37458

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.

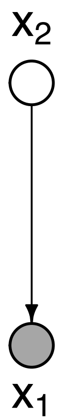
PROBLEM AVOIDANCE IN COMPUTING ASSIGNMENTS AND LABS

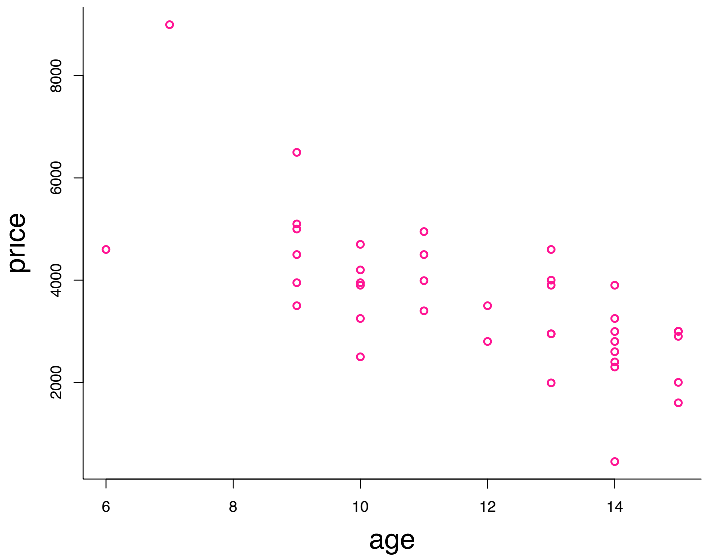
- RULE 1: Never use Microsoft Explorer to download a file.
  Use e.g. Mozilla Firefox of Google Chrome instead.
- RULE 2: Never use Notepad to open a file.
  Use e.g. WordPad instead.

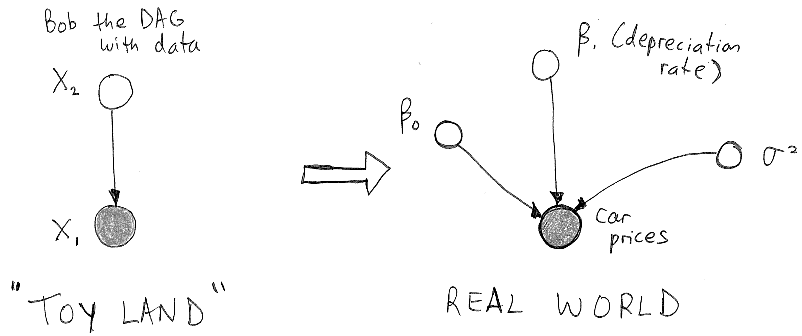
If Rule 1 or Rule 2 accidentally broken then better to delete files and start again.

From Toy Examples to Real Data Analysis

Bob the DAG with data







Bayesian Model For Mitsubishi Price/Age Data

$$\text{price}_i|\beta_0,\beta_1,\sigma^2 \;\sim\; N(\beta_0+\beta_1\,\text{age}_i,\sigma^2)$$

$$p(\beta_0),\quad p(\beta_1),\quad p(\sigma^2)$$

chosen to make ‘prior’ (pre-data) belief about these regression parameters non-informative.

The age values are treated as non-random here.

Interpretation of β_1:

$$\beta_1 = \text{annual depreciation rate of Mitubishi cars}$$

Posterior density function of β_1

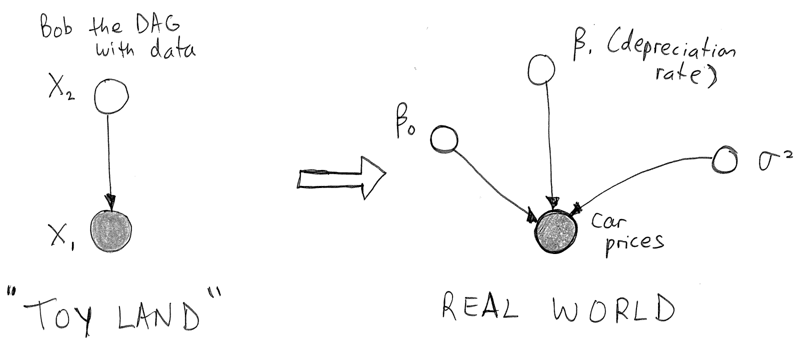
Needed:

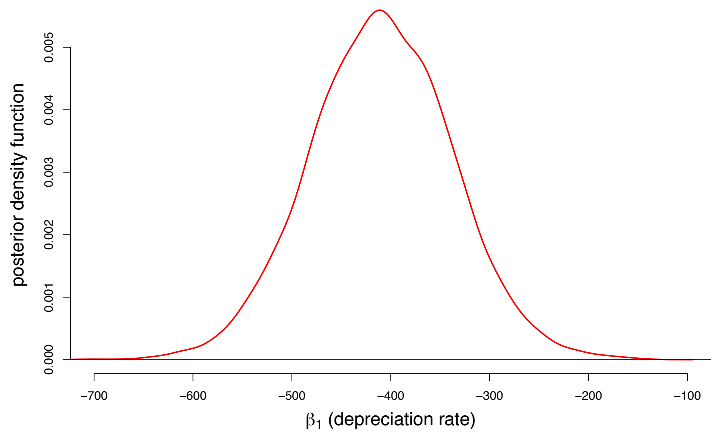
$$p(\beta_1|\text{data on car prices})$$

