Exploiting the quantile optimality ratio in finding confidence intervals for quantiles

Luke A. Prendergast and Robert G. Staudte*

Received 4 January 2016; Accepted 2 February 2016

A standard approach to confidence intervals for quantiles requires good estimates of the quantile density. The optimal bandwidth for kernel estimation of the quantile density depends on an underlying location-scale family only through the quantile optimality ratio (QOR), which is the starting point for our results. While the QOR is not distribution-free, it turns out that what is optimal for one family often works quite well for families having similar shape. This allows one to rely on a single representative QOR if one has a rough idea of the distributional shape. Another option that we explore assumes the data can be modelled by the highly flexible generalized lambda distribution (GLD), already studied by others, and we show that using the QOR for the estimated GLD can lead to more than competitive intervals. Effective confidence intervals for the difference between quantiles from independent populations is a byproduct. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: density quantile; optimal bandwidth; quantile density; two-sample comparisons

1 Introduction

This work is motivated by the need for good interval estimates for quantiles when one has only a rough idea of the underlying shape and builds on substantial theory already in the literature. Here, we concentrate on asymptotic intervals of the simple form (1), usually adding and subtracting two standard errors to the quantile estimate. The challenge is to obtain a good ‘distribution-free’ estimator of the standard error, so as to obtain intervals for quantiles with accurate coverage for all distributions having a similar shape. There is also an extensive literature on completely distribution-free confidence intervals for quantiles and, as a point of departure, we refer the reader to Serfling (1980, Sec. 2.6) for exact and asymptotic intervals based on order statistics and to DasGupta (2006, Ch.29) for bootstrap intervals.

Denote the quantile function associated with a distribution function $F$ by $Q(u) = F^{-1}(u) = \inf\{x : F(x) \geq u\}$, for $0 < u < 1$. Assuming $F$ has a positive derivative $F'(x) = f(x)$ on its domain, define the quantile density function (Parzen, 1979) by $q(u) = Q'(u) = 1/f(Q(u))$; it was earlier dubbed as the “sparsity index” by Tukey (1965), but Parzen’s nomenclature will be followed here.

A point estimate of $Q(u)$ based on a sample $X_1, \ldots, X_n$ of size $n$ from $F$ is the sample quantile $\hat{Q}_n(u) = F_n^{-1}(u)$, where $F_n(x)$ is the usual empirical distribution function. Letting $r_n^2 = u(1-u)q^2(u)$, it can be shown (DasGupta, 2006, Ch.7) that the studentized quantile is asymptotically normal: $\sqrt{n}(\hat{Q}_n(u) - Q(u))/r_n \rightarrow N(0,1)$ in distribution. This leads to a large sample $100(1-\alpha)\%$ confidence interval for $Q(u)$ of the form

Department of Mathematics and Statistics, La Trobe University, Melbourne, Australia, 3086

*Email: r.staudte@latrobe.edu.au
where \( z_u = \Phi^{-1}(u) \) and \( \Phi \) is the standard normal distribution function. In the two sample setting, a large sample 100(1 - \( \alpha \))% confidence interval for the difference between population \( u \)th quantiles \( Q_1(u) - Q_2(u) \) is given by

\[
\hat{Q}_{1,\rho}(u) - \hat{Q}_{2,\rho}(u) \pm \frac{\hat{r}_u}{n^{1/2}} + \frac{\hat{r}_u}{m^{1/2}},
\]

where \( \hat{r}_u = \frac{u(1 - u)q^2(u)}{n} \) and \( m \) denotes the number of observations in the second sample.

To make the one-sample confidence interval for \( Q(u) \) distribution-free, (and similarly for the two-sample case), one needs to replace \( \hat{r}_u \) in (1) by a consistent estimator \( \hat{r}_u \), which in turn requires a consistent estimator of the quantile density \( q(u) \). Thus, there are two sources of error in the interval (1) that due to estimation of the centre and that due to estimation of the width. Sheather & Marron (1990) have compared several asymptotically optimal kernel estimators \( \hat{Q}(u) \) of \( Q(u) \) and found

“... that apart from the extreme quantiles, there is little difference between various quantile estimators (including the sample quantile). Given the well-known distribution-free inference procedures (e.g., easily constructed confidence intervals) associated with the sample quantile, as well as the ease with which it can be calculated, it will often be a reasonable choice as a quantile estimator.” (Sheather & Marron, 1990)

Here, we estimate the quantiles \( x_u = Q(u) \) for \( u \in [0.05, 0.95] \) using a linear combination of adjacent order statistics, the type 8 version of the quantile command on the package (R Core Team, 2014) recommended by Hyndman & Fan (1996). We denote this quantile estimator \( \hat{x}_u \) throughout. Because estimation of the quantile density \( q(u) \) is more difficult, with \( \text{MSE}(\hat{q}(u)) = O(n^{-4/5}) \) and \( \text{MSE}(\hat{Q}(u)) = O(n^{-1}) \), the more important estimate in the confidence interval (1) is not the centre, but the width of the interval, which requires an estimate of \( q(u) \).

