Metrika (1999) 49: 257–262

Metrika

Reverse submartingale property arising from exchangeable random variables*

Wen-Jang Huang[1], **Jyh-Cherng Su**[2]

[1] Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, Taiwan, 80424, R.O.C. (e-mail: huangwj@math.nsysu.edu.tw)
[2] Department of Mathematics, Chinese Military Academy, FengShan, Kaohsiung, Taiwan, 830, R.O.C. (e-mail: sujc@cc.cma.edu.tw)

Received: December 1998

Abstract. In this work, for an exchangeable sequence of random variables $\{X_i,\ i \geq 1\}$, and two nondecreasing sequences of positive integers $\{h_n,\ n \geq 2\}$ and $\{k_n,\ n \geq 2\}$, where $h_n + k_n \leq n$, $\forall n \geq 2$, we prove that $\{R_{n,h_n,k_n}/n,\ n \geq 2\}$ forms a reverse submartingale sequence, where $R_{n,h_n,k_n} = \frac{1}{k_n}\sum_{j=0}^{k_n-1} X_{n-j,n} - \frac{1}{h_n}\sum_{j=1}^{h_n} X_{j,n}$, and $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ are the order statistics based on $\{X_1, \ldots, X_n\}$.

Key words: Exchangeable, order statistics, reverse submartingale

1. Introduction

Exchangeability plays an important role in the theory of order statistics, see Galambos (1982), for instance. In this work we will investigate a reverse submartingale property of order statistics about exchangeable random variables. First we give the following definitions which can be found in books such as Laha and Rohatgi (1979).

Definition 1. *The random variables* $X_1, X_2, \ldots, X_n$ *are said to be exchangeable if the distribution of* $(X_{\pi 1}, X_{\pi 2}, \ldots, X_{\pi n})$ *is the same as that of* $(X_1, X_2, \ldots, X_n)$ *for every permutation* π *of* $\{1, 2, \ldots, n\}$. *An infinite sequence of random variables*

* Support for this research was provided in part by the National Science Council of the Republic of China, Grant No. NSC 86-2115-M-110-014.

$\{X_n,\ n \geq 1\}$ is said to be exchangeable if $X_1, X_2, \ldots, X_n$ are exchangeable for every $n \geq 2$.

Definition 2. *Let $\{X_n,\ n \geq 1\}$ be a sequence of random variables such that $E|X_n| < \infty$ for every $n \geq 1$. We say that $\{X_n,\ n \geq 1\}$ is a reversed submartingale* (RSM) *if*

$$E(X_n | X_{n+1}, X_{n+2}, \ldots) \geq X_{n+1}.$$

Now let $X_1, X_2, \ldots$ be an exchangeable sequence of random variables; $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ be the order statistics based on $\{X_1, X_2, \ldots, X_n\}$ and $R_{n,h,k} = \frac{1}{k}\sum_{j=0}^{k-1} X_{n-j,n} - \frac{1}{h}\sum_{j=1}^{h} X_{j,n}$, where h and k are two positive integers. $R_{n,1,1} = X_{n,n} - X_{1,n}$ is known to be the range of $\{X_1, X_2, \ldots, X_n\}$. Huang and Huang (1994) showed that

(i) if $E|X_1| < \infty$, then $\{n^{-1}R_{n,1,1}\}$ is an RSM sequence;
(ii) if $E|X_1|^l < \infty$, then $E(R^l_{j,1,1}) \leq (j/i)^l E(R^l_{i,1,1})$, $2 \leq i \leq j$, $l \geq 1$.

This is a stronger result than that in Bhattacharyya (1970). In Section 2, we will generalize the above result. We prove that if $R_{n,1,1}$ is replaced by the more general statistics R_{n,h_n,k_n}, where $\{h_n,\ n \geq 2\}$ and $\{k_n,\ n \geq 2\}$ are two non-decreasing sequences of positive integers with $h_n + k_n \leq n$, $\forall n \geq 2$, similar results still hold.

2. Reverse submartingale property

First we have the following lemma.

