Shape–Constrained Maximum Likelihood Estimation for the Proportional Hazards Model

RIHONG HUI
huihong@yorku.ca

Department of Mathematics and Statistics
York University

HANNA JANKOWSKI
hkj@mathstat.yorku.ca

Introduction

We consider the classical proportional hazards model

$$h(t|z) = e^{\beta^T z}\, h_0(t),$$

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 then that we observe t_i (event time), z_i (vector of covariates), and δ_i (equals 0 if the observation was censored) for each of n subjects, and let H_0 denote the cumulative baseline hazard. The full likelihood is

$$\begin{aligned}\mathcal{L}(\beta, h_0) &= \prod_{i=1}^{n} h(t_i|z_i)^{\delta_i}(1 - F(t_i|z_i)) \\ &= \left\{\prod_{i=1}^{n} e^{\delta_i \beta^T z_i}\right\} \times \exp\left\{-\sum_{i=1}^{n} e^{\beta^T z_i} H_0(t_i)\right\} \times \left\{\prod_{i=1}^{n} h_0(t_i)^{\delta_i}\right\}.\end{aligned}$$

Maximising the likelihood is the same as minimising the function ($-\log \mathcal{L}$)

$$\varphi(\beta, h_0) = \sum_{i=1}^{n} e^{\beta^T z_i} H_0(t_i) - \sum_{i=1}^{n} \delta_i \log h_0(t_i) - \sum_{i=1}^{n} \delta_i \beta^T z_i,$$

which is strictly convex in all parameters.

The Algorithm

SET initial value for $\widehat{\beta} = \beta_{\text{ini}}$

  find $\widehat{h}_0 = \text{argmin}_{h_0 \in \mathcal{C}}\, \varphi(\widehat{\beta}, h_0) = \text{argmin}_{h_0 \in \mathcal{C}}\, \varphi_1(\widehat{\beta}, h_0)$

  set $\varphi_{\text{new}} = \varphi(\widehat{\beta}, \widehat{h}_0)$

  WHILE $\delta < \varepsilon$ REPEAT

    set $\varphi_{\text{old}} = \varphi_{\text{new}}$

    find $\widehat{\beta} = \text{argmin}_{\beta \in \mathbb{R}^k}\, \varphi(\beta, \widehat{h}_0)$

    find $\widehat{h}_0 = \text{argmin}_{h_0 \in \mathcal{C}}\, \varphi(\widehat{\beta}, h_0)$

    set $\varphi_{\text{new}} = \varphi(\widehat{\beta}, \widehat{h}_0)$

    set $\delta = |\varphi_{\text{new}} - \varphi_{\text{old}}|$

  END WHILE LOOP

To find $\widehat{\beta} = \text{argmin}_{\beta \in \mathbb{R}^k}\, \varphi(\beta, \widehat{h}_0)$, we use the Newton-Raphson algorithm. The solution to $\widehat{h}_0 = \text{argmin}_{h_0 \in \mathcal{C}}\, \varphi(\widehat{\beta}, h_0)$ 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 indicies such that $\delta_i = 1$. For $j = 1, \ldots, m$, define

$$s_{i_j} = \sum_{j=0}^{i_j - 1} \left(\sum_{l=j+1}^{n} e^{\widehat{\beta}^T z_l} \right) (t_{j+1} - t_j).$$

Then $\widehat{h}_0$ is piecewise constant with jumps restricted to $t_{i_1}, \ldots, t_{i_m}$, and the height of the hazard on $(t_{i_k}, t_{i_{k+1}}]$ is found as the inverse of the slope on $(k, k+1]$ of the greatest convex minorant of $\{(0, 0,), (1, s_{i_1}), \ldots, (m, s_{i_m})\}$.

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 $\mathbb{I}_{[0,\tau]}(t)$. As τ varies, these functions form a basis for any decreasing function. Since the score must be equal to zero at $\widehat{h}_0$, 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.

