Accelerating Research With a Personal Parallel Cloud

Kai Du

University of Queensland

MATLAB EXPO 2022

Overview

1. Introduction
2. Parallel Computing
3. MATLAB Toolbox
4. Wrap Up

Introduction

- Productivity is the key source of economic growth and competitiveness (of a country, a firm and even a person).
- My research focuses on the determinant of efficiency gain, such as the effect of the health care system on a country's productivity.
- I am also interested in how to estimate the change in productivity and efficiency

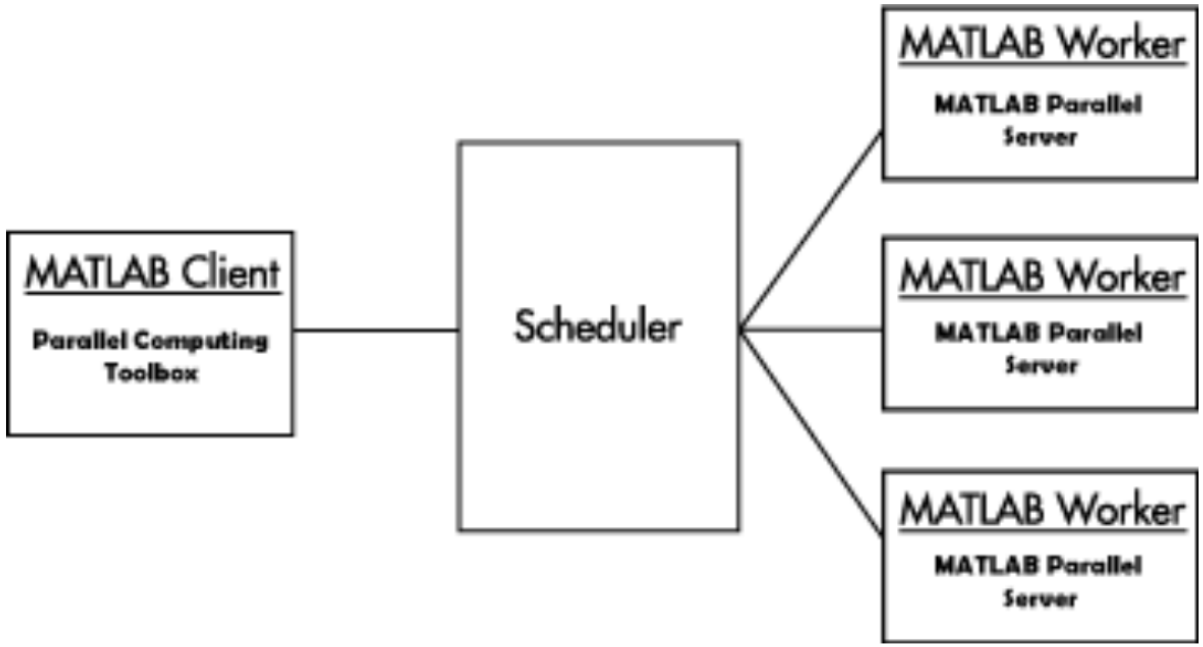
Monte Carlo Simulation

• A critical part of academic rigour is having a systematic and formally designed model that gives reliability and confidence to the inferences drawn.

• Monte Carlo (MC) simulation is an essential ingredient in my research, which needs a lot of computing power and is very time-consuming, especially when bootstrapping is used.