Lemma 1. *Given $n+1$ numbers $x_1 \leq x_2 \leq \cdots \leq x_{n+1}$, let $x_1^{(i)} \leq x_2^{(i)} \leq \cdots \leq x_n^{(i)}$, $i = 1, 2, \ldots, n+1$, be the set of all possible subsets of n-tuples that can be formed from the $n+1$ x's. Let h, k be fixed positive integers, and*

$$L^{(i)}_{n,h,k} = \frac{1}{k}\sum_{j=0}^{k-1} x^{(i)}_{n-j} - \frac{1}{h}\sum_{j=1}^{h} x^{(i)}_j. \tag{1}$$

For $n \geq \max\{h,k\}$,

$$\sum_{i=1}^{n+1} L^{(i)}_{n,h,k} = n\left(\frac{1}{k}\sum_{j=0}^{k-1} x_{n+1-j} - \frac{1}{h}\sum_{j=1}^{h} x_j\right) + (x_{n+1-k} - x_{h+1}). \tag{2}$$

Note that if for each $i = 1, 2, \ldots, n+1$, let $x_j^{(i)} = x_j$, $1 \leq j < i$, and $x_j^{(i)} = x_{j+1}$, $i \leq j \leq n$, then $x_1^{(i)} \leq x_2^{(i)} \leq \cdots \leq x_n^{(i)}$, $i = 1, 2, \ldots, n+1$, is the set of all possible subsets of n-tuples that can be formed from the $n+1$ x's, and

$$L_{n,h,k}^{(i)} = \begin{cases} \dfrac{1}{k}\displaystyle\sum_{j=0}^{k-1} x_{n+1-j} - \dfrac{1}{h}\left(\sum_{j=1}^{h+1} x_j - x_i\right), & \text{if } 1 \le i \le h, \\ \dfrac{1}{k}\displaystyle\sum_{j=0}^{k-1} x_{n+1-j} - \dfrac{1}{h}\sum_{j=1}^{h} x_j, & \text{if } h+1 \le i \le n-k+1, \\ \dfrac{1}{k}\left(\displaystyle\sum_{j=0}^{k} x_{n+1-j} - x_i\right) - \dfrac{1}{h}\sum_{j=1}^{h} x_j, & \text{if } n-k+2 \le i \le n+1. \end{cases} \tag{3}$$

Using this representation, (2) can be obtained immediately.

Using Lemma 1, we can obtain an extension of Huang and Huang (1994).

Theorem 1. *For any fixed positive integers h, k, if $E|X_1| < \infty$, then*

(i) $E(R_{n,h,k}/n) = E(R_{n+1,h,k}/(n+1)) + \dfrac{1}{n(n+1)} E(X_{n+1-k,n+1} - X_{h+1,n+1})$, $n \ge \max\{h,k\}$;
(ii) $\{R_{n,h,k}/n,\ n \ge h+k\}$ *forms an RSM.*

Proof. Let $X_{1,n}^{(i)} \le X_{2,n}^{(i)} \le \cdots \le X_{n,n}^{(i)}$, $i = 1, 2, \ldots, n+1$, be the $n+1$ ordered subsets of n-tuples that can be formed from the random variables $\{X_1, X_2, \ldots, X_{n+1}\}$. Also let $R_{n,h,k}^{(i)} = \frac{1}{k}\sum_{j=0}^{k-1} X_{n-j,n}^{(i)} - \frac{1}{h}\sum_{j=1}^{h} X_{j,n}^{(i)}$. As the random variables $X_1, X_2, \ldots, X_{n+1}$ are exchangeable, we have that the distributions of $(X_{1,n}^{(i)}, X_{2,n}^{(i)}, \ldots, X_{n,n}^{(i)})$ and $R_{n,h,k}^{(i)}$ are the same as those of $(X_{1,n}, X_{2,n}, \ldots, X_{n,n})$ and $R_{n,h,k}$ respectively. From Lemma 1, we have

$$\sum_{i=1}^{n+1} R_{n,h,k}^{(i)} = nR_{n+1,h,k} + (X_{n+1-k,n+1} - X_{h+1,n+1}), \quad n \ge \max\{h,k\}. \tag{4}$$

