Introduction

Regression models for counting processes are often specified by modeling the *intensity*:

$$\lambda_i(t) = Y_i(t)\alpha(t|X_i).$$

Here $\lambda_i(t)$ is the intensity process for subject i, $Y_i(t) = \mathbb{I}(U_i > t)$ is the at-risk process, $\alpha(t|X_i)$ is the subject-specific hazard, and X_i is a vector of covariates for subject i. Censoring and truncation are assumed independent of the time to event. Equivalently, $\alpha(t|X_i)$ is the same hazard that would be observed in their absence. While potentially time-dependent, the covariates should be *predictable*, meaning that their value is known immediately prior to the occurrence of an event or censoring $X_i(t) = X_i(t-)$.

Relative Rate Model

In a relative rate model:

$$\alpha(t|X_i) = \alpha_0(t)r(X_i;\beta).$$

Here $\alpha_0(t)$ is an unspecified baseline hazard, $r(X_i;\beta)$ is the relative rate function for a subject with covariate vector X_i, and β is a vector of regression coefficients. $r(0;\beta)$ is normalized such that the relative rate for a subject with the null covariate vector is one:

$$r(0;\beta) = 1.$$

$r(X_i;\beta)$ is the hazard ratio, comparing a subject with covariate X_i to baseline:

$$r(X_i;\beta) = \frac{\alpha(t|X_i)}{\alpha_0(t)}.$$

2.1 Partial Likelihood

Let $\tau_1 < \cdots < \tau_K$ denote the distinct observed event times, and let $\mathcal{R}_k = \mathcal{R}(\tau_k)$ denote the *risk set*, consisting of the of those subjects who remain at risk for an event at time τ_k. Suppose subject i in fact experiences an event at time τ_k. The *partial likelihood contribution* of subject i is the probability that subject i experienced the event given that some subject in $\mathcal{R}_k$ experienced an event:

$$L_i(\beta) = \frac{\alpha(\tau_k|X_i)}{\sum_{j\in\mathcal{R}_k}\alpha(\tau_k|X_j)} = \frac{\alpha_0(\tau_k)r(X_i;\beta)}{\sum_{j\in\mathcal{R}_k}\alpha_0(\tau_k)r(X_j;\beta)} = \frac{r(X_i;\beta)}{\sum_{j\in\mathcal{R}_k}r(X_j;\beta)}.$$

The overall **partial likelihood** is:

$$L(\beta) = \prod_{k=1}^{K}\frac{r(X_k;\beta)}{\sum_{j\in\mathcal{R}_k}r(X_j;\beta)}, \tag{2.1.1}$$

where X_k is the covariate vector for the subject who experiences the event at time k.

Although suppressed for simplicity, here and subsequently, if $X_i(t)$ is time-varying, then $X_i = X_i(\tau_k)$. For instance, in the partial likelihood:

$$L(\beta) = \prod_{k=1}^{K} \frac{r\{X_i(\tau_k);\beta\}}{\sum_{j\in\mathcal{R}_k} r\{X_j(\tau_k);\beta\}}.$$

The *partial log likelihood* is:

$$\ell(\beta) = \sum_{k=1}^{K} \left[\ln r(X_k;\beta) - \ln\left\{\sum_{j\in\mathcal{R}_k} r(X_j;\beta)\right\}\right].$$

Specializing to the **Cox model**:

$$r(X_k;\beta) = \exp\big(X_k\beta\big),$$

the partial log likelihood takes the form:

$$\ell(\beta) = \sum_{k=1}^{K}\left\{X_k\beta - \ln\left(\sum_{j\in\mathcal{R}_k} e^{X_j\beta}\right)\right\}. \tag{2.1.2}$$

2.2 Partial Score

The **partial score equation** for the Cox model is:

$$\mathcal{U}_\beta = \sum_{k=1}^{K}\left\{X_k - \frac{\sum_{j\in\mathcal{R}_k} X_j e^{X_j\beta}}{\sum_{j\in\mathcal{R}_k} e^{X_j\beta}}\right\}, \tag{2.2.3}$$

which follows directly from taking the gradient of (2.1.2).