Examples

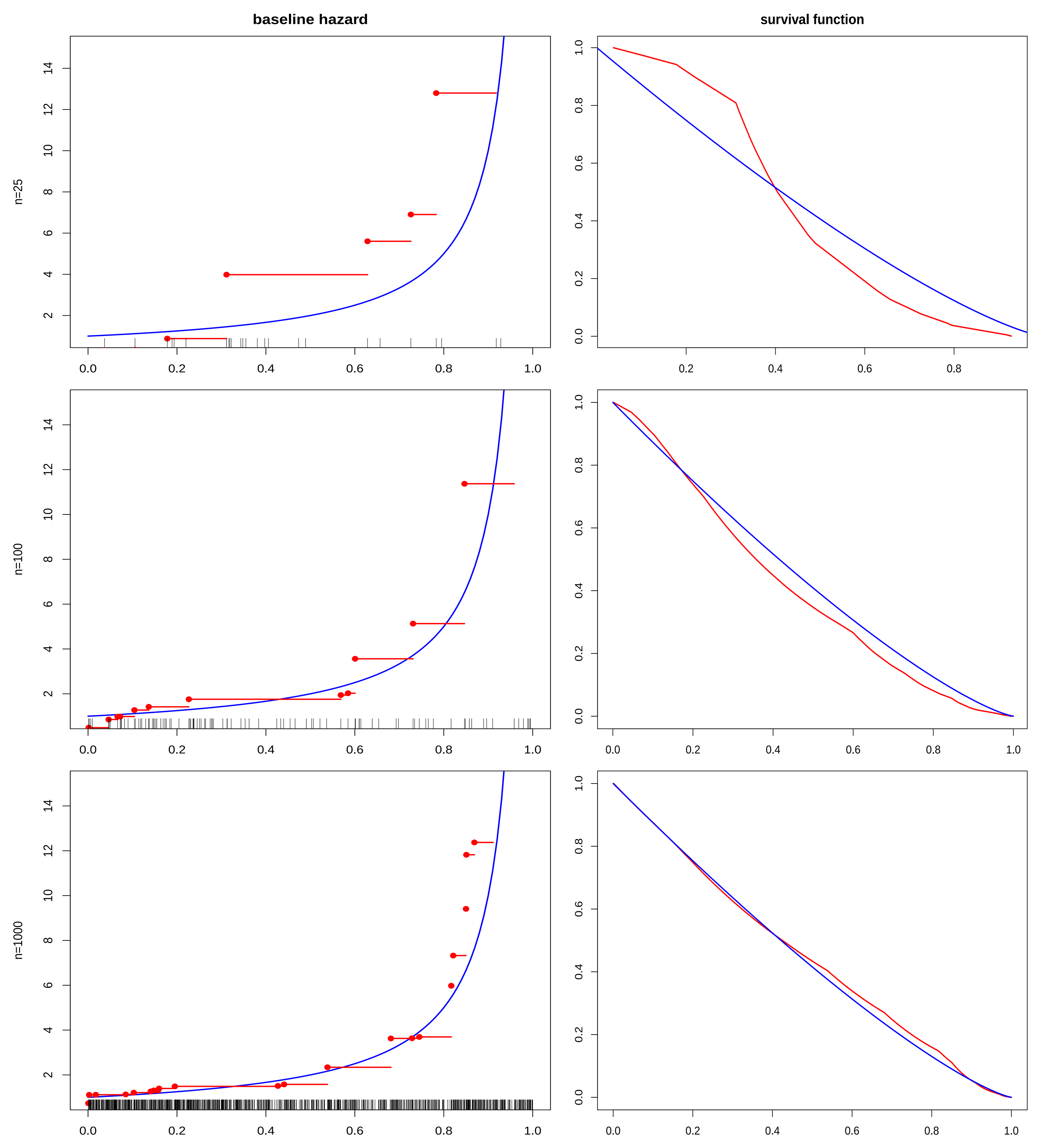

Fig. 1: Examples of the estimator for the increasing assumption when sampling from the uniform distribution without censoring. The baseline hazard is shown on the right, and the survival function at the average covariate is shown on the left. The blue line is the truth, and the estimator is shown in red. From top to bottom the sample sizes are $n = 25, 100,$ and 1000.

