LIMIT BEHAVIORS OF THE DEVIATION BETWEEN THE SAMPLE QUANTILES AND THE QUANTILE

Shoufang Xu and Yu Miao

Abstract

In this article, we discuss the limit behaviors for the deviation between the sample $p$-quantile $\hat{\xi}_{np}$ and the $p$-quantile $\xi_p$ by sampling from a sequence of independent and identically distributed samples of size $n$. The moderate deviation, large deviation and Bahadur asymptotic efficiency for $(\hat{\xi}_{np} - \xi_p)$ are established under some weak conditions.

1 Introduction

It is well known that the time series data is particularly important in the research on economic, finance, biostatistics and so on. However, time series data often are heavy-tailed, so in this case, the classical statistical analysis can not be used because of the restrictions of moment conditions. While quantile can be used for describing some properties of random variables, and there are not the restrictions of moment conditions. As a result, it is being widely employed in diverse problems in finance, such as, quantile-hedging, optimal portfolio allocation, risk management, and so on. In practice, the large sample theory which can give the asymptotic properties of sample estimator is an important method to analyze statistical problems.

To describe the results of the paper, suppose that we have an independent and identically distributed sample of size $n$ from a distribution function $F(x)$ with a continuous probability density function $f(x)$. Let $\xi_p$ denote an $p$-th quantile of $F$, i.e.,

$$\xi_p = \inf\{x : F(x) \geq p\}, \quad p \in (0, 1).$$

It is well known that there are two important estimators to estimate the $p$-th quantile: sample quantiles and order statistics. Let $F_n(x)$ denote the sample distribution function

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x\}}, \quad -\infty < x < \infty.$$

2010 Mathematics Subject Classifications. 62G30, 60F10.

Key words and Phrases. $p$-quantile, sample $p$-quantile, large deviation, moderate deviation, Bahadur asymptotic efficiency.

Received: August 28, 2010

Communicated by Miroslav M. Ristić

This work is supported by the National Natural Science Foundation of China (No. 11001077).
and \( \hat{\xi}_{np} \) denote the sample quantile, thus \( \hat{\xi}_{np} \) can be represented as

\[
\hat{\xi}_{np} = \inf\{x : F_n(x) \geq p\}, \quad p \in (0, 1).
\]

Let \( X_{(1)} \leq \cdots \leq X_{(n)} \) denote the order statistics of \( X_1, \ldots, X_n \).

There are numerous literatures to study the order statistics and the sample quantiles. Adler [1, 2] obtained some limit theorems for maximal and minimal order statistics. Park [12, 13] gave the asymptotic Fisher information in order statistics. Suppose 0 < \( p < 1 \), \( (k/n) \to p \), \( p \leq (k/n) \leq p + (1/n) \) for \( n \to \infty \), \( k \to \infty \), then in the latter limit \( X_{(k)} \) converges in probability to \( \xi_p \) (see [5, 15]). In addition, if \( F \) has a continuous first derivative \( f \) in the neighborhood of \( \xi_p \) and \( f(\xi_p) > 0 \), then

\[
\frac{\sqrt{n}f(\xi_p)(X_{(k)} - \xi_p)}{\sqrt{p(1-p)}} \to N(0,1), \quad n \to \infty,
\]

where \( N(0,1) \) denotes the standard normal random variable (see [5, 10, 15]). Assume that \( F(x) \) is twice differentiable at \( \xi_p \), with \( F'(\xi_p) = f(\xi_p) > 0 \). If \( k = np + o(\sqrt{n}\log n)^\delta \) for some \( \delta \geq 2 \), then Bahadur [3] proved

\[
X_{(k)} = \xi_p + \frac{k/n - \sum_{i=1}^n 1_{X_i \leq \xi_p}}{f(\xi_p)} + R_n, \quad \text{a.e.}
\]

where \( R_n = O(n^{-3/4}\log n)^{(1/2)(\delta+1)} \), a.e. as \( n \to \infty \). In his paper, he also raised the question of finding the exact order of \( R_n \). Further analysis by Eicker [7] revealed that \( R_n = o_p(n^{-3/4}g(n)) \) if and only if \( g(n) \to \infty \). Kiefer [8] obtained very precise details. Lahiri and Sun [9] gave a Berry-Esseen theorem for sample quantiles of strongly-mixing random variables under a polynomial mixing rate. Very recently, Miao, Chen and Xu [11] studied some precise asymptotic properties of the deviation between the order statistics and the \( p \)-quantile.

