Introduction to Stan @ ISBA 2016

Hands-on Session

Sebastian Weber, Novartis Pharma

12th June 16, ISBA16

NOVARTIS

Warming Up

Hierarchical Models

Mixture Models

Complex Models

The End

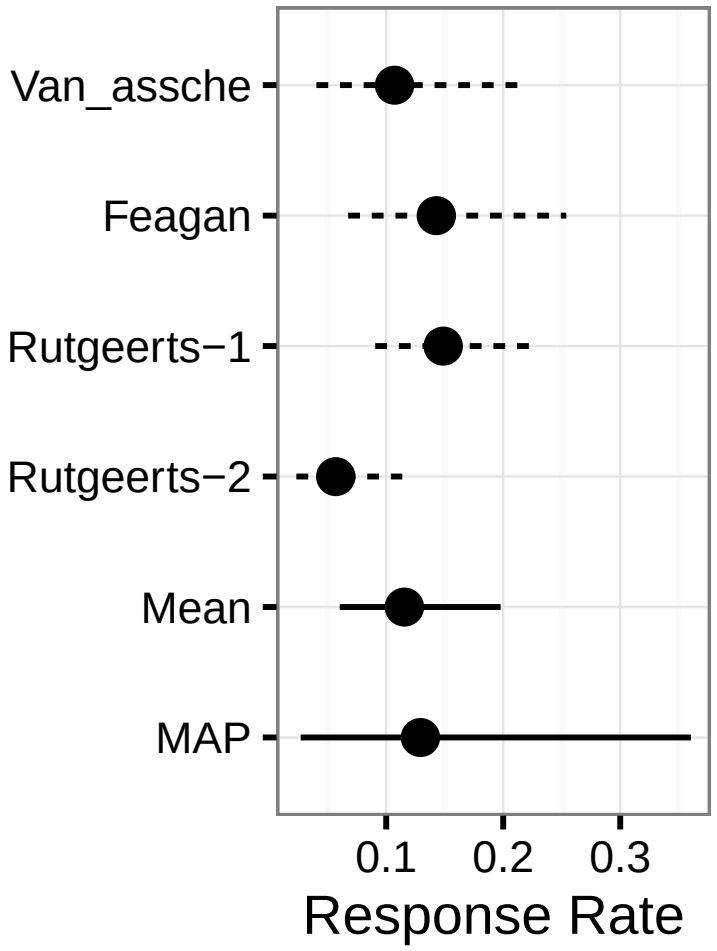
Warming Up

Using Historical Control Information in Clinical Trials

Disease: Ulcerative colitis

- Binary Endpoint:
  clinical remission at week 8
- New study:
  parallel group (Placebo vs. Test treatment)
- Placebo data:
  four external trials of similar design

Goal: Reduce placebo patient group size in new trial by using an informative prior derived from literature data.



NOVARTIS

Ulcerative Colitis Literature Data

Literature data `colitis_data.R`

study	n	r
Van_assche	56	6
Feagan	63	9
Rutgeerts-1	121	18
Rutgeerts-2	123	7

Task 1: Complete Pooling Approach

Ignoring between-trial heterogeniety

Use Stan to do a complete pooling analysis using logistic regression

$$Y_h|\theta_h, n_h \sim \text{Binomial}(\theta, n_h)$$
$$\text{logit}(\theta)|\mu \sim \text{Normal}(0, 2^2)$$

In your stan program you should

- Use constraints on input data
- Write a variant using only a `for` loop and then **vectorize**
- Use the `binomial_logit` instead of the `binomial` statement - why?

Cross-check the 95% credible interval from Stan using a simple Beta-Binomial approach.

Even more efficient is to just use the overall summary statistics.

Hierarchical Models

The Generalized Meta-Analytic-Predictive Approach

Hierarchical model accounting for between-trial heterogeneity

Y is the (control) group summary data for H historical trials

$$Y_h|\theta_h \sim f(\theta_h) \qquad \forall\, h \in [1, H]$$
$$Y_*|\theta_* \sim f(\theta_*) \qquad \text{for new trial}$$

Exchangeability assumption:

$$g(\theta_h)|\mu, \tau \sim \text{Normal}(\mu, \tau^2) \qquad \forall\, h \in [1, H]$$
$$g(\theta_*)|\mu, \tau \sim \text{Normal}(\mu, \tau^2) \qquad \text{for new trial}$$

- f likelihood and g link function
  Binomial/logit, Normal (fixed σ)/identity or Poisson/log
- μ population mean with prior Normal(m_μ, s_μ^2)
- τ between-trial heterogeneity with prior $P(\tau)$

Task 2: Implement MAP for the Binomial Case

Accounting for between-trial heterogeneity

Statistical model

$$Y_h|\theta_h, n_h \sim \text{Binomial}(\theta_h, n_h)$$
$$\text{logit}(\theta_h)|\mu, \tau \sim \text{Normal}(\mu, \tau^2)$$

Priors: $\mu \sim \text{Normal}(0, 2^2)$ and $\tau \sim \text{HalfNormal}(0, 1^2)$