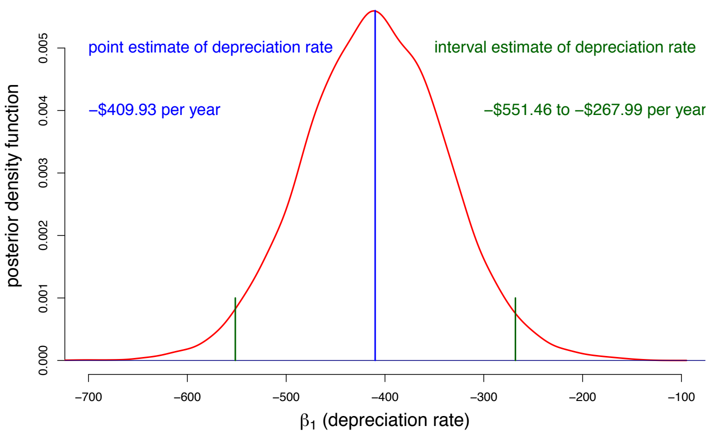
which is COMPLETELY ANALOGOUS to

$p(x_2|\mathring{x}_1)$ in Assignment 2

Show BRugs script and run.







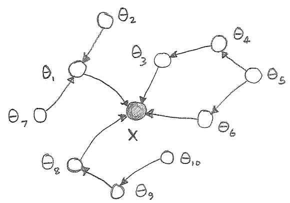
Bayesian Inference

Data: $\boldsymbol{x}$
Parameters: $\theta_1, \ldots, \theta_k$

Bayesian inference is based on the posterior density functions

$$p(\theta_1 | \, \boldsymbol{x}), \ldots, p(\theta_k | \, \boldsymbol{x})$$

But these are simply marginal conditional density functions as in Assignments 1 and 2.



Inverse Gamma Notation

$$x \sim \text{Inverse-Gamma}(A, B)$$

$$\iff \quad p(x) = \frac{B^A}{\Gamma(A)} x^{-A-1} \exp(-B/x), \quad x > 0.$$

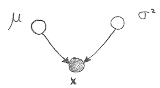
Bayesian Inference for Normal Data

Consider the Bayesian version of inference for a univariate normal sample:

$$x_1, \ldots, x_n \overset{\text{ind.}}{\sim} N(\mu, \sigma^2).$$

A Bayesian model is:

$$x_i | \mu, \sigma^2 \overset{\text{ind.}}{\sim} N(\mu \mathbf{1}, \sigma^2 \boldsymbol{I})$$
$$\mu \sim N(0, \sigma_\mu^2) \quad \sigma^2 \sim \text{Inverse-Gamma}(A, B)$$



(Jessica the DAG again!)

We need:

$$p(\mu | \boldsymbol{x}) \quad \text{and} \quad p(\sigma^2 | \boldsymbol{x}).$$

Posterior Distributions

$$p(\mu|\boldsymbol{x}) = \frac{e^{-\mu^2/(2\sigma_\mu^2)}\Gamma(A+\frac{n}{2})}{\sqrt{2\pi}\sigma_\mu(B+\frac{1}{2}\|\boldsymbol{x}-\mu\mathbf{1}\|^2)^{A+\frac{n}{2}}\int_0^\infty(\sigma^2+\sigma_\mu^2)^{-n/2}(\sigma^2)^{-A-1}e^{-\frac{\|\boldsymbol{x}\|^2}{2(\sigma^2+\sigma_\mu^2)}-\frac{B}{\sigma^2}}d\sigma^2}$$

$$p(\sigma^2|\boldsymbol{x}) = \frac{(\sigma^2+\sigma_\mu^2)^{-n/2}(\sigma^2)^{-A-1}e^{-\frac{\|\boldsymbol{x}\|^2}{2(\sigma^2+\sigma_\mu^2)}-\frac{B}{\sigma^2}}}{\int_0^\infty(\sigma^2+\sigma_\mu^2)^{-n/2}(\sigma^2)^{-A-1}e^{-\frac{\|\boldsymbol{x}\|^2}{2(\sigma^2+\sigma_\mu^2)}-\frac{B}{\sigma^2}}d\sigma^2}$$

($\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v}^T\boldsymbol{v}}$ is norm of the vector $\boldsymbol{v}$).

Even in the

univariate normal data setting

the posterior distributions involve

intractable integrals!!!

How About the Full Conditionals?