For the sample quantiles, let \( p \in (0, 1) \), if \( \xi_p \) is the unique solution \( x \) of \( F(x) = p \), then \( \hat{\xi}_{np} \xrightarrow{a.e.} \xi_p \) (see [14]). In addition, if \( F(x) \) possesses a continuous density function \( f(x) \) in a neighborhood of \( \xi_p \) and \( f(\xi_p) > 0 \), then

\[
\frac{n^{1/4}f(\xi_p)(\hat{\xi}_{np} - \xi_p)}{[p(1-p)]^{1/2}} \to N(0,1), \quad n \to \infty,
\]

where \( N(0,1) \) denotes the standard normal variable (see [15, 14]). Suppose that \( F(x) \) is twice differentiable at \( \xi_p \), with \( F''(\xi_p) = f(\xi_p) > 0 \), then Bahadur [3] proved

\[
\hat{\xi}_{np} = \xi_p + \frac{p - F_n(\xi_p)}{f(\xi_p)} + \tilde{R}_n, \quad \text{a.e.,}
\]

where \( \tilde{R}_n = O(n^{-1/4}\log n)^{1/4} \), a.e. as \( n \to \infty \).

In this paper, we are interested in the exponential convergent rate of the deviation \( (\hat{\xi}_{np} - \xi_p) \) under some suitable conditions. The moderate deviation, large deviation and Bahadur’s asymptotic efficiency of the deviation \( (\hat{\xi}_{np} - \xi_p) \) will be stated in the next section, and their proofs will be given in Section 3.
2 Main Results

First we give the following moderate deviation principle.

**Theorem 1.** Let $X_1, \cdots, X_n$ be independent identically distributed random variables with a continuous distribution $F(x)$, and let $\xi_p$ be a $p$-th quantile of $F$ for $p \in (0, 1)$. Corresponding to the sample $\{X_1, \cdots, X_n\}$, the sample $p$-th quantile which is denoted by $\hat{\xi}_{pn}$ is defined as the $p$-th quantile of the sample distribution function $F_n(x)$. Assume that $F(x)$ has a continuous density function $f(x)$ in the neighborhood of $\xi_p$ and $f(\xi_p) > 0$. In addition, let $\{b_n\}$ be a positive sequence satisfying $b_n \to \infty$ and $\frac{b_n}{\sqrt{n}} \to 0$, as $n \to \infty$.

Then for any $r > 0$, we have

$$
\lim_{n \to \infty} \frac{1}{b_n} \log P \left( \frac{\sqrt{n}}{b_n} |\hat{\xi}_{pn} - \xi_p| \geq r \right) = -\frac{f(\xi_p)^2 r^2}{2p(1 - p)}.
$$

The following result is the large deviation principle.

**Theorem 2.** Let $X_1, \cdots, X_n$ be independent identically distributed random variables with a continuous distribution $F(x)$, and let $\xi_p$ be a $p$-th quantile of $F$ for $p \in (0, 1)$. Corresponding to the sample $\{X_1, \cdots, X_n\}$, the sample $p$-th quantile which is denoted by $\hat{\xi}_{pn}$ is defined as the $p$-th quantile of the sample distribution function $F_n(x)$. Then for any $r > 0$, we have

$$
\lim_{n \to \infty} \frac{1}{n} \log P \left( |\hat{\xi}_{pn} - \xi_p| \geq r \right) = -\inf_{x \geq 1 - p} \Lambda^*_+(x)
$$

and

$$
\lim_{n \to \infty} \frac{1}{n} \log P \left( |\hat{\xi}_{pn} - \xi_p| \leq -r \right) = -\inf_{x \geq p} \Lambda^*_-(x)
$$

where

$$
\Lambda^*_+(x) = x \log \frac{x}{1 - F(\xi_p + r)} + (1 - x) \log \frac{1 - x}{F(\xi_p + r)}
$$

and

$$
\Lambda^*_-(x) = x \log \frac{x}{F(\xi_p - r)} + (1 - x) \log \frac{1 - x}{1 - F(\xi_p - r)}
$$

for $x \in [0, 1]$, and $\Lambda^*_-(x) = \Lambda^*_+(x) = \infty$, for $x \notin [0, 1]$. In particular, it follows that

$$
\lim_{n \to \infty} \frac{1}{n} \log P \left( |\hat{\xi}_{pn} - \xi_p| \geq r \right) = -\min \left\{ \inf_{x \geq 1 - p} \Lambda^*_+(x), \inf_{x \geq p} \Lambda^*_-(x) \right\}.
$$

