
§4.1. Major procedures

Graphical comparison.

One-sample and Multiple-sample tests.

Tests for trend, Stratified tests and Test for matched pairs.

More powerful tests for detecting crossing hazard rate functions.

Tests at fixed time point.

§4.2. An example: Bone Marrow Transplantation for Leukemia Patients.

In a clinical trial, a total of 137 leukemia patients (99 with acute myelocytic leukemia, AML, 38 with acute lymphoblastic leukemia, ALL) went
bone marrow transplantation and then were followed up until relapse, death or end of study. The AML patients are further grouped as AML low-risk (54 patients) and AML high-risk (45 patients). Except relapse, death, other events are also observed. Information other covariates of the patients are also collected. For detail of the trial, see Section 1.3 of Klein and Moeschberger (2003). For the data and its description, see the file bmt.pdf on www.stat.nus.edu.sg/stachenz~.

The following R-session produces a graphical comparison of the survival functions of the three groups of patients:

```r
bmt=read.table("bmt.txt")
bmt.fit=survfit(Surv(T2,d3)~g, data=bmt)
plot(bmt.fit)
```
The estimated survival functions of the three groups
§4.3. One-sample tests

• **Data**

  Distinct event times: \( t_1 < \cdots < t_D \).
  Number of events: \( d_1, \ldots, d_D \).
  Number at risk: \( Y(t_1), \ldots, Y(t_D) \).

• **Hypothesis**

  \( H_0 : h(t) = h_0(t) \) for all \( t \leq \tau \), where \( h_0(t) \) is known.

• **Test statistic**

Let

\[
Z(\tau) = O(\tau) - E(\tau)
\]

\[
= \sum_{i=1}^{D} W(t_i)\frac{d_i}{Y(t_i)} - \int_{0}^{\tau} W(s)h_0(s)ds.
\]
The test statistic is defined as
\[ T(\tau) = \frac{Z(\tau)}{\sqrt{V[Z(\tau)]}} , \]
where
\[ V[Z(\tau)] = \int_0^\tau W^2(s) \frac{h_0(s)}{Y(s)} ds , \]
and \( W(\cdot) \) is a weight function which reflects different emphases over different parts of the survival time.

Estimate of \( V[Z(\tau)] \):
\[ \hat{V}[Z(\tau)] = \sum_{i=1}^{D} W^2(t_i) \frac{h_0(t_i)}{Y(t_i)} \]
or
\[ \sum_{i=1}^{D} W^2(t_i) \frac{d_i}{Y^2(t_i)} . \]

Under \( H_0 \),
\[ T(\tau) \rightarrow N(0, 1) . \]
• The tests

One-sided: \( H_1 : h(t) > h_0(t) \).
Reject \( H_0 \) at level \( \alpha \), if \( T(\tau) > z_{1-\alpha} \).

Two-sided: \( H_1 : h(t) \neq h_0(t) \).
Reject \( H_0 \) at level \( \alpha \), if \( |T(\tau)| > z_{1-\alpha/2} \).

• Weight function

Constant weight:
\[ W(t) \equiv 1. \]

Log-rank weight:
\[ W(t) = Y(t). \]

Harrington-Fleming weight:
\[ W(t) = Y(t)S_0^{\mu}(t)[1 - S_0(t)]^q. \]

When \( W(t) \) is chosen as \( Y(t) \), the test is referred to as **Log-rank test**.
§4.4. Multi-sample tests

• Data:

$K (\geq 2)$ independent censoring samples.

Pooled distinct event times: $t_1 < \cdots < t_D$.

Number of events in $j$th sample: $d_{1j}, \ldots, d_{Dj}$.

Number at risk in $j$th sample: $Y_{1j}, \ldots, Y_{Dj}$.

$$d_i = \sum_{j=1}^{K} d_{ij}, \quad Y_i = \sum_{i=1}^{K} Y_{ij}.$$  

• Hypotheses:

$H_0 : h_1(t) = h_2(t) = \cdots = h_K(t)$, for all $t \leq \tau$,

$H_1 :$ at least one of the $h_j(t)$'s is different from others at some $t \leq \tau$. 

