MATH20802 Computer Lab 1

Term Week 5

The topic of this computer lab is to explore maximum likelihood inference in R.

Note that for this tutorial the solutions are provided along with the questions. The reason is to accommodate different levels of prior programming experience:

1) If you are not very experienced in R please study the example solutions carefully and run them in R yourself. Try to understand what each command is doing so that you are able to run similar analyses for a different model yourself!

2) If you do have some prior experience in R programming please try to code the solutions in R yourself first before looking up the answers!

Recommended books for learning how to program in R

1. G. Grolemund and H. Wickham. 2017. R for data science. https://r4ds.had.co.nz/ (see also https://www.tidyverse.org/)
2. R. D. Peng. 2016, R Programming for Data Science. https://bookdown.org/rdpeng/rprogdatascience/
3. H. Wickham. 2018. Advanced R. https://adv-r.hadley.nz/
4. Y. Xie, J. J. Allaire, G. Grolemund. 2019. R Markdown: The Definitive Guide. https://bookdown.org/yihui/rmarkdown/
5. K. Browman. hipsteR (old-style R vs. new style R). https://kbroman.org/hipsteR/

Formulas:

Poisson distribution:

If a random variable $X = 0, 1, 2, \ldots$ has this distribution we write $X \sim Pois(\lambda)$. The corresponding probability mass function is:

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

for $\lambda > 0$. The expected value is $\text{E}(X) = \lambda$ and the variance is $\text{Var}(X) = \lambda$.

Tasks:

1. Work out the details for maximum likelihood estimation of the parameter of the Poisson distribution:
   i. Write down the log-likelihood function and score function.
   ii. Derive analytically the MLE.
   iii. Compute the expected Fisher information and the asymptotic variance of the MLE.

2. Following the R example for the exponential model in the lecture implement a complete numerical maximum likelihood analysis for the Poisson model:
   i. Create an R function for the log-likelihood function that can be applied to a arbitrary data set. You can do this in two ways:
      - either you use the built-in R function dpois() for the probability mass function (don't forget to Vectorize() the resulting function!) or
      - you use the analytic expression obtained in Question 1.
   ii. Set n=10 and lambda.true = 4 and simulate example data x using the function rpois().
   iii. Create a new function (e.g. loglkl2 that depends only on the parameter λ and includes the data.
   iv. Plot the log-likelihood function between 1 and 10 given the data x.
   v. Find the MLE, observed Fisher information and standard deviation of the MLE and compare the numerica and analytic results.
   vi. Find the 95% normal-based confidence interval.
   vii. Find the 95% likelihood-based confidence interval.
   viii. Plot both intervals.

3. Use maximum likelihood to numerically estimate the mean vector of the multivariate normal distribution:
   i. Use the R package "mnormt" to simulate data from a d-dimensional multivariate normal distribution. Specifically, set $d = 10$, the true mean vector to $(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$ (1:d in R) and the scale (covariance) parameter (varcov) to the diagonal matrix (diag(d) in R), and then use the function rmnorm() to simulate $n = 1000$ data points.
   ii. Write a function to compute the negative log-likelihood for the mean parameter vector using dmnorm().
   iii. Optimise (minimise) the negative log-likelihood function using optim() and three different methods: Nelder-Mead, BFGS, and CG. Any conclusions?
   iv. Compute numerically the observed Fisher information matrix.