PhD course in "Advanced survival analysis"

Non-parametric hypothesis tests. Parametric models.

Per Kragh Andersen: 4 October 2021

One-sample tests

(ABGK, sect. V.1.1)

- Counting process $N(t)$
- Intensity process $\lambda(t) = \alpha(t)Y(t)$ satisfying Aalen's multiplicative intensity model.
- Usually, $N(t)$ is obtained by aggregating individual counting processes, $\alpha(t)$ is the hazard function, and $Y(t)$ is the number at risk.

We will test the null hypothesis

$$H_0 : \alpha(t) = \alpha_0(t),$$

with $\alpha_0(t)$ a *known* hazard function, e.g. a population mortality.

Idea

Compare Nelson-Aalen increments

$$\mathrm{d}\widehat{A}(s) = \frac{\mathrm{d}N(s)}{Y(s)} \quad \text{with} \quad \alpha_0(s)\mathrm{d}s.$$

Formally: introduce predictable, non-negative weights, $K(s)$, and consider the process:

$$Z(t) = \int_0^t K(s)\left(\mathrm{d}\widehat{A}(s) - \alpha_0(s)\mathrm{d}s\right).$$

Then $Z(t)$ tends to be large (small) if $\alpha(s)$ is larger (smaller) than $\alpha_0(s)$ for all $s \in [0, t]$.

Note that we always assume that $K(s) = 0$ whenever $Y(s) = 0$.

Properties

If H_0 is true then $\mathrm{d}N(s) = \alpha_0(s)Y(s)\mathrm{d}s + \mathrm{d}M(s)$ and $Z(t)$ equals the stochastic integral

$$Z(t) = \int_0^t \frac{K(s)}{Y(s)} \mathrm{d}M(s).$$

In particular, $Z(t)$ is a mean zero martingale with (under H_0)

$$\langle Z \rangle(t) = \int_0^t \left(\frac{K(s)}{Y(s)} \right)^2 \mathrm{d}\langle M \rangle(s) = \int_0^t \frac{K^2(s)}{Y(s)} \alpha_0(s) \mathrm{d}s.$$

Conditions can be found under which, by the martingale CLT, the test statistic

$$Z(t)/\sqrt{\langle Z \rangle(t)}$$

is asymptotically standard normal under H_0.

Choice of weights

Common choice is $K(t) = Y(t)$, cf. Ex. V.1.1. Then

$$Z(t) = \int_0^t Y(s) \left(\frac{\mathrm{d}N(s)}{Y(s)} - \alpha_0(s)\mathrm{d}s \right) = N(t) - E(t),$$

$$\langle Z \rangle(t) = \int_0^t \frac{Y^2(s)}{Y(s)} \alpha_0(s)\mathrm{d}s = E(t)$$

with

$$E(t) = \int_0^t Y(s)\alpha_0(s)\mathrm{d}s.$$

The *one-sample logrank* test statistic is then

$$(N(t) - E(t))/\sqrt{E(t)}$$

and $E(t)$ is the "expected number" of events in $[0, t]$ under H_0 (since then $N(t) - E(t)$ is a martingale).

Example: Fyn diabetics

For female diabetics (Ex. V.1.3) we have, for $t = 100$ years, $N(t) = 237$ and using published Danish life-tables we get $E(t) = 80.1$.

The one-sample logrank statistic takes the (highly significant) value:

$$\frac{237 - 80.1}{\sqrt{80.1}} = 17.5.$$

k-sample tests

(ABGK sec. V.2.1)

k-variate counting process: $(N_1(t), \ldots, N_k(t))$ where, typically, each component, $N_h(t)$, is obtained by aggregating individual processes.

Intensity process $(\lambda_1(t), \ldots, \lambda_k(t))$ of the form:

$$\lambda_h(t) = \alpha_h(t) Y_h(t), h = 1, \ldots, k.$$

We will test the null hypothesis:

$$H_0 : \alpha_1(t) = \cdots = \alpha_k(t) \text{ for all } t,$$

e.g., if mortality is the same in k groups.

Properties of test statistic

If H_0 is true then $N_{\cdot}(t) = \sum_{h=1}^{k} N_h(t)$ is a univariate counting process with intensity process:

$$\lambda_{\cdot}(t) = \sum_{h=1}^{k} \lambda_h(t) = \alpha(t) Y_{\cdot}(t)$$

where $\alpha(t)$ is the common value of the $\alpha_h(t)$'s under H_0.

