Conditional Tests for comparing $k$ Competing Risks

Test condizionati per il confronto di $k$ Rischi Competitivi

Livio Finos$^1$, Fortunato Pesarin$^2$, Luigi Salmaso$^3$, Aldo Solari$^2$

$^1$ 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


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 $\delta$ denote the cause of failure, that is, $\{\delta = j\}$ is the event that the failure is due to risk $j$, $j = 1, \ldots, k$. 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 $(X_i, \delta_i), i = 1, \ldots, n$ of $(X, \delta)$, 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

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

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

$$H_0 : \lambda_1(t) = \ldots = \lambda_k(t) \quad \forall t \geq 0,$$  \hspace{1cm} (1)
against the ordered alternative

\[ H_2 : \lambda_1(t) \leq \ldots \leq \lambda_k(t), \quad \forall \; t \geq 0, \]  

(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

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

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

\[ H_1 : F_1(t) \leq \ldots \leq F_k(t), \forall \; t \geq 0 \]  

with strict inequality for some \( t \) \( \quad (3) \)

but not vice versa.

2. Conditional distribution

Under the null hypothesis, \( T \) and \( \delta \) are independent, thus

\[ \Pr\{ T \leq t \} \Pr\{ \delta = 1 \} = \ldots = \Pr\{ T \leq t \} \Pr\{ \delta = k \} \quad \forall \; t \geq 0. \]

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 \( k^\nu \) 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 \( \nu \) uncensored subjects when times are held fixed. Roughly speaking, we allow \( \delta_i \) 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-Aelen 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 \( H_0 \),

\[ n^{1/2}D_{3n} \overset{d}{\to} \sup_{0 \leq y \leq 1} W(y), \quad n^{1/2}D_{4n} \overset{d}{\to} \sup_{0 \leq y \leq 1} |W(y)|, \]

\[ \text{– 182 –} \]
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

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

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 \( \beta \). 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>
<thead>
<tr>
<th>( \beta )</th>
<th>Proportional Hazard (light censored)</th>
<th>Proportional Hazard (heavy censored)</th>
<th>Linearly Increasing Failure Rate (light censored)</th>
<th>Linearly Increasing Failure Rate (heavy censored)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>3.6 48.5 90.3</td>
<td>3.0 35.6 75.7</td>
<td>3.0 38.5 58.2</td>
<td>1.9 13.2 22.7</td>
</tr>
<tr>
<td>0.5</td>
<td>4.5 52.0 91.1</td>
<td>3.7 41.3 80.0</td>
<td>4.7 48.0 66.6</td>
<td>4.5 24.2 36.6</td>
</tr>
<tr>
<td>1.0</td>
<td>3.4 43.8 86.4</td>
<td>2.4 29.5 69.7</td>
<td>2.7 38.4 59.5</td>
<td>1.0 10.4 19.5</td>
</tr>
<tr>
<td>( D_{3n} )</td>
<td>4.4 49.4 88.7</td>
<td>3.9 37.9 76.3</td>
<td>5.1 55.2 72.1</td>
<td>4.6 28.1 38.7</td>
</tr>
<tr>
<td>( D_{4n} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

$$H_0 : \bigcap_{j=2}^{k} \{H_{0j}\} = \bigcap_{j=2}^{k} \{F_1 = \ldots = F_{j-1} = F_j = \ldots = F_k\} ,$$

$$H_1 : \bigcup_{j=2}^{k} \{H_{1j}\} = \bigcup_{j=2}^{k} \{F_1 = \ldots = F_{j-1} \leq F_j = \ldots = F_k\} ,$$

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

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

for testing $H_{0j}$ against $H_{1j} - H_{0j}$, where $\tilde{\Lambda}_{r,s} = \sum_{j=r+1}^{s} \tilde{\Lambda}_j$. 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 $\theta$ and $\beta$ move closer to 0, the departure from $H_0$ is larger. We choose several pairs of $(\theta, \beta)$ 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%.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>1.0</th>
<th>1.0</th>
<th>1.0</th>
<th>1.0</th>
<th>0.4</th>
<th>0.3</th>
<th>0.2</th>
<th>0.7</th>
<th>0.6</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>1.0</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>$D_n''$</td>
<td>4.5</td>
<td>19.5</td>
<td>48.9</td>
<td>87.1</td>
<td>26.4</td>
<td>45.9</td>
<td>85.6</td>
<td>17.2</td>
<td>37.0</td>
<td>79.4</td>
</tr>
<tr>
<td>$D_T''$</td>
<td>4.5</td>
<td>21.5</td>
<td>49.8</td>
<td>86.0</td>
<td>29.0</td>
<td>55.4</td>
<td>92.5</td>
<td>18.4</td>
<td>38.0</td>
<td>79.3</td>
</tr>
</tbody>
</table>

### References


