ON ASYMPTOTIC REPRESENTATIONS FOR REDUCED QUANTILES IN SAMPLING FROM A LENGTH-BIASED DISTRIBUTION

by

Pranab Kumar Sen

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1419

October 1982
ON ASYMPTOTIC REPRESENTATIONS FOR REDUCED QUANTILES IN SAMPLING FROM A LENGTH-BIASED DISTRIBUTION *

BY PRANAB KUMAR SEN

University of North Carolina, Chapel Hill.

Nonparametric estimation of the quantiles of a distribution based on a sample from the corresponding length-biased distribution is considered. Along with some representations of this estimator in terms of averages of independent random variables, some limiting results are established. The case of the reduced quantile process is treated briefly.

AMS Subject Classification : 60F05, 62E20, 62G05

Key Words and Phrases : Asymptotic normality, Bahadur-representation, empirical process, invariance principles, reduced quantiles and quantile process, weak convergence.

* Work partially supported by the National Heart, Lung and Blood Institute, Contract NIH-NHLBI-71-2243-L from the National Institutes of Health.
1. Introduction. Let $Y_1, \ldots, Y_n$ be $n$ independent and identically distributed (i.i.d.) nonnegative random variables (r.v.) from a distribution $G$, defined on $\mathbb{R}^+ = [0, \infty)$. $G = G_F$ is called a length-biased distribution corresponding to a distribution $F$ (also defined on $\mathbb{R}^+$), if

$$G_F(y) = \mu^{-1} \int_0^y x dF(x), \quad \text{for every } y \in \mathbb{R}^+, \tag{1.1}$$

where

$$\mu = \int_0^\infty x dF(x) \text{ is assumed to be finite.} \tag{1.2}$$

The distribution function (d.f.) $G_F$ arises naturally in many fields; see for example, Cox (1969), Patil and Rao (1977, 1978) and Coleman (1979) where many life applications are considered. We assume that the underlying d.f. $F$ admits a unique $p$-quantile $\xi_p = \xi_p(F)$, i.e.,

$$F(x) = p \text{ has a unique solution } \xi_p, \quad \text{where } 0 < p < 1. \tag{1.3}$$

Our main interest is to provide a nonparametric estimate of $\xi_p$ based on $Y_1, \ldots, Y_n$, and to study its various properties. For a somewhat related problem (based on a mixture of an ordinary and a length-biased distribution), we may refer to Vardi (1982a, b).

Let $Y_{n:1} \leq \ldots \leq Y_{n:n}$ be the order statistics corresponding to $Y_1, \ldots, Y_n$. Then, the sample $p$-quantile corresponding to $\xi_p$ is defined by $Y_{n:k}$ where $k$ is a suitably chosen (random) integer, depending on all the order statistics. This estimator is formally defined in Section 2, where the basic regularity conditions are also introduced. Section 3 is devoted to the study of the large sample properties of this estimator based on (i) the weak convergence of some related empirical processes, (ii) Bahadur (1966) representation of sample quantiles and (iii) some strong invariance principles for the empirical distributions. Some general remarks are appended in the concluding section.

2. The sample estimator. Note that by (1.1) and (1.2),

$$\mu^{-1} P = \mu^{-1} \int_0^{\xi_p} dF(x) = \mu^{-1} \int_0^{\xi_p} x^{-1} x dF(x) = \int_0^{\xi_p} x^{-1} dG(x), \tag{2.1}$$

so that

$$\mu^{-1} = \int_0^\infty x^{-1} dG(x). \tag{2.2}$$
Let $G_n(y) = n^{-1} \sum_{i=1}^{n} I(Y_i \leq y), y \in \mathbb{R}^+$ be the empirical d.f. based on $Y_1, \ldots, Y_n$. In (2.1) and (2.2), we replace the true d.f. $G$ by its nonparametric estimator $G_n$ and for $\mu^{-1}$, we obtain the estimator $\hat{\mu}^{-1}$ given by

$$\hat{\mu}^{-1} = n^{-1} \sum_{i=1}^{n} Y_i^{-1} = \int_0^\infty x^{-1} dG_n(x).$$