The idea is now to compare the Nelson-Aalen increments a priori:

$$\mathrm{d}\widehat{A}_h(t) = \frac{\mathrm{d}N_h(t)}{Y_h(t)}$$

with those under H_0:

$$\mathrm{d}\widehat{A}(t) = \frac{\mathrm{d}N_{\cdot}(t)}{Y_{\cdot}(t)}.$$

Properties of test statistic

Formally, follow the approach from the one-sample situation: introduce non-negative weights, $K(s)$, and consider the processes (for $h = 1, \ldots, k$):

$$Z_h(t) = \int_0^t K(s) Y_h(s) \left(\mathrm{d}\widehat{A}_h(s) - \mathrm{d}\widehat{A}(s) \right)$$

$$= \int_0^t K(s) \mathrm{d}N_h(s) \text{ (observed (weighted))}$$

$$- \int_0^t K(s) \frac{Y_h(s)}{Y_\cdot(s)} \mathrm{d}N_\cdot(s) \text{ (expected (weighted))}.$$

Note that $\sum_{h=1}^k Z_h(t) = 0$.

We assume that $K(s) = 0$ whenever $Y_\cdot(s) = 0$.

Properties of test statistic

Under H_0, the $Z_h(t)$'s are mean zero martingales.

From their predictable (co-)variation processes, we obtain the (co-)variance estimators:

$$\sigma_{hj}(t) = \int_0^t K^2(s) \frac{Y_h(s)}{Y_\cdot(s)} \left(\delta_{hj} - \frac{Y_j(s)}{Y_\cdot(s)} \right) \mathrm{d}N_\cdot(s).$$

Introduce

$$\mathbf{Z}(t) = (Z_1(t), \ldots, Z_{k-1}(t))^\mathsf{T} \text{ and } \widehat{\mathbf{\Sigma}}(t) = (\sigma_{hj}(t))_{h,j=1}^{k-1}$$

and use the quadratic form as test statistic:

$$\mathbf{Z}(t)^\mathsf{T} \widehat{\mathbf{\Sigma}}(t)^{-1} \mathbf{Z}(t)$$

which is $\sim \chi^2_{k-1}$ under H_0 under suitable regularity conditions (Th. V.2.1).

Choices of weights

General choices:

