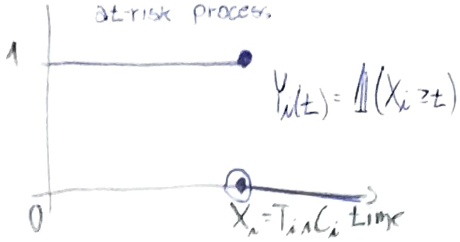
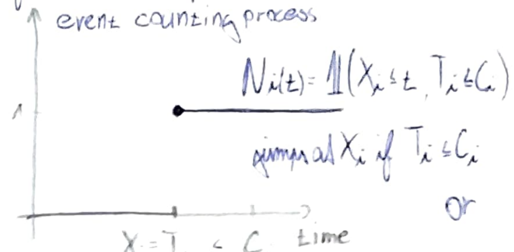
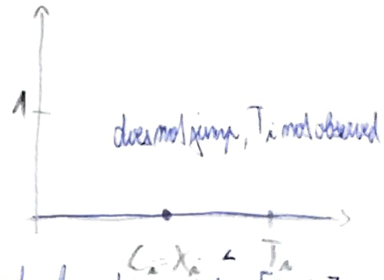
SEMIPARAMETRIC REGRESSION for the MEAN and RATE FUNCTIONS OF RECURRENT EVENTS

Cox model. Data: $(Y_i(t), N_i(t), Z_i(t))$ independent for $i = 1, \dots, n$ form filtration $\mathcal{F}(t) = \sigma\{N_i(s), Y_i(s), Z_i(s), 0 \le s \le t, i=1, \dots, n\}$

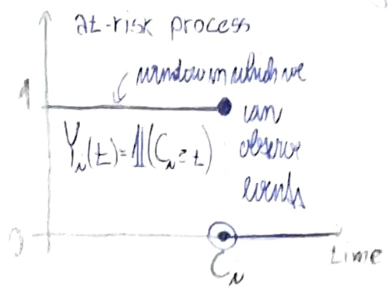


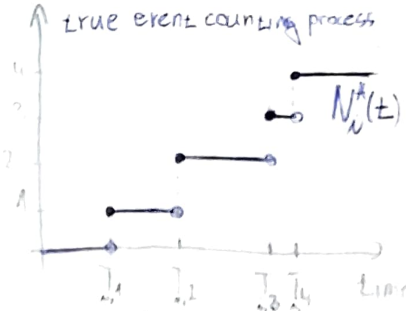


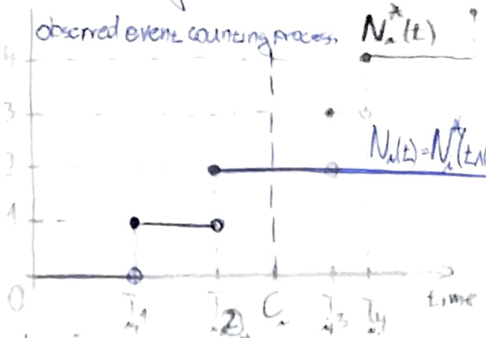
or



Extension: Events can be recurring (e.g. repeated infection, but not death). However, can be observed only in time window $[0, C_i]$.







hazard function $\lambda(t|Z) \equiv \lim_{h \searrow 0} \frac{1}{h} P[t \le T < t+h \,|\, T \ge t, Z(t)] \overset{\text{independent censoring cond.}}{=} \lim_{h \searrow 0} \frac{1}{h} P[t \le T < t+h \,|\, T \ge t, C \ge t, Z(t)]$

intensity function $\lim_{h \searrow 0} \frac{1}{h} P[N^*(t+h) - N^*(t) = 1 \,|\, \overset{\text{whole process}}{Z}] = \lambda(t|Z) + \lim_{h \searrow 0} \frac{o(h)}{h}$

