EFFICIENT ESTIMATION OF THE RECIPROCAL OF THE DENSITY QUANTILE FUNCTION AT A POINT

Gutti Jogesh BABU

Indian Statistical Institute, Calcutta, India *

Received March 1985
Revised November 1985

Abstract: Consistent estimators for the reciprocal of the density at a quantile point are considered. Optimal rates of convergence of these estimators, depending on the smoothness properties of the density, are obtained. An asymptotically efficient estimator sequence, in the mean square error sense, is found from among this class; unlike density estimators, these density quantile estimators do not require knowledge of the actual values of the density and its derivatives.

AMS 1980 Subject Classification: 60F05, 62G05, 62G20.

Keywords: density quantile function, quantile density function, weak convergence, asymptotic normality, strong approximations, order statistics.

1. Introduction

The importance of the density quantile function in statistical data modelling was realized by Parzen (1979). According to him greater insight will be obtained by formulating conclusions in terms of the density quantile function. So there is need for efficient estimators of the density at a quantile point. Study of an entirely unrelated problem – estimation of the bootstrap variance of sample median – led us to construct consistent estimators for the reciprocal of the density quantile function (what Parzen calls quantile density function) at a point. Stone (1980) has shown that the exponent of n in Theorem 1 is the best one attainable for nonparametric density estimators. He optimizes the ‘Window’ selection rule in density estimation. In this sense the rates of convergence of these quantile density estimators are optimal. We find an asymptotically efficient estimator sequence, in the mean square error sense, among this class. That is, we choose that sequence from this class whose expected mean square error is the smallest possible, asymptotically. These results differ from those of known density estimates. For example the asymptotically efficient estimator obtained by Krieger and Pickands (1981) depends in some fashion on the unknown density f and its derivatives. To get around this problem they find some preliminary estimates for f and its derivatives and substitute them in their efficient estimator; then they show that the resultant sequence of this two stage estimation is asymptotically efficient. In the case of quantile density estimation, no such two stage procedures are necessary.

To motivate the unusual, but natural, form of the estimator considered here, note that if F has a derivative \( f(\xi_p) > 0 \) at \( \xi_p \), where \( F(\xi_p) = p \), then

\[
(F^{-1}(v) - F^{-1}(p))/(v - p) = (f(\xi_p))^{-1} + o(1)
\]

* Present address: Pennsylvania State University, 219 Pond Lab University Park, PA 16802, USA

as \( v \to p \). Since the limit exists, both the right-hand side as well as the left-hand side limits exist and are both equal to \( 1/f(\xi_p) \). To get an approximation for \( 1/f(\xi_p) \), it is enough to consider \( F^{-1}(v) - F^{-1}(p) \) for \( v \geq p \) near \( p \). As \( F^{-1}(v) \) is not known, we replace it by a \( v \)th sample quantile \( F_n^{-1}(v) \). Since (for \( v \geq p \)) all the quantities \( (F_n^{-1}(v) - F_n^{-1}(p))/(v - p) \) are close to \( 1/f(\xi_p) \), a linear combination of these should also be near the value \( 1/f(\xi_p) \). If we add the values for too many \( v \)'s we might introduce some bias. So one has to have an optimal choice. The idea is to select the best possible weight function.

Finally we obtain a weak convergence result, the argument of the process being a parameter of estimation procedure, in particular the window width.

Csörgő and Révész (1981) and Csörgő (1983) consider the confidence bounds for \( 1/f(\xi_p) \). But these estimators do not have the minimum variance.

2. Preliminaries

Let \( k \geq 2 \) be an integer. Let \( h \) be a function on the positive real line such that \( h(y)e^y \) is a polynomial of degree \( \leq k \) and

\[
\int_0^\infty y^j h(y) \, dy = \begin{cases} 1 & \text{if } j = 0, 1, \\ 0 & \text{if } j = 2, \ldots, k. \end{cases}
\]

Existence of such an \( h \) is shown in the Lemma, given at the end of the last section. It is also shown that, for this \( h \),

\[
\sigma^2(x) = \int_0^x \int_0^x h(u)h(v) \min\{u, v\} \, du \, dv > 0
\]

for all \( x > 0 \). Note that

\[
0 < \sigma^2 = \int_0^\infty \int_0^\infty h(u)h(v) \min\{u, v\} \, du \, dv = \sigma^2(\log n) + O((\log n)^{k-1}).
\]

For any distribution function \( G \), \( 0 < u < 1 \), let \( G^{-1}(u) = \inf\{ y : G(y) \geq u \} \) be the least \( u \)th quantile. Let \( 0 < p < 1 \) and \( \epsilon > 0 \) be such that \( J = (p - \epsilon, p + \epsilon) \subset (0, 1) \).

