Resampling methods for randomly censored survival data

Markus Pauly

Lehrstuhl für Mathematische Statistik und Wahrscheinlichkeitstheorie
Mathematisches Institut
Heinrich-Heine-Universität Düsseldorf

Wien, 5 June 2014

Thanks to

- J. Beyersmann (Ulm)
- M. Brendel (Boehringer Ingelheim, Biberach)
- D. Dobler (Düsseldorf)
- A. Janssen (Düsseldorf)
- C.-D. Mayer (BioSS, Aberdeen)
- S. di Termini (Freiburg)

Outline

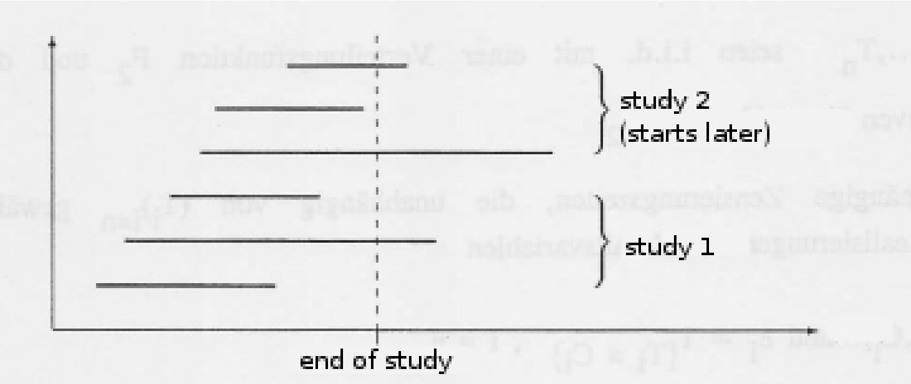
- Typcial study in medicine: 30-40 patients

$\rightarrow$ Sample sizes are only sufficient if efficiently analyzed!

- Problems in practice
    - Incompleteness of observations;
    - e.g. right-censored data

- Observed are 2-types of data
    - Complete observations, so called "survival times"
    - Incomplete observations; last observed survival time or drop out of study for other reasons

Censoring



Source: Janssen

- Typcial study in medicine: 30-40 patients

$\rightarrow$ Sample sizes are only sufficient if efficiently analyzed!

- Problems in practice
  - Incompleteness of observations;
  - e.g. right-censored data

- Observed are 2-types of data
  - Complete observations, so called “survival times”
  - Incomplete observations; last observed survival time or drop out of study for other reasons

- Example (“kidney data”):
  Infection times in two groups of dialysis patients

Different catheterization procedures: percutaneous and surgical placements

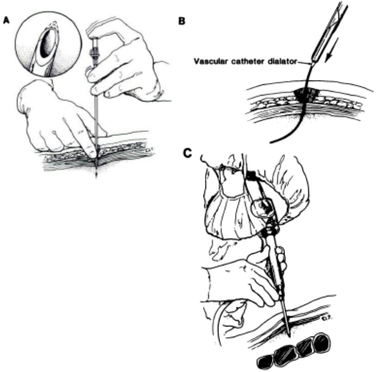


Figure 1. Technical considerations (see Text for details). (A) The tip of a 6-inch, 14-gauge angiocatheter is dulled (insert), and the catheter is used to enter the peritoneal cavity. (B) The peritoneoscopic guide, precannulated with a 7 French vascular catheter dilator, is advanced over a flexible guidewire. (C) Peritoneoscopic visualization of the peritoneal cavity.

Source: Klein/Moeschberger

Infection times of dialysis patients

TABLE 1.2

Times to infection (in months) of kidney dialysis patients with different catheterization procedures

Surgically Placed Catheter
Infection Times: 1.5, 3.5, 4.5, 4.5, 5.5, 8.5, 8.5, 9.5, 10.5, 11.5, 15.5, 16.5, 18.5, 23.5, 26.5
Censored Observations: 2.5, 2.5, 3.5, 3.5, 3.5, 4.5, 5.5, 6.5, 6.5, 7 5, 7.5, 7.5, 7.5, 8.5, 9.5, 10.5, 11.5, 12.5, 12.5, 13.5, 14.5, 14.5, 21.5, 21.5, 22.5, 22.5, 25.5, 27.5

Percutaneous Placed Catheter
Infection Times: 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 2.5, 2.5, 3.5, 6.5, 15.5
Censored Observations: 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.5, 1.5, 1.5, 1.5, 2.5, 2.5, 2.5, 2.5, 2.5, 3.5, 3.5, 3.5, 3.5, 3.5, 4.5, 4.5, 4.5, 5.5, 5.5, 5.5, 5.5, 5.5, 6.5, 7.5, 7.5, 7.5, 8.5, 8.5, 8.5, 9.5, 9.5, 10.5, 10.5, 10.5, 11.5, 11.5, 12.5, 12.5, 12.5, 12.5, 14.5, 14.5, 16.5, 16.5, 18.5, 19.5, 19.5, 19.5, 20.5, 22.5, 24.5, 25.5, 26.5, 26.5, 28.5

