

Semiparametric estimation for hazard rates

Version 0.12, 31-v-2022

Nils Lid Hjort¹ and Emil Aas Stoltenberg²

¹Department of Mathematics, University of Oslo

²Norwegian Business School, Oslo

Abstract

Suppose hazard rate $\alpha(s)$ should become close to a parametric $\alpha(s, \theta)$ for growing s ; there are e.g. cases where independent arguments imply that the hazard rate should reach a constant level. Here are some ideas, perhaps to be merged to a Nils Emil paper, for estimating the cumulative hazard rate $A(t)$. These partly involve estimating the parametric part for data above a threshold only. There are at the outset as many as three weight functions involved in our method, though they may be taken to be equal: (i) $w_n(s)$, used for the construction of our $\hat{A}(t)$; (ii) $v_n(s)$, used for the weighted log-likelihood function for estimating θ ; and (iii) $r_n(s)$, used in a monitoring process, to check for adequacy of the parametric model above a threshold.

Key words: hazard rates, parametric above threshold, weighted log-likelihoods

1 Introduction

Consider survival data $(t_1, \delta_1), \dots, (t_n, \delta_n)$ over a time window $[0, \tau]$, with $\alpha(s)$ the underlying hazard rate. We wish to estimate the cumulative hazard rate $A(t) = \int_0^t \alpha(s) ds$, exploiting also potential closeness to a parametric $\alpha(s, \theta)$. There are in particular situations where $\alpha(s)$ should become close to such a parametric model for high s .

(xx more on motivation here: $A(t)$ harder to estimate nonparametrically for higher t , hence benefits if we may do that parametrically. also various further developments. xx)

Let as usual $N(t) = \sum_{i=1}^n I(t_i \leq t, \delta_i = 1)$ and $Y(s) = \sum_{i=1}^n I(t_i \geq s)$ be the counting process and at-risk process, with $A^*(t) = \int_0^t dN(s)/Y(s)$ the nonparametric Nelson–Aalen estimator. Our class of estimators takes the form

$$\hat{A}(t) = \int_0^t [\{1 - w_n(s)\} dN(s)/Y(s) + w_n(s) \alpha(s, \hat{\theta}) ds], \quad (1.1)$$

with $\hat{\theta}$ an estimator for the parameter and $w_n(s) \in [0, 1]$ a weight function. The idea is that $w_n(s)$ should or might converge to 1 for growing s . The Nelson–Aalen estimator corresponds to $w_n = 0$, whereas $w_n = 1$ leads to the fully parametric $A(t, \hat{\theta}) = \int_0^t \alpha(s, \hat{\theta}) ds$.

A special case is that of

$$w_n(s) = I(s \geq t_0) \quad \text{for some threshold } t_0 \quad (1.2)$$

above which the parametric model is assumed to hold. Then $\widehat{A}(t)$ is equal to the Nelson–Aalen estimator $A^*(t)$ up to this t_0 , and

$$\widehat{A}(t) = A^*(t_0) + \int_{t_0}^t \alpha(s, \widehat{\theta}) ds = A^*(t_0) + A(t, \widehat{\theta}) - A(t_0, \widehat{\theta}) \quad \text{for } t \geq t_0. \quad (1.3)$$

Here $\widehat{\theta}$ might be the ML estimator, or something involving its own weight function; it might e.g. be computed using data above s only. Estimating $A(t)$ leads of course to estimates also of the survival curve $S(t) = P(T \geq t)$, via $\widehat{S}(t) = \exp\{-\widehat{A}(t)\}$, etc. A slight variation, for the above case of assuming parametrics above t_0 , is

$$\widehat{S}(t) = \widehat{S}_{\text{km}}(t_0) \exp[-\{A(t, \widehat{\theta}) - A(t_0, \widehat{\theta})\}] \quad \text{for } t \geq t_0,$$

with $\widehat{S}_{\text{km}}(t) = \prod_0^t \{1 - dN/Y\}$ the Kaplan–Meier estimator.

2 Analysis

Based on the N and Y processes, let $M(t) = N(t) - \int_0^t Y(s)\alpha(s) ds$ be the associated martingale. Under natural assumptions, $Y(s)/n \rightarrow_{\text{pr}} y(s)$, and $Z_n(t) = M(t)/\sqrt{n}$ converges in distribution to a Gaussian martingale $Z(t)$ with $\text{Var } dZ(s) = y(s)\alpha(s) ds$. It is well known that

$$\sqrt{n}\{A^*(t) - A(t)\} \rightarrow_d Z^*(t) = \int_0^t \frac{dZ(s)}{y(s)},$$

another Gaussian martingale, with incremental variances $\text{Var } dZ^*(s) = \alpha(s) ds/y(s)$.

