Non-parametric tests

STK4080 H16

1. Two-sample tests
2. Test for several samples
3. Stratified tests
4. Intro to general event histories

Two samples, notation

Counting processes

- No. events in group 1 with treatment A: $N_1(t)$
- No. events in group 2 with treatment B: $N_2(t)$

In each group, $j = 1, 2,$,

- No. at risk $Y_j(t)$
- hazard $\alpha_j(t)$
- Martingale $M_j(t) = N_j(t) - \int_0^t Y_j(s)\alpha_j(s)ds$

Null hypothesis

$$\mathrm{H}_0 : \alpha_1(t) = \alpha_2(t)$$

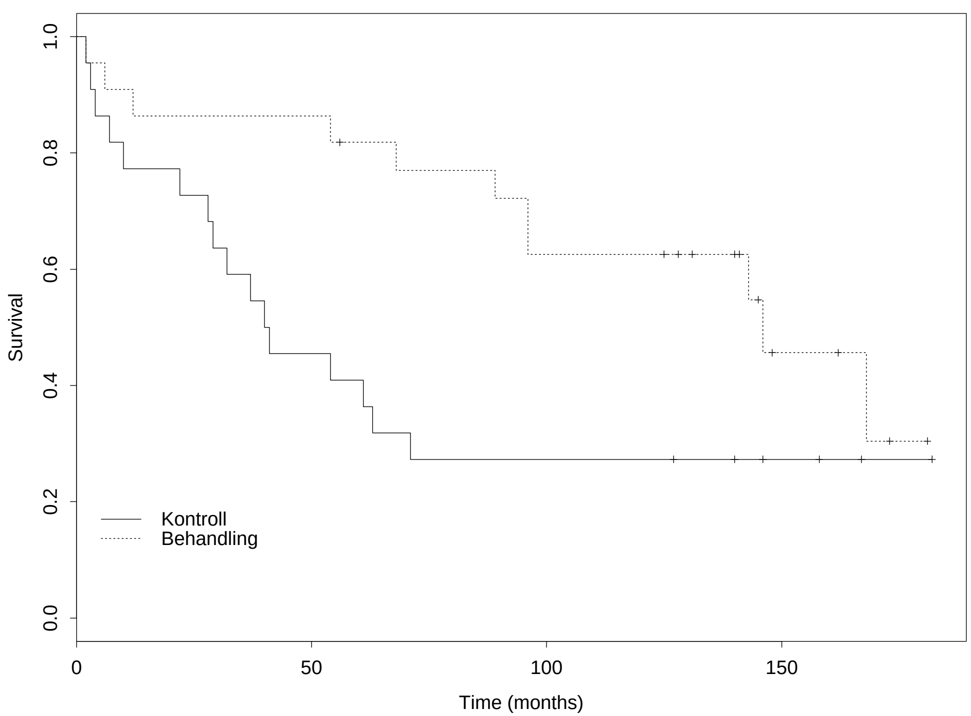
Alternative hypothesis

$$\mathrm{H}_0 : \alpha_1(t) \neq \alpha_2(t)$$

Two samples, Graphical check

Let $\hat{A}_j(t) = \int_0^t dN_j(s)/Y_j(s)$ be the Nelson-Aalen estimators of $A_j(t)$ and $\hat{S}_j(t) = \prod_0^t (1 - dN_j(s)/Y_j(s))$ the Kaplan-Meier estimator of $S_j(t)$.

Under the null we would expect $\hat{A}_1(t) \approx \hat{A}_2(t)$ and $\hat{S}_1(t) \approx \hat{S}_2(t)$



Two samples, Naive test

We could test the null by comparing $\hat{A}_1(t) - \hat{A}_2(t)$ through

$$Z = \frac{\hat{A}_1(t) - \hat{A}_2(t)}{\sqrt{\widehat{\text{Var}}(\hat{A}_1(t)) + \widehat{\text{Var}}(\hat{A}_2(t))}} \sim \text{N}(0,1) \text{ under H}_0$$

where $\widehat{\text{Var}}(\hat{A}_j(t)) = \int_0^t dN_j(s)/Y_j(s)^2$ are the variance estimators of the $\hat{A}_j(t)$.