Noting that the d.f. $G_n$ is a step function, we now define a (random) integer $k (= K_n)$ by

$$K_n = \max \{ k : \sum_{i=1}^{k} Y_i^{-1} \leq p(\sum_{i=1}^{n} Y_i^{-1}) \}.$$  

Note that depending on the value of $p$ ($0 < p < 1$) and the $Y_i, i=1, \ldots, n$, the inequality in (2.4) may not hold even for $k=1$. This is particularly true when $n$ may be small and $p$ is close to zero. We may, however, remove this technical difficulty by letting $K_n = 1$ whenever $Y_1^{-1} > p(\sum_{i=1}^{n} Y_i^{-1})$. Thus, $K_n$ is a positive integer valued random variable, and, for every $n$, $P\{ 1 \leq K_n \leq n \} = 1$. The sample estimator of $\xi_p$ is then taken as

$$\hat{\xi}_{n,p} = Y_{n:K_n}, \text{ where } K_n \text{ is defined by (2.4)}. \tag{2.5}$$

We may remark that in the classical case, the sample $p$-quantile is defined by the $k$th largest order statistics, where $k$ is a positive integer such that $k/n$ is closest to $p$. However, in (2.5), the estimator comes out as an order statistic with a random index $K_n$ whose properties are not all that known. In fact, $K_n$ in (2.4) is defined in terms of partial sums of reciprocals of order statistics which are neither independent nor identically distributed. Hence, the classical renewal theory results may not apply directly to $K_n$. Note that by (2.3), $\mu^{-1} = n^{-1} \sum_{i=1}^{n} Y_i^{-1}$ has expectation $\mu^{-1}$ and variance $n^{-1} \gamma^2$, where

$$\gamma^2 = \int_0^\infty y^2 dG(y) - \mu^{-2} = \int_0^\infty x^{-1} d\mathbb{F}(x) - \mu^{-2}, \tag{2.6}$$

where we assume that $E_G Y^{-2} = E_F X^{-1} = \nu^2 < \infty$. Then, by the classical central limit theorem, as $n \to \infty$,

$$n^{\frac{1}{2}}(\hat{\mu}^{-1} - \mu^{-1}) / \gamma \overset{\mathcal{D}}{\to} \mathcal{N}(0, 1), \tag{2.7}$$

so that

$$n^{\frac{1}{2}}(\hat{\mu}^{-1} - \mu^{-1}) = O_p(1). \tag{2.8}$$
Further, note that \( a(x) = x^{-1} \) has a continuous derivative \( a'(x) = -x^{-2} \) for all \( x \in (0, \infty) \), so that by (2.7) and (2.8),

\[
(2.9) \quad \beta n^\frac{1}{2} (\mu - \mu_n) = \mu n^\frac{1}{2} (\mu_n - \mu) + o_p(1) \xrightarrow{D} \mathcal{N}(0, \mu^2 \gamma^2).
\]

These results will be useful in our subsequent analysis. We also assume that the d.f. \( F \) has a continuous probability density function (p.d.f.) \( f \) in some neighbourhood of \( \xi_p \) and \( f(\xi_p) \) is strictly positive. Note that \( g \), the p.d.f. of \( G \), is given by \( g(x) = xf(x) \), so that noting that \( \xi_p \in (0, \infty) \), we have

\[
(2.10) \quad 0 < g(\xi_p) = \xi_p f(\xi_p) < \infty.
\]

We may need some other regularity conditions which will be introduced as the occasions arise.

