Logrank Test

Chao Cheng

July 15, 2024

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.

	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

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.

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

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

And it's easy to know that

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

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

- E_j can be viewed as a conditional expectation for each j, but strictly speaking, it is not clear that E has such an interpretation.
- The construction of Z seems to imply that the distributions from each τ_j are independent $N(0,1)$ under H_0. Is this true/accurate?
- Note that $Y_0(\tau_1), Y_0(\tau_2), \cdots$ and $Y_1(\tau_1), Y_1(\tau_2), \cdots$ are nonincreasing, and as soon as one reaches 0, it **must** follow that $O_j = E_j$ and $V_j = 0$ **at and beyond** that time. Thus there is no contribution for the data after that time point when computing Z.
- The logrank test is a **directional** test oriented towards alternatives where $S_1(t) = (S_0(t))^{\theta}$, or equivalently, when $h_1(t)/h_0(t) = \theta$, the constant hazard ratio alternative.
- The logrank statistic arise as a score test from a partial likelihood function from Cox's proportional hazards model.

- This construction of logrank test in this notes is an intuitive way. Some more serious technical details, including how does the test behave as a function of the amount of censoring or the hazard functions, will be discussed in other notes.

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. And the test can be constructed as follow:

1. Seperate data into L groups according to your strata variables.
2. Compute $O^{(l)}$, $E^{(l)}$ and $V^{(l)}$ within each stratified level, just as with ordinary logrank test.
3. Compute the test statistics

$$Z = \frac{\sum_{l=1}^{L} \left(O^{(l)} - E^{(l)}\right)}{\sqrt{\sum_{l=1}^{L} V^{(l)}}}.$$

   and under H_0, we have

$$Z \overset{apx}{\sim} N(0,1).$$

Some comments about the stratified logrank tests:

- Intuitively, I think this test assumes $\left\{O^{(l)}, E^{(l)}, V^{(l)}\right\}$ are (approximately) uncorrelated, so the test statistic will be $N(0,1)$.
- If there are too many strata, this test will have poor power. That's because there is no contribution for any 2×2 table once one of the $Y_l(\tau_j)$ becomes zero.
- The stratified logrank test also arises as a score test from Cox's model. This relationship will also clarify the types of alternatives to H_0 for which the stratified logrank test is directed.

3.2 Weighted logrank test

Note that in logrank test, $O_j - E_j$ measures of how $h_0(\tau_j)$ and $h_1(\tau_j)$ differ. And in ordinary logrank test, each time point has the contribution (weights) to the final test statistic.

Suppose we want to compare groups, but in a way that **emphasizes** certain times more than others. Let $w_1, w_2, \cdots, w_k$ be known, non-negative constants. The weighted logrank test is given by

$$Z_w = \frac{\sum_{j=1}^{K} w_j (O_j - E_j)}{\sqrt{\sum_{j=1}^{K} w_j^2 V_j}},$$

and under H_0, $Z_w \overset{apx}{\sim} N(0,1)$.

Some comments about the weighted logrank test:

- Choosing $w_i = 1$ for all i is just the ordinary logrank test.
- One can choose to place larger weights at those τ_js where larger difference is anticipated. But what dose "difference" refer to? $h_0(\tau_j) - h_1(\tau_j)$, $h_0(\tau_j)/h_1(\tau_j)$, $S_0(\tau_j)/S_1(\tau_j)$, $\cdots$
- Choosing $w_j = Y(\tau_j)$ yields the **generalized wilcoxon** test. Since $Y(\tau_1) > Y(\tau_2) > \cdots$, this generalized wilcoxon test places greater weights on **early** differences between $h_0(\cdot)$ and $h_1(\cdot)$.