- Use constraints on parameters where applicable - why?
- Use the centered parametrization
- Use the non-centered parametrization
- Create a `generated quantities` block and simulate the MAP prior which can be used in a new trial for θ_*

Wrap-Up: Best Practices Basic Models

- Use constraints on data, parameters, transformed parameters
- **Ensure that the support of parameters matches their constraints**
- (Not Stan specific) Use priors to your best knowledge
- **Vectorize** your statements
- Use numerically stable formulations
  - `binomial_logit` instead of `binomial`
  - `poisson_log` instead of `poisson`
  - calculate in `log` space directly, i.e. use `log_sum_exp`, `log1m_exp`, ...
- **Hierarchical models** usually benefit from a **non-centered parametrization** (aka Matt trick)

Mixture Models

The Story of Discreteness & Mixtures

Stan cannot sample discrete parameters! What now? Marginalize!

Discrete parameters (`int` typed parameters) must be *marginalized* out from the log-likelihood, i.e. for a mixture model with K components we get

$$\log(f(y|\theta)) = \log\left[\sum_{k=1}^{K} \pi_k \, f(y|\theta_k)\right]$$

Note:

- Summation on the natural space
  $=>$ use `log_sum_exp` or `log_mix`
- Vectorization must be reconsidered
- Posterior mixture weights can be calculated

Detour: Posterior Mixture Weights

$$p(\theta|y) = f(y|\theta)\, p(\theta)/p(y) = \sum_{k=1}^{K} \frac{\pi_k}{p(y)} f(y|\theta)\, p_k(\theta)$$

$$= \sum_{k=1}^{K} \underbrace{\frac{\pi_k}{p(y)} \left(\int p_k(\theta) f(y|\theta)\, d\theta \right)}_{=\pi'_k} \underbrace{\frac{p_k(\theta) f(y|\theta)}{\int p_k(\theta) f(y|\theta)\, d\theta}}_{p_k(\theta|y)}$$

$$= \sum_{k=1}^{K} \pi'_k\, p_k(\theta|y)$$

Posterior mixture weights, π'_k, determined by the marginal likelihood of the data under the prior for component k

$$p(y) = \int f(y|\theta)\, p(\theta) d\theta = \sum_{k=1}^{K} \pi_k \int f(y|\theta)\, p_k(\theta) d\theta$$

Example Mixture Model: Robust MAP Analysis

Relaxing the exchangeability assumption

Allowing each trial $h \in [1, H]$ to be non-exchangable with the others with a probability $1 - \pi_h$

- $Y_h \in$ EXchangeable set with probability π_h, then

$$Y_h|\theta_h, n_h \sim \text{Binomial}(\theta_h, n_h)$$
$$\text{logit}(\theta_h)|\mu, \tau \sim \text{Normal}(\mu, \tau^2)$$

- $Y_h \notin$ EXchangeable set with probability $1 - \pi_h$, then

$$Y_h|\theta_{NEX,h}, n_h \sim \text{Binomial}(\theta_{NEX,h}, n_h)$$
$$\text{logit}(\theta_{NEX,h}) \sim \text{Normal}(0, 2^2)$$

Complex Models

Bayesian Inference for Complex Models with Stan

- Stan facilitates *complex* modeling
  - Non-linear models
  - Ordinary Differential Equations (ODE)
  - High efficiency of HMC allows short warmup and sampling
  - Programming constructs
  - Debugging facilities

  All you need to do is to code the log-likelihood!

- Example: pharmacokinetic model
  - Longitudinal model describing drug concentration in a human body over time after administration of a dose
  - Standard technique is a compartemental approach which is conveniently described by ODEs

Pharmacokinetic Model for an Oral Drug

Compartemental modeling goverend by mass action kinetics

- $A_1(t)$ is the drug amount in the dosing compartement (gut)

$$\frac{dA_1}{dt} = -k_a A_1$$

- $A_2(t)$ is the drug amount in the central compartement (blood)

$$\frac{dA_2}{dt} = k_a A_1 - k_{10} A_2$$

Mass action kinetics law describes **amounts**, but we measure **concentrations**, so:

