Second order improvements of sample quantiles using subsamples

Keith Knight*
Department of Statistics
University of Toronto

Gilbert W. Bassett, Jr.
Department of Finance
University of Illinois, Chicago

Abstract: Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with distribution function F. It is well known that if F is differentiable at the α-quantile $q(\alpha)$ with $F'(q(\alpha)) > 0$ then the sample quantile is asymptotically normal. In this note we compare this standard quantile estimator to one obtained by averaging the sample quantiles of non overlapping subsamples. It is straightforward to show that this "average-of-subsample" quantiles is first-order equivalent to the standard estimator. Second order properties however differ in an interesting fashion. While the standard estimate would be intuitively expected best, it is shown that a convex combination of the two (correlated) estimates outperforms either one in a certain sense. We also indicate connections to recently considered methods averaging of estimates from bootstrap samples (bagging) and from without replacement subsamples (subagging). Finally, we show how results generalize when the standard differentiability condition on F is relaxed.

1 Introduction

Suppose that $X_1, \cdots, X_n$ are independent, identically distributed (i.i.d.) random variables with distribution function F. The α quantile of F, $q(\alpha)$ can estimated non-parametrically by $\widehat{q}_n(\alpha)$ minimizing the objective function

$$h_n(t) = \sum_{i=1}^{n} \rho_\alpha(X_i - t) \tag{1}$$

*Research supported by a grant from the Natural Sciences and Engineering Research Council of Canada

where $\rho_\alpha(x) = x[\alpha - I(x < 0)]$. In the case where the minimizer of (1) is not unique, we can define $\widehat{q}_n(\alpha)$ to be the mid-point of the set of minimizers or, alternatively, one of the end-points of this set. In any event, if F is differentiable at $q(\alpha)$ with $F'(q(\alpha)) = f(q(\alpha)) > 0$ then

$$\sqrt{n}\,(\widehat{q}_n(\alpha) - q(\alpha)) \xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f^2(q(\alpha))}\right)$$

with the result holding for any sequence $\{\widehat{q}_n(\alpha)\}$ of minimizers of (1).

In this article, we will attempt to "improve" the asymptotic linearity of the estimator $\widehat{q}_n(\alpha)$ by combining quantile estimators from subsamples. More precisely, we can write

$$\sqrt{n}(\widehat{q}_n(\alpha) - q(\alpha)) = \frac{1}{f(q(\alpha))\sqrt{n}} \sum_{i=1}^{n} [\alpha - I(X_i < q(\alpha)] + \widehat{R}_n(\alpha). \tag{2}$$

The remainder term $\widehat{R}_n(\alpha)$ in the Bahadur-Kiefer (Bahadur, 1966; Kiefer, 1967) representation (2) can be thought of as representing the deviation from linearity of $\widehat{q}_n(\alpha)$. It is straightforward to construct estimators $\widetilde{q}_n(\alpha)$ such that

$$\sqrt{n}(\widetilde{q}_n(\alpha) - q(\alpha)) = \frac{1}{f(q(\alpha))\sqrt{n}} \sum_{i=1}^{n} [\alpha - I(X_i < q(\alpha)] + \widetilde{R}_n(\alpha) \tag{3}$$

where $\widetilde{R}_n(\alpha) \neq \widehat{R}_n(\alpha)$. A natural question to ask is whether or not it is possible to construct an estimator $\widetilde{q}_n(\alpha)$ satisfying (3) such that its remainder $\widetilde{R}_n(\alpha)$ is "smaller" than $\widehat{R}_n(\alpha)$.

2 Combining information from subsamples

We start by dividing $X_1, \cdots, X_n$ into k non-overlapping subsamples of length $n_1, \cdots, n_k$. Define $\widehat{q}_n^{(1)}(\alpha)$ to be the sample α quantile of $X_1, \cdots, X_{n_1}$, $\widehat{q}_n^{(2)}(\alpha)$ to be the sample α quantile of $X_{n_1+1}, \cdots, X_{n_1+n_2}$ and so on.

The quantile estimators $\widehat{q}_n^{(1)}(\alpha), \cdots, \widehat{q}_n^{(k)}(\alpha)$ can be combined in a number of ways. For example, we could define an estimator $\widetilde{q}_n(\alpha)$ to be a weighted average of the subsample estimators or the median (or some other order statistic); the latter type of estimator might be viewed as a generalization of Tukey's (1978) "ninther".

The following result gives a convolution-type theorem for estimators that combine the information in $\widehat{q}_n^{(1)}(\alpha), \cdots, \widehat{q}_n^{(k)}(\alpha)$.

THEOREM 1. Suppose that $\widetilde{q}_n(\alpha) = g_n\left(\widehat{q}_n^{(1)}(\alpha), \cdots, \widehat{q}_n^{(k)}(\alpha)\right)$ where
(a) $n_i/n \to \lambda_i > 0$ as $n \to \infty$ for $i = 1, \cdots, k$,
(b) $g_n(\boldsymbol{x}_n) \to g_0(\boldsymbol{x}_0)$ for all sequences $\{\boldsymbol{x}_n\}$ converging to $\boldsymbol{x}_0$, and

(c) $\{g_n\}$ and g_0 are location and scale equivariant: $g_n(a\,\boldsymbol{x} + b\mathbf{1}) = a\,g_n(\boldsymbol{x}) + b$.
Then

$$\sqrt{n}(\tilde{q}_n(\alpha) - q(\alpha)) \xrightarrow{d} \frac{1}{f(q(\alpha))}(W + V)$$

where $W \sim \mathcal{N}(0, \alpha(1-\alpha))$ and V is independent of W.

