Introducing Bayesian Data Analysis Using R

Mark Andrews
Psychology Department, Nottingham Trent University

mark.andrews@ntu.ac.uk
@xmjandrews
https://github.com/lawsofthought/notts_rmeetup-jul2018

July 25, 2018

Bayesian data analysis

- Bayesian statistics is an alternative school of statistics to the *classical* or *frequentist* or school.
- Classical statistics is defined by concepts like *sampling distributions*, *(null) hypothesis significance tests*, *confidence intervals*, etc.
- These concepts per se do not exist in Bayesian approaches to statistics (though there are analogous concepts).
- Bayesian statistics arguably originated with a single essay by Reverend Thomas Bayes in 1763, though Bayes's main idea was independently discovered and developed much further by Pierre-Simon Laplace at the end of the 18th century.
- With the origin of frequentist approaches to statistics in the early 20th century, Bayesian method were sidelined.
- Beginning in the late 20th century, Bayesian methods have regained popularity.

Bayesian methods: A definition

- Classical and Bayesian approaches to data analysis both begin by positing a probabilistic *generative model* of the data being analysed.
- For example, if our data was $x_1, x_2 \ldots x_n$, we might assume or propose that this data arose from the following process:

  $$x_i \sim N(\mu, \sigma^2), \quad \text{for } i \in 1 \ldots n,$$

  where μ and σ^2 are the fixed but unknown mean and variance of a Normal distribution.
- The fundamental difference between classical and Bayesian methods are that Bayesian methods describes our state of uncertainty about μ and σ^2 by a probability distribution (known as *priors*).
- In general, all unknown variables in the probabilistic are described by priors.

Posterior distributions

- In general, given observed data D and a model Ω, the posterior distribution over the parameters θ of the model is

$$P(\theta|D,\Omega) = \frac{\overbrace{P(D|\theta)}^{\text{Likelihood}}\ \overbrace{P(\theta|\Omega)}^{\text{Prior}}}{\underbrace{\int P(D|\theta)P(\theta|\Omega)\ d\theta}_{\text{Marginal likelihood}}}.$$

  where the *marginal likelihood* gives the likelihood of the model given the observed data.
- Given the posterior distribution $P(\theta|D,\Omega)$, our aim is often to characterise this distribution in terms of e.g. its mean, variance, etc.
- Likewise, we may aim to calculate *posterior predictive* distributions such as

$$P(x_{\text{new}}|D,\Omega) = \int P(x_{\text{new}}|\theta,\Omega)P(\theta|D,\Omega)\ d\theta.$$

Sampling from posterior distributions

- In only rare situations can we determine the characteristics of the posterior distribution, or calculate posterior predictive distributions, in closed form.
- However, in general, if we can draw samples from $P(\theta|D, \Omega)$ then we can approximate, e.g., the mean of the distribution by

$$\langle\theta\rangle = \int \theta P(\theta|D, \Omega) \approx \frac{1}{N}\sum_{i=1}^{N} \tilde{\theta}_i,$$

or the posterior predictive distribution by

$$P(x_{\text{new}}|D, \Omega) = \int P(x_{\text{new}}|\theta, \Omega)P(\theta|D, \Omega)\, d\theta \approx \frac{1}{N}\sum_{i=1}^{N} P(x_{\text{new}}|\tilde{\theta}_i, \Omega),$$

where

$$\tilde{\theta}_1, \tilde{\theta}_2 \ldots \tilde{\theta}_N$$

are samples from $P(\theta|D, \Omega)$.

Markov Chain Monte Carlo (MCMC)

- Markov Chain Monte Carlo (MCMC) methods provide with general purpose methods for drawing samples from the posterior distributions of Bayesian models.
- Until recently the two most widely used MCMC methods were
  - Metropolis Hastings
  - Gibbs sampling
- More recently Hamiltonian Monte Carlo (HMC) has become very popular.

Probabilistic programming languages

- The most general purpose approach to Bayesian data analysis is to use a *probabilistic programming language* like
  - Bugs
  - Jags
  - PyMC
  - Stan
- In all these case, your work entails defining your probabilistic model, including priors, and then a MCMC sampling algorithm is compiled.

