SOME LIMIT BEHAVIOR FOR LINEAR COMBINATIONS OF ORDER STATISTICS

YU MIAO AND MENGYAO MA

In the present paper, we establish the moderate and large deviations for the linear combinations of uniform order statistics. As applications, the moderate and large deviations for the $k$-th order statistics from uniform distribution, Gini mean difference statistics and the $k$-th order statistics from general continuous distribution are obtained.

Keywords: linear combinations of order statistics, large deviation, moderate deviation, Gini mean difference statistics

Classification: 62G30

1. INTRODUCTION

Consider independent observations $X_1, X_2, \ldots, X_n$ on a distribution function $F$, and denote the ordered values by $X_{n,1} \leq X_{n,2} \leq \cdots \leq X_{n,n}$. Many important statistics may be expressed as linear combinations of the ordered values, that is, in the form

$$T_n = \sum_{i=1}^{n} c_{n,i} X_{n,i}$$

(1.1)

for some choice of constants $c_{n,1}, c_{n,2}, \ldots, c_{n,n}$. $T_n$ is also called $L$-statistics, which may have the following more general form:

$$T_n = \sum_{i=1}^{n} c_{n,i} h(X_{n,i})$$

where $h$ is some measurable function.

The asymptotic behavior and applications of $T_n$ have been studied widely. Law of large numbers: Wellner [30, 31] proved a strengthened version of the Glivenko–Cantelli theorem for the uniform empirical distribution function, by which the law of large numbers for linear functions of order statistics is established. Sen [23] gave the almost sure convergence of certain functions of order statistics having some special properties. van Zwet [28] obtained a strong law of large numbers for linear combinations of order statistics under integrability conditions only, which generalized previous results of Wellner.
Some limit behavior for linear combinations of order statistics

Mason [18] gave some necessary and sufficient conditions for strong law of large numbers to hold for certain classes of linear functions of order statistics. Helmers et al. [14] obtained the strong convergence of generalized $L$-statistics. Aaronson et al. [1] studied the strong law of large numbers for $L$-statistics for ergodic stationary processes. Central limit theorem: Chernoff et al. [7] gave the asymptotic distribution of linear combinations of functions of order statistics. Stigler [25, 26] proved a central limit theorem by using Hájek projections. Bjerve [4], Helmers [12, 13] established the Berry-Esseen-type bounds for linear combinations of order statistics. A very complete version of the central limit theorem with necessary and sufficient conditions is proved in Mason and Shorack [19, 20], via empirical processes theory. For weaker conditions on the function $h$, a central limit theorem and a law of the iterated logarithm can be found in Li et al. [15].


In the present paper, we consider the moderate and large deviations for the linear combinations of uniform order statistics. As applications, the moderate and large deviations for the $k$-th order statistics from uniform distribution, Gini mean difference statistics and the $k$-th order statistics from general continuous distribution are obtained. In Section 2, we give some known results and lemmas. Our main results are stated in Section 3 and the applications are given in Section 4.

2. SOME KNOWN RESULTS AND LEMMAS

In this section, we give some known results and lemmas for order statistics. Let $U_1, U_2, \ldots, U_n$ be a sequence of random variables with the uniform distribution on the interval $(0, 1)$ and $U_{n,1} \leq U_{n,2} \leq \cdots \leq U_{n,n}$ be the corresponding order statistics. Let $\beta(p,q)$, where $p, q > 0$, denote the Beta distribution with parameters $(p,q)$, i.e., the density function of $\beta(p,q)$ is

$$f_{\beta(p,q)}(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}x^{p-1}(1-x)^{q-1}, \quad x \in (0,1).$$

From (2.1) and the following equality,

$$\int_0^1 x^{m+k-1}(1-x)^{n-k} \, dx = \frac{(n-k)!(m+k-1)!}{(m+n)!},$$

where $m$ is a positive integer, we can obtain the $m$-th moment of $U_{n,k}$

$$\mathbb{E}U_{n,k}^m = \frac{n!(m+k-1)!}{(k-1)!(m+n)!},$$
which yields the expectation and the variance of $U_{n,k}$,

$$
\mathbb{E}U_{n,k} = \frac{k}{n+1}, \quad \text{Var}(U_{n,k}) = \frac{k(n-k+1)}{(n+1)^2(n+2)},
$$

(2.2)

Let us recall the Gamma distribution with parameters $(a, \lambda)$, i.e., if $X \sim \Gamma(a, \lambda)$, then its density function is

$$
f_{\Gamma(a,\lambda)}(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, \quad x > 0
$$

(2.3)

and its characteristic function is

$$
\phi(t) = \left( \frac{\lambda}{\lambda - it} \right)^a.
$$

It is well known that

$$
U_{n,k} \overset{d}{=} \frac{X_1 + \cdots + X_k}{X_1 + \cdots + X_{n+1}}
$$

and

$$
U_{n,k} \overset{d}{=} \frac{Y_1^2 + \cdots + Y_{2k}^2}{Y_1^2 + \cdots + Y_{2(n+1)}^2},
$$

where $X_1, X_2, \ldots, X_{n+1}$ is a sequence of i.i.d. random variables with standard exponential distribution and $Y_1, Y_2, \ldots, Y_{n+1}$ is a sequence of i.i.d. random variables with standard normal distribution.

The following results, which was obtained by Plachky and Steinebach [22], show that the one-sided large deviation principle holds.