However,

- choice of t is arbitrary
- $t = t_0 =$ largest time with ind. at risk in both groups may be a bad choice
- few individual at risk toward t_0, large uncertainty

Two-sample log-rank test:

Define weight process

$$L(t) = \frac{Y_1(t)Y_2(t)}{Y_1(t) + Y_2(t)} = \frac{Y_1(t)Y_2(t)}{Y_\bullet(t)}$$

The log-rank test statistic is then defined as

$$Z_1 = \int_0^{t_0} L(s)[d\hat{A}_1(s) - d\hat{A}_2(s)] = \int_0^{t_0} L(s)[\frac{dN_1(s)}{Y_1(s)} - \frac{dN_2(s)}{Y_2(s)}]$$

However, note Z_1 is not N(0,1) under the null, needs to be standardized.

Furthermore, with Z_2 defined similarly, we have

$$Z_1 + Z_2 = 0$$

need only consider Z_1.

Two-sample log-rank test:

With $N_\bullet(t) = N_1(t) + N_2(t)$ we have

$$\begin{aligned} Z_1 &= \int_0^{t_0} [\tfrac{Y_2(s)}{Y_\bullet(s)} dN_1(s) - \tfrac{Y_1(s)}{Y_\bullet(s)} dN_2(s)] \\ &= \int_0^{t_0} [(1 - \tfrac{Y_1(s)}{Y_\bullet(s)}) dN_1(s) - \tfrac{Y_1(s)}{Y_\bullet(s)} dN_2(s)] \\ &= N_1(t_0) - \int_0^{t_0} \tfrac{Y_1(s)}{Y_\bullet(s)} [dN_1(s) + dN_2(s)] \\ &= O_1 - E_1 \end{aligned}$$

where $O_1 = N_1(t_0)$ the observed no. events in group 1 and

$$E_1 = \int_0^{t_0} Y_1(s) \frac{dN_\bullet(s)}{Y_\bullet(s)}$$

can be interpreted as the "expected" number under the null $\alpha_1(t) = \alpha_2(t) = \alpha(t)$. Note: the common Nelson-Aalen under the null will have increments $dN_\bullet(s)/Y_\bullet(s)$.

Z_1 as martingale under $\mathbf{H}_0$

In general we may write, with $M_j(t) = N_j(t) - \int_0^t Y_j(s)\alpha_j(s)ds$,

$$Z_1 = \int_0^{t_0} L(s)[\frac{dM_1(s)}{Y_1(s)} - \frac{dM_2(s)}{Y_2(s)}] + \int_0^{t_0} L(s)(\alpha_1(s) - \alpha_2(s))ds$$

Thus under the null, $\alpha_1(t) = \alpha_2(t)$, we have

$$Z_1 = \int_0^{t_0} \frac{L(s)}{Y_1(s)} dM_1(s) - \int_0^{t_0} \frac{L(s)}{Y_2(s)} dM_2(s)$$

the difference between two integrals with respect to martingales.

In particular $\mathrm{E}[Z_1] = 0$ under the null.

Variance of Z_1

The two terms in Z_1 under the null as expressed on previous slide are uncorrelated, thus

$$\begin{aligned}\mathrm{Var}[Z_1] &= \mathrm{E}[\int_0^{t_0} \tfrac{Y_1(s)^2}{Y_\bullet(s)^2} Y_2(s)\alpha(s)ds] + \mathrm{E}[\int_0^{t_0} \tfrac{Y_2(s)^2}{Y\bullet(s)^2} Y_1(s)\alpha(s)ds] \\ &= \mathrm{E}[\int_0^{t_0} \tfrac{Y_1(s)Y_2(s)}{Y_\bullet(s)} \alpha(s)ds]\end{aligned}$$

which may be estimated by

$$V_{11} = \int_0^{t_0} \frac{Y_1(s)Y_2(s)}{Y_\bullet(s)^2} dN_\bullet(s)$$

The log-rank test is thus to reject $\mathrm{H}_0 : \alpha_1(t) = \alpha_2(t)$ at 5% level when $|Z_1/\sqrt{V_{11}}| > 1.96$

Why uncorrelated?