Using Stan

- We define a probabilistic model in Stan as follows:

```
...
transformed parameters {
  vector[N] mu;
  mu = a + b * x;
}

model {
  a ~ normal(0, 10);
  b ~ normal(0, 10);
  sigma ~ cauchy(0, 5);
  y ~ normal(mu, sigma);
}
```

Call Stan from R

```
M_stan <- stan(file="regression.stan",
               data = model_data,
               pars = c('a', 'b', 'sigma'),
               warmup = 5000,
               iter=10000)
```

- Here, `model_data` is artificial data with $N = 50$ data points and where the true values of the parameters were: $a = 0.5$, $b = 2.25$, $\sigma = 1.75$.

Summarize

```
S <- rstan::summary(M_stan)
pander::pander(S$summary)
```

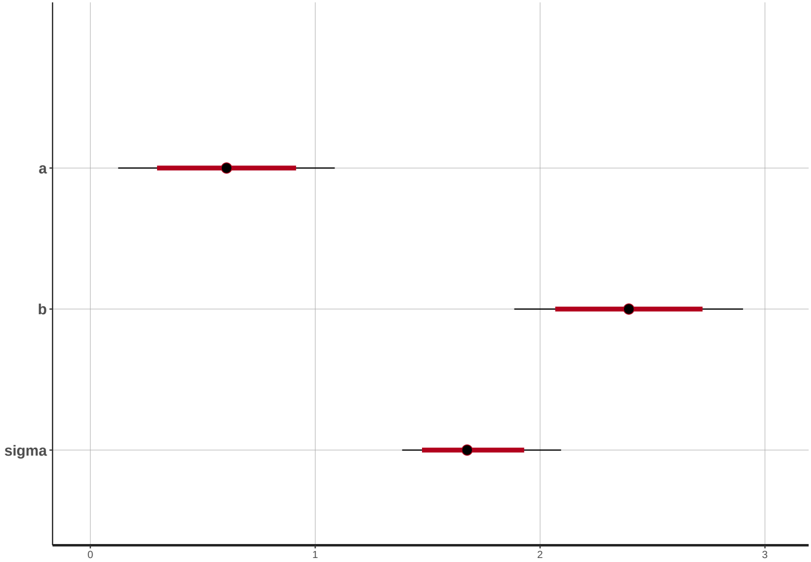
Table 1: Table continues below

	mean	se_mean	sd	2.5%	25%	50%	75%
a	0.6058	0.001801	0.2436	0.1238	0.445	0.6057	0.7656
b	2.395	0.001959	0.259	1.885	2.223	2.395	2.569
sigma	1.692	0.001362	0.1808	1.387	1.563	1.675	1.803
lp__	-50.26	0.01322	1.271	-53.55	-50.83	-49.93	-49.33

	97.5%	n_eff	Rhat
a	1.087	18282	0.9999
b	2.903	17481	1
sigma	2.093	17621	0.9999
lp__	-48.83	9249	1

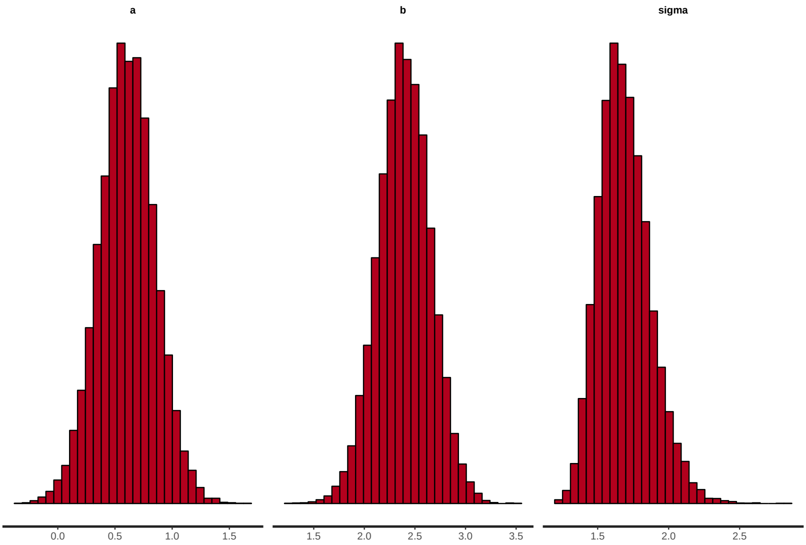
Visualize

```
rstan::plot(M_stan)
```



Visualize

```r
rstan::stan_hist(M_stan)
```



Conclusions

- Bayesian methods are vital where you need to build complex probabilistic models.
- In general, inference is based on MCMC sampling.
- Probabilistic modelling language like Stan make this easy and efficient.