Proof. Let $W_1, \cdots, W_k$ be independent $\mathcal{N}(0, \alpha(1-\alpha))$ random variables. Then

$$\begin{pmatrix} \sqrt{n}(\hat{q}_n^{(1)}(\alpha) - q(\alpha)) \\ \vdots \\ \sqrt{n}(\hat{q}_n^{(k)}(\alpha) - q(\alpha)) \end{pmatrix} \xrightarrow{d} \frac{1}{f(q(\alpha))} \begin{pmatrix} \lambda_1^{-1/2} W_1 \\ \vdots \\ \lambda_k^{-1/2} W_k \end{pmatrix}$$

and so

$$\begin{aligned} \sqrt{n}(\tilde{q}_n(\alpha) - q(\alpha)) &= g_n\left(\sqrt{n}(\hat{q}_n^{(1)}(\alpha) - q(\alpha)), \cdots, \sqrt{n}(\hat{q}_n^{(k)}(\alpha) - q(\alpha))\right) \\ &\xrightarrow{d} \frac{1}{f(q(\alpha))} g_0\left(W_1/\lambda_1^{1/2}, \cdots, W_k/\lambda_k^{1/2}\right). \end{aligned}$$

Finally, suppose that $W_1', \cdots, W_k'$ are independent random variables with

$$W_i' \sim \mathcal{N}(\theta, \lambda_i^{-1}\alpha(1-\alpha))$$

where θ is unknown. Then $S = \sum_{i=1}^k \lambda_i W_i'$ is a sufficient and complete statistic for θ while $g_0(W_1', \cdots, W_k') - S$ is ancillary. Thus by Basu's Theorem, S is independent of $g_0(W_1', \cdots, W_k') - S$ (for all θ). The conclusion of the theorem follows by noting that when $\theta = 0$, $g_0(W_1/\lambda_1^{1/2}, \cdots, W_k/\lambda_k^{1/2})$ has the same distribution as $g_0(W_1', \cdots, W_k')$ and S has the same distribution as W. $\square$

It is easy to construct examples of estimators satisfying the conditions of Theorem 1 whose limiting distribution has $V \neq 0$ and $W + V$ is not normally distributed. For example, take $\tilde{q}_n(\alpha)$ to be some fixed order statistic of $\hat{q}_n^{(1)}(\alpha), \cdots, \hat{q}_n^{(k)}(\alpha)$; this is discussed further in Knight (2002).

3 Some second order theory

Under the conditions of Theorem 1, an optimal estimator of $q(\alpha)$ based on $\hat{q}_n^{(1)}(\alpha), \cdots, \hat{q}_n^{(k)}(\alpha)$ is

$$\bar{q}_n(\alpha) = \sum_{i=1}^k \frac{n_i}{n} \hat{q}_n^{(i)}(\alpha) \approx \sum_{i=1}^k \lambda_i \hat{q}_n^{(i)}(\alpha), \tag{4}$$

which satisfies $\sqrt{n}(\bar{q}_n(\alpha) - \hat{q}_n(\alpha)) = o_p(1)$.

Intuitively, the sample quantile $\widehat{q}_n(\alpha)$ should be a better estimator of $q(\alpha)$. To investigate this, we look at the second order representations of both estimators:

$$\sqrt{n}(\widehat{q}_n(\alpha) - q(\alpha)) = \frac{1}{f(q(\alpha))\sqrt{n}} \sum_{i=1}^{n} [\alpha - I(X_i < q(\alpha))] + \widehat{R}_n(\alpha) \tag{5}$$

$$\sqrt{n}(\bar{q}_n(\alpha) - q(\alpha)) = \frac{1}{f(q(\alpha))\sqrt{n}} \sum_{i=1}^{n} [\alpha - I(X_i < q(\alpha))] + \bar{R}_n(\alpha). \tag{6}$$

The following theorem gives the joint asymptotic behaviour of $\widehat{R}_n(\alpha)$ and $\bar{R}_n(\alpha)$.

THEOREM 2. Define $\bar{q}_n(\alpha)$ as in (4) and suppose that
(a) $F(q(\alpha) + t) - \alpha = t\, f(q(\alpha)) + o(t^{3/2})$ as $t \to 0$, and
(b) $n_i/n = \lambda_i + o(n^{-1/4})$ for $i = 1, \cdots, k$.
Then for $\widehat{R}_n(\alpha)$ and $\bar{R}_n(\alpha)$ defined in (5) and (6), we have

$$n^{1/4} \begin{pmatrix} \widehat{R}_n(\alpha) \\ \bar{R}_n(\alpha) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \widehat{R}_0(\alpha) \\ \bar{R}_0(\alpha) \end{pmatrix}$$

where

$$\widehat{R}_0(\alpha) = \frac{1}{f(q(\alpha))} \sum_{i=1}^{k} \lambda_i^{1/4} B_i \left(\frac{\lambda_i^{1/2} W}{f(q(\alpha))} \right)$$

$$\bar{R}_0(\alpha) = \frac{1}{f(q(\alpha))} \sum_{i=1}^{k} \lambda_i^{1/4} B_i \left(\frac{W_i}{f(q(\alpha))} \right);$$

