The Korean Communications
in Statistics Vol. 3, No. 3, 1996
pp. 195-204

Rank Test for Ordered Alternatives under Random Censorship[1)]

Gyu-Jin Jeong[2)], Sang-Gue Park[3)]

Abstract

Some rank tests for comparing r treatments against ordered alternatives are proposed when some of data are randomly censored, which are the weighted logrank tests based on pairwise-ranking scheme. The covariances of the proposed test statistics are explicitly obtained from the results of the counting process theory and the test procedures are illustrated by a numerical example. Simulation studies are also performed for comparing with the other well-known tests.

1. Introduction

When comparing r treatments, the experimenter frequently wants to know whether r treatment effects have a increasing(or decreasing) order. For example, when effects of r different dosages for a medicine are compared in clinical trials, one usually expects that taking more medicine will be more effective. This problem has been approached with the ordered alternatives.

There have been many studies for the complete data case. Bartholomew(1961) and Abelson and Tukey(1963) presented some important results under the normal location models. On the other hand, Jonckheere(1954), Tryon and Hettmansperger(1973), Shirahata(1980), Fairly and Fligner(1987) and Chakraborti and Desu(1988) approached this problem nonparametrically. Particularly Jonckheere proposed a nonparametric test based on two sample Wilcoxon-Mann-Whitney statistic, which has been used most popularly in this area.

Censored data are frequently obtained in medical experiments. Several methods for analyzing such data had been studied in testing ordered alternatives. Patel and Hoel(1973) extended Jonckheere's test to censored data by using two sample Gehan(1965) statistic, a censored version of Wilcoxon-Mann-Whitney statistic. Tarone (1975) studied a trend test with r sample logrank test which was originated by Mantel(1966). Recently, Liu et al.(1993) discussed about

1) This paper was supported by 1994 NONDIRECTED RESEARCH FUND, Korea Research Foundation.
2) Professor, Department of Applied Statistics, Hannam University, Taejon, 300-791, Korea
3) Associate Professor, Department of Applied Statistics, Chung-Ang University, Seoul, 156-756, Korea

the ordered weighted log rank test. The weighted logrank test was introduced by Tarone and Ware(1977). They showed that it is the same to the logrank and Gehan test with different weights.

In this paper, the asymptotically distribution-free tests based on pairwise ranking scheme are proposed for the censored data case. The proposed tests would be expressed as some combinations of two sample weighted logrank statistics.

In section 2, the tests are proposed and the asymptotic distributions are examined. The implementation of the test is illustrated in section 3 by a real data example. Further, some simulation results are also discussed in section 4.

2. Test statistics

Let F_i $(i=1,\cdots,r)$ denote the cdf of failure time corresponding to the treatment group i, then the hypotheses of the form,

$$H_0 : F_1 = \cdots = F_r \quad vs \quad H_1 : F_1 \geq \cdots \geq F_r , \tag{1}$$

are of interest, where at least one inequality is strict.

Assume that r=2. Consider the two sample weighted logrank statistics through a conditioning argument based on the number at risk at each observed distinct failure time in the combined sample. Let $t_1 < t_2 < \cdots < t_d$ denote the ordered observed distinct failure times in the combined sample, and let D_{ik} and Y_{ik} $(i=1,2;\ k=1,\cdots,d)$ denote respectively the number of observed failures and number at risk in sample i at time t_k. The data at time t_k can be summarized as follows;

	samples		
Failure	1	2	Total
Yes	D_{1k}	D_{2k}	D_k
No	$Y_{1k}-D_{1k}$	$Y_{2k}-D_{2k}$	Y_k-D_k
Total	Y_{1k}	Y_{2k}	Y_k

From the results of conditional argument, D_{1k} has the expected value $E_{1k}=D_kY_{1k}/Y_k$ and the variance $V_{1k}=D_kY_{1k}Y_{2k}(Y_k-D_k)/\{Y_k^2(Y_k-1)\}$. Then the two sample weighted logrank statistic with the weight function $w(\cdot)$ is expressed as

$$W = \sum_{k=1}^{d} w(t_k)(D_{1k}-E_{1k}) \Big/ \left\{ \sum_{k=1}^{d} w^2(t_k) V_{1k} \right\}^{1/2} .$$

There are three types of weights to be used widely. The logrank statistic W comes to the Gehan statistic if $w(t_k)=Y_k$, the logrank statistic if $w(t_k)=1$ and the Peto-Prentice(Peto and Peto 1972, Prentice 1978) statistic if $w(t_k)=\hat{S}(t_k-)$, where $\hat{S}$ is the Kaplan-Meier estimator of the common survival function under the null hypothesis.

