

Conditional Tests for comparing k Competing Risks

Test condizionati per il confronto di k Rischi Competitivi

Livio Finos¹, Fortunato Pesarin², Luigi Salmaso³, Aldo Solari²

¹ Dipartimento di Scienze Biomediche e Terapie Avanzate, Università di Ferrara

² Dipartimento di Scienze Statistiche, Università di Padova

³ Dipartimento di Tecnica e Gestione dei Sistemi Industriali, Università di Padova

e-mail: solari@stat.unipd.it

Riassunto: Nell'ambito dei rischi competitivi, può risultare d'interesse verificare se i rischi considerati risultino ordinati per gravità. Nel caso di soli due rischi, viene suggerito l'utilizzo della distribuzione condizionata delle statistiche test presenti in letteratura. Nel caso di più di due rischi, viene proposta una procedura di verifica multipla basata sulla combinazione di test di permutazione. Per valutare le prestazioni dei test condizionati sono stati condotti degli studi di simulazione.

Keywords: combining dependent permutation tests, competing risks, conditional inference, multiple testing, ordered alternatives.

1. Introduction

In the competing risks model, a subject is exposed to several risks at the same time, but the eventual failure results from just one of these. Suppose that there are k competing risks and that they may be dependent on each other. Let T denote the continuous lifetime of a subject, with distribution function F and survival function S , and let δ denote the cause of failure, that is, $\{\delta = j\}$ is the event that the failure is due to risk j , $j = 1, \dots, k$. Furthermore, we allow for the possibility of right-censoring. Denote the censoring time by C and its survivor function by S_C . We assume that $S_C(t) > 0$ for all t and C is independent of T . Let $\{\delta = 0\}$ denote the event that the subject was censored; we observe n independent, identically distributed copies (X_i, δ_i) , $i = 1, \dots, n$ of (X, δ) , where $X = \min(T, C)$.

To quantify the risks of failure from the various causes, the notion of cause-specific hazard rate (CSHR) is used. The CSHR due to cause j is defined by

$$\lambda_j(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \Pr\{t \leq T < t + \Delta t, \delta = j | T \geq t\} \quad j = 1, \dots, k,$$

and the overall hazard rate satisfies the relation $\lambda(t) = \sum_{j=1}^k \lambda_j(t)$.

Recently, attention has been paid to the problem of investigating possible differences in mortality from different causes: the more fatal the cause, the more that outcome will drive clinical interventions. Hence it is important to know whether the various risks under consideration are equally serious or whether they are ordered in severity, i.e. we consider the problem of testing the null hypothesis

$$H_0 : \lambda_1(t) = \dots = \lambda_k(t) \quad \forall t \geq 0, \quad (1)$$

against the ordered alternative

$$H_2 : \lambda_1(t) \leq \dots \leq \lambda_k(t), \quad \forall t \geq 0, \quad (2)$$

with at least one strict inequality holding for some t .

In many applications involving competing risks, such comparisons have also been made in terms of the cumulative probability of occurrence by time t for a particular cause in the presence of other risks. The cumulative incidence (sub-distribution) function (CIF) due to cause j is defined by

$$F_j(t) = \Pr\{T \leq t, \delta = j\} = \int_0^t \lambda_j(u) S(u) du, \quad j = 1, \dots, k,$$

with $F(t) = \sum_{j=1}^k F_j(t)$. Note that (1) is equivalent to the equality of k CIFs and (2) implies

$$H_1 : F_1(t) \leq \dots \leq F_k(t), \quad \forall t \geq 0 \text{ with strict inequality for some } t \quad (3)$$

but not *vice versa*.

2. Conditional distribution

Under the null hypothesis, T and δ are independent, thus

$$\Pr\{T \leq t\} \Pr\{\delta = 1\} = \dots = \Pr\{T \leq t\} \Pr\{\delta = k\} \quad \forall t \geq 0.$$

This means that, under H_0 , each uncensored time is randomly attributed to one of the k possible causes of failure. It follows that, under H_0 , there are k^ν equally likely “permutations” of the observed data set, such that each permutation is one of all possible ways of attributing causes of failure to the ν uncensored subjects when times are held fixed. Roughly speaking, we allow δ_i to vary from 1 to k for all i such that $\delta_i \neq 0$.