**Remark 1.** It is not difficult to check that

$$
\inf_{x \geq 1 - p} \Lambda^*_+(x) = (1 - p) \log \frac{1 - p}{1 - F(\xi_p + r)} + p \log \frac{p}{F(\xi_p + r)}
$$
and
\[ \inf_{x \geq p} \Lambda^*(x) = p \log \frac{p}{F(\xi_p - r)} + (1 - p) \log \frac{1 - p}{1 - F(\xi_p - r)}. \]

**Theorem 3.** Under the conditions of Theorem 2, assume that \( F(x) \) has a continuous first derivative \( f(x) \) in the neighborhood of \( \xi_p \) and \( f(\xi_p) > 0 \), then we have the following Bahadur’s asymptotic efficiency
\[
\lim_{r \to 0} \lim_{n \to \infty} \frac{1}{r^2} \frac{1}{n} \log P(\hat{\xi}_{pn} - \xi_p \geq r)
= -\lim_{r \to 0} \frac{1}{r^2} \inf_{x \geq 1 - p} \Lambda^*(x) = -\frac{f(\xi_p)^2}{2p(1 - p)}
\]
and
\[
\lim_{r \to 0} \lim_{n \to \infty} \frac{1}{r^2} \frac{1}{n} \log P(\hat{\xi}_{pn} - \xi_p \leq -r)
= -\lim_{r \to 0} \frac{1}{r^2} \inf_{x \geq p} \Lambda^*(x) = -\frac{f(\xi_p)^2}{2p(1 - p)}.
\]

### 3 Proofs of Main Results

The following lemma will be applied in our proof.

**Lemma 1.** [14] Let \( F \) be a distribution function. The function \( F^{-1}(t), 0 < t < 1 \), is nondecreasing and continuous, and satisfies
\[ F^{-1}(F(x)) \leq x, \quad x \in (-\infty, +\infty), \]

and
\[ F(F^{-1}(t)) \geq t, \quad t \in (0, 1). \]

Hence we have
\[ F(x) \geq t \iff x \geq F^{-1}(t). \]

**Proof of Theorem 1.** For any \( r > 0 \), we have
\[
P\left( \frac{\sqrt{n}}{b_n} [\hat{\xi}_{pn} - \xi_p] \geq r \right) = P\left( \hat{\xi}_{pn} \geq \xi_p + \frac{b_n r}{\sqrt{n}} \right) + P\left( \hat{\xi}_{pn} \leq \xi_p - \frac{b_n r}{\sqrt{n}} \right). \tag{1}
\]
By Lemma 1,
\[
P\left( \hat{\xi}_{pn} \geq \xi_p + \frac{b_n r}{\sqrt{n}} \right) = P\left( \hat{p} \geq F_n\left( \xi_p + \frac{b_n r}{\sqrt{n}} \right) \right). \tag{2}
\]
Thus, we give the following form firstly

\[
P\left(\frac{\sqrt{n}}{b_n}(\hat{\xi}_{pn} - \xi_p) \geq r\right) = P\left(\hat{\xi}_{pn} \geq \xi_p + \frac{b_n r}{\sqrt{n}}\right)
\]

\[
= P\left(\frac{1}{n} \sum_{i=1}^{n} I\{X_i \leq \frac{b_n r}{\sqrt{n}} + \xi_p\} \leq p\right)
\]

\[
= P\left(\sum_{i=1}^{n} I\{X_i \geq \frac{b_n r}{\sqrt{n}} + \xi_p\} \geq n(1 - p)\right) = P\left(\sum_{i=1}^{n} [I\{X_i \geq \frac{b_n r}{\sqrt{n}} + \xi_p\} - EI\{X_i \geq \frac{b_n r}{\sqrt{n}} + \xi_p\}] \geq n(1 - p) - nEI\{X_i \geq \frac{b_n r}{\sqrt{n}} + \xi_p\}\right)
\]

