Intuitive Joint Priors for Bayesian Multilevel Models

ICSDS 2022, Florence, Italy

Javier Enrique Aguilar, Paul-Christian Bürkner

16.12.2022

Cluster of Excellence SimTech , University of Stuttgart
javier.aguilar-romero@simtech.uni-stuttgart.de

Priors

Prior specification

Question: How to make it easier for the user to specify priors?

Prior specification

Question: How to make it easier for the user to specify priors?

- Build upon **intuition**.
- Propose priors on **predictive quantities** that are better understood by the user and move uncertainty.
- Propose semi automatic and parsimonious priors.

Prior specification

Question: How to make it easier for the user to specify priors?

- Build upon **intuition**.
- Propose priors on **predictive quantities** that are better understood by the user and move uncertainty.
- Propose semi automatic and parsimonious priors.

Context: Bayesian Linear Multilevel models.

$$\mu_n = \sum_{i=1}^{p} x_{ni} b_i + \sum_{i=0}^{p} x_{ni} \left(\sum_{g \in G_i} u_{ig_{j[n]}} \right)$$
$$y_n \sim N(\mu_n, \sigma)$$

- b_i Overall coefficients
- u_{igj} Varying coefficients s.t $\sim N(0, \Sigma_u)$.

Priors in Linear Regression

How to select the prior for b?

1. **Usual setting:**

$$b|\sigma \sim N(0, \sigma^2 \Sigma_b), \quad \sigma \sim p(\sigma)$$

2. **Shrinkage priors**

$$b_i|\Psi_i \sim N(0, \Psi_i), \quad \Psi_i \sim G(\cdot)$$
$$p(b_i) = \int N(b_i|0, \Psi_i) dG(\Psi_i)$$

The form of G gives rise to famous priors!

Implied prior on R^2

Consider

- b_i have **weakly informative priors**

$$b_i \sim N(0, 1), \quad i = 1, ..., p$$

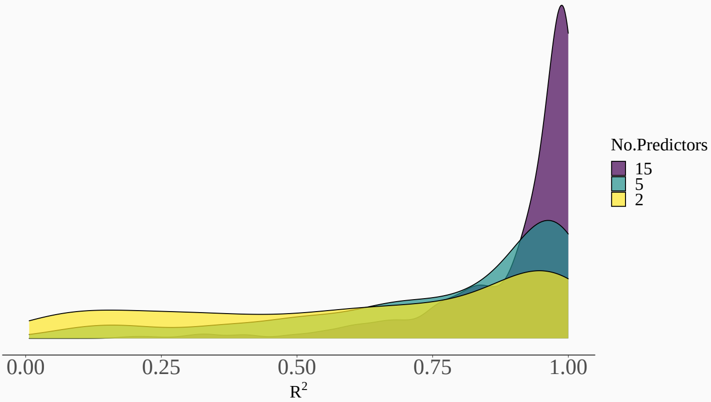
- $\sigma \sim \text{Exp}(1)$
- The **proportion of explained variance** R^2 is given by

$$R^2 = \frac{\text{var}(\mu)}{\text{var}(\mu) + \sigma^2}$$

Curiosity: What is the effect on R^2?

Implied prior on R^2

The implied prior distribution is highly informative!



Implied prior on R^2

Idea: Specify a prior on R^2 and decompose the **explained variance** via a *Dirichlet Decomposition* (R2D2) [2]

- **The good:** Important theoretical properties. (Posterior consistency and contraction, heavy tails, mass near origin)
- **Limitation:** Single level models.
- **Goal**: Generalize R2D2 prior to Multilevel Models setting.

Our prior

The R2D2M2 prior

Define a **global proportion of explained variance** R^2 measure for multilevel models as

$$R^2 := \mathrm{corr}^2(y, \mu) = \frac{\mathrm{var}(\mu)}{\mathrm{var}(\mu) + \sigma^2},$$

- **Main objective:** Jointly regularize
- Multiple measures of R^2 are defined in the literature (No consensus?)

