Liet. mat. rink. LMD darbai, **50**, 2009, 287–292

Orthogonal decomposition of finite population L-statistics

Andrius ČIGINAS
Vilnius University, Faculty of Mathematics and Informatics
Naugarduko 24, LT-03225 Vilnius, Lithuania
e-mail: andrius.ciginas@mif.vu.lt

Abstract. In this paper we study orthogonal decomposition of finite population L-statistics. We propose quite simple form of first two terms of such decomposition.

Keywords: finite population, sampling without replacement, L-statistic, Hoeffding decomposition.

1. Introduction

Consider the population $\mathcal{X} = \{x_1, \dots, x_N\}$ of size N and assume that $x_1 < \dots < x_N$. Let $X_1, \dots, X_n$ is simple random sample of size $n < N$ drawn without replacement from $\mathcal{X}$ and let $X_{(1)} < \dots < X_{(n)}$ denote the order statistics of $X_1, \dots, X_n$. Then for arbitrary real numbers $c_1, \dots, c_n$ define L-statistic $L = c_1 X_{(1)} + \dots + c_n X_{(n)}$.

This statistic can be decomposed into the sum

$$L = \mathbf{E}L + U_1 + \dots + U_n, \tag{1}$$
$$U_m = \sum_{1 \leqslant i_1 < \dots < i_m \leqslant n} g_m(X_{i_1}, \dots, X_{i_m}), \quad m = 1, \dots, n.$$

Decomposition (1) is called orthogonal (called also Hoeffding) decomposition and U_m are called U-statistics. The symmetric kernels $g_m, m = 1, \dots, n$ are linear combinations of conditional expectations

$$h_j(x_{k_1}, \dots, x_{k_j}) = \mathbf{E}(L - \mathbf{E}L | X_1 = x_{k_1}, \dots, X_j = x_{k_j}), \quad 1 \leqslant j \leqslant m. \tag{2}$$

The decompositon (1) and its applications for finite population symmetric statistics were studied in [1]. In that paper we can find coefficients of linear combinations of (2).

The main interest of present work is conditional expectations (2) for $j = 1, \dots, n$ and functions

$$g_1(x) = \frac{N-1}{N-n} h_1(x), \tag{3}$$

$$g_2(x, y) = \frac{N-2}{N-n} \frac{N-3}{N-n-1} \left(h_2(x, y) - \frac{N-1}{N-2} (h_1(x) + h_1(y)) \right), \tag{4}$$

which can be useful for various applications.

In the case where random variables $X_1, \ldots, X_n$ are independent and identically distributed the orthogonal decomposition of L-statistics was studied in [2]. We shall adopt some ideas from that paper to get a similar form of orthogonal decomposition. Note that for samples without replacement from finite population variables $X_1, \ldots, X_n$ are identically distributed, but they are not independent.

In Section 2 we state our results, which are proved in Section 3.

2. Results

For the given n let fix $0 \leqslant m \leqslant n$ and define a set of conditions $A_m = \{X_1 = x_{k_1}, \ldots, X_m = x_{k_m}\}$, where $1 \leqslant k_1 < \ldots < k_m \leqslant N$. Let $k_0 = 0$, $k_{m+1} = N+1$ and define $X_{(0)} = x_0$, $X_{(n+1)} = x_{N+1}$ where $x_0 = x_1$, $x_{N+1} = x_N$. Consider statistics $X_{(r+1)} - X_{(r)}$, $r = 0, \ldots, n$.

LEMMA 1. *For any $m = 0, \ldots, n$ and $r = 0, \ldots, n$ we have*

$$\mathbf{E}(X_{(r+1)} - X_{(r)} | A_m) = \sum_{s=1}^{m+1} \sum_{i=k_{s-1}}^{k_s - 1} \Delta_{m,s,i}(r)(x_{i+1} - x_i), \tag{5}$$

where we denote

$$\Delta_{m,s,i}(r) = \binom{N-m}{n-m}^{-1} \binom{i-s+1}{r-s+1} \binom{N-i-m+s-1}{n-r-m+s-1}.$$

We shall use differences $X_{(r+1)} - X_{(r)}$, $r = 0, \ldots, n$ to get convenient expression of (2).