7
• Construction of test statistic

Let $W(t)$ be the weight function.

Measure of the difference of $j$th population from the others:

$$Z_j(\tau) = \sum_{i=1}^{D} W(t_i) \left[ d_{ij} - Y_{ij} \left( \frac{d_i}{Y_i} \right) \right].$$

Note $\sum_{j=1}^{K} Z_j(\tau) = 0$.

Variances and covariances:

$$\hat{\sigma}_{jj} = \hat{V}[Z_j(\tau)]$$

$$= \sum_{i=1}^{D} W^2(t_i) \frac{Y_{ij}}{Y_i} \left[ 1 - \frac{Y_{ij}}{Y_i} \right] \frac{Y_i - d_i}{Y_i - 1} d_i,$$

$$\hat{\sigma}_{jg} = \text{Cov}[Z_j(\tau), Z_g(\tau)]$$

$$= - \sum_{i=1}^{D} W^2(t_i) \frac{Y_{ij} Y_{ig}}{Y_i Y_i} \left( \frac{Y_i - d_i}{Y_i - 1} \right) d_i.$$
The test statistic:

\[ \Sigma_{K-1} = (\hat{\sigma}_{jg})_{j,g=1,...,K-1}, \]
\[ Z = (Z_1(\tau), \ldots, Z_{K-1}(\tau))^t, \]
\[ \chi^2 = Z' \Sigma_{K-1}^{-1} Z. \]

Under \( H_0 \), as sample sizes go to infinity,
\[ \chi^2 \to \chi^2_{K-1}. \]

**A heuristic argument on the statistic.**
Consider the contingency table:

<table>
<thead>
<tr>
<th></th>
<th>Sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure</td>
<td>( j )</td>
<td>others</td>
</tr>
<tr>
<td>Yes</td>
<td>( d_{i,j} )</td>
<td>( d_i )</td>
</tr>
<tr>
<td>No</td>
<td>( y_{i,j} - d_{i,j} )</td>
<td>( Y_i - d_i )</td>
</tr>
<tr>
<td>Total</td>
<td>( Y_{ij} )</td>
<td>( Y_i )</td>
</tr>
</tbody>
</table>
Conditioning on all the marginals, $d_{ij}$ follows the hypergeometric distribution

$$k(d_{ij}, n, N, k) = \frac{k^d_{ij}}{n^{d_{ij}}(N-k)^{N-n}}$$

where $n = Y_{ij}$, $N = Y_i$, $k = d_i$, with mean $\mu_d$ and variance $\sigma^2_d$ given by

$$\mu_d = \frac{nk}{N} = Y_{ij} \frac{d_i}{Y_i}$$

$$\sigma^2_d = \frac{nk(N-k)(N-n)}{N^2(N-1)}$$

$$= \frac{Y_{ij}}{Y_i} \left[ 1 - \frac{Y_{ij}}{Y_i} \right] \frac{Y_i - d_i}{Y_i - 1} d_i.$$

The above conditional distribution motivates the form of the test statistic.
• One-sided test for two samples

\[ Z = \frac{Z_1(\tau)}{\sqrt{\hat{\sigma}_{11}}} = \frac{Z_2(\tau)}{\sqrt{\hat{\sigma}_{22}}} = \frac{\sum_{i=1}^{D} W(t_i)[d_{i1} - Y_{i1}(\frac{d_i}{Y_i})]}{\sum_{i=1}^{D} W^2(t_i)\frac{Y_{i1}}{Y_i}[1 - \frac{Y_{i1}}{Y_i}]\frac{Y_i - d_i}{Y_i - 1}d_i}. \]

Under \( H_0 \), asymptotically,

\[ Z \sim N(0, 1). \]

For alternative \( H_1 : h_1(t) > h_2(t) \), \( H_0 \) is rejected if \( Z > z_{1-\alpha} \).

For alternative \( H_1 : h_1(t) < h_2(t) \), \( H_0 \) is rejected if \( Z < z_\alpha \).

