Stan for the people

Two day introductory workshop
on Bayesian modeling

McGill University
January 26th 2019

mc-stan.org

VI

Model parametrization

- We will do parts of Bob Carpenter's case study on *Pooling with Hierarchical Models for Repeated Binary Trials*.
- Link: `https://mc-stan.org/users/documentation/case-studies/pool-binary-trials.html`
- This case study is in part based on work by [Efron and Morris, 1975] and [Gelman et al., 2013].

We want to model player's hitting successes at Baseball.

	FirstName	LastName	Hits	At.Bats
1	Roberto	Clemente	18	45
2	Frank	Robinson	17	45
3	Frank	Howard	16	45
4	Jay	Johnstone	15	45
5	Ken	Berry	14	45
6	Jim	Spencer	14	45

- The data is `data/efron-moris-75-data.tsv`.

Complete pooling model

We want to estimate the chance of of a success,

$$\phi \in (0, 1)$$

An appropriate likelihood would be:

$$y \sim \text{Binomial}(N, \phi)$$

Again, three options:

- complete pooling
- no pooling
- partial pooling

Complete pooling

Result:

- ϕ has a mean of 0.27.
- A 95% confidence interval is [0.24, 0.30]

No pooling model

This time, we estimate $\phi_i \in (0, 1)$, the individual chance of success:

$$y_i \sim \mathrm{Binomial}(K_i, \phi_i)$$

No pooling model

Results:

- This time, we get individual estimates for each player.
- These estimates range from $\phi_{18} = 0.17$ to $\phi_1 = 0.40$.
- The 95% confidence intervals are much wider.
  For example $\phi_1 \in [0.28, 0.54]$ and $\phi_{18} \in [0.08, 0.29]$.

Using expert knowledge

- Ted Williams, 30 years prior to this data, is the only player with a recorded hitting success of 40%.
- On the other hand, a hitting success below 20% is very unusual.

What could be some issues with our data and our model?

Partial pooling model

Proposition: construct a hierarchical model, using the log odds.

Let θ_i be the odds of success for the i^{th} player.

Then let

$$\alpha_i = \mathrm{logit}(\theta_i) = \log \frac{\theta_i}{1 - \theta_i}$$

One can then show that

$$\theta_i = \mathrm{logit}^{-1}(\alpha_i) = \frac{1}{1 + e^{-\alpha_i}}$$

Remark:

$$\theta_i \in (0, 1)$$
$$\alpha_i \in \mathbb{R}$$

Our data generating process is:

$$\alpha_i \sim \mathrm{Normal}(\mu, \sigma)$$
$$y_i \sim \mathrm{Binomial}(K_i, \mathrm{logit}^{-1}(\alpha_i))$$

Priors:

$$\mu \sim \mathrm{Normal}(-1, 1)$$

This prior places a 95% probability mass on (-3, 1).

This interval corresponds on the odds scale to (0.05, 0.73).

For σ, use

$$\sigma \sim \mathrm{halfNormal}(0, 1)$$

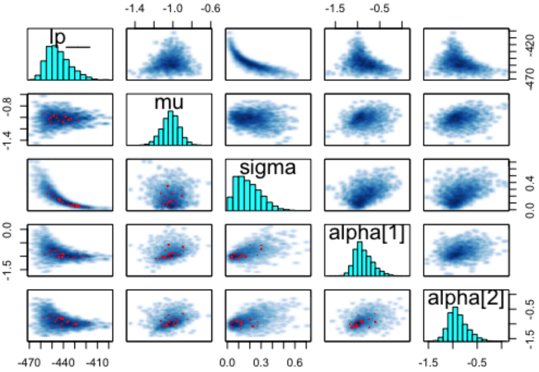
Exercise: write and fit the partial pooling model.

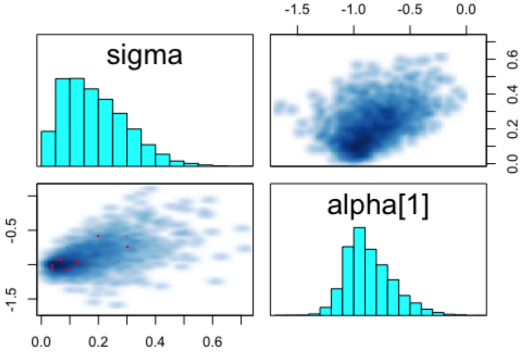
$$\begin{aligned}
\mu &\sim \mathrm{Normal}(-1, 1) \\
\sigma &\sim \mathrm{Normal}(0, 1) \\
\alpha_i &\sim \mathrm{Normal}(\mu, \sigma) \\
y_i &\sim \mathrm{Binomial}(K_i, \mathrm{logit}^{-1}(\alpha_i))
\end{aligned}$$

- Use `baseball_fit.r`
- Compute each chain in parallel.
- Report any warning messages.

