MATH3701 Statistical Modelling
School of Mathematics, University of Leeds

Supplementary Proof: The Cox Partial Likelihood

This note justifies why the partial likelihood provides a valid basis for inference on $\boldsymbol{\beta}$ in the Cox model, without requiring specification of the baseline hazard $h_0(t)$.

Setup

The Cox proportional hazards model is

$$h(t \mid \mathbf{x}_i) = h_0(t)\exp(\boldsymbol{\beta}^T\mathbf{x}_i), \qquad i = 1, \ldots, n.$$

We observe $(t_i, \delta_i, \mathbf{x}_i)$ where t_i is the observed time, δ_i the event indicator, and $\mathbf{x}_i \in \mathbb{R}^p$ the covariate vector.

Let $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$ be the ordered distinct event times (assuming no ties for simplicity), and let $\mathcal{R}(t_{(j)}) = \{i : t_i \geq t_{(j)}\}$ be the risk set just before $t_{(j)}$.

Derivation of the partial likelihood

Cox (1972) proposed to condition on the event times $t_{(1)}, \ldots, t_{(r)}$ and use only the information about which individual experienced the event at each time.

Lemma 1. *Under the Cox model, the conditional probability that individual $i_{(j)}$ is the one who dies at time $t_{(j)}$, given that exactly one death occurs in $\mathcal{R}(t_{(j)})$ at that time, is*

$$\mathbb{P}\big(T_{i_{(j)}} = t_{(j)} \mid \textit{one death in } \mathcal{R}(t_{(j)})\big) = \frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_{i_{(j)}})}{\displaystyle\sum_{\ell \in \mathcal{R}(t_{(j)})} \exp(\boldsymbol{\beta}^T\mathbf{x}_\ell)}.$$

Proof. For a small interval $[t, t + \delta t)$, the probability that individual ℓ dies in this interval, given they are at risk, is approximately $h(t \mid \mathbf{x}_\ell)\,\delta t = h_0(t)\exp(\boldsymbol{\beta}^T\mathbf{x}_\ell)\,\delta t$.

The probability that individual $i_{(j)}$ dies at $t_{(j)}$ while all others in the risk set survive past $t_{(j)}$ is proportional to $h_0(t_{(j)})\exp(\boldsymbol{\beta}^T\mathbf{x}_{i_{(j)}})$. The probability that *some* individual in $\mathcal{R}(t_{(j)})$ dies at $t_{(j)}$ is proportional to $h_0(t_{(j)})\sum_{\ell \in \mathcal{R}} \exp(\boldsymbol{\beta}^T\mathbf{x}_\ell)$.

Dividing:

$$\mathbb{P}(i_{(j)} \text{ dies} \mid \text{one death in } \mathcal{R}) = \frac{h_0(t_{(j)})\exp(\boldsymbol{\beta}^T\mathbf{x}_{i_{(j)}})}{h_0(t_{(j)})\sum_{\ell \in \mathcal{R}(t_{(j)})}\exp(\boldsymbol{\beta}^T\mathbf{x}_\ell)} = \frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_{i_{(j)}})}{\sum_{\ell \in \mathcal{R}(t_{(j)})}\exp(\boldsymbol{\beta}^T\mathbf{x}_\ell)}.$$

The baseline hazard $h_0(t_{(j)})$ cancels completely. □

Remark 1. This cancellation is the key insight: by conditioning on the observed event times, we eliminate the nuisance function $h_0(t)$ entirely. The remaining expression depends only on $\boldsymbol{\beta}$.