Basic Survival Analysis 1-4

- $T \geq 0$ *survival time*
- $S(t) = P(T > t)$ *survival function* with ν-density f
- $\Lambda(t) = \int_0^t f(t)/S(t-)d\nu(t)$ *cumulative hazard function*
- $\lambda = f/S$ *hazard rate*
- Interpretation (for smooth f)

$$P(T \in (t, t+\epsilon] | T \geq t) \approx \epsilon\lambda(t).$$

Basic Survival Analysis 2-4

- Random Censoring by $C \sim G$;
- C and T independent
- We only observe $X := \min(T, C)$ and $\Delta = \mathbf{1}\{T \leq C\}$

- Estimator for S in this case:

$\Rightarrow$ *Kaplan-Meier estimator* $\hat{S}$

- Estimator for Λ:

$\Rightarrow$ *Nelson-Aalen estimator* $\hat{\Lambda}$

Basic Survival Analysis 3-4

"Easiest" case:

- Let $(T_i)_i$ be i.i.d. discrete r.v.s; indep. from $(C_i)_i$ (also i.i.d.)
- Define (e.g. failure rates in actuarial sciences)[1]

$$r(t) = P(T_1 = t | T_1 \geq t) = \frac{P(T_1 = t)}{P(T_1 \geq t)}$$

- Theorem:

$$\Lambda(t) = \Lambda_T(t) = \sum_{0 \leq s \leq t} r(s) \text{ and } S(t) = S_T(t) = \prod_{0 \leq s \leq t} (1 - r(s))$$

- Kaplan-Meier and Nelson-Aalen estimators by plug-in:

$$\hat{r}(s) = \frac{\text{number of failures at time } s}{\text{number under risk at } s-}$$

[1] $r(t) = 0$ for $P(T_1 \geq t) = 0$

Basic Survival Analysis 4-4

In counting process notation:

- Define counting processes

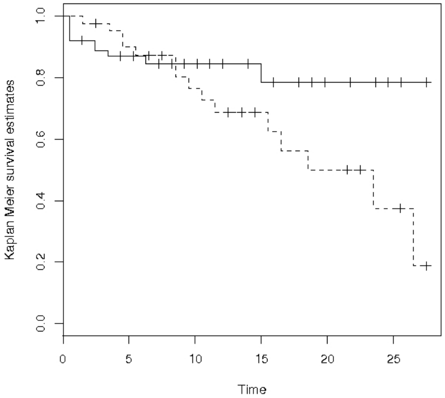
$$N(t) = \sum_{i=1}^{n} \mathbf{1}\{X_i \leq t, \Delta_i = 1\}, \;\; Y(t) = \sum_{i=1}^{n} \mathbf{1}\{X_i \geq t\}$$

- $\Delta N(t) = N(t) - N(t-)$
- Nelson-Aalen and Kaplan-Meier are given by

$$\hat{\Lambda}(t) = \sum_{0 \leq s \leq t} \frac{\Delta N(s)}{Y(s)} \text{ and } \hat{S}(t) = \prod_{0 \leq s \leq t} \left(1 - \frac{\Delta N(S)}{Y(s)}\right)$$

- Uncensored case: $1 - \hat{S} =$ e.d.f.
- Properties (under reg): Consistent and asymptotic Gaussian

Plots of Kaplan Meier estimators for the two sample kidney data



solid line = percutaneous

- Here: Randomly right censored data
- Patients that survive without infection are censored
- Standard model: Cox proportional hazard model[2]

[2] $\lambda_\vartheta(t) = e^\vartheta \lambda_0(t)$

Questions:

- Procedures significant different?
- Is one survival time (or hazard rate) stochastic greater?

Answer of classical log-rank test (for proportional hazards)

- p-value > 0.05
- Reason: time depending hazard ratios!
- Log rank test has problems detecting them!

Classical two sample problem

- Observe independent random variables
  $X_i := \min(T_i, C_i)$ and $\Delta_i = \mathbf{1}\{T_i \leq C_i\}$
  for $1 \leq i \leq n$ within two groups
- $T_i \overset{i.i.d.}{\sim} F_1,\ 1 \leq i \leq n_1,\ T_{n_1+i} \overset{i.i.d.}{\sim} F_2,\ 1 \leq i \leq n_2,$ (continuous)
- $C_i \overset{i.i.d.}{\sim} = G_1,\ 1 \leq i \leq n_1,\ \ C_{n_1+i} \overset{i.i.d.}{\sim} G_2,\ 1 \leq i \leq n_2.$ (continuous)
- Null hypothesis

$$H_0 : \{T_1, \ldots, T_n \text{ i.i.d.}\} = \{F_1 = F_2\} = \{\Lambda_1 = \Lambda_2\}$$

  against 2-sided or 1-sided alternatives

$$H_1^0 : \{\Lambda_1 \neq \Lambda_2\}, \quad H_1^1 : \{\lambda_1 \gneqq \lambda_2\},\ H_1^2 : \{\Lambda_1 \gneqq \Lambda_2\}.$$

