Homework 6

ISyE 6420

Spring 2022

Use of unsolicited electronic and printed resources is allowed except the communication that violates Georgia Tech Academic Integrity Rules (e.g., direct communication about solutions with a third party, use of HW-solving sites, and similar). [1]

1. Rinderpest Virus in Rabbits with Missing Data. Temperatures (`temp`) were recorded in a rabbit at various times (`time`) after the rabbit was inoculated with rinderpest virus (the data modified from Carter and Mitchell, 1958). Rinderpest (RP) is an infectious viral disease of cattle, domestic buffalo, and some species of wildlife; it is commonly referred to as cattle plague. It is characterized by fever, oral erosions, diarrhea, lymphoid necrosis, and high mortality.

Time after injection (`time` in hrs)	Temperature (`temp` in ° F)
24	102.8
32	104.5
48	106.5
56	107.0
NA	107.1
70	105.1
72	103.9
75	NA
80	103.2
96	102.1

(a) Using an MCMC modeling library such as BUGS or PyMC and properly accounting for the missing data, demonstrate that a linear regression with one predictor (`time`) gives relatively low Bayesian R^2. What are estimators of the missing data? Does the 95% Credible Set for the slope contain 0? Comment on how you chose your priors and the results of your model.

(b) Include `time2` (squared time) as the second predictor, making the regression quadratic in variables, but still linear in coefficients. What are the estimators of missing data? Do the 95% Credible Sets for parameters in the quadratic model contain 0? Comment on how you

[1] Course Material for ISyE6420 by Brani Vidakovic is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

chose your priors and the results of your model. Is the Bayesian R^2 better or worse than the model from 1(a)?

Hint: If using BUGS, it is recommended to do the modeling in (a) and (b) in two separate BUGS programs. The linear regression is not good for part (a). Given the missing data, the estimator for σ^2 is large, and $SSE = (n-p) \cdot \sigma^2$ is bigger than SST, making the R^2 negative. For the single predictor model, please consider BR2=max(0, 1-SSE/SST) in your BUGS code. The quadratic regression is fine without any modification of BR2.

2. Bladder Cancer Data. An exercise in the book Pagano and Gauvreau (2000) [2] features data on 86 patients who after surgery were assigned to placebo or chemotherapy (thiopeta). Endpoint was the time to cancer recurrence (in months).

Variables are: `time`, `group` (0 - placebo, 1- chemotherapy), and `observed` (0 - recurrence not observed, 1 - recurrence observed). This data is given in files `bladerc.csv|dat|xlsx`. Data are given in WinBUGS format `bladderBUGS.csv|dat|xlsx`. The starter file `bladderc0.odc` contains data and also initial values for parameters and censored observations. Students should understand that these formats are equivalent, and be able to convert one into the other as needed.

Assume that observed times are exponentially distributed with the rate parameter λ_i depending on the covariate `group`, as

$$\lambda_i = \exp\{\beta_0 + \beta_1 \times \text{group}_i\}$$

After β_0 and β_1 are estimated, since the variable `group` takes values 0 or 1, the means for the placebo and treatment times become

$$\begin{aligned} \mu_0 &= \frac{1}{\exp\{\beta_0\}} = \exp\{-\beta_0\} \\ \mu_1 &= \frac{1}{\exp\{\beta_0 + \beta_1\}} = \exp\{-\beta_0 - \beta_1\}, \end{aligned}$$

respectively. The censored data are modeled as exponentials left truncated by the censoring time. Use noninformative priors on β_0 anmd β_1.

(a) Is the 90% Credible Set for $\mu_1 - \mu_0$ all positive?

(b) What is the posterior probability of hypothesis $H : \mu_1 > \mu_0$?

(c) Comment on the benefits of the treatment (a paragraph).

[2] Bladder cancer data from M Pagano and K Gauvreau, "Principles of Biostatistics, 2nd Ed. Duxbury 2000. Chapter 21, Exercise 9, page 512.