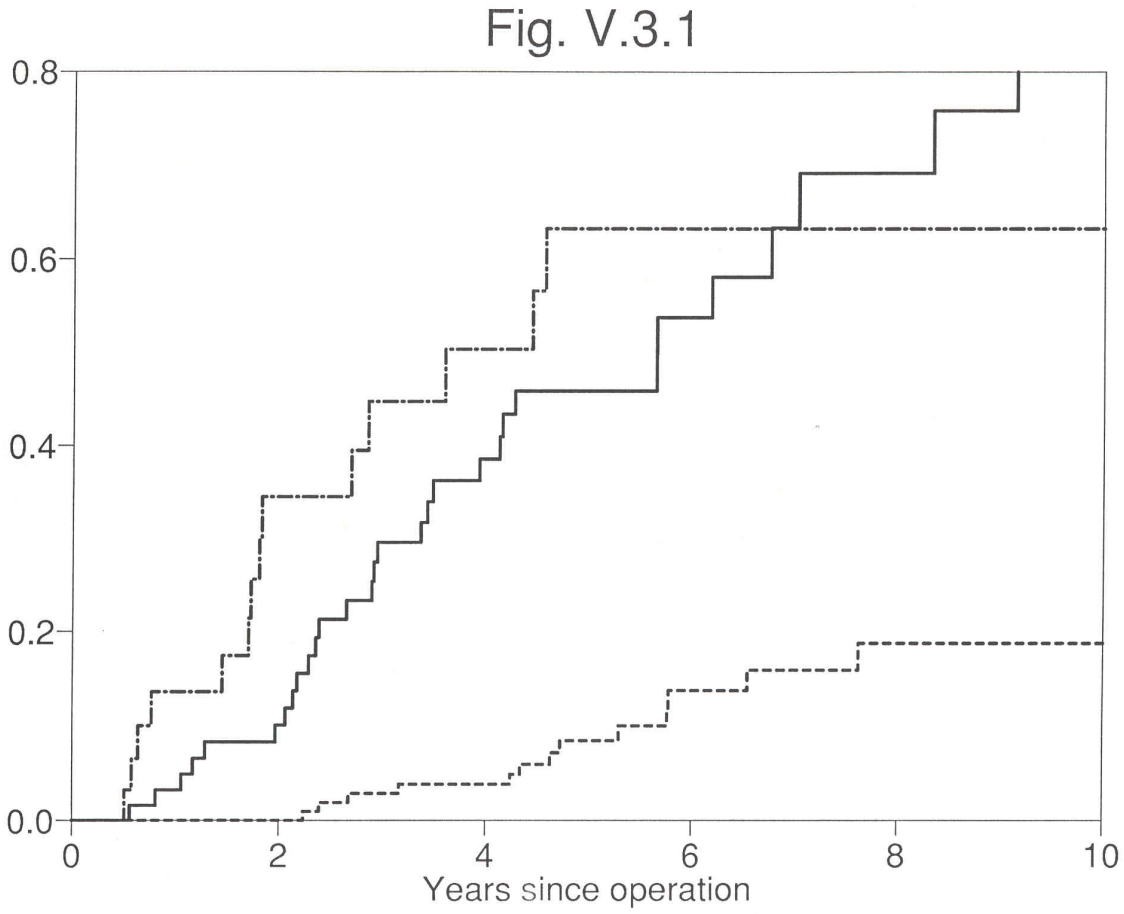
- Log-rank (Savage): $K(s) = I(Y_{\cdot}(s) > 0)$.
- Gehan-Breslow (Wilcoxon, Kruskal-Wallis): $K(s) = Y_{\cdot}(s)$.
- Tarone-Ware: $K(s) = Y_{\cdot}(s)^{\rho}$.

Survival data:

- Peto-Prentice (Wilcoxon, Kruskal-Wallis):
  $K(s) = \prod_{u<s} \left(1 - \frac{\Delta N_{\cdot}(u)}{Y_{\cdot}(u)+1}\right) \frac{Y_{\cdot}(s)}{Y_{\cdot}(s)+1}$
- Harrington-Fleming: $K(s) = I(Y_{\cdot}(s) > 0)\widehat{S}(s-)^{\rho}$.

The log-rank test has optimality properties against proportional hazards alternatives.

Nelson-Aalen estimates by tumor thickness



Categories: 0-2, 2-5, 5+ *mm*

Melanoma data

Compare mortality from the disease among tumor thickness groups: below 2*mm*, 2-5*mm*, above 5*mm*.

Log-rank test (at t =14 years): $X^2(t) = 31.6$, d.f.=2

Harrington-Fleming test (at t =14 years, $\rho = 1$): $X^2(t) = 33.9$, d.f.=2

The two-sample case

When $k = 2$ the test statistic takes the form $X^2(t) = \frac{(Z_1(t))^2}{\widehat{\sigma}_{11}(t)}$ and an alternative, standard normally distributed (under H_0), statistic is:

