

Computer-Intensive Statistics

Assessment Exercises

Students should talk to their supervisors to find out whether or not their department requires this work as part of any formal accreditation process. It is anticipated that departments will decide on the appropriate level of assessment locally, and perhaps select a pertinent subset of these exercises (it is not intended that individual students should attempt all of them).

Ex 1 There is another type of bootstrap-based confidence interval not discussed in the lectures, called *ABC*. Use the references to find out about it, and write up a description of the method and a comparison of its strengths and weaknesses (compared to the methods included in the lecture notes).

Ex 2 Practical 9 analysed only part of the Australian AIDS survival data. Extend the analysis to the complete dataset: you will need to take ‘state’ into account (for a classical analysis see Venables & Ripley, 2002, 379ff), and should allow for many minutes of computation.

Ex 3 Consider a probit model for LD50 for the following data

dose	animals	deaths
50	5	1
100	10	4
200	5	2
300	10	6
2000	5	4

with a suitable vague prior. Find a suitable HPD region for LD50. How does it compare to a logit model and to a classical GLM analysis such as that shown for the budworm data in Venables & Ripley?

Ex 4 If you have not yet done so, finish the practical exercises 7 and 8 on the SAT-V coaching data—this exercise builds on those exercises.

(a) We saw evidence in the plots that some aspects of this problem converge much more slowly than others. See Gelman *et al* (2004) for some discussion, and especially sections 11.8 and C.3/4 for palliative measures using auxiliary variables.

Either write 1–2 pages on the methodology *or* program up in R or BUGS/JAGS one of these schemes and report on how it seemed to help the apparent rate of convergence (or not!).

- (b) ν , the number of degrees of freedom of the t_ν model for the school means, is a rather awkward quantity for which to summarize uncertainty. See if you can come up with something more convincing than the default plots in CODA.
- (c) The convenience prior chosen for ν was uniform on $[0, 1]$. This is unlikely to be a realistic representation of your beliefs (it isn't for mine). Come up with something more plausible and either modify the MCMC sampler or use SIR to study the effect of the change of prior.

Ex 5 (For those with computer-intensive projects.)

Think about how you could organize the algorithms that are used in your field to make use of multiple CPUs, either a cluster of conventional CPUs or of a GPU-based 'personal supercomputer'. Prepare a page or two (no more) of writeup for a non-specialist audience on which approaches are most promising in such an environment (and why).

Ex 6 (More ambitious, for those with previous experience of BUGS or JAGS.)

Both BRugs and JAGS support models with censoring (as does WinBUGS), so program assessment 2 in the BUGS dialect you use.

(There are examples in the BUGS examples manuals which will help you, e.g. those on 'mice' and 'kidney'. This is a large dataset for BUGS and you might want to experiment with a subset.)