\[
= P\left(\sum_{i=1}^{n} W_{ni} \geq b_n \sqrt{n} \delta_1\right),
\]

where

\[
W_{ni} = I\{X_i \geq \frac{b_n r}{\sqrt{n}} + \xi_p\} - EI\{X_i \geq \frac{b_n r}{\sqrt{n}} + \xi_p\}
\]

and

\[
\delta_1 = \frac{n(1 - p) - nEI\{X_i \geq \frac{b_n r}{\sqrt{n}} + \xi_p\}}{b_n \sqrt{n}}.
\]

Hence it is easy to check

\[
EI\{X_i \geq \frac{b_n r}{\sqrt{n}} + \xi_p\} = P\left(X_i \geq \frac{b_n r}{\sqrt{n}} + \xi_p\right) = 1 - F\left(\frac{b_n r}{\sqrt{n}} + \xi_p\right)
\]

and by utilizing Taylor’s theorem we have

\[
F\left(\frac{b_n r}{\sqrt{n}} + \xi_p\right) = F(\xi_p) + F'(\xi_p) \frac{b_n r}{\sqrt{n}} + o\left(\frac{b_n}{\sqrt{n}}\right)
\]

\[
= p + f(\xi_p) \frac{b_n r}{\sqrt{n}} + o\left(\frac{b_n}{\sqrt{n}}\right).
\]

From (4), (5), we have

\[
\delta_1 = \frac{n(1 - p) - n(1 - p) + f(\xi_p) b_n \sqrt{n} r + o(b_n \sqrt{n})}{b_n \sqrt{n}} = f(\xi_p) r + o(1)
\]

and

\[
W_{ni} = I\{X_i \geq \xi_p + \frac{b_n r}{\sqrt{n}}\} - 1 + F\left(\xi_p + \frac{b_n r}{\sqrt{n}}\right),
\]

so it is easy to have

\[
E(W_{ni}) = 0, \ Var(W_{ni}) = p(1 - p) + O(b_n / \sqrt{n}).
\]
Through the above discussions, the equation (3) can be rewritten as follows

\[ P \left( \hat{\xi}_p + b_n r \geq \sqrt{n} \right) = P \left( \frac{1}{b_n} \sum_{i=1}^{n} W_{ni} \geq f(\xi_p)r + o(1) \right). \]

Now we need give Cramér function of the random variable \( \sum_{i=1}^{n} W_{ni} \), i.e., for any \( \lambda \in \mathbb{R} \), by Taylor’s theorem, we have

\[
A(\lambda) = \lim_{n \to \infty} \frac{1}{b_n^2} \log E \exp \left\{ \frac{\lambda b_n}{\sqrt{n}} \sum_{i=1}^{n} W_{ni} \right\} = \lim_{n \to \infty} \frac{1}{b_n^2} \log \left( E \exp \left\{ \frac{\lambda b_n^2}{n} \frac{2}{2} + o \left( \frac{b_n^2}{n} \right) \right\} \right)^n = \frac{\lambda^2 p(1 - p)}{2}.
\]

By the Gärtner-Ellis theorem (see [4, 6]), we have

\[
\lim_{n \to \infty} \frac{1}{b_n^2} \log P \left( \sum_{i=1}^{n} W_{ni} \geq b_n \sqrt{n} \delta_1 \right) = -\frac{f(\xi_p)^2r^2}{2p(1 - p)}
\]

which implies the following result

\[
\lim_{n \to \infty} \frac{1}{b_n^2} \log P \left( \sqrt{n} b_n (\hat{\xi}_p - \xi_p) \geq r \right) = -\frac{f(\xi_p)^2r^2}{2p(1 - p)}. \tag{10}
\]