$P[N(t+h) - N(t) = 1|Z] = h\,\lambda(t|Z) + o(h)$

$P[N(t+h) - N(t) \ge 2|Z] = o(h)$

→ λ models intensity of Poisson process, which is non-homogeneous if Z time-invariant $N^*(t+h) - N(t) \sim \mathrm{Pois}(\Lambda(t+h|Z) - \Lambda(t|Z))$

cumulative hazard function (rates)

mean functions

$\Lambda(t|Z) = \int_0^t \lambda(s|Z)\,ds$ and $E[N^*(t+h) - N(t) \,|\, Z] = \int_t^{t+h} \lambda(s|Z)\,ds = \Lambda(t+h|Z) - \Lambda(t|Z)$

Cox model

- (*) no heterogeneity among patients: $\lambda(t|Z) = \lambda_0(t) \exp\{\beta_0^T Z(t)\}$ — unknown baseline hazard (corresponds to (I) with $\mathrm{var}\,\xi = 0$)
- (I) patient-specific random effect ξ: $\lambda(t|Z,\xi) = \xi\, \lambda_0(t) \exp\{\beta_0^T Z(t)\}$ $\xi \ge 0$, $E\xi = 1$, $\xi \perp Z$

→ both (*) and (I) can be hidden in model given by $\mu(t|Z) = E[N^*(t)|Z(t)] \overset{\text{model}}{=} \int_0^t \exp\{\beta_0^T Z(u)\}\, d\mu_0(u)$

(only if Z of external covariates) unknown continuous

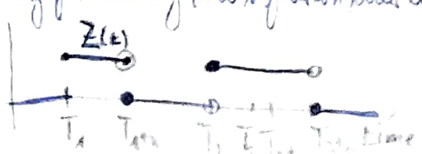
more formally: $d\mu(t|Z) = E[\Delta N^*(t)|Z(t)] \overset{\text{model}}{=} \exp\{\beta_0^T Z(t)\}\, d\mu_0(t)$

Censoring mechanism is assumed to be independent in the following sense $E[\Delta N^*(t)|Z(t), C \ge t] = E[\Delta N^*(t)|Z(t)]$ $\forall t$

Previous attempts for modelling heterogeneity on patients level considered $Z(t)$ such that it used previous history of $N_i(t)$

For example, if patient was infected in last x days, then higher intensity for observing new infection was expected.

$$\Rightarrow Z(t) = \begin{cases} 1, & \text{if infection appeared in } [t-x, t) \\ 0, & \text{otherwise} \end{cases}$$



However, inclusion of such process lead to biased estimates of effects of other covariates

Partial likelihood $L(\beta) = \prod_{i=1}^{n} \prod_{s>0} \left[\frac{Y_i(s) \cdot \exp\{\beta^T Z_i(s)\}}{\sum_{j=1}^{n} Y_j(s) \exp\{\beta^T Z_j(s)\}} \right]^{\Delta N_i(s)}$ ← now jumps at more times than just one (or zero)

constructed the same way as for classical Cox model

log-partial likelihood: $\ell(\beta) = \sum_{i=1}^{n} \sum_{s>0} \Delta N_i(s) \left[\log Y_i(s) \; \beta^T Z_i(s) - \log\left(\sum_{j=1}^{n} Y_j(s) \exp\{\beta^T Z_j(s)\} \right) \right]$

always 1 when N_i jumps

$= \sum_{i=1}^{n} \int_0^{\infty} \left[\beta^T Z_i(s) - \log\left(n S_n^{(0)}(\beta, s) \right) \right] dN_i(s)$ where $S^{(k)}_{n}(\beta,s) = \frac{1}{n} \sum_{i=1}^{n} Y_i(s) \, Z_i(t)^{\otimes k} \exp\{\beta^T Z_i(s)\}$

maximum log-partial likelihood estimator $\hat{\beta} := \underset{\beta}{\operatorname{argmax}} \, \ell(\beta)$

score functions $U_n(\beta, t) = \sum_{i=1}^{n} \int_0^{t} [Z_i(s) - \bar{Z}_n(\beta, s)] \, dN_i(s)$ where $\bar{Z}_n = \frac{S^{(1)}}{S^{(0)}}$

Let there be predetermined τ satisfying $P(C_i \geq \tau) > 0$ (observational window can still be longer than τ and $N_i(\tau)$ bounded)

Then, define $\hat{\beta}$ to be a solution to $U_n(\beta, \tau) \overset{!}{=} 0$.