For estimating the parametric part, consider in general terms the weighted log-likelihood function

$$\ell_n(\theta) = \int_0^\tau v_n(s) \{\log \alpha(s, \theta) dN(s) - Y(s)\alpha(s, \theta) ds\}, \quad (2.1)$$

with $v_n(s)$ a previsible weight function, scaled such that there is a limit in probability function $v_n(s) \rightarrow_{\text{pr}} v(s)$. The ordinary log-likelihood function corresponds to weight function $v_n(s) = 1$. Our estimator $\widehat{\theta}$ is the maximiser of this weighted log-likelihood; see Hjort (1985, 1986).

To analyse this estimator, assume the model holds, for position θ_0 , and let $\psi(s, \theta) = \partial \log \alpha(s, \theta) / \partial \theta$. The first and second order derivatives of the $\ell_n(\theta)$ are

$$\begin{aligned} \ell'_n(\theta) &= \int_0^\tau v_n(s) \psi(s, \theta) \{dN(s) - Y(s)\alpha(s, \theta) ds\}, \\ \ell''_n(\theta) &= - \int_0^\tau v_n(s) Y(s) \alpha(s, \theta) \psi(s, \theta) \psi(s, \theta)^t + \int_0^\tau v_n(s) \psi'(s, \theta) \{dN(s) - Y(s)\alpha(s, \theta) ds\}. \end{aligned}$$

One learns (i) that

$$J_n = -(1/n) \ell''_n(\theta_0) \rightarrow_{\text{pr}} J_v = \int_0^\tau v(s) y(s) \alpha(s, \theta_0) \psi(s, \theta_0) \psi(s, \theta_0)^t ds;$$

(ii) that

$$U_n = (1/\sqrt{n}) \ell'_n(\theta_0) = \int_0^\tau v_n(s) \psi(s, \theta_0) dZ_n(s) \rightarrow_d U_v = \int_0^\tau v(s) \psi(s, \theta_0) dZ(s),$$

with variance matrix $K_v = \int_0^\tau v(s)^2 y(s) \alpha(s, \theta_0) \psi(s, \theta_0) \psi(s, \theta_0)^t ds$; and (iii) that

$$\sqrt{n}(\widehat{\theta} - \theta_0) \rightarrow_d J_v^{-1} U_v \sim N_p(0, J_v^{-1} K_v J_v^{-1}). \quad (2.2)$$

A special case is that of $v(s) = I(s \geq t_0)$, for some interval $[t_0, \tau]$, i.e. with the same weight function (1.2) used for estimation of θ as for the semiparametric cumulative hazard rate estimator (1.3).

We can already read off the limiting distribution of our semiparametric estimator $\hat{A}(t)$ of (1.1) for one particular case, namely where t_0 is decided on as a value above which the parametric model holds; with data above that value being used for θ estimation, i.e. with $v(s) = I(s \geq t_0)$ in (2.1); and with the same $w(s) = I(s \geq t_0)$ used in (1.1). We then have $\sqrt{n}\{\hat{A}(t) - A(t)\} \rightarrow_d Z^*(t)$ for $t \leq t_0$, and

$$\sqrt{n}\{\hat{A}(t) - A(t)\} \rightarrow_d Z^*(t_0) + \int_{t_0}^t \alpha(s, \theta_0) \psi(s, \theta_0)^t J_v^{-1} U_v ds$$

for $t \geq t_0$. The variance of this limit is

$$\int_0^{t_0} \frac{\alpha(s) ds}{y(s)} + B(t)^t J_v^{-1} K_v J_v^{-1} B(t) = \int_0^{t_0} \frac{\alpha(s) ds}{y(s)} + \text{Tr}\{J_v^{-1} K_v J_v^{-1} B(t) B(t)^t\}$$

for $t \geq t_0$, where $B(t) = \int_{t_0}^t \alpha(s, \theta_0) \psi(s, \theta_0) ds$. We should demonstrate that this is smaller than for the Nelson–Aalen.

Special case: $\alpha(s)$ tends to a constant θ . Here $\ell_n(\theta) = \int_0^\tau v_n(s) \{\log \theta dN(s) - Y(s)\theta ds\}$, with maximiser

$$\hat{\theta} = \int_0^\tau v_n dN / \int_0^\tau v_n Y ds. \quad (2.3)$$

Furthermore, the limit variance for $\sqrt{n}(\hat{\theta} - \theta_0)$, under model conditions, is

$$\frac{K_v}{J_v^2} = \frac{\int_0^\tau v^2 y ds / \theta_0}{(\int_0^\tau v y ds / \theta_0)^2} = \frac{\int_0^\tau v^2 y ds}{(\int_0^\tau v y ds)^2} \theta_0.$$