Jones (1992) makes a strong case for estimating \( q(u) \) directly by kernel methods, rather than by taking the reciprocal of a kernel estimate of \( f(x_u) \). It turns out that an asymptotically optimal choice of bandwidth for estimating \( q(u) \) only depends on what we choose to call the \textit{quantile optimality ratio} \( \text{QOR}(u) = q(u)/q''(u) \), so this is the object of our attention in Section 2. Welsh (1988) suggests estimating \( q(u) \) and \( q''(u) \) separately and taking the ratio of these estimates to find the optimal bandwidth; and Cheng & Sun (2006) do so in computing the mean-squared error of estimators of \( Q(u) \). While ideally one would consistently estimate the QOR at \( u \), it turns out that one only needs a rough estimate of it to obtain good confidence intervals for \( x_u \).

Our goal is finding conceptually and computationally simple distribution-free confidence intervals for \( x_u \) and propose using either one of the optimal bandwidth kernel estimators for a representative location-scale family or an adaptive estimator for the generalized Tukey \( \lambda \) families. For the latter, our work is motivated by recent research.

In particular, we briefly describe in Section 2.4 the methods of Su (2009). We compare the finite sample performance of their confidence intervals with ours in Section 3.1 and also demonstrate the effectiveness of our intervals in two-population comparisons in Section 3.2. An R script for implementing the optimal QOR bandwidth and finding the associated intervals described in the next section is provided as Supporting Information. We conclude with a discussion and summary in Section 4.

## 2 Quantile optimality ratios for optimal bandwidths

For background material on kernel density estimation, see Wand & Jones (1995).
2.1 Quantile density estimators

An appealing and simple to implement quantile density estimator is a kernel density estimator, which can be expressed as a linear combination of order statistics:

\[ \hat{q}(u; b) = \sum_{i=1}^{n} X_{(i)} \left\{ k_{b} \left( u - \frac{(i-1)}{n} \right) - k_{b} \left( u - \frac{i}{n} \right) \right\}, \tag{3} \]

where \( b \) is a bandwidth and \( k_{b}(-) = k(-/b)/b \) for some kernel function \( k \), which is an even function on \([-1, 1]\] that has variance \( \sigma_{k}^2 = \int x^2 k(x)dx \) and roughness \( \kappa = \int k''(y)dy \). The asymptotic properties of this quantile density estimator have been studied by Falk (1986), Welsh (1988) and Jones (1992), who give conditions for consistency, including use of the Epanechnikov (1969) kernel, which we adopt.

Before proceeding, in Table I, we list simulated coverage probabilities of intervals for \( X_{u} \), using (1) with the aid of (3) and fixed bandwidths to estimate \( \tau_{u} \). The data are generated from the \( \text{LN}(0,1) \) and \( x^2_{1} \) distributions. We consider \( u \in \{0.05, 0.1, 0.5, 0.9, 0.95\} \), \( b \in \{0.05, 0.1, 0.2, 0.3\} \) and two different sample sizes \( n = 100 \) and \( n = 500 \). The nominal coverage is 0.95, and clearly, the performance of interval estimators depend on the choice of bandwidth, \( b \), and the best choice depends on \( u \) and the underlying distribution. These observations will come as no surprise to those who study kernel estimators, but proposals for kernel estimators using fixed bandwidths still appear in the literature. Jones (1992) gives the asymptotic MSE of \( \hat{q}(u) = \hat{q}(u; b) \) as

\[ \text{MSE}[\hat{q}(u)] = \frac{b^4 \sigma_{k}^4 \left\{ q''(u) \right\}^2}{4} + \frac{\kappa q^2(u)}{bn}. \tag{4} \]

By differentiating (4) with respect to \( b \), one finds that \( \text{MSE}[\hat{q}(u)] \) has a minimum when the bandwidth \( b(u) = A(u)/n^{1/5} \), where

\[ A(u) = \left( \frac{\kappa}{\sigma_{k}^4} \right)^{1/5} \left\{ \frac{q(u)}{q''(u)} \right\}^{2/5}. \tag{5} \]

Therefore, an asymptotically optimal choice of bandwidth (5) for estimating \( q(u) \) only depends on the underlying distribution through the QOR \( u = q(u)/q''(u) \). Note that if \( F_{a,b}(x) = F((x-a)/b) \), for unknown \( a \) and \( b > 0 \), is the
location-scale family generated by $F = F_{a,1}$, then the quantile function for $F_{a,b}$ is $Q_{a,b}(u) = a + b Q(u)$ and the quantile density is $q_{a,b}(u) = b q(u)$; thus, the quantile density is location invariant and scale equivariant, and the QOR is both location and scale invariant.

