Logrank Test

Chao Cheng

December 10, 2022

Contents

1 Introduction

The log-rank test is one of the most commonly used test for comparing two or more survival distributions. To simplify the discussion, let's assume there are two groups of subjects, coded by 0 and 1. In group j, there are n_j i.i.d. underlying survival times with common c.d.f. denoted by $F_j(\cdot)$. And the corresponding hazard, cummulative hazard and survival functions are denoted by $h_j(\cdot)$, $H_j(\cdot)$ and $S_j(\cdot)$, respectively.

As usual, we assume the non-informative right censoring. So in each group, T_i and C_i are independent.

Here we want to test the null hypothesis $F_1(\cdot) = F_2(\cdot)$. If we know the parametric form of $F_1(\cdot)$ and $F_2(\cdot)$, e.g. the exponential distribution family, then this test can be reduced to test against a point/region in a Eucilidean parameter space. However, here we want a non-parametric test; that is, a test whose validity dose not depend on any parametric assumptions.

Clearly, a UMP test can not exist for this type of hypothesis. And there are two options in this case:

- **Directional test:** These are oriented towards a specific type of difference, e.g. $S_1(t) = S_0(t)^\theta$ for some θ.
- **Omnibus test:** These test are designed to have some power against all types of difference, e.g. a test based on $\int |S_1(t) - S_0(t)| \, \mathrm{d}t$ over some time interval.

The Pros-and-Cons of these two options of tests are summarised in Table 1. And a chioce between these two types of tests in real application involves several factors. Here we just point out that log-rank test is a directional test, and the specific type is the constant hazard ratio over time.

	Pros	Cons
Directional test	Strong power against the specified type of difference	(often) poor power against other types of difference
Omnibus test	have some power against most types of difference	lower power compared to a directional test for certain types of difference

Table 1: Pros and cons for different types of tests

2 Log-rank test

Log-rank test can be viewed as modification for the contingency table test to allow censoring in the data. Now let's consider these 2 groups, and denote the distinct times of observed failures as $0 < \tau_1 < \cdots < \tau_k$. We also define

$$\begin{aligned} Y_i(\tau_j) &= \text{number at risk (including events) for group } i \text{ at } \tau_j \\ Y(\tau_j) &= Y_0(\tau_j) + Y_1(\tau_j) \\ d_{ij} &= \text{number of events for group } i \text{ at } \tau_j \\ d_j &= d_{0j} + d_{1j} \end{aligned}$$

Then the information at time τ_j can be summarized in the following 2×2 table(Table 2):

Group	event	no event	number at risk
Group 0	d_{0j}	$Y_0(\tau_j) - d_{0j}$	$Y_0(\tau_j)$
Group 1	d_{1j}	$Y_1(\tau_j) - d_{1j}$	$Y_1(\tau_j)$
Overall	d_j	$Y(\tau_j) - d_j$	$Y(\tau_j)$

Table 2: Information at τ_j

Note that $d_{0j}/Y_0(\tau_j)$, $d_{1j}/Y_1(\tau_j)$ and $d_j/Y(\tau_j)$ are the estimates of $h_0(\tau_j)$, $h_1(\tau_j)$ and $h(\tau_j)$. To test the difference between $F_0(\cdot)$ and $F_1(\cdot)$ at this time point τ_j, one can consider the χ^2-test (details of χ^2-test can be found in other notes). But here we use the Fisher exact test, which is conditional on the marginal counts $Y_0(\tau_j)$, $Y_1(\tau_j)$, d_j and $Y(\tau_j) - d_j$. (This is more suitable in survival scenario because we know that the estimates are always conditional on the previous results. And this is just my personal opinion.)

Now, given those four marginal counts and $H_0 : F_0(\cdot) = F_1(\cdot)$, one can see that d_{1j} determines the whole table and actually d_{1j} follows a hypergeometric distribution

$$P(D_{1j} = d) = \frac{C_{Y_0(\tau_j)}^{d_{0j}} C_{Y_1(\tau_j)}^{d_j - d_{0j}}}{C_{Y(\tau_j)}^{d_j}},$$

where d ranges such that

$$\begin{aligned} d &\geq 0 \\ d_j - d &\geq 0 \\ Y_1(\tau_j) - d &\geq 0 \\ Y_0(\tau_j) - (d_j - d) &\geq 0 \end{aligned}$$

Therefore

$$\max\left(0, d_j - Y_0\left(\tau_j\right)\right) \leq d \leq \min\left(d_j, Y_1\left(\tau_j\right)\right).$$

And it's easy to know that

$$\begin{aligned}
\mathrm{E}_j = \mathrm{E}\left(D_{1j}\right) &= \frac{Y_1\left(\tau_j\right) d_j}{Y\left(\tau_j\right)} \\
\mathrm{V}_j = \mathrm{Var}\left(D_{1j}\right) &= \frac{Y\left(\tau_j\right) - Y_1\left(\tau_j\right)}{Y\left(\tau_j\right) - 1} \cdot Y_1\left(\tau_j\right)\left(\frac{d_j}{Y\left(\tau_j\right)}\right)\left(1 - \frac{d_j}{Y\left(\tau_j\right)}\right) \\
&= \frac{Y_0\left(\tau_j\right) Y_1\left(\tau_j\right) d_j \left(Y\left(\tau_j\right) - d_j\right)}{Y\left(\tau_j\right)^2 \left(Y\left(\tau_j\right) - 1\right)}
\end{aligned}$$

And denote the observation $O_j = d_{1j}$. And we can define for over the whole time points

$$\mathrm{O} = \sum_{j=1}^{k} O_j$$

$$\mathrm{E} = \sum_{j=1}^{k} \mathrm{E}_j$$

$$\mathrm{V} = \sum_{j=1}^{k} \mathrm{V}_j$$

And the test statistic is argued to follow under H_0:

$$Z = \frac{O - \mathrm{E}}{\sqrt{\mathrm{V}}} \overset{apx}{\sim} N(0,1).$$

Some comments about this test are

•

3 Extensions of Log-rank test

3.1 Stratified logrank test

For example in this case, we want to take into account (adjust for) some covariates, such as gender, in addition to the treatment-control (group1-group0) group. Then actually we will have 4 groups of subjects. In general, if we have overall L stratified levels (L is often the product of levels from each stratified covariates), then we want to test

$$H_0: \; S_0^{(l)}(\cdot) = S_1^{(l)}(\cdot), \quad l = 1, \cdots, L.$$

This stratified test is useful when the distributions of the stratum variable in the treatment-control groups are different, but the distribution of these relavent covariates within each stratum is the same between the treatment-control groups.

3.2 Weighted logrank test

3.3 Logrank test for multiple groups

3.4 Logrank trend test

References