37457

Advanced Bayesian Methods

Professor Matt Wand

University of Technology Sydney

Major Goal for this Subject

By Week 6, tool you up to be able to do

ANY*

data analysis, no matter how complex,

– and understand the underlying maths.

* To be qualified later.

Path to Achieving Goal

- Draft book chapters of GRAPH THEORY AND STATISTICS by M.P. Wand.
- The R computing environment.
  Starting today!
- The JAGS inference engine and its interfaces with R, known as rjags.
  Starting today! (later will use a newer engine named Stan).
- IMPORTANT: no assumed knowledge of any of the above!

MARGINALISATION

and

CONDITIONAL MARGINALISATION

Connection to Assignment 2

Remember the Bob the DAG questions.

- Question 4 was on MARGINALISATION
- Question 5 was on CONDITIONAL MARGINALISATION

QUESTION: Why is (conditional) marginalisation important?

QUESTION: Why is (conditional) marginalisation important?

SOME ANSWERS:

- It is THE mathematical problem that has to be solved for Bayesian statistical inference, which is TOPIC 2 for this subject.

QUESTION: Why is (conditional) marginalisation important?

SOME ANSWERS:

- It is THE mathematical problem that has to be solved for Bayesian statistical inference, which is TOPIC 2 for this subject.
- Many modern Machine Learning algorithms (e.g. speech recognition, Internet searching, robot vision) require marginalisation over big probabilistic graphs.

Beyond Bob

The Bob the DAG questions had:

- A two-node DAGs.
- Binary random variables $\Longrightarrow$ simple Bernoulli distributions.

Beyond Bob

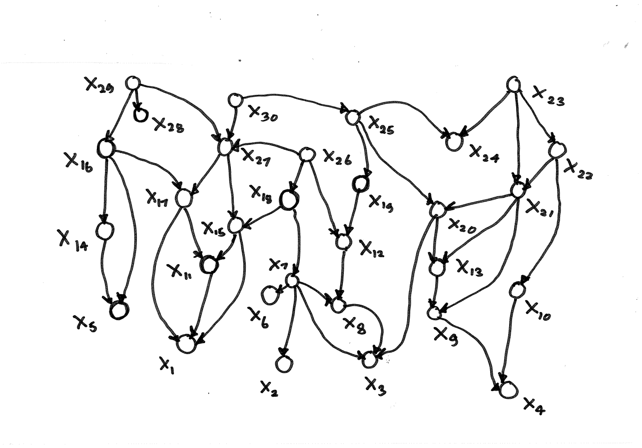
The Bob the DAG questions had:

- A two-node DAG.
- Binary random variables $\Longrightarrow$ simple Bernoulli distributions.

"Real" Statistics problems have DAGs with

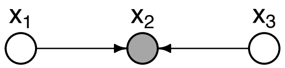
- nodes numbering in dozens, hundreds, thousands....
- more complicated (continuous) distributions.

Real Statistics (and Real Machine Learning)



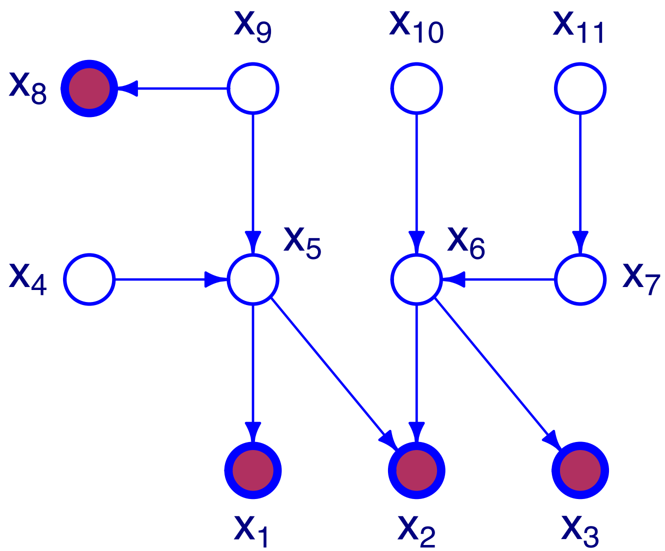
Project and look at marginalisation section of notes.