### 2.2 Examples of the quantile optimality ratio

Now we find the QORs for four families, two of which will be utilized throughout. Plots of them (not shown) are roughly similar in shape to the respective density quantiles $f(Q(u)) = 1/q(u)$, a phenomenon, which we have not yet been able to explain; see many more examples and comments in the Appendix.

#### 2.2.1 Normal

Letting $z_u = Q_{\Phi}(u) = \Phi^{-1}(u)$, the quantile density function is $q_{\Phi}(u) = 1/\varphi(z_u)$, with first two derivatives

\[
q'(u) = q(u) q_{\Phi}(u) + Q(u) q'_{\Phi}(u)
\]

\[
q''(u) = q'(u) q_{\Phi}(u) + 2q(u) q'_{\Phi}(u) + Q(u) q''_{\Phi}(u).
\]

The QOR for the log-normal is just the ratio of (6) to (7).

#### 2.2.2 Lognormal

For $x > 0$, let $F(x) = \Phi(\ln(x))$ be the log-normal distribution function, and let $q_{\Phi}(u) = 1/\varphi(z_u)$ be the quantile density of the normal distribution. It follows that $f(x) = F'(x) = \varphi(\ln(x))/x$. Therefore, $x_u = Q(u) = \exp(z_u)$ for $0 < u < 1$, and the quantile density and its first two derivatives are

\[
q(u) = \frac{x_u}{\varphi(z_u)} = Q(u) q_{\Phi}(u)
\]

\[
q'(u) = q(u) q_{\Phi}(u) + Q(u) q'_{\Phi}(u)
\]

\[
q''(u) = q'(u) q_{\Phi}(u) + 2q(u) q'_{\Phi}(u) + Q(u) q''_{\Phi}(u).
\]

The QOR can be written $QOR_{\Phi}(u) = 0.4 \varphi(6(u - 0.5))$ for $0 < u < 1$.

#### 2.2.3 Generalized lambda distribution

We adopt the parameterization of Freimer et al. (1988) (often referred to as the FKML parameterization), which has quantile function determined by a location parameter $\lambda_1$, an inverse scale parameter $\lambda_2 > 0$ and two shape parameters $\lambda_3$ and $\lambda_4$:

\[
Q(u) = \lambda_1 + \frac{1}{\lambda_2} \left\{ \frac{u^{\lambda_3} - 1}{\lambda_3} - \frac{(1-u)^{\lambda_4} - 1}{\lambda_4} \right\}.
\]

The quantile optimality ratio $QOR(u)$ is location and scale free, so without loss of generality, we can take $\lambda_1 = 0$ and $\lambda_2 = 1$ in (8). The QOR for the FKML parameterization is then

\[
QOR(u) = \frac{u^{\lambda_3} - 1 + (1-u)^{\lambda_4} - 1}{u^{\lambda_3-3}(\lambda_3 - 2)(\lambda_3 - 1) + (1-u)^{\lambda_4-3}(\lambda_4 - 2)(\lambda_4 - 1)}.
\]

The RS parameterization for the GLD (RS Ramberg & Schmeiser, 1974) is also sometimes used. The QOR for this parameterization is

\[
QOR(u) = \frac{\lambda_3 u^{\lambda_3-1} + \lambda_4 (1-u)^{\lambda_4-1}}{u^{\lambda_3-3}(\lambda_3 - 2)\lambda_3 + (1-u)^{\lambda_4-3}(\lambda_4 - 2)\lambda_4}.
\]
This generalized Tukey family of distributions can be used to approximate a large number of others, as described in Freimer et al. (1988), Karian & Dudewicz (2000) and Gilchrist (2000). Consequently, for appropriate choices of \( \lambda_3 \) and \( \lambda_4 \), the GLD QORs in (9) and (10) may be used to approximate the QORs for many other distributions.