Upon dividing by $n(n+1)$ and then taking the expectations on both sides, the assertion (i) follows. Next, if $n \ge h+k$, then (4) implies

$$\frac{1}{n+1}\sum_{i=1}^{n+1} \frac{1}{n} R_{n,h,k}^{(i)} \ge \frac{1}{n+1} R_{n+1,h,k}. \tag{5}$$

It is easy to see that for every $l \ge 1$, given $R_{n+1,h,k}, R_{n+2,h,k}, \ldots, R_{n+l,h,k}$, the random variables $X_1, X_2, \ldots, X_{n+1}$ are also exchangeable. Hence the distribution of $R_{n,h,k}^{(i)} | R_{n+1,h,k}, R_{n+2,h,k}, \ldots, R_{n+l,h,k}$ is the same as that of $R_{n,h,k} | R_{n+1,h,k}, R_{n+2,h,k}, \ldots, R_{n+l,h,k}$. Now taking the conditional expectations on both sides of (5) given $R_{n+1,h,k}, R_{n+2,h,k}, \ldots, R_{n+l,h,k}$, we have, for $l \ge 1$,

$$\frac{1}{n} E(R_{n,h,k} | R_{n+1,h,k}, R_{n+2,h,k}, \ldots, R_{n+l,h,k}) \ge \frac{1}{n+1} R_{n+1,h,k}. \tag{6}$$

The theorem follows by letting l go to infinity. □

Before proving Corollary 1, we give the following trivial lemma.

Lemma 2. *Assume* $x_1 \le x_2 \le \cdots \le x_n$. *For* $1 \le m \le l \le n$, *we have*

(i) $\frac{1}{m}\sum_{j=0}^{m-1} x_{n-j} \ge \frac{1}{l}\sum_{j=0}^{l-1} x_{n-j}$;
(ii) $\frac{1}{l}\sum_{j=1}^{l} x_j \ge \frac{1}{m}\sum_{j=1}^{m} x_j$.

As a consequence of Theorem 1, we now show that the more general sequence $\{R_{n,h_n,k_n},\ n \ge 2\}$ also forms an RSM.

Corollary 1. *Let* $\{h_n,\ n \ge 2\}$ *and* $\{k_n,\ n \ge 2\}$ *be two nondecreasing sequences of positive integers with* $h_n + k_n \le n$, $\forall n \ge 2$. *If* $E|X_1| < \infty$, *then* $\{R_{n,h_n,k_n}/n,\ n \ge 2\}$ *is an RSM.*

Proof. Again from (5) we have, for $n \ge 2$ and $n \ge h_n + k_n$,

$$\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{n}R^{(i)}_{n,h_n,k_n} \ge \frac{1}{n+1}R_{n+1,h_n,k_n}. \tag{7}$$

Since $h_n \le h_{n+1}$, $k_n \le k_{n+1}$, Lemma 2 yields

$$\begin{aligned} R_{n+1,h_n,k_n} &= \frac{1}{k_n}\sum_{j=0}^{k_n-1} X_{n+1-j,n+1} - \frac{1}{h_n}\sum_{j=1}^{h_n} X_{j,n+1} \\ &\ge \frac{1}{k_{n+1}}\sum_{j=0}^{k_{n+1}-1} X_{n+1-j,n+1} - \frac{1}{h_{n+1}}\sum_{j=1}^{h_{n+1}} X_{j,n+1} \\ &= R_{n+1,h_{n+1},k_{n+1}}. \end{aligned}$$

Hence

$$\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{n}R^{(i)}_{n,h_n,k_n} \ge \frac{1}{n+1}R_{n+1,h_{n+1},k_{n+1}}.$$

From this the assertion follows. □

Finally, we have the following immediate consequences of Corollary 1.

Corollary 2. *If* $E|X_1|^l < \infty$, *then* $E(R^l_{n_2,r_2,s_2}) \le (n_2/n_1)^l E(R^l_{n_1,r_1,s_1})$, $r_1 \le r_2$, $s_1 \le s_2$, $r_1 + s_1 \le n_1$, $r_2 + s_2 \le n_2$, $n_1 \le n_2$ *and* $l \ge 1$.

