# Specifying Priors in a Bayesian Workflow

Paul Bürkner

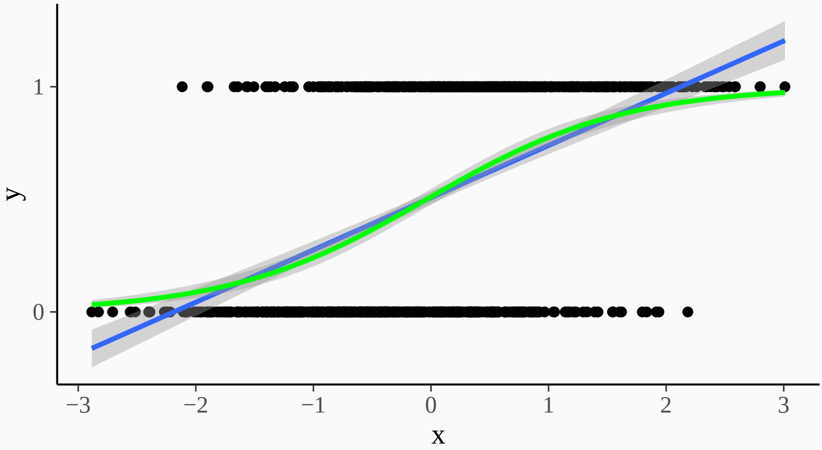
Cluster of Excellence SimTech, University of Stuttgart

I don't have many good answers (yet)

## Bayesian inference

$$p(\theta|y) \propto p(y|\theta)p(\theta) = p(y, \theta)$$

# The prior can only be understood in the context of the model




Further reading:

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*.

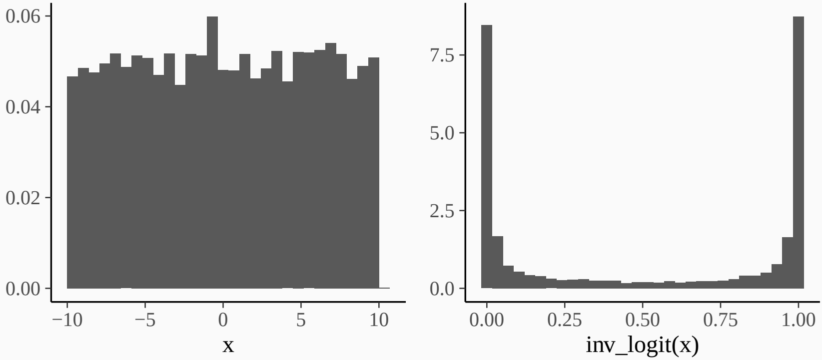
What are your reasons for using priors?

# Some reasons for using priors

- Make a-priori implausible values unlikely (weakly informative priors)
- Incorporate specific expert information into the model ("subjective" priors)
- Mimic frequentist methods (uninformative/"objective" priors)
- Represent known data structure (multilevel priors)
- Regularize the model to avoid overfitting (shrinkage/sparsifying priors)
- Enable hypothesis testing via Bayes factors
- Ensure unimodal posteriors
- Facilitate convergence
- ...

# Some observations about priors

# Uniformity is informative



# Prior tails matter

What is the posterior?