- Unknown nuisance param: G_1, G_2 and under the null $\mathcal{L}(T_1)$

Class of weighted logrank test statistics

- $w_n(t) := \tilde{w}(\widehat{F}_n(t-))$, $\widehat{F}_n$ Kaplan Meier estimator of the pooled sample
- $\tilde{w} : [0,1] \to \mathbb{R}$ weight function
- Test statistics (ABGK 1993) are $\frac{T_n(w_n)}{\sigma(w_n)}$ or $\frac{T_n(w_n)}{\sigma(w_n)}\mathbf{1}\{T_n(w_n) > 0\}$ with

$$T_n(w_n) = \sqrt{\frac{n}{n_1 n_2}} \int_0^\infty w_n(s) \frac{Y_1(s) Y_2(s)}{Y(s)} \left\{ d\hat{\Lambda}_{1n}(s) - d\hat{\Lambda}_{2n}(s) \right\}$$

- $\sigma^2(w_n)$ adequate variance estimator[3]
  - $\tilde{w} = 1$ for classical logrank statistic
  - $\tilde{w}(u) = 1 - u$ for Prentice-Wilcoxon statistic

[3] Gill (1980): $\sigma^2(w_n) = \frac{n}{n_1 n_2} \int_0^\infty w_n^2 \frac{Y_1 Y_2}{Y} d\hat{\Lambda}_n$

Remarks

- Intuitively: $\tilde{w}$ determines power behaviour
- Example: Prentice-Wilcoxon with decreasing $\tilde{w}(u) = 1 - u$

$\Rightarrow$ More weight on early times!

$\Rightarrow$ Should have good power against early hazard differences!

- In comparison: Classical logrank: Equal weight on all time points!
- Now: ARE comparison via reparametrization!

Local parametrization (scores)

Classical approach:

- classical likelihood based approach
- parametric path $\vartheta \mapsto P_\vartheta$ in a nonparametric model $\mathcal{P}$ of distributions
- score functions $\frac{d}{d\vartheta} \log \frac{dP_\vartheta}{d\mu} = g(\vartheta)$

$\Rightarrow$ score test

Local parametrization (hazards)

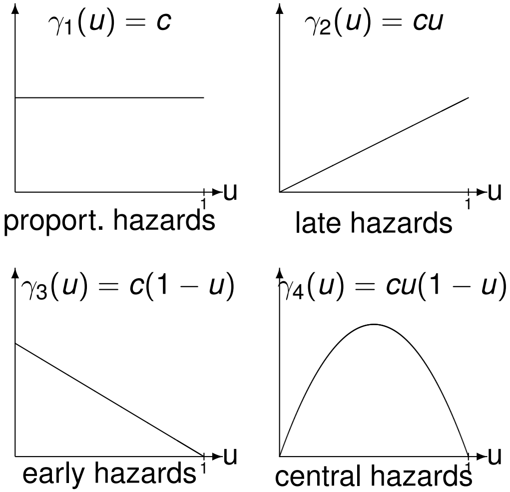
- approach based on hazards:
- $F = 1 - S \longleftrightarrow \Lambda$ cumulative hazard function
- Now: hazard rates $\lambda(u) = \frac{f(u)}{S(u)}$ rather than densities $f(u)$
- parameters: relative risk = hazard rate ratio
- represented by $\gamma : [0, 1] \to \mathbb{R}$ under ass

$$\lim_{\vartheta \to 0} \frac{1}{\vartheta}\left(\frac{d\Lambda_\vartheta}{d\Lambda_0} - 1\right) = \gamma \circ F_0.$$

- Rem: Holds for ($\vartheta \ll 1$)

$$\Lambda_\vartheta(t) := \int_0^t (1 + \vartheta\gamma \circ F) d\Lambda_0, \quad \Lambda_0 \text{ baseline hazard.}$$

Examples of semiparametric models



- Consider model given by $\gamma \circ F_0$
- Recall $w_n(t) := \tilde{w}(\widehat{F}_n(t-))$, $\tilde{w} : [0,1] \to \mathbb{R}$ weight function
- Janssen (1991) and Neuhaus (2000): ARE of $\frac{T_n(w_n)}{\sigma(w_n)} \mathbf{1}\{T_n(w_n) > 0\}$ for local alternatives[4] in direction of $\gamma \circ F_0$ given by

$$\text{ARE} = \frac{\langle \tilde{w}, \gamma \rangle_\mu^2}{\|\tilde{w}\|_\mu^2 \|\gamma\|_\mu^2} = \cos^2(\beta)$$

- $\langle \tilde{w}, \gamma \rangle_\mu = \int_0^1 \tilde{w} \gamma d\mu$, β angle between $\tilde{w}$ and γ
- μ measure on $\mathbb{R}_+$ depending on G_i, F_0 and $\lim_n \frac{n_1}{n}$

$\Rightarrow$ Bad ARE for some directions γ!

[4] $\mathcal{L}(T_1, \ldots, T_n) = \bigotimes_i P_{c_{ni}}$

On the ARE

- ARE (ψ): asymptotic relative efficiency of a test ψ

$$\text{ARE}(\psi) \approx \frac{N_{opt}}{N(\psi)} \quad \text{(in the limit)}$$

- $N(\psi) =$ No. of needed obs. for ψ to achieve a given power
- $N_{opt} =$ Minimum no. of needed obs.
- Pitman's interpretation: $100(1 - \text{ARE})\%$ of observations are wasted by using ψ

$\Rightarrow$ Use of the full data information only for ARE $= 1$!