Proposition 2.2.1. The partial score equation from the Cox model is expressible as:

$$\mathcal{U}_\beta(\tau) = \sum_{i=1}^{n}\int_0^\tau \left\{X_i - \frac{\sum_{j=1}^n X_jY_j(s)e^{X_j\beta}}{\sum_{j=1}^n Y_j(s)e^{X_j\beta}}\right\} dN_i(s).$$

♦

Proof. Starting from the counting process representation:

$$\begin{aligned}
&\sum_{i=1}^{n}\int_0^\tau \left\{X_i - \frac{\sum_{j=1}^n X_jY_j(s)e^{X_j\beta}}{\sum_{j=1}^n Y_j(s)e^{X_j\beta}}\right\} dN_i(s) \\
&\quad = \sum_{i=1}^{n}\sum_{k=1}^{K}\left\{X_i - \frac{\sum_{j=1}^n X_jY_j(s)e^{X_j\beta}}{\sum_{j=1}^n Y_j(s)e^{X_j\beta}}\right\}\mathbb{I}(U_i = \tau_k, \delta_i = 1).
\end{aligned}$$

Assuming only a single individual experiences an event at each time point, only the individual whose event time is τ_k contributes to the sum over i. Therefore:

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{k=1}^{K}\left\{X_i - \frac{\sum_{j=1}^{n} X_j Y_j(s) e^{X_j\beta}}{\sum_{j=1}^{n} Y_j(s) e^{X_j\beta}}\right\}\mathbb{I}(U_i = \tau_k, \delta_i = 1) \\
&\quad = \sum_{k=1}^{K}\left\{X_k - \frac{\sum_{j=1}^{n} X_j Y_j(\tau_k) e^{X_j\beta}}{\sum_{j=1}^{n} Y_j(\tau_k) e^{X_j\beta}}\right\} = \sum_{k=1}^{K}\left\{X_k - \frac{\sum_{j\in\mathcal{R}_k} X_j e^{X_j\beta}}{\sum_{j\in\mathcal{R}_k} e^{X_j\beta}}\right\}.
\end{aligned}$$

■

Proposition 2.2.2. The partial score equation from the Cox model is expressible as:

$$\mathcal{U}_\beta(\tau) = \sum_{i=1}^{n}\int_0^\tau \left\{X_i - \frac{\sum_{j=1}^{n} X_j Y_j(s) e^{X_j\beta}}{\sum_{j=1}^{n} Y_j(s) e^{X_j\beta}}\right\} dM_i(s),$$

where $M_i(s) = N_i(s) - \int_0^s Y_i(s)\alpha_0(s)e^{X_i\beta}ds$ is the counting process martingale. ♦

Proof. Substituting $dN_i(s) = dM_i(s) + Y_i(s)\alpha_0(s)e^{X_i\beta}ds$ into the previous counting process expression for $\mathcal{U}_\beta(\tau)$:

$$\mathcal{U}_\beta(\tau) = \sum_{i=1}^{n}\int_0^\tau \left\{X_i - \frac{\sum_{j=1}^{n} X_j Y_j(s) e^{X_j\beta}}{\sum_{j=1}^{n} Y_j(s) e^{X_j\beta}}\right\}\left\{dM_i(s) + Y_i(s)\alpha_0(s)e^{X_i\beta}ds\right\}.$$

The conclusion follows if the second term that arises from distributing the differential is identically zero. To see this:

$$\begin{aligned}
&\sum_{i=1}^{n}\int_0^\tau \left\{X_i - \frac{\sum_{j=1}^{n} X_j Y_j(s) e^{X_j\beta}}{\sum_{j=1}^{n} Y_j(s) e^{X_j\beta}}\right\} Y_i(s)\alpha_0(s)e^{X_i\beta}ds \\
&\quad = \int_0^\tau \alpha_0(s)\left\{\sum_{i=1}^{n} X_i Y_i(s) e^{X_i\beta} - \frac{\sum_{j=1}^{n} X_j Y_j(s) e^{X_j\beta}}{\sum_{j=1}^{n} Y_j(s) e^{X_j\beta}}\sum_{i=1}^{n} Y_i(s) e^{X_i\beta}\right\} ds \\
&\quad = \int_0^\tau \alpha_0(s)\left\{\sum_{i=1}^{n} X_i Y_i(s) e^{X_i\beta} - \sum_{j=1}^{n} X_j Y_j(s) e^{X_j\beta}\right\} ds = 0.
\end{aligned}$$

■

2.3 Partial Information

Proposition 2.3.3. The partial information for β from the Cox model is:

$$\mathcal{I}_{\beta\beta'} = \sum_{k=1}^{K}\left\{\frac{\sum_{j\in\mathcal{R}_k} X_j^{\otimes 2} e^{X_j\beta}}{\sum_{j\in\mathcal{R}_k} e^{X_j\beta}} - \frac{\left(\sum_{j\in\mathcal{R}_k} X_j e^{X_j\beta}\right)^{\otimes 2}}{\left(\sum_{j\in\mathcal{R}_k} e^{X_j\beta}\right)^2}\right\}.$$