We assume that the distribution function \( F \) is \( k \) times continuously differentiable at \( F^{-1}(u) \) for \( u \in J \) and that the first derivative of \( f \) of \( F \) at \( F^{-1}(p) \) is positive. Let \( a_j(u) \) denote the \( j \)th derivative of \( F^{-1} \) at \( u \in J \), \( b_j = a_j(p)/j! \) for \( j = 1, \ldots, k \) and let \( b(u) = \sum_{j=1}^k b_j u^j \).

Let \( X_1, \ldots, X_n \) be independent random variables with common distribution function \( F \). Let \( F_n \) denote the empirical distribution function. Note that \( F_n^{-1}(u) \), for \( 0 < u < 1 \), is the \( \lfloor un \rfloor \)th order statistic. Put \( \delta = (2k - 1)^{-1} \) and \( \beta = (1 - \delta)/2 = (k - 1)/(2k - 1) \). Define, for \( x > 0 \),

\[
D^*(x, n) = n^{2\delta - 1} \sum_{1 \leq i \leq n^{1/\delta}} (F_n^{-1}(p + i/n) - F_n^{-1}(p)) h(in^{\delta - 1})
\]

and

\[
D(x, n) = D(x, n, p) = n^\delta \int_0^x (F_n^{-1}(p + vn^{-\delta}) - F_n^{-1}(p)) h(v) \, dv = b_1(F(x, n) + L(x, n)),
\]

where

\[
L(x, n) = n^{\delta} b_1^{-1} \int_0^x b(vn^{-\delta}) h(v) \, dv \quad \text{and} \quad F(x, n) = \frac{1}{b_1} \{ D(x, n) - b_1 L(x, n) \}.
\]
Note that \( b_1 = 1/f(F^{-1}(p)) \). In what follows for any two functions \( g(y) \) and \( l(y) \), by \( g(y) \ll l(y) \) we mean there exists a constant \( C > 0 \) such that \( |g(y)| < C |l(y)| \) for all \( y \) in the domain of \( g \) and \( l \). Since, for \( 1 \leq i \leq n \),

\[
\left| h(in^{\delta-1}) \int_{(i-1)/n}^{i/n} nh(un^{\delta}) \, du \right| \ll \int_{(i-1)n^{\delta-1}}^{in^{\delta-1}} (1 + u^k) e^{-u} \, du,
\]

we have, by P7 of the next section, that if \( EX_1^2 < \infty \), then

\[
E(D(x, n) - D^*(x, n))^2 \ll n^{2\beta-2}(\log n)^{2k}
\]

uniformly in \( x \). So the statistics \( D(x, n) \) and \( D^*(x, n) \) are equivalent as far as the asymptotic results considered in this paper are concerned. So we study only \( D(x, n) \).

### 3. Results

We are now ready to state the main result.

**Theorem 1.** Let \( k \geq 2 \), \( 0 < p < 1 \) and \( \varepsilon > 0 \) be such that \( J = (p - \varepsilon, p + \varepsilon) \subset (0, 1) \). Let the distribution function \( F \) of \( X_1 \) be \( k \)-times continuously differentiable at \( F^{-1}(p) \) such that the first derivative \( f \) of \( F \) at \( F^{-1}(p) \) is positive. If \( E(X_1^2) < \infty \), then, uniformly in \( x > 0 \),

\[
n^{2\beta}E(D(x, n) f(F^{-1}(p)) - 1)^2 = \sigma^2(x) + o(1) + n^{2\beta}(1 - L(x, n))^2.
\]

Since by (1), \( L(log n, n) = 1 + O(n^{-1}(\log n)^k) \), it follows from (3) that

\[
n^{2\beta}E(D(log n, n) f(F^{-1}(p)) - 1)^2 = \sigma^2 + o(1).
\]

Let

\[
f_j(x) = \int_x^\infty y^j h(y) \, dy
\]

and

\[
(H_k): f_1 \text{ and } f_j \text{ do not have common positive roots for any } 2 \leq j \leq k - 1.
\]

It is easy to check directly that \((H_k)\) holds for \( k \leq 6 \). The following Theorem is immediate from Theorem 1, if we note that \( 1 - L(x, n) \gg n^{-\beta} \log n \) for any \( x \) in a bounded interval.

