Stat4DS / Homework 03 - Part A

Pierpaolo Brutti

Due Thursday, January 31, 2019, 23:00 PM on Moodle

General Instructions

I expect you to upload your solutions on Moodle as a **single running** `R Markdown` file (`.rmd`) + its `html` output, named with your surnames.

You will give the commands to answer each question in its own code block, which will also produce plots that will be automatically embedded in the output file. Your responses must be supported by both textual explanations and the code you generate to produce your results. *Just examining your various objects in the "Environment" section of RStudio is insufficient – you must use scripted commands and functions.*

R Markdown Test

To be sure that everything is working fine, start `RStudio` and create an empty project called `HW1`. Now open a new `R Markdown` file (`File > New File > R Markdown...`); set the output to `HTML mode`, press `OK` and then click on `Knit HTML`. This should produce a web page with the knitting procedure executing the default code blocks. You can now start editing this file to produce your homework submission.

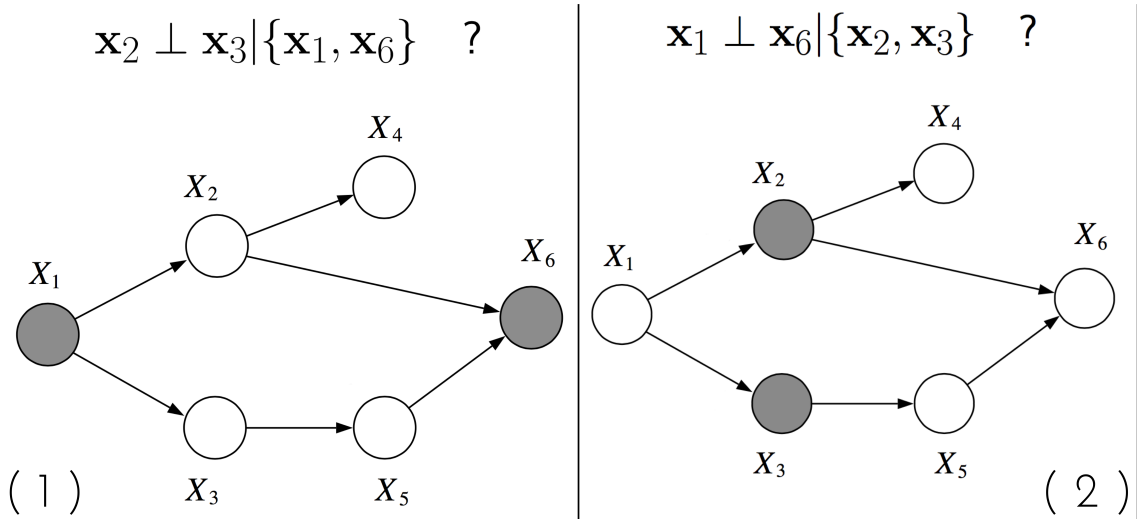
Please Notice

- For more info on `R Markdown`, check the support webpage that explains the main steps and ingredients: R Markdown from RStudio.
- For more info on how to write math formulas in LaTex: Wikibooks.
- Remember our **policy on collaboration**: *Collaboration on homework assignments with fellow students is **encouraged**. However, such collaboration should be clearly acknowledged, by listing the names of the students with whom you have had discussions concerning your solution. You may **not**, however, share written work or code after discussing a problem with others. The solutions should be written by **you**.*

Exercise I: Ultra-fast exercise

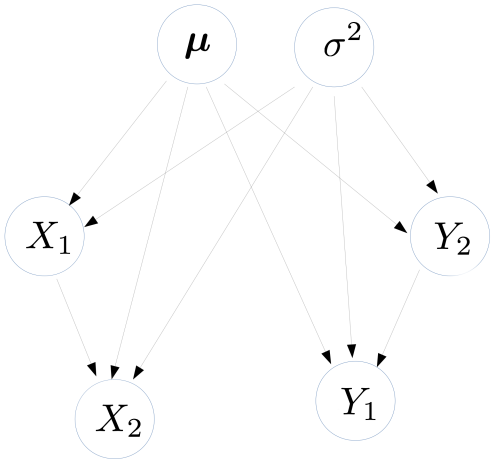
Look at the two DAGs below and use *d*-separation + Markov condition to check whether the indicated *conditional independence* relationships are satisfied.

Explain your answer and, for the conditional independence relation in (2), provide also a direct check by showing whether or not $p(x_6 \,|\, \mathbf{x}_{1:3}) \stackrel{?}{=} p(x_6 \,|\, \mathbf{x}_{2:3})$.



Exercise II: Sampling a DAG

Consider the following (Bayesian) model in DAG format



where, conditionally on $\boldsymbol{\mu} = [\mu_1, \mu_2]$ and σ^2, we assume that

- $\boldsymbol{X} = [X_1, X_2]^\intercal \sim \mathrm{N}_2(\boldsymbol{\mu}, \sigma^2 \mathbb{I}_2)$
- $\boldsymbol{Y} = [Y_1, Y_2]^\intercal \sim \mathrm{N}_2(\boldsymbol{\mu}, \sigma^2 \mathbb{I}_2)$

1. Use what we studied about conditional distributions of multivariate Normal vectors to write down in formula the joint distribution corresponding to this DAG.
2. Choose two suitable (prior) distributions for $\boldsymbol{\mu}$ and σ^2 and write the corresponding R code to simulate a sample of $n = 10000$ random vectors from the joint distribution by forward/ancestral sampling.
3. Show a suitable plot with the empirical distribution of each component of the $\boldsymbol{X}$ vector. What theoretical distribution are you approximately describing?