PROPOSITION 1. *For chosen $m = 1, \ldots, n$ we have*

$$\mathbf{E}(L - \mathbf{E}L | A_m) = \sum_{j=1}^{n} c_j \sum_{r=0}^{j-1} \big\{ \mathbf{E}\big(X_{(r+1)} - X_{(r)} | A_m\big) - \mathbf{E}\big(X_{(r+1)} - X_{(r)}\big) \big\}. \tag{6}$$

Next we shall propose simple form of kernels (3) and (4).

THEOREM 1. (i) *For $1 \leqslant k \leqslant N$*

$$g_1(x_k) = -\sum_{j=1}^{n} c_j \sum_{i=1}^{N-1} \varphi_k(i) \frac{\binom{i-1}{j-1}\binom{N-i-1}{n-j}}{\binom{N-2}{n-1}} (x_{i+1} - x_i), \tag{7}$$

where

$$\varphi_k(i) = \begin{cases} -\frac{i}{N}, & \text{if } 1 \leqslant i < k, \\ 1 - \frac{i}{N}, & \text{if } k \leqslant i < N. \end{cases}$$

(ii) *For* $1 \leqslant k < l \leqslant N$

$$g_2(x_k, x_l) = -\sum_{j=2}^{n}(c_j - c_{j-1})\sum_{i=1}^{N-1}\phi_{k,l}(i)\frac{\binom{i-2}{j-2}\binom{N-i-2}{n-j}}{\binom{N-4}{n-2}}(x_{i+1} - x_i), \tag{8}$$

where

$$\phi_{k,l}(i) = \begin{cases} \frac{i(i-1)}{(N-1)(N-2)}, & \text{if } 1 \leqslant i < k, \\ -\frac{(i-1)(N-i-1)}{(N-1)(N-2)}, & \text{if } k \leqslant i < l, \\ \frac{(N-i)(N-i-1)}{(N-1)(N-2)}, & \text{if } l \leqslant i < N. \end{cases}$$

Remark 1. While we do not talk about applications, which are beyond the scope of this paper, the Theorem 1 is just a theoretical result, i.e., the kernels (7) and (8) are just the formal functions.

3. Proofs

Proof of Lemma 1. For any $m = 0, \ldots, n$ and $r = 0, \ldots, n+1$ straightforward combinatorial calculations give

$$\mathbf{E}(X_{(r)}|A_m) = \binom{N-m}{n-m}^{-1}\left[\sum_{s=1}^{m+1}\sum_{i=k_{s-1}+1}^{k_s-1}\binom{i-s}{r-s}\binom{N-i-m+s-1}{n-r-m+s-1}x_i \right.$$
$$\left. + \sum_{s=0}^{m+1}\binom{k_s-s}{r-s}\binom{N-k_s-m+s}{n-r-m+s}x_{k_s}\right].$$

The key idea is for $r = 0, \ldots, n$ write

$$\mathbf{E}(X_{(r+1)}|A_m) = \binom{N-m}{n-m}^{-1}\left[\sum_{s=1}^{m+1}\sum_{i=k_{s-1}+1}^{k_s-1}\binom{i-s}{r-s+1}\delta'_{m,s,i}(r)x_i \right.$$
$$\left. + \sum_{s=0}^{m+1}\binom{k_s-s}{r-s+1}\binom{N-k_s-m+s}{n-r-m+s-1}x_{k_s}\right],$$

where

$$\delta'_{m,s,i}(r) = \binom{N-i-m+s}{n-r-m+s-1} - \binom{N-i-m+s-1}{n-r-m+s-1}$$

and

$$\mathbf{E}(X_{(r)}|A_m) = \binom{N-m}{n-m}^{-1}\left[\sum_{s=1}^{m+1}\sum_{i=k_{s-1}+1}^{k_s-1}\delta''_{m,s,i}(r)\binom{N-i-m+s-1}{n-r-m+s-1}x_i\right.$$