♦

Proof. Finding the Hessian of (2.1.2) with respect to β:

$$\mathcal{H}_{\beta\beta'} = -\sum_{k=1}^{K}\left\{\frac{\sum_{j\in\mathcal{R}_k} X_j^{\otimes 2} e^{X_j\beta}}{\sum_{j\in\mathcal{R}_k} e^{X_j\beta}} - \frac{\left(\sum_{j\in\mathcal{R}_k} X_j e^{X_j\beta}\right)^{\otimes 2}}{\left(\sum_{j\in\mathcal{R}_k} e^{X_j\beta}\right)^2}\right\}.$$

Since $\mathcal{H}_{\beta\beta'}$ does not depend on (U_i, δ_i), the expected information:

$$\mathcal{I}_{\beta\beta'} = -\mathbb{E}\big(\mathcal{H}_{\beta\beta'}\big) = -\mathcal{H}_{\beta\beta'}.$$

■

Corollary 2.3.1. The partial information for β from the Cox model is:

$$\mathcal{I}_{\beta\beta'} = \int_0^{\tau} \nu(s) dN(s),$$

where:

$$\nu(s) = \frac{\sum_{i=1}^{n} X_i^{\otimes 2} Y_i(s) e^{X_i\beta}}{\sum_{i=1}^{n} Y_i(s) e^{X_i\beta}} - \frac{\left\{\sum_{i=1}^{n} X_i Y_i(s) e^{X_i\beta}\right\}^{\otimes 2}}{\left\{\sum_{i=1}^{n} Y_i(s) e^{X_i\beta}\right\}^2}$$

and $\tau \geq \tau_K$, the last observed event. ♣

Proposition 2.3.4. The predictable variation of the partial score equation form the Cox model is:

$$\langle \mathcal{U}_\beta(\tau) \rangle = \int_0^{\tau} \nu(s) \left\{\sum_{i=1}^{n} Y_i(s) e^{X_i\beta}\right\} \alpha_0(s) ds.$$

♦

Proof. Introduce the notation:

$$S^{(0)}(s) = \sum_{i=1}^{n} Y_i(s) e^{X_i\beta}, \quad S^{(1)}(s) = \sum_{i=1}^{n} X_i Y_i(s) e^{X_i\beta}, \quad S^{(2)}(s) = \sum_{i=1}^{n} X_i^{\otimes 2} Y_i(s) e^{X_i\beta}.$$

Define the predictable process:

$$H_i(s) = X_i - \frac{S^{(1)}(s)}{S^{(0)}(s)},$$

such that the partial score of the Cox model is:

$$\mathcal{U}_\beta(\tau) = \sum_{i=1}^{n} \int_0^{\tau} H_i(s) dM_i(s).$$

Since the counting process martingales of independent subjects are orthogonal, the predictable variation of $\mathcal{U}_\beta(\tau)$ is:

$$\begin{aligned}\langle \mathcal{U}_\beta(\tau)\rangle &= \sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^\tau H_i(s)\otimes H_j(s)d\langle M_i, M_j\rangle(s)\\ &= \sum_{i=1}^{n}\int_0^\tau H_i^{\otimes 2}(s)d\langle M_i\rangle(s) = \sum_{i=1}^{n}\int_0^\tau H_i^{\otimes 2}(s)\lambda_i(s)ds,\end{aligned}$$

where $\lambda_i(s) = Y_i(s)\alpha(s|X_i) = Y_i(s)\alpha_0(s)e^{X_i\beta}$ is the intensity process for subject i. Expanding the outer product of the predictable process:

$$\langle \mathcal{U}_\beta(\tau)\rangle = \sum_{i=1}^{n}\int_0^\tau \left\{X_i^{\otimes 2} + \frac{\{S^{(1)}\}^{\otimes 2}}{\{S^{(0)}\}^2} - 2\frac{X_i\otimes S^{(1)}}{S^{(0)}}\right\}\alpha_0(s)Y_i(s)e^{X_i\beta}ds.$$

Bringing the sum inside the integral:

$$\begin{aligned}\langle \mathcal{U}_\beta(\tau)\rangle &= \int_0^\tau \left\{S^{(2)} + S^{(0)}\frac{\{S^{(1)}\}^{\otimes 2}}{\{S^{(0)}\}^2} - 2\frac{\{S^{(1)}\}^{\otimes 2}}{S^{(0)}}\right\}\alpha_0(s)ds\\ &= \int_0^\tau \left\{S^{(2)} - \frac{\{S^{(1)}\}^{\otimes 2}}{S^{(0)}}\right\}\alpha_0(s)ds = \int_0^\tau \left\{\frac{S^{(2)}}{S^{(0)}} - \frac{\{S^{(1)}\}^{\otimes 2}}{\{S^{(0)}\}^2}\right\}S^{(0)}\cdot\alpha_0(s)ds.\end{aligned}$$