### 2.3 Interval estimators for quantiles based on the quantile optimality ratio

As noted in Section 2.1, the asymptotic MSE of the quantile density estimator \( \hat{q}(u) \) in (3) is minimized when the bandwidth is chosen by (5), which itself depends on the QOR. For the Epanechnikov (1969) kernel, we have \( A(u) = 1.718 \) QOR\((u)^{2/5}, \) and the optimal bandwidth is \( b^*(u) = A(u)/n^{1/5} \). Then our nominal 100\((1 - \alpha)%\) confidence interval for \( x_u \) is \( \hat{x}_u = zx_{1 - \alpha/2} \frac{\hat{e}}{\sqrt{n}} \), where \( \hat{e}^2 = u(1 - u)\hat{q}(u) \). The interval estimators for differences of quantiles based on independent estimators are similarly obtained.

For the special case of symmetric unimodal distributions, we found via simulation studies that the QOR for the Cauchy distribution, see the Appendix, led to excellent coverage of interval estimates of quantiles for the normal, logistic, Laplace distributions and the Cauchy. However, the intervals based on the QORs for these other distributions did not lead to quantile intervals with good coverage for distributions other than the one for which they were optimal.

**Important boundary correction:** Preliminary simulations showed that the previously mentioned intervals had unsatisfactory coverage for smallish \( u < 0.2 \) when the support of \( F \) was positive, because the “optimal” bandwidths were extending beyond the lower boundary of \([0, +\infty)\). This problem was successfully resolved by taking the bandwidth to be \( \min(u, b^*(u)) \). Hereafter, we make this correction to optimal QOR bandwidths for distributions with support \([0, +\infty)\).

With this correction, it turns out that the lognormal QOR results in excellent coverage probabilities for interval estimators of quantiles \( x_u \) for \( 0.05 \leq u \leq 0.95 \) for many skewed distributions.

Alternatively, if one prefers to employ the GLD model and estimates the parameters \( \lambda_3, \lambda_4 \), one obtains an estimated GLD QOR that leads to a good bandwidth selection and competitive confidence intervals.

### 2.4 Other adaptive interval estimators of quantiles

Our considering to use the GLD QOR as an adaptive QOR for optimal bandwidth selection was motivated largely by the work of Su (2009) who also used the GLD distribution to obtain interval estimators for quantiles. Given that Su (2009) shows that the approaches provide good results in the small to moderate sample size setting when compared with bootstrap methods, we use these GLD methods as comparisons.

**2.4.1 Generalized lambda distribution method 1.** Su (2009) suggests taking \( f \) to be the density of the GLD which depends on the location parameter \( \lambda_1 \), inverse scale parameter \( \lambda_2 \) and the shape parameters \( \lambda_3 \) and \( \lambda_4 \). Given estimates of \( \lambda_1, \ldots, \lambda_4 \), which give rise to estimated GLD quantile \( \hat{Q}_C(u) \), estimated GLD density \( \hat{f}_C \) and estimated standard error \( \hat{e}_C/\sqrt{n} = \hat{u}(1 - \hat{u}) \hat{f}_C(\hat{Q}_C(u))/\sqrt{n} \) of the GLD quantile density, one can construct a confidence interval (1) for \( x_u \). Su (2009) calls this approach the Normal-GLD method, and we will similarly refer to it as such.

**2.4.2 Generalized lambda distribution method 2.** Su (2009) also introduced a method for estimating \( x_u = Q(u) \) based on the estimated GLD distribution function \( \hat{F} \). Let \( m = \text{floor}(nu) \) and \( F_{\hat{F}}(\cdot; m + 1, n - m) \) be the distribution function for the Beta \( B(m + 1, n - m) \) distribution. One then defines a \( 1 - \alpha \) confidence interval \( [L, U] \) by taking the solutions to \( \alpha/2 = F_{\hat{F}}(L; m + 1, n - m) \) and \( 1 - \alpha/2 = F_{\hat{F}}(U; m + 1, n - m) \). Such root-finding is routine on most statistical packages. For further justification of this approach, see Su (2009), who calls this the Analytical-GLD method.
3 Simulation studies

In this section, we present the results of simulations assessing the coverage probabilities of intervals for quantiles based on the methods of Su (2009) (Section 2.4) with those proposed by us, approximating the optimal bandwidth for the estimation of \( q(u) \) by using the QOR (Section 2.3).