**Lemma 2.1.** (Plachky and Steinebach [22]) Let $\{W_n, n \geq 1\}$ be a sequence of real-valued random variables on a probability space with probability measure $\mathbb{P}$, which satisfies the following assumptions:

1. for all $t \in [0, T_1), T_1 > 0$,

$$
\int e^{tW_n} \, d\mathbb{P} < \infty;
$$

2. for $0 \leq T_0 < T_1$ and for all $t \in (T_0, T_1)$,

$$
\frac{1}{n} \log \int e^{tW_n} \, d\mathbb{P} < \infty \to c_0(t) \in \mathbb{R}.
$$

Then for any real sequence $\{d_n, n \geq 1\}$ with $d_n \to d \in A$, where

$$
A = \left\{ c'_0(h) : c'_0(h) \text{ exists and is continuous on} \right.
$$

the right and strictly monotonic for $h \in (T_0, T_1) \left\}, \right.
$$

it holds that

$$
\frac{1}{n} \log \mathbb{P}(W_n > nd_n) \to \exp (c_0(h) - hd)
$$

where the limit is equal to $\inf_{t > 0} \{\exp(c_0(t) - td)\}.$
Lemma 2.2. (Boistard [3]) Let $b$ be a continuous function on $[0, 1]$ and $b_{n,k}$ be some coefficients such that
\[
\lim_{n \to \infty} \max_{1 \leq k \leq n} \left| b_{n,k} - b \left( \frac{k}{n} \right) \right| = 0.
\]
Suppose that $1 - b(t) > 0$ for all $t$. Then for large $n$, $b_{n,k} < 1$, $1 \leq k \leq n$ and
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \log(1 - b_{n,k}) = \int_{0}^{1} \log(1 - b(t)) \, dt.
\]

3. MODERATE AND LARGE DEVIATION PRINCIPLES FOR UNIFORM RANDOM VARIABLES

In this section, we establish the moderate and large deviation principles for the linear combinations of uniform order statistics.

Theorem 3.1. Let $U_1, U_2, \ldots, U_n$ be a sequence of random variables with the uniform distribution on the interval $(0, 1)$ and $U_{n,1} \leq U_{n,2} \leq \cdots \leq U_{n,n}$ be the corresponding order statistics. Let $T_n = \sum_{i=1}^{n} c_{n,i} U_{n,i}$, where $\{c_{n,i}, 1 \leq i \leq n\}$ is an array of constants. Suppose that $\{b_n, n \geq 1\}$ is a sequence of positive numbers such that
\[
b_n \to \infty, \quad \frac{b_n}{\sqrt{n}} \to 0, \quad \frac{b_n}{\sqrt{n}} \max_{1 \leq k \leq n} |B_{n,k} - D_n| \to 0,
\]
and
\[
\frac{1}{n} \sum_{k=1}^{n} B_{n,k}^2 - D_n^2 \to \sigma^2,
\]
where
\[
B_{n,k} = \sum_{i=k}^{n} c_{n,i}, \quad D_n = \frac{1}{n+1} \sum_{k=1}^{n} B_{n,k}
\]
and $\sigma^2$ is a positive constant. Then for any $r > 0$, we have
\[
\lim_{n \to \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \frac{\sqrt{n}}{b_n} |T_n - ET_n| > r \right) = -\frac{r^2}{2\sigma^2}.
\]

Proof. It is easy to check
\[
T_n = \sum_{i=1}^{n} c_{n,i} U_{n,i} = \sum_{i=1}^{n} c_{n,i} \frac{X_1 + \cdots + X_i}{X_1 + \cdots + X_{n+1}} = \sum_{k=1}^{n} \frac{B_{n,k} X_k}{X_1 + \cdots + X_{n+1}}.
\]

By defining $B_{n,n+1} = 0$, then from (2.2), we have
\[
ET_n = \frac{1}{n+1} \sum_{i=1}^{n} i c_{n,i} = \frac{1}{n+1} \sum_{i=1}^{n+1} B_{n,i}.
\]
So we can rewrite $\frac{\sqrt{n}}{b_n}(T_n - ET_n)$ as follows

$$\frac{\sqrt{n}}{b_n}(T_n - ET_n) = \frac{\sqrt{n}}{b_n(n + 1)} \frac{\sum_{k=1}^{n+1} B_{n,k}X_k - \sum_{k=1}^{n+1} B_{n,k} \frac{\sum_{k=1}^{n+1} X_k}{n+1}}{\sum_{k=1}^{n+1} X_k}. \quad (3.2)$$

Let us define

$$\tilde{T}_n := \sum_{k=1}^{n+1} (B_{n,k} - D_n)X_k$$

then for any $\lambda \in \mathbb{R}$, from the condition

$$\frac{1}{b_n} \log \mathbb{E} \exp \left( \frac{\lambda b_n \sqrt{n} \tilde{T}_n}{\sqrt{n}} \right) \to \lambda^2 \sigma^2$$

we have

$$\frac{1}{b_n^2} \log \mathbb{E} \exp \left( \frac{\lambda b_n \sqrt{n} \sum_{k=1}^{n+1} (B_{n,k} - D_n)X_k}{\sqrt{n}} \right)$$

$$= \frac{1}{b_n^2} \log \mathbb{E} \exp \left( \frac{\lambda b_n \sqrt{n} \sum_{k=1}^{n+1} (B_{n,k} - D_n)X_k}{\sqrt{n}} \right)$$

$$= \frac{1}{b_n^2} \sum_{k=1}^{n+1} \log \mathbb{E} \exp \left( \frac{\lambda b_n (B_{n,k} - D_n)X_k}{\sqrt{n}} \right)$$

$$= - \frac{1}{b_n^2} \sum_{k=1}^{n+1} \log \left( 1 - \frac{\lambda b_n (B_{n,k} - D_n)}{\sqrt{n}} \right)$$

$$= \frac{\lambda}{b_n \sqrt{n}} \sum_{k=1}^{n+1} (B_{n,k} - D_n) + \frac{\lambda^2}{2n} \sum_{k=1}^{n+1} (B_{n,k} - D_n)^2$$

$$\quad + O \left( \frac{b_n}{n^{3/2}} \sum_{k=1}^{n+1} (B_{n,k} - D_n)^3 \right).$$