- Two independent groups

- The intensity process of $N_1(t) + N_2(t)$ equals $\Lambda_1(t) + \Lambda_2(t) = \int_0^t Y_1(s)\alpha_1(s)ds + \int_0^t Y_2(s)\alpha_2(s)ds$
- thus, $\text{Var}(M_1(t) + M_2(t)) = \text{E}[\Lambda_1(t) + \Lambda_2(t)] = \text{Var}(M_1(t)) + \text{Var}(M_1(t))$,
- thus $\text{Cov}(M_1(t), M_2(t)) = 0$

- For such uncorrelated martingales and predictable processes $K_j(t)$ we also get $\text{Var}[\int_0^t K_1(s)dM_1(s) + \int_0^t K_2(s)dM_2(s)] = \text{E}[\int_0^t (K_1^2(s)d\Lambda_1(s) + K_2^2(s)d\Lambda_2(s)]$

The standardized log-rank statistic

is given as

$$\frac{Z_1}{\sqrt{V_{11}}} = \frac{O_1 - E_1}{\sqrt{V}} \sim \mathrm{N}(0,1)$$

or as

$$\left[\frac{Z_1}{\sqrt{V_{11}}}\right]^2 = \frac{(O_1 - E_1)^2}{V_{11}} \sim \chi^2_1$$

approximately for "large" data sets.

Note that V_{11} is symmetric wrt groupe 1 and 2 and the choice of reference group is arbitrary.

Example Log-rank: Data on kidney transplantation

The data can be found in the `R`-library `KMsurv` on `Cran`

```
> library(KMsurv); data(kidtran); attach(kidtran)

> eldre<-(age>49)

> survdiff(Surv(time,delta)~eldre)

                N Observed Expected (O-E)^2/E (O-E)^2/V
eldre=FALSE 574       73    100.3      7.44      26.5
 eldre=TRUE 289       67     39.7     18.81      26.5

 Chisq= 26.5  on 1 degrees of freedom, p= 2.64e-07
```

Approximation log-rank test

Often good approximation

$$\frac{(O_2 - E_2)^2}{V_{11}} \approx \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

In general the left hand side is larger or equal to the right hand side and the approximation is close when

- Same censoring pattern in both groups
- Small (moderate) difference in mortality

When these assumptions hold we have, for some q,

$$\frac{Y_1(t)}{Y_\bullet(t)} \approx q \quad \text{og} \quad \frac{Y_2(t)}{Y_\bullet(t)} \approx 1 - q$$

for all t.

This gives

$$V_{11} = \int_0^{t_0} \frac{Y_1(t)Y_2(t)}{Y_\bullet(t)^2} dN_\bullet(t) \approx q(1-q)N_\bullet(\tau)$$

and

$$\frac{1}{V_{11}} \approx \frac{1}{q(1-q)N_\bullet(t_0)} = \frac{1}{qN_\bullet(t_0)} + \frac{1}{(1-q)N_\bullet(t_0)} \approx \frac{1}{E_1} + \frac{1}{E_2}$$

since $qN_\bullet(t_0) \approx \int_0^{t_0} \frac{Y_1(t)}{Y_\bullet(t)} dN_\bullet(t) = E_1$ and corresp. for E_2.

Thus

$$\begin{aligned} \frac{(O_1-E_1)^2}{V_{11}} &\approx \frac{(O_1-E_1)^2}{E_1} + \frac{(O_1-E_1)^2}{E_2} \\ &= \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} \end{aligned}$$

since $O_1 - E_1 = E_2 - O_2$.

A class of two-sample test:

Let $L(t)$ be a general predictable weight function and define a general two-sample test by

$$Z_1 = \int_0^{t_0} L(s)[d\hat{A}_1(s) - d\hat{A}_2(s)] = \int_0^{t_0} L(s)[\frac{dN_1(s)}{Y_1(s)} - \frac{dN_2(s)}{Y_2(s)}]$$

Again, Z_1 is not N(0,1) under the null, needs to be standardized.

Furthermore, with $L(t) = L_{1,2}(t) = L_{2,1}(t)$, that is indices $j = 1$ and 2 can be interchanged, and Z_2 defined similarly, we have

$$Z_1 + Z_2 = 0$$

need only consider Z_1.

Some weight functions

Write $L(t) = K(t)\frac{Y_1(t)Y_2(t)}{Y_\bullet(t)}$ for some function $K(t)$. Different choices of $K(s)$ give some standard variations on the log-rank test.