For the case of $v_n(s) = I(s \geq t_0)$, we have $\hat{\theta} = \int_{t_0}^\tau dN / \int_{t_0}^\tau Y ds$, which may be seen as the ML estimator based only on data above t_0 . Here the limit variance is

$$\frac{K_v}{J_v^2} = \frac{\theta_0}{\int_{t_0}^\tau y ds}.$$

Importantly, here we only need to assume that $\alpha(s)$ is constant for $t \geq t_0$. Also, the limit variance of $\sqrt{n}(\hat{A} - A)$ is

$$\int_0^{t_0} \frac{\alpha ds}{y} + (t - t_0)^2 \theta_0 / \int_{t_0}^\tau y ds.$$

(xx we include a bit more here; how much have we earned? xx)

3 Setting the threshold above which parametrics hold

So how should we decide on a threshold t_0 , above which the parametric model holds? Such problems are worked with in Clauset (2018) and in Cunen, Hjort & Nygård (2020), but in a setting with complete data. Here we may generalise the goodness-of-fit tests of Hjort (1990) to encompass weighted likelihood estimators. We need to work with $dA^*(s) - \alpha(s, \hat{\theta}) ds$, which under model conditions may be expressed as

$$\begin{aligned} dA^*(s) - dA(s) &= \{\alpha(s, \hat{\theta}) - \alpha(s, \theta_0)\} ds \\ &= dM(s)/Y(s) - \alpha(s, \theta_0) \psi(s, \theta_0)^t (\hat{\theta} - \theta_0) + O_{\text{pr}}(1/\sqrt{n}), \end{aligned}$$

which leads to

$$\sqrt{n}\{dA^*(s) - \alpha(s, \hat{\theta}) ds\} \rightarrow_d dQ(s) = dZ(s)/y(s) - \alpha(s, \theta_0)\psi(s, \theta_0)^t J_v^{-1} U_v.$$

With $r_n(s)$ our third weight function in this paper, scaled to have a limit in probability $r(s)$, we have a monitoring process

$$H_n(t) = \sqrt{n} \int_0^t r_n(s) \{dA^*(s) - \alpha(s, \hat{\theta}) ds\} \rightarrow_d H(t) = \int_0^t r(s) dQ(s). \quad (3.1)$$

Once again a default choice for our paper, tied to the notion of the parametric model holding above a threshold, is to use $r_n(s) = I(s \geq t_0)$, then test this, for the $\hat{\theta}$ computed for data above this threshold, and use a KS version for H_n .

Then check a few special cases: For weight function $r_n(s) = Y(s)/n$, we have the direct monitoring of observed $N(t)$ with the predicted $\int_0^t Y(s)\alpha(s, \hat{\theta}) ds$:

$$H_n(t) = (1/\sqrt{n})\{N(t) - \int_0^t Y(s)\alpha(s, \hat{\theta}) ds\} \rightarrow_d H(t) = \int_0^t y(s) dQ(s).$$

We then check all of this when the parametric model is simply the constant hazard θ .

All of this leads to a KS type test for the hypothesis that $\alpha(s) = \alpha(s, \theta)$ above a threshold, or on an interval.

Can also use a more flexible weight function $w_n(s)$ in (1.1). We may attempt to choose $w_n(s)$ to minimise an estimated mse, or mise; this again is in the spirit of FIC and smoothed FIC, see Jullum & Hjort (2017, 2019).

(xx ranting on a bit; to be edited. xx) Nils 1990 explains how the choice $r_n(s) = \{Y(s)/n\}\psi(s, \hat{\theta})$ leads to simplifications. For the one-dimensional case the limit is a time-transformed Brownian bridge, making subsequent testing simple. here we need such things for $[t_0, \tau]$. let's check versions of this. with $r_n = Y/n$, for the constant hazard case, we have

$$\begin{aligned} y(s) dQ(s) &= dZ(s) - y(s) J_v^{-1} \int_0^\tau v dZ/\theta_0 \\ &= dZ(s) - y(s) \frac{1}{\int_0^\tau v y ds} \int_0^\tau v dZ. \end{aligned}$$

multiplying with our weight function $v_n(s) = I(s \geq t_0)$, consider the monitoring process

$$H_n(t) = (1/\sqrt{n}) \left\{ \int_{t_0}^t dN(s) - Y(s) \hat{\theta} ds \right\} = (1/\sqrt{n}) \left\{ N(t) - N(t_0) - \hat{\theta} \int_{t_0}^t Y(s) ds \right\}$$

for $t \geq t_0$, using the same weight function for estimation of θ , i.e. using $\hat{\theta} = \int_{t_0}^\tau dN(s) / \int_{t_0}^\tau Y(s) ds$. This H_n starts and ends at zero, and converges in distribution to

$$H(t) = Z(t) - Z(t_0) - \frac{\int_{t_0}^t y(s) ds}{\int_{t_0}^\tau y(s) ds} \{Z(\tau) - Z(t_0)\} \quad \text{for } t \in [t_0, \tau].$$

Here $Z(t) - Z(t_0)$ has variance $\theta_0 \int_{t_0}^t y(s) ds$.