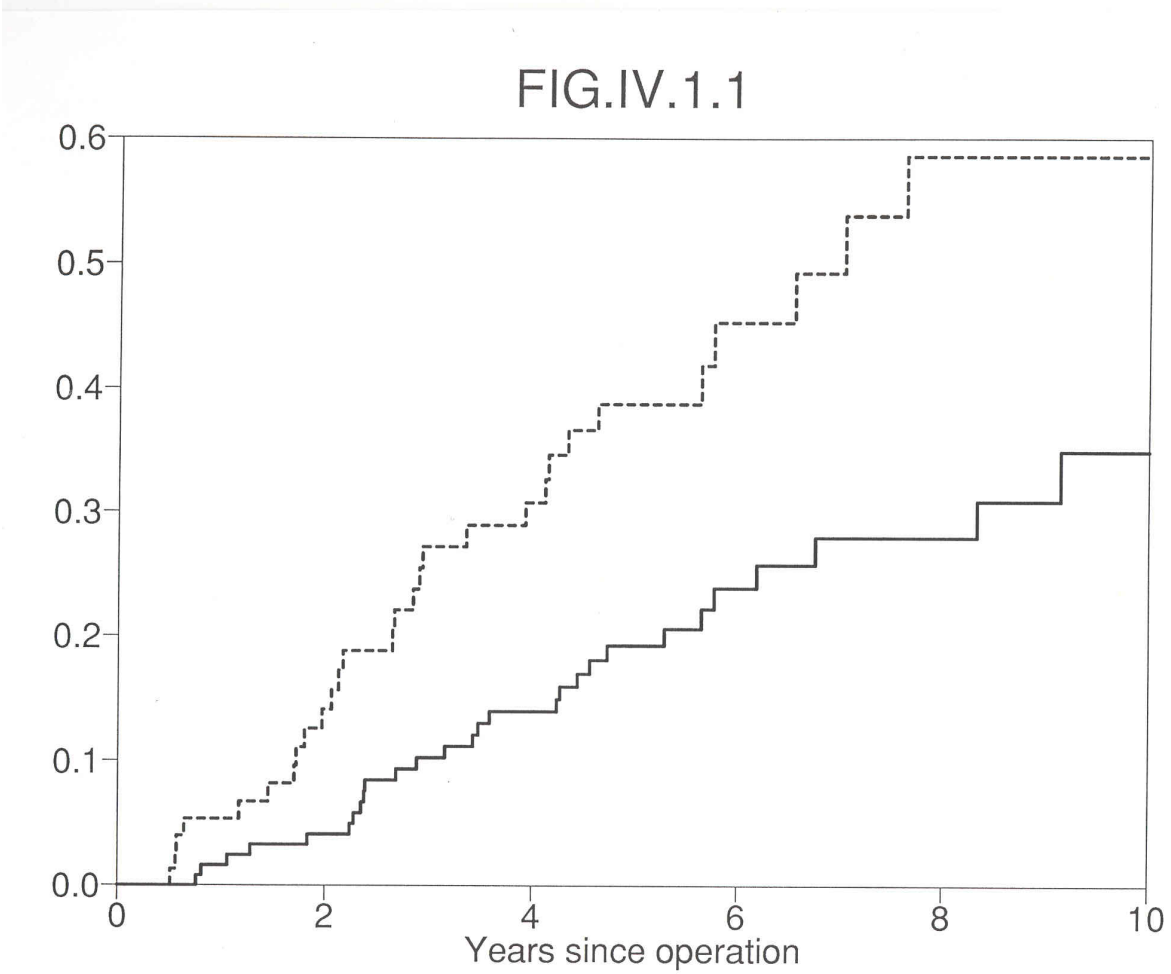
$$U(t) = \frac{Z_1(t)}{\sqrt{\widehat{\sigma}_{11}(t)}}.$$

Note that $U(t)$ has a direction.

Example: melanoma data, mortality from disease compare males and females.

- Log-rank: $U(t) = -2.54, P = 0.011,$
- Gehan-Breslow: $U(t) = -2.72, P = 0.0065,$
- Peto-Prentice: $U(t) = -2.66, P = 0.0077.$

Nelson-Aalen estimates by sex



Males: dashed, females: solid

Exercise

1. Show that the 2-sample logrank test (i.e., $K(s) = I(Y_{\cdot}(s) > 0)$) can be written in the form

$$Z_1(t) = \int_0^t \frac{Y_1(s)Y_2(s)}{Y_1(s) + Y_2(s)} (d\widehat{A}_1(s) - d\widehat{A}_2(s)).$$

2. Show that, under $H_0 : \alpha_1(t) = \alpha_2(t)$, Z_1 is a martingale.

3. Find the predictable variation process $\langle Z_1 \rangle$ and the variance estimate $\widehat{\sigma}_{11}(t)$.

Parametric models

Set-up: $N_i(t), i = 1, \ldots, n$ are counting processes with intensity processes $\lambda_i(t), i = 1, \ldots, n$ w.r.t. filtration $(\mathcal{F}_t)$.