It is easy to check that

$$\sum_{k=1}^{n+1} (B_{n,k} - D_n) = 0 \quad \text{and} \quad \frac{1}{n} \sum_{k=1}^{n+1} (B_{n,k} - D_n)^2 = \frac{1}{n} \sum_{k=1}^{n+1} B_{n,k}^2 - \frac{n+1}{n} D_n^2 \rightarrow \sigma^2.$$

So we have

$$\frac{1}{b_n^2} \log \mathbb{E} \exp \left( \frac{\lambda b_n \sqrt{n} \tilde{T}_n}{\sqrt{n}} \right) \to \frac{\lambda^2 \sigma^2}{2}$$

which implies, by the Gärtner-Ellis theorem [9], that for any $r > 0$,

$$\frac{1}{b_n^2} \log \mathbb{P} \left( \frac{1}{b_n \sqrt{n}} \left| \tilde{T}_n \right| \geq r \right) \to - \frac{r^2}{2\sigma^2}. \quad (3.3)$$
Furthermore, for any \( \varepsilon > 0 \), we give the following exponential inequalities: for any \( 0 < \lambda < 1 \),
\[
\mathbb{P} \left( \frac{1}{n+1} \sum_{k=1}^{n+1} (X_k - 1) > \varepsilon \right) \leq \left( \frac{1}{1 - \lambda} e^{-\lambda(1+\varepsilon)} \right)^{n+1}
\]
which, by taking \( \lambda = \frac{\varepsilon}{1+\varepsilon} \), yields
\[
\mathbb{P} \left( \frac{1}{n+1} \sum_{k=1}^{n+1} (X_k - 1) > \varepsilon \right) \leq \left( (1 + \varepsilon) e^{-\varepsilon} \right)^{n+1}.
\tag{3.4}
\]
By similar proofs, for any \( 0 < \varepsilon < 1 \) and \( \lambda > 0 \), we have
\[
\mathbb{P} \left( \frac{1}{n+1} \sum_{k=1}^{n+1} (X_k - 1) < -\varepsilon \right) \leq \left( \frac{1}{1 + \lambda} e^{\lambda(1-\varepsilon)} \right)^{n+1}
\]
which, by taking \( \lambda = \frac{\varepsilon}{1-\varepsilon} \), yields
\[
\mathbb{P} \left( \frac{1}{n+1} \sum_{k=1}^{n+1} (X_k - 1) < -\varepsilon \right) \leq \left( (1 - \varepsilon) e^{\varepsilon} \right)^{n+1}.
\tag{3.5}
\]
From the inequalities (3.4), (3.5) and the following inequalities: for any \( 0 < \varepsilon < 1 \)
\[
\log(1 + \varepsilon) < \varepsilon - \frac{\varepsilon^2}{2(1+\varepsilon)}, \quad \log(1 - \varepsilon) < -\varepsilon - \frac{\varepsilon^2}{2},
\]
we have
\[
\frac{1}{b_n^2} \log \mathbb{P} \left( \frac{1}{n+1} \sum_{k=1}^{n+1} (X_k - 1) > \varepsilon \right) \rightarrow -\infty.
\tag{3.6}
\]
Hence for any \( r > 0 \) and \( 0 < \varepsilon < 1 \), we get
\[
\mathbb{P} \left( \frac{\sqrt{n}}{b_n} |T_n - ET_n| > r \right) \\
\leq \mathbb{P} \left( \frac{1}{n+1} \sum_{k=1}^{n+1} (X_k - 1) > \varepsilon \right) \\
+ \mathbb{P} \left( \frac{\sqrt{n}}{b_n} |T_n - ET_n| > r, \frac{1}{n+1} \sum_{k=1}^{n+1} (X_k - 1) \leq \varepsilon \right) \\
\leq \mathbb{P} \left( \frac{1}{n+1} \sum_{k=1}^{n+1} (X_k - 1) > \varepsilon \right) + \mathbb{P} \left( \frac{1}{b_n \sqrt{n}} \tilde{T}_n \geq r(1 - \varepsilon) \right)
\]
which implies by (3.3) and (3.6),
\[
\limsup_{n \to \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \frac{\sqrt{n}}{b_n} |T_n - ET_n| > r \right) \leq - \frac{r^2(1 - \varepsilon)^2}{2\sigma^2}.
\tag{3.7}
\]
Similarly we have
\[
\liminf_{n \to \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \frac{\sqrt{n}}{b_n} \left| T_n - \mathbb{E} T_n \right| > r \right) \\
\geq \liminf_{n \to \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \frac{\sqrt{n}}{b_n} \left| T_n - \mathbb{E} T_n \right| > r, \frac{1}{n + 1} \sum_{k=1}^{n+1} (X_k - 1) \leq \varepsilon \right) \\
\geq \liminf_{n \to \infty} \frac{1}{b_n^2} \log \left[ \mathbb{P} \left( \frac{1}{b_n} \sqrt{n} \left| \tilde{T}_n \right| \geq r(1 + \varepsilon) \right) - \mathbb{P} \left( \frac{1}{n + 1} \sum_{k=1}^{n+1} (X_k - 1) > \varepsilon \right) \right] \\
= - \frac{r^2 (1 + \varepsilon)^2}{2\sigma^2}. 
\]

From (3.7) and (3.8) and by the arbitrariness of \( \varepsilon \), Theorem 3.1 can be proved. \( \square \)

The following results are about the moderate deviation principle for the trimmed sums.

