An introduction to Bayesian computation & evidence synthesis using Stan

mc-stan.org
About the speaker

Robert Grant is senior lecturer in health & social care statistics at Kingston University & St George’s, University of London, UK

Wrote the StataStan interface

Interested in Bayesian latent variable models

robertgrantstats.co.uk
Bayesian computation

Computer-intensive methods
Simulation

Metropolis algorithm (40s)
Metropolis-Hastings algorithm (70s)
Gibbs sampler (80s)
Hamiltonian Monte Carlo (80s)
Bayesian software

M-H / Gibbs: BUGS, JAGS, JASP, SAS (proc mcmc), Stata (bayesmh)

Hamiltonian MC: Stan
Hamiltonian Monte Carlo

- Speed (rotation-invariance + convergence + mixing)
- Flexibility of priors
- Stability to initial values

See Radford Neal’s chapter in the "Handbook of MCMC"
Hamiltonian Monte Carlo

Tuning is tricky

One solution is the No U-Turn Sampler (NUTS)

Stan is a C++ library for NUTS (and variational inference, and L-BFGS)
Figure 7: Samples generated by random-walk Metropolis, Gibbs sampling, and NUTS. The plots compare 1,000 independent draws from a highly correlated 250-dimensional distribution (right) with 1,000,000 samples (thinned to 1,000 samples for display) generated by random-walk Metropolis (left), 1,000,000 samples (thinned to 1,000 samples for display) generated by Gibbs sampling (second from left), and 1,000 samples generated by NUTS (second from right). Only the first two dimensions are shown here.
Some Stan model code

data {
  int N;
  real y[N];
  real x[N];
}

parameters {
  real beta[2];
  real<lower=0> sigma;
}

model {
  real mu[N];
  beta ~ normal(0,50);
  sigma ~ normal(0,20);
  for(i in 1:N) {
    mu[i] <- beta[1] + beta[2]*x[i];
  }
  y ~ normal(mu,sigma);
}
```r
stan(file = 'model.stan',
     data = list.of.data,
     chains = 4,
     iter = 10000,
     warmup = 2000,
     init = list.of.initial.values,
     seed = 1234)
```
CmdStan

make "C:\model.exe"

model.exe sample data file="mydata.R"

stansummary.exe output.csv
global cmdstandir "C:/cmdstan-2.9.0"

quietly count
global N=r(N)

stan y x1 x2 x3, modelfile("model.stan") ///
cmd("$cmdstandir") globals("N")
Some simulations

Collaboration with Furr, Carpenter, Rabe-Hesketh, Gelman

arxiv.org/pdf/1601.03443v1.pdf
rstan v Stata Stan v JAGS v Stata

More recently: rstan v rjags
robertgrantstats.co.uk/rstan_v_jags.R
Rasch model (item-response)

\[ \Pr(y_{ip} = 1|\theta_p, \delta_i) = \logit^{-1}(\theta_p + \delta_i) \]

\[ \theta_p \sim N(0, \sigma^2) \]

Hierarchical Rasch model (includes hyperpriors)

\[ \Pr(y_{ip} = 1|\mu, \theta_p, \delta_i) = \logit^{-1}(\mu + \theta_p + \delta_i) \]

\[ \theta_p \sim N(0, \sigma^2) \]

\[ \delta_i \sim N(0, \tau^2) \]
StataStan vs Stata vs rjags
# rstan vs rjags

## Seconds:

<table>
<thead>
<tr>
<th></th>
<th>Rasch</th>
<th>H-Rasch</th>
</tr>
</thead>
<tbody>
<tr>
<td>rstan</td>
<td>180</td>
<td>210</td>
</tr>
<tr>
<td>rjags</td>
<td>558</td>
<td>1270</td>
</tr>
</tbody>
</table>

## ESS (sigma):

<table>
<thead>
<tr>
<th></th>
<th>Rasch</th>
<th>H-Rasch</th>
</tr>
</thead>
<tbody>
<tr>
<td>rstan</td>
<td>22965</td>
<td>21572</td>
</tr>
<tr>
<td>rjags</td>
<td>7835</td>
<td>8098</td>
</tr>
</tbody>
</table>

## ESS (theta1):

<table>
<thead>
<tr>
<th></th>
<th>Rasch</th>
<th>H-Rasch</th>
</tr>
</thead>
<tbody>
<tr>
<td>rstan</td>
<td>32000</td>
<td>32000</td>
</tr>
<tr>
<td>rjags</td>
<td>19119</td>
<td>19637</td>
</tr>
</tbody>
</table>
rstan vs rjags
rstan vs rjags
Evidence synthesis

Bayesian models can go beyond crude approximations

- Different statistics
- Different metrics
- Different scales
- Other uncertainty & bias
Coarsened data

See Heitjan & Rubin 1990

Given proportion achieving a threshold at endpoint, and baseline statistics, we can work out a posterior conditional distribution for the endpoint means.

We may have to assume, model or simulate SDs, correlation...
Test case

Cochrane review of tricyclic antidepressants in children (latest update: Mizraei et al 2013)

13 trials, sample size between 6 and 173

8 trials: mean differences & responders, one responder only

Mostly relative-ratio, but some ambiguity
Simulation study

Relative-ratio, 1000 simulations

Bias

Studies with dichotomised data
Simulation study
Simulation study

Relative-ratio, 1000 simulations

![Graph showing the relative ratio with 95% CI width against studies with dichotomised data. The graph suggests a mild increase in the 95% CI width with the number of studies.](image)
Cochrane review results

From Mizraei et al:
mean reduction (SMD) of 0.32
CI 0.04 to 0.59
risk ratio for responding: 1.07
CI 0.91 to 1.26

From the Bayesian model:
mean reduction (on CDRS scale) of 3.8 points
CI 2.4 to 5.4
A more complex setting

Review of psycho-social benefits of exercise in osteoarthritis

Lots of differences among studies
Change from baseline vs endpoint
Duration of intervention
A structural equation model
Getting started

mc-stan.org

stan-users Google Group