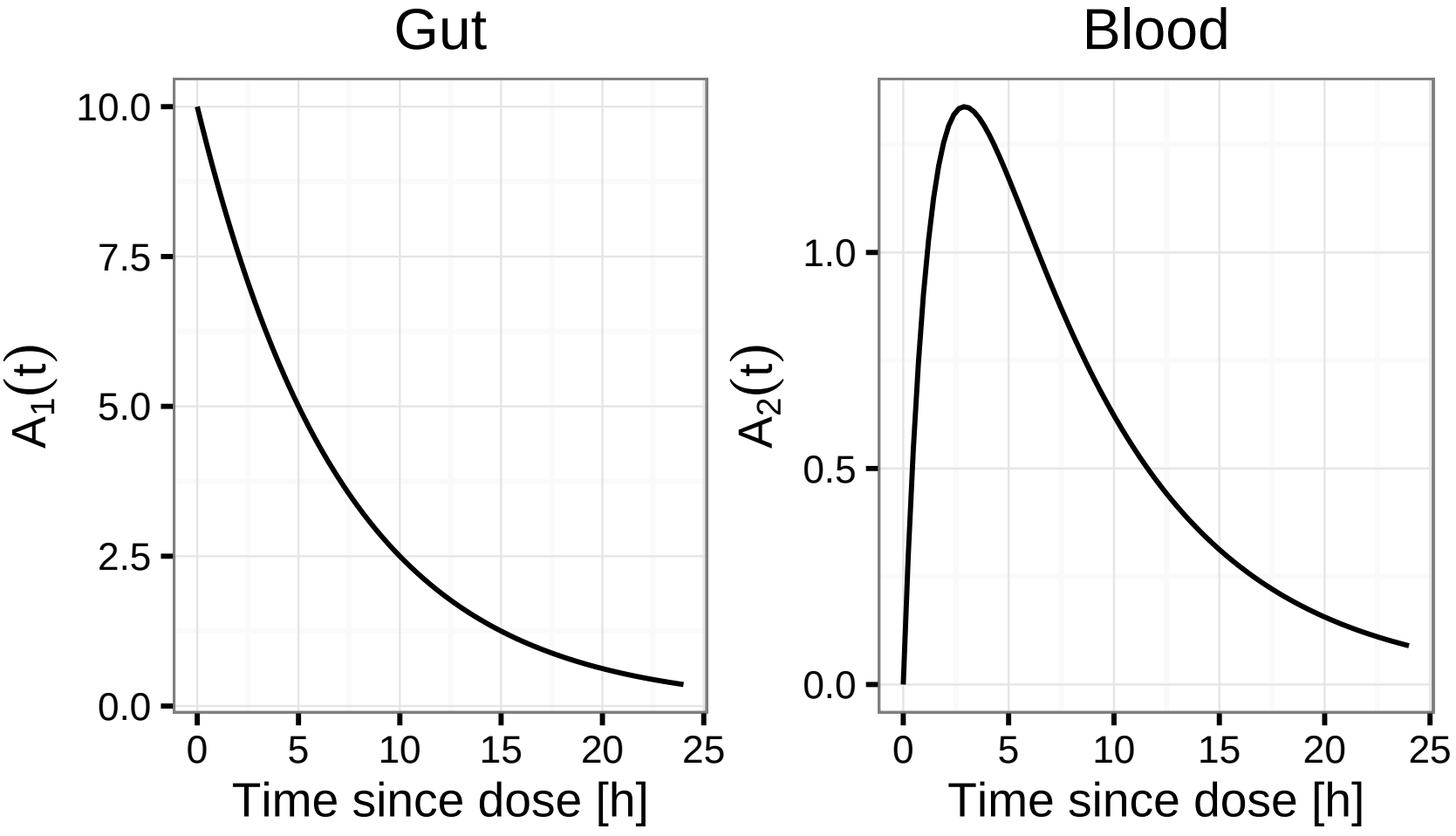
$$C(t) = \frac{A_2(t)}{V} \qquad\qquad k_{10} = \frac{CL}{V}$$

Initial condition: $A_1(t = 0) = d$ and $A_2(t = 0) = 0$

Patients differ $\Rightarrow$ *hierarchical parameter modeling*

Pharmacokinetic Model for an Oral Drug

Example for $d = 10$ and $V = 1$



Real Example Next: Warfarin Pharmacokinetics

NOVARTIS

Wrap-Up: Best Practices Complex Models

- Build complex models in **small building blocks**, i.e. use the `functions` block
- **Test functions** with `expose_stan_functions` in R
- Use `print` for debugging and `reject` to abort early
- Choose **appropiate parametrizations**
- Use unit scaled variables
- For non-regular data use `ragged arrays`
- Employ multi-variate non-centered parametrization when needed for random effects
- Consider the use of the **cholesky multi-variate normal version**, i.e. with `cholesky_factor_corr` and LKJ priors

The End

Stan for Bayesian Inference

- **Highly efficient and scalable Hamiltonian Monte Carlo**
  - Sampling as geometrical problem
  - Constrained types $\rightarrow$ efficient parametrizations
  - Often only needs a **few hundred iterations** to warmup
  - **Transparent** $\rightarrow$ sampler diagnostics
- **Rich user-friendly Stan language**
  - Linear algebra
  - Various data structures
  - User defined functions
  - Debugging facilities
- User manual, case studies, `stan-users` mailinglist, ...

Many thanks to Michael Betancourt

References

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. (2013) *Bayesian Data Analysis*, third. CRC Press.

Neuenschwander, B., Capkun-Niggli, G., Branson, M. and Spiegelhalter, D.J. (2010) Summarizing historical information on controls in. *Clin Trials*, **7**, 5–18.

O'reilly, R.A. and Aggeler, P.M. (1968) Studies on Coumarin Anticoagulant Drugs Initiation of Warfarin Therapy Without a Loading Dose. *Circulation*, **38**, 169–177.

O'Reilly, R.A., Aggeler, P.M. and Leong, L.S. (1963) Studies on the Coumarin Anticoagulant Drugs: The Pharmacodynamics of Warfarin in Man. *J Clin Invest*, **42**, 1542–1551.