Handout on Marginal and Partial Likelihood

FOLLOWING KALBFLEISCH AND PRENTICE SEC. 4.7.1

As in class, assume that (X_i, C_i, Z_i) are independent triples, with Z_i non-time-dependent $k-vector$ covariates, and X_i, C_i conditionally independent given Z_i, and

$$h_{X_i|Z_i}(t|z) \; = \; \lambda_0(t)\, e^{\beta' z} \quad , \qquad \text{all} \quad t > 0\,, \;\; z \in \mathbf{R}^k$$

where the unknown parameters are $\beta \in \mathbf{R}^k$ and λ_0 fall in the class of hazard intensity functions strictly positive on some (fixed) interval of support, with $\Lambda_0(t) \equiv \int_0^t \lambda_0(s)\, ds \to \infty$ as $t \to \infty$.

Denote $T_i \equiv X_i \wedge C_i$, $\Delta_i \equiv I_{[X_i \le C_i]}$, $N(t) \equiv \sum_{i=1}^n \Delta_i\, I_{[T_i \le t]}$, and also let t_j, $1 \le j \le N(\infty)$, denote the ordered distinct observed death-times, i.e. the jump-points for $N(\cdot)$, and for all j,

$$\mathcal{R}_j \equiv \{i = 1, \ldots, n \; : \quad T_i \ge t_j\} \;\; , \qquad L_j \equiv \text{ index } i \text{ such that } \quad T_i = t_j,\; \Delta_i = 1$$

Denote by $\underline{R}$ the rank-data consisting of all inequalities

$$X_l < X_a \quad \text{where} \quad l \in \{L_j \,:\; 1 \le j \le N(\infty)\}\,, \quad \text{and if} \quad l = L_j,\; a \in \mathcal{R}_j \backslash \{l\}$$

Note that $\underline{R}$ is less information than we have from $\{(\mathcal{R}_j,\, L_j),\; j = 1, \ldots, N(\infty)\}$ because $\underline{R}$ drops information about censoring times for $\mathcal{R}_j$ individuals being greater than X_{L_j}, but the $\mathcal{R}_j$, L_j information does imply the $\underline{R}$ information.

Proposition. Either regard the covariates Z_i as fixed, or condition on them. Then the likelihood for the data $\underline{R}$ coincides with the Cox Partial Likelihood $\prod_{j=1}^{N(\infty)} (e^{\beta' Z_{L_j}} / \sum_{k \in \mathcal{R}_j} e^{\beta' Z_k})$.

Proof. *Step 1.* The rank information $\underline{R}$ would be unchanged if the time axis (the interval of positivity of λ_0) were transformed by any strictly monotone transformation; and the mapping $t \mapsto \Lambda_0(t)$ provides $X_i^* \equiv \Lambda_0(X_i)$ with hazards $h_{X_i^*|Z_i}(s|z) = e^{\beta' z}$. Thus there is no loss of generality in assuming (conditionally) $X_i \sim \text{Expon}(e^{\beta' Z_i})$.

Step 2. The calculation now relies on the memoryless property of exponentials together with the fact that if $U \sim \text{Expon}(\mu)$, $V \sim \text{Expon}(\nu)$ are independent, then $P(U < V \,|\, \min(U,V) = s) = \mu/(\mu + \nu)$. We note that $\underline{R}$ does incorporate knowledge of all $\mathcal{R}_j$. Since $t_j = \min\{X_a \,:\, a \in \mathcal{R}_j\}$ for all j, with $t_0 \equiv 0$ by convention, the memorlyess property implies that given t_{j-1}, for all $a \in \mathcal{R}_j$ the r.v.'s $X_a - t_{j-1}$ are independent and $\text{Expon}(e^{\beta' Z_a})$ distributed, so that $\min_{a \in \mathcal{R}_j \backslash \{l\}} (X_a - t_{j-1}) \sim \text{Expon}(\sum_{a \in \mathcal{R}_j \backslash \{l\}} e^{\beta' Z_a})$, and

$$P(X_l - t_{j-1} < \min_{a \in \mathcal{R}_j \backslash \{l\}} (X_a - t_{j-1}) \,|\, \mathcal{R}_j,\, t_{j-1}) \; = \; e^{\beta' Z_l} / \prod_{k \in \mathcal{R}_j} e^{\beta' Z_k}$$