Recall that

$$\lambda_i(t)\mathrm{d}t \approx P(\mathrm{d}N_i(t) = 1 \mid \mathcal{F}_{t-}).$$

Examples (mainly survival data) of parametric models (where $Y_i(t)$ is at-risk indicator):

- $\lambda_i(t) = \alpha Y_i(t)$, constant hazard
- $\lambda_i(t) = \alpha\rho(\alpha t)^{\rho-1} Y_i(t)$, Weibull
- $\lambda_i(t) = \theta\mu_i(t)Y_i(t)$, relative mortality ($\mu_i(t)$ population hazard)
- $\lambda_i(t) = (\gamma + \mu_i(t))Y_i(t)$, excess mortality
- $\lambda_i(t) = \alpha_j Y_i(t), \quad t_j < t \leq t_{j+1}$ piecewise constant hazard

Likelihood: Jacod's formula

ABGK; p.58. Consider a general parametric model

$$\lambda_i(t) = \alpha_i(t;\theta)Y_i(t),$$

where $\theta = (\theta_1, \ldots, \theta_p)$ is the parameter vector.

To derive the likelihood for θ, let $N_\cdot = \sum_i N_i, Y_\cdot = \sum_i Y_i$ and note that

$$P(\mathrm{d}N_\cdot(t) = 0 \mid \mathcal{F}_{t-}) \approx 1 - \lambda_\cdot(t;\theta)\mathrm{d}t.$$

With τ being the terminal time of observation we have

$$P(\text{data}) = \prod_{0<t\leq\tau} P(\mathrm{d}N_1(t), \ldots, \mathrm{d}N_n(t) \mid \mathcal{F}_{t-}) P(\text{other events at } t \mid \mathrm{d}\mathbf{N}(t), \mathcal{F}_{t-})$$

$$\propto \prod_{0<t\leq\tau} P(\mathrm{d}N_1(t), \ldots, \mathrm{d}N_n(t) \mid \mathcal{F}_{t-}).$$

Likelihood

$$L(\theta) = P(\text{data}) \propto \prod_{0<t\leq\tau} \lambda_1(t,\theta)^{\mathrm{d}N_1(t)} \cdots \lambda_n(t,\theta)^{\mathrm{d}N_n(t)} (1-\lambda_\cdot(t,\theta)\mathrm{d}t)^{1-\mathrm{d}N_\cdot(t)}$$

$$= \Big(\prod_{0<t\leq\tau} \prod_{i=1}^{n} \lambda_i(t,\theta)^{\mathrm{d}N_i(t)} \Big) \exp\Big(-\int_0^\tau \lambda_\cdot(u,\theta)\mathrm{d}u\Big).$$

Here, $\mathrm{d}N_i(t) = 1$ if N_i jumps at t and 0 otherwise.

One may show (next slide) that the scores $U_j(\theta) = \frac{\partial}{\partial\theta_j} \log L(\theta)$ are stochastic integrals w.r.t. the counting process martingales when evaluated at the true parameter:

$$\log L(\theta) = \sum_{i=1}^{n} \int_0^\tau \log \lambda_i(t,\theta) \mathrm{d}N_i(t) - \int_0^\tau \lambda_\cdot(t) \mathrm{d}t.$$

Score:

$$\frac{\partial}{\partial \theta_j} \log L(\theta) = \sum_{i=1}^{n} \int_0^\tau \frac{\lambda'_{ij}(t,\theta)}{\lambda_i(t,\theta)} \mathrm{d}N_i(t) - \int_0^\tau \lambda'_{\cdot j}(t) \mathrm{d}t.$$

Use the Doob-Meyer decomposition $\mathrm{d}N_i(t) = \lambda_i(t,\theta_0)\mathrm{d}t - \mathrm{d}M_i(t)$ to show that the score evaluated at θ_0 equals:

$$\frac{\partial}{\partial \theta_j} \log L(\theta) = \sum_{i=1}^{n} \int_0^\tau \frac{\lambda'_{ij}(t,\theta)}{\lambda_i(t,\theta)} \mathrm{d}M_i(t).$$

Thereby, using the martingale CLT, asymptotic normality of the scores may be established leading (by a standard Taylor expansion) to a proof that the MLE for $\theta = (\theta_1, \ldots, \theta_q)$ enjoys the "usual" large sample properties.