Can we under regularity conditions (including $A \equiv E\left[\int_0^{\tau} [Z(t) - \bar{Z}(\beta_0, t)]^{\otimes 2} Y(t) \exp\{\beta_0^T Z(t)\} \, d\mu_0(t) \right] > 0$) prove convergence

limit of $\bar{Z}_n(\beta, t)$

$$\boxed{\frac{1}{\sqrt{n}} U_n(\beta_0, \tau) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, A) \quad ?}$$

- under ⊛ use martingale theory + martingale CLT on $M_i(t) = N_i(t) - \int_0^t Y_i(u) \exp\{\beta_0^T Z_i(u)\} \, d\mu_0(u)$ ⇒ weak convergence of $U_n(\beta,t)$
- under ⊡ M_i are no longer martingales! → cannot use martingale CLT

Individual contributions $\int_0^t [Z_i(s) - \bar{Z}_n(\beta, s)] \, dN_i(s)$ depend on other $j \in \{1, \dots, n\}$ through $\bar{Z}_n$ → cannot use Z-estimator theory directly

Combine these two approaches and prove it the same way the asymptotics of Z-estimators is done.

a) $U_n(\beta, t) = \sum_{i=1}^{n} \int_0^t [Z_i(u) - \bar{Z}_n(u)] \, d(M_i + (N_i - M_i))(u) = \dots = \sum_{i=1}^{n} \int_0^t [Z_i(u) - \bar{Z}_n(\beta, u)] \, dM_i(u)$

b) $E[dM_i(t) \mid Z_i(t)] = 0$ since $dM_i(t) = \mathbb{1}(C_i \geq t) \cdot [dN_i^*(t) - \exp\{\beta_0^T Z_i(t)\} d\mu_0(t)]$ → β_0 is identified

c) $\frac{1}{\sqrt{n}} U_n(\beta_0, t) = \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t [Z_i(u) - \bar{Z}(\beta_0, u)] \, dM_i(u)}_{\xrightarrow{\mathcal{D}} N(0, \Sigma(t,t)) \text{ iid CLT}} + \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t [\bar{Z}(\beta_0, u) - \bar{Z}_n(\beta_0, u)] \, dM_i}_{\xrightarrow{P} 0}$

$W(t)$ is gaussian process, $EW(t) = 0$

d) weak convergence of processes $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t [Z_i(u) - \bar{Z}(\beta_0, u)] \, dM_i(u) \xrightarrow{W} W(t)$ on $\mathcal{D}^p[0, \tau]$, and $\Sigma(s,t)$ covariance matrix

$$\Sigma(s, t) = E\left[\int_0^s (Z_1(u) - \bar{Z}(\beta_0, u)) \, dM_1(u) \cdot \int_0^t (Z_1(v) - \bar{Z}(\beta_0, v))^T \, dM_1(v) \right]$$ and denote $\Sigma := \Sigma(\tau, \tau)$

under ⊛ $\Sigma = A$ ⇒ $\frac{1}{\sqrt{n}} U_n(\beta_0, \tau) \xrightarrow{\mathcal{D}} N(0, A)$

under ⊡ $\Sigma \neq A$ in general ⇒ $\frac{1}{\sqrt{n}} U_n(\beta_0, \tau) \xrightarrow{\mathcal{D}} N(0, \Sigma)$

$I_n(\beta^*) \xrightarrow[n\to\infty]{P} A$

e) Taylor expansion $0 = U_n(\hat{\beta}, \tau) = U(\beta_0, \tau) + \left. \frac{\partial U(\beta, \tau)}{\partial \beta^T} \right|_{\beta = \beta^*} (\hat{\beta} - \beta_0)$ ⇒ $\sqrt{n}(\hat{\beta} - \beta_0) = -n \left[\left. \frac{\partial U(\beta, \tau)}{\partial \beta^T} \right|_{\beta = \beta^*} \right]^{-1} \frac{1}{\sqrt{n}} U(\beta_0, \tau)$