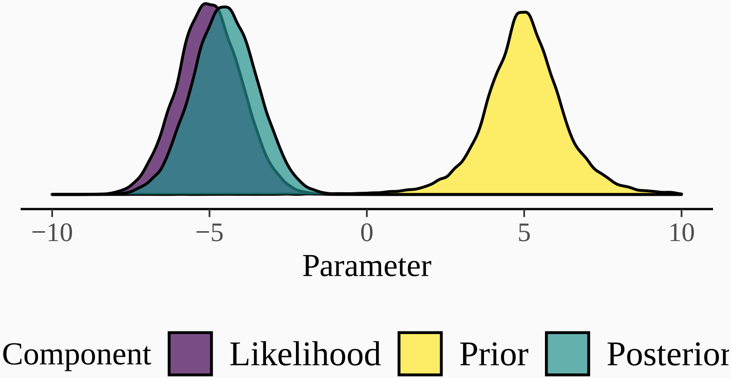
- . . . if the prior is

$$\theta \sim \text{student-t}(4, 5, 1)$$

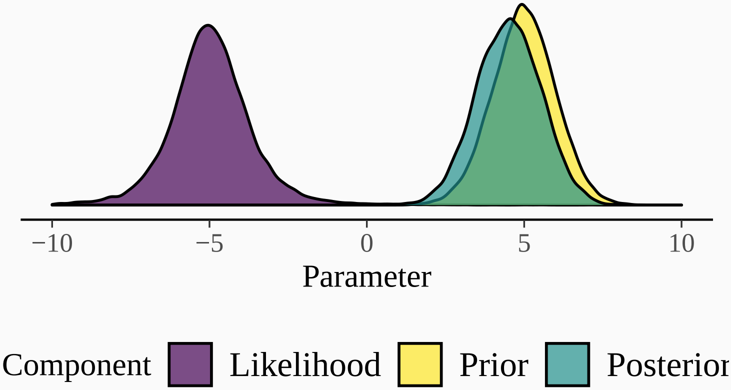
- . . . and the likelihood alone implies a posterior of