The R2D2M2 prior: Construction

- Assumption

$$\mathbb{E}(b_i) = \mathbb{E}(u_{igj}) = 0,\ \text{var}(b_i) = \sigma^2\lambda_i^2,\quad \text{var}(u_{igj}) = \sigma^2\lambda_{ig}^2$$

The R2D2M2 prior: Construction

- Assumption

$$\mathbb{E}(b_i) = \mathbb{E}(u_{igj}) = 0, \ \text{var}(b_i) = \sigma^2 \lambda_i^2, \quad \text{var}(u_{igj}) = \sigma^2 \lambda_{ig}^2$$

- Variance decomposition of the linear predictor μ

$$\text{var}(\mu) = \underbrace{\sum_{i=1}^{p} \sigma^2 \lambda_i^2}_{\text{Overall}} + \underbrace{\sum_{i=1}^{p} \sum_{G_i} \sigma^2 \lambda_{ig}^2}_{\text{Varying}}$$

The R2D2M2 prior: Construction

- Assumption

$$\mathbb{E}(b_i) = \mathbb{E}(u_{igj}) = 0, \ \text{var}(b_i) = \sigma^2 \lambda_i^2, \ \ \text{var}(u_{igj}) = \sigma^2 \lambda_{ig}^2$$

- Variance decomposition of the linear predictor μ

$$\text{var}(\mu) = \underbrace{\sum_{i=1}^{p} \sigma^2 \lambda_i^2}_{\text{Overall}} + \underbrace{\sum_{i=1}^{p} \sum_{G_i} \sigma^2 \lambda_{ig}^2}_{\text{Varying}}$$

- Total explained variance τ^2

$$\tau^2 := \sum_{i=1}^{p} \lambda_i^2 + \sum_{i=1}^{p} \sum_{G_i} \lambda_{ig}^2$$

The R2D2M2 prior: Construction

- Assumption

$$\mathbb{E}(b_i) = \mathbb{E}(u_{igj}) = 0,\ \text{var}(b_i) = \sigma^2\lambda_i^2,\ \ \text{var}(u_{igj}) = \sigma^2\lambda_{ig}^2$$

- Variance decomposition of the linear predictor μ

$$\text{var}(\mu) = \underbrace{\sum_{i=1}^{p} \sigma^2\lambda_i^2}_{\text{Overall}} + \underbrace{\sum_{i=1}^{p}\sum_{G_i} \sigma^2\lambda_{ig}^2}_{\text{Varying}}$$

- Total explained variance τ^2

$$\tau^2 := \sum_{i=1}^{p} \lambda_i^2 + \sum_{i=1}^{p}\sum_{G_i} \lambda_{ig}^2$$

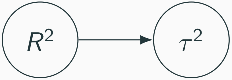
- Rewrite R^2

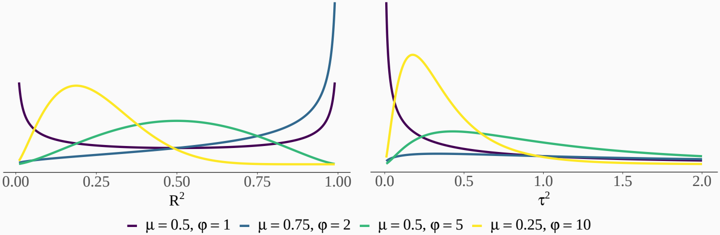
$$R^2 = \frac{\sigma^2\tau^2}{\sigma^2\tau^2 + \sigma^2} = \frac{\tau^2}{\tau^2 + 1}.$$

The R2D2M2 prior: Construction

Set

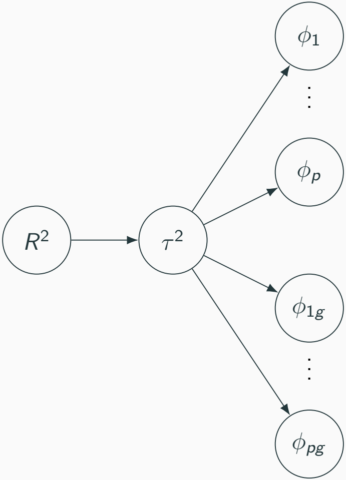
$$R^2 \sim \text{Beta}(\mu_{R^2}, \varphi_{R^2}) \iff \tau^2 \sim \text{BP}(\mu_{R^2}, \varphi_{R^2})$$





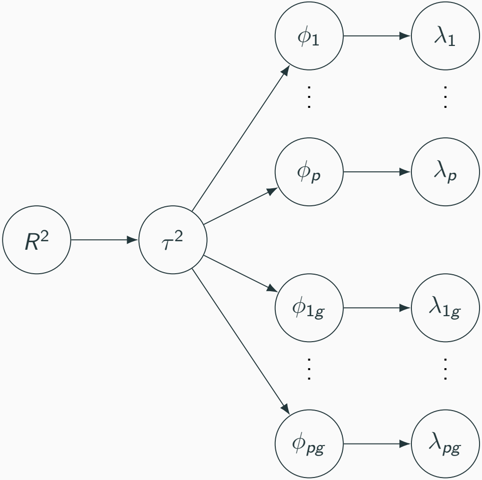
The R2D2M2 prior: Construction

- Partition the variance $\phi \sim \text{Dir}(\alpha)$ **Dirichlet Decomposition**



