Part III. Hypothesis Testing

III.1. Log-rank Test for Right-censored Failure Time Data

Consider a survival study consisting of $n$ independent subjects from $p$ different populations with survival functions $S_1(t), \ldots, S_p(t)$. Suppose that the goal is to test the hypothesis

$$H_0 : S_1(t) = \ldots = S_p(t).$$

based on right-censored failure time data

$$\{ X_i = \min(T_i, C_i), \delta_i = I(X_i = T_i) ; i = 1, \ldots, n \}.$$ 

Let

$$t_1 < t_2 < \ldots < t_k$$ — observed failure times,

$$d_{ij} = \# \text{ of failures at } t_j \text{ from the } i\text{th population}$$

$$r_{ij} = \# \text{ of subjects at risk at } t_j^- \text{ from the } i\text{th population}$$

$$d_j = \# \text{ of failures at } t_j (= d_{1j} + \ldots + d_{pj}),$$

$$r_j = \# \text{ of subjects at risk at } t_j^- (= r_{1j} + \ldots + r_{pj}),$$

$$j = 1, \ldots, k, \ i = 1, \ldots, p.$$
To construct a test statistic, consider what happened at time $t_j$. Conditional on the failure and censoring experience up to time $t_j$, under $H_0$, the conditional distribution of $d_{1j}, ..., d_{pj}$ given $d_j$ is the hypergeometric distribution

$$P(d_{1j}, ..., d_{pj} \mid d_j, r_{1j}, ..., r_{pj}) = \cdot$$

Thus we have

$$w_{ij} = E[d_{ij} \mid d_j] = r_{ij} d_j r_j^{-1},$$

$$V_{ii}^j = Var[d_{ij} \mid d_j] = r_{ij} (r_j - r_{ij}) d_j (r_j - d_j) r_j^{-2} (r_j - 1)^{-1},$$

$$V_{i_1i_2}^j = cov[d_{i_1j}, d_{i_2j} \mid d_j] = -r_{i_1j} r_{i_2j} d_j (r_j - d_j) r_j^{-2} (r_j - 1)^{-1}.$$

Define the statistic

$$\nu_j = (d_{1j} - w_{1j}, ..., d_{pj} - w_{pj})'$$

at $t_j$, which has (conditional) mean zero and covariance matrix $V_j = (V_{i_1i_2}^j)$. The log-rank statistic is defined as the simple summation over failure times

$$\nu = \sum_{j=1}^{k} \nu_j = (D_1 - E_1, ..., D_p - E_p)'$$,
the vector of the observed numbers of failures in each population minus the corresponding vector of the “expected” numbers of failures, where
\[ D_i = \sum_{j=1}^{k} d_{ij}, \quad E_i = \sum_{j=1}^{k} w_{ij}. \]

Or the statistic \( \nu \) can be written as
\[ \nu = D - E, \]
where
\[ D = (D_1, \ldots, D_p), \quad E = (E_1, \ldots, E_p). \]

If the \( \nu_j \)'s are independent, then
\[ E[\nu] = 0, \quad Var[\nu] = V_1 + \ldots + V_k. \]

The hypothesis \( H_0 \) can be tested using the statistic
\[ \chi^2 = \nu' V^{-1} \nu \]
based on a \( \chi^2_{p-1} \) distribution for large samples.