$$\theta \sim \text{normal}(-5, 1)$$

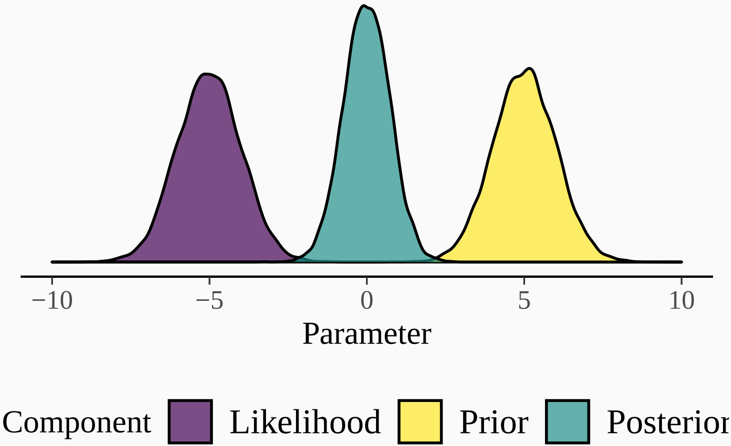
# Prior tails matter



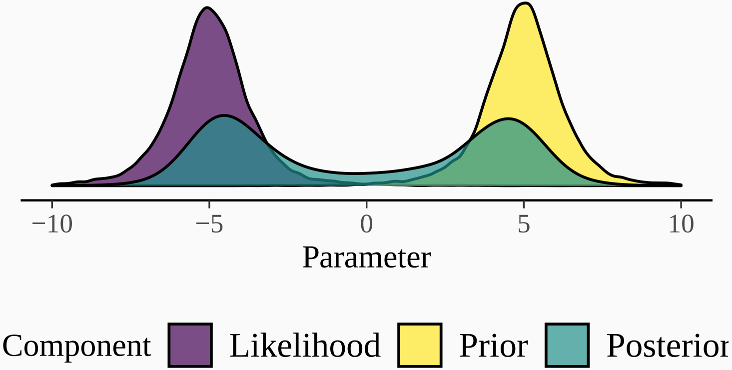
# Prior tails matter



# Prior tails matter



# Prior tails matter



# Priors on high dimensional models are weird

Suppose a linear regression model with

$$y \sim \text{normal}(\mu, \sigma)$$

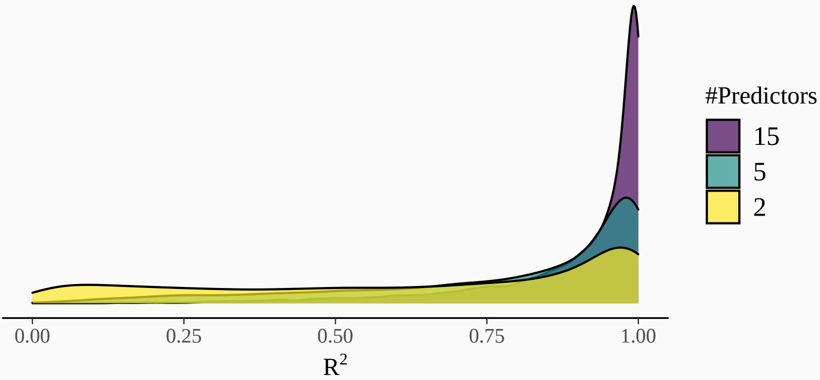
$$\mu = \sum_{k=1}^{K} b_k x_k$$

$$b_k \sim \text{normal}(0, 1)$$

$$\sigma \sim \text{exponential}(1)$$

What happens to the *a-priori* percentage of explained variance $R^2$ as we increase the number of predictors $K$?

# Priors on high dimensional models are weird



# Priors for Simulation-Based Calibration

Under perfect calibration of a posterior approximator, the data averaged posterior equals the prior

$$p(\theta) = \int p(\theta|\tilde{y})\, p(\tilde{y}|\tilde{\theta})\, p(\tilde{\theta})\, d\tilde{y}\, d\tilde{\theta}$$

Further Reading:

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., & Gelman, A. (2018). Validating Bayesian inference algorithms with simulation-based calibration. *arXiv preprint*.

# Bayesian Gamma regression with independent priors

Likelihood:

$$y \sim \text{Gamma}(\text{mean} = \mu, \text{shape} = \alpha)$$

$$\mu = \exp\left(b_0 + \sum_{k=1}^{K} b_k x_k\right)$$

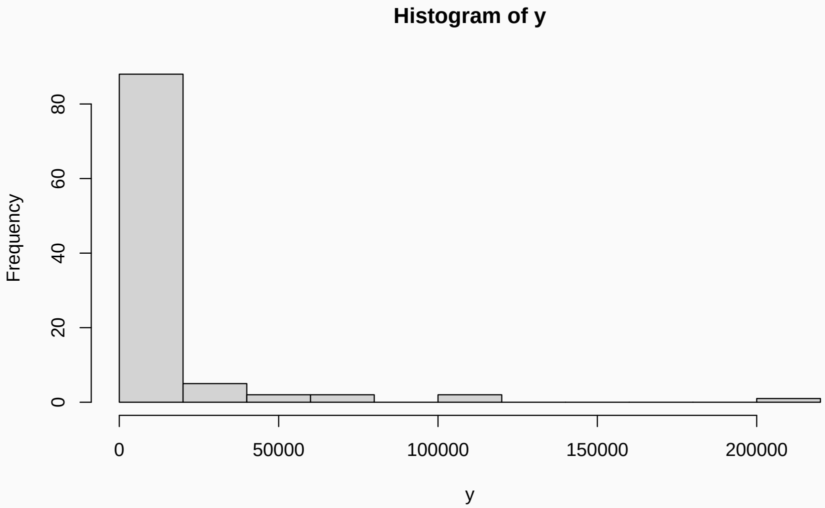
Priors:

$$b_0 \sim \text{normal}(\text{mean} = 2, \text{sd} = 5)$$

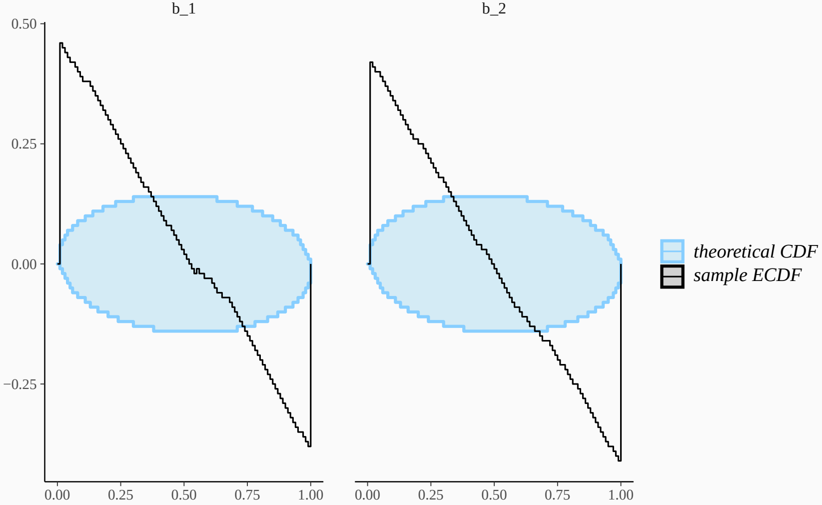
$$b_k \sim \text{normal}(\text{mean} = 0, \text{sd} = 1)$$

