We consider the classical proportional hazards model
\[ h_i(t) = e^{\beta_i} h_0(t), \]
and assume that the baseline hazard \( h_0 \) satisfies a shape constraint: increasing, decreasing, unimodal, or \( u \)-shaped. This approach was originally suggested in Cox (1972, page 190), and later discussed in Mykytyn and Santner (1981). Our goal was to implement this approach in a readily available R package (coming soon!) for continuous and discrete data, while allowing for censoring. To date, we have completed the algorithm for continuous data.

Suppose now that we observe \( t_i \) (event time), \( z_i \) (vector of covariates), and \( \delta_i \) (equals 0 if the observation was censored) for each of \( n \) subjects, and let \( H_t \) denote the cumulative baseline hazard. The full likelihood is
\[
L(\delta, h_0) = \prod_{i=1}^{n} h_i(z_i)(1 - F(t_i))^{\delta_i} \prod_{i=1}^{n} e^{\beta z_i} H_t(z_i) \times \prod_{i=1}^{n} h_0(z_i)^{\delta_i}. 
\]

Maximising the likelihood is the same as minimising the function \((-\log L)\)
\[
\varphi(\beta, h_0) = \sum_{i=1}^{n} e^{\beta z_i} H_t(z_i) - \sum_{i=1}^{n} \delta_i \log h_0(z_i) - \sum_{i=1}^{n} \delta_i \beta z_i, 
\]
which is strictly convex in all parameters.

The Algorithm

To find \( \beta = \arg \min_{\beta} \varphi(\beta, h_0) \), we use the Newton-Raphson algorithm. The solution to \( \hat{h}_0 = \arg \min_{h_0} \varphi(\hat{\beta}, h_0) \) may be found explicitly using a graphical representation. We describe it here under the "decreasing" assumption.

Assume WLOG that the observed times are ordered. Let \( i_j, j = 1, \ldots, m \) denote the indices such that \( \delta_{i_j} = 1 \). For \( j = 1, \ldots, m \), define
\[
s_j = \sum_{t_{i_{j-1}}}^{t_{i_j}} e^{\beta z_i} (t_{i_{j-1}} - t_j).
\]

Then \( \hat{h}_0 \) is piecewise constant with jumps restricted to \( t_{i_{j-1}}, \ldots, t_{i_j} \), and the height of the hazard on \( (t_{i_{j-1}}, t_{i_j}] \) is found as the inverse of the slope on \( (k, k + 1] \) of the greatest convex minorant of \( \{(0, 0), (1, s_0), \ldots, (m, s_m)\} \).

to derive the graphical representation one would need to make the following notes:
1. First, reduce the infinite dimensional problem to a finite dimensional one by noting that the hazard function must be piecewise constant with jumps possible only at the uncensored event times. This is easily seen by examining the log-likelihood.
2. Calculate the score function evaluated at \( \hat{h}_0(t) \). As \( \tau \) varies, these functions form a basis for any decreasing function. Since the score must be equal to zero at \( h_0 \), and be negative otherwise, this yields a set of inequalities often referred to as the Fenchel conditions.
3. Lastly, the Fenchel conditions can be converted into the graphical solution.

A graphical representation of this type was first noted by Grenander (1956), who showed that the MLE of a decreasing density can be found as the (left) derivative of the least concave majorant of the empirical distribution function.

Future Work: Response of Flour–Beetles to DDT

Hewlett (1974) describes an experiment in which the mortality of flour-beetles was measured under varying levels of the insecticide DDT. Adult male and female beetles were sprayed with four different concentration levels (0.20, 0.32, 0.50, and 0.80 mg/cm\(^2\)) of the insecticide at time 0, and their mortality was observed over the next 13 days. We use \( z_i \) to denote the sex and DDT level covariates. For each of the eight treatments, the number of deaths in the intervals 0–1, 1–2, 2–3, …, 12–13 was reported. Quoting Hewlett (1974), “What is required is some single process of computation for estimating simultaneously the parameters of distributions of tolerances and times to death.” Previously, Pack and Morgan (1990) used a mixture model and Chen (2007) used a logistic regression (with dose, sex, and inverse time as covariates) to analyse this data.

Given the empirical plots, it seems natural to analyse this data using the proportional hazards model assuming that the baseline hazard \( h_0(t) \) is unimodal. As the data is grouped, we need to modify our existing algorithms (for continuous data) to handle this setting. Surprisingly, the discrete/grouped model is more difficult to handle, as exact graphical solutions don’t necessarily exist.

References