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_1, z_2 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.

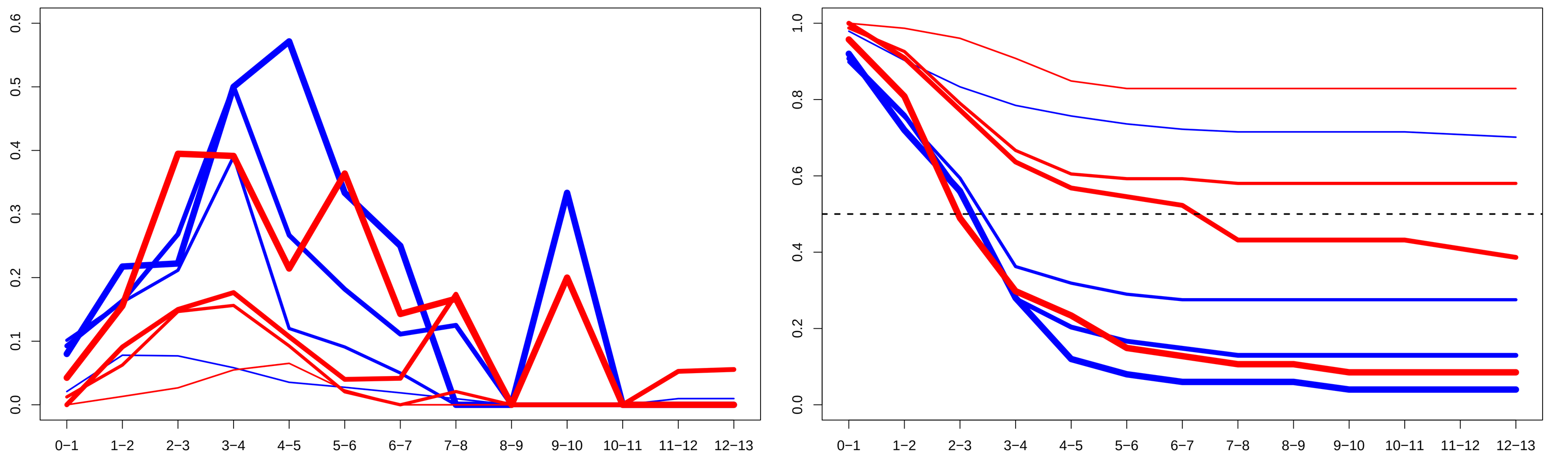

Fig. 2: Empirical hazard (left) and survival (right) functions for the data for females (red) and males (blue). The wider the line the higher the dose level.

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

Chen, D. G. (2007). Dose-time response cumulative multinomial generalized linear model. *J. Biopharm. Statist.* **17** 173–185.

Cox, D. R. (1972). Regression Models and Life Tables. *JRSSB* **34** 187–220.

Grenander, U. (1956). On the theory of mortality measurement. Part II. *Skand. Actuar.* **39** 125–153.

Hewlett, P. S. (1974). Time from dosage to death in beetles *Tribolium castaneum*, treated with pyrethrins or DDT, and its bearing on dose-mortality relations. *Journal of Stored Product Research* **10** 27–41.

Mykytyn, S. W. and Santner, T. J. (1981). Maximum likelihood estimation of the survival function based on censored data under hazard rate assumptions. *Comm. Statist.–Theor. Meth.* **10** 1369–1387.

Pack, S. E. and Morgan, B. J. T. (1990). A mixture model for interval-censored time-to-response quantal assay data. *Biometrics* **46** 749–757.