For alternative \( H_1 : h_1(t) \neq h_2(t) \), \( H_0 \) is rejected if \( |Z| > z_{1-\alpha/2} \).
• The choice of $W(t)$

Let $\hat{S}(t)$ be the PL estimate obtained by pooling the $K$ samples together. Fleming and Harrington’s weight specifies:

$$W_{p,q}(t_i) = \hat{S}^p(t_{i-1})[1 - \hat{S}(t_{i-1})]^q, \quad p, q \geq 0.$$ 

Other weights:

$$W_\alpha(t_i) = Y_i^\alpha,$$

Log-rank: $\alpha = 0$

Tarone-Ware: $\alpha = 1/2$

Gehan: $\alpha = 1$.

Define

$$\tilde{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{Y_i + 1}\right).$$

Peto-Peto: $W(t_i) = \tilde{S}(t_i)$,

Modified Peto-Peto: $W(t_i) = \tilde{S}(t_i)Y_i/(Y_i + 1)$. 

12
Figure 7.2  Relative weights for comparison of observed and expected numbers of deaths for kidney dialysis patients.
• R function for the tests

\texttt{survdiff(formula, data, subset,}
  \texttt{ na.action, rho=0)}

- \texttt{formula}: a formula expression as for other survival models, of the form
  \begin{align*}
  \text{Surv}(\text{time}, \text{status}) \sim & \text{predictors.}
  \end{align*}
  
  For a one-sample test, the predictors must consist of a single \texttt{'offset(sp)'} term, where \texttt{'sp'} is a vector giving the survival probability of each subject.

  For a k-sample test, each unique combination of predictors defines a subgroup.

- \texttt{rho}: the exponent in the weight function:
  \begin{equation}
  W(t) = \hat{S}^\rho(t).
  \end{equation}
Example (cont.): Test the difference among the three groups of Bone Marrow Transplant patients:

```r
bmt=read.table("bmt.txt")
bmt.test=survdiff(Surv(T2,d3)~g, 
    data=bmt)
bmt.test
```

```
N  Obs  Exp  (O-E)^2/E  (O-E)^2/V
  g=1 38  24  21.9     0.211   0.289
  g=2 54  25  40.0     5.604  11.012
  g=3 45  34  21.2     7.756  10.529

Chisq= 13.8 on 2 degrees of freedom,
p= 0.00101
```

```r
bmt.test$var
```

```
[,1]        [,2]        [,3]
[1,]    15.955175  -10.345092  -5.610084
[2,]   -10.345092   20.339789  -9.994697
[3,]   -5.610084  -9.994697  15.604781
```

§4.5. Tests for trend

- **Hypotheses:**
  \[ H_0 : h_1(t) = h_2(t) = \cdots = h_K(t), \text{ for } t \leq \tau, \]
  \[ H_1 : h_1(t) \leq h_2(t) \leq \cdots \leq h_K(t), \text{ for } t \leq \tau, \]
  with at least one strict inequality.
  \[ \Leftrightarrow S_1(t) \geq \cdots \geq S_K(t). \]

- **Test statistic**
  Let
  \[ a_1 < a_2 < \cdots < a_K \]
  be positive scores. In particular, \( a_j = j \).
  The test statistic is given by
  \[ Z = \frac{\sum_{j=1}^{K} a_j Z_j(\tau)}{\sqrt{\sum_{j=1}^{K} \sum_{g=1}^{K} a_j a_g \hat{\sigma}_{jg}}}. \]
Under $H_0$, asymptotically,

$$Z \sim N(0, 1).$$

$H_0$ is rejected if $Z \geq z_{1-\alpha}$.

Remark: The test is more powerful for detecting the trend. However, it should only be used when prior information indicates that the alternatives are ordered.