- Example:
  ARE(ψ)=$\frac{1}{2}$ $\Rightarrow$ twice as many obs. are necessary for ψ
- Clearly: Optimal procedures depend on the model
- which is in general unknown!
- Idea: “Estimate“ the model! (locally)

Closer look to weighted logrank statistics

- Consider emp. subspace $V = \{\beta w_n : \beta \in \mathbb{R}\}$ and cone $V_1^+ = \{\beta w_n : \beta \in [0, \infty)\}$, generated by w_n.

Lemma (Brendel et al., 2014)

We have

$$\frac{T_n(w_n)}{\sigma(w_n)} = \left\|\Pi_{V_1}(\hat{\gamma}_n)\right\|_{\hat{\mu}_n}; \frac{T_n(w_n)}{\sigma(w_n)} \mathbf{1}\{T_n(w_n) > 0\} = \left\|\Pi_{V_1^+}(\hat{\gamma}_n)\right\|_{\hat{\mu}_n},$$

where

- $\hat{\mu}_n \approx$ *emp. estimator of* μ, $\hat{\gamma}_n$ = *emp. estimator of* $\gamma \circ F_0$ *(locally* [a]*)*
- $\Pi_{V_1^{(+)}}$ = $L_2(\hat{\mu}_n)$*-projection into* $V_1^{(+)}$

[a] all estimates depend on NA (and Y); e.g. $\hat{\gamma}_n \approx \sqrt{\frac{n_1 n_2}{n}}(d\widehat{\Lambda}_1 - d\widehat{\Lambda}_2)/d\widehat{\Lambda}_n$

Interpretation

- $\hat{\gamma}_n$ estimates the $\gamma \circ F_0$-model (locally)
- Logrank stat measures distance of its projection into space/cone generated by w_n
- Idea: Use larger spaces/cones to cover larger classes of alternatives!

Projection approach

- Idea: Scientist chooses “relevant“ weights[5]

$$w_{in}(\cdot) := \tilde{w}_i(\widehat{F}_n(\cdot -)) \quad 1 \leq i \leq r,$$

- e.g. to discover differences of the relative risk for
  - proportional hazards (constant over time)
  - early survival times
  - central survival times
  - late survival times.
- These generate larger linear space/cone:

$$V := \left\{ \sum_{i=1}^{r} \beta_i \, w_{in}, \; \beta_i \in \mathbb{R}, \right\} \text{ and } V^+ := \left\{ \sum_{i=1}^{r} \beta_i \, w_{in}, \; \beta_i \geq 0, \right\}$$

[5] $\widehat{F}_n$ Kaplan Meier estimator of the pooled sample

Projection statistics

- As in the one-dimensional case consider test statistics

$$S_{n0} := \|\Pi_V(\hat{\gamma}_n)\|^2_{\hat{\mu}_n} \text{ and } S_{n1} := \|\Pi_{V^+}(\hat{\gamma}_n)\|^2_{\hat{\mu}_n}$$

Theorem (Brendel et al., 2014)

Under regularity conditions we have convergence in distribution under the null

$$\|\Pi_V(\hat{\gamma}_n)\|^2_{\hat{\mu}_n} \xrightarrow{d} \chi^2_{rank(\Sigma)} \text{ and } \|\Pi_{V^+}(\hat{\gamma}_n)\|^2_{\hat{\mu}_n} \xrightarrow{d} F_\Sigma$$

where F_Σ is continuous and $\Sigma = (\langle \tilde{w}_i, \tilde{w}_j \rangle_\mu)_{i,j \le r}$.

Projection-type tests

- Estimate Σ consistently by $\widehat{\Sigma}$
- Results in tests

$$\phi_{0n} = \mathbf{1}\{\|\Pi_V(\hat{\gamma}_n)\|^2_{\hat{\mu}_n} - \chi^2_{\mathrm{rank}(\widehat{\Sigma}),1-\alpha} > 0\},$$

$$\phi_{1n} = \mathbf{1}\{\|\Pi_{V^+}(\hat{\gamma}_n)\|^2_{\hat{\mu}_n} - F^{-1}_{\widehat{\Sigma}}(1-\alpha) > 0\},$$

Theorem (Brendel et al., 2014)

Both tests are

- *asymptotical level α (even for $G_1 \neq G_2$) and*
- *consistent for fixed alternatives if at least one of their associated weighted logrank tests is consistent.*

$\Rightarrow$ Posses broader power functions!

Projection-type permutation tests

- Further modification: Use permutation version of the tests!
- Advantage: Finitely exact under $\{F_1 = F_2, G_1 = G_2\}$.
- For $c_{n0} = \chi^2_{\text{rank}(\widehat{\Sigma}),1-\alpha}$ and $c_{n1} = F^{-1}_{\widehat{\Sigma}}(1-\alpha)$:

$$\phi^*_{nk} := \mathbf{1}\{S_{nk} - c_{nk} > c^*_{nk}\} + k^*_n \mathbf{1}\{S_{nk} - c_{nk} = c^*_{nk}\} \quad k = 0, 1,$$

with

- c^*_{nk} = cond $(1-\alpha)$-quantile of the perm dist of $S_{nk} - c_{nk}$
- Role of $\tilde{S}_{nk} = S_{nk} - c_{nk}(\widehat{\Sigma})$: Studentized-type statistics!