$B_1, \cdots, B_k$ are independent Gaussian processes with $E[(B_i(s) - B_i(t))^2] = f(q(\alpha))|s - t|$, $W_1, \cdots, W_k$ are independent $\mathcal{N}(0, \alpha(1-\alpha))$ random variables that are independent of the B_i's, and

$$W = \sum_{i=1}^{k} \lambda_i^{1/2} W_i.$$

Proof. For notational convenience, relabel the observations X_{ij} for $i = 1, \cdots, k$ and $j = 1, \cdots, n_i$. Define

$$Z_n^{(i)}(u) = \sum_{j=1}^{n_i} \left[\rho_\alpha \left(X_{ij} - q(\alpha) - u/\sqrt{n} \right) - \rho_\alpha \left(X_{ij} - q(\alpha) \right) \right]$$

for $i = 1, \cdots, k$ and

$$Z_n(u) = \sum_{i=1}^{k} Z_n^{(i)}(u).$$

Note that $\sqrt{n}(\widehat{q}_n^{(i)}(\alpha)-q(\alpha))$ minimizes $Z_n^{(i)}$ and $\sqrt{n}(\widehat{q}_n(\alpha)-q(\alpha))$ minimizes Z_n. Moreover, $Z_n^{(i)}$ and Z_n can be approximated by (respectively)

$$\begin{aligned}
\bar{Z}_n^{(i)}(u) &= -\frac{u}{\sqrt{n}}\sum_{j=1}^{n_i}[\alpha - I(X_{ij}<q(\alpha)] + \frac{\lambda_i f(q(\alpha))}{2}u^2 \\
\bar{Z}_n(u) &= -\frac{u}{\sqrt{n}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}[\alpha - I(X_{ij}<q(\alpha)] + \frac{f(q(\alpha))}{2}u^2 \\
&= \sum_{i=1}^{k} Z_n^{(i)}(u).
\end{aligned}$$

Then

$$\begin{aligned}
& n^{1/4}(Z_n^{(i)}(u) - \bar{Z}_n^{(i)}(u)) \\
&= n^{-1/4}\sum_{j=1}^{n_i}\int_0^u \left\{\left[I(X_{ij}\le q(\alpha)+t/\sqrt{n}) - I(X_{ij}\le q(\alpha))\right] - f(q(\alpha))\frac{t}{\sqrt{n}}\right\} dt \\
&= \int_0^u Y_n^{(i)}(t)\,dt
\end{aligned}$$

and

$$n^{1/4}(Z_n(u) - \bar{Z}_n(u)) = \sum_{i=1}^{k}\int_0^u Y_n^{(i)}(t)\,dt$$

where $Y_n^{(i)} \xrightarrow{d} -\lambda_i^{1/4}B_i(\lambda_i^{1/2}\cdot)$ and $B_1,\cdots,B_k$ are independent Gaussian processes. Then following Knight (1998), we have

$$n^{1/4}\widehat{R}_n(\alpha) \xrightarrow{d} \frac{1}{f(q(\alpha))}\sum_{i=1}^{k}\lambda_i^{1/4}B_i\left(\frac{\lambda_i^{1/2}W}{f(q(\alpha))}\right)$$

and

$$\begin{aligned}
n^{1/4}\widehat{R}_n^{(i)}(\alpha) &= n^{1/4}\left\{\sqrt{n}(\widehat{q}_n^{(i)}(\alpha)-q(\alpha)) - \frac{1}{\lambda_i f(q(\alpha))\sqrt{n}}\sum_{j=1}^{n_i}[\alpha - I(X_{ij}<q(\alpha)]\right\} \\
&\xrightarrow{d} \frac{1}{\lambda_i^{3/4}f(q(\alpha))}B_i\left(\frac{W_i}{f(q(\alpha))}\right)
\end{aligned}$$

where

$$\frac{1}{\sqrt{n_i}}\sum_{j=1}^{n_i}[\alpha - I(X_{ij}<q(\alpha)] \xrightarrow{d} W_i$$

(as $n_i\to\infty$) and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}[\alpha - I(X_{ij}<q(\alpha)] \xrightarrow{d} W = \sum_{i=1}^{k}\lambda_i^{1/2}W_i.$$

Thus

$$\begin{aligned} n^{1/4}\bar{R}_n(\alpha) &= n^{1/4}\sum_{i=1}^{k}\lambda_i\widehat{R}_n^{(i)}(\alpha) + o_p(1) \\ &\xrightarrow{d} \frac{1}{f(q(\alpha))}\sum_{i=1}^{k}\lambda_i^{1/4}B_i\left(\frac{W_i}{f(q(\alpha))}\right), \end{aligned}$$

which completes the proof. □

Using properties of the two-sided Brownian motions, we have

$$\begin{aligned} \mathrm{Var}[\widehat{R}_0(\alpha)] &= \frac{[2\alpha(1-\alpha)]^{1/2}}{f^2(q(\alpha))\sqrt{\pi}} \\ \mathrm{Var}[\bar{R}_0(\alpha)] &= \frac{[2\alpha(1-\alpha)]^{1/2}}{f^2(q(\alpha))\sqrt{\pi}}\sum_{i=1}^{k}\lambda_i^{1/2} \\ \mathrm{Cov}[\widehat{R}_0(\alpha),\bar{R}_0(\alpha)] &= \frac{[2\alpha(1-\alpha)]^{1/2}}{2f^2(q(\alpha))\sqrt{\pi}}\sum_{i=1}^{k}\lambda_i^{1/2}\left[1+\lambda_i^{1/2}-(1-\lambda_i)^{1/2}\right]. \end{aligned}$$

Note that $\mathrm{Var}[\widehat{R}_0(\alpha)] < \mathrm{Var}[\bar{R}_0(\alpha)]$. (Duttweiler (1973) shows that if $F(x)$ is twice differentiable at $x = q(\alpha)$ then

$$\sqrt{n}E[\widehat{R}_n^2(\alpha)] = \mathrm{Var}(\widehat{R}_0(\alpha)) + o(n^{-1/4+\delta})$$

for any $\delta > 0$.) Moreover, both $\widehat{R}_0(\alpha)$ and $\bar{R}_0(\alpha)$ are uncorrelated with $W_1,\cdots,W_k$:

$$\begin{aligned} \mathrm{Cov}(\widehat{R}_0(\alpha),W_i) &= 0 \\ \mathrm{Cov}(\bar{R}_0(\alpha),W_i) &= 0 \end{aligned}$$

for $i = 1,\cdots,k$.