Suppose that A is an appropriate test statistic for which, without loss of generality, we assume that large values are significant. The conditional distribution of A can be constructed by calculating A for each of the possible permutations, and the p -value of the permutation test can be calculated as the probability of getting a test statistic as extreme as, or more extreme than, the observed test statistic.

3. Comparing two risks

For testing against H_1 and H_2 , the test statistics considered in Ali *et al.* (1994) are

$$D_{3n} = \sup_{t \geq 0} \{ \phi_n(t) \}, \quad D_{4n} = \sup_{0 \leq s < t < \infty} \{ \phi_n(t) - \phi_n(s) \}$$

respectively, where $\phi_n(t) = \int_0^t \hat{S}(u-) \hat{S}_C(u-)^{1/2} d(\hat{\Lambda}_2 - \hat{\Lambda}_1)(u)$, $\hat{S}$ and $\hat{S}_C$ are the Kaplan-Meier estimators of S and S_C , respectively, whereas $\hat{\Lambda}_j$ is the Nelson-Aalen estimator of the cumulative CSHR $\Lambda_j(t) = \int_0^t \lambda_j(u) du$ for risk j . From Theorem 3.1 in Ali *et al.* (1994), under H_0 ,

$$n^{1/2} D_{3n} \xrightarrow{d} \sup_{0 \leq y \leq 1} W(y), \quad n^{1/2} D_{4n} \xrightarrow{d} \sup_{0 \leq y \leq 1} |W(y)|,$$

where $\{W(y), y \geq 0\}$ is a standard Brownian motion. An unpleasant property of these tests is that when n is not extremely large, the use of the asymptotic critical levels gives conservative tests, and this effect increases as the censoring becomes more severe; in such situations, exact conditional distributions are an interesting alternative.

In order to compare the performance of the conditional and asymptotic tests based on the D_{3n} and D_{4n} statistics, we report some results of a simulation study that consider two kinds of ordered alternatives:

$$H_a : \lambda_2(t) = (\beta + 1)\lambda_1(t), \quad H_b : \lambda_2(t) = [1 + \beta\Lambda_1(t)]\lambda_1(t).$$

For H_a , we used Block and Basu's absolutely continuous bivariate exponential distribution. If (T_1, T_2) has this distribution, then the joint density is

$$f(t_1, t_2) = \begin{cases} \frac{\gamma_1\gamma(\gamma+\gamma_0)}{\gamma_1+\gamma_2} \exp\{-\gamma_1 t_1 - (\gamma_2 + \gamma_0)t_2\}, & t_1 < t_2 \\ \frac{\gamma_2\gamma(\gamma+\gamma_0)}{\gamma_1+\gamma_2} \exp\{-\gamma_2 t_2 - (\gamma_1 + \gamma_0)t_1\}, & t_2 > t_1 \end{cases}$$

where $\gamma_0 \geq 0$, $\gamma_1 > 0$, $\gamma_2 > 0$ and $\gamma = \gamma_0 + \gamma_1 + \gamma_2$ and the failure times are $T = \min(T_1, T_2)$. We consider various values of $\beta = \gamma_2 - 1$ with $\gamma_0 = \gamma_1 = 1$.

For H_b , we consider the case when T_1 and T_2 are independent with T_1 having standard exponential distribution and T_2 having linearly increasing failure rate distribution with hazard rate $\lambda_2(t) = (1 + \beta t)$.

For both alternatives, we carried out the simulations with $n = 100$ for various values of β . The censoring was taken to be exponential with parameter values 1 and 3, corresponding to "light" and "heavy" censoring. Simulations were repeated 1000 times, with 1000 random permutations. The results are shown in Table 1, where the superscript '*' indicates the conditional test.

Table 1: Observed levels and powers of considered tests at a significance level of 5%.