$$+\sum_{s=0}^{m+1}\binom{k_s-s}{r-s}\binom{N-k_s-m+s}{n-r-m+s}x_{k_s}\Bigg],$$

where

$$\delta''_{m,s,i}(r)=\binom{i-s+1}{r-s+1}-\binom{i-s}{r-s+1}.$$

Then it is easy to see, that for $r=0,\dots,n$, $\mathbf{E}(X_{(r+1)}-X_{(r)}|A_m)$ is the same as in lemma's statement.

Proof of Proposition 1. Applying summation by parts we can write

$$L=\sum_{r=1}^{n-1}\alpha_r\big(X_{(r+1)}-X_{(r)}\big)+\bar{c}\sum_{j=1}^{n}X_j,$$

where $\alpha_r=-\sum_{j=1}^{r}(c_j-\bar{c})$ for $r=1,\dots,n-1$ and $\bar{c}=\frac{1}{n}\sum_{j=1}^{n}c_j$.

Then for $m=1,\dots,n$ we have

$$\begin{aligned}\mathbf{E}(L-\mathbf{E}L|A_m)=&-\sum_{j=1}^{n}c_j\sum_{r=j}^{n}\big\{\mathbf{E}\big(X_{(r+1)}-X_{(r)}|A_m\big)-\mathbf{E}\big(X_{(r+1)}-X_{(r)}\big)\big\}\\&+\bar{c}\Bigg[\sum_{r=0}^{n}r\big\{\mathbf{E}\big(X_{(r+1)}-X_{(r)}|A_m\big)-\mathbf{E}\big(X_{(r+1)}-X_{(r)}\big)\big\}\\&+\frac{N-n}{N(N-m)}\sum_{s=1}^{m}\Bigg(\sum_{i=0}^{k_s-1}i(x_{i+1}-x_i)-\sum_{i=k_s}^{N}(N-i)(x_{i+1}-x_i)\Bigg)\Bigg].\end{aligned}$$

Note that the term in brackets vanishes, because using lemma 1 and changing order of summation, for fixed $m=0,\dots,n$, $s=1,\dots,m+1$, $i=k_{s-1},\dots k_s-1$

$$\begin{aligned}\sum_{r=0}^{n}r\Delta_{m,s,i}(r)&=\sum_{r=s-1}^{n-m+s-1}(r-s+1)\Delta_{m,s,i}(r)+(s-1)\sum_{r=s-1}^{n-m+s-1}\Delta_{m,s,i}(r)\\&=\frac{n-m}{N-m}(i-s+1)+s-1,\end{aligned}$$

where

$$\Delta_{m,s,i}(r)=\binom{N-m}{n-m}^{-1}\binom{i-s+1}{r-s+1}\binom{N-m-(i-s+1)}{n-m-(r-s+1)},$$

and the remaining verifying is quite simple.

Applying of Vandermonde's identity completes the proof.

Proof of Theorem 1. (i) For chosen $1 \leqslant k \leqslant N$ using proposition 1 for $m = 1$ and lemma 1 for $m = 0; 1$ from (3) we have

$$g_1(x_k) = \binom{N-2}{n-1}^{-1} \sum_{j=1}^{n} c_j \sum_{r=0}^{j-1} \left\{ \sum_{i=1}^{k-1} \frac{i}{N} \theta_{21}(i,r)(x_{i+1} - x_i) - \sum_{i=k}^{N-1} \left(1 - \frac{i}{N}\right) \theta_{22}(i,r)(x_{i+1} - x_i) \right\},$$

where

$$\theta_{21}(i,r) = \frac{N}{i}\binom{i}{r}\left\{\binom{N-i-1}{n-r-1} - \frac{n}{N}\binom{N-i}{n-r}\right\},$$

$$\theta_{22}(i,r) = -\frac{N}{N-i}\binom{N-i}{n-r}\left\{\binom{i-1}{r-1} - \frac{n}{N}\binom{i}{r}\right\}.$$

It is easy to verify that $\theta_{21}(i,r) = \theta_{22}(i,r)$. Next using principle of mathematical induction it is easy to show that for every $j = 1, \ldots, n$

$$\sum_{r=0}^{j-1} \theta_{22}(i,r) = \binom{i-1}{j-1}\binom{N-i-1}{n-j},$$

and the proof of the part (i) follows.