It's ready to propose the test when there are r treatment groups. Let n_i be the sample size of group i and let $N=\sum_{i=1}^{r} n_i$ be the total sample size. Let W_{ij} $(i,j=1,\cdots,r,\ i<j)$ denote the numerator of two sample weighted logrank statistic comparing group i and j, then we consider a class of test statistics

$$V = \sum_{i=1}^{r-1}\sum_{j=i+1}^{r} W_{ij}. \tag{2}$$

This class of statistics is just a generalization of Jonckheere(1954)'s to censored data. One can have as many statistics as he want with suitable weights $w(\cdot)$. Patel and Hoel(1973)'s test statistic belongs to this class with weights $w(t_k)=Y_k$.

Since Jonckheere proposed his test, so many rank tests have been proposed by modifying his test statistics. One can have the following class of statisitcs by combining Tryon and Hettmansper(1973) and Fairley and Fligner(1987)

$$V_p = \sum_{i=1}^{r-1} a_i W_{i,i+1}\ , \tag{3}$$

where $a_i=\sqrt{s_i(1-s_{i-1})}$ with $s_i=\sum_{j=1}^{i} n_j/N$.

The second test statistic (3) consists of r-1 two sample weighted logrank statistics while the first one (2) has $\sum_{i=1}^{r-1}(r-i)$ statistics. However, Tryon and Hettmansperger showed that statistic (3) is asymptotically efficient as Jonckheere's one when $a_i's$ are equal to 1 and observations are not censored, so one might use the unit weights from the point of simplicity. However, Fairley and Fligner showed that the unit weights could be very bad to detect the real differences among treatment groups in the case of no censored observations. Further they proposed a relative maxmin rank test to have good efficiency with weights $a_i's$. This is the reason why we consider the statistic (3).

Since the two sample weighted logrank statistics have the asymptotic standard normal distri bution, the statistics proposed should do, too. Thus, we obtain an approximate size α tests as

$$\text{reject } H_0 \text{ if } V/\sqrt{var(V)} > z_\alpha\ ,$$

$$\text{reject } H_0 \text{ if } V_p/\sqrt{var(V_p)} > z_\alpha ,$$

where z_α is the upper 100α percentile of the standard normal distribution.

The null variances $var(V)$ is given in Appendix and $var(V_p)$ is easily obtained from it. It was established by the counting process and martingale theory found in, for example, Fleming and harrington(1991), chapter 3.

There are several tests availabe for r sample problem. The trend test considered in Tarone(1975) is based on the r sample logrank statistic with the ordered weights: $\sum_{i=2}^{r}\sum_{k} c_i (D_{ik}-E_{ik})$, where k corresponds to the k-th ordered observed distinct failure time t_k in the combined sample 1 through r, D_{ik} and E_{ik} are respectively the observed number of failures and the expected number of failures in the sample i at t_k. The weights c_i's have the same order with F_i's in H_1, for example, $w_i = N-i$. The variance formula of this statistic can be found in Tarone(1975) or Miller(1981). One can easily generalize the Tarone's test statistics to the following; $\sum_{i=2}^{r}\sum_{k} c_i w(t_k)(D_{ik}-E_{ik})$. It might be called the weighted logrank statistics based on the joint ranking scheme comparing with statistics (2), or weighted Tarone's statistics.

Recently, Liu et al.(1993) proposed a test of the form $\sum_{i=1}^{r-1} W_{i+}$, where W_{i+} is the numerator of a two sample weighted logrank statistic comparing group i to the combined groups i+1 through r. The asymptotic relative efficiencies of these statistics were evaluated under the various schemes in Liu et al.(1993).

3. Illustrated example

Let us consider the problem of determining whether or not there is a monotone dose response relationship in animal survival experiments, cited as Patel and Hoel(1973). Since Patel and Hoel used Gehan weights and they showed the significance for the ordered alternatives, it can be expected that the test using logrank weights is significant either. Table 1 shows the Kaplan-Meier survival function estimates of five treatment groups and Table 2 shows $\widehat{Cov}(W_{ij}, W_{uv})$.