Constant hazard

When the hazard is assumed *constant*, the intensity process is

$$\lambda_i(t;\alpha) = \alpha Y_i(t).$$

If we let $R(t) = \int_0^t Y_\cdot(u)\mathrm{d}u$ be the total time at risk in $[0,t]$ then

$$L(\alpha) = \alpha^{N_\cdot(\tau)} \exp(-\alpha R(\tau))$$

and the MLE is the "occurrence/exposure rate"

$$\widehat{\alpha} = \frac{N_\cdot(\tau)}{R(\tau)}$$

and

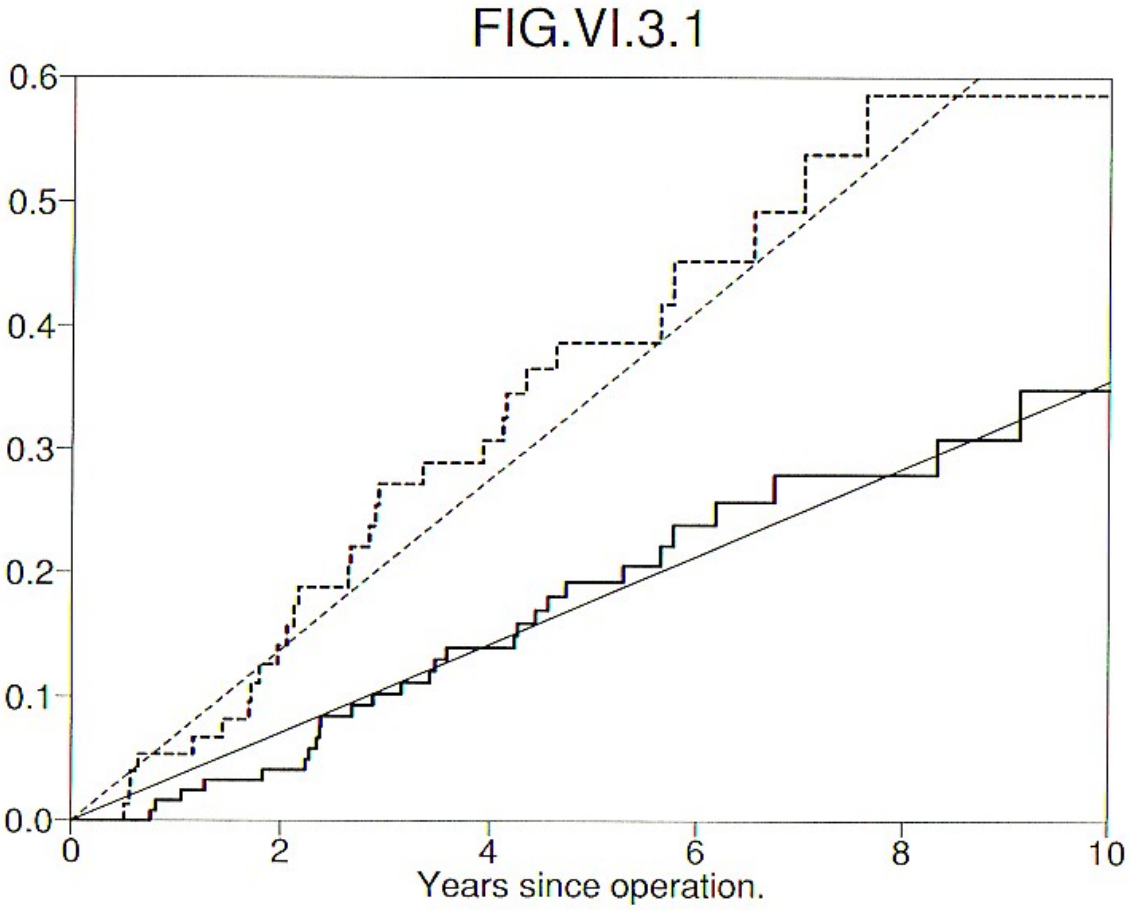
$$\widehat{\mathrm{SD}}(\widehat{\alpha}) = \big(-\frac{\partial^2}{\partial\alpha^2}\log L(\widehat{\alpha})\big)^{-\frac{1}{2}} = \sqrt{\frac{\widehat{\alpha}}{R(\tau)}}.$$

Example: melanoma data

Here we have, for females: $\widehat{\alpha}_0 = \frac{28}{786.9} = 0.0356$ per year with $\widehat{SD} = 0.0067$ while, for males, we have $\widehat{\alpha}_0 = \frac{29}{420.8} = 0.0689$ per year with $\widehat{SD} = 0.0128$. We can compare the two using the approximately χ^2_1 LR statistic which takes the value 6.15 ($P = 0.01$).

