The University of British Columbia

Data Science 581 Modelling and Simulation II

Lab 1

This is a set of group exercises. Please work with the same group members that you worked on to prepare your presentation in DATA 543.

Procedure:

1. Think about the following questions on your own, before the lab, and provide answers as well as you can.
2. Meet with your group in the lab to discuss the questions for 45 minutes.
3. The remainder of the lab will be a discussion of these questions led by the TA.
4. Write up answers to the questions for submission to Canvas.

Questions:

1. I simulated a normal random variate, but I have forgotten whether it was from a standard normal distribution or from a normal distribution with mean 0 and standard deviation 0.5 or from a normal distribution with mean 1 and standard deviation 2.5. If the value was 1.2, calculate the maximum likelihood estimates of μ and σ.
2. Returning to the previous question, suppose I remember that the standard deviation is actually 2.5, but I am not sure that the mean is really 1. Can I use maximum likelihood estimation to estimate μ in this case? If so, what should my estimate of μ be?
3. Suppose I really have no idea what μ and σ should be. Can I still estimate them? If not, can I estimate μ?
4. Suppose there are two ways of experimentally determining a quantity μ. Both produce errors which have mean 0, but the first way has an error standard deviation of 20, and the second way has an error standard deviation of 50. If you have a budget of \$1000, and independent measurements cost \$10 each for the first technique, and only \$2.50 each for the second technique, which technique should be used in order to get the best estimate of μ? What does this tell you about sample size: is bigger always better?
5. Consider the logistic function
$$L(x) = \frac{e^x}{1 + e^x}.$$
   (a) Does $L(x)$ satisfy the properties of a cumulative distribution function?
   (b) In the logistic regression model, we have $P(Y = 1|x) = L(\beta_0 + \beta_1 x)$. Suppose we replace $L(x)$ with the standard normal CDF, often denoted by $\Phi(x)$, giving us a *probit* model:
$$P(Y = 1|x) = \Phi(\beta_0 + \beta_1 x).$$
      i. Suppose $\beta_0 = 2$ and $\beta_1 = 3$. Find $P(Y = 1|x)$ under both the logit and probit models when
         A. $x = 0$
         B. $x = -1$
         C. $x = 0.5$
      ii. Under the conditions of the preceding part, for what value of x is $P(Y = 1|x) = 0.5$?
   (c) Reconsider the missile success data of `p13.1` from the *MPV* package. If you use `family = probit` in the call to `glm()`, how does the output change?