Innocent-Looking Example



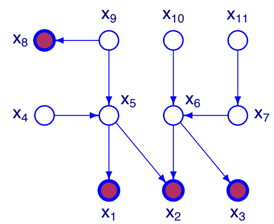
$$x_2|x_1, x_3 \sim N(x_1, 1/x_3), \quad x_1 \sim N(0,1), \quad x_3 \sim \text{Gamma}(1,1).$$

BUT EXACT MARGINALISATION NOT POSSIBLE!



Look at mathematics of innocent-looking example in notes.

Example Distributions on This DAG



$$\begin{aligned}
& x_1|x_5 \sim \text{Poisson}(3\,x_5), && x_2|x_5, x_6 \sim N(2x_6 + 5, 9/x_5), \\
& x_3|x_6 \sim N(x_6, 36), && x_4 \sim \text{Bernoulli}(0.37), \\
& x_5|x_4, x_9 \sim \text{Gamma}(x_4 + 3, 4x_9), && x_6|x_7, x_{10} \sim N(x_7 x_{10}, 16) \\
& x_7|x_{11} \sim \text{Bernoulli}(x_{11}), && x_8|x_9 \sim \text{Poisson}(x_9), \\
& x_9 \sim \text{Gamma}(4, 13), && x_{10} \sim N(0,1) \\
\text{and} \quad & x_{11} \sim \text{Beta}(9, 3). &&
\end{aligned}$$

Example Conditional Marginalisation Problem

$$p(x_6|x_1 = \mathring{x}_1, x_2 = \mathring{x}_2, x_3 = \mathring{x}_3, x_8 = \mathring{x}_8)$$

The required maths is:

$$\begin{aligned} p(x_6|x_1,x_2,x_3,x_8) &= \frac{p(x_1,x_2,x_3,x_6,x_8)}{p(x_1,x_2,x_3,x_8)} \\ &= \frac{\sum_{x_4=0}^{1}\int_0^\infty \sum_{x_7=0}^{1}\int_0^\infty \int_{-\infty}^\infty \int_0^1 p(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10},x_{11})\,dx_5\,dx_9\,dx_{10}\,dx_{11}}{\sum_{x_4=0}^{1}\int_0^\infty \int_{-\infty}^\infty \sum_{x_7=0}^{1}\int_0^\infty \int_{-\infty}^\infty \int_0^1 p(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10},x_{11})\,dx_5\,dx_6\,dx_9\,dx_{10}\,dx_{11}} \\ &= \text{lots of difficult integration (and some integrals may not be tractable).} \end{aligned}$$

Markov Chain Monte Carlo (MCMC) Alternative

This requires that we draw samples from the full conditionals:

$p(x_4|$rest of graph$)$

$p(x_5|$rest of graph$)$

$p(x_6|$rest of graph$)$

$p(x_7|$rest of graph$)$

$p(x_9|$rest of graph$)$

$p(x_{10}|$rest of graph$)$

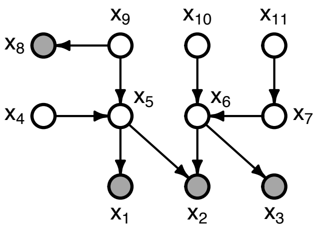
$p(x_{11}|$rest of graph$)$

MARKOV CHAIN MONTE CARLO

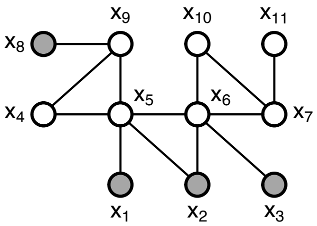
to the rescue!!! ...