Corollary 3. *If* $E|X_1| < \infty$, *then* $E(R_{n+1,r_2,s_2}) \le ((n+1)/n)E(R_{n,r_1,s_1})$, $r_1 \le r_2$, $s_1 \le s_2$, $r_1 + s_1 \le n$, $r_2 + s_2 \le n + 1$.

Again the bound is tight at $r_1 = r_2$, $s_1 = s_2$ and $r_1 + s_1 = n$.

Discussions

Assume r and s are two fixed positive real numbers with $r+s \leq 1$. Let n_0 be the integer such that $\min\{rn_0, sn_0\} \geq 1$. Define $h_n = [rn]$ and $k_n = [sn]$, $n \geq n_0$, where for $u \in R$, $[u]$ denotes the integer part of u, then $h_n \leq h_{n+1}, k_n \leq k_{n+1}$ and $h_n+k_n \leq n$, $\forall n \geq 2$. From Corollary 1, $\{R_{n,h_n,k_n}/n, n \geq n_0\}$ is an RSM. In this case, $\frac{1}{k_n}\sum_{j=0}^{k_n-1} X_{n-j,n}$ and $\frac{1}{h_n}\sum_{j=1}^{h_n} X_{j,n}$ are the sample averages of the largest $100s$ percent and the smallest $100r$ percent, respectively, of the sample $\{X_1, X_2, \ldots, X_n\}$. The statistics R_{n,h_n,k_n} is the difference of the above two sample averages and can be viewed as a generalization of sample range and quasi ranges. This statistics has been used in many occasions, for example in the investigation of the difference between the highest $100s$ percent and the lowest $100r$ percent families income in a certain country.

Next, let $E|X_1| < \infty$ and h and k be two positive integers. According to previous results, it is natural to ask whether similar sequences of statistics, such as $\{(X_{n-k+1,n} - X_{h,n})/n,\ n \geq h+k\}$ also forms an RSM sequence. We give a counterexample in the following. Let $\{X_i,\ i \geq 1\}$ be a sequence of independent and identically distributed Bernoulli random variables with $P(X_1=1)=P(X_1=0)=1/2$. Again let $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ be the order statistics of $\{X_1, X_2, \ldots, X_n\}$. Let $h=k=2$. Then it can be shown that for $n \geq 4$, $X_{n-1,n} - X_{2,n}$ is also a Bernoulli random variable and

$$P(X_{n-1,n}-X_{2,n}=0) = 1/2^n + n/2^n + 1/2^n + n/2^n = (n+1)/2^{n-1}. \tag{8}$$

Thus

$$E((X_{n-1,n} - X_{2,n})/n) = \frac{1}{n}(1-(n+1)/2^{n-1}).$$

From this for $n=4$ or 5, we have

$$E((X_{n-1,n} - X_{2,n})/n) < E((X_{n,n+1} - X_{2,n+1})/(n+1)).$$

Hence $\{(X_{n-1,n} - X_{2,n})/n,\ n \geq 4\}$ does not form an RSM sequence.

The above counterexample suggests the following open problem which is worth further investigation. Let

$$\tilde{R}_{n,h,k} = \sum_{j=0}^{k-1} c_{j+1} X_{n-j,n} - \sum_{j=1}^{h} b_j X_{j,n}$$

with weights c_j, b_j summing up to 1, i.e. $c_1 + c_2 + \cdots + c_k = b_1 + b_2 + \cdots + b_h = 1$. Then under which conditions on the sequences $c_j = c_j(n)$, $b_j = b_j(n)$ is $\tilde{R}_{n,h,k}/n$ still an RSM?

Acknowledgements: The authors are grateful to the referees for many helpful comments and suggestions which improve the presentation of the paper.

References

1. Bhattacharyya BB (1970) Reverse submartingale and some functions of order statistics. Ann. Math. Statist. 41:2155–2157
2. Galambos J (1982) The role of exchangeability in the theory of order statistics. In: Koch G, Spizzichino F (eds.) Exchangeability in probability and statistics. North-Holland, Amsterdam, pp. 75–86
3. Huang JS, Huang WJ (1994) A reverse submartingale property of the range. Ann. Statist. 22:474–477
4. Laha RG, Rohatgi VK (1979) Probability theory. John Wiley, New York