(xx still ranting, and still to be edited. the point is to reach a simple clear $R_n(t_0)$ which is distributed as $\max |W^0(s)|$ under constant rate. xx) let $a(t) = \int_{t_0}^t y(s) ds$. we have $Z(t) - Z(t_0) = W(\theta_0 a(t))$, with time clock running from zero to $\sigma^2 = \theta_0 a(\tau)$. Also,

$$H(t) =_d \sigma [W^*(t) - \{a(t)/a(\tau)\} W^*(1)].$$

this is equal in distribution to σW^* , where time is running from 0 to 1 for W^* . A natural monitoring quantity is therefore

$$R_n(t_0) = \max_{t \geq t_0} \frac{|H_n(t)|}{\hat{\sigma}} = (1/\sqrt{n}) \max_{t \geq t_0} \left| N(t) - N(t_0) - \hat{\theta} \int_{t_0}^t Y(s) ds \right| / \hat{\sigma}. \quad (3.2)$$

Note that also $\hat{\theta}$ depends on t_0 here, as does $\hat{\sigma}^2 = \hat{\theta} \int_{t_0}^{\tau} (Y/n) ds$. The limit distribution of $R_n(t_0)$ is equal to that of $\max |W^0(s)|$ under the hypothesis of a constant rate above t_0 . We select the threshold t_0 such that $R_n(t_0)$ is inside say the 0.90 range under this hypothesis.

4 Application to the Correlates of War dataset

Point to Richardson (1960), Clauset (2018), Cunen, Hjort & Nygård (2020) and to Gleditsch volume. Statements about $P(Z \geq z) \doteq c/z^\theta$ for large z translate to constant hazard for high t for $T = \log Z$.

(xx nils does this. compute and plot $A^*(t)$ and $\hat{A}(t)$, for data $t_i = \log z_i/z_0$, with $z_0 = 1000$. hope to see that the monitoring quantity says ‘ok, data now follow and exponential, from here and up’, for a good t_0 . xx)

The battle death distribution will have a power-law tail, of the form $P(Z \geq z) \doteq c/z^\theta$ for z above a suitably high threshold z_0 . This is equivalent to $\log Z$ having an exponential tail. With $T = \log(Z/1000)$ we should have $P(T \geq t | T \geq t_0) \doteq \exp\{-\theta(t - t_0)\}$ for $t \geq t_0$. Clauset (2018) finds $z_0 = 7061$, which here translates to $t_0 = \log(z_0/1000) = 1.955$. We there employ method (1.1), in its form (1.3), to produce

$$\hat{A}(t) = A^*(t_0) + \hat{\theta}(t_0)(t - t_0) \quad \text{for } t \geq t_0$$

with

$$\hat{\theta}(t_0) = \frac{N(\tau) - N(t_0)}{\int_{t_0}^{\tau} Y(s) ds} = 0.525.$$

See Figure 4.1.

These construction also lend themselves to estimating the hazard rate $\alpha(s)$ itself in a semi-parametric fashion. We perform a kernel smoothing of the nonparametric Nelson–Aalen estimator to the left of t_0 and use the parametric $\alpha(s, \hat{\theta}(t_0))$ to the right of this threshold:

$$\hat{\alpha}(s) = \begin{cases} \int K_b(s - t) dN(t)/Y(t) & \text{if } s \leq t_0, \\ \hat{\theta} & \text{if } s > t_0. \end{cases}$$

Here $K_b(u) = b^{-1}K(b^{-1}u)$ for a kernel function K , like the standard normal, with b a bandwidth. This is illustrated for the CoW data in Figure 4.2, presenting this semiparametric estimate $\hat{\alpha}(s)$ for the hazard rate of $T = \log(Z/1000)$. This may of course be transformed back to the $W = Z/1000$ scale. Here we have $P(W \geq w) = P(T \log w) = \exp\{-A(\log w)\}$, in terms of the cumulative hazard rate function A for T . Hence $B(w) = A(\log w)$ is the cumulative hazard rate for W , with hazard rate function $\alpha(\log w)/w$ on that scale.