- **Computation issue**
  
  Use `survdiff` to obtain
  
  $$\hat{\Sigma}_K = (\hat{\sigma}_{jg}), \ Z = (Z_1(\tau), \ldots, Z_K(\tau))'.$$

  Then compute
  
  $$Z = \frac{a'Z}{\sqrt{a'\hat{\Sigma}a}}.$$

  $$a = (a_1, \ldots, a_K)'.$$
Example: Bone Marrow Transplant (cont.)

\[ Z = \text{bmt.test}\$\text{obs} - \text{bmt.test}\$\text{exp} \]
\[ AA = \text{bmt.test}\$\text{var} \]
\[ a = c(1, 2, 3) \]
\[ TT = t(a) \times Z / \sqrt{t(a) \times AA \times a} \]

\[ Z: \quad 2.148285 \quad -14.966116 \quad 12.817830 \]

\[ AA: \]
\[ \begin{array}{ccc}
    15.955175 & -10.345092 & -5.610084 \\
   -10.345092 & 20.339789 & -9.994697 \\
   -5.610084 & -9.994697 & 15.604781
\end{array} \]

\[ a: \quad 1 \quad 2 \quad 3 \]

\[ TT: \quad 1.631266 \]
§4.6. Stratified tests

Assume the populations are stratified into $M$ strata by covariates.

$s$: index for stratum.
$j, g$: index for population.
$i$: index for individual.

- **Construction of test statistic**

  For each stratum, construct the statistics

  $Z_{js}(\tau) = \sum_{i=1}^{D} W(t_i) \left[ d_{ijs} - Y_{ijs} \left( \frac{d_{is}}{Y_{is}} \right) \right]$,

  $\hat{\sigma}_{jjs} = \sum_{i=1}^{D} W^2(t_i) \frac{Y_{ijs}}{Y_{is}} \left[ 1 - \frac{Y_{ijs}}{Y_{is}} \right] \frac{Y_{is} - d_{is}}{Y_{is} - 1} d_{is},$

  $\hat{\sigma}_{jgs} = -\sum_{i=1}^{D} W^2(t_i) \frac{Y_{ijs}}{Y_{is}} \frac{Y_{igs}}{Y_{is}} \left( \frac{Y_{is} - d_{is}}{Y_{is} - 1} \right) d_{is}.$

  Let

  $Z_j(\tau) = \sum_{s=1}^{M} Z_{js}(\tau), \quad \hat{\sigma}_{jg} = \sum_{s=1}^{M} \hat{\sigma}_{jgs}.$
The test statistic is computed as

$$\chi^2 = (Z_1(\tau), \cdots, Z_{K-1}(\tau))\Sigma^{-1}(Z_1(\tau), \cdots, Z_{K-1}(\tau))',$$

where $$\Sigma = (\hat{\sigma}_{jg})_{j,g=1,\ldots,K-1}.$$ Under $$H_0$$, $$\chi^2$$ has an asymptotic chi square distribution with d.f. $$K - 1$$.

Two-sample case: Test statistic is

$$Z = \frac{\sum_{s=1}^{M} Z_{1s}(\tau)}{\sqrt{\sum_{s=1}^{M} \hat{\sigma}_{11s}}}.$$ Under $$H_0$$, $$Z \sim N(0, 1).$$ Both one- and two-sided alternatives can be tested with $$Z$$.

The stratified tests can be computed by `survdiff` with the specification of `formula` as follows:

`Surv(time, status) ~ . + strata(s.v)`
• **Matched pair tests**

By treating each matched pair as a stratum, the stratified tests can be used to test matched pairs. The two-sample statistic $Z$ reduces to

$$Z = \frac{D_1 - D_2}{D_1 + D_2},$$

$D_1$: the number of matched pairs in which the individual from sample 1 experience the event first.

$D_2$: the number of matched pairs in which the individual from sample 2 experience the event first.

The matched pairs, where the smaller of the two times is a censored observation, make no contribution to the test statistic.
§4.7. More powerful tests for crossing hazard rate functions

When hazard functions cross each other, the tests in previous sections have little power. The tests in this section are designed to test crossing hazard functions with more power.

