

Extension of the Mantel-Haenszel test to bivariate interval censored data

Dong-Hyun Lee^a, Yang-Jin Kim^{1,a}

^aDepartment of Statistics, Sookmyung Women's University, Korea

Abstract

This article presents an independence test between pairs of interval censored failure times. The Mantel-Haenszel test is commonly applied to test the independence between two categorical variables accompanied with a strata variable. Hsu and Prentice (1996) applied a Mantel-Haenszel test to the sequence of 2×2 tables formed at the grids which are composed of failure times. In this article, due to unknown failure times, the suitable grid points should be determined and the status of failure and at risk are estimated at those grid points. We also consider a weighted test statistic to bring a more powerful test. Simulation studies are performed to evaluate the power of test statistics under finite samples. The method is applied to analyze two real data sets, mastitis data from milk cows and an age-related eye disease study.

Keywords: bivariate interval censored data, Clayton model, Gumbel model, independence test, Mantel-Haenszel test

1. Introduction

Interval censored data commonly arise in observational studies where the exact occurrence time of the event is unobservable but two inspection times including it are available. When (T_{i1}, T_{i2}) is a pair of failure times from the i th subject, a bivariate interval censored data is composed of two pairs of inspection times $\{(L_{i1}, R_{i1}] \times (L_{i2}, R_{i2}], i = 1, \dots, n\}$, where the inspection times satisfy $L_{ik} < T_{ik} < R_{ik} (k = 1, 2)$ and are assumed to be independent of the failure times.

For right censored data, several approaches for testing the independence have been proposed. Oakes (1982) developed a concordance test based on the estimated Kendall's τ and Shih and Louis (1996) constructed a test based on the martingale residual and considered its extension by implementing several weights to reflect a time-varying association. Hsu and Prentice (1996) suggested a generalization of the Mantel-Haenszel test by combining the sequence of 2×2 tables formed at the grids on a plane where the grids are composed of a pair of observed failure times.

These above approaches have been extended to test the independence of bivariate interval censored data. Betensky and Finkelstein (1999) applied a multiple imputation technique for missing failure times and then implemented the method Oakes (1982) proposed. In their approach, simulated bivariate failure times are generated from the estimated joint survival distribution and are utilized to calculate Kendall's τ . Sun *et al.* (2006) estimated the association measure of a Clayton copula by extending Wang and Ding (2000)'s two-stage procedures. As a parametric approach, Bogaerts and Lesaffre

This work was supported by a Korea Research Grant (NRF-2020R1A2C1A01100755).

¹ Corresponding author: Department of Statistics, Sookmyung Women's University, Yongsan-Gu, Cheongpa-ro 47-gil 100 (Cheongpa-dong 2ga), Seoul 04310, Korea. E-mail: yjin@sookmyung.ac.kr

Published 31 July 2022 / journal homepage: <http://csam.or.kr>

© 2022 The Korean Statistical Society, and Korean International Statistical Society. All rights reserved.

(2008) assumed a bivariate log-normal distribution for (T_1, T_2) and derived a two stage procedure to calculate a local measure such as a cross ratio function. For testing the association, Ding and Wang (2004) constructed test statistics for bivariate current status data, which is sometimes denoted as interval censoring case I by extending the Hsu and Prentice (1996)'s method. In this article, we extend the Mantel-Haenszel test to bivariate interval censored data by constructing dependent 2×2 tables formed at the emulated grid points and by assigning the estimated number of failures and that of at-risk to the cell of each table. This article is organized as follows. In Section 2, we introduce the model and develop the estimation procedure. Simulation and data analysis are given in Section 3 and 4, respectively. A discussion about the suggested method is provided in Section 5.