Properties

Theorem (Brendel et al., 2014)

The permutation tests are asymptotically equivalent to their corresponding projection tests, i.e.

$$\lim_{n\to\infty} E_{H_0}(|\phi_{0,n} - \phi^*_{0,n}|) = 0,$$

$$\lim_{n\to\infty} E_{H_0}(|\phi_{1,n} - \phi^*_{1,n}|) = 0.$$

- Implies same power for contigouos alternatives!
- More math. details (as asymptotic admissability) in Brendel, Janssen, Mayer & Pauly (2014, SJS).

Set-up 1-2

- Group I: $T_1, \ldots, T_{n_1}$ i.i.d. $F_0(x) = (1 - \exp(-x))\mathbf{1}_{(0,\infty)}(x)$
- For group II: 3 different scenarios; given by directions $\gamma_i = \tilde{w}_i$

$$\tilde{w}_1(u) = 1, \quad \tilde{w}_2(u) = 1 - 2u, \quad \tilde{w}_3(u) = u(1-u), \quad 0 \le u \le 1,$$

- corresponding to proportional, crossing and central hazards
- Group II: $T_{n_1+1}, \ldots, T_n$ i.i.d. with

$$\Lambda_{\vartheta,i}(t) = \int_0^t 1 + \vartheta \tilde{w}_i(F_0(x))dx$$

Set-up 2-2

- Censoring: $C_i \overset{u.i.v.}{\sim} Exp(0.2)$ in both groups.

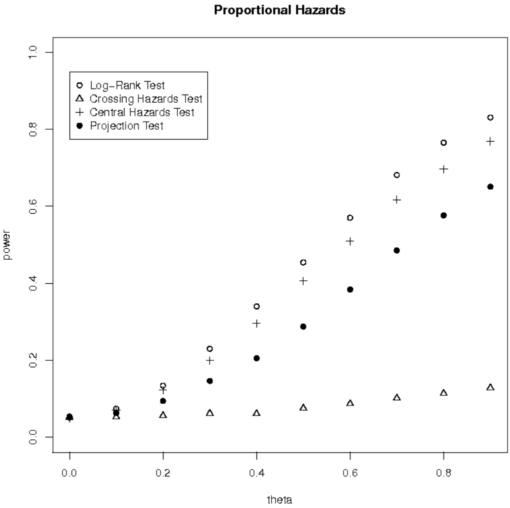
Simulated power of

- 2-sided weighted logrank test in $T_n(\tilde{w}_i)$ (optimal for direction $\tilde{w}_i$)
- Projection test $\phi_{0,n}$ (covering all 3 directions.)

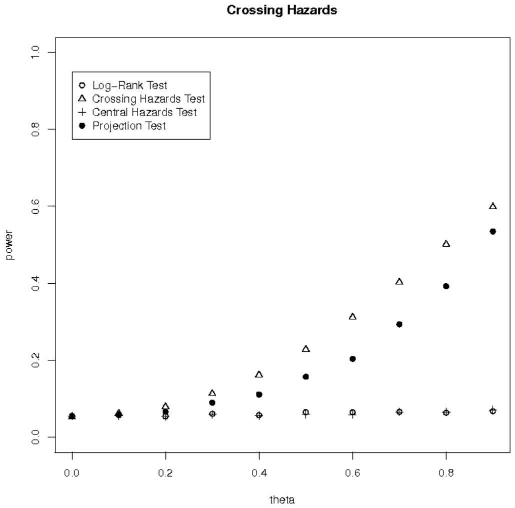
for

- $n_1 = n_2 = 50$, $\alpha = 5\%$ and different values of ϑ
- Realizations of $F_{\vartheta,i}$ by
  von Neuman's "Acceptance-Rejection"-procedure

