Stan:
Probabilistic Modeling & Bayesian Inference

Development Team

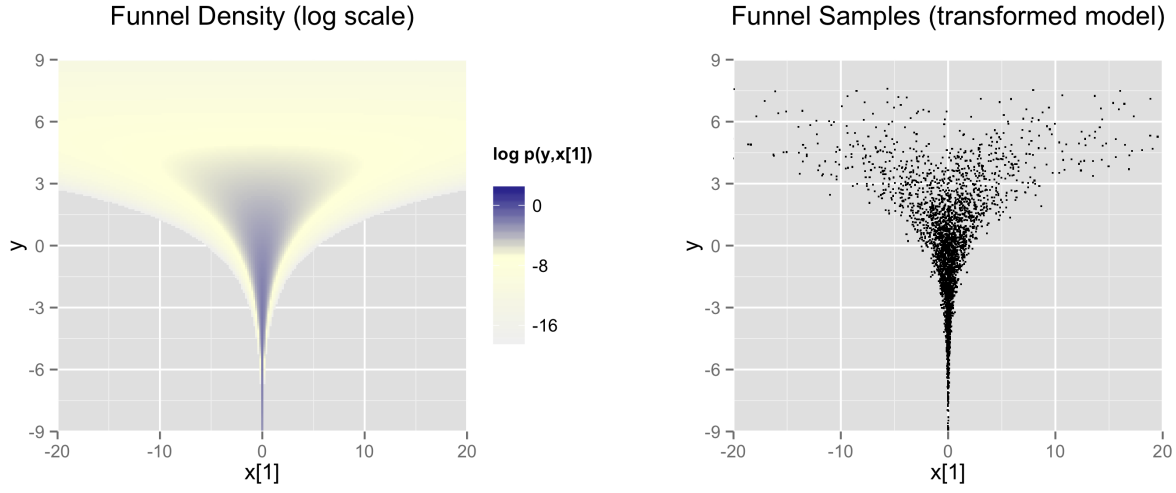
Andrew Gelman, **Bob Carpenter**, Daniel Lee, Ben Goodrich,
Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Allen Riddell,
Marco Inacio, Jeffrey Arnold, **Mitzi Morris**, Rob Trangucci,
Rob Goedman, Brian Lau, Jonah Sol Gabry, Robert L. Grant,
Krzysztof Sakrejda, Aki Vehtari, Rayleigh Lei, Sebastian Weber,
Charles Margossian, Vincent Picaud, Imad Ali, **Sean Talts**,
Ben Bales, Ari Hartikainen, Matthijs Vàkàr, Andrew Johnson,
Dan Simpson

Stan 2.17 (November 2017) http://mc-stan.org

The Funnel

Funnel-Shaped Posteriors (1/2)

- Arise in hierarchical model with no data (Neal 2003)
  - $\log \sigma \sim \text{Normal}(0, 1.5)$ [hierarchical scale]
  - $\beta_n \sim \text{Normal}(0, \sigma)$ for $n \in 1:9$ [low-level coefficients]



β_1 coefficient (x-axis) vs. $\log \sigma$ (y-axis); *left*) density plot (log scale); *right*) 4000 independent draws

Neal, R. M. 2003. Slice sampling. *Annals of Statistics*, **31**(3):705–767

Funnel-Shaped Posteriors (2/2)

- Very **challenging for sampling**
- Need large step size to explore mouth of funnel
- Need small step size to explore neck of funnel
- Even small step sizes lead to divergences
  - numerical failure of Hamiltonian dynamics simulation to conserve the Hamiltonian
- Betancourt and Girolami (2015) analyzed for Hamiltonian Monte Carlo

Betancourt and Girolami. 2015. Hamiltonian Monte Carlo for hierarchical models.