Likewise, by Lemma 1, we have

\[
P \left( \hat{\xi}_p \leq \xi_p - b_n r \sqrt{n} \right) = P \left( \hat{\xi}_p \leq \xi_p - b_n r \right), \tag{11}
\]

so, we can give the following form

\[
P \left( \sqrt{n} b_n (\hat{\xi}_p - \xi_p) \leq -r \right) = P \left( \hat{\xi}_p \leq \xi_p - b_n r \sqrt{n} \right)
\]

\[
= P \left( \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq \xi_p - \frac{b_n r}{\sqrt{n}}\}} \geq p \right)
\]

\[
= P \left( \sum_{i=1}^{n} I_{\{X_i \leq \xi_p - \frac{b_n r}{\sqrt{n}}\}} \geq np \right)
\]

\[
= P \left( \sum_{i=1}^{n} \left[ I_{\{X_i \leq \xi_p - \frac{b_n r}{\sqrt{n}}\}} - EI_{\{X_i \leq \xi_p - \frac{b_n r}{\sqrt{n}}\}} \right] \geq np - nEI_{\{X_i \leq \xi_p - \frac{b_n r}{\sqrt{n}}\}} \right)
\]

\[
= P \left( \sum_{i=1}^{n} V_{ni} \geq b_n \sqrt{n} \delta_2 \right). \tag{12}
\]
The deviation between the sample quantiles and the quantile

\[ V_{ni} = I_{\{X_i \leq \xi_p - \frac{b_n r}{\sqrt{n}}\}} - EI_{\{X_i \leq \xi_p - \frac{b_n r}{\sqrt{n}}\}}, \]

\[ \delta_2 = \frac{np - nEI_{\{X_i \leq \xi_p - \frac{b_n r}{\sqrt{n}}\}}}{b_n \sqrt{n}}. \]

According to a simple calculation, we have

\[ EI_{\{X_i \leq \xi_p - \frac{b_n r}{\sqrt{n}}\}} = P(X_i \leq \xi_p - \frac{b_n r}{\sqrt{n}}) = F(\xi_p - \frac{b_n r}{\sqrt{n}}) \quad (13) \]

and

\[ F(\xi_p - \frac{b_n r}{\sqrt{n}}) = F(\xi_p) - F'(\xi_p) \frac{b_n r}{\sqrt{n}} + o\left(\frac{b_n}{\sqrt{n}}\right). \quad (14) \]

Thus, by (13), (14) we see that

\[ \delta_2 = \frac{np - np + f(\xi_p)b_n \sqrt{n}r + o(b_n \sqrt{n})}{b_n \sqrt{n}} = f(\xi_p)r + o(1), \]

and

\[ V_{ni} = I_{\{X_i \leq \xi_p - \frac{b_n r}{\sqrt{n}}\}} - F(\xi_p - \frac{b_n r}{\sqrt{n}}). \quad (16) \]

Obviously

\[ E(V_{ni}) = 0, \quad Var(V_{ni}) = p(1-p) + O(b_n/\sqrt{n}). \quad (17) \]

So the equation (12) can be rewritten as follows

\[ P\left(\hat{\xi}_{pn} \leq \xi_p - \frac{b_n r}{\sqrt{n}}\right) = P\left(\frac{1}{b_n \sqrt{n}} \sum_{i=1}^{n} V_{ni} \geq f(\xi_p)r + o(1)\right). \]

Then we give Cramér function of the random variable \( \sum_{i=1}^{n} V_{ni} \): for any \( \lambda \in \mathbb{R} \)

\[ \Lambda(\lambda) = \lim_{n \to \infty} \frac{1}{b_n^2} \log E \exp \left\{ \frac{\lambda b_n}{\sqrt{n}} \sum_{i=1}^{n} V_{ni} \right\} \]

\[ = \lim_{n \to \infty} \frac{1}{b_n^2} \log \left( E \exp \left\{ \frac{\lambda b_n}{\sqrt{n}} V_{ni} \right\} \right)^n \]

\[ = \lim_{n \to \infty} \frac{1}{b_n^2} \log \left( 1 + \frac{\lambda^2 b_n^2}{n} E(V_{ni}^2) + o\left(\frac{b_n^2}{n}\right) \right)^n \]

\[ = \lambda^2 p(1-p) \]

(18)

By the similar proof of (10), we have

\[ \lim_{n \to \infty} \frac{1}{b_n^2} \log P\left(\sqrt{n} \left(\hat{\xi}_{pn} - \xi_p\right) \leq -r\right) = -\frac{f(\xi_p)^2 r^2}{2p(1-p)} \]

