Introducing Brms: A R package for Bayesian Regression

Mark Andrews
Psychology Department, Nottingham Trent University

mark.andrews@ntu.ac.uk
@xmjandrews
https://github.com/lawsofthought/intro-to-brms

December 12, 2018

What is Brms?

- Brms is an R package doing Bayesian general and generalized linear models, and general and generalized multilevel variants.
- To understand why Brms is so valuable, we must understand the following facts:
  1. Bayes is best. No further discussion necessary.
  2. Doing Bayesian data analysis, except for when using a prohibitively small set of models, requires Markov Chain Monte Carlo (MCMC) samplers.
  3. Writing your own MCMC is either hard or very hard.
  4. Probabilistic programming languages like Stan essentially write your MCMC sampler for any model you programmatically define.
  5. Although probabilistic programming languages reduce down the time and effort to obtain your sampler by orders of magnitude, they *still* require considerable time and effort relative to writing a single R command.
- Brms allow you to write your Bayesian model (with some restrictions) using standard R regression commands. It then writes Stan code, which then writes and compiles your sampler.

Major features

- Although ultimately more flexibility will be obtained using Stan, Brms is remarkably powerful:
- It includes far more probability models for outcome variables than almost all other regression packages: gaussian, student, binomial, bernoulli, poisson, negbinomial, geometric, Gamma, skew_normal, lognormal, shifted_lognormal, exgaussian, wiener, inverse.gaussian, exponential, weibull, frechet, Beta, von_mises, asym_laplace, gen_extreme_value, categorical, cumulative, cratio, sratio, acat, hurdle_poisson, hurdle_negbinomial, hurdle_gamma, hurdle_lognormal, zero_inflated_binomial, zero_inflated_beta, zero_inflated_negbinomial, zero_inflated_poisson, and zero_one_inflated_beta.
- It also allows for censored data, missing data, measurment error, nonlinear regression, probabilistic mixture models, *distributional* models (whereby all parameters of the outcome variables have predictors), and so on.