3. Weak representation of \( \hat{\xi}_{n,p} \). Let us denote by

\[
(3.1) \quad H_n(x) = \int_0^x y^{-1} dG_n(y) \quad \text{and} \quad H(x) = \int_0^x y^{-1} dG(y), \quad \text{for } x \in (0, \infty).
\]

Then, note that both \( H_n \) and \( H \) are nondecreasing, and, by (2.1)–(2.4),

\[
(3.2) \quad H(\xi_p) = p/\mu = p H(\infty) \quad \text{and} \quad H_n(Y_{n+1}^{\infty}) \leq p H_n(\infty) < H_n(Y_{n+1}^{\infty+1}),
\]

where, conventionally, we let \( Y_{n+1}^{\infty} = \infty \). Further, for every \( x \in (0, \infty) \), we let \( Y_i^{-1}(x) = Y_i^{-1}(Y_i \leq x) \), \( i = 1, \ldots, n \), so that

\[
(3.3) \quad n^\frac{1}{2} [H_n(x) - H(x)] = n^{-\frac{1}{2}} \sum_{i=1}^{n} Y_i^{-1}(x) - F G_i^{-1}(x) \xrightarrow{D} \mathcal{N}(0, \gamma_x^2),
\]

where

\[
(3.4) \quad \gamma_x^2 = \int_0^x y^{-2} dG(y) - (\int_0^x y^{-1} dG(y))^2 \quad \text{exists for every } x \in (0, \infty).
\]

Then, we have the following

Theorem 1. If \( \nu < \infty \) and (2.10) holds, then, as \( n \to \infty \),

\[
(3.5) \quad n^\frac{1}{2} f(\xi_p) + n^\frac{1}{2} \sum_{i=1}^{n} \{y_i^{-1} - \mu^{-1}\} + n^{-\frac{1}{2}} \sum_{i=1}^{n} \{y_i^{-1}(\xi_p) - p\mu^{-1}\} + o_p(1).
\]

Proof. Note that for every \( x > 0 \),

\[
(3.6) \quad \mu H_n(x) - H(x) = n^\frac{1}{2} \sum_{i=1}^{n} \{y_i^{-1} - \mu^{-1}\} + n^{-\frac{1}{2}} \sum_{i=1}^{n} \{y_i^{-1}(\xi_p) - p\mu^{-1}\} + o_p(1).
\]

Also, if we let \( x = \xi_p \pm C n^{-\frac{1}{2}} \), for some finite \( C(>0) \), then, it can be shown that (3.3) holds with the further simplification that \( \gamma_x^2 \) may be replaced by \( \gamma_{\xi_p}^2 \).

As such, by (2.9), (3.3) and (3.6), we conclude that for every \( n > 0 \), there exist
a $c_n (< \infty)$ and an integer $n_0$, such that for every (fixed) $C$ and $n \geq n_0$,

$$P\left\{ n^{1/2} \left| \mu_n H_n (\xi_p \pm Cn^{-1/2}) - \mu H(\xi_p \pm Cn^{-1/2}) \right| > c_n \right\} < \eta/2.$$  \hfill (3.7)

On the other hand, if we let $J_n = \{ x : \xi_p - Cn^{-1/2} \leq x \leq \xi_p + Cn^{-1/2} \}$, then

$$\mu_n H_n (\xi_p - Cn^{-1/2}) \leq \hat{\mu}_n H_n (x) \leq \mu_n H_n (\xi_p + Cn^{-1/2}), \forall x \in J_n,$$  \hfill (3.8)

while, by (2.10), as $n \to \infty$,

$$n^{1/2} \{ \mu H(\xi_p \pm Cn^{-1/2}) - \mu H(\xi_p) \} \to \pm C\mu f(\xi_p).$$  \hfill (3.9)

Hence, if we choose $C$ so large that $\mu C\mu f(\xi_p) > c_n$, we conclude from (3.7), (3.8) and (3.9) that, by virtue of (3.2), namely, $\mu_n H_n (Y_{n:K_n}^k) \leq p < \hat{\mu}_n H_n (Y_{n:K_n}^{k+1})$,

$$P\left\{ Y_{n:K_n}^k \in J_n \right\} \geq 1 - \eta, \text{ for every } n \geq n_0.$$  \hfill (3.10)