$\beta^* = \alpha \beta_0 + (1-\alpha) \hat{\beta}$, $\alpha \in [0, 1]$

Hence, $\boxed{\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, A^{-1} \Sigma A^{-1})}$

Breslow-type estimator for $\mu_0(\Delta,)$ $\quad \hat{\mu}_0(t) = \int_0^t \frac{d\bar{N}(u)}{n S^{(0)}(\hat\beta, u)}$ for $t \in [0, \tau]$

estimator for A: $\cancel{\hat{A}_n}$ $A = E\left[\int_0^\tau [Z(t) - \bar{Z}(\beta_0, t)][Z(t) - \bar{Z}(\beta_0, t)]^T Y(t) \exp\{\beta_0^T Z(t)\} d\mu_0(t)\right]$

$$\hat{A}_n = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau [Z_i(t) - \bar{Z}_n(\hat\beta, t)][Z_i(t) - \bar{Z}_n(\hat\beta, t)]^T Y_i(t) \exp\{\hat\beta_0^T Z_i(t)\} \frac{d\bar{N}(t)}{n S^{(0)}(\hat\beta, t)}$$

$$= \frac{1}{n} \int_0^\tau \sum_{i=1}^{n} [Z_i(t) Z_i^T(t) - 2 Z_i(t) \bar{Z}_n^T(\hat\beta, t) + \bar{Z}_n(\hat\beta, t) \bar{Z}_n^T(\hat\beta, t)] Y_i(t) \exp\{\hat\beta_0^T Z_i(t)\} \frac{d\bar{N}(t)}{n S^{(0)}(\hat\beta, t)}$$

do not depend on i

$$= \frac{1}{n} \int_0^\tau \left[n S_n^{(2)}(\hat\beta, t) - n \left(\frac{S_n^{(1)}(\hat\beta, t)}{S_n^{(0)}(\hat\beta, t)}\right) \left(S_n^{(1)}(\hat\beta, t)\right)^T\right] \frac{d\bar{N}(t)}{n S^{(0)}(\hat\beta, t)}$$

$$= \frac{1}{n} \int_0^\tau \left[\frac{S_n^{(2)}(\hat\beta, t)}{S^{(0)}(\hat\beta, t)} - \bar{Z}_n(\hat\beta, t) \left(\bar{Z}_n(\hat\beta, t)\right)^T\right] d\bar{N}(t)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left[\frac{S_n^{(2)}(\hat\beta, t)}{S^{(0)}(\hat\beta, t)} - \bar{Z}_n(\hat\beta, t) \left(\bar{Z}_n(\hat\beta, t)\right)^T\right] dN_i(t)$$

estimator for $M_i(t)$: $N_i(t) - \int_0^t Y_i(u) \exp\{\beta_0^T Z_i(u)\} d\mu_0(u)$

$$\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(u) \exp\{\hat\beta Z_i(u)\} \frac{d\bar{N}(t)}{n S^{(0)}(\hat\beta, t)}$$

estimator for $\Sigma = \Sigma(\tau, \tau) = E\left[\int_0^\tau (Z_i(u) - \bar{Z}(\beta_0, u)) dM_i(u) \cdot \int_0^\tau (Z_i(u) - \bar{Z}(\beta_0, u))^T dM_i(v)\right]$

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^{n} \left[\int_0^\tau (Z_i(u) - \bar{Z}_n(\hat\beta, u)) d\hat{M}_i(u) \cdot \int_0^\tau (Z_i(u) - \bar{Z}_n(\hat\beta, u))^T d\hat{M}_i(v)\right]$$

naive estimator for asymptotic variance of $\hat\beta$: $[\hat{A}_n]^{-1}$

robust estimator for asymptotic variance of $\hat\beta$: $\hat{\Gamma}_n = [\hat{A}_n]^{-1} \hat{\Sigma}_n [\hat{A}_n]^{-1}$