(ii) For chosen $1 \leqslant k < l \leqslant N$ using Proposition 1 for $m = 1; 2$ and Lemma 1 for $m = 0; 1; 2$ from (4) we have

$$g_2(x_k, x_l) = \binom{N-4}{n-2}^{-1} \sum_{j=1}^{n} c_j \sum_{r=0}^{j-1} \left\{ \sum_{i=1}^{k-1} \frac{i(i-1)}{(N-1)(N-2)} \theta_{31}(i,r)(x_{i+1} - x_i) - \sum_{i=k}^{l-1} \frac{(i-1)(N-i-1)}{(N-1)(N-2)} \theta_{32}(i,r)(x_{i+1} - x_i) + \sum_{i=l}^{N-1} \frac{(N-i)(N-i-1)}{(N-1)(N-2)} \theta_{33}(i,r)(x_{i+1} - x_i) \right\},$$

where

$$\theta_{31}(i,r) = \frac{(N-1)(N-2)}{i(i-1)}\binom{i}{r}\left\{\binom{N-i-2}{n-r-2} - 2\frac{n-1}{N-2}\binom{N-i-1}{n-r-1} + \frac{n(n-1)}{(N-1)(N-2)}\binom{N-i}{n-r}\right\},$$

$$\theta_{32}(i,r) = -\frac{(N-1)(N-2)}{(i-1)(N-i-1)}\Bigg[\binom{i-1}{r-1}\Bigg\{\binom{N-i-1}{n-r-1} - \frac{n-1}{N-2}\binom{N-i}{n-r}\Bigg\}$$
$$- \frac{n-1}{N-2}\binom{i}{r}\Bigg\{\binom{N-i-1}{n-r-1} - \frac{n}{N-1}\binom{N-i}{n-r}\Bigg\}\Bigg],$$
$$\theta_{33}(i,r) = \frac{(N-1)(N-2)}{(N-i)(N-i-1)}\binom{N-i}{n-r}\Bigg\{\binom{i-2}{r-2} - 2\frac{n-1}{N-2}\binom{i-1}{r-1}$$
$$+ \frac{n(n-1)}{(N-1)(N-2)}\binom{i}{r}\Bigg\}.$$

Similarly $\theta_{31}(i,r) = \theta_{32}(i,r) = \theta_{33}(i,r)$. Next using principle of mathematical induction we can show that for every $j = 1, \dots, n$

$$\sum_{r=0}^{j-1}\theta_{33}(i,r) = \binom{i-2}{j-1}\binom{N-i-2}{n-j-1} - \binom{i-2}{j-2}\binom{N-i-2}{n-j}.$$

Then summation by parts completes the proof of the part (ii).

Acknowledgement. I'm grateful to M. Bloznelis for introduction to the problem and valuable discussions.

References

1. M. Bloznelis, F. Götze. Orthogonal decomposition of finite population statistics and its applications to distributional asymptotics. *Ann. Statist.*, 29:899–917, 2001.
2. H. Putter, W.R. van Zwet. Empirical Edgeworth expansions for symmetric statistics. *Ann. Statist.*, 26(4):1540–1569, 1998.

REZIUMĖ

A. Čiginas. Baigtinių populiacijų L-statistikų ortogonalusis skleidinys

Straipsnyje nagrinėjamas baigtinių populiacijų L-statistikų ortogonalusis skleidinys. Pasiūlomos patogios pirmųjų dviejų skleidinio narių išraiškos.

Raktiniai žodžiai: baigtinė populiacija, ėmimas be grąžinimo, L-statistika, Hoeffding'o skleidinys.