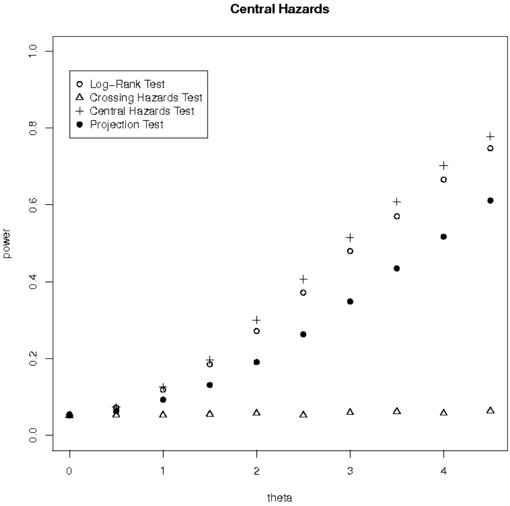
Results for proportional hazards



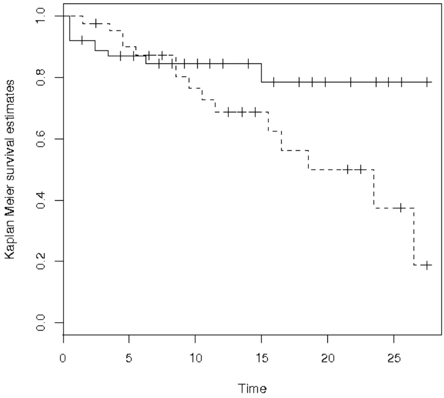
Results for crossing hazards



Resultats for central hazards

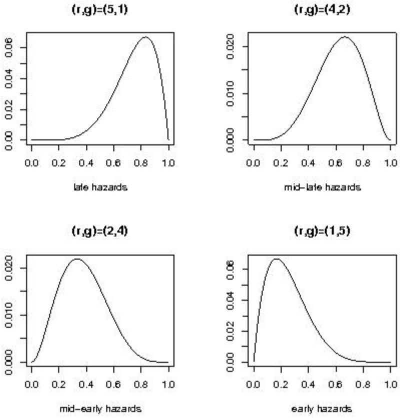


Analysis of the data example



Analysis of the data example

Choose from class of weights $\tilde{w}_{r,g}(x) = x^r(1-x)^g, \ 0 \leq x \leq 1$



Fleming/Harrington (1991)

Results

Kidney Data

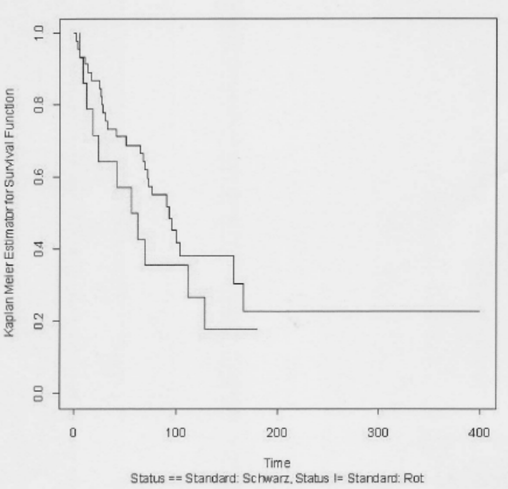
test with weight	p-value
(r,g) = (1,5)	0.0203
(r,g) = (2,4)	0.0089
(r,g) = (4,2)	0.0012
(r,g) = (5,1)	0.0084
(r,g) = (0,0)	0.0549
(logrank test)	
projection test $\phi^*_{1,n}$	0.02

One more example: Survival times of patients with tongue cancer

Differences:

Group 1: euploide cells
Group 2: aneuploide cells (i.e. cells with abnormal number of chromosomes)

Question: Can aneuploide cells be used as a prognostic parameter for survival time?



Source: Klein/Moeschberger

Results

Tongue Data

test with weight	*p*-value
(r,g) = (1,5)	0.2104414
(r,g) = (2,4)	0.2761031
(r,g) = (4,2)	0.1451610
(r,g) = (5,1)	0.05391548
(r,g) = (0,0)	0.0832526
(logrank test)	
projection test ϕ_n^*	0.09

Competing Risks Model (easiest case)

- Often: More than one event of interest!

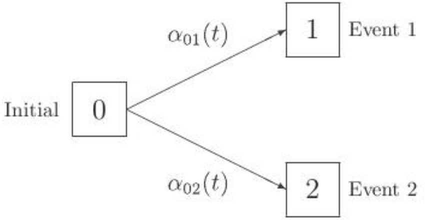
- $(X_t)_{t \geq 0}$ càdlàg, $X_t : \Omega \to \{0, 1, 2\}$.
- 0 = initial state, $P(X_0 = 0) = 1$,
- 1 and 2 absorbing states (competing events).
- $\bar{T} = \inf\{t > 0 \mid X_t \neq 0\}$ (event time)

$\Rightarrow$ $X_{\bar{T}} \in \{1, 2\}$ (event)

Regulated by cause-specific hazard intensities $\alpha_j(t)$

$$\alpha_j(t) = \alpha_{0j}(t) = \lim_{\Delta \searrow 0} \frac{P(\bar{T} \in [t, t + \Delta), X_{\bar{T}} = j \mid \bar{T} \geq t)}{\Delta}, \; j = 1, 2.$$

Cumulative Incidence Functions



- Aim: Statistical inference for CIFs

$$F_j(t) = P(\bar{T} \leq t, X_{\bar{T}} = j) = \int_0^t P(\bar{T} > u-)\alpha_j(u)du, j = 1,\ 2,$$

- Study with $1 \leq i \leq n$ independent patients,
- $(X_t^{(i)})_{t \geq 0}$ independently right-censored and left-truncated[6]
- Counting processes
  - $Y(t) = \sum_{i=1}^n Y_i(t) =$ No. under risk at $t-$
  - $N_j(t) = \sum_{i=1}^n N_{j,i}(t) =$ Observed j-events in $[0, t]$
- Aalen-Johansen estimator for the CIFs:

$$\widehat{F}_j(t) = \int_0^t \frac{\widehat{P}(\bar{T} > u-) dN_j(u)}{Y(u)}, j = 1, 2.$$

- Remark:

$$M_{j,i}(s) = N_{j,i}(s) - \int_0^s Y_i(u) \alpha_j(u) \, du$$

are local L_2-martingales!