β	Proportional Hazard						Linearly Increasing Failure Rate					
	light censored			heavy censored			light censored			heavy censored		
	0.0	0.5	1.0	0.0	0.5	1.0	0.0	2.0	3.0	0.0	2.0	3.0
D_{3n}	3.6	48.5	90.3	3.0	35.6	75.7	3.0	38.5	58.2	1.9	13.2	22.7
D_{3n}^*	4.5	52.0	91.1	3.7	41.3	80.0	4.7	48.0	66.6	4.5	24.2	36.6
D_{4n}	3.4	43.8	86.4	2.4	29.5	69.7	2.7	38.4	59.5	1.0	10.4	19.5
D_{4n}^*	4.4	49.4	88.7	3.9	37.9	76.3	5.1	55.2	72.1	4.6	28.1	38.7

4. Comparing $k \geq 2$ risks

Recently, El Barmi and Mukerjee (2006) extended the results to the case of more than two CIFs, by using an adapted version of the sequential testing procedure developed in Hogg (1965) for testing equality of distribution functions based on independent random samples. We compare this sequential testing procedure to a multiple testing procedure by using the nonparametric combination of dependent permutation tests (Pesarin, 2001). Note that it is straightforward to consider this procedure for the case of $k \geq 2$ CSHRs.

It might be possible to express equivalently the global hypotheses (1) and (3) as

$$\begin{aligned} H_0 &: \bigcap_{j=2}^k \{H_{0j}\} = \bigcap_{j=2}^k \{F_1 = \dots = F_{j-1} = F_j = \dots = F_k\}, \\ H_1 &: \bigcup_{j=2}^k \{H_{1j}\} = \bigcup_{j=2}^k \{F_1 = \dots = F_{j-1} \leq F_j = \dots = F_k\}, \end{aligned}$$

where for testing H_{0j} we pooling together the first $(j - 1)$ causes of failure and the last $(k - j + 1)$ causes of failure, $j = 2, \dots, k$. The multiple hypotheses testing problem is solved by considering the permutation test

$$D_j^* = \sup_{x \geq 0} \left\{ \int_0^t \hat{S}(u-) \hat{S}_C(u-)^{1/2} d \left(\bar{\Lambda}_{j:k} - \bar{\Lambda}_{1:j-1} \right) (u) \right\} \quad j = 2, \dots, k$$

for testing H_{0j} against $H_{1j} - H_{0j}$, where $\bar{\Lambda}_{r:s} = \frac{\sum_{j=r}^s \hat{\Lambda}_j}{s-r+1}$. Hence, the $(k - 1)$ dependent permutation tests are combined into a global test by means of the nonparametric combination method (Pesarin, 2001). For instance, Tippett's combined permutation test is $D_T^* = \max_{2 \leq j \leq k} \left\{ 1 - p_{D_j^*} \right\}$, where $p_{D_j^*}$ is the p -value of the permutation test D_j^* .

To assess the performance of the two procedures, a simulation study is carried out. We consider three independent risks with no censoring; the data are generated using the Lehmann-type alternatives: $F_1 \leq F_1^\theta \leq F_3^\beta$ with $0 < \beta \leq \theta \leq 1$. This gives H_0 when $\beta = \theta = 1$ and H_1 when $\beta \leq \theta \leq 1$ with at least one strict inequality, where as θ and β move closer to 0, the departure from H_0 is larger. We choose several pairs of (θ, β) and we assume that F_1 follows unit exponential. The results are showed in Table 2, where D_n'' is the global test proposed in El Barmi and Mukerjee (2006).

Table 2: Empirical Type I Error Rates and Powers of Tests for Lehmann alternative at a significance level of 5%.

θ	1.0	1.0	1.0	1.0	0.4	0.3	0.2	0.7	0.6	0.5
β	1.0	0.5	0.4	0.3	0.4	0.3	0.2	0.5	0.4	0.3
D_n''	4.5	19.5	48.9	87.1	26.4	45.9	85.6	17.2	37.0	79.4
D_T^*	4.5	21.5	49.8	86.0	29.0	55.4	92.5	18.4	38.0	79.3

References

- Aly, E. A. A., Kochar, S. C., McKeague, I. W. (1994) Some tests for comparing cumulative incidence functions and cause-specific hazard rates. *Journal of the American Statistical Association*, 89, 994-999.
- El Barmi, H., Mukerjee H. (2006) Restricted estimation of the cumulative incidence functions corresponding to k competing risks. *Submitted*.
- Hogg, R. V. (1965) On models and hypotheses with restricted alternatives. *Journal of the American Statistical Association*, 60, 1153-1162.
- Pesarin F. (2001) *Multivariate Permutation Tests with Applications in Biostatistics*. Wiley, Chichester.