(0.0464)	0.0392(0.1514)	0.0234(0.0536)
.4	$q_n(u)$	0.1219(0.0756)	0.0562(0.0672)	0.0664(0.1730)
	$q_n^{j1}(u)$	0.1633(0.1410)	0.0864(0.1367)	0.1537(0.5008)
	$q_n^{j2}(u)$	0.1269(0.1099)	0.0558(0.0806)	0.1126(0.2346)
.6	$q_n(u)$	0.1848(0.0968)	0.1458(0.1123)	0.0868(0.5317)
	$q_n^{j1}(u)$	0.6358(0.3375)	0.6659(0.3067)	0.7477(0.5177)
	$q_n^{j2}(u)$	0.3395(0.1179)	0.2901(0.1155)	0.1989(0.6395)
.8	$q_n(u)$	0.5504(0.4865)	0.2872(0.3592)	0.2520(4.7515)
	$q_n^{j1}(u)$	4.6744(4.2817)	3.2896(2.2621)	4.5683(1.3512)
	$q_n^{j2}(u)$	1.9648(1.5185)	1.0768(0.6908)	3.8116(3.5122)
.9	$q_n(u)$	2.0676(1.4213)	7.2984(5.7775)	8.4150(2.4611)
	$q_n^{j1}(u)$	28.8638(33.3130)	27.4051(24.6925)	11.9094(37.637)
	$q_n^{j2}(u)$	36.4881(54.2897)	98.48(128.592)	116.281 (149.3)

The following table is for GLD ($\lambda_1 = 1, \lambda_2 = -1, \lambda_3 = -1/8, \lambda_4 = -1/8$).

Table 3.2 MSEs for Triangular (Epanechnikov) kernel and n=200, GLD(1,-1,-1/8,-1/8)

		h(n)		
u	estimate	.15	.19	.25
.2	$q_n(u)$	0.0532(0.0586)	0.0242(0.0237)	0.0304(1.2604)
	$q_n^{j1}(u)$	0.1516(0.0960)	0.0374(0.1126)	0.1560(0.2987)
	$q_n^{j2}(u)$	0.0522(0.0650)	0.0326(0.0335)	0.0736(0.1510)
.4	$q_n(u)$	1.2049(1.0868)	1.1482(1.0886)	0.9185(1.3451)
	$q_n^{j1}(u)$	1.2417(1.1382)	0.4704(1.1483)	0.9710(0.1499)
	$q_n^{j2}(u)$	1.2447(1.1835)	1.2701(1.2434)	1.0613(0.1942)
.6	$q_n(u)$	3.7274(3.5064)	3.6923(3.4799)	3.0358(0.0157)
	$q_n^{j1}(u)$	3.7833(3.5871)	1.4812(3.5639)	3.1598(0.2219)
	$q_n^{j2}(u)$	3.8205(3.5324)	3.8532(3.6032)	3.1089(0.0314)
.8	$q_n(u)$	16.6397(15.5172)	15.9542(14.694)	10.0711(14.694)
	$q_n^{j1}(u)$	17.0570(16.2006)	6.7417(16.22454)	15.3410(8.6574)
	$q_n^{j2}(u)$	16.0456(15.0820)	15.1303(13.3583)	21.4279(23.6571)
.9	$q_n(u)$	59.1509(56.6860)	63.2191(50.0144)	52.1443(61.9908)
	$q_n^{j1}(u)$	74.0821(69.3978)	28.541(71.904)	61.213 (63.469)
	$q_n^{j2}(u)$	131.776(155.959)	148.514(164.307)	144.773(178.68)

When the observations follow Exp(1) distribution, we take the censoring distribution as Exp(.25). This ensures that 20 percent of the data is censored. When the observations follow Generalised Lambda with parameters $\lambda_1 = 1, \lambda_2 = -1, \lambda_3 = -1/8, \lambda_4 = -1/8$, the censoring distribution Uniform(0,4.1) ensures 20 percent censoring. Tables 3.3 and 3.4 report the MSEs of $q_n^c(u), q_n^{j1}(u)$ and $q_n^{j2}(u)$ for Exp(1) and GLD distribution respectively