In *Current Trends in Bayesian Methodology with Applications*. CRC

Non-Centered Parameterization

- The non-centered parameterization of the funnel is

$$\begin{aligned} \log \sigma &\sim \text{Normal}(0, 1.5) \\ \beta_n^{\text{std}} &\sim \text{Normal}(0, 1) \\ \beta_n &= \sigma \times \beta_n^{\text{std}} \end{aligned}$$

- Removes dependency of β on σ in prior
- Called it “Matt trick” (after Matt Hoffman) before realizing it was well-known

Adding Data

- Use the β in an intercept-only binomial logistic regression,

$$y_j \sim \text{Binomial}(N, \text{logit}^{-1}(\beta_j))$$

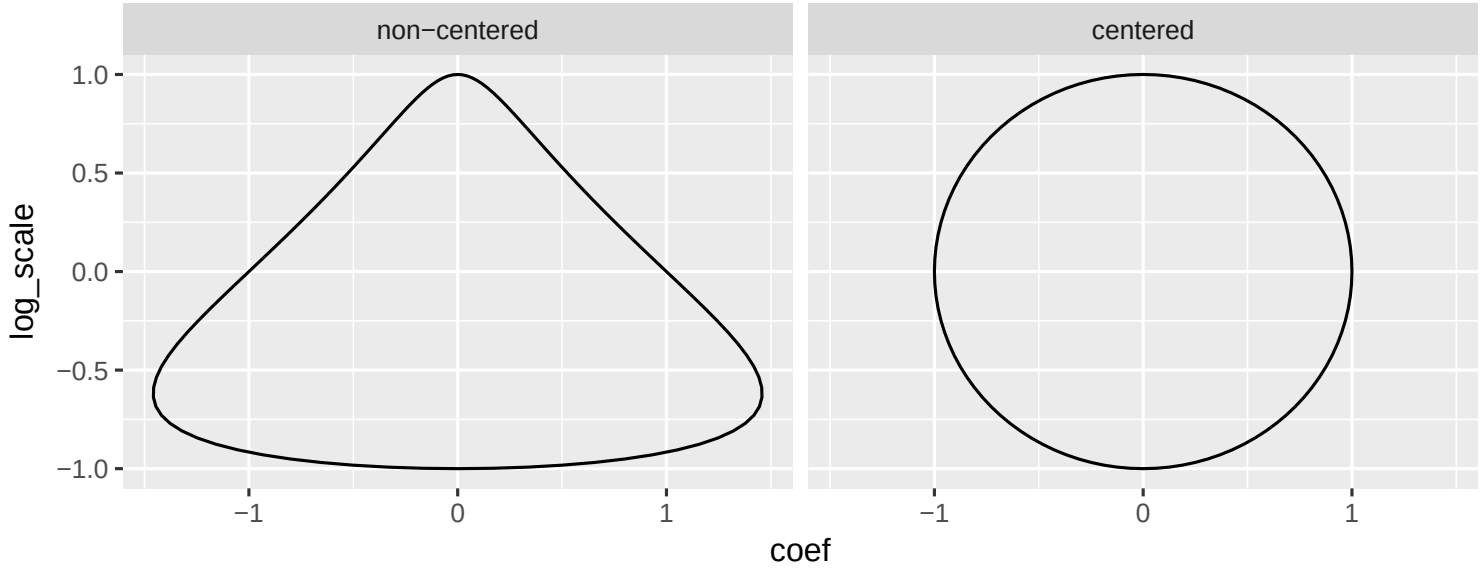
  - i.e., $y_j \in 0:N$ is number of successes in K trials
  - β_n is log odds of success for group j

- More data lessens dependency between β and σ

- With informative enough data, centered parameterization is better
  - not size of data, but how much it constrains posterior

Non-Centered + Data = Funnel

- With more data, centered approaches independent normal
- Non-centering ($\beta^{\text{std}} = \beta / \sigma$) produces a funnel
- Centered parameterizations dominate with lots of data



Funnel + Data in the Wild

centered vs. non-centered parameterization

hierarchical logistic regression with 10 groups, intercept only, 20000 draws, divergences in orange