3.1 One-sample comparisons

For estimation of the GLD parameters, we use maximum likelihood estimates for the FMKL parameterization (e.g. Su, 2007a). While many other estimators of the GLD parameters are available, after making several comparisons, Corlu & Meterelliyoz (2015) find that the MLE estimators are a good choice because they often result in decreased estimator variability. However, they note that there are ongoing improvements to GLD parameter estimation, so that the results that follow could later be improved with the introduction of better estimators. All simulation studies were carried out with the R statistical software package (R Core Team, 2014). Specifically, to obtain the numerical maximum likelihood estimates for the GLD distribution, we used the \texttt{gld} package (King et al., 2014). The package also includes many other estimators and the reader is referred to Corlu & Meterelliyoz (2015) for a discussion of the many available estimators. Another package for estimating the GLD parameters is the \texttt{GLDEX} package (Su, 2007b).

In Table II are shown simulated coverage probabilities for the interval estimators of \( Q(u) \) with \( u \in \{0.05, 0.1, 0.5, 0.9, 0.95\} \) for seven distributions (the log-normal, exponential, Pareto with shape equal to one and two and \( \chi^2 \) distributions with degrees of freedom equal to 1, 2 and 5) and for sample sizes of 50, 100, 200 and 500. The nominal coverage probability is 0.95 and a “grey” shaded cell indicates a coverage probability outside of the range \([0.93, 0.97]\).

As can be seen all of the methods can have difficulty in achieving close-to-nominal coverage for some distributions.
when \( n = 50 \). However, for this sample size, very good coverage probabilities are achieved for the QOR methods for the median \((u = 0.5)\). For the other choices of \( u \), the methods of Su (2009) are preferable.

As \( n \) increases, the QOR methods typically achieve close-to-nominal coverage for each of the distributions considered. With the exception of the \( \chi^2_1 \) distribution, the Analytical-GLD methods typically provide good results when \( n \) is not large. This is likely due to the approximating GLD distribution not being close enough to the true underlying distribution. This is certainly the case for the \( \chi^2_1 \) distribution where the estimated GLD distribution provided a very poor approximation, in particular for smaller \( u \), which explains why the coverage can be as low as zero. Overall, we feel with the exception of \( n = 50 \), either QOR approach works well, followed by the Normal-GLD method. Finally, it can be seen that there does not appear to be an advantage in choosing the GLD QOR because the log-normal QOR provides very good results. On the other hand, the GLD QOR is more adaptive and therefore more likely to provide better results for other types of distributions. This will be considered next.

In Figure 1, we provide the simulated coverage probabilities across a wider range of \( u \) for several distributions. To see whether the log-normal QOR is useful for non-skewed distributions, we also consider the symmetric Cauchy distribution. We can see that for the Cauchy, the log-normal QOR can result in poor coverage of the interval estimator (broken grey line), in particular for smaller choices of \( u \). On the other hand, the GLD QOR adapts to this new distribution well resulting in very good coverage for both \( n = 100 \) and \( n = 500 \) (grey line). In fact, the GLD QOR provides very good coverage for all the distributions considered here for both sample sizes. The log-normal QOR works well for the skewed distributions while the Analytical-GLD method provides reasonable results for \( n = 100 \),

![Figure 1. Simulation coverage probability results for the Cauchy, LN(0, 1) and EXP(1) distributions, plotted as a function of \( u \), for samples sizes \( n = 100 \) (left) and 500 (right).](image-url)
Table III. Simulation coverage probability results the QOR interval estimators for the difference in quantiles from data sampled from two independent populations.

| (n, m) | \( u \) | LN(1,1)-LN(0,2) | EXP(1)-EXP(1,5) | LN(0,1)–\( \chi^2 \) |
|--------|--------|-----------------|-----------------|-----------------|
| (100,100) | 0.05 | 0.983 | 0.950 | 0.973 |
|        | 0.1  | 0.979 | 0.942 | 0.967 |
|        | 0.5  | 0.97 | 0.961 | 0.959 |
|        | 0.9  | 0.972 | 0.963 | 0.967 |
|        | 0.95 | 0.96 | 0.971 | 0.973 |
| (200,200) | 0.05 | 0.982 | 0.952 | 0.973 |
|        | 0.1  | 0.982 | 0.952 | 0.969 |
|        | 0.5  | 0.968 | 0.962 | 0.961 |
|        | 0.9  | 0.967 | 0.950 | 0.962 |
|        | 0.95 | 0.969 | 0.973 | 0.978 |
| (200,500) | 0.05 | 0.989 | 0.947 | 0.982 |
|        | 0.1  | 0.976 | 0.942 | 0.973 |
|        | 0.5  | 0.97 | 0.955 | 0.961 |
|        | 0.9  | 0.974 | 0.955 | 0.969 |
|        | 0.95 | 0.969 | 0.966 | 0.956 |
| (500,500) | 0.05 | 0.978 | 0.943 | 0.968 |
|        | 0.1  | 0.971 | 0.943 | 0.958 |
|        | 0.5  | 0.963 | 0.962 | 0.964 |
|        | 0.9  | 0.962 | 0.955 | 0.971 |
|        | 0.95 | 0.974 | 0.960 | 0.965 |