(xx nils: will then use the monitoring system to find our own $t_0 = \log(z_0/1000)$, perhaps with a z_0 close to Clauset’s 7061. xx)

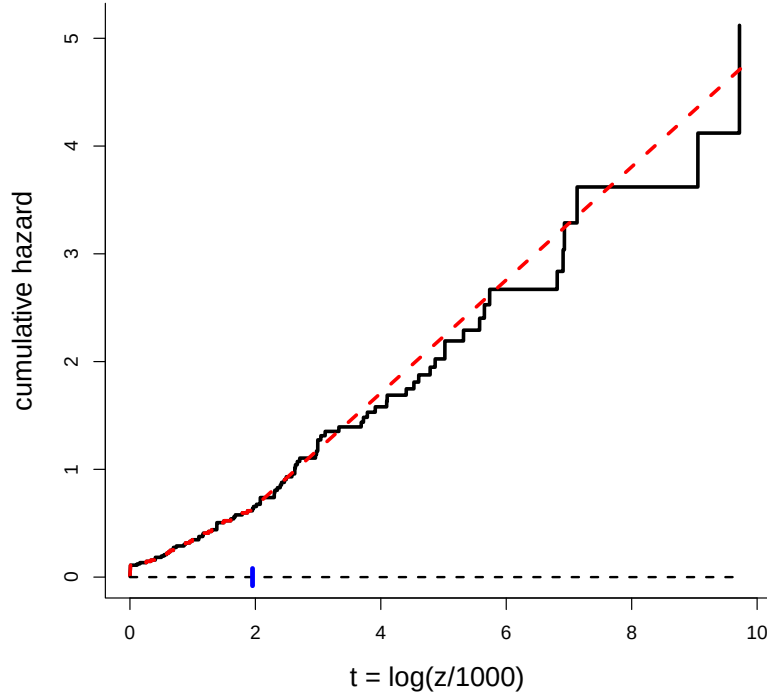


Figure 4.1: Nonparametric Nelson–Aalen estimate (black, full) and semiparametric estimate (red, dashed), for the $t_i = \log(z_i/1000)$ data of the CoW battle deaths data; the two are equal for $t \leq t_0 = 1.955$.

5 More math

Here we derive certain results, limit distributions, etc., and perhaps address a few special cases. Check Hjort (1990).

For our (1.1) estimator, consider its increment

$$\begin{aligned} d\hat{A}(s) &= \{1 - w_n(s)\}dN(s)/Y(s) + w_n(s)\alpha(s, \hat{\theta}) ds \\ &\doteq_d (1 - w)dA(s) + w\alpha(s, \theta_0) ds \\ &\quad + (1/\sqrt{n})\{(1 - w)dZ(s)/y(s) + w\alpha(s, \theta_0)\psi(s, \theta_0)^t J_v^{-1} U_v\}, \end{aligned}$$

writing for simplicity w for $w(s)$. Here θ_0 is the least false parameter associated with the weighted log-likelihood function (2.1). The implied approximate bias is

$$E d\hat{A}(s) - \alpha(s) ds \doteq w\{\alpha(s, \theta_0) - \alpha(s)\},$$

and the approximate variance is $1/n$ times

$$\begin{aligned} &(1 - w)^2 \alpha(s)/y(s) + w^2 \alpha(s, \theta_0)^2 \psi(s, \theta_0)^t J_v^{-1} K_v J_v^{-1} \psi(s, \theta_0) \\ &\quad + 2w(1 - w) \frac{1}{y(s)} \alpha(s, \theta_0) \psi(s, \theta_0)^t J_v^{-1} v(s) \psi(s, \theta_0) y(s) \alpha(s) ds. \end{aligned}$$

(xx check and calibrate with emil note. can we do jullum hjort FIC? or AFIC? xx)

(xx if we select threshold t_0 as above, smallest threshold such that data follows parametrics above that value, what properties for $\hat{t}_0$ have? how does this interact with the rest? xx)