The partial likelihood

Multiplying the conditional probabilities across all r event times, the **partial likelihood** is

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_{(j)})}{\displaystyle\sum_{\ell \in \mathcal{R}(t_{(j)})} \exp(\boldsymbol{\beta}^T\mathbf{x}_\ell)},$$

or equivalently, using the event indicators δ_i:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_i)}{\sum_{\ell \in \mathcal{R}(t_i)} \exp(\boldsymbol{\beta}^T \mathbf{x}_\ell)} \right]^{\delta_i}.$$

The corresponding partial log-likelihood is

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left\{ \boldsymbol{\beta}^T \mathbf{x}_i - \log \sum_{\ell \in \mathcal{R}(t_i)} \exp(\boldsymbol{\beta}^T \mathbf{x}_\ell) \right\}.$$

Score equations

Differentiating $\ell(\boldsymbol{\beta})$ with respect to β_k:

$$\frac{\partial \ell}{\partial \beta_k} = \sum_{i=1}^{n} \delta_i \left\{ x_{ik} - \frac{\sum_{\ell \in \mathcal{R}(t_i)} x_{\ell k} \exp(\boldsymbol{\beta}^T \mathbf{x}_\ell)}{\sum_{\ell \in \mathcal{R}(t_i)} \exp(\boldsymbol{\beta}^T \mathbf{x}_\ell)} \right\} = 0.$$

These p equations are solved numerically (Newton–Raphson) to give the maximum partial likelihood estimator (MPLE) $\hat{\boldsymbol{\beta}}$.

Asymptotic properties

Theorem 1 (Cox 1975; Tsiatis 1981)**.** *Under standard regularity conditions, the MPLE $\hat{\boldsymbol{\beta}}$ satisfies:*

1. ***Consistency:*** $\hat{\boldsymbol{\beta}} \xrightarrow{\mathbb{P}} \boldsymbol{\beta}$ *as* $n \to \infty$.
2. ***Asymptotic normality:***

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}_p\big(\mathbf{0}, \mathcal{I}(\boldsymbol{\beta})^{-1}\big),$$

   where $\mathcal{I}(\boldsymbol{\beta}) = -n^{-1}\, \partial^2 \ell / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T$ *is the (per-observation) observed information matrix evaluated at* $\hat{\boldsymbol{\beta}}$.

3. ***Efficiency:*** *Among all estimators based on the rank ordering of event times,* $\hat{\boldsymbol{\beta}}$ *is asymptotically efficient.*

The asymptotic normality justifies the Wald confidence intervals reported by `coxph()` in R:

$$\hat{\beta}_k \pm z_{1-\alpha/2} \cdot \widehat{\mathrm{sd}}(\hat{\beta}_k),$$

where $\widehat{\mathrm{sd}}(\hat{\beta}_k) = \sqrt{[\mathcal{I}(\hat{\boldsymbol{\beta}})^{-1}]_{kk}}$.

Partial likelihood ratio test

For nested models $M_0 \subset M_1$ with $q - p$ additional parameters in M_1, the partial likelihood ratio test statistic

$$\Lambda = -2(\hat{\ell}_0 - \hat{\ell}_1) \mathrel{\dot\sim} \chi^2_{q-p}$$

has the same asymptotic χ^2 distribution as for standard MLEs, justifying the use of this test in `coxph()` comparisons.

Tied event times: Breslow's approximation

With d_j tied deaths at $t_{(j)}$ with covariate sum $\mathbf{s}_{(j)} = \sum_{k:t_k=t_{(j)},\delta_k=1} \mathbf{x}_k$, Breslow's approximation replaces the exact combinatorial expression with

$$L(\boldsymbol{\beta}) \approx \prod_{j=1}^{r} \frac{\exp(\boldsymbol{\beta}^T \mathbf{s}_{(j)})}{\left(\sum_{\ell \in \mathcal{R}(t_{(j)})} \exp(\boldsymbol{\beta}^T \mathbf{x}_\ell)\right)^{d_j}}.$$

This approximation is accurate when ties are rare (the typical situation in continuous survival data).

References

Cox, D.R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.

Cox, D.R. (1975). Partial likelihood. *Biometrika*, **62**, 269–276.

Tsiatis, A.A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics*, **9**, 93–108.