**Theorem 2.** Let \((H_k)\) hold. Suppose the \( j \)-th derivative of \( F^{-1} \) at \( p \) is non-zero for some \( 2 \leq j \leq k - 1 \). Then we have, uniformly in \( x \), that

\[
n^{2\beta}E\left(D(x, n) f(F^{-1}(p)) - 1\right)^2 \geq \sigma^2 + o(1) \tag{4}
\]

and that the equality occurs in (4) at \( x = \log n \).

**Remark.** As a result, in this case, \( D(log n, n) \) is an efficient estimator of \( 1/f(F^{-1}(p)) \), in the mean square error sense, among the class of estimators \( \{ D(x, n): x > 0 \} \).

To prove the results we assume, without loss of generality, that \( X_i = F^{-1}(U_i) \), where \( U_1, \ldots, U_n \) are independent variables, each uniformly distributed on \((0, 1)\). Let \( V_n \) denote the empirical distribution of
Let \( U_1, \ldots, U_n \) and let \( U_{(1)}, \ldots, U_{(n)} \) denote the order statistics. We require the following estimates; some of them are easy to check. The details are omitted in such cases.

P1. For \( p-e < s < t < p+e \),

\[
F^{-1}(t) - F^{-1}(s) = \sum_{j=1}^{k} \frac{a_j(p)}{(j-1)!} \int_s^t (v-p)^{j-1} \, dv + o\left( \int_s^t |v-p|^{k-1} \, dv \right)
\]

where

\[
F^{-1}(t) - F^{-1}(s) = \sum_{j=1}^{k} b_j \left( (t-p)^j - (s-p)^j \right) + o\left( |t-p|^k + |s-p|^k \right).
\]

Note that here and in what follows the error terms depend only on the bounds of derivatives \( a_j \) for \( j = 1, \ldots, k \) and on \( g(\lambda) \), where \( g(\lambda) \) is any function satisfying \( \sup\{ |a_p(p+u) - a_p(p)|; |u| < \lambda \} \leq g(\lambda) \to 0 \) as \( \lambda \to 0 \).

P2. For \( |p-p'| < e/2 \),

\[
P\left( |V_n^{-1}(p') - p| > \epsilon \right) \leq P\left( |V_n^{-1}(p') - p'| > \epsilon/2 \right) \ll n^{-2}.
\]

P3. As \( U_{(m)} \) has density \( n(m-1)u^{m-1}(1-u)^{n-m} \) on \( 0 < u < 1 \), we have \( EU_{(m)} = m/(n+1) \) and

\[
E\left( U_{(m)} - m(n+1)^{-1} \right)^2 = m(n-m+1)/(n+1)^2(n+2) \ll n^{-1}.
\]

P4. Let \( \mu = m/(n+1), \nu = 1 - \mu \). We have

\[
E(U_{(m)} - \mu)^4 = \mu \prod_{j=1}^{3} \frac{m+1}{n+j+1} - 4\mu^2 \frac{m+1}{n+2} \frac{m+2}{n+3} + 6\frac{m+1}{n+2} \mu^3 - 3\mu^4
\]

\[
\ll \prod_{j=1}^{3} \left( \mu + \frac{j\nu}{n+j+1} \right) - 4\mu^2 \prod_{j=1}^{2} \left( \mu + \frac{j\nu}{n+j+1} \right) + 6\mu^2 \left( \mu + \frac{\nu}{n+2} \right) - 3\mu^4
\]

\[
\ll n^{-2}.
\]

P5. For \( 1 < m < r < N < n \),

\[
E(U_{(r)} - U_{(m)})^2 = (r-m)(r-m+1)/(n+1)(n+2) \quad \text{and}
\]

\[
E((U_{(r)} - U_{(m)})(U_{(r)} - U_{(m)})) = (r-m)(N-r)/(n+1)(n+2).
\]