2. Independence test for bivariate interval censored data

2.1. Preliminary

In this paper, our interest is to develop a nonparametric procedure for testing the independence of bivariate interval censored data. Denote $F(t_1, t_2) = \Pr(T_1 \leq t_1, T_2 \leq t_2)$, $F_1(t_1) = \Pr(T_1 \leq t_1)$ and $F_2(t_2) = \Pr(T_2 \leq t_2)$ as the joint distribution and marginal distributions of (T_1, T_2) , respectively. When $H_0 : T_1 \perp T_2$ is true, then $F(t_1, t_2) = F_1(t_1)F_2(t_2)$ is valid. For bivariate right censored data, $\{(X_{i1}, \delta_{i1}), (X_{i2}, \delta_{i2}), i = 1, \dots, n\}$ denotes observation times and censoring indicators, respectively, where $X_{ik} = \min(T_{ik}, C_{ik})$ for failure time T_{ik} and the censoring time C_{ik} and $\delta_{ik} = I(T_{ik} \leq C_{ik})$, $k = 1, 2$. The grid points are formatted using the observed failure times X_{ik} satisfying $\delta_{ik} = 1$ and composed of $v_{11} < v_{12} < \dots < v_{1m_1}$ for the first event and $v_{21} < v_{22} < \dots < v_{2m_2}$ for the second one, respectively. Thus, the information from $m_1 \times m_2$'s tables are evaluated. For each grid point $\{(v_{1r}, v_{2l}), r = 1, \dots, m_1, l = 1, \dots, m_2\}$, $\mathbf{T}(v_{1r}, v_{2l})$ denotes a 2×2 table with a value (0,1) representing a non-failure and a failure, respectively. Then, the four cells at $\mathbf{T}(v_{1r}, v_{2l})$ are denoted as $(D_{00}(v_{1r}, v_{2l}), D_{01}(v_{1r}, v_{2l}), D_{10}(v_{1r}, v_{2l}), D_{11}(v_{1r}, v_{2l}))$, where,

$$D_{ab}(v_{1r}, v_{2l}) = \sum_{i=1}^n Y_{i1}(v_{1r}) Y_{i2}(v_{2l}) I(dN_{i1}(v_{1r}) = a, dN_{i2}(v_{2l}) = b), \quad a = 0, 1; b = 0, 1,$$

and where $Y_{ik}(s) = I(X_{ik} \geq s)$, $k = 1, 2$ denotes a risk process and $N_{ik}(s) = I(T_{ik} < s, \delta_{ik} = 1)$ denotes a counting process with $dN_{ik} = N_{ik}(s) - N_{ik}(s-)$, respectively. Then the margins are defined as $D_{+b}(v_{1r}, v_{2l}) = D_{0b}(v_{1r}, v_{2l}) + D_{1b}(v_{1r}, v_{2l})$ and $D_{a+}(v_{1r}, v_{2l}) = D_{a0}(v_{1r}, v_{2l}) + D_{a1}(v_{1r}, v_{2l})$. For example, $D_{1+}(v_{1r})$ denotes the total number of the first event at v_{1r} and $D_{+1}(v_{2r})$ denotes the total number of the second event at v_{2r} , respectively. Let $\bar{D}_{rl} = D_{00}(v_{1r}, v_{2l}) + D_{01}(v_{1r}, v_{2l}) + D_{10}(v_{1r}, v_{2l}) + D_{11}(v_{1r}, v_{2l})$ be the number of pairs at risk at (v_{1r}, v_{2l}) . Conditional on the margins, $D_{11}(v_{1r}, v_{2l})$ follows a hypergeometric distribution with the following conditional expectation

$$E_{11}(v_{1r}, v_{2l}) = \frac{D_{1+}(v_{1r}, v_{2l}) D_{+1}(v_{1r}, v_{2l})}{\bar{D}_{rl}}.$$

If all tables are independent, the summation of the differences between observed and expected numbers would follow an asymptotic normal distribution. Since all the tables are formed at failure times and are related through the risk processes, they are not independent, and thus, the sum does not have a same asymptotic property as that of an ordinary Mantel-Haenszel test statistic. Hsu and Prentice (1996) rewrote the Mantel-Haenszel test statistic in terms of a counting process in order to derive the