■

Proposition 2.3.5. The difference between the partial information of the Cox model and the predictable variation of the partial score is a mean-zero martingale:

$$\mathcal{I}_{\beta\beta'}(\tau) - \langle \mathcal{U}_\beta(\tau)\rangle = \int_0^\tau \nu(s)dM(s)$$

♦

Proof. Writing $dN(s) = \lambda(s)ds + dM(s)$:

$$\mathcal{I}(\tau) = \int_0^\tau \nu(s)dN(s) = \int_0^\tau \nu(s)\lambda(s)ds + \int_0^\tau \nu(s)dM(s).$$

Expanding the overall intensity:

$$\begin{aligned}\int_0^\tau \nu(s)\lambda(s)ds &= \int_0^\tau \nu(s)\left\{\sum_{i=1}^{n}\lambda_i(s)\right\}ds = \int_0^\tau \nu(s)\left\{\sum_{i=1}^{n}Y_i(s)\alpha_0(s)e^{X_i\beta}\right\}ds\\ &= \int_0^\tau \nu(s)\left\{\sum_{i=1}^{n}Y_i(s)e^{X_i\beta}\right\}\alpha_0(s)ds = \langle \mathcal{U}_\beta(\tau)\rangle.\end{aligned}$$

■

Corollary 2.3.2. In particular, the partial information is the expectation of the predictable variation of the partial score:

$$\mathcal{I}_{\beta\beta'}(\tau) = \mathbb{E}\langle \mathcal{U}_\beta(\tau)\rangle.$$

♣

2.4 Asymptotics

Theorem 2.4.1. Under regularity conditions, the maximum partial likelihood estimator of $\hat{\beta}$ from the Cox model is consistent and asymptotically normal:

$$\sqrt{n}\big(\hat{\beta} - \beta\big) \xrightarrow{d} N\big(0, i_{\beta\beta'}^{-1}\big),$$

where $i_{\beta\beta'}$ is the limit in probability of the information matrix:

$$n^{-1}\mathcal{I}_{\beta\beta'}(\tau) \xrightarrow{p} i_{\beta\beta'},$$

and may be estimated via:

$$\hat{i}_{\beta\beta'} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left[\frac{\sum_{i=1}^{n} X_i^{\otimes 2} Y_i(s) e^{X_i\beta}}{\sum_{i=1}^{n} Y_i(s) e^{X_i\beta}} - \frac{\big\{\sum_{i=1}^{n} X_i Y_i(s) e^{X_i\beta}\big\}^{\otimes 2}}{\big\{\sum_{i=1}^{n} Y_i(s) e^{X_i\beta}\big\}^{2}}\right] dN_i(s).$$

See [1] (VII.2.2). □

2.5 Inference

Inference on β may proceed via the standard Wald, score, and likelihood ratio statistics, each of which is asymptotically χ^2 with $\dim(\beta)$ degrees of freedom.

- Wald statistic:

$$T_W = \big(\hat{\beta} - \beta_0\big)'\mathcal{I}(\hat{\beta})\big(\hat{\beta} - \beta_0\big).$$

- Score statistic:

$$T_S = \mathcal{U}(\beta_0)'\mathcal{I}^{-1}(\beta_0)\mathcal{U}(\beta_0).$$

- Likelihood ratio statistic:

$$T_{LR} = 2\big\{\ell(\hat{\beta}) - \ell(\beta_0)\big\}.$$

2.6 Tied Events

Suppose d_k events occur at time τ_k, and let $\mathcal{D}_k \subseteq \mathcal{R}_k$ denote the indices of those subjects who experience events. *Breslow's partial likelihood* takes the form:

$$L(\beta) = \prod_{k=1}^{K}\frac{\exp\big(\sum_{i\in\mathcal{D}_k} X_i\beta\big)}{\big(\sum_{j\in\mathcal{R}_k} e^{X_j\beta}\big)^{d_k}}$$

Breslow's partial likelihood is simplest to implement and performs well when the number of ties is few. *Efron's partial likelihood* is more accurate but more complex:

$$L(\beta) = \prod_{k=1}^{K} \frac{\exp\left(\sum_{i\in\mathcal{D}_k} X_i\beta\right)}{\prod_{j=1}^{d_k}\sum_{j^*\in\mathcal{R}_k} e^{X_{j^*}\beta} - \frac{j-1}{d_k}\sum_{j^*\in\mathcal{D}_k} e^{X_{j^*}\beta}}.$$

An exact partial likelihood is available, but is computationally intensive and not necessitated unless there are few subjects and few ties. By default, `R`'s `coxph` function uses Efron's partial likelihood.