If \( p = 2 \), the test of the hypothesis \( H_0 \) can be based on the statistic
\[ Z = \frac{\Sigma_{j=1}^{k} (d_{1j} - r_{1j}d_j/r_j)}{[\Sigma_{j=1}^{k} r_{1j} (r_j - r_{1j}) d_j (r_j - d_j) r_j^{-2} (r_j - 1)^{-1}]^{1/2}} \]
with the standard normal distribution for large samples.
Comments:
1. The log-rank test can be seen as censored data generalizations of linear rank statistics such as the Wilcoxon test and Savage exponential score test. It is also referred to as the generalized Savage test.
2. The log-rank test can also be derived as a score test from the marginal or partial likelihood under the proportional hazards model, which means that the hazard or survival functions are proportional to each other. Under this case, it can be shown that the log-rank test is the optimal test or the most efficient test.
3. The log-rank test is derived based on large-sample theory under the assumption that the censoring distribution is independent of the failure distributions.
4. The log-rank test statistic can be rewritten as
   \[ \nu = \sum_{j=1}^{k} \nu_j = (D_1 - E_1, \ldots, D_p - E_p)' \]
   with
   \[ D_i - E_i = \sum_{j=1}^{k} r_{ij} \left( \frac{d_{ij}}{r_{ij}} - \frac{d_j}{r_j} \right) = \sum_{j=1}^{k} r_{ij} (\hat{\lambda}_{ij} - \hat{\lambda}_j) \]
   \[ = \int_0^{\infty} w_i(t) [d\tilde{\Lambda}_i(t) - d\tilde{\Lambda}(t)] , \]
the summation of weighted differences between the estimates of hazard functions for individual populations and the common population under \( H_0 \).
III.2. Other Tests for Right-censored Failure Time Data

As in the previous section, again consider the problem of comparing \( p = 2 \) survival functions based on right-censored data from \( n \) independent subjects, that is, testing \( H_0 : S_1(t) = S_2(t) \).

III.2.1. Weighted log-rank tests:

Note that we can rewrite the log-rank statistic \( \nu_1 \) as

\[
\nu_1 = D_1 - E_1 = \sum_{j=1}^{k} r_{1j} \left( \frac{d_{1j}}{r_{1j}} - \frac{d_j}{r_j} \right)
\]

\[
= \sum_{j=1}^{k} \frac{r_{1j} r_{2j}}{r_{1j} + r_{2j}} \left( \frac{d_{1j}}{r_{1j}} - \frac{d_{2j}}{r_{2j}} \right)
\]

\[
= \int_0^\infty \frac{\bar{Y}_1(t) \bar{Y}_2(t)}{\bar{Y}_1(t) + \bar{Y}_2(t)} \left\{ d\tilde{\Lambda}_1(t) - d\tilde{\Lambda}_2(t) \right\},
\]

where

\( \bar{Y}_1(t) = \# \text{ of subjects from the population 1 at risk at } t^- \),

\( \bar{Y}_2(t) = \# \text{ of subjects from the population 2 at risk at } t^- \).
This motivated the weighted log-rank test statistics

\[ W_{LR} = \int_0^\infty K(s) \left\{ d\tilde{\Lambda}_1(t) - d\tilde{\Lambda}_2(t) \right\} \]

\[ = \int_0^\infty W(t) \frac{\bar{Y}_1(t) \bar{Y}_2(t)}{\bar{Y}_1(t) + \bar{Y}_2(t)} \left\{ d\tilde{\Lambda}_1(t) - d\tilde{\Lambda}_2(t) \right\} \]

\[ = \sum_{j=1}^k W(t_j) \frac{r_{1j}r_{2j}}{r_{1j} + r_{2j}} \left( \frac{d_{1j}}{r_{1j}} - \frac{d_{2j}}{r_{2j}} \right), \]

where \( K(s) \) or \( W(s) \) is a weight process. It can be shown that under \( H_0 \) and some regularity conditions, \( W_{LR} \) has an asymptotic normal distribution with mean zero and variance that can be estimated by

\[ \hat{\sigma}^2 = \sum_{j=1}^k K^2(t_j) \frac{1}{r_{1j}r_{2j}} \frac{r_j - d_j}{r_j - 1} d_j = \sum_{j=1}^k W^2(t_j) \frac{r_{1j}r_{2j}}{r_j^2} \frac{r_j - d_j}{r_j - 1} d_j \]

as \( n \to \infty \).