large sample properties based on the martingale theory,

$$\begin{aligned} \tilde{S} &= \sum_{r=1}^{m_1} \sum_{l=1}^{m_2} \left(D_{11}(v_{1r}, v_{2l}) - \frac{D_{+1}(v_{1r})D_{1+}(v_{2l})}{\bar{D}_{rl}} \right) \\ &= \sum_{i=1}^n \int_0^{\tau_1} \int_0^{\tau_2} \left(N_{i1}(dt_1)N_{i2}(dt_2) - \frac{1}{n} \bar{R}(t_1, t_2) N_{\cdot 1}(dt_1)N_{\cdot 2}(dt_2) \right) \\ &= \sum_{i=1}^n \int_0^{\tau_1} \int_0^{\tau_2} \left(\hat{M}_{i1}(dt_1)\hat{M}_{i2}(dt_2) - \frac{1}{n} \bar{R}(t_1, t_2) \hat{M}_{\cdot 1}(dt_1)\hat{M}_{\cdot 2}(dt_2) \right), \end{aligned} \tag{2.1}$$

where τ_k denotes the maximum follow-up time of a type k event. $\bar{R}(t_1, t_2) = \sum_{i=1}^n I(X_{i1} \geq t_1, X_{i2} \geq t_2)$ denotes the total number of at-risk subjects at (t_1, t_2) and a martingale $\hat{M}_{ik}(dt)$ is defined as $\hat{M}_{ik}(dt) = N_{ik}(dt) - Y_{ik}\hat{\Lambda}_k(dt)$ with the Nelson-Aalen estimate $\hat{\Lambda}_k(t) = \sum_{X_{jk} \leq t} dN_{jk}(t)/Y_{jk}(t)$ where $Y_{jk}(t) = \sum_{l=1}^{m_k} I(X_{jl} \leq t)$.

2.2. Suggested method

The first step to apply the Mantel-Haenszel test to the interval censored data is to format suitable grid points due to unknown exact failure times. In our approach, grid points are composed of the pseudo failure times derived from the interval censored data, $\{(L_{i1}, R_{i1}, L_{i2}, R_{i2}), i = 1, \dots, n\}$. These pseudo failure times $\{s_{k1} < s_{k2} < \dots < s_{km_k}, k = 1, 2\}$ are constructed from the equivalence set $\{(l_1^k, r_1^k)_{l=1}^{m_k}\}$ of the univariate interval censored data (Sun, 2006) and defined as $s_{kl} = (l_1^k + r_1^k)/2$.

The second step is to define the test statistic at the grid points by extending (2.1),

$$S = \sum_{r=1}^{m_1} \sum_{l=1}^{m_2} (D_{11}(s_{1r}, s_{2l}) - E_{11}(s_{1r}, s_{2l})), \tag{2.2}$$

where $D_{11}(s_{1r}, s_{2l})$ denotes the number of pairs with both failures at (s_{1r}, s_{2l}) , and $E_{11}(s_{1r}, s_{2l}) = E(D_{11}(s_{1r}, s_{2l})|H_0)$ represents a conditional expectation of D_{11} under H_0 . Here $E_{11}(s_{1r}, s_{2l})$ is estimated as $R(s_{1r}, s_{2l})\lambda_1(s_{1r})\lambda_2(s_{2l})$, where $R(s_{1r}, s_{2l})$ represents the risk set, $\lambda_1(s_{1r})$ and $\lambda_2(s_{2l})$ marginal hazards, respectively. However, $D_{11}(s_{1r}, s_{2l})$ and $E_{11}(s_{1r}, s_{2l})$ are unknown under the interval censored data and calculated by,

$$\begin{aligned} \hat{D}_{11}(s_{1r}, s_{2l}) &= \sum_{i=1}^n \frac{\alpha_{irl} f_{rl}}{\sum_{u=1}^{m_1} \sum_{v=1}^{m_2} \alpha_{iuv} f_{uv}} = \sum_{i=1}^n \tilde{f}_{irl}, \\ \hat{E}_{11}(s_{1r}, s_{2l}) &= r_{rl} \hat{\lambda}_1(s_{1r}) \hat{\lambda}_2(s_{2l}), \end{aligned} \tag{2.3}$$