**Theorem 3.2.** Let \( U_1, U_2, \ldots, U_n \) be a sequence of random variables with the uniform distribution on the interval \( (0, 1) \) and \( U_{n,1} \leq U_{n,2} \leq \cdots \leq U_{n,n} \) be the corresponding order statistics. Let
\[
T_n = \sum_{i=\alpha_n}^{\beta_n} c_{n,i} U_{n,i},
\]
where \( \{ c_{n,i}, 1 \leq i \leq n \} \) is an array of constants, \( 0 < \alpha_n < \beta_n \leq n \) are some integers. Suppose that \( \{ b_n, n \geq 1 \} \) is a sequence of positive numbers such that
\[
b_n \to \infty, \quad \frac{b_n}{\sqrt{n}} \to 0, \quad \frac{b_n}{\sqrt{n}} \max_{1 \leq k \leq n} |D_{n,k}| \to 0,
\]
and
\[
\frac{1}{n} \left[ (\alpha_n - 1)B_{n,\alpha_n}^2 + \sum_{k=\alpha_n}^{\beta_n} B_{n,k}^2 \right] - \frac{n+1}{n} D_n^2 \to \sigma^2,
\]
where \( B_{n,k} = \sum_{i=k}^{\beta_n} c_{n,i} \) and \( \sigma^2 \) is a positive constant,
\[
D_n = \frac{\sum_{k=\alpha_n}^{\beta_n} B_{n,k} + (\alpha_n - 1)B_{n,\alpha_n}}{n+1}
\]
and
\[
\tilde{D}_{n,k} = \begin{cases} 
B_{n,\alpha_n} - D_n, & 1 \leq k \leq \alpha_n - 1 \\
B_{n,k} - D_n, & \alpha_n \leq k \leq \beta_n \\
-D_n, & \beta_n + 1 \leq k \leq n + 1
\end{cases}.
\]

Then for any \( r > 0 \), we have
\[
\lim_{n \to \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \frac{\sqrt{n}}{b_n} \left| T_n - \mathbb{E} T_n \right| > r \right) = - \frac{r^2}{2\sigma^2}.
\]
Proof. It is easy to check
\[ T_n = \sum_{i=\alpha_n}^{\beta_n} c_{n,i} U_{n,i} = \sum_{i=\alpha_n}^{\beta_n} c_{n,i} \frac{X_1 + \cdots + X_i}{X_1 + \cdots + X_{n+1}} \]
\[ = S_{n+1}^{-1} \left[ B_{n,\alpha_n} \sum_{k=1}^{\alpha_n-1} X_k + \sum_{k=\alpha_n}^{\beta_n} B_{n,k} X_k \right], \tag{3.9} \]
where \( S_{n+1} = X_1 + \cdots + X_{n+1} \), and from (2.2), we have
\[ ET_n = \frac{1}{n+1} \sum_{i=\alpha_n}^{\beta_n} i c_{n,i} = \frac{1}{n+1} \left[ (\alpha_n - 1) B_{n,\alpha_n} + \sum_{k=\alpha_n}^{\beta_n} B_{n,k} \right]. \]

So we can rewrite \( \frac{\sqrt{n}}{b_n} (T_n - ET_n) \) as follows
\[ \frac{\sqrt{n}}{b_n} (T_n - ET_n) = \frac{\sqrt{n}}{b_n (n+1)} \left[ \sum_{k=\alpha_n}^{\beta_n} B_{n,k} X_k + B_{n,\alpha_n} \sum_{k=1}^{\alpha_n-1} X_k - \left( \sum_{k=\alpha_n}^{\beta_n} B_{n,k} + (\alpha_n - 1) B_{n,\alpha_n} \right) \frac{S_{n+1}}{n+1} \right] \]
\[ \tag{3.10} \]
where
\[ D_n = \frac{\sum_{k=\alpha_n}^{\beta_n} B_{n,k} + (\alpha_n - 1) B_{n,\alpha_n}}{n+1}. \]

Let us define
\[ \tilde{T}_n := \sum_{k=1}^{n+1} \tilde{D}_n X_k, \]
where
\[ \tilde{D}_{n,k} = \begin{cases} B_{n,\alpha_n} - D_n, & 1 \leq k \leq \alpha_n - 1 \\ B_{n,k} - D_n, & \alpha_n \leq k \leq \beta_n \\ -D_n, & \beta_n + 1 \leq k \leq n+1 \end{cases}. \]

Then for any \( \lambda \in \mathbb{R} \), from the condition
\[ \frac{b_n}{\sqrt{n}} \max_{1 \leq k \leq n+1} |\tilde{D}_{n,k}| \to 0, \]
we have
\[
\frac{1}{b^2_n} \log \mathbb{E} \exp \left( \frac{\lambda b_n \sqrt{n}}{\sqrt{n}} \tilde{T}_n \right) = 1
\]
\[
= \frac{1}{b^2_n} \log \mathbb{E} \exp \left( \frac{\lambda b_n}{\sqrt{n}} \sum_{k=1}^{n+1} \tilde{D}_{n,k} X_k \right)
\]
\[
= \frac{1}{b^2_n} \sum_{k=1}^{n+1} \log \mathbb{E} \exp \left( \frac{\lambda b_n}{\sqrt{n}} \tilde{D}_{n,k} X_k \right)
\]
\[
= - \frac{1}{b^2_n} \sum_{k=1}^{n+1} \log \left( 1 - \frac{\lambda b_n \tilde{D}_{n,k}}{\sqrt{n}} \right)
\]
\[
= \frac{\lambda}{b_n \sqrt{n}} \sum_{k=1}^{n+1} \tilde{D}_{n,k} + \frac{\lambda^2}{2n} \sum_{k=1}^{n+1} \tilde{D}_{n,k}^2 + O \left( \frac{b_n}{n^{3/2}} \sum_{k=1}^{n+1} \tilde{D}_{n,k}^3 \right).
\]