As expected, all tests are significant at level α = 0.05; p-values of tests based on Tarone's statistic with weights (4, 3, 2, 1, 0), V, and Liu et al.'s statistic are all less than 0.0001, but

the p-value of test based on V_p is $P(V_p > 2.0982) \fallingdotseq 0.0191$. This is rather surprised comparing $P(V > 4.8476) < 0.0001$. The test based on V_p looks good, because less computation comparing with V and same asymptotic efficiency. We think the test based on V_p might lose some efficiency in the small sample sizes. However, from table 2, one can make the direct comparison between the specific two treatments as well as testing of orderness.

Table 1. 5 estimated Kaplan-Meier survival functions

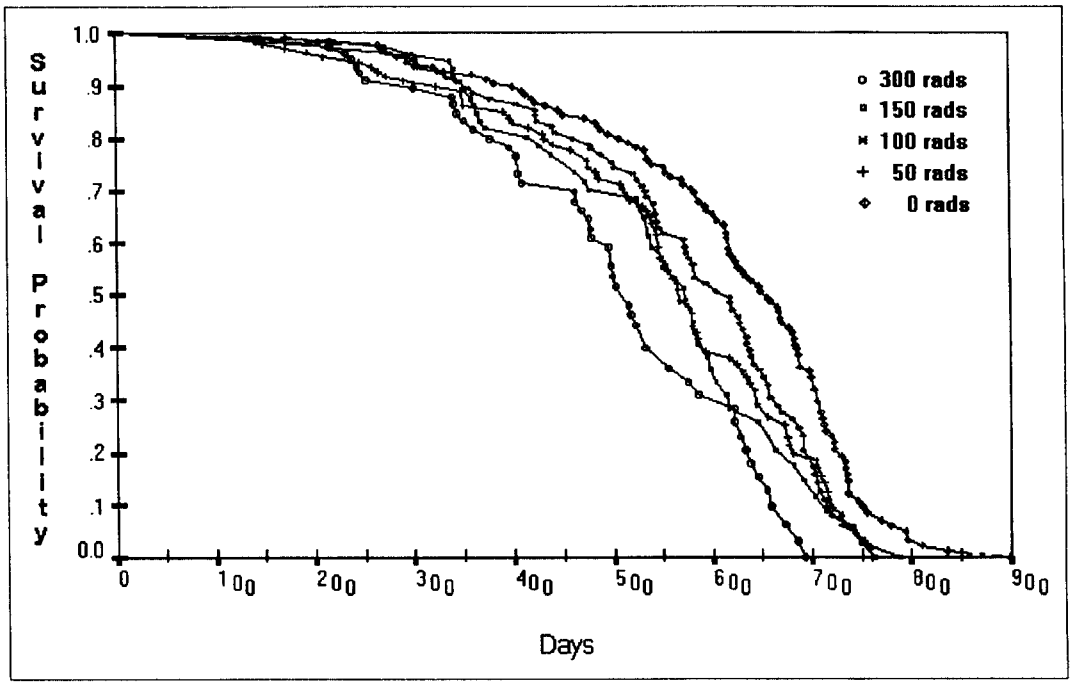

Table 2. $\widehat{Cov}(W_{ij}, W_{uv})$

	W_{12}	W_{13}	W_{14}	W_{15}	W_{23}	W_{24}	W_{25}	W_{34}	W_{35}	W_{45}
W_{12}	22.97	13.04	13.90	11.40	-8.12	-9.12	-5.45	0.00	0.00	0.00
W_{13}		22.87	15.34	13.34	11.95	0.00	0.00	-11.36	-7.54	0.00
W_{14}			25.15	13.96	0.00	12.64	0.00	10.79	0.00	-7.76
W_{15}				18.93	0.00	0.00	10.77	0.00	10.47	11.27
W_{23}					28.50	18.67	17.28	-14.73	-9.67	0.00
W_{24}						30.60	17.81	14.26	0.00	-9.91
W_{25}							24.80	0.00	14.42	15.21
W_{34}								40.76	21.41	-15.20
W_{35}									40.29	21.84
W_{45}										40.62

4. Simulation and Conclusion

The simulation experiments were carried out to evaluate the performance of tests mentioned in section 2 for practical sample sizes. Four test statistics were considered with logrank weights; T_1, Tarone(1975)'s statistics with linear weights (r-1, r-2, ... , 0), T_2, a generalization of Jonckheere's statistics V, T_3, a generalization of Tryon and Hettmansperger(1973)'s statistics V_p with fairley and fligner's weights, T_4, Liu et al.'s statistics.