where $\alpha_{irl} = I(L_{i1} < s_{1r} \leq R_{i1}, L_{i2} < s_{2l} \leq R_{i2})$ and $f_{rl} = \Pr(T_1 = s_{1r}, T_2 = s_{2l})$ is the joint probability mass function. $r_{rl} = \hat{R}(s_{1r}, s_{2l})$ denotes the estimated risk set size at (s_{1r}, s_{2l}) and is obtained by,

$$r_{rl} = \sum_{i=1}^n \sum_{r \leq u} \sum_{l \leq v} \tilde{f}_{iuv}.$$

Let $g_r = \Pr(T_1 = s_{1r})$ and $h_l = \Pr(T_2 = s_{2l})$ be the marginal probability mass functions. Denote $\beta_{ir} = I(L_{i1} < s_{1r} \leq R_{i1})$ and $\gamma_{il} = I(L_{i2} < s_{2l} \leq R_{i2})$. Then λ_1 and λ_2 are estimated by

$$\hat{\lambda}_1(s_{1r}) = \frac{\hat{D}_1(s_{1r})}{Y_1(s_{1r})}, \quad \hat{\lambda}_2(s_{2l}) = \frac{\hat{D}_2(s_{2l})}{Y_2(s_{2l})},$$

where for the first event

$$\hat{D}_1(s_{1r}) = \sum_{i=1}^n \frac{\beta_{ir}g_r}{\sum_{u=1}^{m_1} \beta_{iu}g_u} = \sum_{i=1}^n \tilde{g}_{ir}, \quad Y_1(s_{1r}) = \sum_{i=1}^n \sum_{r \leq u} \tilde{g}_{iu} \tag{2.4}$$

and for the second event,

$$\hat{D}_2(s_{2l}) = \sum_{i=1}^n \frac{\gamma_{il}h_l}{\sum_{v=1}^{m_2} \beta_{iv}h_v} = \sum_{i=1}^n \tilde{h}_{il}, \quad Y_1(s_{2l}) = \sum_{i=1}^n \sum_{l \leq v} \tilde{h}_{iv}. \tag{2.5}$$

Finally, the joint probability and marginal probabilities defined at (2.3), (2.4) and (2.5) are estimated by a well-known self-consistency algorithm (Sun, 2006; Maathuis, 2013). The likelihood function

$$L(f) = \prod_{i=1}^n \sum_{r=1}^m \sum_{l=1}^m \alpha_{irl} f_{rl},$$

where the NPMLE of f is determined by maximizing the $L(f)$ subject to $f_{rl} > 0$ and $\sum_{rl} f_{rl} = 1$. After obtaining \hat{f}_{rl} , the marginal probability mass functions are defined as $\hat{g}_r = \sum_l \hat{f}_{rl}$ and $\hat{h}_l = \sum_r \hat{f}_{rl}$, respectively.

The explicit estimation for the variance of test statistic (2.2) is technically complicated since it is related with the complexity of the nonparametric maximum likelihood estimators (NPMLEs) of f_{kl} , g_k and h_l . Therefore, the bootstrap method is adopted to obtain the variance estimation in this article. In detail, we generated a bootstrap sample $\{(L_{i1}^*, R_{i1}^*, L_{i2}^*, R_{i2}^*, \delta_{i1}^*, \delta_{i2}^*), i = 1, \dots, n\}$ from the original sample. The procedure is repeated B times and brings $S^{*b} = \sum_{k=1}^{m_1} \sum_{l=1}^{m_2} (D_{11}^*(s_{1k}, s_{2l}) - E_{11}^*(s_{1k}, s_{2l}))$, $b = 1, \dots, B$ as the bootstrap statistics for (2.2). Then, the variance was calculated by the sample variance of S^{*b} , $b = 1, \dots, B$. That is,

$$\sigma^{2*} = \sum_{b=1}^B \frac{(S^{*b} - \bar{S}^*)^2}{B-1}, \quad \bar{S}^* = \sum_{b=1}^B \frac{S^{*b}}{B}.$$

The resulting test statistic $Q^* = S / \sqrt{\sigma^{2*}}$ is assumed to follow a standard normal distribution under the null hypothesis. Here $B = 200$ is implemented.

