Homework 11

Problem 1

A data set (x_i, y_i), for $i = 1, \ldots, 50$, is available in "HW11_data_1.csv" for the logistic regression model;

$$P(Y_i = j \mid \beta, x_i) = \frac{e^{x_i \beta_j}}{1 + \sum_{j=1}^{3} e^{x_i \beta_j}}, \quad j = 1, 2, 3,$$

and

$$P(Y_i = 4 \mid \beta, x_i) = \frac{1}{1 + \sum_{j=1}^{3} e^{x_i \beta_j}}.$$

(1) (10 pts) Write down the log–likelihood function

$$L(\beta) = \log \prod_{i=1}^{n} P(y_i \mid \beta, x_i).$$

(2) (10 pts) Find the expressions of the partial derivatives

$$\frac{\partial L}{\partial \beta_j} \quad \text{and} \quad \frac{\partial^2 L}{\partial \beta_j \partial \beta_k}$$

for $j = 1, 2, 3$ and $k = 1, 2, 3$.

(3) (10 pts) Using the partial derivatives just found, write and run a Newton–Raphson algorithm to obtain the maximum likelihood estimator $\widehat{\beta}$. State the algorithm and the final result.

(4) (10 pts) Find the predictive probabilities for y with a new predictor at $x = \bar{x}$.

(5) (10 pts) Hence, what would be the predicted outcome for y at this x.

Problem 2

A data set (x_i, t_i) for $i = 1, \ldots, 10$ (in "HW11_data_2.csv") is concerned with a survival regression model, whereby

$$h(t \mid x) = h_0(t)\, e^{x\beta},$$

where $h_0(t)$ is a baseline hazard function and $h(t \mid x)$ is the hazard function for an individual with predictor variable x. In effect, the baseline hazard is the hazard for an individual with $x = 0$. This is also known as a proportional hazards model. Recall that if $f_0(t)$ is the baseline density function for survival time T then

$$h_0(t) = \frac{f_0(t)}{S_0(t)} \quad \text{where} \quad S_0(t) = \int_t^{\infty} f_0(s)\, ds.$$

(1) (10 pts) If $f_0(t) = \theta \exp(-t\theta)$ for $t > 0$ for some $\theta > 0$, find $h_0(t)$.

(2) (10 pts) Hence, write down $f(t \mid x, \theta, \beta)$ corresponding to the hazard function $h(t|x)$.

(3) (10 pts) Following on from part (2), write down the expression of the log–likelihood function in terms of the data and (θ, β).

(4) (10 pts) Assuming $\theta = 1$, find the maximum likelihood estimator for β; i.e. $\widehat{\beta}$.

(5) (10 pts) Use parametric Bootstrap to get an approximation of the variance of $\widehat{\beta}$, and find the appropriate normal approximation to the distribution of $\widehat{\beta}$.