Choices of $K(t)$ up or down-weights early and late events:

- $K(t) = 1$ gives log-rank
- $K(t) = Y_\bullet(t)$ Gehan's generalisation of the Wilcoxon-test
- $K(t) = \hat{S}(t)$ Peto and Prentice generalization of Wilcoxon

  $(\hat{S}(t) = \prod_{s \leq t}(1 - dN_\bullet(s)/Y_\bullet(s)))$
- $K(t) = \hat{S}(t)^p(1 - \hat{S}(t))^q$ Fleming and Harrington generalization of Prentice' test.
- $K(t) = \hat{S}(t)^p$ implementered in R (and referred to as Fleming-Harrington in ABG).

Properties two-sample tests

Under the null, $\alpha_1(t) = \alpha_2(t)$ we have

$$Z_1 = \int_0^{t_0} L(s)[\frac{dM_1(s)}{Y_1(s)} - \frac{dM_2(s)}{Y_2(s)}]$$

and thus as an integral with respect to martingales and has expectation zero and variance

$$\mathrm{E}[\int_0^{t_0} L(s)^2(\frac{1}{Y_1(s)} + \frac{1}{Y_2(s)})\alpha(s)ds]$$

which may be estimated

$$V_{11} = \int_0^{t_0} \frac{L(s)^2}{Y_1(s)Y_2(s)} dN_\bullet(s)$$

Example: Kidney transpl. data

```
> eldre<-(age>49)

> survdiff(Surv(time,delta)~eldre)

              N Observed Expected (O-E)^2/E (O-E)^2/V
eldre=FALSE 574       73    100.3      7.44      26.5
 eldre=TRUE 289       67     39.7     18.81      26.5

 Chisq= 26.5  on 1 degrees of freedom, p= 2.64e-07
> survdiff(Surv(time,delta)~eldre,rho=0)

              N Observed Expected (O-E)^2/E (O-E)^2/V
eldre=FALSE 574       73    100.3      7.44      26.5
 eldre=TRUE 289       67     39.7     18.81      26.5

 Chisq= 26.5  on 1 degrees of freedom, p= 2.64e-07
```

Example, contd.

```
> survdiff(Surv(time,delta)~eldre,rho=1)

               N Observed Expected (O-E)^2/E (O-E)^2/V
eldre=FALSE 574     65.6     90.3      6.74        26
 eldre=TRUE 289     60.9     36.2     16.80        26

 Chisq= 26  on 1 degrees of freedom, p= 3.45e-07

> survdiff(Surv(time,delta)~eldre,rho=0.5)

               N Observed Expected (O-E)^2/E (O-E)^2/V
eldre=FALSE 574     69.2     95.1      7.08      26.3
 eldre=TRUE 289     63.8     37.9     17.77      26.3

 Chisq= 26.3  on 1 degrees of freedom, p= 2.98e-07
```

Test - More than two samples

$k > 2$ groups. For group h define

- $Y_h(t) =$ no. at risk at t,
- $N_h(t) =$ no. events in $[0, t]$.

Totally let $Y_\bullet(t) = \sum_{h=1}^{k} Y_h(t)$ and $N_\bullet(t) = \sum_{h=1}^{k} N_h(t)$.

Model: Hazard $\alpha_h(t)$ in group h.

Null hypothesis: $\mathrm{H}_0 : \alpha_1(t) = \alpha_2(t) = \cdots = \alpha_k(t)$.

The tests are defined from for $h = 1, 2, \ldots, k$, and predictable processes $K(t)$

$$Z_h \quad = \int_0^{t_0} K(s) dN_h(t) - \int_0^{t_0} K(s) Y_h(s) \frac{dN_\bullet(s)}{Y_\bullet(s)} = O_h - E_h$$

Test - more than two samples

Can derive estimates for $\text{Var}[Z_h]$ as

$$V_{hh} \quad = \int_0^{t_0} K(s)^2 \frac{Y_h(s)(Y_\bullet(s) - Y_h(s))}{Y_\bullet(s)^2} dN_\bullet(s)$$

and for $\text{Cov}(Z_h, Z_j)$ as

$$V_{hj} = -\int_0^{t_0} K(s)^2 \frac{Y_h(s) Y_j(s)}{Y_\bullet(s)^2} dN_\bullet(s)$$

