Partial Likelihood as Profile Likelihood

Consider generally with time-dependent covariates:

$$\lambda_i(t) = \lambda_0(t) \exp\{\beta'Z_i(t)\},$$

where $i = 1, \ldots, n$ denotes the subjects.

The likelihood is

$$L = \prod_i \lambda_i(X_i)^{\delta_i} S_i(X_i)/S_i(Q_i),$$

where $X_i$ is the possibly censored event time, $\delta_i$ is the event indicator, and $Q_i$ is the entry time.

The log-likelihood is then [ex]

$$l = \sum_i \left[ \delta_i \{ \log \lambda_0(X_i) + \beta'Z_i(X_i) \} - \int_{Q_i}^{X_i} \lambda_0(s) \exp\{\beta'Z_i(s)\} ds \right]. \tag{1}$$

Nonparametric likelihood discretizes $\lambda_0(\cdot)$ to mass points at the observed event times $0 < t_1 < \ldots < t_K$, so that the integral in (1) becomes $\sum_{k:Q_i \leq t_k \leq X_i} \lambda_0(t_k) \exp\{\beta'Z_i(t_k)\}$.

Denote $\lambda_k = \lambda_0(t_k)$. 
For given $\beta$, we want to maximize $l$ over $\lambda_0(\cdot)$, i.e. over $\lambda_1, ..., \lambda_K$.

Set
\[
\frac{\partial l}{\partial \lambda_k} = \frac{d_k}{\lambda_k} - \sum_{i:Q_i \leq t_k \leq X_i} \exp\{\beta'Z_i(t_k)\} = 0,
\]
where $d_k$ is the number of events at $t_k$. [ex]

We have
\[
\hat{\lambda}_k = \frac{d_k}{\sum_{R_k} \exp\{\beta'Z_i(t_k)\}}, \tag{2}
\]
where $R_k$ is the risk set at $t_k$.

We can verify that $\partial^2 l / \partial \lambda_k^2 = -d_k / \lambda_k^2 < 0$, so $\hat{\lambda}_k$ is the maximum.
Plugging $\hat{\lambda}_k$ back into (1), assuming no ties ($d_k = 1$) we have

$$
\begin{align*}
  l &= \sum_i \left[ \delta_i \left\{ \log \frac{1}{\sum_{j \in R_i} \exp\{\beta' Z_j(X_i)\}} + \beta' Z_i(X_i) \right\} \\
  &\quad - \sum_{k:Q_i \leq t_k \leq X_i} d_k \exp\{\beta' Z_i(t_k)\} \right] \\
  &= \sum_i \left[ \delta_i \log \frac{\exp\{\beta' Z_i(X_i)\}}{\sum_{j \in R_i} \exp\{\beta' Z_j(X_i)\}} \right] - A, \quad (3)
\end{align*}
$$

where

$$
A = \sum_i \sum_{k:Q_i \leq t_k \leq X_i} \frac{\eta_i(t_k)}{\sum_{j \in R_k} \eta_j(t_k)} = \sum_{k=1}^K \sum_{i: X_i \geq t_k} \frac{\eta_i(t_k)}{\sum_{j \in R_k} \eta_j(t_k)} = K,
$$

and $\eta_i(t_k) = \exp\{\beta' Z_i(t_k)\}$.

The first term in (3) is the partial log-likelihood.
Profile Likelihood Theory

In general, if $\beta$ and $\lambda$ are two sets of parameters, the profile likelihood for $\beta$ is

$$pL(\beta) = \sup_{\lambda} L(\beta, \lambda).$$

Often $\beta$ is finite dimensional, referred to as the parameter of interest, and $\lambda$ the *nuisance parameter* (finite or infinite dim).

Then $\hat{\beta}$ that maximizes $pL(\beta)$ also maximizes $L(\beta, \lambda)$ (why).

In order for a profile likelihood to have asymptotic properties of a regular likelihood, it suffices if $pl(\beta) = \log pL(\beta)$ behaves as a quadratic function asymptotically near the true $\beta_0$.

**Review:** asymptotic theory of a regular parametric likelihood function (consistency, AN, lik ratio test).
Murphy and van der Vaart (2000, *JASA*) showed that under suitable conditions, for any random sequence $\beta_n$ such that $\|\beta_n - \beta_0\| = O_p(1/\sqrt{n})$,

$$
\frac{1}{n} \{pl(\beta_n) - pl(\beta_0)\} = (\beta_n - \beta_0)'S - \frac{1}{2}(\beta_n - \beta_0)'I(\beta_n - \beta_0) + o_p\left(\frac{1}{n}\right),
$$

where $S$ is called the *efficient score*, $I$ - the *efficient Fisher information*.

M+V (2000) verified that the required conditions are met for the Cox model (this part can be very technical).

In practice, Maples, Murphy and Axinn (2002) recommend to plot the contours of the profile likelihood to see if it’s approximately quadratic.

**Note that the profile likelihood does not always have a closed form like the partial likelihood.**