The full conditionals in the univariate normal example are:

$$p(\mu|\sigma^2,\boldsymbol{x}) \quad \text{and} \quad p(\sigma^2|\mu,\boldsymbol{x})$$

$$p(\mu|\sigma^2,\boldsymbol{x}) \sim N\left(\frac{\overline{x}}{1+\sigma^2/(n\sigma_\mu^2)}, \frac{\sigma^2}{n+\sigma^2/\sigma_\mu^2}\right)$$

$$p(\sigma^2|\mu,\boldsymbol{x}) \sim \text{Inverse-Gamma}\left(A+\tfrac{n}{2}, B+\tfrac{1}{2}\textstyle\sum_{i=1}^n(x_i-\mu)2\right)$$

Central MCMC Theoretical Result

Theory says that successive draws from

$$p(\mu|\sigma^2,\boldsymbol{x}) \quad \text{and} \quad p(\sigma^2|\mu,\boldsymbol{x})$$

eventually leads to samples from

$$p(\mu,\sigma^2|\boldsymbol{x})!$$

This is known as Gibbs sampling

and is a special case of

Markov Chain Monte Carlo (MCMC)

Live R Demonstration

We will now do a live

MCMC

demonstration using

R

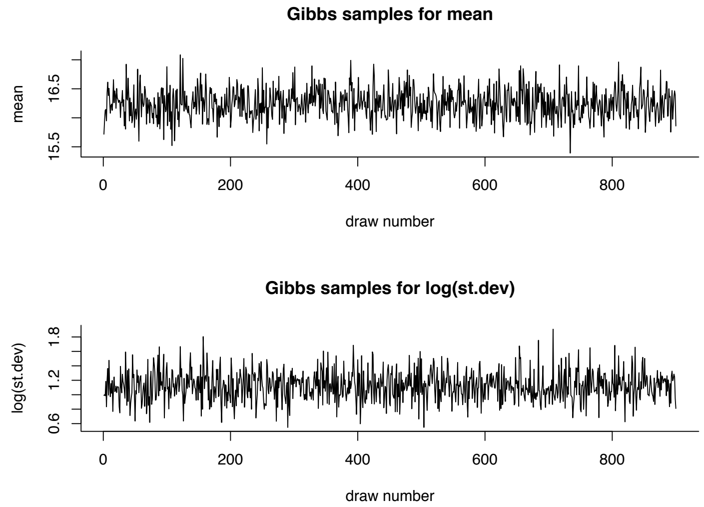
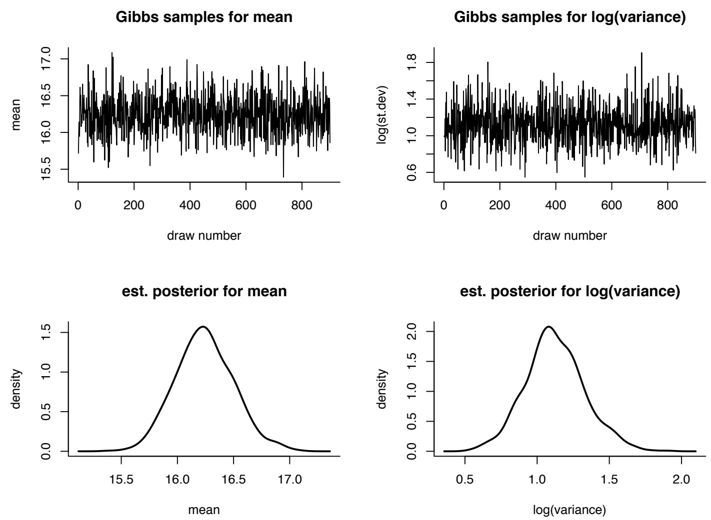


Illustration of MCMC

The following slides shows how MCMC works for the univariate normal example.

The first plots are successive draws from the full conditionals.



Practical MCMC

MCMC is a young (1990+) branch of Statistics and has several practical issues; e.g.

- partitioning of parameters,
- starting values,
- correlation between successive draws,
- convergence to required posteriors $\Longleftrightarrow$ length of burn-in.
- number of draws.

MCMC Software

Until late 2012, the most sophisticated MCMC software was from

The BUGS Project (www.mrc-bsu.cam.ac.uk/bugs)

BUGS is an acronym for

Bayesian inference Using Gibbs Sampling

New Kid on the Block: Stan

2013 has seen the emergence of a new MCMC software product named

Stan

Starting in Laboratory 2 next week we will starting using and learning Stan via its R interface:

the rstan package

Please see the new link on the subject web-site under the heading

Advice about Laptop Preparation for Computer Laboratories
THIS IS NEEDED FOR NEXT WEDNESDAY'S LABORATORY 2

Look at both BRugs and Stan script for mitsub example.

Week 4 evaluations debriefing.

Some (Allowable) Screen Time!

SCREEN-FREE ZONE

REPRIEVE COMING UP

(you may start your screens!)

Task for You During the Break

Go to the web-site

http://distancecalculator.globefeed.com/australia_distance_calculator.asp

and determine

shortest road distance in kilometres from

your Sydney residence to UTS (to nearest 0.1km)

A good UTS address to use is:
15 Broadway, Ultimo, New South Wales (Tower street entrance)