In order to improve the power of a test, the above test statistics can be generalized by implementing a weight function $\hat{W}(t_1, t_2)$,

$$S_W = \sum_{k=1}^{m_1} \sum_{l=1}^{m_2} \hat{W}(s_{1k}, s_{2l}) (D_{11}(s_{1r}, s_{2l}) - E_{11}(s_{1r}, s_{2l})), \tag{2.6}$$

where the weight function can be chosen to emphasize the difference at the specific regions during follow-up times. For example, $\hat{W}(t_1, t_2) = \{\hat{S}_1(t)\hat{S}_2(t_2)\}^\rho \{(1 - \hat{S}_1(t_1))(1 - \hat{S}_2(t_2))\}^\lambda$, where the $G^{\rho,\lambda}$ class was introduced by Harrington and Fleming (1982).

3. Simulation study

We performed a series of simulations to evaluate the finite sample performance of the proposed tests. Bivariate failure times (T_1, T_2) are generated from the Clayton family and the Gumbel family using

Table 1: Empirical powers of Unweighted (U) and Weighted (W) test under Clayton model

n	Censoring	W	τ : Association								
			0	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
50	0%	U	0.050	0.086	0.202	0.284	0.420	0.552	0.640	0.785	0.830
		W	0.040	0.062	0.130	0.270	0.432	0.632	0.797	0.868	0.936
	40%	U	0.044	0.088	0.124	0.147	0.256	0.331	0.522	0.616	0.774
		W	0.038	0.086	0.120	0.238	0.328	0.504	0.702	0.824	0.918
100	0%	U	0.058	0.124	0.283	0.488	0.766	0.900	0.980	0.986	0.997
		W	0.056	0.108	0.286	0.510	0.774	0.938	1.000	1.000	1.000
	40%	U	0.054	0.110	0.142	0.288	0.446	0.530	0.723	0.912	0.996
		W	0.050	0.102	0.200	0.368	0.636	0.816	0.932	0.974	0.998
200	0%	U	0.047	0.198	0.400	0.748	0.916	1.000	1.000	1.000	1.000
		W	0.056	0.188	0.494	0.866	0.986	1.000	1.000	1.000	1.000
	40%	U	0.060	0.110	0.250	0.470	0.714	0.910	0.974	0.995	1.000
		W	0.058	0.132	0.362	0.688	0.916	0.984	1.000	1.000	1.000

Table 2: Empirical powers of Unweighted (U) and Weighted (W) test under Gumbel model

n	Censoring	W	τ : Association								
			0	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
50	0%	U	0.069	0.098	0.098	0.204	0.244	0.346	0.468	0.584	0.656
		W	0.044	0.086	0.092	0.148	0.272	0.372	0.536	0.644	0.726
	40%	U	0.056	0.058	0.056	0.092	0.108	0.140	0.184	0.260	0.330
		W	0.046	0.072	0.112	0.204	0.360	0.524	0.672	0.818	1.000
100	0%	U	0.068	0.100	0.140	0.266	0.426	0.562	0.680	0.850	0.928
		W	0.048	0.110	0.192	0.368	0.468	0.748	0.832	0.954	0.988
	40%	U	0.052	0.056	0.074	0.092	0.148	0.250	0.295	0.378	0.478
		W	0.048	0.088	0.208	0.368	0.640	0.818	0.922	0.974	0.998
200	0%	U	0.050	0.140	0.273	0.426	0.700	0.847	0.943	0.997	1.000
		W	0.070	0.090	0.350	0.626	0.856	0.954	1.000	1.000	1.000
	40%	U	0.050	0.060	0.065	0.190	0.227	0.390	0.437	0.590	0.750
		W	0.066	0.138	0.382	0.676	0.918	0.982	0.998	1.000	1.000

R package **CopulaCenR** (Sun and Ding, 2019). (i) A Clayton family is known as a gamma frailty model, where the cross ratio $\theta(t_1, t_2) = \lambda_1(t_1|T_2 = t_2)/\lambda_1(t_1|T_2 > t_2) = \lambda_2(t_2|T_1 = t_1)/\lambda_2(t_2|T_1 > t_1)$ is constant over the time and (ii) a Gumbel family has a decreasing cross ratio with increasing t_1 or t_2 .