2.7 Adjusted Hazard and Survival Curves

Recall that, in the absence of covariates, the Nelson-Aalen estimator of the cumulative hazard is:

$$\hat{A}(\tau) = \int_0^\tau \frac{dN(t)}{Y(t)} = \int_0^\tau \frac{dN(t)}{\sum_{i=1}^n Y_i(t)}.$$

Under the relative rate model, *Breslow's estimate* of the baseline cumulative hazard is:

$$\hat{A}_0(t) = \int_0^\tau \frac{dN(t)}{\sum_{i=1}^n Y_i(t)r(X_i;\beta)} = \sum_{\{k:\tau_k\le t\}} \frac{dN(\tau_k)}{\sum_{j\in\mathcal{R}_k} r(X_j;\beta)}.$$

For a given covariate vector X_{new}, the predicted cumulative hazard curve is:

$$\hat{A}(t|X_{\text{new}}) = \hat{A}_0(t)r(X_{\text{new}};\hat{\beta}).$$

Similarly, the predicted survival curve is:

$$\hat{S}(t|X_{\text{new}}) = \prod_{u\le t}\left\{1 - d\hat{A}(u|X_{\text{new}})\right\} = \prod_{\{k:\tau_k\le t\}}\left\{1 - \frac{r(X_{\text{new}};\hat{\beta})\cdot dN(\tau_k)}{\sum_{j\in\mathcal{R}_k} r(X_j;\beta)}\right\}.$$

2.8 Martingale Residuals

Definition 2.8.1. The **cumulative intensity process** for a subject with covariate process $X_i(t)$ is:

$$\Lambda_i(t) = \Lambda(t, X_i) = \int_0^t Y_i(s)\alpha(s|X_i)ds = \int_0^t Y_i(s)r(X_i;\beta)\cdot\alpha_0(s)ds.$$

■

Breslow's estimate of the cumulative intensity process is:

$$\hat{\Lambda}_i(t) = \int_0^t Y_i(s)r(X_i;\hat{\beta})\cdot d\hat{A}_0(s) = \sum_{\{k:\tau_k\le t\}} \frac{Y_i(\tau_k)r(X_i;\hat{\beta})}{\sum_{j\in\mathcal{R}_k} r(X_j;\hat{\beta})}$$

Definition 2.8.2. The **martingale residual process** is:

$$\hat{M}_i(t) = N_i(t) - \hat{\Lambda}_i(t).$$

The *martingale residuals* typically refer to the collection of residuals $\hat{M}_i$ evaluated at the end of study follow-up τ. ■

Additive Rate Models

For an **additive rate model**, the intensity process takes the form:

$$\lambda_i(t) = Y_i(t)\big\{\beta_0(t) + \beta_1(t)X_{1i}(t) + \cdots + \beta_J(t)X_{Ji}(t)\big\} \equiv Y_i(t)\big\{X_i(t)\beta(t)\big\}.$$

- Estimation focuses on the cumulative regression functions $B_j(t) = \int_0^t \beta_j(t)dt$. The regression functions $\beta_j(t)$ may be estimated from the cumulative functions via kernel smoothing.
- The additive rate model allows each regression function $\beta_j(t)$ to change across time, and is fully non-parametric. This flexibility is in contrast to the Cox model, which suppose the effect of a covariate on the relative rate is constant over time.
- Linearity allows the additive model to easily accommodate measurement error or omission of key covariates. In contrast, the Cox model is non-collapsible, and may provide misleading estimates under measurement error or omitted covariates.

Parametric Models

In parametric counting process models, a parametric form is placed on the subject-specific hazard rate. The intensity takes the form:

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

4.1 Likelihood

Let $0 = \tau_1 < \cdots < \tau_K = \tau$ denote a partition of the observation period $[0,\tau]$ into increments of length dt. The full likelihood of the data $\mathcal{D}$ expressible as:

$$L(\mathcal{D}) = \prod_{k=0}^{K-1} \mathbb{P}\big\{\mathcal{D}_k | \mathscr{F}_{\tau_k -}\big\},$$

where $\mathcal{D}_k$ is the data occurring in the interval $[\tau_k, \tau_k + dt) = [\tau_k, \tau_{k+1})$, and $\mathscr{F}_{\tau_k -}$ is the observed data immediately before time τ_k. Partition $\mathcal{D}_k = \mathcal{E}_k \cup \mathcal{R}_k$ where $\mathcal{E}_k$ is the data on events, occurring in the interval $[\tau_k, \tau_{k+1})$, and $\mathcal{R}_k$ is the remaining data. Then the full likelihood decomposes as:

$$L(\mathcal{D}) = \prod_{k=0}^{K-1} \mathbb{P}\big\{\mathcal{E}_k | \mathscr{F}_{\tau_k -}\big\} \cdot \mathbb{P}\big\{\mathcal{R}_k | \mathcal{E}_k, \mathscr{F}_{\tau_k -}\big\}.$$

