Amazon EC2 assignment (presentation by Trevor Hefley)

STAT 992

Spring 2014

The purpose of this assignment is to use Amazon Elastic Compute Cloud (EC2) to reproduce parts of assignment #8. Because the EC2 requires an account with a valid credit card, this assignment is optional.

1. Create an Amazon Web Services account.
2. Using the pre-made Amazon Machine Images on Louis Aslett (<http://www.louisaslett.com/RStudio_AMI/>) login to RStudio (see video tutorial). Provide a screen shot showing that you have successfully logged into RStudio.
3. Create your own Amazon Machine Image and install R. I recommend using Windows Server 2012 (see video tutorial). Provide a screen shot showing that you have installed R onto your Amazon Machine Image.
4. The purpose of this problem is to reproduce parts of Assignment #1 and #8 using the EC2. Please note that you can use the code you created from assignment #8. You can use the Amazon Machine Image created in either #2 or #3.
   1. Simulate the R = 1000 data sets of size n = 20 in exactly the same manner as in the Assignment #1 answer key. Note that depending on how you do the parallel processing, you may or may not be using this exact data for the project.
   2. Complete the following table (B = 4999, α = 0.05):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Package | Instance | Number of cores | Number of Threads | Time in minutes |
| Parallel or  foreach (your choice) | t1.micro |  |  |  |
|  |  |
|  |  |
| Your choice |  |  |  |
|  |  |
|  |  |
| Your choice |  |  |  |
|  |  |
|  |  |

where the time in minutes corresponds to the actual Monte Carlo simulation part of your program. Set a seed of 8881 before EACH implementation of the parallel and foreach packages using clusterSetRNGStream().