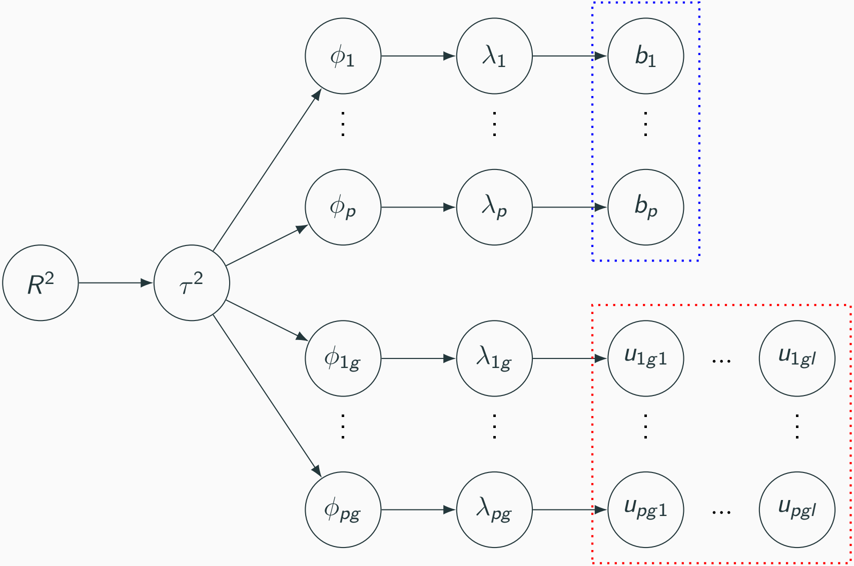
The R2D2M2 prior: Construction

- Assign variances $\lambda^2 = \phi\tau^2$



The R2D2M2 prior: Construction

Make b_i, u_{ig_j} Gaussian



The R2D2M2 prior

Our prior is

$$R^2 \sim \text{Beta}(\mu_{R^2}, \varphi_{R^2})$$
$$\phi \sim \text{Dir}(\alpha), \quad \sigma \sim p(\sigma)$$
$$b_i|\sigma,\phi,\tau \sim N(0, \sigma^2 \phi_i \tau^2), \quad u_{igj}|\sigma,\phi,\tau \sim N(0, \sigma^2 \phi_{ig} \tau^2)$$

The R2D2M2 prior

Our prior is

$$R^2 \sim \text{Beta}(\mu_{R^2}, \varphi_{R^2})$$
$$\phi \sim \text{Dir}(\alpha), \quad \sigma \sim p(\sigma)$$
$$b_i|\sigma, \phi, \tau \sim N(0, \sigma^2 \phi_i \tau^2), \quad u_{igj}|\sigma, \phi, \tau \sim N(0, \sigma^2 \phi_{ig} \tau^2)$$

- Global-local continuous shrinkage prior.
- Low number of prior hyperparameters.
- In practice $\alpha = (a_\pi, ..., a_\pi)'$ where $a_\pi > 0$.
- **Prior beliefs** about model fit are **propagated** in an **intuitive wa**y by use of the prior mean μ_{R^2} and prior precision φ_{R^2}.
- Joint regularization is taking place

Properties

1. **Spike behavior** If $a_\pi \leq 1/2$ then marginal priors are unbounded near the origin.
2. **Heavy tails** When $(1 - \mu_{R^2})\varphi_{R^2} \leq 1/2$ the marginal priors have heavier tails than the Cauchy distribution.
3. Bounded influence and regularised version.
4. Possible to treat high dimensionality in both overall coefficients p and varying terms q. (Not common)

Examples

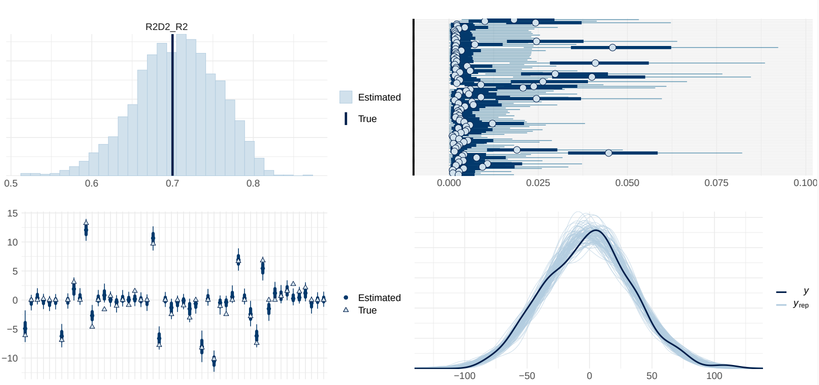
Example: Simulations

- Large scale simulations where we test the performance of the prior under different conditions and hyperparameter values.
- Maximal models, 48 datasets, $T = 4000$ posterior draws with Stan.