[6] can be relaxed as explained in Andersen et al. (1993, Chapter III); only multiplicative intensity model needed

Martingale Representation

Let $t < \tau$. Under

$$\sup_{u \in [0,t]} |Y(u)/n - y(u)| \xrightarrow{p} 0$$

with $\inf_u y(u) > 0$, it follows

$$\begin{aligned} W_n(t) &= n^{1/2}\{\widehat{F}_1(t) - F_1(t)\} \\ &= n^{1/2} \sum_{i=1}^{n} \Big\{ \int_0^t \frac{S_2(u) dM_{1,i}(u)}{Y(u)} + \int_0^t \frac{F_1(u) dM_{2,i}(u)}{Y(u)} \\ &\quad - F_1(t) \int_0^t \frac{d(M_{1,i} + M_{2,i})(u)}{Y(u)} \Big\} + o_P(1) \end{aligned}$$

for $t < \tau$, where $S_2 = 1 - F_2$, see Andersen et al. (1993).

- Consequence: $W_n \xrightarrow{\mathcal{D}} U$ on $D[0, t]$ by Rebolledo

- Problems
  - Covariance function ζ of the Gaussian process U unknown
  - and lacks independent increments
- Solution: Apply 'Resampling version' of W_n

$\Rightarrow$ 1. Possibility: Wild Bootstrap:

- Concrete: $G_{j,i}$, $1 \leq i \leq n$, $1 \leq j \leq 2$, i.i.d. and $\perp\!\!\!\perp$ data with $(\mu, \sigma^2) = (0, 1)$. Approx $\mathcal{L}(W_n)$ by cond dist of

$$\widehat{W}_n(t) = n^{1/2} \sum_{i=1}^{n} \Big\{ \int_0^t \frac{\widehat{S}_2(u) G_{1,i} dN_{1,i}(u)}{Y(u)} + \int_0^t \frac{\widehat{F}_1(u) G_{2,i} dN_{2,i}(u)}{Y(u)} - \widehat{F}_1(t) \int_0^t \frac{G_{1,i} dN_{1,i}(u) + G_{2,i} dN_{2,i}(u)}{Y(u)} \Big\}.$$

- Lin's resampling technique as special case for $G_{j,i} \overset{i.i.d.}{\sim} N(0, 1)$

CCLT for $\widehat{W}_n$

Theorem (Beyersmann et al., 2013)

We have cond conv on $D[0, t]$ *given data*

$$\widehat{W}_n \xrightarrow{\mathcal{D}} U \quad \textit{in probability}$$

- Results in various inference proc based on W_n and crit. values from $\widehat{W}_n$ as
  - simultaneous CBs for F_j, see e.g. Beyersmann et al. (2013)
  - or differences of CIFs from 2 ind groups
  - tests for

    $$"\leq" \quad \text{or} \quad "="$$

    in the 2-sample case, see Bajrounaite and Klein (2007, 2008) and later...
- Q: Can we also apply other resampling techniques, e.g. the classical or even weighted bootstrap?

Answers

- A1: Application of Efron's or even the weighted bootstrap depends on the inference problem!

$\Rightarrow$ Ex1: Testing for ordered CIFs in unpaired 2-sample problem works via studentization! (details in Dobler and Pauly, 2013)

$\Rightarrow$ Ex2: Testing for equality does not work in general!
Reason: "Centering" and "too" complicated limit distribution (e.g., weighted χ^2).

- A2: Yes, e.g. the Weird Bootstrap or the more general, data-dependent Wild Bootstrap (DDWB)
Details:...

Rewrite

$$\widehat{W}_n(t) = n^{1/2} \sum_{i=1}^{n} \Big\{ G_{1,i} \int_0^t \frac{\widehat{S}_2(u) dN_{1,i}(u)}{Y(u)} + G_{2,i} \int_0^t \frac{\widehat{F}_1(u) dN_{2,i}(u)}{Y(u)}$$
$$- \widehat{F}_1(t) \int_0^t G_{1,i} \frac{dN_{1,i}(u)}{Y(u)} + G_{2,i} \frac{dN_{2,i}(u)}{Y(u)} \Big\}$$
$$= \sqrt{2n} \sum_{i=1}^{2n} G_{2n,i} Z_{2n,i}(t).$$

for i.i.d. $(G_{2n,i})_i$ with $(\mu, \sigma^2) = (0, 1)$.

- DDWB-weights $D_{2n,i}$ conditionally independent given **Z**
- DDWB vers of AJE: $\widehat{W}_n^D = \sqrt{2n} \sum_{i=1}^{2n} D_{2n,i} Z_{2n,i}$.

Theorem (Dobler and Pauly)

If the weights satisfy a cond Lindeberg ass (and some regularity conditions), then we have cond conv on $D[0, t]$

$$\widehat{W}_n^D \xrightarrow{\mathcal{D}} U \quad \textit{in probability.}$$

Examples

- Lin's Resampling technique and
- the (independent) Wild Bootstrap.
- The Weird Bootstrap of Andersen et al. (1993) corresponds to independent weights $(D_{2n,i})_{i \le 2n} = ((D_{i,j})_{j=1,2})_{i \le n}$ with[7]

$$D_{i,j} = \Big(B\big(Y(\tilde{T}_i), J(\tilde{T}_i)/Y(\tilde{T}_i)\big) - 1\Big).$$

$\Rightarrow$ Close in spirit to Wild Bootstrap with Poisson weights

- ...