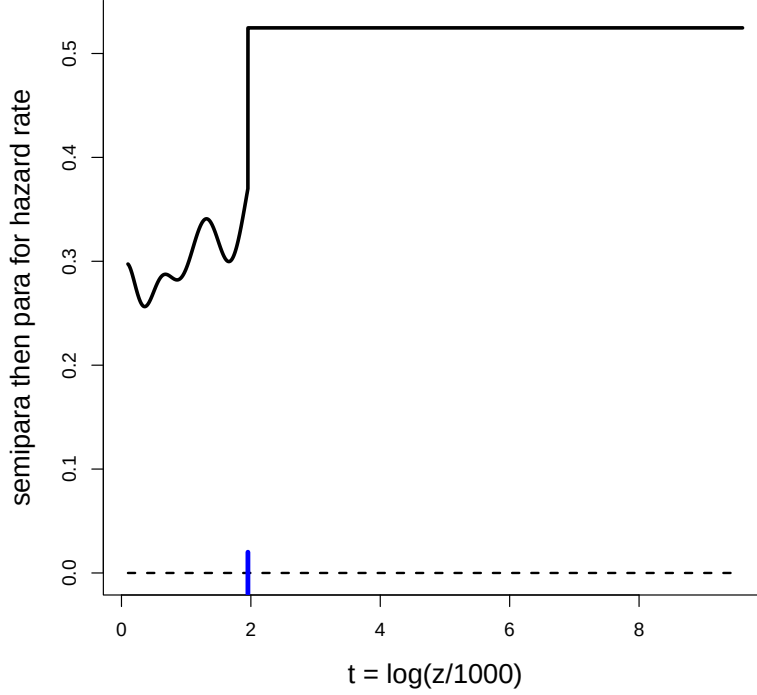


Figure 4.2: Nonparametric kernel estimate of $\alpha(s)$ for $[0, t_0]$, followed by the constant $\hat{\theta}$ after t_0 , with $t_0 = \log(z_0/1000) = 1.955$.

6 More math: the $\delta/\sqrt{n}$ setup

(xx nils rant; not polished. xx) Let's see what happens in the $O(1/\sqrt{n})$ type framework of Claeskens & Hjort (2003, 2008). Assume data are generated from the model

$$\begin{aligned}\alpha_{\text{true}}(s) &= \alpha(s, \theta_0, \gamma_0 + \delta/\sqrt{n}) = \alpha(s, \theta_0) + \beta(s)/\sqrt{n} + o(1/\sqrt{n}) \\ &= \alpha(s, \theta_0) \{1 + V(s, \theta_0)^t \delta/\sqrt{n}\} + o(1/\sqrt{n}).\end{aligned}\tag{6.1}$$

Here $\gamma = \gamma_0 + \delta/\sqrt{n}$ has dimension q , and $\gamma = \gamma_0$ brings us back to $\alpha(s, \theta_0)$; also, to the relevant order of approximation above, we have $\beta(s) = \alpha(s, \theta_0) V(s, \theta_0)^t \delta$, with $V(s, \theta_0) = \partial \log \alpha(s, \theta_0, \gamma_0) / \partial \gamma$.

I do think $Y(s)/n \rightarrow_{\text{pr}} y(s)$, the same limit as when $\gamma = \gamma_0$.

First, let's see what happens with $Z_n(t) = M(t)/\sqrt{n}$, where we now more carefully let $M(t) = N(t) - \int_0^t Y(s) \alpha_{\text{true}}(s) ds$. We have $\langle M_n, M_n \rangle(t) = \int_0^t Y(s) \alpha_{\text{true}}(s) ds$, so $\text{Var } Z_n(t) = \text{E} \{Y(s)/n\} \alpha_{\text{true}}(s) ds \rightarrow y(s) \alpha(s, \theta_0) ds$. So $Z_n(t) = M(t)/\sqrt{n} \rightarrow_d Z(t)$, a Gaussian martingale with $\text{Var } dZ(t) = y(s) \alpha(s, \theta_0) ds$.

Second, what happens to $\hat{\theta}$, our weighted likelihood estimator. Here J_n has the same limit J_v as above, but

$$\begin{aligned}U_n &= \ell'_n(\theta_0)/\sqrt{n} = \int_0^\tau v_n(s) \psi(s, \theta_0) \{dN(s) - Y(s) \alpha(s, \theta_0) ds\} / \sqrt{n} \\ &= \int_0^\tau v_n(s) \psi(s, \theta_0) [dM(s) + Y(s) \{\alpha_{\text{true}}(s) - \alpha(s, \theta_0)\} ds] / \sqrt{n} \\ &\rightarrow_d \int_0^\tau v(s) \psi(s, \theta_0) \{dZ(s) + y(s) \beta(s) ds\}.\end{aligned}$$

We learn from this that

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d J_v^{-1}(U_v + \eta), \quad \text{where } \eta = \int_0^\tau v(s)\psi(s, \theta_0)y(s)\beta(s) \, ds.$$

It has the same sandwich variance matrix as under $\alpha(s, \theta_0)$ model conditions, but a non-zero bias, thanks to the $\beta(s)/\sqrt{n}$ term of (6.1).