$$\alpha \sim \text{Gamma}(\text{shape} = 0.01, \text{rate} = 0.01)$$

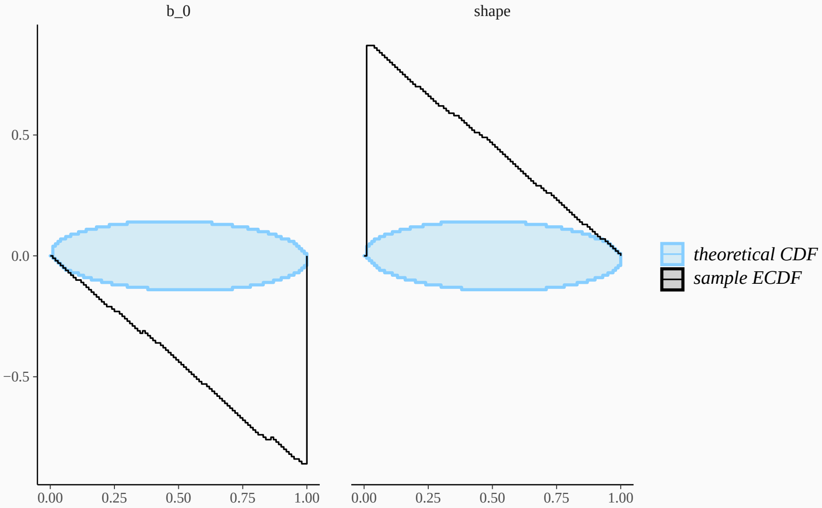
# Example: Data generated with Gamma regression



# SBC for the Gamma regression: b-coefficients



# SBC for the Gamma regression: intercept and shape



# Prior predictive performance is always prior sensitive

Marginal likelihood of model $M$:

$$p(y \mid M) = \int p(y \mid \theta, M)\, p(\theta \mid M)\, d\,\theta$$

$$BF = \frac{p(y \mid M_1)}{p(y \mid M_2)}$$

For Bayes factor-related analysis, priors *always* matter

Further Reading:


Schad D. J., Nicenboim B., Bürkner P. C., Betancourt M., & Vasishth S. (in review). Workflow Techniques for the Robust Use of Bayes Factors. *ArXiv preprint*.

Heck, D., Boehm, U., Böing-Messing, F., Bürkner P. C., . . . , & Hoijtink, H. (accepted). A Review of Applications of the Bayes Factor in Psychological Research. *Psychological Methods*.

But what now?

# (Hierarchical) Joint Priors

For a vector of parameters $\theta = (\theta_1, \ldots, \theta_K)$ set

$$\theta_k \sim \text{prior}(\lambda)$$

with hyperparameters $\lambda = (\lambda_1, \ldots, \lambda_L)$

$$\lambda_l \sim \text{prior}(\tau)$$

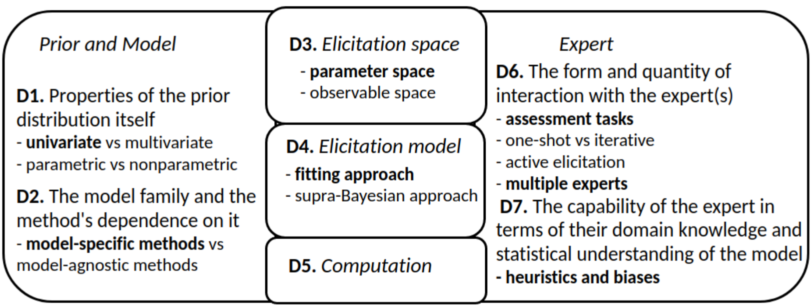
where $\tau = (\tau_1, \ldots, \tau_M)$ is low-dimensional and user choosable

Further reading:


Piironen, J., & Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11(2), 5018-5051.