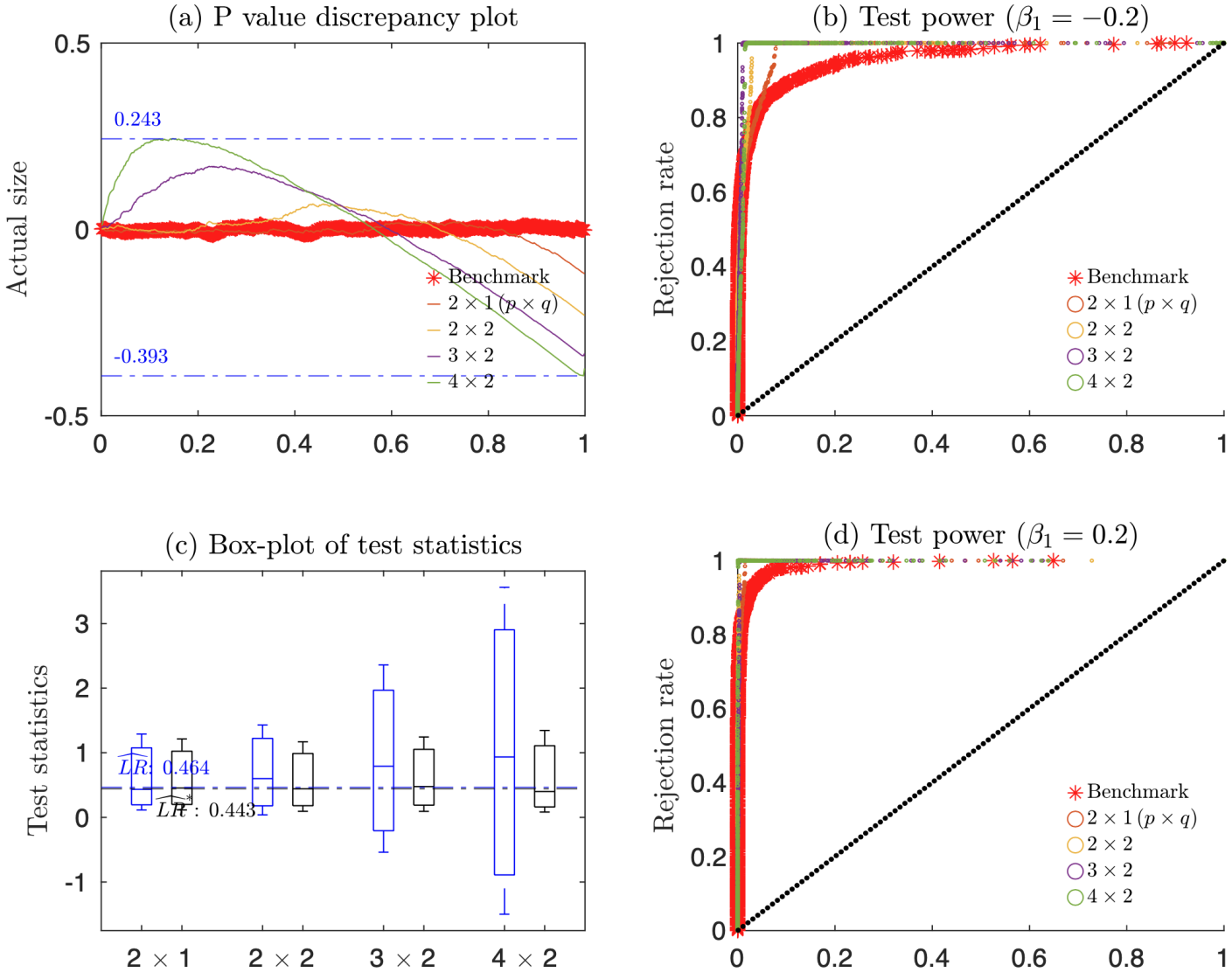
• **Parallel computing** is a natural solution since it can carry out many calculations simultaneously to save the computing time of Monte Carlo simulation and bootstrapping.

MATLAB Parallel Computing



- I run 1,000 MC trials and 1,000 bootstrapping iteration in each trial for an experience, which means 1k $\times$ 1k MATLAB *linprom*.
- Meanwhile, the sample size increased from $n = 100$ to $n = 1,000$.

An Example



- The computing time for this graph is 165.144 hours (or almost a week) using 48 virtual CPUs.