3.3 Logrank test for multiple groups

So now we want to compare the survival functions among several(> 2) groups. Specifically, there are $p+1$ groups, indexed by $0, 1, 2, \cdots, p$ and the hypothesis is

$$H_0 : \ S_0(\cdot) = S_1(\cdot) = \cdots = S_p(\cdot)$$

Then an extension of the ordinary logrank test can be constructed from Table 3:

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}	$Y_1(\tau_j) - d_{1j}$	$Y_1(\tau_j)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
Group p	d_{pj}	$Y_j(\tau_j) - d_{pj}$	$Y_p(\tau_j)$
Overall	$d_j = \sum_{i=0}^{p} d_{ij}$	$Y(\tau_j) - d_j$	$Y(\tau_j) = \sum_{i=0}^{p} Y_i(\tau_j)$

Table 3: Information at τ_j for multiple groups

And like before, we can construct

$$\boldsymbol{O}_j = \begin{pmatrix} d_{1j} \\ d_{2j} \\ \vdots \\ d_{pj} \end{pmatrix}_{p\times 1}, \quad \boldsymbol{E}_j = \begin{pmatrix} E_{1j} \\ E_{2j} \\ \vdots \\ E_{pj} \end{pmatrix}_{p\times 1}, \quad \text{where } E_{ij} = \frac{Y_i(\tau_j)}{Y(\tau_j)} \cdot d_j$$

and

$$\boldsymbol{V}_j = \begin{pmatrix} V_{11,j} & \cdots & V_{1p,j} \\ \vdots & \ddots & \vdots \\ V_{p1,j} & \cdots & V_{pp,j} \end{pmatrix}_{p\times p}, \quad \text{where } V_{kl,j} = \frac{d_j \left(Y\left(\tau_j\right) - d_j\right) Y_k\left(\tau_j\right) \left(Y\left(\tau_j\right) \cdot 1\left(k=l\right) - Y_k\left(\tau_j\right)\right)}{Y\left(\tau_j\right)^2 \left(Y\left(\tau_j\right) - 1\right)}$$

Then with

$$\boldsymbol{O} = \sum_{j=1}^{K} \boldsymbol{O}_j, \quad \boldsymbol{E} = \sum_{j=1}^{K} \boldsymbol{E}_j, \quad \boldsymbol{V} = \sum_{j=1}^{K} \boldsymbol{V}_j$$

we have under H_0:

$$Q_p = \left(\boldsymbol{O} - \boldsymbol{E}\right)^T \boldsymbol{V}^{-1} \left(\boldsymbol{O} - \boldsymbol{E}\right) \overset{apx}{\sim} \chi^2_p.$$

Some comments about this test

- This test is **omnibus** in terms of how it combines the $p+1$ groups; i.e., it is not directed towards a dose-response, in contrast to the trend test below.

3.4 Logrank trend test

How can we test for a **trend** in the survival functions in the $p+1$ groups? For example, we want to test for a dose-response effect, where higher dose usage implies more significant response (the risk of failure to be monotone with exposure/dose). So we want to design a test that is especially oriented towards this type of alternative to H_0.

Let $\boldsymbol{c}$ = any $p \times 1$ vector of constants. If $\boldsymbol{O} - \boldsymbol{E} \overset{apx}{\sim} N\left(\boldsymbol{0}, \boldsymbol{V}\right)$ under H_0, then

$$\boldsymbol{c}^T \left(\boldsymbol{O} - \boldsymbol{E}\right) \overset{apx}{\sim} N\left(\boldsymbol{0}, \boldsymbol{c}^T \boldsymbol{V} \boldsymbol{c}\right),$$

which means

$$Z_{tr} = \frac{\boldsymbol{c}^T \left(\boldsymbol{O} - \boldsymbol{E}\right)}{\sqrt{\boldsymbol{c}^T \boldsymbol{V} \boldsymbol{c}}} \overset{apx}{\sim} N\left(0, 1\right), \quad \text{under } H_0.$$

Some comments are

- How to choose $\boldsymbol{c}$? One might consider setting $c_j = D_j$ for $j = 1, \cdots p$ where D_j is the dose usage for group j. But in summary the choice of $\boldsymbol{c}$ depends in part on the setting. Sometimes we want to test for monotone trend, other times we want to distinguish a linear versus superlinear(e.g., quadratic) dose-response.