Zhang, Y. D., Naughton, B. P., Bondell, H. D., & Reich, B. J. (2020). Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association*, 1-13.

# Prior Elicitation




Mikkola, Klami, et al. (in progress). Prior knowledge elicitation: Principles and practice.

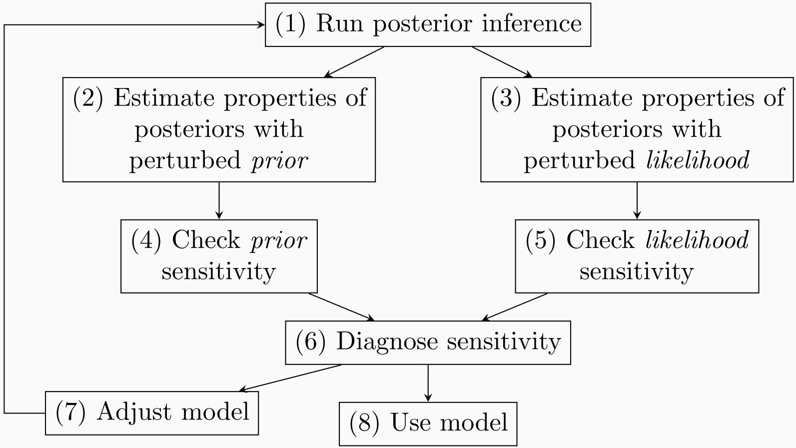
# Prior Elicitation in Observable Space

Given an elicited prior distribution $\hat{p}(y)$ in the observable space, find a sensible prior $\hat{p}(\theta)$ such that

$$\hat{p}(y) = \int p(y \mid \theta)\, \hat{p}(\theta)\, d\theta$$

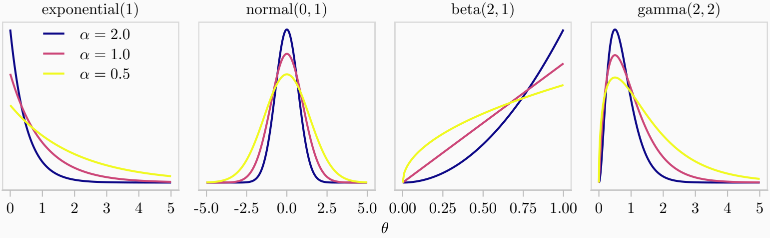
Hartmann M., Agiashvili G., Bürkner P. C., & Klami A. (2020). Flexible Prior Elicitation via the Prior Predictive Distribution. *Uncertainty in Artificial Intelligence (UAI) Conference Proceedings*.

# Prior Sensitivity of the Posterior

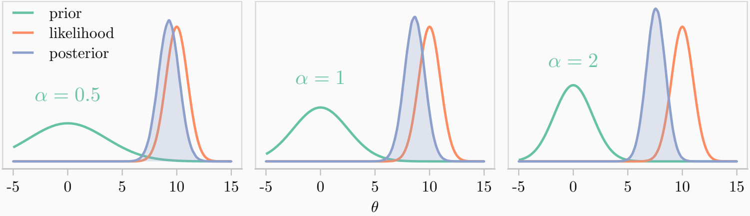



Kallioinen N., Paananen T., Bürkner P. C., & Vehtari A. (in review). Detecting and diagnosing prior and likelihood sensitivity with power-scaling. *ArXiv preprint*.

# Power Scaling of Priors



# Sensitivity to power scaled priors



# Learn more about me and my research

- Website: https://paul-buerkner.github.io/
- Email: `paul.buerkner@gmail.com`
- Twitter: @paulbuerkner