From the weighted logrank test statistics proposed, various test statistics with suitable weights can be easily available. So our attention was restricted to exponential case with logrank weights. It was repeated 1,000 times to compare approximate powers. The subroutine IMSL GGUBS was used to generate the exponential random numbers and assumed 10% censoring rate(β=10%).

In the simulation studies, It was found that the empirical sizes of tests had some variation so that the experiment was repeated 5 times and then they were averaged. The empirical sizes relatively maintained well the given significance level α=0.05. Some results are summarized in Table 3 and 4 for equal sample sizes 30 and 40(n=30, 40). The upper lines represent for sample size 30.

It seems that most recommendable one is the test based on T_2, a generalization of Jockheere's test. Tarone's test based on T_1 seems to depend on weights a lot and Liu et al.'s test based on T_4 was insensitive to the case that many treatments are similarly effective. It was expected that the test based on T_3 performed well, but it did not work like the example in section 3. It seemed to be good to the case that treatments are similarly effective, but it showed poor empirical powers in other cases. Though Tryon and Hettmansperger showed asymptotically equivalence, it was most likely there were surely some loss of efficiency in small sample sizes.

A class of tests based on the pairwise-ranking scheme was proposed. Many researcher prefer the rank tests based on pairwise ranking scheme to ones in joint ranking scheme, since they give some information of differences between two treatments as well as differences among all treatments. The test procedures proposed contain more complex computation relative to others. One might argue that this procedure is a kind of complicated one and needs times to obtian the test statistics, but any procedures for testing the ordered alternatives depend on computer, especially the case of censored data. We hope it is useful for some researches. The Fortran programs are available from the second author.

Table 3. Simulation results α = 0.05, r = 4, β = 10%, n=30, 40
(Population distribution : exponential(λ_i), Censoring distribution : exponential)

(λ_1, λ_2, λ_3, λ_4) \ Tests	T_1	T_2	T_3	T_4
0.01 0.01 0.01 0.01	0.060	0.062	0.059	0.060
	0.050	0.052	0.053	0.054
0.01 0.01 0.01 0.009	0.123	0.114	0.120	0.130
	0.114	0.117	0.118	0.131
0.01 0.01 0.009 0.009	0.139	0.140	0.112	0.132
	0.139	0.139	0.114	0.140
0.01 0.009 0.009 0.009	0.110	0.105	0.104	0.099
	0.102	0.109	0.100	0.100
0.01 0.009 0.009 0.008	0.202	0.196	0.209	0.207
	0.227	0.224	0.223	0.228
0.01 0.009 0.008 0.008	0.238	0.240	0.203	0.204
	0.271	0.269	0.218	0.237
0.01 0.009 0.008 0.007	0.419	0.416	0.367	0.417
	0.459	0.450	0.423	0.432

Table 4. Simulation results α = 0.05, r = 5, β = 10%, n=30, 40
(Population distribution : exponential(λ_i), Censoring distribution : exponential)

(λ_1, λ_2, λ_3, λ_4, λ_5) \ Tests	T_1	T_2	T_3	T_4
0.01 0.01 0.01 0.01 0.01	0.059	0.059	0.052	0.057
	0.055	0.052	0.045	0.047
0.01 0.01 0.01 0.01 0.009	0.114	0.110	0.106	0.121
	0.114	0.116	0.112	0.146
0.01 0.01 0.01 0.009 0.009	0.142	0.146	0.110	0.134
	0.149	0.149	0.115	0.151
0.01 0.01 0.009 0.009 0.009	0.139	0.141	0.102	0.120
	0.167	0.167	0.107	0.142
0.01 0.009 0.009 0.009 0.009	0.104	0.105	0.092	0.087
	0.134	0.135	0.112	0.128
0.01 0.009 0.009 0.009 0.008	0.191	0.187	0.169	0.187
	0.216	0.210	0.195	0.225
0.01 0.01 0.009 0.008 0.008	0.226	0.226	0.170	0.213
	0.285	0.284	0.238	0.254
0.01 0.009 0.008 0.008 0.007	0.381	0.380	0.316	0.357
	0.468	0.457	0.399	0.450
0.01 0.009 0.008 0.007 0.006	0.673	0.663	0.569	0.662
	0.784	0.777	0.682	0.766

