Homework 10. Due by 11:59pm on Sunday 11/23.

Statistical computing on greatlakes

From homework 9, we understand the relevance of being able to implement embarrassingly parallel calculations. In the context of statistical computing in R, we are motivated to learn to use `foreach` on the greatlakes cluster. Despite the growing popularity of Python, R is currently the dominant language at the *Journal of Statistical Software*. Some recent packages come in both R and Python.

In this homework we carry out a simple example, but once you can run a simple parallel for loop on greatlakes you are prepared for most simulation experiments that arise in statistics PhD research. Work through the questions below and write brief answers to the following questions, by editing the tex file available at `https://github.com/ionides/810f25`, and submit the resulting pdf file via Canvas.

1. R and Python have a reputation of being convenient for data analysis but slow in some situations. Loops can be slow in these interpreted languages, though careful coding may avoid this, e.g.,
   `https://adv-r.hadley.nz/control-flow.html#common-pitfalls`
   `https://lectures.scientific-python.org/advanced/optimizing/`
   If run time starts limiting your research—i.e., whenever there is a numerical result you would like to see but you don't have the patience to wait for it to be computed—you have various options, including:

   (a) Use smaller examples.

   (b) Write the critical part of your code in C or C++. In R, this can make use of the C functions underpinning the R language
   (`https://cran.r-project.org/doc/manuals/R-exts.html#The-R-API`).
   Alternatively, the Rcpp package facilitates combining R with compiled C++ code.

   (c) Work on rewriting your code to make it more efficient.

   (d) Run your code on a more powerful machine, maybe moving from your laptop to greatlakes.

   Comment on situations when each approach might be suitable. Do you have advice, or relevant experiences?

   YOUR ANSWER HERE.

2. Follow the instructions at `https://ionides.github.io/810f25/gl/slides.pdf` to run a job testing foreach on greatlakes. Report back on any problems that arose.

   YOUR ANSWER HERE.

3. How do you explain the differences in results between running `test.sbat` and `test2.sbat`? You will need to inspect the code in `test.R` and `test2.R`.

YOUR ANSWER HERE.