Let V be the $(k-1)$x$(k-1)$ matrix with V_{hh} along the diagonal and V_{hj} outside for $h, j = 1, 2, \ldots, k-1$. Furthermore, let $Z = (Z_1, Z_2, \ldots, Z_{k-1})^\top$. Then under the null hypothesis

$$(Z)^\top V^{-1} Z \sim \chi^2_{k-1}$$

Ex: Kidney transpl.

```
> agegr<-trunc(age/20)
> table(agegr)
  0   1   2   3
 29 304 429 101
> survdiff(Surv(time,delta)~agegr)
Call:
survdiff(formula = Surv(time, death) ~ agegr)

          N Observed Expected (O-E)^2/E (O-E)^2/V
agegr=0  29        1     5.65      3.82      3.99
agegr=1 304       21    56.76     22.53     38.17
agegr=2 429       88    65.45      7.77     14.63
agegr=3 101       30    12.15     26.24     28.97

 Chisq= 61.2  on 3 degrees of freedom, p= 3.26e-13
```

Ex: Kidney transpl., contd.

```
> survdiff(Surv(time,delta)~agegr,rho=1)
Call:
survdiff(formula = Surv(time, death) ~ agegr, rho = 1)

            N Observed Expected (O-E)^2/E (O-E)^2/V
agegr=0  29    0.999     5.04      3.24      3.76
agegr=1 304   18.870    50.92     20.17     37.50
agegr=2 429   79.281    59.36      6.68     13.88
agegr=3 101   27.382    11.20     23.36     27.92

 Chisq= 59.3  on 3 degrees of freedom, p= 8.48e-13
```

Stratified tests

Ex. Kidney transplants: La $\alpha_{\text{BM}}(t)$ be the hazard for black men and $\alpha_{\text{BF}}(t), \alpha_{\text{WM}}(t)$ and $\alpha_{\text{WF}}(t)$ hazards for black females, white males and white females.

May be interested in testing difference between races *irrespectively* of differences between sexes, i.e.

$$\text{H}_0 : \alpha_{\text{BM}}(t) = \alpha_{\text{WM}}(t) \text{ and } \alpha_{\text{BF}}(t) = \alpha_{\text{WF}}(t)$$

We can immediately apply tests separately for men and women based on

$$O_{\text{W|M}} - E_{\text{W|M}} \text{ and } O_{\text{W|F}} - E_{\text{W|F}}$$

A combined (or stratified test) will use

$$(O_{\text{W|M}} - E_{\text{W|M}}) + (O_{\text{W|F}} - E_{\text{W|F}}).$$

Stratified tests more generally

Hazard in group $h = 1, 2, \ldots, k$ and strata $s = 1, 2, \ldots, S$ is given by $\alpha_{hs}(t)$. Would like to test, over all strata s,

$$\mathrm{H}_0 : \alpha_{1s}(t) = \alpha_{2s}(t) = \cdots = \alpha_{ks}(t).$$

With stratum spesific tests based on $Z_{hs} = (O_{h|s} - E_{h|s})$, $h = 1, 2, \ldots, k$ we may construct a stratified test based on $(k-1)$ of

$$Z_h = \sum_{s=1}^{S} Z_{hs}.$$

In R:

- Add `+strata(gender)` to model formula

Ex: Separat and stratified test on kidney transplant data

```
> survdiff(Surv(time,delta)˜race,subset=(gender==1))
          N Observed Expected (O-E)ˆ2/E (O-E)ˆ2/V
race=1 432       73     71.9    0.0168     0.097
race=2  92       14     15.1    0.0801     0.097

 Chisq= 0.1  on 1 degrees of freedom, p= 0.755

> survdiff(Surv(time,delta)˜race,subset=(gender==2))
          N Observed Expected (O-E)ˆ2/E (O-E)ˆ2/V
race=1 280       39    44.79     0.748      4.85
race=2  59       14     8.21     4.076      4.85

 Chisq= 4.8  on 1 degrees of freedom, p= 0.0277

> survdiff(Surv(time,delta)˜race+strata(gender))
          N Observed Expected (O-E)ˆ2/E (O-E)ˆ2/V
race=1 712      112    116.7     0.188      1.13
race=2 151       28     23.3     0.942      1.13

 Chisq= 1.1  on 1 degrees of freedom, p= 0.287
```