- Entries in $\boldsymbol{c}$ should be **monotone**, but against what specific alternative is a paticular choice of $\boldsymbol{c}$ optimal and what are the consequences of selecting the wrong value of $\boldsymbol{c}$?

4 Power analysis

The power of logrank test under alternative $H_1 : h_1\left(t\right) = h_0\left(t\right) e^{\beta}$ is approximately

$$\Phi\left(\left|\beta\right| \sqrt{D \pi_0 \left(1 - \pi_0\right)} - 1.96\right),$$

where D is the expected number of failures and π_0 is the proportion of patient in groups 0.

To be added. The weighted logrank test, Z_w, follows $N(\theta, 1)$, where $\theta = \xi/\sigma_w$, where

$$\xi = \sqrt{p(1-p)} \int_0^\infty \frac{f_0(s)(1-G_0(s))(1-G_1(s))w(s)g(s)}{(1-p)(1-G_0(s)) + p(1-G_1(s))} \mathrm{d}s, \tag{1}$$

and

$$\sigma_w^2 = \int_0^\infty \frac{f_0(s)(1-G_0(s))(1-G_1(s))w^2(s)}{(1-p)(1-G_0(s)) + p(1-G_1(s))} \mathrm{d}s \tag{2}$$

In later analysis, we assume $P(z_i = 1) = p$ for all i. And define $p(t)$ to be the null probability that some one at risk at time t is in treatment group 1; that is

$$p(t) = P(Z_i = 1 | U_i \geq t)$$

under H_0. Then we have

$$p(t) = \frac{p(1-G_1(t))}{(1-p)(1-G_0(t)) + p(1-G_1(t))}.$$

Also let $v(t)$ denote (under H_0) the density for observing a failure at time t; that is

$$v(t) = ((1-p)(1-G_0(t)) + p(1-G_1(t)))f_0(t).$$

Then we wan re-express ξ and σ_w^2 and

$$\theta = \frac{\xi}{\sigma} = \frac{\int_0^\infty p(t)(1-p(t))w(t)g(t)v(t)\mathrm{d}t}{\sqrt{\int_0^\infty p(t)(1-p(t))w^2(t)v(t)\mathrm{d}t}} \tag{3}$$

Also this NCP arise from the sequence of contiguous alternatives given by

$$H_{A,n}: \quad \log\left(\frac{\lambda_1(t)}{\lambda_2(t)}\right) = n^{-1/2}g(t);$$

that is $g(t)$ is proportional to the log hazard ratio.

For a simple and special case, assume $G_0(\cdot) = G_1(\cdot)$ (i.e., same censoring distribution in each group); $w(t) = 1$ (i.e., ordinary logrank test); $\lambda_1(t)/\lambda_0(t) = \rho$ (i.e., constant hazard ratio). Then we have $p(t) = p$ and (3) simplifies to

$$\theta = \sqrt{p(1-p)} \cdot \sqrt{n}\log\rho \cdot \sqrt{P(\delta = 1)} = \log\rho\sqrt{p(1-p)nP(\delta = 1)}. \tag{4}$$

Here we use the fact, which can be found in previous survival notes that

$$\int_0^\infty v(t)\mathrm{d}t = \int_0^\infty (1-G(t))f(t)\mathrm{d}t = P(U < \infty, \delta = 1) = P(\delta = 1),$$

which is the probability to observe an event. And (4) can tell us how many **events** need to be observed to achieve the target power.

Alternatively, we would fix n and vary the length of study (i.e., vary G_0 and G_1) to given the desired numerical value of θ.

Note: Here we use a simple case, but (3) can be used for arbitrary case. The difficulty is to correctly specify those parameters.

While for parametric test for exponential data, the NCP θ_{exp} is

$$\theta_{exp}^2 = \frac{p(1-p)d_0d_1}{pd_1 + (1-p)d_0},$$

where $d_j = P(\delta_i = 1 | z_i = j, H_0) = \int_0^\infty f_0(t)(1-G_j(t))\mathrm{d}t$, $j = 0, 1$.