Cells coloured in “grey” indicate coverage below 0.93 or above 0.97.

but can perform poorly for \( n = 500 \) for various \( u \). As we saw previously in Table II, overall, the Normal-GLD method outperforms the Analytical-GLD method. In summary, the QOR approach to interval estimation provides the most reliable results for the distributions considered here.

3.2 Two-sample comparisons

Given that the QOR interval estimators resulted in good coverage in the one-sample setting (which indicates a reliable estimator of the variance), we expect then that good results will also be found in the two-sample setting. We consider this in Table III where we report the simulated coverage for three different distribution scenarios, and each for varying sample size combinations and different \( u \). For example, in the column labelled LN(1,1)-LN(0,2), we consider the difference in quantiles for a set \( u \) from the LN(1,1) and LN(0,2) distributions where there are \( n \) and \( m \) observations sampled, respectively. Again, a “grey” cell indicates coverage outside of \([0.93, 0.97]\). For this simulation, we used the log-normal QOR for simplicity because similar results would be expected for the GLD QOR. As can be seen, coverage is typically very good for all scenarios.

4 Summary and discussion

A kernel estimator of the quantile density \( q(u) \) that has been the subject of investigation by several authors is known to have optimal (in the sense of minimizing the asymptotic mean squared error at \( u \)) bandwidth that depends on the underlying location-scale family only through the quantile optimality ratio \( \text{QOR}(u) = q(u)/q''(u) \). By examining numerous examples of this ratio, we found it to be relatively well behaved (compared with \( q \) and its derivatives) with
graph similar in shape to the square of the density quantile $f(x_u)$. The consequence for estimation of $q(u)$ needed in construction of the simple confidence intervals (1) is that $q''(u)$ need not be estimated. Rather, a representative QOR that is optimal for one family turns out to be more than adequate for many similarly shaped families. For example, the log-normal QOR provided good results for the setting of skewed distributions that we considered for our simulations. On the other hand, the QOR for the flexible GLD distribution may be useful because the GLD distributions is able to approximate a wide number of distributions. We found that by estimating the parameters first and then using the optimal QOR for this GLD led to relatively good coverage for the interval estimators, compared with GLD-based methods of Su (2009) when $n \geq 100$.

In summary, examination of the QOR for many parametric families shows that the optimal bandwidth based on it need not be very accurate to obtain a good estimate of the quantile density required for simple distribution-free confidence intervals, for moderate and large sample sizes. This methodology may enable one to obtain interval estimates of quantiles for mixtures of normal distributions, because parameter estimates are readily available. One would need an expression for the QOR in this case. More generally, an adaptive bandwidth estimator would utilize the fact that the QOR has approximately the shape of the density quantile $1/q(u)$, and alternately estimate the quantile density $q(u)$ with a new QOR bandwidth, for each of several iterations.

**Appendix. More examples of the quantile optimality ratio**

The distributions considered in the succeeding discussions are described in Johnson et al. (1994, 1995). Before presenting the examples, some remarks are in order. The first two derivatives of the quantile density $q$ are

$$q'(u) = \frac{f'(x_u)}{f^3(x_u)} = -\frac{f''(x_u)}{f^3(x_u)}q^3(u) \tag{A1}$$

$$q''(u) = 3\{f'(x_u)\}^2q^5(u) - f''(x_u)q^4(u).$$

In many cases, $f'(x) = g(x)f(x)$. (For example, in the Pearson systems of distributions, see Johnson et al. (1994), $g(x)$ is a rational function whose numerator is linear in $x$ and denominator is quadratic in $x$.) For such $f$, we have $f''(x) = \{g'(x) + g^2(x)\}f(x)$ so that from (A1) one finds $q'(u) = -g(x_u)q^2(u)$ and $q''(u) = \{2g^2(x_u) - g'(x_u)\}q^3(u)$; thus, the quantile optimality ratio becomes

$$\text{QOR}(u) = \frac{q(u)}{q''(u)} = \frac{f^2(x_u)}{2g^2(x_u) - g'(x_u)}. \tag{A2}$$

In this case, $\sqrt{\text{QOR}(u)}$ is the product of the density quantile function $f(x_u) = 1/q(u)$ defined by Parzen (1979) and a function depending only on $g$ and its derivative composed with the quantile function. Perhaps, it is worth noting that the score function of classical nonparametric statistics is $J(u) = -f'(x_u)/f(x_u) = -g(x_u).