“There were 29 divergent transitions after warmup.”

- A divergent transition occurs when we fail to accurately compute a Hamiltonian trajectory.
- This is because we *approximate* trajectories.
- Our sampler may not be refined enough to explore the entire typical set.


lp__
mu
sigma
alpha[1]
alpha[2]
-1.4 -1.0 -0.6
-1.5 -0.5
-420 -470
-0.8 -1.4
0.0 0.4
0.0 -1.5
-0.5 -1.5
-470 -440 -410
0.0 0.3 0.6
-1.5 -0.5


sigma
alpha[1]
-1.5
-1.0
-0.5
0.0
0.6
0.4
0.2
0.0
-0.5
-1.5
0.0
0.2
0.4
0.6

- This geometric shape is known as Neil's funnel [Neil, 2003].
- Its interaction with HMC is described in [Betancourt and Girolmi, 2015].
- It occurs in hierarchical models when we have sparse data and a centered prior.

Proposition: reparametrize the model to avoid the funnel shape. We will do so by standardizing α.

$$\alpha_{\mathrm{std},i} = \frac{\alpha_i - \mu}{\sigma}$$

Then

$$\alpha_{\mathrm{std}} \sim \mathrm{Normal}(0, 1)$$

Then

$$\alpha_i = \mu + \sigma\alpha_{\mathrm{std},i}$$

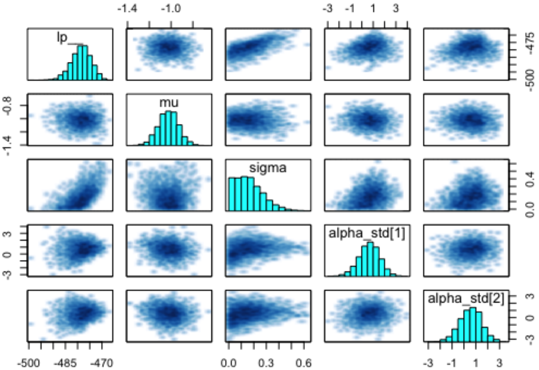
Hence

$$y_i \sim \mathrm{Binomial}(K_i, \mathrm{logit}^{-1}(\mu + \sigma\alpha_{\mathrm{std},i}))$$

Exercise: write and fit the reparametrized partial pooling model.

$$
\begin{aligned}
\mu &\sim \mathrm{Normal}(-1, 1) \\
\sigma &\sim \mathrm{Normal}(0, 1) \\
\alpha_{\mathrm{std},i} &\sim \mathrm{Normal}(0, 1) \\
y_i &\sim \mathrm{Binomial}(K_i, \mathrm{logit}^{-1}(\mu + \sigma \alpha_{\mathrm{std},i}))
\end{aligned}
$$

- Report any warning messages.
- Plot a pairs plot and examine the geometry of the posterior.
- Look at estimates: how do the *odds* compare to our results with the no pooling model?



lp__
mu
sigma
alpha_std[1]
alpha_std[2]
-1.4
-1.0
-3
-1
1
3
-475
-500
-0.8
-1.4
0.4
0.0
3
0
-3
-500
-485
-470
0.0
0.3
0.6

There is more to that example:

- An alternative hierarchical model
- More discussion on priors.
- A extensive treatment of posterior predictive checks.

References I

[Betancourt and Girolmi, 2015] Betancourt, M. and Girolmi, M. (2015).
Hamiltonian monte carlo for hierarchical models.
Current trends in Bayesian methodology with applications, 79.

[Efron and Morris, 1975] Efron, B. and Morris, C. (1975).
Data analysis using stein?s estimator and its generalizations.
Journal of the American Statistical Association, 70.

[Gelman et al., 2013] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013).
Bayesian Data Analysis.
Chapman & Hall.

[Neil, 2003] Neil, R. M. (2003).
Slice sampling.
Annals of Statistics, 31.