(19)
According to (10), (19), we have the following moderate deviation principle
\[
\lim_{n \to \infty} \frac{1}{b_n^2} \log P \left( \frac{\sqrt{n}}{b_n} |\hat{\xi}_p - \xi_p| \geq r \right) = -\frac{f(\xi_p)^2 r^2}{2p(1-p)}.
\]

**Proof of Theorem 2.** As the same as the proof of Theorem 1, we have
\[
P(\hat{\xi}_p n - \xi_p \geq r) = P \left( p \geq F_n(\xi_p + r) \right) = P \left( \sum_{i=1}^{n} I_{\{x_i \geq \xi_p + r\}} \geq n(1-p) \right),
\]
where
\[
U_{ni} = I_{\{x_i \geq \xi_p + r\}}.
\]
For any \( \lambda \in \mathbb{R} \), the Cramér functional of \( U_{ni} \) is
\[
\Lambda(\lambda) = \log E e^{\lambda U_{ni}} = \log(e^{\lambda [1 - F(\xi_p + r)]} + F(\xi_p + r))
\]
and the Fenchel-Legendre transform of \( \Lambda(\lambda) \) is
\[
\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} \{ \lambda x - \Lambda(\lambda) \}
\]
if \( x \in [0, 1] \), then
\[
\Lambda^*(x) = x \log \frac{x}{1 - F(\xi_p + r)} + (1 - x) \log \frac{1 - x}{F(\xi_p + r)},
\]
and, if \( x \notin [0, 1] \), then \( \Lambda^*(x) = \infty \).

Then by Cramér Theorem (see [4, 6]), we have
\[
\lim_{n \to \infty} \frac{1}{b_n^2} \log P \left( \sum_{i=1}^{n} U_{ni} \geq n(1-p) \right) = -\inf_{x \geq p} \Lambda^*_+(x),
\]
that is
\[
\lim_{n \to \infty} \frac{1}{b_n^2} \log P \left( \hat{\xi}_p n - \xi_p \geq r \right) = -\inf_{x \geq 1-p} \Lambda^*_+(x).
\]
Likewise,
\[
\lim_{n \to \infty} \frac{1}{b_n^2} \log P \left( \hat{\xi}_p n - \xi_p \leq -r \right) = -\inf_{x \geq p} \Lambda^*_+(x).
\]
So the proof of the theorem is completed.
The deviation between the sample quantiles and the quantile

**Proof of Theorem 3.** By Theorem 2, it is enough to show

\[ \lim_{r \to 0} \frac{1}{r^2} \inf_{x \geq 1-p} \Lambda^*_r(x) = \frac{f(\xi_p)^2}{2p(1-p)} \]  

(21)

and

\[ \lim_{r \to 0} \frac{1}{r^2} \inf_{x \geq p} \Lambda^*_r(x) = \frac{f(\xi_p)^2}{2p(1-p)} \]  

(22)

Here we only give the proof of (21) and the proof of (22) is similar. From Remark 1, we know

\[
\frac{1}{r^2} \inf_{x \geq 1-p} \Lambda^*_r(x) = \frac{1}{r^2} \left( (1 - p) \log \frac{1 - p}{1 - F(\xi_p + r)} + p \log \frac{p}{F(\xi_p + r)} \right).
\]

Since

\[ F(\xi_p + r) = F(\xi_p) + F'(\xi_p)r + o(r) = p + f(\xi_p)r + o(r), \]

then by Hospital’s rule, we can obtain

\[
\lim_{r \to 0} \frac{1}{r^2} \left( (1 - p) \log \frac{1 - p}{1 - F(\xi_p + r)} + p \log \frac{p}{F(\xi_p + r)} \right) = \frac{f^2(\xi_p)r}{2p(1 - p)}.
\]

Hence we obtain our desired results.  

\[ \square \]

**References**


Shoufang Xu:
Department of Mathematics, Xinxiang University, Henan Province, 453000, China.
E-mail: xuxu8803@yahoo.cn

Yu Miao:
College of Mathematics and Information Science, Henan Normal University, Henan Province, 453007, China.
E-mail: yumiao728@yahoo.com.cn