Multistate models

Will consider stochastic processes $X(t)$ that

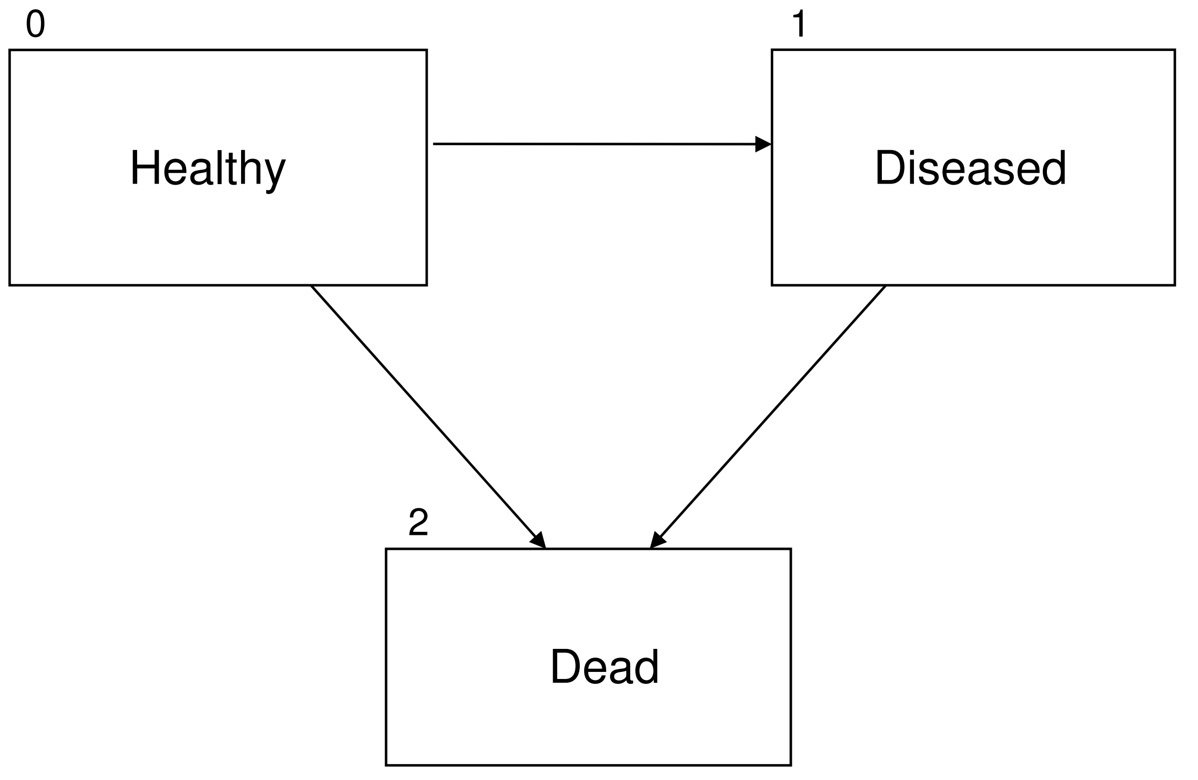
- can move between states $\{0, 1, \ldots, k\}$ with
- transition rates $\alpha_{ij}(t) = \lim_{h \downarrow 0} \frac{1}{h} \mathrm{P}(X(t+h) = j | X(t) = i)$
- under a Markov assumption, $\mathcal{F}_s$ is the history up to time s, $\mathrm{P}(X(t+h) = j | X(s) = i) = \mathrm{P}(X(t+h) = j | \mathcal{F}_s)$

We will then derive the Aalen-Johansen estimator of $\mathrm{P}_{ij}(s,t) = \mathrm{P}(X(t) = j | X(s) = i)$.