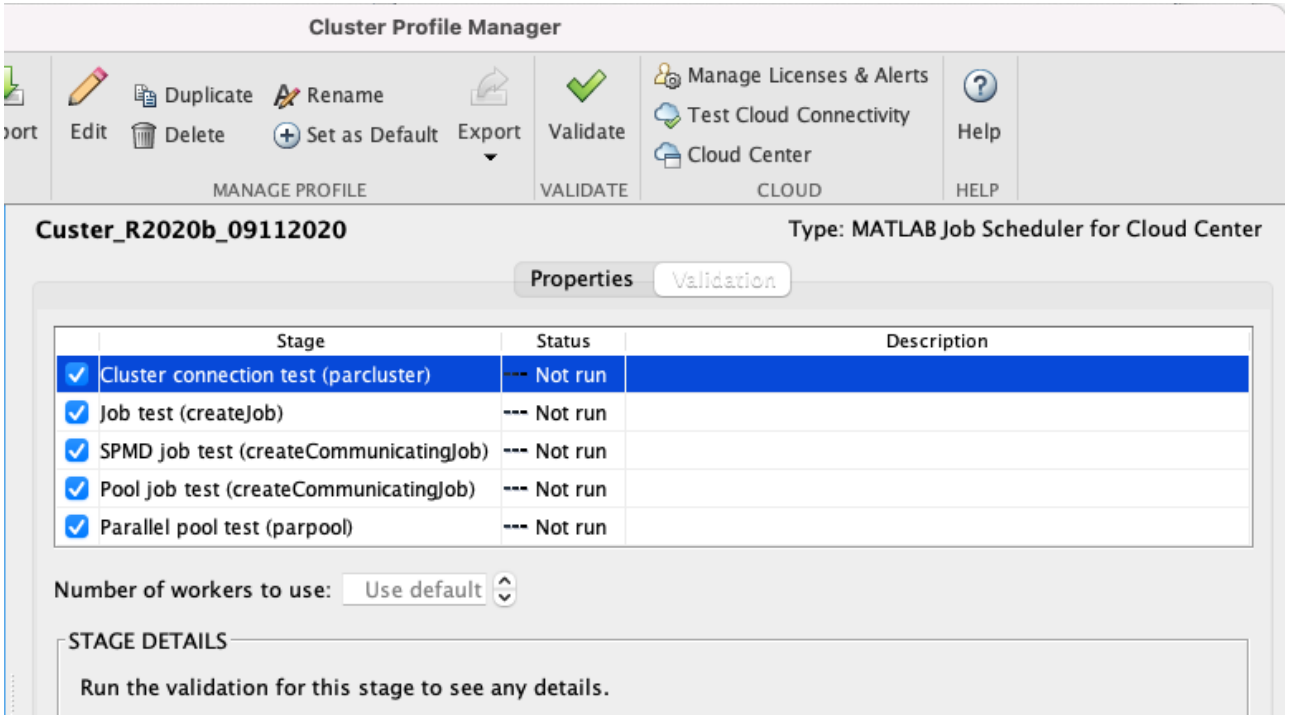
Batch Processing

```
%% Force files to attach
fName = dir('*.m');
myFiles = {fName(:).name};
%myFiles{end+1} = 'RegreVariable.mat';
```

```
%%
j = batch(c, 'DEA_TrN_Boot', ...
    'Pool', 22, ...
    'AttachedFiles', myFiles);
```

- I use batch jobs to off-load the execution of long-running computations in the background.
- *Also*, it supports local machine and cloud computing.

