Logrank Test

Chao Cheng

December 6, 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 $\int |S_1(t) - S_0(t)| \, dt$ 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

$$\begin{aligned} Y_i\left(\tau_j\right) &= \\ Y\left(\tau_j\right) &= Y_0\left(\tau_j\right) + Y_1\left(\tau_j\right) \\ d_{ij} &= \\ d_j &= d_{0j} + d_{1j} \end{aligned}$$

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

References