Bayesian Cumulative Probability Models for Continuous Response Variables

Nathan T. James, ScM

Dept. of Biostatistics, Vanderbilt University

July 30, 2018

Introduction

- ▶ Ordinal Cumulative Probability Model (CPM)
  - ▶ $G[P(Y \leq y_i|X)] = \alpha_i - \beta^T X$
  - ▶ y_i - ordered, continuous outcome
 - ▶ X - matrix of covariates
  - ▶ $G(\cdot)$ - link function

Why use a Bayesian CPM with a **continuous** outcome?

- ▶ Invariant to monotonic transformation of outcome

- ▶ Directly model full conditional CDF

- ▶ Handles any ordered outcome including mixed discrete/continuous distributions (e.g., continuous outcome with lower limit of detection)

- ▶ Inference using posterior probabilities

- α_i estimate posterior CDF for $X = 0$
- β measure association between X and distribution of Y; interpretation depends on link function
- Mean and quantiles calculated from posterior distribution of full conditional CDF using single model
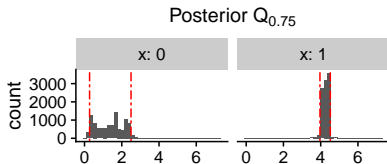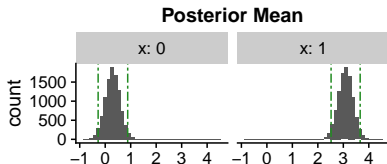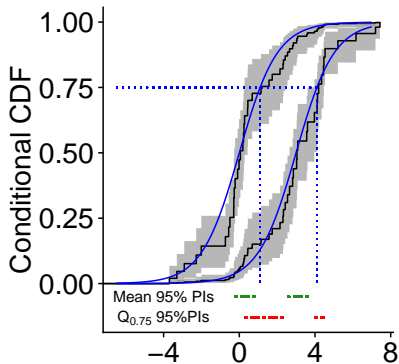
- ▶ Implemented using `brms` and `rstanarm`; both call `Rstan`
- ▶ Different parameterizations; using default priors `rstanarm` more accurate in simulations
- ▶ Model convergence depends on package and link function
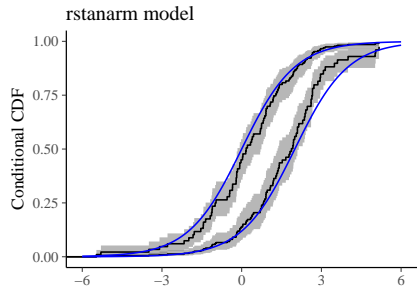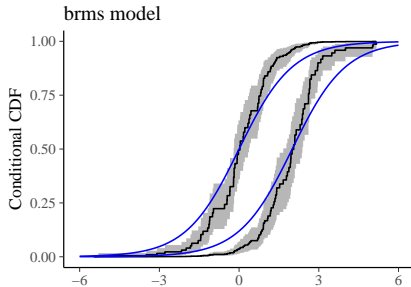
- ▶ `brms` needs to compile C++ code, `rstanarm` pre-compiled
- ▶ Major differences in computation time based on link function
- ▶ For datasets up to ~ 1000 distinct y values computation time is approximately linear for both packages; for larger datasets compute time increases at a faster rate for `brms`
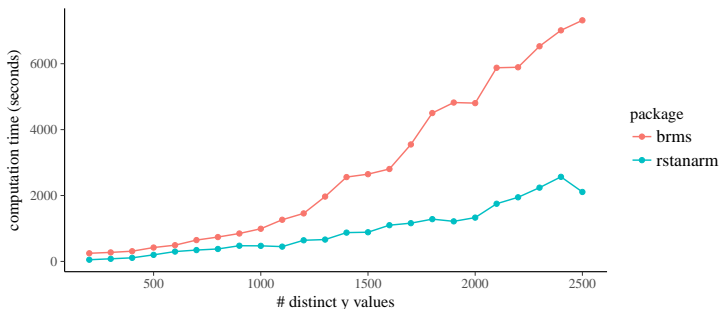
- With moderate sample size, reasonably robust to misspecification of link function
- Uncertainty in link function can be accounted for using a mixture of links
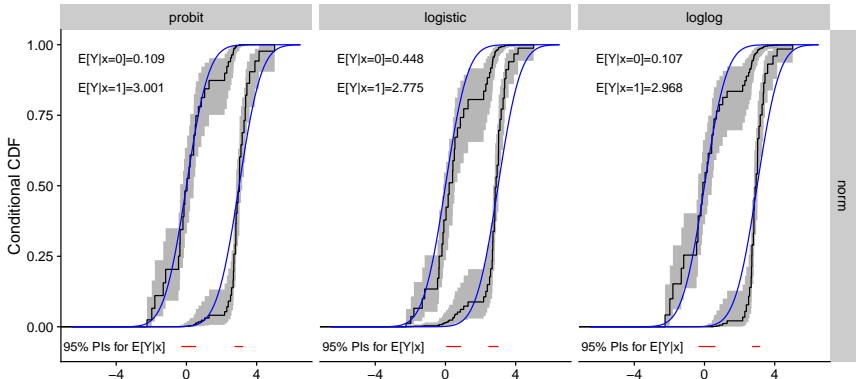
Contact

email: nathan.t.james AT vanderbilt.edu

web: ntjames.com

twitter: @nalhsyjones