We generated 10's random numbers $h_1 < \dots < h_{10}$ from the uniform distribution $U(0, q)$ to construct the inspection times satisfying $L_k = h_{l-1} < T_k < h_l = R_k$. Here, q is determined to set a given censoring rate. Two cases were considered, no right censoring (0%) and 40% right censoring on each margin. We calculated test statistics and corresponding variance in order to evaluate the performance by the empirical power, which is defined as the relative frequency by which the test rejects the null hypothesis at the 0.05 nominal level based on 500 replicates. The power pattern was investigated under the combination of nine τ values ranging from 0 to 0.4 and three sample sizes ($n = 50, 100, 200$). The results of two test statistics (unweighted and weighted test ($W(t_1, t_2) = S_1(t_1)S_2(t_2)$)) are presented in Table 1 (Clayton family) and in Table 2 (Gumbel family). Figure 1 shows the Q-Q plot of test statistics under (i) unweighted (ii) weighted versions with W to check the normality under $n = 100$ and 40% censoring and those plots show they satisfy the normal distribution.

According to the result of Table 1 (Clayton family with constant cross hazard), the type I errors of the unweighted test are very close to the 0.05 nominal level. Type I errors of the weighted test are smaller than the nominal size at small sample sizes but have similar values with the nominal value as the sample size increases.

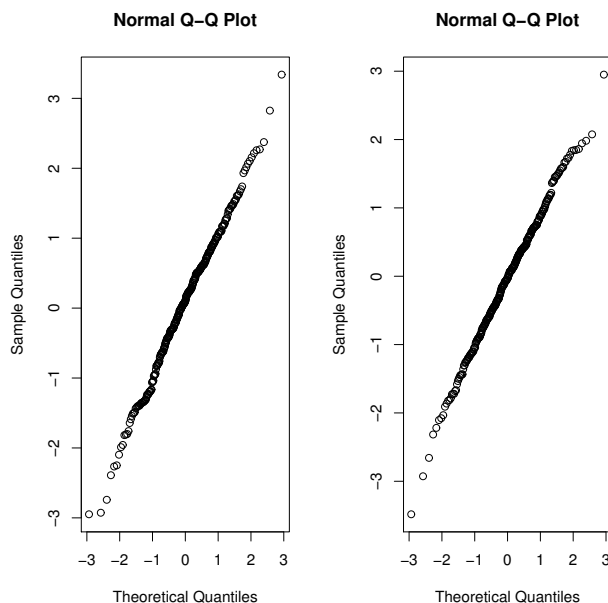


Figure 1: Q-Q plot of test statistic (unweighted and weighted) at $n = 100$.

The powers of both test statistics increase as the association and sample size become greater. Under the 40% censoring rate, two test statistics show smaller powers compared with the no censoring case. However, the weighted test statistic shows a higher power than the unweighted one as the association gets stronger and the sample size increases.

Table 2 shows the result related with the Gumbel family and indicates an overestimated type I error with a small sample size and no censoring. In particular, the weighted test statistic gives a remarkable power improvement even when the censoring rate is large. Although the sample size increases, the unweighted test statistic does not show a significant improvement of power under heavy censoring.

4. Data analysis

4.1. Mastitis data

Mastitis in dairy cattle is an inflammation of the udder and reduces the milk production and the quality of the milk (Bogaert *et al.*, 2018). In the analysis, 100 cows were screened monthly at the udder quarter level for bacteria infection. Monthly visits were independently performed and results in interval-censored data. When there no infection occurs before the end of study, it is regarded as right-censored. Denote T_1, T_2, T_3 and T_4 as the infection times of the rear and front udders at side 1 and those of the rear and front udders at side 2, respectively. Right censoring rates are 19, 20, 22 and 22%, respectively. Table 3 shows the result of three test statistics; unweighted S , S_{w1} with $W(s, t) = S_1(s)S_2(t)$ and S_{w2} with $W(s, t) = (1 - S_1(s))(1 - S_2(t))$ at (2.6). According to the results, the unweighted test statistic S and the weighted test statistic S_{w2} with weight emphasizing the latter association yield insignificant result while the weighted test statistic S_{w1} with weight emphasizing the early dependency yields a significant result. Figure 2 shows the rectangle plot of (T_1, T_2) from use of **MLEcens** package (2013) and indicates most rectangles posit at the left lower part, which means the

