1 Partial Likelihood

To obtain estimates of the covariate parameters, Cox (1972, 1975) developed a nonparametric method he called *partial likelihood*. Estimation of the parameter values is then obtained by use of maximum partial likelihood estimation. Let's see it conceptually.

Table 1: Data Sorted by Ordered Failure Time

Case Number	Duration Time	Censored Case
7	7	No
4	15	No
5	21	No
2	28	Yes
9	30	Yes
3	36	No
8	45	Yes
1	46	No
6	51	No

Data are sorted by the duration time. The duration time for censored cases denotes the time of last observation.

What are the main features of these data?

- Events can be ordered.
- At t_0 all cases are at risk of failing.
- After the first failure, the risk set decreases by 1.
- The risk set successively dwindles as events occur.

Now to motivate the partial likelihood estimator, let $\psi = \exp(\beta' \mathbf{x_i})$ (this notation is from Collett, 1994, p. 64).

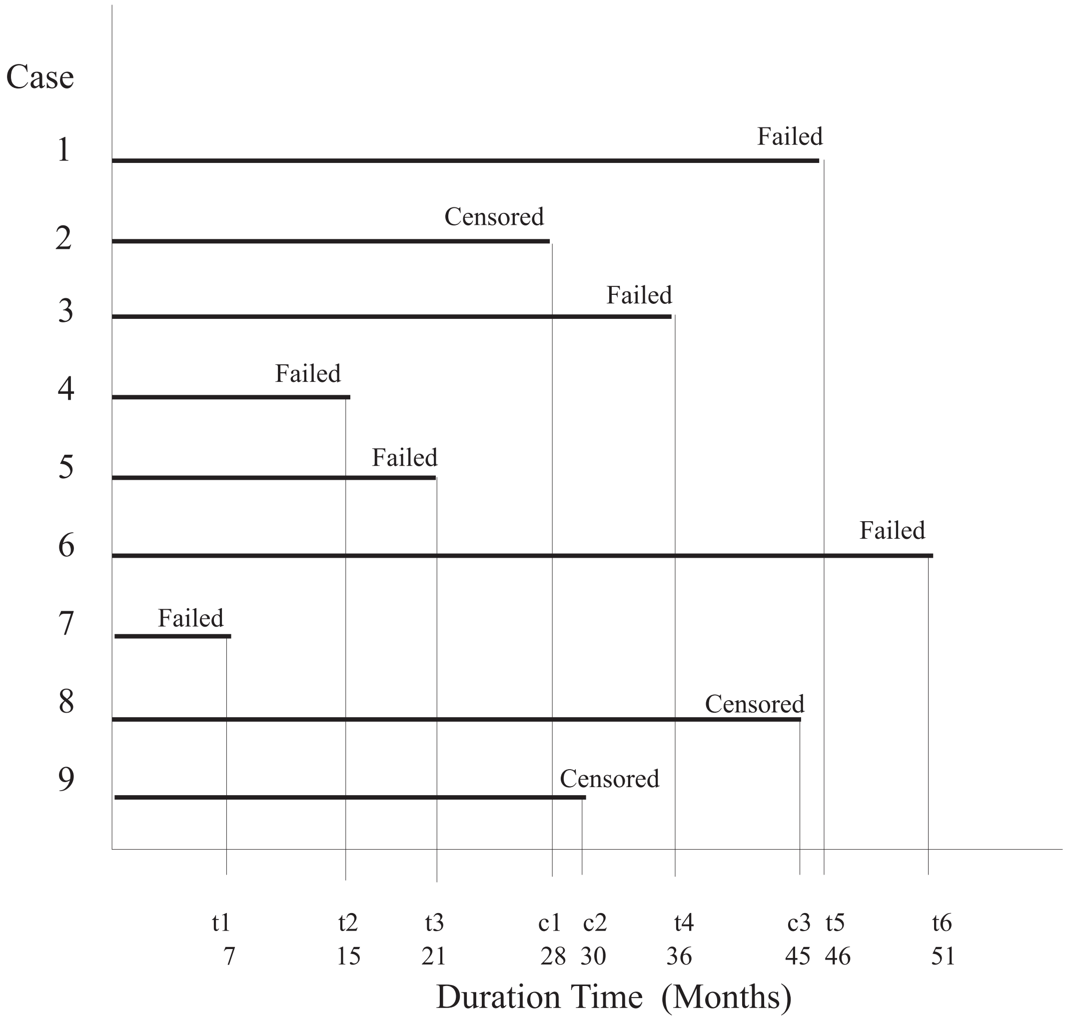

Figure 1: *Duration times for nine censored and uncensored (failed) cases.*

The partial likelihood function for these data would be equivalent to,

$$\mathcal{L}_p = \frac{\psi(7)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)+\psi(5)+\psi(6)+\psi(7)+\psi(8)+\psi(9)} \times$$
$$\frac{\psi(4)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)+\psi(5)+\psi(6)+\psi(8)+\psi(9)}$$
$$\frac{\psi(5)}{\psi(1)+\psi(2)+\psi(3)+\psi 5+\psi(6)+\psi(8)+\psi(9)} \times$$
$$\frac{\psi(3)}{\psi(1)+\psi(3)+\psi(6)+\psi(8)} \times$$
$$\frac{\psi(1)}{\psi(1)+\psi(6)} \times$$
$$\frac{\psi(6)}{\psi(6)}.$$

To be a little more formal, suppose we have a data set with n observations and k distinct failure (event) times. Cox estimation first proceeds by sorting the ordered failure times, such that

$$t_1 < t_2 < \ldots < t_k,$$

where t_i denotes the failure time for the ith individual. For censored cases, we define δ_i to be 0 if the case is right-censored, and 1 if the case is uncensored. Finally, the ordered event times are modeled as a function of covariates, $\mathbf{x}$.

The partial likelihood function is derived by taking the product of the conditional probability of a failure at time t_i, given the number of cases that are at risk of failing at time t_i.

That is to say, given that some event has occurred, what is the probability the event occurred to the ith individual from a risk set of size n?

We define $R(t_i)$ to denote the number of cases that are at risk of experiencing an event at time t_i, that is, the "risk set," then the probability that the jth case will fail at time T_i is given by

$$\Pr(t_j = T_i \mid R(t_i)) = \frac{e^{\beta' \mathbf{x_i}}}{\sum_{j \in R(t_i)} e^{\beta' \mathbf{x_j}}}, \tag{1}$$

where the summation operator in the denominator is summing over all individuals in the risk set. Taking the product of the conditional probabilities in (1) yields the partial likelihood function,

$$\mathcal{L}_p = \prod_{i=1}^{K} \Big[\frac{e^{\beta' \mathbf{x_i}}}{\sum_{j \in R(t_i)} e^{\beta' \mathbf{x_j}}}\Big]^{\delta_i}, \tag{2}$$

with corresponding log-likelihood function,

$$\log L_p = \sum_{i=1}^{K} \delta_i \Big[\beta' \mathbf{x_i} - \log \sum_{j \in R(t_i)} e^{\beta' \mathbf{x_j}}\Big]. \tag{3}$$

By maximizing the log-likelihood in (3), estimates of the β may be obtained.

What is the importance of this result?

- Specifying the baseline hazard, $h_0(t)$ is unnecessary.
- The interval between events does not inform the PL function.
- Censored cases contribute information only pertinent to the risk set (i.e. the denominator, not the numerator)

The critical thing here is to note that no assumptions about the shape of the baseline hazard need to be made. Another way to see this is to think about the heuristic partial likelihood function above. All we need to know to compute a probability is ψ (or $\exp(\beta'\mathbf{x_i})$.