It is easy to check that
\[
\sum_{k=1}^{n+1} \tilde{D}_{n,k} = 0
\]
and
\[
\frac{1}{n} \sum_{k=1}^{n+1} \tilde{D}_{n,k}^2 = \frac{1}{n} \left[ (\alpha_n - 1)B_{n,\alpha_n} + \sum_{k=\alpha_n}^{\beta_n} B_{n,k}^2 \right] - \frac{n+1}{n} D_n^2 \to \sigma^2.
\]

So we have
\[
\frac{1}{b^2_n} \log \mathbb{E} \exp \left( \frac{\lambda b_n \sqrt{n}}{\sqrt{n}} \tilde{T}_n \right) \to \frac{\lambda^2 \sigma^2}{2}
\]
which implies, by the Gärtner-Ellis theorem \([9]\), that for any \(r > 0\),
\[
\frac{1}{b^2_n} \log \mathbb{P} \left( \frac{1}{b_n \sqrt{n}} | \tilde{T}_n | \geq r \right) \to - \frac{r^2}{2\sigma^2}.
\]

Next, by the similar proofs in Theorem 3.1, the desired result can be obtained. \(\square\)

The following result is the large deviation principles for the linear combinations of uniform order statistics.

**Theorem 3.3.** Let \(U_1, U_2, \ldots, U_n\) be a sequence of random variables with the uniform distribution on the interval \((0, 1)\) and \(U_{n,1} \leq U_{n,2} \leq \cdots \leq U_{n,n}\) be the corresponding order statistics. Let \(T_n = \sum_{i=1}^{n} c_{n,i} U_{n,i}\), where \(\{c_{n,i}, 1 \leq i \leq n\}\) is an array of constants. Assume that
\[
B_{n,k} = \sum_{i=k}^{n} c_{n,i}, \quad D_n = \frac{1}{n+1} \sum_{i=1}^{n+1} B_{n,i}.
\]

(1) Suppose that there is a positive number \(a_1 > 0\) such that
\[
\max_{1 \leq k \leq n} \left( B_{n,k} - D_n \right) \leq a_1,
\]
and there exists a function $\Lambda_1(\lambda, r)$, such that for any $0 < r < a_1$, $\lambda \in [0, (a_1 - r)^{-1}]$,
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \log \frac{1}{1 - \lambda (B_{n,k} - D_n - r)} = \Lambda_1(\lambda, r).
\]
Assume that $0 \in A_1 r$, where
\[
A_{1r} = \left\{ \Lambda'_1(h, r) : \Lambda'_1(h, r) \text{ exists and is continuous on} \right. \nonumber
\]
the right and strictly monotonic for $h \in [0, (a_1 - r)^{-1}] \right\}.
\]
Then for any $0 < r < a_1$, we have
\[
\lim_{n \to \infty} \frac{1}{n} \log P(T_n - ET_n > r) = \inf_{\lambda > 0} \Lambda_1(\lambda, r).
\]

(2) Suppose that there is a positive number $a_2 > 0$ such that
\[
\max_{1 \leq k \leq n} (D_n - B_{n,k}) \leq a_2,
\]
and there exists a function $\Lambda_2(\lambda, r)$, such that for any $0 < r < a_2$, $\lambda \in [0, (a_2 - r)^{-1}]$,
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \log \frac{1}{1 - \lambda (D_n - B_{n,k} - r)} = \Lambda_2(\lambda, r).
\]
Assume that $0 \in A_2 r$, where
\[
A_{2r} = \left\{ \Lambda'_2(h, r) : \Lambda'_2(h, r) \text{ exists and is continuous on} \right. \nonumber
\]
the right and strictly monotonic for $h \in [0, (a_2 - r)^{-1}] \right\}.
\]
Then for any $0 < r < a_2$, we have
\[
\lim_{n \to \infty} \frac{1}{n} \log P(T_n - ET_n < -r) = \inf_{\lambda > 0} \Lambda_2(\lambda, r).
\]

P r o o f. It is easy to check
\[
T_n = \sum_{i=1}^{n} c_{n,i} U_{n,i} = \sum_{i=1}^{n} c_{n,i} \frac{X_1 + \cdots + X_i}{X_1 + \cdots + X_{n+1}} = \sum_{k=1}^{n} \frac{B_{n,k} X_k}{X_1 + \cdots + X_{n+1}}.
\]
(3.12)
By defining $B_{n,n+1} = 0$, then from (2.2), we have
\[
ET_n = \frac{1}{n+1} \sum_{i=1}^{n} i c_{n,i} = \frac{1}{n+1} \sum_{i=1}^{n+1} B_{n,i} = D_n.
\]
Moreover, for any \(0 < r < a_1\), we have
\[
\mathbb{P}(T_n - \mathbb{E}T_n > r) = \mathbb{P}\left(\sum_{k=1}^{n} (B_{n,k} - D_n - r)X_k > 0\right)
\]
and for any \(0 < r < a_2\), we get
\[
\mathbb{P}(T_n - \mathbb{E}T_n < -r) = \mathbb{P}\left(\sum_{k=1}^{n} (D_n - B_{n,k} - r)X_k > 0\right).
\]

It is not difficult to show that for any \(\lambda \in [0, (a_1 - r)^{-1}]\),
\[
\Lambda_{1n}(\lambda, r) := \frac{1}{n} \log \mathbb{E} \exp\left(\sum_{k=1}^{n} \lambda (B_{n,k} - D_n - r)X_k\right) \to \Lambda_1(\lambda, r)
\]
and for any \(\lambda \in [0, (a_2 - r)^{-1}]\),
\[
\Lambda_{2n}(\lambda, r) := \frac{1}{n} \log \mathbb{E} \exp\left(\sum_{k=1}^{n} \lambda (D_n - B_{n,k} - r)X_k\right) \to \Lambda_2(\lambda, r).
\]

From the theorem of Plachky and Steinebach (see Lemma 2.1), the desired results can be obtained.