Table 3.3 MSEs for Triangular (Epanechnikov) kernel, n=200

		h(n)		
u	estimate	.15	.19	.25
.2	$q_n^c(u)$	0.0060(0.1244)	0.00003(0.011)	0.0006(0.0269)
	$q_n^{j1}(u)$	0.0584(0.3152)	0.0810(0.1540)	0.0435(0.0295)
	$q_n^{j2}(u)$	0.0480(0.3451)	0.0763(0.1644)	0.0668(0.0518)
.4	$q_n^c(u)$	0.0015(0.0354)	0.0622(0.0889)	0.0067(0.0010)
	$q_n^{j1}(u)$	0.1145(0.2072)	0.3550(0.3111)	0.2186(0.3305)
	$q_n^{j2}(u)$	0.0661(0.0570)	0.3215(0.2633)	0.1527(0.2532)
.6	$q_n^c(u)$	0.0002(0.1095)	0.1759(0.0037)	0.0023(0.3975)
	$q_n^{j1}(u)$	0.7925(0.5516)	0.7615(0.8331)	0.4849(0.5872)
	$q_n^{j2}(u)$	0.2228(0.4045)	0.5524(0.3999)	0.1230(0.1986)
.8	$q_n^c(u)$	0.2098(0.2511)	0.0004(0.0010)	0.8528(3.8652)
	$q_n^{j1}(u)$	3.4570(3.9458)	5.9596(3.1700)	2.6181(2.7327)
	$q_n^{j2}(u)$	0.6548(1.3919)	0.9848(1.0021)	4.7585(3.8531)
.9	$q_n^c(u)$	7.5869(14.1619)	8.3516(9.9614)	1.9265(6.1782)
	$q_n^{j1}(u)$	28.3032(23.3419)	23.2899(19.6989)	28.4622(18.481)
	$q_n^{j2}(u)$	56.2880(71.9684)	99.5988(98.0264)	119.118(112.89)

Table 3.4 MSEs for Triangular (Epanechnikov) kernel, n=200

		h(n)		
u	estimate	.15	.19	.25
.2	$q_n^c(u)$	0.16186(0.2040)	0.3569(0.3620)	0.8398(1.3901)
	$q_n^{j1}(u)$	0.1950(0.2523)	0.1803(0.2055)	0.1644(0.1669)
	$q_n^{j2}(u)$	0.4579(0.3880)	0.5917(0.4799)	0.6309(0.6873)
.4	$q_n^c(u)$	0.7828(0.4690)	0.7298(0.4943)	0.6638(0.6296)
	$q_n^{j1}(u)$	0.6959(0.6397)	0.6868(0.6491)	0.6632(0.6466)
	$q_n^{j2}(u)$	0.6217(0.5623)	0.6145(0.5485)	0.5711(0.5511)
.6	$q_n^c(u)$	3.3586(3.5921)	3.2812(3.3630)	3.1353(3.0568)
	$q_n^{j1}(u)$	3.2500(3.5421)	3.2020(3.4300)	3.1026(3.0473)
	$q_n^{j2}(u)$	3.2002(3.5994)	3.2214(3.5361)	3.1978(3.1878)
.8	$q_n^c(u)$	16.875(17.0534)	15.2707(16.2932)	11.2619(9.0418)
	$q_n^{j1}(u)$	17.5286(18.3496)	17.2305(18.0225)	16.9206(16.6974)
	$q_n^{j2}(u)$	17.0753(17.7818)	15.4134(16.6909)	25.4494(29.2862)
.9	$q_n^c(u)$	50.9776(70.1115)	49.8754(73.8416)	52.3556(48.7843)
	$q_n^{j1}(u)$	75.8818(78.2903)	75.9366(79.1654)	75.998(75.955)
	$q_n^{j2}(u)$	131.485(156.797)	148.0253(164.663)	151.405(159.924)

From the tables, we conclude that

(i) in majority of the cases, the MSEs for $q_n(u)$ and $q_n^c(u)$ are smaller than those of $q_n^{j1}(u)$ and $q_n^{j2}(u)$,

(ii) the MSEs increase with an increase in u for all cases,

(iii) the MSEs decrease as the sample size increases,

(iv) the MSEs for the censored case are more than those in uncensored case,

(v) none of the two kernels give a uniformly better result than the other. For some cases, the MSE is smaller for the Triangular kernel and for others, it is smaller for Epanechikov kernel,

(vi) for u in the tails (u=0.9), the MSEs for $q_n^{j2}(u)$ are much higher than those of $q_n^{j1}(u)$, $q_n(u)$ and $q_n^c(u)$,

(vii) the estimator is not performing well in the tails.

5 Conclusions

This article proposes a smooth estimator of the quantile density function under censored and uncensored models. The proposed estimator is consistent and asymptotically follows normal distribution. The estimator is compared with those given by Jones (1992) via graphs and MSEs. For most of the cases, the proposed estimator gives smaller MSEs in censored and uncensored case as compared to Jones (1992) estimators, when the underlying distributions are Exponential and Generalised Lambda. The figures show that our estimators $q_n(u)$ and $q_n^c(u)$ are closer to the unknown quantile density function $q(u)$.

The estimator of the quantile density function proposed here is not good at the tails. Several solutions for correction at tails have already been proposed in literature. Some of these are reflection method proposed by Silverman (1986) and Cline and Hart (1991), the boundary kernel method by Gasser, Miller and Mammitzsch (1985), Jones (1993), Zhang and Karunamuni (2000), the transformation method by Marron and Ruppert (1994), the local linear method by Loader (1996), Hjort and Jones (1996) and Wand and Jones (1998). We propose to work on modification of our estimator at the tails in future.