Since $\widehat{q}_n(\alpha)$ and $\bar{q}_n(\alpha)$ are equivalent to first order, any convex combination of the two will have the same first order representation although the variance of the second order term will vary. For some $t \in [0,1]$ define

$$\widetilde{q}_n(\alpha) = t\,\widehat{q}_n(\alpha) + (1-t)\,\bar{q}_n(\alpha)$$

and note that

$$\sqrt{n}(\widetilde{q}_n(\alpha) - q(\alpha)) = \frac{1}{f(q(\alpha))\sqrt{n}}\sum_{i=1}^{n}[\alpha - I(X_i < q(\alpha))] + t\,\widehat{R}_n(\alpha) + (1-t)\,\bar{R}_n(\alpha)$$

with

$$n^{1/4}\left[t\,\widehat{R}_n(\alpha) + (1-t)\,\bar{R}_n(\alpha)\right] \xrightarrow{d} t\,\widehat{R}_0(\alpha) + (1-t)\,\bar{R}_0(\alpha).$$

Since the remainder term above is asymptotically uncorrelated with the first order term, we can try to minimize the asymptotic variance of the remainder term; note that $\mathrm{Var}[t\,\widehat{R}_0(\alpha) + (1-t)\,\bar{R}_0(\alpha)]$ is minimized (for given $\lambda_1, \cdots, \lambda_k$) at

$$t(\lambda_1, \cdots, \lambda_k) = \frac{\mathrm{Var}[\bar{R}_0(\alpha)] - \mathrm{Cov}[\widehat{R}_0(\alpha), \bar{R}_0(\alpha)]}{\mathrm{Var}[\bar{R}_0(\alpha)] + \mathrm{Var}[\widehat{R}_0(\alpha)] - 2\mathrm{Cov}[\widehat{R}_0(\alpha), \bar{R}_0(\alpha)]}.$$

We can also minimize $\mathrm{Var}[t\,\widehat{R}_0(\alpha) + (1-t)\,\bar{R}_0(\alpha)]$ over all t, k, and non-negative λ_i's satisfying $\lambda_1 + \cdots + \lambda_k = 1$.

THEOREM 3. $\mathrm{Var}[t\,\widehat{R}_0(\alpha) + (1-t)\,\bar{R}_0(\alpha)]$ is minimized at $k = 2$, $\lambda_1 = \lambda_2 = 1/2$ and $t = 1/\sqrt{2}$.

Proof. For fixed k and t, $\mathrm{Var}[t\,\widehat{R}_0(\alpha) + (1-t)\,\bar{R}_0(\alpha)]$ is a symmetric function of $\lambda_1, \cdots, \lambda_k$ so we can assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$. Moreover, we can focus on local minima for which $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$. For fixed t and k, the local minima of $\mathrm{Var}[t\,\widehat{R}_0(\alpha) + (1-t)\,\bar{R}_0(\alpha)]$ occur at $\lambda_1 = \cdots = \lambda_k = 1/k$, in which case

$$\mathrm{Var}[t\,\widehat{R}_0(\alpha) + (1-t)\,\bar{R}_0(\alpha)] = \frac{[2\alpha(1-\alpha)]^{1/2}}{f^2(q(\alpha))\sqrt{\pi}} \left[t^2 + (1-t)^2\sqrt{k} + t(1-t)\left(1 + \sqrt{k} - \sqrt{k-1}\right)\right].$$

For given k, this is minimized over t at $t_k = (\sqrt{k} + \sqrt{k-1} - 1)/(2\sqrt{k-1})$, yielding

$$\mathrm{Var}[t_k\,\widehat{R}_0(\alpha) + (1-t_k)\,\bar{R}_0(\alpha)] = \frac{[2\alpha(1-\alpha)]^{1/2}}{f^2(q(\alpha))\sqrt{\pi}} \frac{\sqrt{k(k-1)} + \sqrt{k} + \sqrt{k-1} - k}{2\sqrt{k-1}}$$

which in turn is minimized over $k = 1, 2, \cdots$ at $k = 2$ with $t_2 = 1/\sqrt{2}$. □

At the optimal values $\lambda_1 = \lambda_2 = 1/2$ and $t = 1/\sqrt{2}$, we have

$$\frac{\mathrm{Var}[t\,\widehat{R}_0(\alpha) + (1-t)\,\bar{R}_0(\alpha)]}{\mathrm{Var}[\widehat{R}_0(\alpha)]} = \sqrt{2} - \frac{1}{2} \approx 0.9142.$$

The phenomenon noted above typically does not occur for "smoother" estimators. For example, suppose that $\widehat{\theta}_n$ minimizes

$$\sum_{i=1}^{n} \rho(X_i; t)$$

where $\rho(x;t)$ is a twice differentiable, convex function in t for each x. If $\psi(x;t)$ and $\psi'(x;t)$ are the first and second derivatives (with respect to t) of $\rho(x;t)$ then (under appropriate regularity conditions), we have

$$\sqrt{n}(\widehat{\theta}_n - \theta) = -\frac{1}{E[\psi'(X_1;\theta)]\sqrt{n}} \sum_{i=1}^{n} \psi(X_i;\theta) + R_n.$$

For the remainder term R_n, we have

$$n^{1/2} R_n \xrightarrow{d} \frac{1}{E[\psi'(X_1;\theta)]} W\, V$$

where

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{i=1}^n \psi(X_i;\theta) \\ \sum_{i=1}^n \{\psi'(X_i;\theta) - E[\psi'(X_i;\theta)]\} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} W \\ V \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, C).$$