**Cauchy**

For $F(x) = 0.5 + \text{arctan}(x)/\pi$, the quantile function is $x_u = Q(u) = \tan(\pi(u - 0.5))$. Hence, $q(u) = \pi \sec^2\{\pi(u - 0.5)\}, q'(u) = 2\pi^2[\tan\{\pi(u - 0.5)\} + \tan^3\{\pi(u - 0.5)\}], q''(u) = 2\pi^3 \sec^2\{\pi(u - 0.5)\}[1 + 3 \tan^2\{\pi(u - 0.5)\}]$, and elementary
calculations show that the QOR equals \( 1/(2\pi \sqrt{3}) \) times the Cauchy density having median 0 and scale parameter \( 1/\sqrt{3} \), evaluated at the quantile function.

**Laplace**

If \( f(x) = e^{-|x|}/2 \), the quantile function is \( Q(u) = \ln(2u) \) for \( 0 < u < 1/2 \) and \( Q(u) = -\ln(2(1-u)) \) for \( 1/2 < u < 1 \), so the quantile optimality ratio is \( QOR(u) = u^2/2 \) for \( 0 < u < 1/2 \) and \((1-u)^2/2 \) for \( 1/2 < u < 1 \). Elementary calculations show that \( QOR(u) = f_{1/2}(x_u)/8 \), a scale multiple of the Laplace distribution with median 0 and scale \( 1/2 \), evaluated at the quantile function.

**Logistic**

The density function \( f(x) = e^{-x}(1 + e^{-x})^2 \) for all \( x \) and the quantile function \( Q(u) = \ln(u) - \ln(1-u) \), for \( 0 < u < 1 \). It follows after taking its derivatives that the QOR is given by \( 2((1-u)^3 - u^3)/(u(1-u))^2 \).

**Bimodal with constant ratio**

Because at least one publication (Soni et al., 2012) used a constant ratio in the definition of the optimal bandwidth (5), we looked for a density \( f \) which leads to a constant QOR. One possibility is \( f(x) = (2e-|x|)^{-1} \) for all \( |x| < e-1 \). It has constant QOR(\(u\)) = 1/4. However, when the QOR in the bandwidth is constant, the ensuing estimator of \( q(u) \) is poor for most \( u \) compared with many others, so we did not consider it further.

**Tukey \( \lambda \)**

Recall that the Tukey(\( \lambda \)) distributions have quantile density function given by \( q(u) = u^{\lambda-1} + (1-u)^{\lambda-1} \) for all \( \lambda \). There is no closed form for the density \( f \) itself except when \( \lambda = 0 \) and the distribution \( F \) is logistic, or when \( \lambda = 1 \) or 2 and the distribution is uniform. When \( \lambda = 2.5 \), the quantile optimality ratio \( q(u)/q''(u) \) is approximately constant over the range \( 0.1 < u < 0.9 \).

**Pareto**

For \( x > 0 \) let, \( F(x;a) = 1 - (1 + x)^{-a} \), where \( a > 0 \) is a shape parameter; such distributions are called Lomax or Pareto type II. The quantile function \( x_u = Q(u) = (1 - u)^{-1/a} - 1 \), for \( 0 < u < 1 \); and, the quantile optimality ratio \( QOR(u) = a^2(1-u)^2/\{(1+a)(1+2a)\} \).

**Gamma**

Recall the Gamma distribution has density \( f(x;\alpha) = x^{\alpha-1}e^{-x}/\Gamma(\alpha) \) for \( x > 0 \), where \( \alpha > 0 \) is the shape parameter and \( \Gamma(\alpha) \) is the Gamma function. Using (A2) and the fact that \( f'(x;\alpha) = g(x)f(x;\alpha) \), where \( g(x) = (\alpha-1)/x-1 \), one finds the quantile optimality ratio:

\[
QOR_\alpha(u) = \frac{x_u^2 f^2(x_u;\alpha)}{(2\alpha-1)(\alpha-1) - 4(\alpha-1)x_u + 2x_u^2}.
\]  

(A.3)

Note that this ratio is negative for some \( u \) if \( 0.5 < \alpha < 1 \).
Weibull