Having obtained this weaker bound for $Y_{n:K_n}^k$, we like to extend a result of Ghosh (1971) on (a weaker) Bahadur representation of sample quantiles to our $H_n$ and employ the same for the proof of (3.5). Note that by (3.6), for every $x, y \in J_n$,

$$\hat{\mu}_n H_n (x) - \mu H(\mu) = \{ \hat{\mu}_n - \mu \} \{ H(x) - H(y) \},$$  \hfill (3.11)

where, by (2.9), (2.10) and the definition of $J_n$,

$$\sup \left\{ \left| \{ \hat{\mu}_n - \mu \} \{ H(x) - H(y) \} \right| : x, y \in J_n \right\} = O_p (n^{-1}), \quad \hat{\mu}_n = O_p (1),$$  \hfill (3.12)

while, using the identity that for $x \geq y$,

$$H_n (x) - H_n (y) = \int_y^x u^{-1} d [G_n (u) - G(u)] = x^{-1} [G_n (x) - G(x)] - y^{-1} [G_n (y) - G(y)] + \int_y^x u^{-2} [G_n (u) - G(u)] \, du$$

along with (2.10) and the Bahadur (1966) representation:

$$\sup \left\{ \left| G_n (x) - G_n (y) - G(x) + G(y) \right| : x, y \in J_n \right\} = O_p (n^{-3/4} \log n),$$  \hfill (3.13)

we conclude that

$$\sup_{x, y \in J_n} \left\{ n^{1/2} \left| H_n (x) - H_n (y) + H(y) \right| \right\} = O_p (n^{-1/4} \log n) = o_p (1), \text{ as } n \to \infty.$$  \hfill (3.14)

By virtue of (3.10) and (3.15), we conclude that

$$n^{1/2} \left\{ H_n (Y_{n:K_n}^k) - H(Y_{n:K_n}^{k+1}) \right\} = \mu_n H_n (\xi_p) - H(\xi_p) + o_p (1),$$  \hfill (3.16)

so that, by (3.6), (3.15) and (3.16), we have
\[(3.17) \quad n^{1/2} \{ \hat{Y}_n - \mu Y_n \} = n^{1/2} \{ H_n(\xi_p) - H(\xi_p) \} + H(\xi_p) n^{1/2} (\hat{\mu}_n - \mu) + o_p(1) \]

where \(H(\xi_p) = \mu^{-1} p\) and \(\hat{Y}_n = p + o_p(n^{-1})\). Hence, as \(n \to \infty\),
\[(3.18) \quad n^{1/2} \{ \hat{Y}_n - \mu Y_n \} = n^{1/2} \{ H_n(\xi_p) - H(\xi_p) \} + \frac{1}{n^{1/2}} (\hat{\mu}_n - \mu) + o_p(1) \]

Finally, (3.5) follows from (3.18), (2.7), (2.9) and the fact that for \(x \in J_n\),
\[H'(x) = x^{-1} g(x) = f(x) \rightarrow f(\xi_p)\]
as \(n \to \infty\). Q.E.D.

Note that \(\{ pY^{-1}_i - p \mu^{-1}_i - Y_i^{-1}(\xi_p) + p \mu^{-1}_i = pY^{-1}_i - Y_i^{-1}(\xi_p), i \geq 1 \}\) form a sequence of i.i.d.r.v.'s with mean zero and variance
\[(3.19) \quad \sigma^2 = (1-p)^2 \int_0^\xi p y^{-2} dG(y) + p^2 \int_\xi^\infty y^{-2} dG(y) < \nu^2 \]
and hence, by (3.5) and the central limit theorem, we arrive at the following.

**Theorem 2.** Under the hypothesis of Theorem 1,
\[(3.20) \quad n^{1/2} \{ Y_n - \xi_p \} \xrightarrow{D} \mathcal{N}(0, \sigma^2 / f^2(\xi_p)) \]

This result as well as (3.5) can be extended to several quantiles under no extra regularity conditions.