Description	Hyperparameter	Values
True value of R^2	R_0^2	$\{0.25, 0.75\}$
Groups	K	$\{1, 3\}$
Levels	L	20
Covariates	p	$\{10, 100, 300\}$
Prior mean of R^2	μ_{R^2}	$\{0.1, 0.5\}$
Prior precision of R^2	φ_{R^2}	$\{0.5, 1\}$
Concentration parameter	a_π	$\{0.5, 1\}$
Covariance matrix of X	Σ_x	$\{I_p, AR(\rho)\}$ where $\rho \in \{0.5\}$
Level of induced sparsity	v, z	$\{0.5, 0.95\}$

Single simulation:

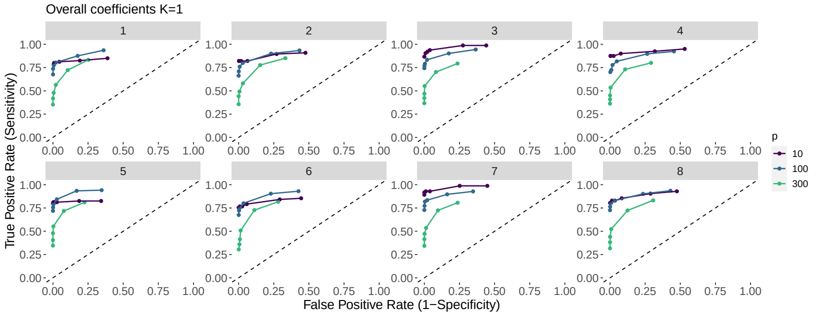
$p = 50, L = 25, n = 500, R_0^2 = 0.7, (\mu_{R^2}, \varphi_{R^2}, a_\pi)' = (0.5, 0.5, 0.5)'$



Total number of parameters: 1480

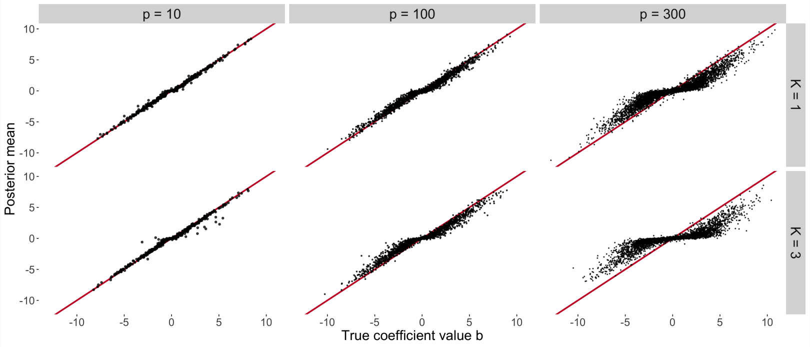
Example: Simulations

Classical hypothesis testing for overall coefficients $K = 1$



Example: Simulations

Posterior shrinkage of b , $n = 100, N_{\text{sims}} = 48$



Number of coefficients to estimate ranges from 231 to 18361.

Takeaway message

1. User friendly prior specification is hard but it can be done!
2. Joint regularization offers new opportunities in how to specify priors.
3. Intuitive priors give the ability to nonexpert users to apply Bayesian Modeling.
4. Two step procedure:
   4.1 Follow theoretical and practical guidelines to specify priors.
   4.2 Simulations allow us to evaluate priors.

References

[1] Javier Enrique Aguilar and Paul-Christian Bürkner. Intuitive joint priors for bayesian linear multilevel models: The r2d2m2 prior. *arXiv preprint arXiv:2208.07132*, 2022.

[2] Yan Dora Zhang, Brian P. Naughton, Howard D. Bondell, and Brian J. Reich. Bayesian regression using a prior on the model fit: The r2-d2 shrinkage prior. *Journal of the American Statistical Association*, 0(0):1–13, 2020. doi: 10.1080/01621459.2020.1825449. URL `https://doi.org/10.1080/01621459.2020.1825449`.

Contact

Thank you for your attention!

- ✉ javier.aguilar-romero@simtech.uni-stuttgart.de