$G_0(\cdot) = G_1(\cdot)$: In the censoring distribution in the 2 groups are identical, $d_0 = d_1$ and $p(t) = p$. Therefore

$$ARE = \frac{\theta_{LR}}{\theta_{exp}} = 1,$$

where θ_{LR} is the NCP for logrank test.

Note: this NCP takes different form than the previous approximated one for sample size calculation.

For more reference about sample size calculation of logrank test, one can refer to Lu [2020], Jung and Chow [2012], Lakatos and Lan [1992], Tekindal [2018], Wu [2014].

5 Independent Increments

According to Theorem 1 in Scharfstein et al. [1997], the logrank test statistics are RAL semiparametric efficient and satisfies the independent increments property. To be specific, but not that accurate since I'm not good at stochastic process, (also related to Cox regression), we have asymptotic normality for the estimates of the log hazard ratio

$$\hat{\beta} = \log(HR) \sim N\left(\beta, \frac{1}{p(1-p)n_{evt}}\right),$$

where p is the proportion of subjects being allocated in treatment group and n_{evt} is the number of events observed at analysis. The fisher information of the accumulated data is

$$I(\beta) = p(1-p)n_{evt},$$

Please refer to the fisher information notes in this repo for more details about this concept. The logrank test statistics follows

$$z = \hat{\beta} * \sqrt{p(1-p)n_{evt}} \sim N\left(\beta * \sqrt{p(1-p)n_{evt}}, 1\right).$$

The score? statistic

$$s = I^{1/2}(\beta) z \sim N\left(\beta * I^{1/2}(\beta), \quad I(\beta)\right),$$

has the independent increments property. About this increment, let's remind ourselves that these statistics can be computed along with time, therefore we can have $s(t_1), s(t_2), \cdots, s(t_n)$ as time process, and we know that

$$\left(s(t_1),\ s(t_2) - s(t_1),\ s(t_3) - s(t_2),\ \cdots,\ s(t_n) - s(t_{n-1})\right)^T$$

are independent, that's what we say about independent increments. Therefore

$$\begin{aligned} Var(s(t_i)) &= I(t_i), & i = 1, \cdots, n \\ Cov(s(t_i), s(t_j) - s(t_i)) &= 0, & i \le j \\ Cov(s(t_i), s(t_j)) &= I(t_i), & i \le j. \end{aligned}$$

Hence

$$Cov(z(t_i), z(t_j)) = \sqrt{n_{evt,\,t_i}/n_{evt,\,t_j}}, \quad i \le j.$$

References

Sin-Ho Jung and Shein-Chung Chow. On sample size calculation for comparing survival curves under general hypothesis testing. *Journal of Biopharmaceutical Statistics*, 22(3): 485–495, mar 2012. doi: 10.1080/10543406.2010.550701.

Edward Lakatos and K. K. Gordon Lan. A comparison of sample size methods for the logrank statistic. *Statistics in Medicine*, 11(2):179–191, 1992. doi: 10.1002/sim.4780110205.

Kaifeng Lu. Sample size calculation for logrank test and prediction of number of events over time. *Pharmaceutical Statistics*, 20(2):229–244, sep 2020. doi: 10.1002/pst.2069.

Daniel O. Scharfstein, Anastasios A. Tsiatis, and James M. Robins. Semiparametric efficiency and its implication on the design and analysis of group-sequential studies. *Journal of the American Statistical Association*, 92(440):1342–1350, December 1997. ISSN 1537-274X. doi: 10.1080/01621459.1997.10473655.

Mustafa Agah Tekindal. Power analysis and sample size determination in log-rank (lakatos) test. *Biostatistics and Biometrics Open Access Journal*, 6(1), apr 2018. doi: 10.19080/bboaj.2018.06.555677.

Jianrong Wu. Sample size calculation for the one-sample log-rank test. *Pharmaceutical Statistics*, 14(1):26–33, oct 2014. doi: 10.1002/pst.1654.