4. Some general remarks. In (3.5), we have considered a weak representation for \(\hat{\xi}_n, p\) where the precise order of the remainder term has not been studied. If, in addition to (2.10), we assume that in some neighbourhoud of \(\xi_p\), the p.d.f. \(f\) has a bounded first derivative \(f'\), then, noting that \(g'(x) = (d/dx)\{xf(x)\} = f(x) + xf'(x)\), we conclude that \(g'\) is also bounded in the same neighbourhoud. As such, we may let \(J_n^* = \{x : x - \xi_p < n^{-3/4} (\log n)^{1/2}\}\) and parallel to (3.10), we may claim that \(Y_n \in J_n^*\) almost surely (a.s.), as \(n \to \infty\), while in (3.14), we may not only use the order as \(n^{-3/4} \log n\) a.s., but also, by Taylor's expansion, we obtain that as \(n \to \infty\),
\[(4.1) \quad \sup \{|g(\xi_p) - G_n(x) + G_n(\xi_p); x \in J_n^*\} = O(n^{-3/4} \log n)\) a.s.

As such, we obtain that under this additional regularity condition, in (3.5), we have an almost sure representation with a remainder term \(O(n^{-3/4} \log n)\).
Secondly, instead of the behaviour of a single quantile, if we like to study the same for the entire process \( \{ n^{\frac{1}{2}} ( \hat{\xi}_{n,p} - \xi_p ) ; 0 < p < 1 \} \), we may need a more stringent moment condition. Note that \( \nu < \infty \) implies that as \( x \downarrow 0 \), \( x^{-2}G(x) \rightarrow 0 \), though the rate of this convergence is not precisely known. If we assume that

\[
\nu_{2+\delta} = \int_0^\infty y^{-2-\delta} \, dG(y) < \infty \quad \text{for some } \delta > 0,
\]

then, we obtain that

\[
x^{-2-\delta} G(x) \rightarrow 0 \text{ as } x \downarrow 0.
\]

Also, by (3.1), for every \( x > 0 \),

\[
\frac{1}{n^2} \{ H_n(x) - H(x) \} = \int_0^x y^{-1} d\{ n^{\frac{1}{2}} (G_n(y) - G(y)) \}
\]

\[
= x^{-1} n^{\frac{1}{2}} [ G_n(x) - G(x) ] + \int_0^x n^{\frac{1}{2}} [ G_n(y) - G(y) ] y^{-2} dG(y).
\]

Further, it follows from O'Reilly (1974) that for every \( \varepsilon > 0 \),

\[
\sup \{ \frac{1}{n^2} [G_n(x) - G(x)] / [G(x)[1-G(x)]] \}^{1-\varepsilon} : x \in \mathbb{R}^+ \} = O_p(1),
\]

so that by (4.3) and (4.5), for every \( x \in (0, \infty) \), as \( n \uparrow \infty \),

\[
x^{-1} n^{\frac{1}{2}} [ G_n(x) - G(x) ] = \left( \frac{1}{n^2} [G_n(x) - G(x)] / [G(x)[1-G(x)]] \right)^{1-\varepsilon} x \]

converges to 0, in probability,

where we choose \( \varepsilon \) so small that \( (2+\delta)(1-\varepsilon) > 1 \). As such, the representation in (4.4) is valid for every \( x > 0 \), and, using (3.6) and (4.4), we have

\[
n^{\frac{1}{2}} [ \hat{\mu}_n H_n(x) - \mu H(x) ] = \mu x^{-1} n^{\frac{1}{2}} [G_n(x) - G(x)] + \mu \int_0^x y^{-2} n^{\frac{1}{2}} [G_n(y) - G(y)] dy
\]

\[
+ H(x)n^{\frac{1}{2}} ( \hat{\mu}_n - \mu ) + o_p(1) \text{ a.e.}
\]

\[
= \mu x^{-1} n^{\frac{1}{2}} [G_n(x) - G(x)] + \mu \int_0^x y^{-2} n^{\frac{1}{2}} [G_n(y) - G(y)] dy +
\]

\[
- H(x)\mu^2 \int_0^\infty y^{-1} d\{ n^{\frac{1}{2}} [G_n(y) - G(y)] \} + o_p(1) \text{ a.e.}
\]