Parzen (2004) defined the conditional quantile as

$Q_{Y/X}(u) = inf\{x : F_{Y/X}(y) \geq u\},\ 0 \leq u \leq 1$, where $F_{Y/X}(y)$ is the conditional distribution function corresponding to the random vector (X, Y). For functional data, Ferraty and Vieu (2006) have discussed the estimation of conditional quantile function. The definition of conditional quantile implies that $F_{Y/X}(Q_{Y/X}(u)) = u$. On differentiating partially w.r.t u, we get
$f_{Y/X}(Q_{Y/X}(u))\frac{\partial}{\partial u}(Q_{Y/X}(u)) = 1$.
Hence the conditional quantile density function can be written as

$$\begin{aligned} q_{Y/X}(u) &= \frac{\partial}{\partial u}(Q_{Y/X}(u)) \qquad \text{(Ref. Xiang (1995))} \\ &= \frac{1}{f_{Y/X}(Q_{Y/X}(u))}. \end{aligned}$$

The estimation of the conditional quantile density function can be done on similar lines as for $q(u)$, the quantile density function.

Acknowledgement

Research work of the first author is supported by Council of Scientific and Industrial Research, Government of India. We thank the referees for their constructive suggestions. Their ideas will lead to more work in quantile density estimation.

References

Anderson, P.K., Borgan, O., Gill, R.D. and Keiding N., 1993. *Statistical methods based on counting processes.* Springer, New York.

Cline D.B.H. and Hart J.D., 1991. Kernel estimation of densities of discontinuous derivative. Statistics, **22**, 69-84.

Ferraty, F. and Vieu, P., 2006. *Nonparametric functional data analysis.* Springer , New York.

Friemer, M, Mudholkar, G.S., Kollia, G.D. and Lin, T.C., 1988. A study of Generalized Lambda family. Communications in Statistics Theory and Methods, **17**, 3547-3567.

Gasser, T., Muller, H.G. and Mammitzsch V., 1985. Kernels for nonparametric curve estimation. Journal of Royal Statistical Society, Series B, **47**, 238-256.

Gilchrist,W., 2000. *Statistical modelling with quantile functions.* Chapman and Hall, New York.

Hjort N. L. and Jones M. C. 1996. Locally parametric nonparametric density estimation. The Annals of Statistics, **24**, 1619-1647.

Jeong, J.H. and Fine, J.P., 2009. Parametric regression on cumulative incidence function. Biostatistics, **8**, 184-196.

Jones, M.C., 1992. Estimating densities, quantiles, quantile densities and density quantiles. Ann. Inst. Statist. Math., **44**, 721-727.

Jones M.C., 1993. Simple boundary correction for kernel density estimation. Statistics and Computing, **3**, 135-146.

Karian, A. and Dudewicz, E. J., 2000. *Fitting statistical distributions* . CRC Press. London.

Loader C. R. 1996. Local likelihood density estimation. The Annals of Statistics, **24**, 1602-1618.

Marron J.S. and Ruppert D., 1994. Transformations to reduce boundary bias in kernel density estimation. Journal of Royal Statistical Society, Series B **56**, 653-671.

Nair, N.U., Sankaran, P.G. and Kumar, B.V, 2008. Total time on test transforms of order n and its implications in reliability analysis. Journal of Applied Probability, **45**, 1126-1139.

Nair, N.U. and Sankaran, P.G., 2009. Quantile based reliability analysis. Communications in Statistics Theory and Methods, **38**, 222-232.

Parzen, E., 1979. Nonparametric statistical data modelling. Journal of American Statistical Association, **74**, 105-122.

Parzen, E., 2004. Quantile probability and statistical data modeling. Statistical Science, **19**, 652-662.

Peng, L. and Fine, J.P., 2007. Nonparametric quantile inference with competing risks data. Biometrika, **94**, 735-744.

Prakasa Rao, B.L.S., 1983. *Nonparametric functional estimation.* Academic Press, New York.

Reid, N., 1981. Estimating the median survival time. Biometrika, **68**, 601-608.

Sankaran, P.G. and Nair N.U., 2009. Nonparametric estimation of hazard quantile function. Journal of Nonparametric Statistics, **21**, 757-767.

Sankaran, P.G., Nair, N.U. and Sreedevi E.P., 2010. A quantile based test for comparing cumulative incidence functions of competing risks models. Statistics and Probability Letters, **80**, 886-891.

Serfling R.J., 1980. *Approximation theorem of mathematical statistics.* John Wiley, New York.

Silverman B.W., 1986. *Density estimation for statistics and data analysis.* Chapman and Hall, London.

Slud, E.V., Byar, D.P. and Green, S.B., 1984. A comparison of reflected versus

testbased confidence intervals for the median survival time, based on censored data. Biometrics, **40**, 587-600.
Su, J.Q. and Wei, L.J., 1993. Nonparametric estimation for the difference or ratio of median failure times. Biometrics, **49**, 603-607.
Wand M.P. and Jones M.C., 1995. *Kernel smoothing.* Chapman and Hall, London.
Xiang, X., 1995. Estimation of conditional quantile density function. Journal of Nonparametric Statistics, **4**, 309-316.
Zhang S. and Karunamuni R.J., 2000. On nonparametric density estimation at the boundary. Journal of Nonparametric Statistics, **12**, 197-221.