We will consider

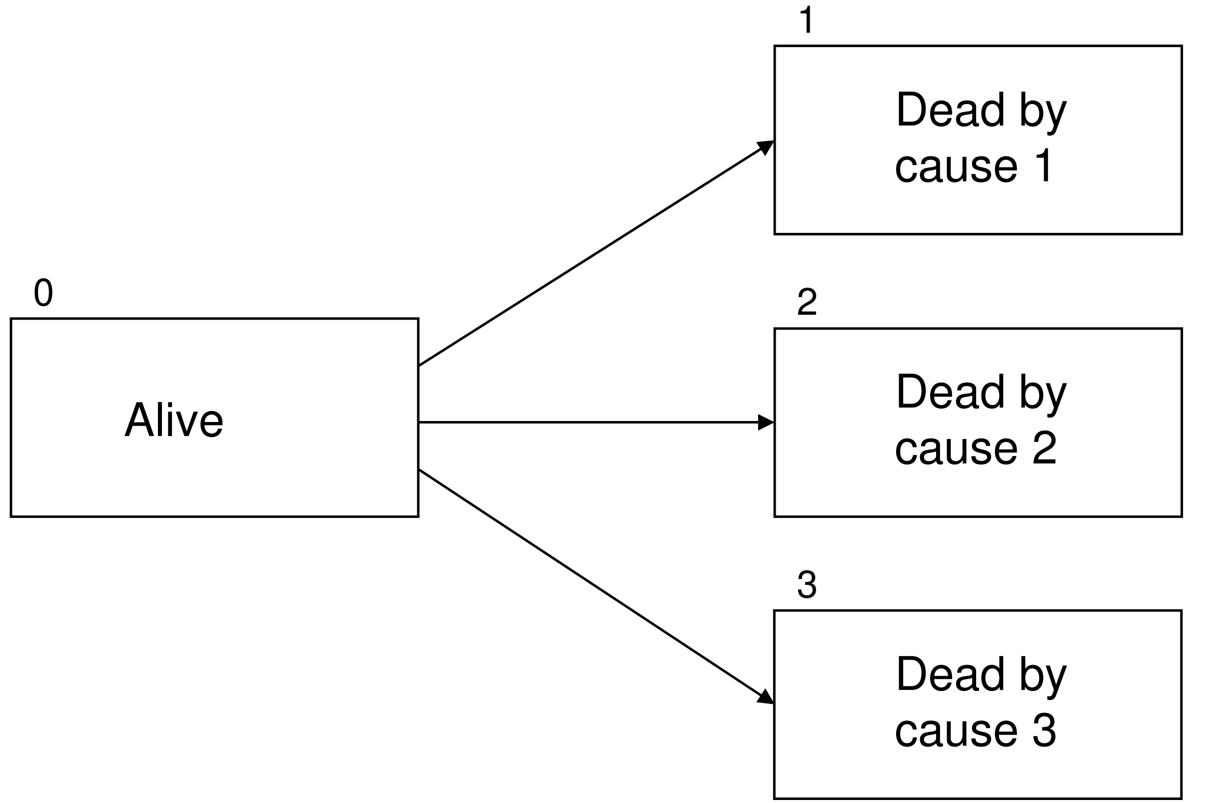
- The competing risk setting (model)
- The illness-death (e.g. healthy-illness-death)) model
- The general case

(Healthy-)Illness-Death (ID)



Hazards (intensities) $\alpha_{gh}(t)$ for transition from state g to state h at time t with $\alpha_{01}(t) > 0$, $\alpha_{02}(t) > 0$ and $\alpha_{12}(t) > 0$ (at least for some t) and all other $\alpha_{gh}(t) = 0$.

Competing risks



Denotes the state "alive" by 0 and state "death from cause h" by $h(= 1, 2, \ldots, k)$

- Transition (hazard) rate $\alpha_{0h}(t)$ of cause h
- k causes of death (can only observe one cause)

Competing risks, random variables

Postulates T_h = time until death of cause $h = 1, 2, \ldots, k$.

Assumes that the T_h are independent with hazards $\alpha_{0h}(t)$.

Note: Observes only $T = \min(T_1, T_2, \ldots, T_k)$ and $D = h$ if $T = T_h$, not the different T_h's.

However, the framework is suitable for describing the model for the process $X(t)$ on states $\{0, 1, \ldots, k\}$ with transition rates $\alpha_{0h}(t)$ from the "alive" state 0 to cause of death states $h = 1, \ldots, k$.