4. SOME APPLICATIONS

In this section, we give some applications, such as, the \(k\)th order statistics, Gini mean difference statistics.

4.1. The \(k\)th uniform order statistics

In this subsection, we consider the large and moderate deviations for the \(k\)th uniform order statistics. Some related properties for the ratios of order statistics from uniform distributions are discussed recently (see Xu et al. [33], Xu and Miao [35]).

Proposition 4.1. Let \(U_1, U_2, \ldots, U_n\) be a sequence of random variables with the uniform distribution on the interval \((0, 1)\) and \(U_{n,1} \leq U_{n,2} \leq \cdots \leq U_{n,n}\) be the corresponding order statistics. Let \(\{k_n, n \geq 1\}\) be a sequence of positive constants such that \(k_n = np + o(n)\), where \(p \in (0, 1)\). Suppose that \(\{b_n, n \geq 1\}\) is a sequence of positive numbers such that
\[
b_n \to \infty, \quad \frac{b_n}{\sqrt{n}} \to 0.
\]

Then for any \(r > 0\), we have
\[
\lim_{n \to \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{n}}{b_n} \left| U_{n,k_n} - \frac{k_n}{n+1} \right| > r\right) = -\frac{r^2}{2p(1-p)}.
\]
Proof. In fact, we choose the constants in the definition of the random variable $T_n$ in such a way: $c_{n,k_n} = 1, c_{n,i} = 0$ ($i \neq k_n$). In this case, we have $B_{n,1} = B_{n,2} = \cdots = B_{n,k_n} = 1, B_{n,k_n+1} = \cdots = B_{n,n} = 0$, $T_n = U_{n,k_n}$, $\mathbb{E}T_n = \frac{k_n}{n+1}$ and
\[
\frac{1}{n} \sum_{k=1}^{n} B_{n,k}^2 \left( \frac{1}{n+1} \sum_{k=1}^{n} B_{n,k} \right)^2 \to p(1-p).
\]
Hence from Theorem 3.1, Proposition 4.1 can be obtained. □

Proposition 4.2. Let $U_1, U_2, \ldots, U_n$ be a sequence of random variables with the uniform distribution on the interval $(0,1)$ and $U_{n,1} \leq U_{n,2} \leq \cdots \leq U_{n,n}$ be the corresponding order statistics. Let $\{k_n, n \geq 1\}$ be a sequence of positive constants such that $k_n = np + o(n)$, where $p \in (0,1)$. Then for any $0 < r < (1-p)$, we have
\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(U_{n,k_n} - p > r) = p \log \frac{p+r}{p} + (1-p) \log \frac{1-p-r}{1-p}
\]
and for any $0 < r < p$, we have
\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(U_{n,k_n} - p < -r) = p \log \frac{p-r}{p} + (1-p) \log \frac{1-p+r}{1-p}.
\]

Proof. As the proof of Proposition 4.1, we choose the constants in the definition of the random variable $T_n$ in such a way: $c_{n,k_n} = 1, c_{n,i} = 0$ ($i \neq k_n$). In this case, we have $B_{n,1} = B_{n,2} = \cdots = B_{n,k_n} = 1, B_{n,k_n+1} = \cdots = B_{n,n} = 0$, $T_n = U_{n,k_n}$, $\mathbb{E}T_n = \frac{k_n}{n+1} \to p$. Hence for any $\lambda \in [0, (1-p-r)^{-1}]$, we have
\[
\Lambda_1(\lambda, r) = - \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \log [1 - \lambda(B_{n,i} - D_{n,i} - r)]
\]
\[
= - \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{k_n} \log \left[ 1 - \lambda \left( 1 - \frac{k_n}{n+1} - r \right) \right]
\]
\[
- \lim_{n \to \infty} \frac{1}{n} \sum_{i=k_n+1}^{n} \log \left[ 1 + \lambda \left( \frac{k_n}{n+1} + r \right) \right]
\]
\[
= - p \log [1 - \lambda(1-p-r)] - (1-p) \log [1 + \lambda(p+r)].
\]
and for any $\lambda \in [0, (p-r)^{-1}]$, we have
\[
\Lambda_2(\lambda, r) = - \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \log [1 - \lambda(D_{n,i} - B_{n,i} - r)]
\]
\[
= - \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{k_n} \log \left[ 1 - \lambda \left( \frac{k_n}{n+1} - 1 - r \right) \right]
\]
\[
- \lim_{n \to \infty} \frac{1}{n} \sum_{i=k_n+1}^{n} \log \left[ 1 - \lambda \left( \frac{k_n}{n+1} - r \right) \right]
\]
\[
= - p \log [1 + \lambda(1+r)] - (1-p) \log [1 - \lambda(p-r)].
\]
By simple calculation, we have
\[
\frac{\partial \Lambda_1(\lambda, r)}{\partial \lambda} = 0 \implies \lambda_1 = \frac{r}{(p + r)(1 - p - r)}
\]
and
\[
\frac{\partial \Lambda_2(\lambda, r)}{\partial \lambda} = 0 \implies \lambda_2 = \frac{r}{(p - r)(1 + r - p)}.
\]
From Theorem 3.3, Proposition 4.2 can be obtained.