On the other hand, if $\widehat{\theta}_n^{(1)}, \cdots, \widehat{\theta}_n^{(k)}$ are estimators based on subsamples of lengths $n_1, \cdots, n_k$ and

$$\bar{\theta}_n = \sum_{i=1}^k \frac{n_i}{n} \widehat{\theta}_n^{(i)}$$

then

$$\sqrt{n}(\bar{\theta}_n - \theta) = -\frac{1}{E[\psi'(X_1;\theta)]\sqrt{n}} \sum_{i=1}^n \psi(X_i;\theta) + R_n'$$

where

$$n^{1/2} R_n' \xrightarrow{d} \frac{1}{E[\psi'(X_1;\theta)]} \sum_{i=1}^k \lambda_i^{1/2} W_i\, V_i$$

and $(W_1, V_1), \cdots, (W_k, V_k)$ are i.i.d. pairs of random variables have the same distribution as (W, V) with

$$\begin{pmatrix} W \\ V \end{pmatrix} = \lambda_1^{1/2} \begin{pmatrix} W_1 \\ V_1 \end{pmatrix} + \cdots + \lambda_k^{1/2} \begin{pmatrix} W_k \\ V_k \end{pmatrix}.$$

Note that

$$\begin{aligned} \mathrm{Cov}\left(W\, V, \sum_{i=1}^k \lambda_i^{1/2} W_i\, V_i \right) &= \sum_{i=1}^k \lambda_i \mathrm{Var}(W_i\, V_i) \\ &= \mathrm{Var}(W\, V) \end{aligned}$$

which implies that the variance of the remainder term for the estimator $t\,\widehat{\theta}_n + (1-t)\,\bar{\theta}_n$ is minimized when $t = 1$. Likewise,

$$\left| \sum_{i=1}^k \lambda_i^{1/2} E(W_i V_i) \right| \geq |E(WV)|$$

with strict inequality if $E(WV) \neq 0$.

4 A step beyond: bagging and subagging

In the previous section, we saw that we could achieve a second order improvement of a sample quantile by combining it with an estimator constructed by taking averages of subsample quantiles.

Suppose that $\widehat{q}_n^*(\alpha)$ minimizes

$$h_n^*(t) = \sum_{i=1}^{n} \Delta_{ni}^* \rho_\alpha(X_i - t) \tag{7}$$

where $\mathbf{\Delta}_n^* = (\Delta_{n1}^*, \cdots, \Delta_{nn}^*)$ is a random vector independent of the X_i's. We will construct an estimator of $q(\alpha)$ by averaging the $\widehat{q}_n^*(\alpha)$'s over the distribution of $\mathbf{\Delta}_n^*$:

$$\widetilde{q}_n(\alpha) = E^*[\widehat{q}_n^*(\alpha)].$$

If $\mathbf{\Delta}_n^*$ has a multinomial distribution with $\Delta_{n1}^* + \cdots + \Delta_{nn}^* = n$ then $\widetilde{q}_n(\alpha)$ is Breiman's (1996) "bagged" estimator. If the Δ_{ni}^*'s are exchangeable 0/1 random variables with $P^*(\Delta_{ni}^* = 1) = \lambda_n \to \lambda > 0$ and $\Delta_{n1}^* + \cdots + \Delta_{nn}^* = n\lambda_n$ then $\widetilde{q}_n(\alpha)$ is the average of all possible subsample quantiles from subsamples of length $n\lambda_n$; we will call these latter estimators "subagged estimators" after Bühlmann and Yu (2002).

If $\widehat{q}_n^*(\alpha)$ minimizes (7) then we have the Bahadur-Kiefer representation

$$\sqrt{n}(\widehat{q}_n^*(\alpha) - q(\alpha)) = \frac{1}{f(q(\alpha))\sqrt{n}} \left\{ \sum_{i=1}^{n} [\alpha - I(X_i \le q(\alpha))] + \sum_{i=1}^{n} [\Delta_{ni}^* - E(\Delta_{ni}^*)] [\alpha - I(X_i \le q(\alpha))] \right\}$$
$$+ \widehat{R}_n^*(\alpha)$$

where

$$P^* \left[n^{1/4} \widehat{R}_n^*(\alpha) \in A \right] \xrightarrow{d} P^* \left[\frac{1}{f(q(\alpha))} B(W + W^*) \in A \right]$$

where B is a two-sided Brownian motion, W and W^* are independent normal random variables (independent of B), and the limiting probability is evaluated conditionally on B and W.

Second order representations for the bagged and subagged estimators can be obtained by averaging over the distribution of W^* conditional on B and W. Both the bagged and subagged estimators of $q(\alpha)$ satisfy

$$\sqrt{n}\left(\widetilde{q}_n(\alpha) - q(\alpha)\right) = \frac{1}{f(q(\alpha))\sqrt{n}} \sum_{i=1}^{n} [\alpha - I(X_i \le q(\alpha))] + \widetilde{R}_n(\alpha).$$

For the bagged estimator, we have

$$n^{1/4} \widetilde{R}_n(\alpha) \xrightarrow{d} \int_{-\infty}^{\infty} B(W + u)\, \phi_\alpha\left(f(q(\alpha))\, u\right)\, du$$

where $\phi_\alpha(x)$ is the density function of a $\mathcal{N}(0, \alpha(1-\alpha))$ random variable, $W \sim \mathcal{N}(0, \alpha(1-\alpha)/f^2(q(\alpha)))$, and B is a two-sided Brownian motion with $E[(B(s) - B(t))^2] = f(q(\alpha))|s-t|$.