• Renyi Type tests

\[ H_0 : h_1(t) = h_2(t), t < \tau, \]

Notation:

\[ t_1 < \cdots < t_D : \text{pooled distinct event times.} \]
\[ d_{ij}, Y_{ij} : \text{# of events and at risk for group } j. \]
\[ d_i = d_{i1} + d_{i2}, \quad Y_i = Y_{i1} + Y_{i2}. \]

Weight function: \( W(t_i) = 1 \) or \( Y_i. \)
Construction of test statistic

For each \( t_i, i = 1, \ldots, D \), compute

\[
Z(t_i) = \sum_{t_k \leq t_i} W(t_k) \left[ d_{k1} - Y_{k1} \left( \frac{d_k}{Y_k} \right) \right].
\]

Let

\[
\sigma^2(\tau) = \sum_{t_k \leq \tau} W^2(t_k) d_k \frac{Y_{k1} Y_{k2} (Y_k - d_k)}{Y_k^2 (Y_k - 1)}.
\]

To test the two-sided alternative: \( h_1(t) \neq h_2(t) \), the test statistic is given by

\[
Q = \max_{1 \leq i \leq D} \{ |Z(t_i)| \} / \sigma(\tau).
\]

Under \( H_0 \),

\[
Q \sim \sup \{ |B(x)|, 0 \leq x \leq 1 \},
\]

where \( B(x) \) is the standard Brownian motion process. Critical values are given in Table C.5 of Klein and Moeschberger (2003).
To test one-sided alternative: \( h_1(t) > h_2(t) \),
the test statistic is given by

\[
Q^* = \max_{1 \leq i \leq D} \{ Z(t_i) \} / \sigma(\tau).
\]

Under \( H_0 \),
\[
Q^* \sim \sup\{ B(x), 0 \leq x \leq 1 \}.
\]
The \( p \)-value is determined as

\[
Pr(\sup B(x) > Q^*) = 2[1 - \Phi(Q^*)],
\]
where \( \Phi \) is the CDF of the standard normal distribution.

**A note on computation**

The R function `survdiff` can be adapted
to obtain \( Z(t_i) \) and \( \sigma(\tau) \).
§4.8. Tests at fixed point in time

It is needed sometimes to compare the survival functions or cumulative incidence functions at a fixed time point. The general method of multiple comparison can be applied.

● A review of general multiple comparison

A vector \( \mathbf{c} = (c_1, \ldots, c_K) \) is said to be a contrast vector if
\[
\sum_{j=1}^{K} c_j = 0.
\]

Let \( \mathbf{C} \) be a \( q \times K \) matrix consists of \( q \) contrast row vectors, i.e.,
\[
\mathbf{C} = \begin{pmatrix}
\mathbf{c}_1' \\
\vdots \\
\mathbf{c}_K'
\end{pmatrix}.
\]
To test $H_0 : C\theta = 0$, where $\theta = (\theta_1, \ldots, \theta_K)'$, the test statistic is formed as

$$
\chi^2 = \hat{\theta}' C'[C\hat{\Sigma}_\theta C'][-1]C\hat{\theta},
$$

where $\hat{\theta}$, which is asymptotically normally distributed, is a consistent estimator of $\theta$, and $\hat{\Sigma}_\theta$ is the estimated variance matrix of $\hat{\theta}$.

Under $H_0$, $\chi^2$ has an asymptotic $\chi^2$-distribution with d.f. $q$.

The hypothesis: $\theta_1 = \cdots = \theta_K$ is equivalent to $C'\theta = 0$ with

$$
C = \begin{pmatrix}
1 & 0 & \cdots & 0 & -1 \\
0 & 1 & \cdots & 0 & -1 \\
\ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & \cdots & 1 & -1 \\
\end{pmatrix}.
$$
• Tests for survival functions

Suppose there are $K$ groups with survival functions $S_1(t), \ldots, S_K(t)$. It is to test whether the survival functions differ at a particular point $t_0$.

$$\theta = (S_1(t_0), \ldots, S_K(t_0))'$$

$$\hat{\theta} = (\hat{S}_1(t_0), \ldots, \hat{S}_K(t_0))'$$

$$\hat{\Sigma}_\theta = \text{Diag}(\hat{V}[\hat{S}_1(t_0)], \ldots, \hat{V}[\hat{S}_K(t_0)])$$

$\hat{S}_j(t_0)$ : PL estimate of $S_j(t_0)$. 
• Tests for cumulative incidence functions

The purpose of the tests is to see whether the incidence functions of a certain risk among \( K \) different groups differ at a particular time point \( t_0 \). Suppose the risk factor of concern is indexed by \( r \).