In partial likelihood, the contribution of the remaining data $\mathbb{P}\big\{\mathcal{R}_k | \mathcal{E}_k, \mathscr{F}_{\tau_k -}\big\}$ is omitted, and inference is based on:

$$L(\mathcal{D}) = \prod_{k=0}^{K-1} \mathbb{P}\big\{\mathcal{E}_k | \mathscr{F}_{\tau_k -}\big\}$$

For event-time data with subject-specific intensity $\lambda_i(t)$, the partial likelihood is:

$$\begin{aligned} L(\mathcal{D}) &= \lim_{dt \to 0} \prod_{k=0}^{K-1} \mathbb{P}\{dN_i(t) = 1 | \mathscr{F}_{\tau_k -}\}^{dN_k(t)} \mathbb{P}\{dN(t) = 0 | \mathscr{F}_{\tau_k -}\}^{1-dN(t)} \\ &= \prod_{u \le \tau} \left\{ \prod_{i=1}^{n} \lambda_i(t)^{dN_i(t)} \right\} \big\{1 - \lambda(t)dt\big\}^{1-dN(t)} \\ &= \left\{ \prod_{i=1}^{n} \prod_{u \le \tau} \lambda_i(t)^{dN_i(t)} \right\} \cdot \exp\left\{ - \int_0^{\tau} \lambda(t)dt \right\}. \end{aligned}$$

where $\lambda(t) = \sum_{i=1}^{n} \lambda_i(t)$ and $dN(t) = \sum_{i=1}^{n} dN_i(t)$ are the aggregated intensity and counting process increments. The log partial likelihood is:

$$\ell(\mathcal{D}) = \sum_{i=1}^{n} \left\{ \int_0^{\tau} \ln \lambda_i(t) dN_i(t) \right\} - \int_0^{\tau} \lambda(t)dt. \tag{4.1.4}$$

4.2 Asymptotics

Suppose $\lambda_i(t) = \lambda_i(t; \theta)$ is parameterized by θ. Taking the gradient of (4.1.4) with respect to θ, the score vector is:

$$\mathcal{U}_{\theta}(\tau) = \sum_{i=1}^{n} \int_0^{\tau} \frac{\partial}{\partial \theta} \ln \lambda_i(t; \theta) \cdot dN_i(t) - \int_0^{\tau} \frac{\partial}{\partial \theta} \lambda(t; \theta) \cdot dt$$

Proposition 4.2.1. $\mathcal{U}_{\theta}$ is expressible as a mean-zero martingale:

$$\mathcal{U}_{\theta}(\tau) = \sum_{i=1}^{n} \int_0^{\tau} \frac{\partial}{\partial \theta} \ln \lambda_i(t; \theta) \cdot dM_i(t).$$

♦

Proof. Using $dN_i(t) = dM_i(t) + \lambda_i(t;\theta)dt$, the first term of the score vector is:

$$\sum_{i=1}^{n}\int_0^{\tau}\frac{\partial}{\partial\theta}\ln\lambda_i(t;\theta)\cdot dN_i(t)$$
$$= \sum_{i=1}^{n}\int_0^{\tau}\frac{\partial}{\partial\theta}\ln\lambda_i(t;\theta)\cdot dM_i(t) + \sum_{i=1}^{n}\int_0^{\tau}\frac{\partial}{\partial\theta}\ln\lambda_i(t;\theta)\cdot\lambda_i(t;\theta)dt.$$

Moving the derivative in the second term past the logarithm:

$$\sum_{i=1}^{n}\int_0^{\tau}\frac{\partial}{\partial\theta}\ln\lambda_i(t;\theta)\cdot\lambda_i(t;\theta)dt = \sum_{i=1}^{n}\int_0^{\tau}\frac{\frac{\partial}{\partial\theta}\lambda_i(t;\theta)}{\lambda_i(t;\theta)}\cdot\lambda_i(t;\theta)dt$$
$$= \sum_{i=1}^{n}\int_0^{\tau}\frac{\partial}{\partial\theta}\lambda_i(t;\theta)dt = \int_0^{\tau}\frac{\partial}{\partial\theta}\lambda(t;\theta)dt.$$

Overall:

$$\mathcal{U}_\theta(\tau) = \sum_{i=1}^{n}\int_0^{\tau}\frac{\partial}{\partial\theta}\ln\lambda_i(t;\theta)\cdot dM_i(t) + \int_0^{\tau}\frac{\partial}{\partial\theta}\lambda(t;\theta)dt - \int_0^{\tau}\frac{\partial}{\partial\theta}\lambda(t;\theta)dt.$$