References

[1] Abelson, R.P. and Tukey, J. W. (1963). Efficient utilization of non-numerical information in quantitive analysis: General theory and the case of simple order, *Annals of Mathematical Statistics,* 34, 1347-1369.

[2] Bartholomew, D.J. (1961). A test of homogeneity of means under restricted alternatives (with discussion), *Journal of the Royal Statistical Society, B,* 23, 239-281.

[3] Chakraborti, S. and Desu, M. M. (1988). A class of distribution-tests for testing homogeneity against ordered alternatives, *Statistics and Probability Letters,* 6, 251-256.

[4] Fairley, D. and Fligner, M. (1987). Linear rank statistics for the ordered alternatives problem, *Communications in Statistics, Theory and Methods,* 16, 1-16.

[5] Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis,* John Wiley and Son, Inc. New York.

[6] Gehan, E.A. (1965). A generalized Wicoxon test for comparing arbitrarily single censored sample, *Biometrika,* 52, 203-23.

[7] Joncheere, A.R. (1954). A distribution-free k-sample test against ordered alternatives, *Biometrika,* 41, 133-145.

[8] Liu, P.Y., Green, S., Wolf, M. and Crowley, J. (1993). Testing against ordered alternatives for censored survival data, *Journal of the American Statistical Association,* 88, 153-160.

[9] Mantel N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemotherapy Report,* 50, 163-170.

[10] Patel K.M. and Hoel D.G. (1973). A generalized Jonckheere k-sample test against ordered alternatives when observations are subject to arbitrary right censorship, *Communications in Statistics, Theory and Methods,* 2, 373-380.

[11] Peto, R. and Peto, R. (1972). Asymptotically efficient rank invariant test procedures (with discussion), *Journal of the Royal Statistical Society,* A, 135, 185-206.

[12] Prentice, R.L. (1978). Linear rank tests with right censored data. Biometrika, 65, 169-79.

[13] Shirahata, S. (1980). Rank tests for the k-sample problems with restricted alternatives, *Communications in Statistics, Theory and Methods,* 9, 1071-1086.

[14] Tarone, R.E. (1975). Tests for trend in life table analysis, *Biometrika,* 62, 679-682.

[15] Tarone, R.E. and Ware, J. (1977). On distribution-free tests for equality of survival distributions, *Biometrika,* 64, 156-60.

[16] Tryon, P. V. and Hettmansperger, T. P. (1973). A class of nonparametric tests for homogeneity against ordered alternatives, *Annals of Statistics,* 1, 1061-1070.

Appendix : computation of the null variance estimator $var(V)$

Let $var(V)$ denote the null variance of V, then it is represented with the variances and covariances of $W_{ij}'s$ as follows;

$$\begin{aligned} Var(V) = & \sum_{i=1}^{r-1}\sum_{j=i+1}^{r} Var(W_{ij}) + 2\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\sum_{v=j+1}^{r} Cov(W_{ij}, W_{iv}) \\ & + 2\sum_{i=1}^{r-1}\sum_{v=i+1}^{r-1}\sum_{j=v+1}^{r} Cov(W_{ij}, W_{uj}) + 2\sum_{i=1}^{r-1}\sum_{j=i+1}^{r-1}\sum_{v=j+1}^{r} Cov(W_{ij}, W_{jv}) \end{aligned} \tag{A.1}$$

In order to obtain the expression of $Var(V)$, thus, we have to consider the sample combined two or three groups. These combined groups will be indicated through the superscripts or subscripts. Let $t_1^{ijv} < \cdots < t_{d_{ijv}}^{ijv}$ be the ordered observed failure times in the sample formed by combining groups i, j and v, then the data at time t_k can be summarized as follows;

	samples			
Failure	i	j	v	Total
Yes	D_{ik}^{ijv}	D_{jk}^{ijv}	D_{vk}^{ijv}	D_k^{ijv}
No	$Y_{ik}^{ijv} - D_{ik}^{ijv}$	$Y_{jk}^{ijv} - D_{jk}^{ijv}$	$Y_{vk}^{ijv} - D_{vk}^{ijv}$	$Y_k^{ijv} - D_k^{ijv}$
Total	Y_{ik}^{ijv}	Y_{jk}^{ijv}	Y_{vk}^{ijv}	Y_k^{ijv}

