STA2201 Winter 2024

Sample mid-term questions

1

Explain why the logit link is appropriate for a binomial generalized linear model.

2

A political election is being held next year between two candidates, Mr Green and Ms Blue. Ms Blue is the current sitting member and is up for re-election.

You have data from a representative survey with 1,000 respondents, where the main question asked was "if the election were held tomorrow, who would you vote for?". You also have the following information on the respondents

- age (binary category, <35, 35+)

- their location of residence (there are 4 different districts)

- who they voted for last election.

It is also known that people don't tend to change their voting preferences, i.e. it's relatively likely that they will vote for who they did last time.

a) Describe how you would build a model in a frequentist framework to study the association between voting for Ms Blue and voters age, location of residence and past voting behaviour. You should define:

- The outcome of interest and the likelihood
- The regression model for the outcome of interest

b) Describe how you would build a regression model in a Bayesian framework to study the association between voting for Ms Blue and voters age, location of residence and past voting behaviour. You should define

- The outcome of interest and the likelihood
- The regression model for the outcome of interest
- The priors on any parameters to be estimated in the model.

c) Based on a regression model similar to that described above, it was estimated that the probability of Ms Blue winning the election was 0.65 with a 95% uncertainty interval of (0.57, 0.69).

- Interpret this uncertainty interval as a frequentist confidence interval.
- Interpret this uncertainty interval as a Bayesian credible interval (you may assume it has been calculated based on taking the 2.5th and 97.5th quantiles of the posterior distribution).
- Comment briefly on the intuition of each interpretation and how it relates to the context of predicting an election.

3

Gibbs sampling is more efficient than the Metropolis algorithm. How does it achieve this extra efficiency? Are there any limitations to the Gibbs sampling strategy?

4

The posterior predictive distribution gives

a) predictions of data y before any data are observed

b) predictions of posterior parameters θ after data are observed

c) predictions of data y after data are observed

d) predictions of posterior parameters θ before any data are observed

5

In Bayesian inference, parameters θ (circle all that apply)

a) are assumed to be fixed but unknown values to be estimated

b) are assumed to be random variables

c) can be assigned a prior probability distribution to encode our knowledge before seeing any data

d) can be estimated using a posterior predictive distribution