3

The Cox Model

3.1 Introduction and notation

The Cox proportional hazards model [36] has become by a wide margin the most used procedure for modeling the relationship of covariates to a survival or other censored outcome.

Let $X_{ij}(t)$ be the jth covariate of the ith person, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$. It is natural to think of the set of covariates as forming an $n \times p$ matrix, and we use X_i to denote the covariate vector for subject i, that is, the ith row of the matrix. When all covariates are *fixed* over time X_i is just a vector of covariate values, familiar from multiple linear regression. For other data sets one or more covariates may vary over time, for example a repeated laboratory test. We use X_i for both time-fixed and time-varying covariate processes, employing $X_i(t)$ when we wish to emphasize the time-varying structure.

The Cox model specifies the hazard for individual i as

$$\lambda_i(t) = \lambda_0(t) e^{X_i(t)\beta}, \tag{3.1}$$

where λ_0 is an unspecified nonnegative function of time called the *baseline hazard*, and β is a $p \times 1$ column vector of coefficients. Event rates cannot be negative (observed deaths can not unhappen), and the exponential thus plays an important role in ensuring that the final estimates are a physical possibility.

Because the hazard ratio for two subjects with fixed covariate vectors X_i and X_j,

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)e^{X_i\beta}}{\lambda_0(t)e^{X_j\beta}} = \frac{e^{X_i\beta}}{e^{X_j\beta}},$$

T. M. Therneau et al., *Modeling Survival Data: Extending the Cox Model*

is constant over time, the model is also known as the *proportional hazards* model.

Estimation of β is based on the partial likelihood function introduced by Cox [36]. For untied failure time data it has the form

$$PL(\beta) = \prod_{i=1}^{n} \prod_{t \geq 0} \left\{ \frac{Y_i(t) r_i(\beta, t)}{\sum_j Y_j(t) r_j(\beta, t)} \right\}^{dN_i(t)}, \tag{3.2}$$

where $r_i(\beta, t)$ is the *risk score* for subject i, $r_i(\beta, t) = \exp[X_i(t)\beta] \equiv r_i(t)$. The log partial likelihood can be written as a sum

$$l(\beta) = \sum_{i=1}^{n} \int_0^\infty \left[Y_i(t) X_i(t)\beta - \log\left(\sum_j Y_j(t) r_j(t) \right) \right] dN_i(t), \tag{3.3}$$

from which we can already foresee the counting process structure.

Although the partial likelihood is not, in general, a likelihood in the sense of being proportional to the probability of an observed dataset, nonetheless it can be treated as a likelihood for purposes of asymptotic inference. Differentiating the log partial likelihood with respect to β gives the $p \times 1$ score vector, $U(\beta)$:

$$U(\beta) = \sum_{i=1}^{n} \int_0^\infty [X_i(s) - \bar{x}(\beta, s)] \, dN_i(s), \tag{3.4}$$

where $\bar{x}(\beta, s)$ is simply a weighted mean of X, over those observations still at risk at time s,

$$\bar{x}(\beta, s) = \frac{\sum Y_i(s) r_i(s) X_i(s)}{\sum Y_i(s) r_i(s)}, \tag{3.5}$$

with $Y_i(s) r_i(s)$ as the weights.

The negative second derivative is the $p \times p$ information matrix

$$\mathcal{I}(\beta) = \sum_{i=1}^{n} \int_0^\infty V(\beta, s) dN_i(s), \tag{3.6}$$

where $V(\beta, s)$ is the weighted variance of X at time s:

$$V(\beta, s) = \frac{\sum_i Y_i(s) r_i(s) [X_i(s) - \bar{x}(\beta, s)]' [X_i(s) - \bar{x}(\beta, s)]}{\sum_i Y_i(s) r_i(s)}. \tag{3.7}$$

The maximum partial likelihood estimator is found by solving the partial likelihood equation:

$$U(\hat{\beta}) = 0.$$

The solution $\hat{\beta}$ is consistent and asymptotically normally distributed with mean β, the true parameter vector, and variance $\{\mathcal{E}\mathcal{I}(\beta)\}^{-1}$, the inverse