■

Corollary 4.2.1. Using orthogonality of the martingales for independent subjects:

$$\langle\mathcal{U}_\theta(\tau)\rangle = \sum_{i=1}^{n}\int_0^{\tau}\left\{\frac{\partial}{\partial\theta}\ln\lambda_i(t;\theta)\right\}^{\otimes 2}\cdot\lambda_i(t;\theta)dt.$$

♣

The observed information is:

$$\mathcal{I}_{\theta\theta'}(\tau) = -\sum_{i=1}^{n}\int_0^{\tau}\frac{\partial^2}{\partial\theta\partial\theta'}\ln\lambda_i(t;\theta)\cdot dN_i(t) + \int_0^{\tau}\frac{\partial^2}{\partial\theta\partial\theta'}\lambda(t;\theta)\cdot dt.$$

Proposition 4.2.2. The difference between the observed information at the predictable variation of the score vector is a mean-zero martingale:

$$\mathcal{I}_{\theta\theta'}(\tau) - \langle\mathcal{U}_\theta(\tau)\rangle = -\sum_{i=1}^{n}\int_0^{\tau}\frac{\partial^2}{\partial\theta\partial\theta'}\ln\lambda_i(t;\theta)\cdot dM_i(t).$$

♦

Proof. Substituting $dN_i(t) = dM_i(t) + \lambda_i(t;\theta)dt$ into the observed information, the first term becomes:

$$\sum_{i=1}^{n}\int_0^{\tau}\frac{\partial^2}{\partial\theta\partial\theta'}\ln\lambda_i(t;\theta)\cdot dN_i(t)$$
$$= \sum_{i=1}^{n}\int_0^{\tau}\frac{\partial^2}{\partial\theta\partial\theta'}\ln\lambda_i(t;\theta)\cdot dM_i(t) + \sum_{i=1}^{n}\int_0^{\tau}\frac{\partial^2}{\partial\theta\partial\theta'}\ln\lambda_i(t;\theta)\cdot\lambda_i(t;\theta)dt.$$

Moving derivatives past the logarithm in the second term:

$$
\begin{aligned}
&\sum_{i=1}^{n}\int_0^\tau \frac{\partial^2}{\partial\theta\partial\theta'}\ln\lambda_i(t;\theta)\cdot\lambda_i(t;\theta)dt \\
&\quad= \sum_{i=1}^{n}\int_0^\tau \frac{\partial}{\partial\theta'}\left\{\frac{\frac{\partial}{\partial\theta}\lambda_i(t;\theta)}{\lambda_i(t;\theta)}\right\}\cdot\lambda_i(t;\theta)dt \\
&\quad= \sum_{i=1}^{n}\int_0^\tau \left\{\frac{\lambda_i(t;\theta)\cdot\frac{\partial^2}{\partial\theta\theta'}\lambda_i(t;\theta)-\frac{\partial}{\partial\theta}\lambda_i(t;\theta)\frac{\partial}{\partial\theta'}\lambda_i(t;\theta)}{\lambda_i^2(t;\theta)}\right\}\cdot\lambda_i(t;\theta)dt \\
&\quad= \sum_{i=1}^{n}\int_0^\tau \frac{\partial^2}{\partial\theta\partial\theta'}\lambda_i(t;\theta)dt-\sum_{i=1}^{n}\int_0^\tau\left\{\frac{\frac{\partial}{\partial\theta}\lambda_i(t;\theta)\frac{\partial}{\partial\theta'}\lambda_i(t;\theta)}{\lambda_i^2(t;\theta)}\right\}\cdot\lambda_i(t;\theta)dt \\
&\quad= \int_0^\tau \frac{\partial^2}{\partial\theta\partial\theta'}\lambda(t;\theta)dt-\sum_{i=1}^{n}\int_0^\tau\left\{\frac{\partial}{\partial\theta}\ln\lambda_i(t;\theta)\right\}^{\otimes 2}\cdot\lambda_i(t;\theta)dt
\end{aligned}
$$