Then the crux of the matter, writing for simplicity w_n for $w_n(s)$: for $D_n(s) = d\hat{A}(s) - \alpha_{\text{true}}(s) \, ds$ we find

$$\begin{aligned} D_n(s) &= (1 - w_n)dN(s)/Y(s) + w_n\alpha(s, \hat{\theta}) \, ds - \alpha_{\text{true}}(s) \, ds \\ &\doteq (1 - w_n)\{\alpha_{\text{true}}(s) \, ds + dM(s)/Y(s)\} \\ &\quad + w_n\{\alpha(s, \theta_0) + \alpha(s, \theta_0)\psi(s, \theta_0)^t(\hat{\theta} - \theta_0)\} \, ds - \alpha_{\text{true}}(s) \, ds \\ &\doteq -w_n\{\alpha_{\text{true}}(s) - \alpha(s, \theta_0)\} \, ds + (1 - w_n) \, dM(s)/Y(s) \\ &\quad + w_n\alpha(s, \theta_0)\psi(s, \theta_0)^t J_v^{-1}U_n/\sqrt{n}, \end{aligned}$$

leading in its turn to

$$\sqrt{n}D_n(s) \rightarrow_d -w\beta(s) \, ds + (1 - w) \, dZ(s)/y(s) + w\alpha(s, \theta_0)\psi(s, \theta_0)^t J_v^{-1}(U_v + \eta).$$

We have (modulo smaller details) shown

$$\begin{aligned} \sqrt{n}\{\hat{A}(t) - A_{\text{true}}(t)\} &\rightarrow_d \int_0^t \{1 - w(s)\} \, dZ(s)/y(s) \\ &\quad + \int_0^t w(s)\alpha(s, \theta_0)\psi(s, \theta_0)^t \, ds J_v^{-1}(U_v + \eta) - \int_0^t w(s)\beta(s) \, ds. \end{aligned}$$

(xx we check that $\delta = 0$, which means $\beta(s) = 0$ and $\eta = 0$, really gives the limit of $\sqrt{n}(\hat{A} - A_{\text{true}})$ unde the conditions of the parametric $\alpha(s, \theta_0)$. xx)

7 Discussion

Diskusjon av hjertets lyst.

Mention the Catharina model of Stoltenberg (2021), has the property that hazard rate $\alpha(s)$ comes close to the parametric $\alpha(s, \theta)$ for growing s . Specifically, consider a Dirichlet process $D = \{D(t) : t \geq 0\}$ with parameter $aG(\cdot, \theta)$, with $G(t, \theta) = 1 - \exp\{-A(t, \theta)\}$ and a the mass parameter. For a threshold $b \in (0, 1)$, define $T = \min\{t > 0 : D(t) \geq b\}$. This leads to a well-defined distribution, with hazard rate say $h(t, \theta, a, b)$. The first point to note here is that the distribution tends to that of T_0 , with hazard rate $\alpha(s, \theta)$, as $a \rightarrow 0$; thus the Catharina model can be seen as as two-parameter extension of any given parametric hazard rate model. More pertinently, in the present context, is that for any fixed (a, b) , the hazard rate $h(t, \theta, a, b)/\alpha(t, \theta)$ tends to 1 as t grows.

Also look in Aalen and Gjessing things, for quasi-stationarity.

For a Gamma (a, θ) , regardless of a , the hazard rate tends to the constant θ as time grows.

8 Concluding remarks

A few remarks.

Covariates, perhaps in the Cox type model $\alpha(s) \exp(x_i^t \beta)$, where $\alpha(s)$ is estimated semiparametrically.

Check that it becomes clear for readers that this is sufficiently different from Hjort (1992b), ‘semiparametric estimation of parametric hazard rates’.

Acknowledgements

We are grateful.

ToDo, as of 31-v-2022

A good deal, as this is just a start note, as of 31-v-2022.

* writing out in full how $\hat{A}(t) = \int_0^t \{(1 - w_n) dN/Y + w_n \alpha(s, \hat{\theta}) ds\}$ behaves, in terms of an underlying $\alpha(s)$, using least false KL from Hjort (1992a), etc. we need a generalisation from ML to maximum weighted likelihood. the natural class has $w_n(s)$ previsible and tending to a $w(s)$. this is partly done in Emil’s December 2021 note.

* nils does the monitoring thing more carefully, with a $R_n(t_0)$ with a clear limiting distribution, for given t_0 , under parametric assumptions. in particular, set up $R_n(t_0)$ such that its limit is that of $\max |W^0(s)|$ for the case of a one-dimensional $\alpha(s, \theta)$. then apply it to the CoW.

* we should find another dataset, perhaps with a non-constant $\alpha(s, \theta)$ reached for high s .

* with covariates; to be dealt with when things are ok for the no-covariate setup.

* choosing w_n . start with the threshold. then logistic idea. can we minimise estimated risk, over all nondecreasing functions $w_n(s)$ starting at zero and ending at one?

* more on the kernel smoothing for hazard rate estimation, with

$$\hat{\alpha}(s) = \{1 - w_n(s)\} \int K_b(s - t) dN(t)/Y(t) + w_n(s) \alpha(s, \hat{\theta}).$$

this is illustrated in Figure 4.2, for CoW data on $T = \log(Z/1000)$ scale, but may need more polish.

