Partial Likelihood estimation for the proportional hazards model

Let's now discuss how the **β** parameters of a proportional hazards regression model can be estimated using the idea of partial likelihood. Suppose we have n individuals with i ranging from 1 to n. Each individual has three characteristics: $\mathbf{x}_i$ is the vector of covariates, t_i is the time of the event or censoring, and δ_i is 1 if the individual is uncensored and 0 if the individual is censored. We start by ranking all the events of the non-censored subjects (t_1 up to t_k). Given the fact that one subject has event time t_i, the probability that this subject has inputs x_j is then given by

$$\frac{h(t_i, \boldsymbol{x}_j)\Delta t}{\sum_{l \,\epsilon\, R(t_i)} h(t_i, \boldsymbol{x}_l)\Delta t}$$

whereby $R(t_i)$ represents the subjects that are at risk at time t_i . We can enter the proportional hazards terms, which produces the second expression. Notice that the baseline hazard $h_0(t_i)$ occurs in both the numerator and denominator and thus cancels it out. Hence, this arrives at the expression

$$\frac{\exp(\boldsymbol{\beta}^T \boldsymbol{x}_j)}{\sum_{l \epsilon R(t_i)} \exp(\boldsymbol{\beta}^T \boldsymbol{x}_l)}$$

which is independent of the baseline hazard.

The partial likelihood function then becomes $\prod_{j=1}^{k} \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{x}_j)}{\sum_{l \epsilon R(t_i)} \exp(\boldsymbol{\beta}^T \boldsymbol{x}_l)}$. Note that, for ease of notation, we assumed that individual j with covariates $\mathbf{x}_j$ has event time t_j. The **β** parameters can again be optimized using the Newton-Raphson algorithm. It is important to observe how the censored observations enter the partial likelihood function; they will be included in the risk sets $R(t_j)$ until their censoring time.

Also, it is important to note that the **β** parameters can be estimated without having to specify the baseline hazard $h_0(t)$. Furthermore, it can be shown that the partial likelihood estimates are consistent, which means that they converge to the true values as the sample increases, and are asymptotically normal. Moreover, the partial likelihood estimates depend only on the ranks of the event times and not on the numerical values. An important assumption that was made in deriving the partial likelihood function is that there are no tied event times. However, in many real-life settings, time is measured in a discrete way so that ties are likely to occur.

There are four ways to deal with tied event times: the exact method, two approximations, and the discrete method. In case there are no ties, all four methods give the same estimates. The exact method assumes that ties occur because of imprecise time measurements and treats time as continuous. Hence, it considers all possible orderings of the event times and constructs a likelihood term for each ordering. When there are three ties, six orderings are possible, and thus six terms are added to the partial likelihood function. Obviously, this procedure is very time consuming for heavily tied data. It is recommended only when few ties occur. Two popular approximations are the Breslow and Efron likelihood methods. The Breslow likelihood method works well if ties occur rarely. Empirical evidence has shown that the Efron approximation is often superior to the Breslow approximation. A fourth option is the discrete method whereby time is treated in a discrete way so that multiple events can occur at the same time. Although this can also be addressed using the partial likelihood approach, it will become very time consuming for big, heavily tied survival data sets. Another, more interesting approach to work with discrete time data and use discrete survival analysis as we discuss in what follows.