Through the martingale and counting process formulation, which can be found in Fleming and Harrington(1991), the estimators of four terms in (A.1) are obtained as

$$Var(W_{ij}) = E_0 \int_0^{\infty} w_{ij}^2(s) \frac{Y_i(s)\, Y_j(s)}{Y_i(s) + Y_j(s)} \{1 - \Delta\Lambda(s)\} d\Lambda(s), \quad i < j$$

where $w_{ij(s)}$ is the weight at time s, $Y_i(s)$ and $Y_j(s)$ are the numbers at risk in sample i and j, respectively, at time s, and Λ is the cumulative hazard function.

Replacing the integral with the sum over the observed distinct failure times and $\{1 - \Delta\Lambda(s)\} d\Lambda(s)$ with

$$(1-\frac{D_{ik}^{ij}+D_{jk}^{ij}-1}{Y_{ik}^{ij}+Y_{jk}^{ij_1}})\frac{D_{ik}^{ij}+D_{jk}^{ij}}{Y_{ik}^{ij}+Y_{jk}^{ij}}=(1-\frac{D_k^{ij}-1}{Y_k^{ij}-1})\frac{D_k^{ij}}{Y_k^{ij}},$$

the estimator $\widehat{Var}(W_{ij})$ of $Var(W_{ij})$ is given by

(1) $$\widehat{Var}(W_{ij})=\sum_{k=1}^{d_{ij}}w_{ij}^2(t_k^{ij})\frac{Y_{ik}^{ij}Y_{jk}^{ij}}{Y_k^{ij}}(1-\frac{D_k^{ij}-1}{Y_k^{ij}-1})\frac{D_k^{ij}}{Y_k^{ij}}.$$

By considering the three-sample estimator for $\{1-\Delta\Lambda(s)\}d\Lambda(s)$, the covariance estimators are of the following forms;

(2) for $i<j<v$,

$$\widehat{Cov}(W_{ij},W_{iv})=\sum_{k=1}^{d_{ijv}}w_{ij}(t_k^{ijv})w_{iv}(t_k^{ijv})\frac{Y_{ik}^{ijv}Y_{jk}^{ijv}Y_{vk}^{ijv}}{(Y_{ik}^{ijv}+Y_{jk}^{ijv})(Y_{ik}^{ijv}+Y_{vk}^{ijv})}(1-\frac{D_k^{ijv}-1}{Y_k^{ijv}-1})\frac{D_k^{ijv}}{Y_k^{ijv}}.$$

(3) for $i<v<j$,

$$\widehat{Cov}(W_{ij},W_{vj})=\sum_{k=1}^{d_{ijv}}w_{ij}(t_k^{ijv})w_{vj}(t_k^{iju})\frac{Y_{ik}^{ijv}Y_{jk}^{ijv}Y_{vk}^{ijv}}{(Y_{ik}^{ijv}+Y_{jk}^{ijv})(Y_{vk}^{ijv}+Y_{jk}^{ijv})}(1-\frac{D_k^{ijv}-1}{Y_k^{ijv}-1})\frac{D_k^{ijv}}{Y_k^{ijv}}.$$

(4) for $i<j<v$,

$$\widehat{Cov}(W_{ij},W_{jv})=-\sum_{k=1}^{d_{ijv}}w_{ij}(t_k^{ijv})w_{jv}(t_k^{ijv})\frac{Y_{ik}^{ijv}Y_{jk}^{ijv}Y_{vk}^{ijv}}{(Y_{ik}^{ijv}+Y_{jk}^{ijv})(Y_{jk}^{ijv}+Y_{vk}^{ijv})}(1-\frac{D_k^{ijv}-1}{Y_k^{ijv}-1})\frac{D_k^{ijv}}{Y_k^{ijv}}.$$

Plugging the estimators (1) through (4) into (A.1), the estimator $var(V)$ can be obtained.