GRAPH THEORY HELPS US TO SIMPLIFY THESE!

a probabilistic DAG



corresponding moral graph



full conditional density functions
functions of unshaded nodes

$p(x_4 \mid \text{rest}) = p(x_4 \mid x_5,x_9)$

$p(x_5 \mid \text{rest}) = p(x_5 \mid x_1,x_2,x_4,x_6,x_9)$

$p(x_6 \mid \text{rest}) = p(x_6 \mid x_2,x_3,x_5,x_7,x_{10})$

$p(x_7 \mid \text{rest}) = p(x_7 \mid x_6,x_{10},x_{11})$

$p(x_9 \mid \text{rest}) = p(x_9 \mid x_4,x_5,x_8)$

$p(x_{10} \mid \text{rest}) = p(x_{10} \mid x_6,x_7)$

$p(x_{11} \mid \text{rest}) = p(x_{11} \mid x_7)$

The rjags Computer Package

rjags

allows running of MCMC within the R computing environment.

All we have to do is specify the distributions that make up the DAG.

You get to learn some rjags before starting your lunch today! (Laboratory 1 – starting at around 11:05)

Run demoDAGandMCMC.Rs ...

The Stan Computing Environment

R version: rstan

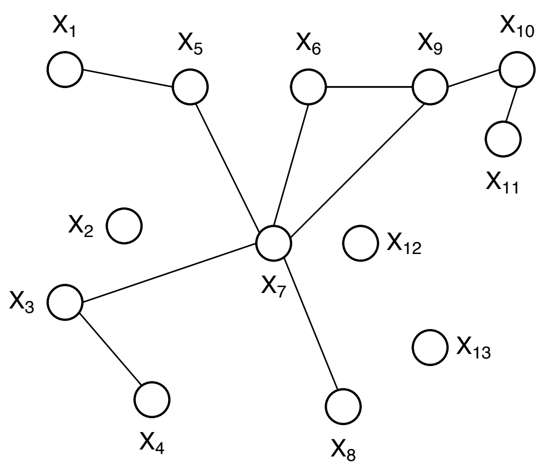
Since late 2013 also allows running of MCMC within the R computing environment.

Has started to replace rjags as main package of type, (but rjags still important for some models).

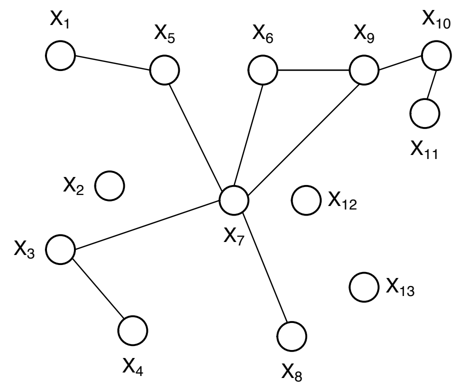
We will use rstan a lot in the second half of this subject.

One last sub-topic...

CONDITIONAL INDEPENDENCE THEOREMS

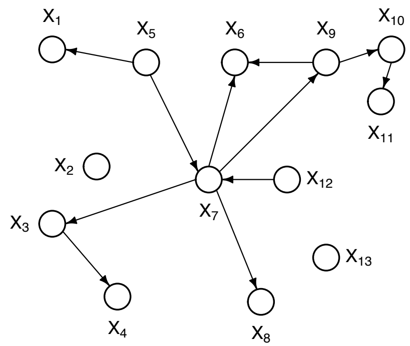


Class Exercise



Write down two other conditional independence statements for this undirected graph.

The DAG Case



PROBLEM: Is $x_3 \perp\!\!\!\perp x_{10} | \{x_7, x_9\}$?

Class 4 Location

Recall that during Weeks 1 – 6 we have

ROOM CHANGES EVERY WEEK!

Class 4 is in

Room 190, Level 4, Building 2

(opposite the quadrangle,
up the main Building 2 staircase,
on the Jones Street side of Level 4).

Laboratory 1

Time: 11:05–12:00

- For Windows users: a safe choice editor to use is WordPad. The editor NotePad can cause problems with Laboratory 1 (due to text-type files being prepared on a Mac computer).
- Upper-case versus lower-case.
- Must save files with exactly the same name as appears in Laboratory 1.
- Room CB07.02.015 table space problem: facing the front is not necessary for Laboratory 1. During the break we will rearrange the furniture.