For the subagged estimators, the limiting distribution of $n^{1/4}\widetilde{R}_n(\alpha)$ depends on the subagging fraction λ; in particular, we have

$$n^{1/4}\widetilde{R}_n(\alpha) \xrightarrow{d} \frac{\lambda^{1/2}}{(1-\lambda)^{1/2}}\int_{-\infty}^{\infty} B(W+u)\phi_\alpha\left(\frac{\lambda^{1/2}f(q(\alpha))}{(1-\lambda)^{1/2}}u\right)du = \widetilde{R}_0(\alpha;\lambda)$$

where B and ϕ_α are as before. Note that the limit for the bagged estimator is the same as that of a subagged estimator with $\lambda = 1/2$ and that $\widetilde{R}_0(\alpha;1) = \widehat{R}_0(\alpha)$, where $\widehat{R}_0(\alpha)$ is defined in the previous section.

The variance of $\widetilde{R}_0(\alpha;\lambda)$ is

$$\mathrm{Var}[\widetilde{R}_0(\alpha;\lambda)] = \frac{[2\alpha(1-\alpha)]^{1/2}}{f^2(q(\alpha))\pi^{1/2}}\left[\frac{1}{\lambda^{1/2}} - \frac{(1-\lambda)^{1/2}}{(2\lambda)^{1/2}}\right]$$

and

$$\begin{aligned}\mathrm{Cov}[\widetilde{R}_0(\alpha;\lambda_1),\widetilde{R}_0(\alpha;\lambda_2)] &= \frac{[2\alpha(1-\alpha)]^{1/2}}{2f^2(q(\alpha))\pi^{1/2}}\left[\frac{1}{\lambda_1^{1/2}} + \frac{1}{\lambda_2^{1/2}} - \left(\frac{\lambda_1+\lambda_2-2\lambda_1\lambda_2}{\lambda_1\lambda_2}\right)^{1/2}\right] \\ &= K(\lambda_1,\lambda_2).\end{aligned}$$

The variance above is minimized at $\lambda = 1/2$ and so bagging or subagging with fraction $1/2$ is optimal in this sense. In fact, we can go a step further. Defining $\widetilde{q}_n(\alpha;\lambda)$ to be the subagged estimator with fraction λ, if we define the estimator

$$\breve{q}_n(\alpha;\mu) = \int_{[0,1]} \widetilde{q}_n(\alpha;\lambda)\,\mu(d\lambda)$$

for a probability measure μ then

$$\sqrt{n}(\breve{q}_n(\alpha;\mu) - q(\alpha)) = \frac{1}{f(q(\alpha))\sqrt{n}}\sum_{i=1}^{n}[\alpha - I(X_i \le q(\alpha))] + \breve{R}_n(\mu)$$

where

$$n^{1/4}\breve{R}_n(\mu) \xrightarrow{d} \int_{[0,1]} \widetilde{R}_0(\alpha;\lambda)\,\mu(d\lambda) = \breve{R}_0(\alpha;\mu).$$

Defining

$$\begin{aligned}\varphi(\mu) &= \mathrm{Var}[\breve{R}_0(\alpha;\mu)] \\ &= \int_0^1\int_0^1 K(\lambda_1,\lambda_2)\,\mu(d\lambda_1)\,\mu(d\lambda_2),\end{aligned}$$

it follows that φ is minimized at the measure putting all its mass at $1/2$ (since $K(\lambda,1/2) = K(1/2,\lambda) = K(1/2,1/2)$ for $0 < \lambda < 1$). This suggests that bagging (or subagging with $\lambda = 1/2$) may be optimal. Note that

$$\frac{\mathrm{Var}[\widetilde{R}_0(\alpha;1/2)]}{\mathrm{Var}[\widehat{R}_0(\alpha)]} = \frac{1}{\sqrt{2}} \approx 0.7071,$$

which is smaller than the best ratio given in section 3.

Both the bagged and subagged estimators can be thought of as L-estimators of $q(\alpha)$ in the sense that they are weighted averages of order statistics. Suppose we define

$$\widetilde{q}_n(\alpha) = \int \widehat{q}_n(\alpha + t/\sqrt{n})\, \nu_n(dt) \tag{8}$$

where $\{\nu_n\}$ is a sequence of probability measures such that $\nu_n(A) \to \nu(A)$ for A such that $\nu(\partial A) = 0$, that is, $\{\nu_n\}$ converges weakly to ν; for bagging and subagging, the measure ν corresponds to a normal distribution whose variance depends on α and λ. For the estimators defined in (8), if $\int u\, \nu_n(du) \to 0$, we have

$$\sqrt{n}(\widetilde{q}_n(\alpha) - q(\alpha)) = \frac{1}{f(q(\alpha))\sqrt{n}} \sum_{i=1}^{n} [\alpha - I(X_i \le q(\alpha))] + \widetilde{R}_n(\alpha)$$

where

$$n^{1/4} \widetilde{R}_n(\alpha) \xrightarrow{d} \frac{1}{f(q(\alpha))} \int B(W + u/f(q(\alpha)))\, \nu(du)$$

where B and W are defined as above. In this case, the asymptotic variance

$$\mathrm{Var}\left[\frac{1}{f(q(\alpha))} \int B(W + u/f(q(\alpha)))\, \nu(du)\right]$$

is minimized over probability measures ν at the measure corresponding to a normal distribution with mean 0 and variance $\alpha(1-\alpha)$; note that this optimal measure ν is the same as the limiting measure used for bagging.