P6. For \( m < r < N \), using P3 and P5 we get

\[
E\left[ (U_{(N)} - U_{(m)})(U_{(r)} - U_{(m)})(U_{(r)} - U_{(m)} - (r-m)/(n+1)) \right]
\]

\[
= E\left[ (U_{(N)} - U_{(r)})(U_{(r)} - U_{(m)}) + (U_{(r)} - U_{(m)})^2 \right] - \frac{(r-m)(N-m)}{(n+1)^2}
\]

\[
= \frac{(r-m)(N-r) + (r-m)(r-m+1)}{(n+1)(n+2)} - \frac{(N-m)(r-m)}{(n+1)^2} = \frac{r-m}{n+1} \frac{N-m+1}{n+2} - \frac{N-m}{n+1}.
\]

\[
= ((r-m)/n(n+1)) + O((N-m)(r-m)n^{-3}) + O(n^{-2}).
\]
P7. Let $I$ denote the indicator function. Let $p', p'' \in (p - \varepsilon/2, p + \varepsilon/2)$. If $\int y^2 dF(y) < \infty$, then using P2 we have

\[ E\left(F_n^{-1}(p')^2 I\left(|V_n^{-1}(p'') - p| > \varepsilon\right)\right) \leq \int_0^\infty yP\left(\left(|F_n^{-1}(p')| > y\right) \cap \left(|V_n^{-1}(p'') - p| > \varepsilon\right)\right) dy
\]

\[ \leq P\left(|V_n^{-1}(p'') - p| > \varepsilon\right) + \int_{F^{-1}(p + \varepsilon)}^\infty \left(yP\left(F_n^{-1}(p') > y\right)\right) dy
\]

\[ + \int_{F^{-1}(p - \varepsilon)}^\infty \left(yP\left(F_n^{-1}(p') < -y\right)\right) dy
\]

\[ \leq n^{-2} + \int_0^\infty y \left[P\left(|F_n(y) - F(y)| > \varepsilon/2\right) + P\left(|F_n(-y) - F(-y)| > \varepsilon/2\right)\right] dy
\]

\[ \leq n^{-2}\left(\int_0^\infty [F(y)(1 - F(y)) + F(-y)(1 - F(-y))] dy + 1\right)
\]

\[ \leq n^{-2}.
\]

**Proof of Theorem 2.** Let $A(v) = V_n^{-1}(p + v) - V_n^{-1}(p) - v$. On the set $p - \varepsilon < V_n^{-1}(p) \leq V_n^{-1}(p') < p + \varepsilon$, $| p - p' | < \varepsilon/2$, we have by P1 that

\[ B(p', p) = F_n^{-1}(p') - F_n^{-1}(p) - b(p' - p) - A(p' - p) b_1
\]

\[ \leq (p' - p) |V_n^{-1}(p') - p'| + (V_n^{-1}(p'') - p'' + (V_n^{-1}(p'') - p) + o((p' - p)^k)).
\]

Using P3, P4 and P7 we obtain, uniformly in $x$, that

\[ E\left(\int_0^X |B(p + vn^{-\delta}, p) h(v)| dv\right)^2 = o(n^{1-\delta}). \tag{5}
\]

As a consequence of (5), P3 and P6 we have, uniformly in $x$,

\[ E(F(x, n)) = o(n^{-\beta}), \quad E(F(x, n))^2 = n^{2\beta} E\left(\int_0^X A(vn^{-\delta}) h(v) dv\right)^2 + O(n^{-2})
\]

\[ = n^{-2\beta}((\sigma^2(x) + o(1))
\]

and

\[ E\left(D(x, n) f\left(F_n^{-1}(p)\right) - 1\right)^2 = E\left[\left(D(x, n) - E(D(x, n))/b_1\right)^2 + (E(D(x, n)/b_1) - 1\right]^2
\]

\[ = n^{-2\beta}((\sigma^2(x) + o(1)) + \left(n^\delta \int_0^X b(vn^{-\delta}) h(v) dv - b_1\right)^2 \left(b_1 n^\beta\right)^{-2}). \tag{6}
\]

Note that by Theorem 4.4.1 of Csörgő and Révész (1981), there exists a Wiener process $W$ such that

\[ n^\beta F(x, n) = n^{\delta + \beta} \int_0^X A(vn^{-\delta}) h(v) dv + o_p(1)n^{-\beta}
\]

\[ = n^{\delta/2} \int_0^X (W(p + vn^{-\delta}) - W(p)) h(v) dv + o_p(1)n^{-\beta}.
\]

where $\overset{\sharp}{=} \overset{\exists}$ denotes the equality in distribution and $o_p(1)$ is a sequence of random variables tending to zero in probability. Since