Overall:

$$
\begin{aligned}
\mathcal{I}_{\theta\theta'}(\tau) &= -\sum_{i=1}^{n}\int_0^\tau \frac{\partial^2}{\partial\theta\partial\theta'}\ln\lambda_i(t;\theta)\cdot dM_i(t) \\
&\quad -\int_0^\tau \frac{\partial^2}{\partial\theta\partial\theta'}\lambda(t;\theta)dt+\sum_{i=1}^{n}\int_0^\tau\left\{\frac{\partial}{\partial\theta}\ln\lambda_i(t;\theta)\right\}^{\otimes 2}\cdot\lambda_i(t;\theta)dt \\
&\quad +\int_0^\tau \frac{\partial^2}{\partial\theta\partial\theta'}\lambda(t;\theta)dt \\
&= -\sum_{i=1}^{n}\int_0^\tau \frac{\partial^2}{\partial\theta\partial\theta'}\ln\lambda_i(t;\theta)\cdot dM_i(t)+\sum_{i=1}^{n}\int_0^\tau\left\{\frac{\partial}{\partial\theta}\ln\lambda_i(t;\theta)\right\}^{\otimes 2}\cdot\lambda_i(t;\theta)dt.
\end{aligned}
$$

Conclude by identifying the second term as the predictable variation of the score. ■

Theorem 4.2.1. Under regularity conditions, the maximum (partial) likelihood estimator from a counting process model is consistent and asymptotically normal:

$$
\sqrt{n}\big(\hat{\theta}-\theta\big)\xrightarrow{d}N\big(0,i_{\theta\theta'}^{-1}\big),
$$

where $i_{\theta\theta'}$ is the limit in probability of the information matrix:

$$
n^{-1}\mathcal{I}_{\theta\theta'}(\tau)\xrightarrow{p}i_{\theta\theta'},
$$

and may be estimated via:

$$
\hat{i}_{\theta\theta'}=\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\left\{\frac{\partial}{\partial\theta}\ln\lambda_i(t;\theta)\right\}^{\otimes 2}dN_i(t).
$$

See [1] VI.1.2. □

4.3 Poisson Regression

Example 4.3.1. Consider the intensity process:

$$\lambda_i(t;\theta) = Y_i(t)\alpha_0(t;\theta)e^{X_i\beta}.$$

Let $0 = \tau_1 < \cdots < \tau_K = \tau$ denote a partition of the observation period $[0, \tau]$ and suppose the baseline hazard is piecewise-constant:

$$\alpha_0(t;\theta) = \sum_{k=1}^{K} \theta_k \mathbb{I}_{[\tau_k,\tau_{k+1})}(t)$$

To form the partial likelihood, define:

$$E_{ik} = \int_0^\tau \mathbb{I}_{[\tau_k,\tau_{k+1})}(t)dN_i(t)$$

as the number of events subject i experienced in interval $[\tau_k, \tau_{k+1})$, and define:

$$R_{ik} = \int_0^\tau \mathbb{I}_{[\tau_k,\tau_{k+1})}(t)Y_i(t)dt$$

as the duration of the $[\tau_k, \tau_{k+1})$ over which subject i was at risk.

Observe that the integrated intensity for subject i is expressible as:

$$\begin{aligned}\int_0^\tau \lambda_i(t;\theta)dt &= \int_0^\tau Y_i(t)\left(\sum_{k=1}^{K}\theta_k\mathbb{I}_{[\tau_k,\tau_{k+1})}\right)e^{X_i\beta}dt \\ &= \sum_{k=1}^{K}\theta_k e^{X_i\beta}\int_0^\tau \mathbb{I}_{[\tau_k,\tau_{k+1})}Y_i(t)dt = \sum_{k=1}^{K}\theta_k e^{X_i\beta}R_{ik}.\end{aligned}$$

Now the partial likelihood is expressible as:

$$\begin{aligned}L(\theta,\beta) &= \left\{\prod_{i=1}^{n}\prod_{k=1}^{K}(\theta_k e^{X_i\beta}R_{ik})^{E_{ik}}\right\}\exp\left\{-\sum_{i=1}^{n}\sum_{k=1}^{K}\theta_k e^{X_i\beta}R_{ik}\right\} \\ &= \prod_{i=1}^{n}\prod_{k=1}^{K}\left\{(\theta_k e^{X_i\beta}R_{ik})^{E_{ik}}e^{-(\theta_k e^{X_i\beta}R_{ik})}\right\} = \prod_{i=1}^{n}\prod_{k=1}^{K}\left\{\mu_{ik}^{E_{ik}}e^{-\mu_{ik}}\right\}.\end{aligned}$$

This likelihood is proportional to a Poisson likelihood on E_{ik} with regression function:

$$\ln\mu_{ik} = \ln\theta_k + X_i\beta + \ln R_{ik}.$$

To fit this model, K tuples of pseudo-data $\{(E_{ik}, R_{ik}, X_i)\}_{k=1}^{K}$ are generated for *each* subject i, one tuple for each interval, and a Poisson GLM is specified, with logarithmic link function and offset $\ln R_{ik}$. ♠

References

[1] PK Andersen et al. *Statistical Models Based on Counting Processes*. 2nd. Springer-Verlag, 1997.