For $\beta > 0$, the Weibull distribution function is defined for each $x$ by $F(x; \beta) = 1 - e^{-x^\beta}$. Clearly, $f'(x; \beta) = g(x)f(x; \beta)$, where $g(x) = (\beta - 1)/x - \beta/x^{\beta-1}$. Again applying (A2) yields the following result:

$$QOR_{\beta}(u) = \frac{x_u^2 f^2(x_u; \beta)}{(2\beta - 1)(\beta - 1) - 3\beta(\beta - 1)x_u^\beta + 2\beta^2 x_u^{2\beta}}. \quad (A.4)$$

The special case of the exponential, Weibull with $\beta = 1$, is of interest because its ratio $QOR(u) = (1-u)^2/2$, which is the same form as the ratio for the Pareto II family. However, for any $a > 0$, the coefficient of $(1-u)^2$ for the Pareto II (a) distribution is less than 1/2.

References

Cheng, M-Y & Sun, S (2006), ‘Bandwidth selection for kernel quantile estimation’, Journal of the Chinese Statistical Association, 44, 271–295.

Corlu, CG & Meterelliyoz, M (2015), ‘Estimating the parameters of the generalized lambda distribution: which method performs best’, Communications in Statistics - Simulation and Computation. Article In Press, DOI: 10.1080/03610918.2014.901355.

Das Gupta, A (2006), Asymptotic Theory of Statistics and Probability, Springer. DOI: 10.1007/978-0-387-75971-5.

Epanechnikov, VA (1969), ‘Nonparametric estimation of a multivariate probability density’, Theory of Probability and its Applications, 14, 153–158.

Falk, M (1986), ‘On the estimation of the quantile density function’, Statistics and Probability Letters, 4, 69–73.

Freimer, M, Mudholkar, GS, Kollia, G & Lin, CT (1988), ‘A study of the generalized Tukey lambda family’, Communications in Statistics - Theory and Methods, 17, 3547–3567.

Gilchrist, W (2000), Statistical Modelling with Quantile Functions, Chapman and Hall/CRC, London.

Hyndman, RJ & Fan, Y (1996), ‘Sample quantiles in statistical packages’, The American Statistician, 50, 361–365.

Johnson, NL, Kotz, S & Balakrishnan, N (1994), Continuous Univariate Distributions, Vol. 1, Wiley, New York.

Johnson, NL, Kotz, S & Balakrishnan, N (1995), Continuous Univariate Distributions, Vol. 2, Wiley, New York.

Jones, MC (1992), ‘Estimating densities, quantiles, quantile densities and density quantiles’, Annals of Institute of Statistical Mathematics, 44(4), 721–727.

Karian, ZA & Dudewicz, EJ (2000), Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods, Chapman and Hall.

King, R, Dean, B & Klinke, S (2014), gld: Estimation and use of the generalised (Tukey) lambda distribution. R package version 2.2.1.

Parzen, E (1979), ‘Nonparametric statistical data modeling’, Journal of the American Statistical Association, 7, 105–131.

R, Core Team (2014), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Ramberg, JS & Schmeiser, BW (1974), ‘An approximate method for generating asymmetric random variables’, Communications of the ACM, 17, 78–82.
Serfling, RJ (1980), *Approximation Theorems of Mathematical Statistics. Probability and Statistics*, Wiley, New York.

Sheather, SJ & Marron, JS (1990), ‘Kernel quantile estimators’, *Journal of the American Statistical Association*, 85, 410–416.

Soni, P, Dewan, I & Jain, K (2012), ‘Nonparametric estimation of quantile density function’, *Computational Statistics and Data Analysis*, 56, 3876–3886.

Su, S (2007a), ‘Numerical maximum log likelihood estimation for generalized lambda distributions’, *Computational Statistics and Data Analysis*, 51(8), 3983–3998.

Su, S (2007b), ‘Fitting single and mixture of generalized lambda distributions to data via discretized and maximum likelihood methods: GLDEX in R’, *Journal of Statistical Software*, 21(9), 1–17.

Su, S (2009), ‘Confidence intervals for quantiles using generalized lambda distributions’, *Computational Statistics and Data Analysis*, 53, 3324–3333.

Tukey, JW (1965), ‘Which part of the sample contains the information?’, *Proceedings of the Mathematical Academy of Science USA*, 53, 127–134.

Wand, MP & Jones, MC (1995), *Kernel Smoothing. Monographs on Statistics and Applied Probability*, Vol. 60, Chapman and Hall/CRC, London.

Welsh, AH (1988), ‘Asymptotically efficient estimation of the sparsity function at a point’, *Statistics and Probability Letters*, 6, 427–432.

**Supporting information**

Additional supporting information may be found in the online version of this article at the publisher’s web-site.