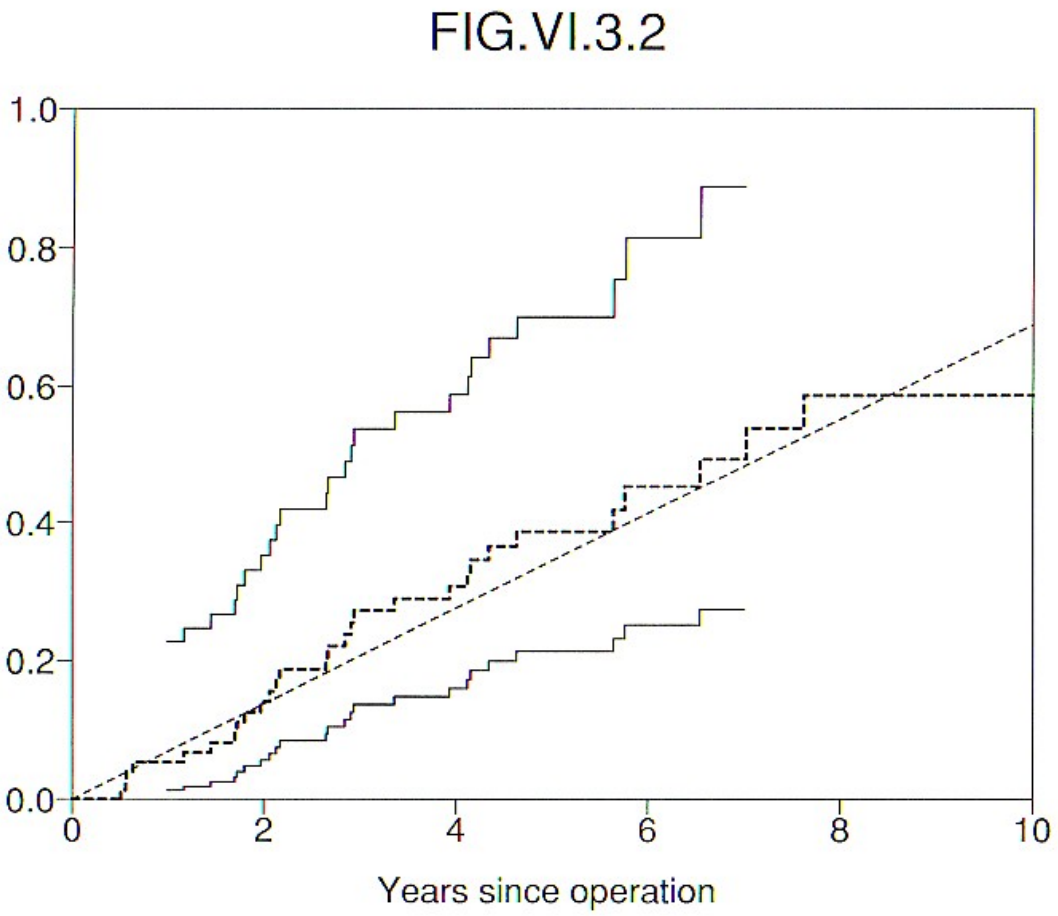
The constant hazard hypothesis may be evaluated using confidence bands for the Nelson-Aalen estimates or by embedding the model in the larger Weibull model. For the Weibull model the estimated shape parameter for females is $\widehat{\rho} = 1.197(0.202)$ while for males it is $\widehat{\rho} = 1.017(0.166)$ and for both sexes the Wald statistic $(\widehat{\rho} - 1)/\widehat{\mathrm{SD}}$ is quite insignificant.

N-Aa estimates for male and female melanoma patients



Males: dashed, females: solid

Male melanoma patients: Neslon-Aalen with bands



Piecewise constant hazards

The constant hazard model is often quite restrictive, however, a simple device offers a lot of flexibility, relax the assumption to *piecewise constant* hazards: $\lambda_i(t) = \alpha_j Y_i(t), t_j < t \leq t_{j+1}$, where $0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = +\infty$ are pre-specified interval cut-points.