Table 3: Independence test of infection times between udder quarters

	S	se	p -value	S_{w1}	se_{w1}	p -value	S_{w2}	se_{w2}	p -value
(T_1, T_2)	-7.198	4.565	0.115	-5.073	1.260	<0.001	-3.374	1.946	0.083
(T_1, T_3)	-2.695	4.110	0.512	-3.494	1.071	<0.001	-2.562	1.955	0.098
(T_1, T_4)	-7.510	4.979	0.131	-5.591	1.000	<0.001	-2.867	1.709	0.935
(T_2, T_3)	-1.962	4.776	0.681	-4.290	1.011	<0.001	-3.159	2.061	0.126
(T_2, T_4)	-3.195	4.564	0.484	-6.038	1.479	<0.001	-2.187	1.584	0.167
(T_3, T_4)	1.164	4.135	0.778	-4.813	1.059	<0.001	-1.618	1.358	0.233

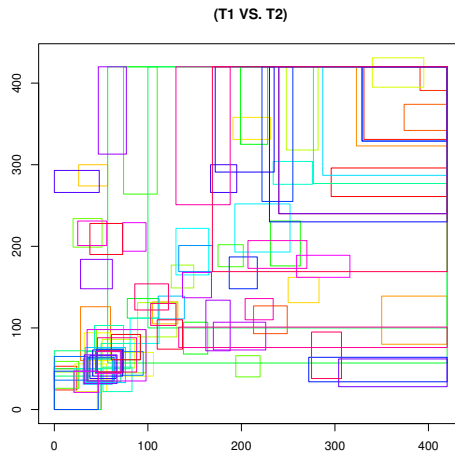


Figure 2: Rectangular plot of inflammation times at two udders.

S_{w1} focusing on early association has a more significant p -value.

4.2. AREDS data

For the second real data, an age-related eye disease study (AREDS) was analyzed (Sun and Ding, 2019). Our interest is to investigate the association of the time to late age-related molecular disease (AMD) between two eyes. The dataset includes 629 Caucasian patients who had at least one eye in moderate AMD stage at baseline. Each patient was examined every six months in the first six year and then every year after year 6. For the j ($= 1, 2$)th eye of the subject i , L_{ij} denotes the last assessment time when the eye was free from AMD and R_{ij} denotes the first assessment time when the eye was diagnosed with AMD. The censoring rate is 46.7% and 44.5% of the left and right eyes, respectively. For investigating the association between the two eyes, the unweighted test statistics yielded $S = 60.83$ with a corresponding standard error $se(S) = 7.08$ and p -value < 0.001 , which means the two eyes were correlated in term of the progression of the disease. When two weighted test statistics are applied, (i) $S_{w1} = 45.668$, $se_{w1} = 3.780$ and p -value < 0.001 , and the (ii) $se_{w2} = 2.857$, $se_{w2} = 0.706$ and p -value < 0.001 , respectively. Figure 3 shows the rectangles of the bivariate interval censored late AMD time for between two eyes. Compared with the mastitis data set in Figure 2, the rectangle is uniformly displayed.

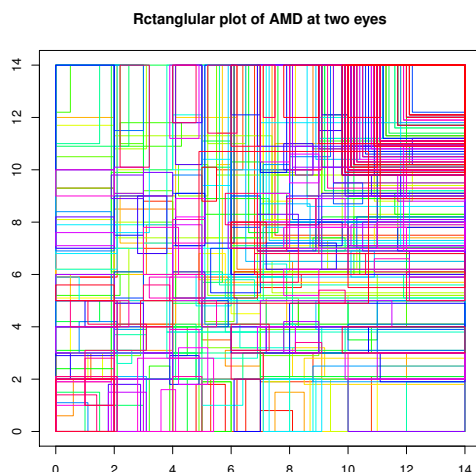


Figure 3: Rectangular plot of AMD occurrence times of two eyes.