* the emil things for cure rates.

* a bit from the Aalen Gjessing literature for quasi-stationarity, Aalen & Gjessing (2001); Gjessing et al. (2003).

* demonstrate that it works, and works well, in simulations, which should mean better than fully nonparametric. nils does a bit of variance comparison.

* let’s check variances, for the nonparametric and parametric, regarding estimation above threshold t_0 . First, the Nelson–Aalen increment has

$$n \text{Var} \{A^*(t) - A^*(t_0)\} \doteq \int_{t_0}^t \frac{\alpha(s)}{y(s)} ds.$$

Second, in the case of a constant θ above threshold, we use $\hat{\theta} = \int_{t_0}^\tau dN(s) / \int_{t_0}^\tau Y ds$, with large-sample variance $(1/n)\theta^2 / \int_{t_0}^\tau y(s) \theta ds$. which ...

$$n \text{Var} \{A(t, \hat{\theta}) - A(t_0, \hat{\theta})\} \doteq n(t - t_0)^2 \text{Var} \hat{\theta} \doteq (t - t_0)^2 \frac{\theta}{\int_{t_0}^\tau y(s) ds}.$$

The variance ratio, for growing n , is hence

$$r = \frac{(t - t_0)^2}{\int_{t_0}^\tau y(s) ds} \frac{1}{\int_{t_0}^t 1/y(s) ds}.$$

Take the natural case with random exponential censoring, leading to $y(s) = P(T \geq s)P(C \geq s) = \exp\{-(\theta + \gamma)s\}$, say. With $c = \theta + \gamma$ we then have

$$\begin{aligned} r &= (t - t_0^2) \frac{c}{\int_{t_0}^{\tau} c \exp(-cs) \, ds} \frac{c}{\int_{t_0}^t c \exp(cs) \, ds} \\ &= c^2(t - t_0)^2 \frac{1}{\exp(-ct_0) - \exp(-c\tau)} \frac{1}{\exp(ct) - \exp(ct_0)}. \end{aligned}$$

For large τ this is

$$r = c^2(t - t_0)^2 \frac{\exp(ct_0)}{\exp(ct) - \exp(ct_0)},$$

which becomes quite small with growing $t > t_0$. (xx should do more here. also, simple asymptotics for $\hat{\theta}$ also outside model. xx)

References

- AALEN, O. O. & GJESSING, H. (2001). Understanding the shape of the hazard rate: a process point of view. *Statistical Science* **16**, 1–22.
- CLAESKENS, G. & HJORT, N. L. (2003). The focused information criterion [with discussion and a rejoinder]. *Journal of the American Statistical Association* **98**, 900–916.
- CLAESKENS, G. & HJORT, N. L. (2008). *Model Selection and Model Averaging*. Cambridge: Cambridge University Press.
- CLAUSET, A. (2018). Trends and fluctuations in the severity of interstate wars. *Science Advances* **4**, 1–9.
- CUNEN, C., HJORT, N. L. & NYGÅRD, H. M. (2020). Statistical sightings of better angels. *Journal of Peace Research* **57**, 221–234.
- GJESSING, H. K., AALEN, O. O. & HJORT, N. L. (2003). Frailty models based on Lévy processes. *Advances in Applied Probability* **35**, 532–550.
- HJORT, N. L. (1985). Discussion contribution to P. K. Andersen and Ø. Borgan’s article ‘counting process models for life history data: A review’. *Scandinavian Journal of Statistics* **12**, 97–158.
- HJORT, N. L. (1986). Bayes estimators and asymptotic efficiency in parametric counting process models. *Scandinavian Journal of Statistics* **13**, 63–85.
- HJORT, N. L. (1990). Goodness of fit tests in models for life history data based on cumulative hazard rates. *Annals of Statistics* **18**, 1221–1258.
- HJORT, N. L. (1992a). On inference in parametric survival data models. *International Statistical Review* **60**, 355–387.
- HJORT, N. L. (1992b). Semiparametric estimation of parametric hazard rates [with discussion]. In *Survival Analysis: State of the Art*, J. P. Klein & P. Goel, eds. Heidelberg: Springer, pp. 211–236.

- JULLUM, M. & HJORT, N. L. (2017). Parametric of nonparametric: The FIC approach. *Statistica Sinica* **27**, 951–981.
- JULLUM, M. & HJORT, N. L. (2019). What price semiparametric Cox regression? *Lifetime Data Analysis* **25**, 406–438.
- STOLTENBERG, A. C. (2021). A class of parametric survival models via dirichlet processes [Master thesis]. Tech. rep., Department of Mathematics, University of Oslo.