MTH5131 Actuarial Statistics

Coursework 7

To be turned in by 9:00 Monday 15 April 2024

This coursework counts 20% of your final mark. You should work alone on this course-work. You must upload a single R ontaining your solutions by 9:00 Monday 15 April 2024 to get credit for your work. Your R script should be well commented and it should be easy to see from your script what are your answers to the questions.

Exercise 1. This exercise is about the Normal/Normal model, in which the conditional distribution of data samples X is $\mathsf{N}(\theta, \sigma_1^2)$ and the prior distribution of θ is $\mathsf{N}(\mu, \sigma_2^2)$, where μ, σ_1 and σ_2 are known.

1. Suppose that $\mu = 5$, $\sigma_1 = 0.5$, and $\sigma_2 = 1$.
   (a) Using a seed of 60 and formula (5) of the typed lecture notes, find the 10 Bayesian estimates of θ under quadratic loss each time using samples of size $n = 7$. Store the estimates of θ in a vector theta1.mean and the values of $\bar{x}$ in another vector xbar1.vec.
   (b) Why is it unnecessary to consider either absolute error loss or all-or nothing loss for the Normal/Normal model?
   (c) Make histograms of theta1.mean and xbar1.vec. What do you observe?
   (d) Calculate and display the credibility factor Z. Were the samples given much credibility?
   (e) Show that formula (6) of the lecture notes gives the same estimates of θ as formula (5) by subtracting the two vectors of estimates, and making a vector of length 10 which contains FALSE whenever the absolute error of the difference of the estimates is less than 0.01.
2. Suppose that $\mu = 5$, $\sigma_1 = 3$, and $\sigma_2 = 1$.
   (a) Using a seed of 60, find the 10 Bayesian estimates of θ under quadratic loss each time using samples of size $n = 7$. Store the estimates of θ in a vector theta2.mean and the values of $\bar{x}$ in another vector xbar2.vec.
   (b) Make histograms of theta2.mean and xbar2.vec. What do you observe?
   (c) Calculate and display the credibility factor Z. Were the samples given much credibility?
3. Explain the difference in Z in 1(d) and 2(c).
4. Explain the difference in your answers to 1(c) and 2(b).

Exercise 2. The file CreditCard.csv contains cross-section data on the credit history for a sample of applicants for a type of credit card.

1. Read CreditCard.csv into the data frame DF1.
2. Make a second data frame DF2 just containing card, age, income and dependents. Here are the descriptions of these four variables.

   ```
   card is a factor. Was the application for a credit card accepted?

   age is in years plus twelfths of a year.

   income is yearly income (in USD 10,000).

   dependents isumber of dependents.
   ```

   An entry of the vector card in DF2 should be 1 whenever the corresponding entry of the vector card in DF1 is "yes" and be 0 whenever the vector whenever the corresponding entry of the vector card in DF1 is "no".
3. We are going to use the other variables to predict card. What family of distributions do you think is suitable for modelling this data?
4. Starting from the linear predictor with all three variables age, income and dependents and all their interactions, use the Backwards process to find the best model. Use the chi square test at every step. (I noticed that to use the chi square test, I had to use type="Chisq" and not type="chi", as it says to do in the lecture notes.
5. Find confidence intervals at the 95% level for the coefficients of the models you obtained. Are all coefficients significant? What is the most significant non-intercept coefficient?
6. Make a new model by adding the square of the varible with the most significant linear coefficient to the one found by the backwards method. Decide whether this model is better by looking at the AIC.
7. Check the deviance residuals of the model that you decided fit the data best. Are they satisfactory?
8. With your choice of the model that you decided fit the data best, predict the probability that an application for a credit card will be accepted if age=23.25000, income=4.3200, and dependents=2.