5. Concluding remarks

We have suggested a generalized Mantel-Haenszel test to test the independence between two interval censored failure times. The proposed test procedures are based on the pseudo occurrence times derived from the equivalence set and the corresponding estimated occurrence numbers and the risk set sizes under H_0 . The variance of the suggested test statistic was constructed using a bootstrap technique. The weighted test statistics give a somewhat different result according to the censoring rate, sample size and association pattern. Simulation results show that the test with more weights at the early time region seems to have desirable powers even at high censoring rates. Considering the relation between censoring and weight, we analyzed the mastitis data, which shows a very different result between the two different weighted and unweighted version. Therefore, the application of a suitable weight is very essential procedure for testing the independence of bivariate failure times. In our data analysis, in Figure 2, most rectangles are located at the lower right part where S_{w1} yields significant p -values while the other test statistics are insignificant. Meanwhile, Figure 3 shows uniformly distributed rectangles and these rectangle plots would be a good indicator to select suitable weights. As for related studies, the weighted Mantel-Haenszel test that combines the subject-specific covariate can be considered. Alternatively, the copula based score test reflecting the subject-specific covariate is suggested and is extended to the bivariate case (Sun *et al.*, 2019; Sun and Ding, 2021).

References

- Betensky R and Finkelstein DF (1999). An extension of Kendall's coefficient of concordance to bivariate interval censored data, *Statistics in Medicine*, **18**, 3101–3109.
- Bogaerts K, Komárek A, and Lesaffre E (2018). *Survival Analysis with Interval-Censored Data: A Practical Approach with Examples in R, SAS, and BUGS*, CRC-Press, New York.
- Bogaerts K and Lesaffre E (2008). Estimating local and global measures of association for bivariate interval censored data with a smooth estimate of density, *Statistics in Medicine*, **27**, 5941–5955.
- Ding AA and Wang W (2004). Testing independence for bivariate current status data, *Journal of the American Statistical Association*, **99**, 145–155.

- Harrington DP and Fleming TR (1982). A class of rank test procedures for censored survival data, *Biometrika*, **69**, 553–566.
- Hsu L and Prentice RL (1996). A generalisation of the Mantel-Haenszel tests to bivariate failure time data, *Biometrika*, **83**, 905–911.
- Maathuis Marloes (2013). **MLEcens**, R package.
- Oakes D (1982). A concordance test for independence in the presence of censoring, *Biometrics*, **38**, 451–455.
- Shih JH and Louis TA (1996). Tests of independence for bivariate survival data, *Biometrics*, **52**, 1440–1449.
- Sun L, Wang L, and Sun J (2006). Estimation of the association for bivariate interval censored failure time data, *Scandinavian Journal of Statistics*, **33**, 637–649.
- Sun J (2006). *The Statistical Analysis of Interval-censored Failure Time Data*, Springer, New-York.
- Sun L, Wang L, and Sun J (2006). Estimation of the association for bivariate interval censored failure time data, *Scandinavian Journal of Statistics*, **33**, 637–649.
- Sun T and Ding Y (2019). **CopulaCenR**: Copula-Based Regression Models for Bivariate Censored Data, R package.
- Sun T and Ding Y (2021). Copula-based semiparametric regression method for bivariate data under general interval censoring, *Biostatistics*, **22**, 315–330.
- Sun T, Liu Y, Cook RJ, Chen W, and Ding Y (2019). Copula-based score test for bivariate time-to-event data, with application to a genetic study of AMD progression, *Life Time Data Analysis*, **25**, 546–568.
- Wang W and Ding AA (2000). On assessing the association for bivariate current status data, *Biometrika*, **87**, 879–893.

Received November 18, 2021; Revised March 20, 2022; Accepted April 24, 2022