If we let 

\[
W_n^0(G(x)) = n^{\frac{1}{2}} [G_n(x) - G(x)] , \ x \in \mathbb{R}^+ \text{ and define the } \rho_q \text{ metric by}
\]

\[
\rho_q(x,y) = \sup \{ |x(t) - y(t)| / q(t) ; 0 < t < 1 \} ; q(t) = \{ t(1-t) \}^{1-\varepsilon},
\]

then, the weak convergence of 

\[
W_n^0 = \{ W_n^0(t) ; 0 < t < 1 \}
\]

to a tied-down Wiener process 

\[
W^0, \quad \text{in the } \rho_q \text{ metric in (4.8), follows from O'Reilly(1974), and hence, using (4.3) and (4.7), the weak convergence of } n^{\frac{1}{2}} [ \hat{\mu}_n H_n(x) - \mu H(x) ] , \ x \in \mathbb{R}^+ \text{ to a Gaussian process follows readily. Thus, if we define } W_n^* = \{ W_n^*(t) ; 0 < t < 1 \} \text{ by letting}
\]

\[
W_n^*(G(x)) = n^{\frac{1}{2}} [ \hat{\mu}_n H_n(x) - \mu H(x) ] , \ x \in \mathbb{R}^+ \text{ and } W^* = \{ W^*(t) ; 0 < t < 1 \} \text{ by letting}
\]

\[ W^*(t) = \mu b(t)W^0(t) + \int_0^t b^2(s)W^0(s)db(s) - H(b(t))\mu \int_0^1 W^0(s)b^2(s)db(s), \]
for \(0 < t < 1\) where \(b(t) = G^{-1}(t), \ t \in (0,1),\) then, we have

\[ (4.9) \quad W^*_n \xrightarrow{D} W^*, \text{ in the Skorokhod } J_1\text{-topology on } D[0,1]. \]

Note that the covariance function of \(W^*\) is given by

\[ (4.10) \quad E W^*(s)W^*(t) = (1-s)(1-t) \nu_{sAt} + s t (\nu_{1} - \nu_{sAt}) - (sAt)(\nu_{svt} - \nu_{sAt}), \quad \forall (s,t) \in (0,1)^2, \]

where

\[ (4.11) \quad \nu_t = \int_0^t y^{-2}dG(y), \text{ for } t \in (0,1). \]

Also, in (2.4), we denote the solution \(K_n^p\) by \(K_n^p(p)\) and assume that \(f(\xi_p)\) is
strictly positive (and finite) for all \(p : 0 < p < 1.\) Then, instead of the Bahadur-
representation in (3.14), if we use its generalization by Kiefer(1967) and
use (4.9), then, we are able to show that for every \(c > 0,\) as \(n \to \infty,\)

\[ (4.12) \quad \sup_{p \in (0,1)} n^{\frac{1}{2}} \left| \hat{Y}_{n:K_n^p} - \xi_p \right| / (\log n)^{\frac{1}{2}} < c \quad \text{a.s.,} \]

and further, (3.18) holds simultaneously for all \(p \in (0,1).\) As such, the weak
invariance principle for \(\{ n^{\frac{1}{2}}[\hat{\xi}_{n,p} - \xi_p ] ; 0 < p < 1\}\) follows from this extended
version of (3.18), (4.9) and the random change of time results in Billingsley
(1968, pp.144-150). Further, using the results of Csaki (1977), (4.7) may be
strengthened to an a.s. representation, and hence, strong invariance principles
(as well as the law of iterated logarithm) for the reduced quantile process
also follow under the additional regularity conditions (4.2) and that

\[ \sup\{ |f'(x)| : x \in R^+\} < \infty. \]
REFERENCES