Let \( \hat{S} \) denote the the Kaplan-Meier estimator of the survival function under \( H_0 \) based on pooled samples. A common class of weight processes is given by

\[ W(t) = \{ \hat{S}(t-) \}^\rho \{ 1 - \hat{S}(t-) \}^\gamma \]

(Harrington and Fleming, 1982), where \( \rho \) and \( \gamma \) are non-negative constants. In this case, the test statistics \( W_{LR} \) are referred to as \( G^{\rho,\gamma} \) statistics.
III.2.2. Weighted Kaplan-Meier statistics:

To test $H_0$, we could also employ the weighted Kaplan-Meier statistics

$$WKM = \sqrt{\frac{n_1 n_2}{n}} \int_0^\tau W(t) [\hat{S}_1(t) - \hat{S}_2(t)] dt,$$

where $\tau$ is the largest observation time, $W(t)$ is a weight process and $\hat{S}_1$ and $\hat{S}_2$ are the Kaplan-Meier estimators of the survival functions $S_1$ and $S_2$ based on separate samples, respectively.

Suppose that the weight process $W(t)$ is small when $t$ is close to $\tau$. Then it can be shown that as $n \to \infty$, the distribution of the statistics $WKM$ can be approximated by a normal distribution with mean zero and variance

$$\hat{\sigma}^2 = - \int_0^\tau \left[ \int_t^\tau W(u) \hat{S}(u) du \right]^2 \hat{S}^2(t) \hat{C}(t) \hat{S}(t) dt,$$

where $\hat{S}$ and $\hat{C}$ are the Kaplan-Meier estimators of the common survival function under $H_0$ and the survival function of the censoring variable based on the pooled samples, respectively.

Comments

1. The test statistics $WLR$, the integrated weighted differences of the estimated hazard functions, are most sensitive to the alternative of ordered hazard functions

\[ H^1_a : \lambda_2(t) \geq \lambda_1(t) \text{ for all } t. \]

In contrast, the test statistics $WKM$, the integrated weighted difference between Kaplan-Meier estimates of the survival functions, are most sensitive to the alternative of ordered survival functions

\[ H^2_a : S_2(t) \geq S_1(t) \text{ for all } t. \]

$H^2_a$ does not imply $H^1_a$.

2. The test statistics $WLR$ are constructed based on ranks and thus invariant under all monotone transformations of time. That is, they do not depend on the scale in which time is measured. This is not true for $WKM$. 
III.3. Log-rank Test for Interval-censored Data

As in the previous sections, consider a survival study which involves $n$ independent subjects from $p$ populations and in which the goal is to test the hypothesis $H_0$. Instead of observing right-censored data, suppose that only interval-censored data are available. Also suppose that the survival time takes discrete values

$$0 = t_0 < t_1 < \ldots < t_k < t_{k+1} = \infty.$$ 

For subject $i$, let $A_i = \{L_i, L_i + 1, \ldots, U_i\} \epsilon \{t_1, \ldots, t_{k+1}\}$ denote the interval within which the $i$th individual fails. Then observed data have the form

$$\{A_i \ ; \ i = 1, \ldots, n\}.$$ 

Also let $0 = s_0 < \cdots < s_{m+1} = k + 1$ denote the smallest subset of $\{t_0, t_1, \ldots, t_{k+1}\}$ such that each $L_i$ and $U_i$ is contained in the subset and $\Delta_j = \{s_{j-1} + 1, \ldots, s_j\}, j = 1, \ldots, m$. Define $\alpha_{ij}$ as the indicator of the event $\Delta_j \subseteq A_i$. Note that if (i) the intervals not including $k + 1$ are not overlapping and (ii) for each interval with $U_i = k + 1$, its left endpoint coincides with a left endpoint of an interval that does not include $k + 1$, then the observed data can be treated as right-censored data by treating each interval as a single point.
To test $H_0$, we will follow the idea behind the log-rank test for right-censored data and determine the death and risk numbers. Let $S = (S_0, \ldots, S_m)$ denote the common survival function of the $p$ populations under $H_0$ ($S_j = Pr\{T > s_j\}$) and $\hat{S} = (\hat{S}_0, \ldots, \hat{S}_m)$ the maximum likelihood estimator of $S$.