\[
\theta = (F_{1r}(t_0), \ldots, F_{Kr}(t_0))'
\]

\[
\hat{\theta} = (\hat{F}_{1r}(t_0), \ldots, \hat{F}_{Kr}(t_0))'
\]

\[
\hat{\Sigma}_\theta = \text{Diag}(\hat{V}[\hat{F}_{1r}(t_0)], \ldots, \hat{V}[\hat{F}_{Kr}(t_0)]).
\]

where \( F_{jr}(t_0) \) is the \( r \)th cumulative incidence function in the \( j \)th group.
**Pairwise comparison with Bonferroni adjustment**

Test $H_{ij0} : \theta_i = \theta_j$ for $M$ pairs.

Test statistics:

$$Z_{ij} = \frac{\hat{\theta}_i - \hat{\theta}_j}{\sqrt{V[\hat{\theta}_i - \hat{\theta}_j]}}.$$ 

Simultaneous critical value:

$$c_\alpha = z_{1-\alpha/M}.$$ 

Reject $H_{ij0}$, if $|Z_{ij}| > c_\alpha$.

**Computation**

Use `survfit` and `cumu.inci` to obtain $\hat{\Theta}, \hat{\Sigma}_\theta$ respectively for testing the survival functions and cumulative incidence functions.
• **Example:** Bone Marrow Transplant (cont.)

The estimated **relapse** cumulative incidence functions at $t_0 = 365$ days for ALL, AML low risk and AML high risk groups and their corresponding variances are obtained using `cumu_inci` as

$$
\hat{F}_{\text{ALL}} = 0.238, \quad \hat{V}(\hat{F}_{\text{ALL}}) = 0.0048;
\hat{F}_{\text{AML}} = 0.093, \quad \hat{V}(\hat{F}_{\text{AML}}) = 0.0015;
\hat{F}_{\text{AMLH}} = 0.378, \quad \hat{V}(\hat{F}_{\text{AMLH}}) = 0.0056.
$$

Contrast matrix

$$
C = \begin{pmatrix}
1 & 0 & -1 \\
0 & 1 & -1
\end{pmatrix}.
$$

$$
C\hat{\Sigma}C' = \begin{pmatrix}
0.01037 & 0.00559 \\
0.00559 & 0.00713
\end{pmatrix}.
$$

$$
\chi^2 = 12.59, \quad p\text{-value} = 0.0018
$$
R-session for the computation:

```r
bmt = read.table("bmt.txt")
tmp.dat = data.frame(time=bmt$T2, status=bmt$d2+bmt$d3, g=bmt$g)
g = tmp.dat$g

inci1 = cumu.inci(K=2, data=tmp.dat, subset=(g==1), plot=T)
inci2 = cumu.inci(K=2, data=tmp.dat, subset=(g==2), plot=T)
inci3 = cumu.inci(K=2, data=tmp.dat, subset=(g==3), plot=T)

e.time1 = inci1$event.time
e.time2 = inci2$event.time
e.time3 = inci3$event.time
n1 = length(e.time1)
n2 = length(e.time2)
n3 = length(e.time3)
```
\[ t_1 = \max(c(1:n_1)[e.\text{time1} \leq 365]) \]
\[ t_2 = \max(c(1:n_2)[e.\text{time2} \leq 365]) \]
\[ t_3 = \max(c(1:n_3)[e.\text{time3} \leq 365]) \]

\[ \text{snv} = \text{rbind}(\text{inci1}[t_1, c(3,5)], \text{inci2}[t_2, c(3,5)], \text{inci3}[t_3, c(3,5)]) \]

\[ \text{C} = \text{matrix}(c(1,0,-1,0,1,-1), \text{ncol}=3, \text{byrow}=T) \]
\[ \text{chi} = \text{t}(\text{C}^\ast \text{snv}[,1]) \text{solve}(\text{C}^\ast \text{diag(snv[,2])} \text{t(C)}) \text{C}^\ast \text{snv}[,1] \]
\[ 1 - \text{pchisq(chi,2)} \]