In particular we find that

$$\begin{aligned} \mathrm{P}_{00}(s,t) &= \mathrm{P}(X(t) = 0 | X(s) = 0) = \mathrm{P}(T > t | T > s) \\ &= \exp(-\textstyle\sum_{h=1}^{k} \int_s^t \alpha_{0h}(u) du) \end{aligned}$$

Cumulative incidence functions

The transition probabilities for competing risks are given as

$$P_{0h}(s,t) = P(X(t) = h|X(s) = 0) = \int_s^t P_{00}(s,u)\alpha_{0h}(u)du$$

with the reasoning that to be in state h at t one have

- stayed in state 0 from time s until some time u where $0 \leq s < u < t$
- and then moved to state h at time u.
- After u the process has to stay in h.

We refer to $P_{0h}(s,t)$ as **cumulative incidence functions**.

Necessarily $P_{00}(s,t) + P_{01}(s,t) + \cdots + P_{0k}(s,t) = 1$.

Competing risks, censoring

The process $X(t)$ is observed up to some censoring time C.

Observation is summarized by $\tilde{T} = \min(T, C)$, where T is the event time of $X(t)$ without censoring, and $D = h$ if $X(\tilde{T}) = h$. In particular $D = 0$ if $\tilde{T} = C$.

With n independent individual processes $X_i(t)$ we observe $(\tilde{T}_i, D_i)$.

We get the following counting process framework:

- $Y_0(t) = \#\{\tilde{T}_i \geq t)\}$ = no. still at risk
- $N_{0h}(t) = \#\{\tilde{T}_i \leq t, D_i = h)\}$ = counts deaths of cause h
- $N(t) = \sum_{h=1}^{k} N_{0h}(t)$ total no. of deaths before t

Intensity processes and martingales

Assume independent censoring. Then we have the following intensity processes for the counting processes:

- Intensity process of $N_{0h}(t)$ becomes $\lambda_{0h}(t) = Y_0(t)\alpha_{0h}(t)$
- Intensity process of $N_0(t)$ becomes $\lambda_0(t) = Y_0(t)\sum_{h=1}^{k}\alpha_{0h}(t) = Y_0(t)\alpha_0(t)$

We then obtain martingales

- $M_{0h}(t) = N_{0h}(t) - \Lambda_{0h}(t) = N_{0h}(t) - \int_0^t Y_0(s)\alpha_{0h}(s)$
- $M_0(t) = N_0(t) - \Lambda_0(t) = N_0(t) - \int_0^t Y_0(s)\alpha_0(s)ds$

In particular the $M_{0h}(t)$ are orthogonal (uncorrelated) because $N_0(t)$ only jumps with step 1, implying that the $N_{0h}(t)$ will not jump at the same time t.

Estimation of survival function $\mathrm{P}_{00}(s,t)$

Since $N_0(t)$ has intensity process $Y_0(t)\alpha_0(t)$ we get that an almost unbiased estimator of the survival function $\mathrm{P}_{00}(s,t) = \exp(-\int_s^t \alpha_0(u)du)$ is given by

$$\widehat{\mathrm{P}}_{00}(s,t) = \prod_{s<u\leq t}\left[1 - \frac{dN_0(u)}{Y_0(u)}\right],$$

corresponding directly to the Kaplan-Meier estimator.

The properties: expectation, variance and asymptotical normality, follows exactly in the same way as for Kaplan-Meier.

In particular $\hat{P}_{00}(s,t)$ has expectation, ($J(s) = I(Y_0(s) > 0)$),

$$\mathrm{E}[\exp(-\int_s^t J(u)\alpha_0(u)du)] = \mathrm{E}[P_{00}^{\star}(s,t)].$$

Estimate cum. incidence function

We noted
$$P_{0h}(s,t) = \int_s^t P_{00}(u)\alpha_{0h}(s,u)du,$$
thus by plug-in estimates of the cumulative incidence functions are given by
$$\hat{P}_{0h}(s,t) = \int_s^t \hat{P}_{00}(s,u-)\frac{dN_{0h}(u)}{Y_0(u)}.$$

We get

$$\mathrm{E}[\hat{P}_{0h}(s,t)] = \int_s^t \mathrm{E}[P^{\star}_{00}(s,u-)J(u)]\alpha_{0h}(u)du,$$

thus close to unbiased.

Variances are somewhat more tricky here, a variance formula is given in ABG, eq. (3.89).)