Define
\[
d'_j = \sum_{i=1}^{n} \{\alpha_{ij} [\hat{S}_{j-1} - \hat{S}_j] / \sum_{u=1}^{m+1} \alpha_{iu} [\hat{S}_{u-1} - \hat{S}_u]\}
\]
and
\[
r'_j = \sum_{r=j}^{m+1} \sum_{i=1}^{n} \{\alpha_{ir} [\hat{S}_{r-1} - \hat{S}_r] / \sum_{u=1}^{m+1} \alpha_{iu} [\hat{S}_{u-1} - \hat{S}_u]\}.
\]

Also define
\[
d'_{jl} = \sum_{i} \{\alpha_{ij} [\hat{S}_{j-1} - \hat{S}_j] / \sum_{u=1}^{m+1} \alpha_{iu} [\hat{S}_{u-1} - \hat{S}_u]\}
\]
and
\[
r'_{jl} = \sum_{r=j}^{m+1} \sum_{i} \{\alpha_{ir} [\hat{S}_{r-1} - \hat{S}_r] / \sum_{u=1}^{m+1} \alpha_{iu} [\hat{S}_{u-1} - \hat{S}_u]\},
\]
where $\sum'_i$ denotes the summation over subjects $i$ in the population $l$. The $d'_j$’s, $r'_j$’s, $d'_{jl}$’s and $r'_{jl}$’s possess the similar meanings to the $d_j$’s, $j$’s, $d_{jl}$’s and $r_{jl}$’s respectively, the numbers of failures and the numbers of risks.
Motivated by the log-rank test statistic for right-censored data, we can construct a test statistic \( T = (T_1, ..., T_p)^t \) for testing \( H_0 \), where

\[
T_l = \sum_{j=1}^{m} \left( d'_{jl} - \frac{r'_{jl}d''_j}{n'_j} \right).
\]

If an estimate, \( V \), of the variance of \( T \) is available, then the test of \( H_0 \) can be based on the approximation

\[
T^t V^{-1} T \sim \chi^2_{p-1}.
\]

To obtain an estimate for the covariance matrix of \( T \) or \( V \), see Sun (1996), *Statistics in Medicine*, Vol. 15, 1387-1395.

III.4. Weighted Survival Test for Interval-censored Data

In this section, we will consider two sample comparison problem \((p = 2)\) and use the notation given in the previous section. To test \(H_0\), similar to the weighted Kaplan-Meier test statistics for right-censored data, we can construct a class of test statistics as

\[
W = \sum_{j=1}^{k} w(t_j) \left[ \hat{S}_1(t_j) - \hat{S}_2(t_j) \right] \Delta_j,
\]

where \(w\) is a weight function, \(\hat{S}_1\) and \(\hat{S}_2\) are the maximum likelihood estimates of the two survival functions \(S_1\) and \(S_2\) based on separate samples, and \(\Delta_j = t_j - t_{j-1}\).

The statistic \(W\) can be rewritten as

\[
W = \sum_{j=1}^{k} w(t_j) \left[ \sum_{l=1}^{j} \hat{p}_l^{(2)} - \sum_{l=1}^{j} \hat{p}_l^{(1)} \right] \Delta_j,
\]

where \(p_l^{(i)} = S_i(t_{l-1}) - S_i(t_l), i = 1, 2\). That is, \(W\) is a function of estimates of parameters \(\{ \hat{p}_l^{(1)}, \hat{p}_l^{(2)} ; l = 1, ..., m \}\), whose covariance can be estimated using the Fisher information matrix. Also under \(H_0\), the distribution of \(W\) can be approximated by the normal distribution with mean zero.