4.2. The \( k \)th order statistics for continuous random variables

A number of results for order statistics corresponding to continuous distribution functions can be obtained by means of Smirnov’s transformation from the results for uniform order statistics. For a random variable \( X \) with arbitrary distribution function \( F \), let \( F^{-1}(u) = \inf \{x : F(x) \geq u\} \), \( 0 < u < 1 \).

By the right continuity of \( F \), it follows that \( F(F^{-1}(u)) \geq u \) and \( F^{-1}(F(x)) \leq x \). Hence for \( 0 \leq F(x) \leq 1 \),
\[
P(X \leq x) = F(x) = P(U \leq F(x)) = P(F^{-1}(U) \leq x)
\]
where \( U \) is a standard uniform distribution. Let \( X_1, X_2, \ldots, X_n \) be a sequence of i.i.d. random variables with a continuous and strictly increasing distribution function \( F \), and denote the ordered values by \( X_{n,1} \leq X_{n,2} \leq \cdots \leq X_{n,n} \). Let \( U_1, U_2, \ldots, U_n \) be a sequence of random variables with the uniform distribution on the interval \((0, 1)\) and \( U_{n,1} \leq U_{n,2} \leq \cdots \leq U_{n,n} \) be the corresponding order statistics. Hence for any \( x \in \mathbb{R} \) we have
\[
P(X_{n,k} \leq x) = P(F(X_{n,k}) \leq F(x)) = P(U_{n,k} \leq F(x)).
\]
Furthermore, it is well known that (see David and Nagaraja [8, (2.3.7)]),
\[
(X_{n,1}, \ldots, X_{n,n}) \overset{d}{=} (F^{-1}(U_{n,1}), \ldots, F^{-1}(U_{n,n})). \tag{4.1}
\]

There are some references to consider the large deviations, moderate deviations and Bahadur’s asymptotic efficiency for the \( k \)th order statistics (see Miao et al. [21], Xu et al. [32], Xu and Miao [34], Yao et al. [36]).

**Proposition 4.3.** Let \( X_1, X_2, \ldots, X_n \) be a sequence of i.i.d. random variables having common continuous and strictly increasing distribution function \( F \) with the density \( f(x) \) such that \( \sup_{x:f(x)>0} |f'(x)| \leq M \), where \( M \) is a positive constant, and \( X_{n,1} \leq X_{n,2} \leq \cdots \leq X_{n,n} \) be the corresponding order statistics. Let \( \{k_n, n \geq 1\} \) be a sequence of positive constants such that \( k_n = np + o(n) \), where \( p \in (0, 1) \). Suppose that \( \{b_n, n \geq 1\} \) is a sequence of positive numbers such that
\[
b_n \to \infty, \quad \frac{b_n}{\sqrt{n}} \to 0.
\]
Then for any $r > 0$, we have
\[
\lim_{n \to \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \left| X_{n,k_n} - G \left( \frac{k_n}{n+1} \right) \right| > r \right) = -\frac{r^2 [f(G(p))]^2}{2p(1-p)}
\]
where $G(x) = F^{-1}(x)$.

**Proof.** Because of $X_{n,k_n} \overset{d}{=} F^{-1}(U_{n,k_n})$, we have
\[
\mathbb{P} \left( X_{n,k_n} > \frac{rb_n}{\sqrt{n}} + G \left( \frac{k_n}{n+1} \right) \right) = \mathbb{P} \left( U_{n,k_n} > F \left( \frac{rb_n}{\sqrt{n}} + G \left( \frac{k_n}{n+1} \right) \right) \right)
\]
and
\[
\mathbb{P} \left( X_{n,k_n} < -\frac{rb_n}{\sqrt{n}} + G \left( \frac{k_n}{n+1} \right) \right) = \mathbb{P} \left( U_{n,k_n} < F \left( -\frac{rb_n}{\sqrt{n}} + G \left( \frac{k_n}{n+1} \right) \right) \right)
\]
where $|\theta_1| \leq M$ and $|\theta_2| \leq M$. From Proposition 4.1, the desired results can be obtained. 

**Proposition 4.4.** Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables having common continuous and strictly increasing distribution function $F$, and $X_{n,1} \leq X_{n,2} \leq \cdots \leq X_{n,n}$ be the corresponding order statistics. Let \( \{k_n, n \geq 1\} \) be a sequence of positive constants such that $k_n = np + o(n)$, where $p \in (0,1)$. Then for any $r > 0$, we have
\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P} \left( X_{n,k_n} - G \left( \frac{k_n}{n+1} \right) > r \right) = p \log \frac{F(r + G(p))}{p} + (1-p) \log \frac{1 - F(r + G(p))}{1-p}
\]
and
\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P} \left( X_{n,k_n} - G \left( \frac{k_n}{n+1} \right) < -r \right) = p \log \frac{F(G(p) - r)}{p} + (1-p) \log \frac{1 - F(G(p) - r)}{1-p}
\]
where $G(x) = F^{-1}(x)$. 

Proof. Because of $X_{n,k} \overset{d}{=} F^{-1}(U_{n,k_n})$, we have
\[
\mathbb{P}(X_{n,k} > r + G\left(\frac{k_n}{n+1}\right)) = \mathbb{P}(U_{n,k} > F\left(r + G\left(\frac{k_n}{n+1}\right)\right)) = \mathbb{P}(U_{n,k} - p > F\left(r + G\left(\frac{k_n}{n+1}\right)\right) - p)
\]
and
\[
\mathbb{P}(X_{n,k} < -r + G\left(\frac{k_n}{n+1}\right)) = \mathbb{P}(U_{n,k} < F\left(-r + G\left(\frac{k_n}{n+1}\right)\right)) = \mathbb{P}(U_{n,k} - p < F\left(-r + G\left(\frac{k_n}{n+1}\right)\right) - p).
\]
From Proposition 4.2, the desired results can be obtained. □

4.3. Gini statistics

Let us consider the following Gini’s mean difference
\[
G_n = \frac{\sum_{i,j=1}^{n} |U_i - U_j|}{n(n-1)},
\]
where $\{U_i, i \geq 1\}$ be i.i.d. random variables with the uniform distribution on the interval $(0,1)$. Let $U_{n,1} \leq U_{n,2} \leq \cdots \leq U_{n,n}$ be the corresponding order statistics.