5 Non-standard conditions

The differentiability condition on F assumed to this point can be generalized; the techniques used to determine limiting distributions remain more or less the same although the limiting distributions themselves will change. Define $\psi_\delta(t)$ to be a non-decreasing function satisfying (for some $\delta \ge 0$)

$$\begin{aligned} \psi_\delta(t) &\to \pm\infty \quad \text{as } t \to \pm\infty \\ \psi_\delta(a\,t) &= a^{1/\delta}\psi_\delta(t) \quad \text{for } a > 0. \end{aligned} \tag{9}$$

When $\delta > 0$, ψ_δ has the form

$$\psi_\delta(t) = \begin{cases} c_+ t^{1/\delta} & \text{for } t > 0, \\ -c_-(-t)^{1/\delta} & \text{for } t < 0, \end{cases} \tag{10}$$

where $0 < c_+, c_- \leq \infty$ and at most one of c_+ and c_- is infinite. ψ_0 has the form

$$\psi_0(t) = \begin{cases} 0 & \text{for } -a_- < t < a_+, \\ +\infty & \text{for } t > a_+, \\ -\infty & \text{for } t < -a_+, \end{cases} \tag{11}$$

where $0 \leq a_+, a_- < \infty$ and at most one of a_+ and a_- is 0.

Given $\psi_\delta(t)$, suppose that for some sequence of constants $\{a_n\}$, we have

$$\lim_{n\to\infty} \sqrt{n}(F(q(\alpha) + t/a_n) - \alpha) \to \psi_\delta(t). \tag{12}$$

where ψ_δ is defined as in (10) or (11); the constants a_n are of the form $a_n = n^{\delta/2}L(n)$ where $L(n)$ is a slowly varying function. If F has a density f then $\delta < 1$ implies that $f(q(\alpha)) = 0$ while if $\delta > 1$ then $f(x) \uparrow \infty$ as $x \to q(\alpha)$.

The form of $\psi_\delta(t)$ and $\{a_n\}$ determine, respectively, the limiting distribution and the convergence rate of the sample quantile $\widehat{q}_n(\alpha)$ minimizing (1); in particular, we have

$$a_n(\widehat{q}_n(\alpha) - q(\alpha)) \xrightarrow{d} \psi_\delta^{\leftrightarrow}(W)$$

where $W \sim \mathcal{N}(0, \alpha(1-\alpha))$ and

$$\psi_\delta^{\leftrightarrow}(x) = \begin{cases} \inf\{t \leq 0 : \psi_\delta(t) \geq x\} & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ \sup\{t \geq 0 : \psi_\delta(t) \leq x\} & \text{if } x > 0. \end{cases} \tag{13}$$

When $\psi_\delta(t)$ is continuous (and strictly increasing) $\psi_\delta^{\leftrightarrow} = \psi_\delta^{-1}$. The scaling condition (9) for ψ_δ implies that each $\psi_\delta^{\leftrightarrow}$ satisfies a scaling condition: For each $a > 0$,

$$\psi_\delta^{\leftrightarrow}(a\,x) = a^\delta\,\psi_\delta^{\leftrightarrow}(x) \tag{14}$$

for some $\delta \geq 0$.

THEOREM 4. Suppose that $F(x)$ satisfies (12) at $x = q(\alpha)$ and

$$\widetilde{q}_n(\alpha) = g_n\left(\widehat{q}_n^{(1)}(\alpha), \cdots, \widehat{q}_n^{(k)}(\alpha)\right)$$

where

(a) $n_i/n \to \lambda_i > 0$ as $n \to \infty$ for $i = 1, \cdots, k$,

(b) $g_n(\boldsymbol{x}_n) \to g_0(\boldsymbol{x}_0)$ for all sequences $\{\boldsymbol{x}_n\}$ converging to $\boldsymbol{x}_0$, and

(c) $\{g_n\}$ and g_0 are equivariant under monotone transformations:

$$g_n(\phi(x_1), \cdots, \phi(x_k)) = \phi(g_n(x_1, \cdots, x_k))$$

for all monotone transformations ϕ.

Then

$$\sqrt{n}(\tilde{q}_n(\alpha) - q(\alpha)) \xrightarrow{d} \psi_\delta^{\leftrightarrow}(W + V)$$

where $\psi_\delta^{\leftrightarrow}$ is defined in (13), $W \sim \mathcal{N}(0, \alpha(1-\alpha))$ and V is independent of W.

The proof of Theorem 4 follows along the same lines as that of Theorem 1 noting that

$$a_n(\widehat{q}_n^{(i)}(\alpha) - q(\alpha)) \xrightarrow{d} \psi_\delta^{\leftrightarrow}\left(\lambda_i^{-1/2} W_i\right)$$

for independent $\mathcal{N}(0, \alpha(1-\alpha))$ random variables $W_1, \cdots, W_k$.

The weighted estimator $\bar{q}_n(\alpha)$ defined in (4) does not satisfy the equivariance condition in Theorem 3. For this estimator (under the assumptions of Theorem 3), we have

$$\begin{aligned} a_n(\bar{q}_n(\alpha) - q(\alpha)) &\xrightarrow{d} \sum_{i=1}^k \lambda_i \psi_\delta^{\leftrightarrow}\left(\lambda_i^{-1/2} W_i\right) \\ &= \sum_{i=1}^k \lambda_i^{1-\delta/2} \psi_\delta^{\leftrightarrow}(W_i) \end{aligned}$$

using the scaling condition (14). When $\psi_\delta^{\leftrightarrow}$ is an odd function (that is, $c_+ = c_-$ in (10) or $a_+ = a_-$ in (11)) then $E[\psi_\delta^{\leftrightarrow}(W_i)] = 0$, in which case, we can say that $\widehat{q}_n(\alpha)$ (and $\bar{q}_n(\alpha)$) are asymptotically unbiased to first order (that is, to order $O_p(a_n^{-1})$); otherwise, the first order asymptotic bias of $\widehat{q}_n(\alpha)$ is given by

$$\mu = E[\psi_\delta^{\leftrightarrow}(W)] = E[\psi_\delta^{\leftrightarrow}(W_i)] \neq 0.$$