\[ \{ Z(v) = (W(p + vn^{-\delta}) - W(p))n^{\delta/2}, v > 0 \} \overset{\exists}{=} \{ W(v); v > 0 \}.
\]
it follows that the process \( \{ n^\theta F(x, n) : x > 0 \} \) converges weakly to \( \{ \int_0^x W(v) h(v) \, dv : x > 0 \} \). So the process \( \{ n^\theta (D(x, n) - b_\theta) : x > 0 \} \) is asymptotically distributed as
\[
\left \{ b_1 \int_0^x W(v) h(v) \, dv + n^{\theta + \delta} \sum_{j=2}^{k} b_j n^{-\delta j} \int_0^x v^j h(v) \, dv : x > 0 \right \}.
\]

As a consequence we have, by (3), \( n^\theta(D(\log n, n) f(F^{-1}(p)) - 1) \) is asymptotically distributed as the normal distribution with mean zero and variance \( \sigma^2 > 0 \) (see the lemma below).

Similar arguments lead to its multidimensional version. That is, if \( 0 < p_1 < \cdots < p_m < 1 \) and \( F \) satisfies similar conditions at \( p_1, \ldots, p_m \) as at \( p \), then
\[
n^\theta(D(\log n, n, p_1) f(F^{-1}(p_1)) - 1) \sigma^{-1}, \ldots, n^\theta(D(\log n, n, p_m) f(F^{-1}(p_m)) - 1) \sigma^{-1}
\]
are asymptotically independent and each coordinate converges weakly to the standard normal variable.

Finally we prove the existence of \( h \) mentioned in Section 2.

**Lemma.** Let \( k > 2 \) be an integer. There exists a function \( h \) such that \( h(y)e^y \) is a polynomial of degree \( \leq k \) and satisfying (1). Further, \( \sigma^2(x) \), \( \sigma^2 \) defined in (2) and (3) are positive for all \( x > 0 \).

**Proof.** Let \( A \) be the square matrix with \( (k + 1) \) columns, whose \((i, j)\)th element is \((i + j)!\), \( i, j = 0, \ldots, k \). For any real numbers \( s_0, \ldots, s_k \), not all zero, we have
\[
\sum_{i=0}^{k} \sum_{j=0}^{k} (i+j)!s_i e_j = \sum_{i=0}^{k} s_i \sum_{j=0}^{k} e_j \int_0^\infty y^{i+j}e^{-y} \, dy = \int_0^\infty \left( \sum_{j=0}^{k} y^j e_j \right)^2 e^{-y} \, dy > 0.
\]

So \( A \) is positive definite. Put
\[
(u_0, \ldots, u_k) = (1, 1, 0, \ldots, 0) A^{-1}.
\]
Clearly \( h(y)e^y \) is a polynomial of degree \( \leq k \) and (1) holds if we take \( h(y) = (\sum_{j=0}^{k} u_j y^j)e^{-y} \). It is easy to show that
\[
u_i = \frac{(-1)^{i+1}}{i!} \left( k \begin{pmatrix} k+1 \cr i+1 \end{pmatrix} - \begin{pmatrix} k+2 \cr i+2 \end{pmatrix} \right)
\]
for \( i = 0, 1, \ldots, k \) satisfy equation (7). Note that
\[
h(y)e^y = 2y - \frac{1}{2}y^2 \quad \text{for } k = 2,
\]
\[
= 2 + 8y - \frac{7}{2}y^2 + \frac{y^3}{3} \quad \text{for } k = 3,
\]
\[
= -5 + 20y - \frac{35}{2}y^2 + \frac{7}{2}y^3 - \frac{y^4}{8} \quad \text{for } k = 4.
\]

To prove \( \sigma^2(x) > 0 \) for all \( x \), first note that
\[
\int_0^\infty \int_0^\infty u \left| h(u) h(v) \right| \, du \, dv < \infty.
\]
Observe that, for $x > 0$,
\[
\sigma^2(x) = \lim_{y \to \infty} \int_0^x \int_0^x h(u) h(v) \int_0^y I(w < u) I(w < v) \, dw \, dv = \lim_{y \to \infty} \int_0^y \left( \int_w^x h(u) \, du \right)^2 \, dw
\]
\[
= \int_0^\infty \left( \int_w^x h(u) \, du \right)^2 \, dw > 0,
\]
as $H(w) = \int_w^x h(u) \, du \neq 0$. Similar arguments show that $\sigma^2 > 0$. This completes the proof of the lemma.

References