Proposition 4.5. Assume that $\{b_n, n \geq 1\}$ is a sequence of positive numbers such that
\[
b_n \to \infty, \quad \frac{b_n}{\sqrt{n}} \to 0.
\]
Then for any $r > 0$, we have
\[
\lim_{n \to \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{n}}{b_n} |G_n - \mathbb{E}G_n| > r\right) = -\frac{45r^2}{2}.
\]

Proof. It is easy to see that
\[
\sum_{i,j=1}^{n} |U_i - U_j| = 2 \sum_{1 \leq i < j \leq n} |U_i - U_j| = 2 \sum_{1 \leq i < j \leq n} |U_{n,i} - U_{n,j}| = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (U_{n,j} - U_{n,i}) = 2 \sum_{i=1}^{n} (2i - n - 1)U_{n,i}.
\]
Let
\[ c_{n,i} = \frac{2(2i - n - 1)}{n(n-1)} \]
and
\[ B_{n,k} = \sum_{i=k}^{n} c_{n,i}, \]
then we have
\[ G_n = \sum_{i=1}^{n} c_{n,i} U_{n,i}. \]

It is not difficult to check that
\[ D_n = \frac{1}{n+1} \sum_{k=1}^{n} B_{n,k} = \frac{1}{n} \sum_{k=1}^{n} \frac{2(k-1)(n-k+1)}{n(n-1)} = \frac{1}{3}, \]
\[ \mathbb{E}G_n = \sum_{i=1}^{n} c_{n,i} \frac{i}{n+1} = \frac{1}{3}, \]
\[ \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{n} B_{n,k}^2 = \frac{2}{3} \frac{2n-1}{n-1} - 2 + \frac{4}{30} \frac{(2n-1)(3n^2-3n-1)}{n^2(n-1)} \to \frac{2}{15}, \]
\[ \rho_n \max_{1 \leq k \leq n} |B_{n,k} - D_n| \to 0, \]
and
\[ \frac{1}{n} \sum_{k=1}^{n} B_{n,k}^2 - D_n^2 \to \frac{1}{45}. \]
So, from Theorem 3.1, Proposition 4.5 can be obtained. □

**Proposition 4.6.** Then for any 0 < r < \( \frac{1}{6} \), we have
\[ \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(G_n - \mathbb{E}G_n > r) = \inf_{\lambda > 0} \int_{0}^{1} \frac{1}{1 - \lambda(2x - 2x^2 - \frac{1}{3} - r)} \, dx. \]

and for any 0 < r < \( \frac{1}{3} \), we have
\[ \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(G_n - \mathbb{E}G_n < -r) = \inf_{\lambda > 0} \int_{0}^{1} \frac{1}{1 - \lambda(\frac{1}{3} - 2x + 2x^2 - r)} \, dx. \]

**Proof.** It is easy to see that
\[ G_n - \mathbb{E}G_n = \sum_{i=1}^{n} c_{n,i} (U_{n,i} - \mathbb{E}U_{n,i}) = \sum_{i=1}^{n} c_{n,i} \left( U_{n,i} - \frac{i}{n+1} \right) \]
where
\[ c_{n,i} = \frac{2(2i - n - 1)}{n(n-1)}. \]
It is not difficult to check that for any $\varepsilon > 0$, and for all $n$ large enough, we have

$$\max_{1 \leq k \leq n} (B_{n,k} - D_n) \leq \frac{1}{6} + \varepsilon, \quad \max_{1 \leq k \leq n} (D_n - B_{n,k}) < \frac{1}{3},$$

where

$$B_{n,k} = \sum_{i=k}^{n} c_{n,i} = \frac{2(k-1)(n-k+1)}{n(n-1)}, \quad D_n = \frac{1}{3}.$$

Now by Lemma 2.2, for any $0 < r < \frac{1}{6}$, $\lambda \in [0, (\frac{1}{6} - r)^{-1}]$, we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \log \frac{1}{1 - \lambda (B_{n,k} - D_n - r)} = \int_{0}^{1} \log \frac{1}{1 - \lambda (2x - 2x^2 - \frac{1}{3} - r)} \, dx$$

and for any $0 < r < \frac{1}{3}$, $\lambda \in [0, (\frac{1}{3} - r)^{-1}]$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \log \frac{1}{1 - \lambda (D_n - B_{n,k} - r)} = \int_{0}^{1} \log \frac{1}{1 - \lambda (\frac{1}{3} - 2x + 2x^2 - r)} \, dx.$$

From Theorem 3.3, Proposition 4.6 can be obtained. □

ACKNOWLEDGEMENT

This work was partially supported by National Natural Science Foundation of China (11971154).

(Received March 26, 2021)

REFERENCES


Some limit behavior for linear combinations of order statistics 987


Yu Miao, College of Mathematics and Information Science, Henan Engineering Laboratory for Big Data Statistical Analysis and Optimal Control, Henan Normal University, Henan Province, 453007. P. R. China.
e-mail: yumiao728@gmail.com

Mengyao Ma, College of Mathematics and Information Science, Henan Normal University, Henan Province, 453007. P. R. China.
e-mail: mengyaoma17@126.com