[7] $\tilde{T}_i = \inf\{s \le t : N_i(s) > 0\} \wedge t$, where $N_i = N_{1,i} + N_{2,i}$

Applications: 2-sample tests for CIFs

- 2 independent groups $k = 1, 2$, each with competing risks $j = 1, 2$.
- Null hypotheses [8]:
  $H_{\leq} : \{F_1^{(1)} \leq_{st} F_1^{(2)}\}$ or $H_{=} : \{F_1^{(1)} = F_1^{(2)}\}$
- Typical test statistic: Functional of

$$W_{n_1 n_2}(t) = \sqrt{\frac{n_1 n_2}{n}} \{\widehat{F}_1^{(1)}(t) - \widehat{F}_1^{(2)}(t)\}$$

- Special case: $T_{1,n} = \int_I W_{n_1 n_2}(t) dt$ or $T_{2,n} = \int_I W^2_{n_1 n_2}(t) dt$

[8] on interval $I \subset [0, \tau)$

Applications: 2-sample tests for ordered CIFs

- We have
  $T_{1,n} = \int_I W_{n_1 n_2}(t)dt \xrightarrow{\mathcal{D}} T_1 \sim N(0, \sigma_\zeta^2)$, σ_ζ^2 unknown[9]
- Solution: Direct resampling with DDWB or
- Resampling of a studentized version with

$$T_{1,stud} := T_{1,n}/V_{1,n} \xrightarrow{\mathcal{D}} T_1/\sigma_\zeta \sim N(0,1)$$

- For the general weighted bootstrap (and also for the DDWB) it can be shown that

$$\widehat{T}^*_{1,stud} := \widehat{T}^*_{1,n}/\widehat{V}^*_{1,n} \xrightarrow{\mathcal{D}} N(0,1) \quad \text{in probability.}$$

$\Rightarrow$ Gives plenty of consistent resampling tests for $H_\leq$.
Example: Bootstrap or permutation test

[9] $\sigma_\zeta^2 = \int_I \int_I \zeta(s,t)dsdt$

Applications: 2-sample tests for equality of CIFs

- We have

$$T_{2,n} = \int_I W^2_{n_1 n_2}(t)dt \xrightarrow{\mathcal{D}} \sum_{j=1}^{\infty} \lambda_j Z_j^2$$

  with $Z_j \overset{i.i.d.}{\sim} N(0,1)$, λ_j unknown
- Solution: Direct resampling with DDWB works due to Theorem 4.
- Additional possibilities/extensions:
  - DDWB of standardized test statistic $T^{stan}_{2,n} = (T_{2,n} - \widehat{E}(T_{2,n}))/\widehat{SD}(T_{2,n})$ works as well!
  - Other approximation techniques also!

$\Rightarrow$ Gives plenty of consistent tests for $H_=$.

Remarks and Outlook

- Done: Weighted Bootstrap and DDWB for the AJE in CR

$\Rightarrow$ Generalizing the wild bootstrap and Lin's resampling technique

- In addition:
    - Comparison of the different testing procedures (For $H_{\leq}$ $\surd$)
    - DDWB for more complex Multi-State models (theory $\surd$)
    - …

Literature

Andersen, P.K., Borgan, O., Gill, R.D. & Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

Behnen, K. & Neuhaus, G. (1989). *Rank Tests with Estimated Scores and Their Application*. B.G. Teubner, Stuttgart.

Beyersmann, J., Di Termini, S. & Pauly, M. (2013). Weak convergence of the wild bootstrap for the Aalen-Johansen estimator of the cumulative incidence function of a competing risk. *Scand. J. Stat.* **40**, 387 - 402.

Brendel, M. & Janssen, A. & Mayer, C.-D. & Pauly, M. (2014). Weighted logrank permutation tests for randomly right censored life science data. *Scand. J. Stat.*, Early View.

Dobler, D. & Pauly, M. (2013). How to Bootstrap Aalen-Johansen Processes for Competing Risks? Handicaps, Solutions and Limitations. Submitted.

Ehm, W., Mammen, E. & Mueller, D.W. (1995). Power robustification of approximately linear tests. *J.A.S.A.* **90** 1025-1033.

Fleming, T.R. & Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Literature

Janssen, A. & Mayer, C.-D. (2001). Conditional studentized survival tests for randomly censored models. *Scand.J.Stat.* **28**, 283-293.

Klein, J.P. & Moeschberger, M.L. (2003). *Survival analysis: Techniques for censored and truncated data.* Springer, New York.

Neuhaus, G. (1993). Conditional rank tests for the two-sample problem under random censorship. *Ann.Stat.* **21**, 1760-1779.

Neuhaus, G. (2000). A method of constructing rank tests in Survival Analysis. *J.Stat.Plan.Inf.* **91**, 481-497.

Pauly, M. (2011). Weighted resampling of martingale difference arrays with applications, *Electronic Journal of Statistics*, **5**, 41–52.

Strasser, H. and Weber, C. (1999). *The asymptotic theory of permutation statistics. Math. Methods Stat.* **8**, No. 2, 220-250.

Thank you for your attention!