Define "occurrence", $D_j = N_\cdot(t_{j+1}) - N_\cdot(t_j)$ and "exposure", $R_j = \int_{t_j}^{t_{j+1}} Y_\cdot(u)\mathrm{d}u$. Then

$$L(\alpha) = (\prod_j \alpha_j^{D_j}) \exp(-\sum_j \alpha_j R_j),$$

and the MLE simply becomes

$$\widehat{\alpha}_j = \frac{D_j}{R_j}, \quad \widehat{\mathrm{SD}}(\widehat{\alpha}_j) = \sqrt{\frac{\widehat{\alpha}_j}{R_j}}, \quad \text{as. indep.}$$

The 'Poisson' model

Since the likelihood is proportional to that obtained by formally treating the D_j as independent Poisson variates with mean

$$\alpha_j R_j,$$

this piecewise constant hazard model is some times denoted the 'Poisson' model for survival data.

Estimation may be performed using standard GLM software with $\log(R_j)$ as an 'offset'.

The model is easily extended to a log-linear regression model including other (categorical) explanatory variables.

Reducing data to tables of counts and person-years at risk (sufficiency reduction) may offer worthwhile savings of computing ressources.

Example: suicides among Danish non-manual workers

Ex. I.3.3, p.17. Four categories of non-manual workers:

1. academics
2. advanced non-academic training
3. extensive practical training
4. other non-manual workers

Suicides and person-years 1970-80 tabulated by age group and work category (excerpt):

Suicides	Age	Group 1	Group 2	Group 3	Group 4
	20-24	2	14	31	51
	25-29	10	48	46	46
Males	...				
	60-64	6	10	21	17
	20-24	0	12	29	61
	25-29	5	10	27	65
Females	...				
	60-64	2	3	7	17

Person-years	Age	Group 1	Group 2	Group 3	Group 4
	20-24	13439	67623	153241	225630
	25-29	52787			
Males	...				
	60-64	22965	37160	83207	45255
	20-24	2366	79773	199778	610034
	25-29	14047			
Females	...				
	60-64	4866	18858	32705	69418

Example: suicides among Danish non-manual workers

Age-specific suicide rates (e.g., male academics, 20-24, 2/13439=0.00015; 25-29, 10/52787=0.00019):

Suicide rates (per 10000 years) (SD)

Age	Males		Females	
20-24	1.5	(1.1)	0	(0)
25-29	1.9	(0.6)	3.6	(1.6)
30-34	4.9	(1.0)	4.9	(1.9)
35-39	4.0	(0.9)	7.6	(2.9)
40-44	5.2	(1.0)	6.4	(2.6)
45-49	4.1	(0.9)	4.9	(2.2)
50-54	5.1	(1.1)	4.6	(2.3)
55-59	5.8	(1.3)	9.0	(3.7)
60-64	2.6	(1.1)	4.1	(2.9)

Example: suicides among Danish non-manual workers

The 'Poisson'model was used in this study based on a census of the entire Danish labor force in 1970.

Categorical variables in the sub-study on suicides were: sex (s) (2), age (j) (9), work group (g) (4), marital status (m) 4, and geographical region (r) 3. Then the tables D_{sjgmr} and R_{sjgmr} are sufficient.

The full model with hazard θ_{sjgmr} was reduced to

1. $\theta_{sj}\theta_{sgmr}$
2. $\theta_{sj}\theta_{sg}\theta_m\theta_r, P = 0.14$
3. $\theta_j\theta_{sg}\theta_m\theta_r, P = 0.89$

The work group by sex interaction was highly significant, $P < 0.001$.

Estimates

Age	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
Rate per 10000 years	2.9	4.1	6.8	8.0	8.8	9.8	8.6	8.6	5.3

Work group	1: Academics	2: Advanced nonacad.	3: Extensive pract.	4: Other
Male	1 (ref.)	0.89 (.09)	0.82 (.07)	1.12 (.11)
Female	1.00 (.18)	0.55 (.06)	0.74 (.06)	0.47 (.06)

Geographical region			Marital status			
Capital	Other urban	Rural	Unmarried	Married	Widowhood	Divorced
1 (ref.)	0.76 (.04)	0.70 (.05)	1 (ref.)	0.58 (.03)	0.95 (.13)	1.46 (.10)