In this case, we have

$$E\left[\sum_{i=1}^k \lambda_i^{1-\delta/2} \psi_\delta^{\leftrightarrow}(W_i)\right] = \mu \sum_{i=1}^k \lambda_i^{1-\delta/2}$$

and so the asymptotic bias of $\bar{q}_n(\alpha)$ is worse than that of $\widehat{q}_n(\alpha)$ unless $\mu = 0$ or $\delta = 0$. It follows that

$$\mathrm{Var}\left(\sum_{i=1}^k \lambda_i^{1-\delta/2} \psi_\delta^{\leftrightarrow}(W_i)\right) = \sum_{i=1}^k \lambda_i^{2-\delta} \mathrm{Var}(\psi_\delta^{\leftrightarrow}(W_i)) \quad \begin{cases} < & \mathrm{Var}(\psi_\delta^{\leftrightarrow}(W)) \quad \text{if } \delta < 1 \\ > & \mathrm{Var}(\psi_\delta^{\leftrightarrow}(W)) \quad \text{if } \delta > 1. \end{cases}$$

The fact that the variance of the limiting distribution of $a_n(\bar{q}_n(\alpha) - q(\alpha))$ tends to 0 with k when $\delta < 1$ suggests that we could improve on the sample quantile by averaging a large number of subsample quantiles; this is true provided that the bias of each subsample quantile is not too severe.

In the case of the bagged estimator, we have

$$\sqrt{n}(\tilde{q}_n(\alpha) - q(\alpha)) \xrightarrow{d} \int_{-\infty}^{\infty} \psi_\delta^{\leftrightarrow}(W + u)\phi_\alpha(u)\, du$$

while for subagging (with fraction λ) we have

$$\sqrt{n}(\tilde{q}_n(\alpha) - q(\alpha)) \xrightarrow{d} \frac{\lambda^{1/2}}{(1-\lambda)^{1/2}} \int_{-\infty}^{\infty} \psi_\delta^{\leftrightarrow}(W+u)\phi_\alpha\left(\frac{\lambda^{1/2}u}{(1-\lambda)^{1/2}}\right) du = h_{\lambda,\delta}(W).$$

In this case, we have a first order asymptotic equivalence between bagging and subagging with $\lambda = 1/2$ and note that as $\lambda \uparrow 1$, we have

$$h_{\lambda,\delta}(W) \to \psi_\delta^{\leftrightarrow}(W),$$

which is the limiting distribution of the sample quantile.

It is interesting to look at the case where ψ_δ is an odd function. Identifying $h_{1,\delta}(W) = \psi_\delta^{\leftrightarrow}(W)$, it is possible to show that for $0 < \delta < 1$ and $0 < \lambda < 1$,

$$0 \le h'_{\lambda,\delta}(W) < h'_{1,\delta}(W)$$

while for $\delta > 1$ and $0 < \lambda < 1$, we have

$$h'_{\lambda,\delta}(W) > h'_{1,\delta}(W)$$

where $h'_{\lambda,\delta}$ is the derivative of $h_{\lambda,\delta}$. Since $h_{\lambda,\delta}(0) = 0$ for all λ and δ, this implies that bagging and subagging provide (asymptotically) a contraction of the limiting distribution (compared to not using bagging or subagging) when $\delta < 1$ and an "expansion" when $\delta > 1$; likewise, when $\delta = 0$, it's easy to show that $|h_{\lambda,0}(W)| \le |\psi_0^{\leftrightarrow}(W)|$. In particular, for any non-negative "bowl shaped" function ℓ (that is, the set $\{x : \ell(x) \le y\}$ is symmetric and convex for each $y > 0$), we have for $0 < \lambda < 1$

$$\begin{aligned} E\left[\ell\left(h_{\lambda,\delta}(W)\right)\right] &< E\left[\ell\left(\psi_\delta^{\leftrightarrow}(W)\right)\right] \quad \text{if } \delta < 1, \text{ and} \\ E\left[\ell\left(h_{\lambda,\delta}(W)\right)\right] &> E\left[\ell\left(\psi_\delta^{\leftrightarrow}(W)\right)\right] \quad \text{if } \delta > 1. \end{aligned}$$

This confirms the observation made for estimators obtained by combining subsamples; when $\delta < 1$ (for example, if $f(q(\alpha)) = 0$) then the efficiency of the sample quantile can be improved by combining estimators from subsamples while when $\delta > 1$, this approach does not produce a more efficient estimator.

References

Bahadur, R.R. (1966) A note on quantiles in large samples. *Annals of Mathematical Statistics.* **37**, 577-580.

Breiman, L. (1996) Bagging predictors. *Machine Learning.* **24**, 123-140.

Bühlmann, P. and Yu, B. (2002) Analyzing bagging. *Annals of Statistics.* **30**, 927-961.

Duttweiler, D.L. (1973) The mean-square error of Bahadur's order-statistic approximation. *Annals of Statistics.* **1**, 446-453.

Kiefer, J. (1967) On Bahadur's representation of sample quantiles. *Annals of Mathematical Statistics.* **38**, 1323-1342.

Knight, K. (1998) A delta method approach to Bahadur-Kiefer theorems. *Scandinavian Journal of Statistics.* **25**, 555-568.

Knight, K. (2002) What are the limiting distributions of quantile estimators? In *Statistical Data Analysis Based on the L_1-Norm and Related Methods.* (Y. Dodge, editor) Basel: Birkhäuser.

Smirnov, N.V. (1952) Limit distributions for the terms of a variational series. *American Mathematical Society Translations.* no. 67.

Tukey, J.W. (1978) The ninther: a technique for low-effort robust (resistant) location in large samples. In *Contributions to Survey Sampling and Applied Statistics in Honor of H.O. Hartley*, ed. H.A. David. New York: Academic Press, 251-257.