Parallel Computing Toolbox

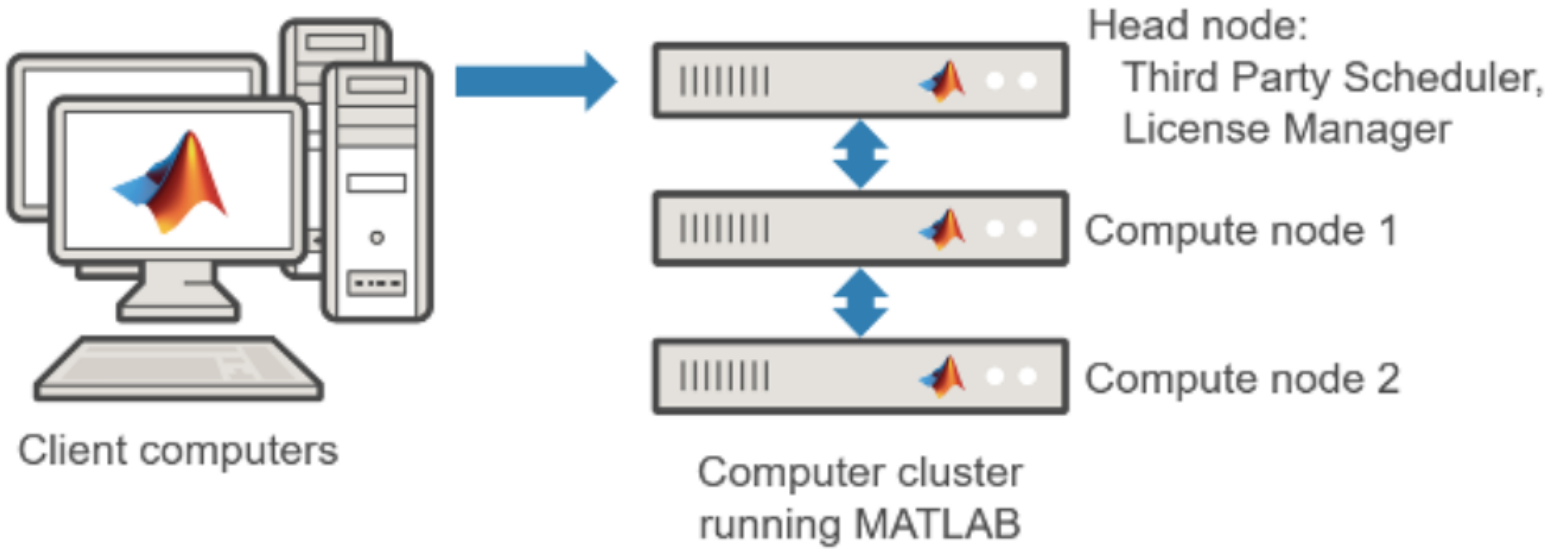


- The process to set up the cluster is very intuitive and there are several options for the user, such as local machine, Cloud Center on Amazon Web Services (AWS) and cluster using other computing facilities.

MATLAB Cloud Center

- It is straightforward to use the Cloud Center, but the cost of AWS instances is its disadvantage.
- The cost could be more than one hundred dollars when several instances are running for a day.
- It is a good solution for small projects but is not suitable for academic research.

Personal Cluster



• With a little bit of help from MathWorks, it is possible to set up a personal cluster using the computing facilities in Australia, such as National Computational Infrastructure (NCI) Australia, ARDC Nectar Research Cloud and Pawsey Supercomputing Centre. Most importantly, **zero financial burden**!

General Procedure

- **Head node**

Using the X server install the software of parallel server

Edit the start-up scripts to point to the server[1]

- **Worker nodes**

Using the X server install the client software

Edit the start-up scripts

[1] The scripts are available upon request (with the permission from the author and Mathworks).

Wrap Up

- With more and more businesses moving to the ‘cloud’, it’s important for researchers to embrace the new technology, so they don’t get stuck in the ‘digital stone age’.
- For our research, we also need to do a cost-benefit analysis of our precious time, financial burden, and outcome.
- The software of MATLAB is a useful tool but their service is more important than the software.

Acknowledgement and Thank you

- Dr. Peter Brady from Mathworks
- Mr. Bradley Horton from Mathworks
- Mr. Morgan Quinn from Mathworks
- UQ Development Fellowship