Setting Priors in brms

Bayesian Mixed Effects Models with brms for Linguists

Job Schepens

2026-01-06

Table of contents

1 Setting Priors in brms (20 min)

1.1 Default vs. Weakly Informative Priors

brms uses weakly informative priors by default (not completely flat). However, for psycholinguistics, domain-specific priors are even better.

1.1.1 What is a Prior?

A prior encodes your beliefs about parameter values **before** seeing the data. In Bayesian inference:

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Types of priors:

- **Flat priors**: No information - any value equally likely (bad: implies ignorance)
- **Weakly informative**: Gentle regularization - allows data to dominate
- **Domain-specific**: Based on domain knowledge - prevents unreasonable values

1.1.2 The Intercept Adapts to Your Data!

This is important: The default intercept prior depends on `mean(y)`. Let's see this in action:

Now let's check what default priors brms suggests:

=== DEFAULT PRIORS: Typical RT data (mean log-RT   6) ===

```
               prior     class      coef   group resp dpar nlpar lb ub tag
              (flat)         b
              (flat)         b conditionB
              lkj(1)       cor
              lkj(1)       cor             subject
 student_t(3, 6, 2.5) Intercept
 student_t(3, 0, 2.5)        sd                                   0
 student_t(3, 0, 2.5)        sd                item               0
 student_t(3, 0, 2.5)        sd  Intercept    item               0
 student_t(3, 0, 2.5)        sd             subject               0
 student_t(3, 0, 2.5)        sd conditionB subject               0
 student_t(3, 0, 2.5)        sd  Intercept subject               0
 student_t(3, 0, 2.5)     sigma                                   0
       source
      default
 (vectorized)
      default
 (vectorized)
      default
      default
 (vectorized)
 (vectorized)
 (vectorized)
 (vectorized)
 (vectorized)
      default
```

=== DEFAULT PRIORS: Extreme RT data (mean log-RT   10) ===

```
                prior     class      coef   group resp dpar nlpar lb ub tag
               (flat)         b
               (flat)         b conditionB
               lkj(1)       cor
               lkj(1)       cor             subject
 student_t(3, 10, 2.5) Intercept
  student_t(3, 0, 2.5)        sd                                   0
  student_t(3, 0, 2.5)        sd                item               0
  student_t(3, 0, 2.5)        sd  Intercept    item               0
  student_t(3, 0, 2.5)        sd             subject               0
  student_t(3, 0, 2.5)        sd conditionB subject               0
  student_t(3, 0, 2.5)        sd  Intercept subject               0
  student_t(3, 0, 2.5)     sigma                                   0
       source
      default
```

```
 (vectorized)
      default
 (vectorized)
      default
      default
 (vectorized)
 (vectorized)
 (vectorized)
 (vectorized)
 (vectorized)
      default
```

```
=== Intercept Comparison ===

Typical data prior:   student_t(3, 6, 2.5)

Extreme data prior:   student_t(3, 10, 2.5)

→ The intercept prior CHANGES with data scale!
```

Key insight: The intercept prior automatically scales with your data. This is convenient but has a problem: **if you don't specify priors, your prior assumptions implicitly depend on how you code your variables!**

1.2 Default brms Priors

When you don't specify priors, brms assigns defaults:

- **Intercept**: `student_t(3, mean(y), 2.5)` - **DATA-DEPENDENT!** Centers at your data mean
  - Adapts to your data scale automatically
  - For RT data with mean(log_rt) = 6: allows roughly 150ms-1100ms range
- **Slopes (b)**: `(flat)` - improper uniform prior over ($-\infty$, $+\infty$)
  - No information: any effect size equally likely
  - Technically improper (doesn't integrate to 1)
- **Sigma** (residual SD): `student_t(3, 0, 2.5)` with lower bound 0
  - Weakly informative for residual variance
- **SD** (random effects): `student_t(3, 0, 2.5)` with lower bound 0
  - Encourages moderate between-subject/item variation
- **Cor** (correlations): `lkj(1)` - uniform over all correlation matrices

1.3 Setting Weakly Informative Priors for Reaction Times

For psycholinguistics, it's better to specify priors based on domain knowledge:

```
Our weakly informative priors:

          prior     class coef group resp dpar nlpar   lb   ub tag source
 normal(6, 1.5) Intercept                               4 <NA>       user
 normal(0, 0.5)         b                            <NA> <NA>       user
 exponential(1)     sigma                            <NA> <NA>       user
 exponential(1)        sd                            <NA> <NA>       user
```

```
        lkj(2)       cor                             <NA> <NA>       user
```

1.3.1 Why These Numbers?

1.3.1.1 `normal(6, 1.5)` for Intercept with `lb = 4`

- **Mean = 6** on log scale → exp(6)   403ms (typical RT)
- **SD = 1.5** → 95% prior interval spans [3, 9] on log scale
- **Lower bound = 4** → exp(4)   55ms (minimum physiologically possible motor response)
  - Without `lb`, the prior puts only ~9% probability below this threshold
  - Adding `lb` makes our assumption explicit: we don't entertain impossibly fast RTs
  - In practice, data will dominate anyway, but bounded priors clarify our theoretical assumptions
- But 95% of prior **mass** is between $\pm 1.96 \times 1.5$ around the mean
- This puts most probability on reasonable RTs (100-1500ms), while still allowing outliers

1.3.1.2 `normal(0, 0.5)` for Effects

- **Mean = 0** (no directional assumption)
- **SD = 0.5** on log scale
- 95% prior interval: [-1, 1] on log scale
- Translates to effect sizes of roughly ±65% of baseline (multiplicative)
- Or approximately ±100-150ms for typical RTs around 400-600ms
- **Why not flat?** Flat priors prefer extreme effect sizes (counterintuitive!)

1.3.1.3 `exponential(1)` for Sigma and SD

- Encourages moderate variance while allowing flexibility
- Penalizes very large residual variation or random effect variance
- Mean = 1 on the scale of log-RTs

1.3.1.4 `lkj(2)` for Correlations

- = 2: slight preference for correlations near 0
- Skeptical of strong correlations (like perfect intercept-slope correlation)
- If you truly expected strong correlations, you could use `lkj(1)` or lower

1.3.2 Comparison: Normal vs. Student-t

brms defaults use `student_t(3,  , 2.5)` which has heavier tails than `normal()`. Why switch?

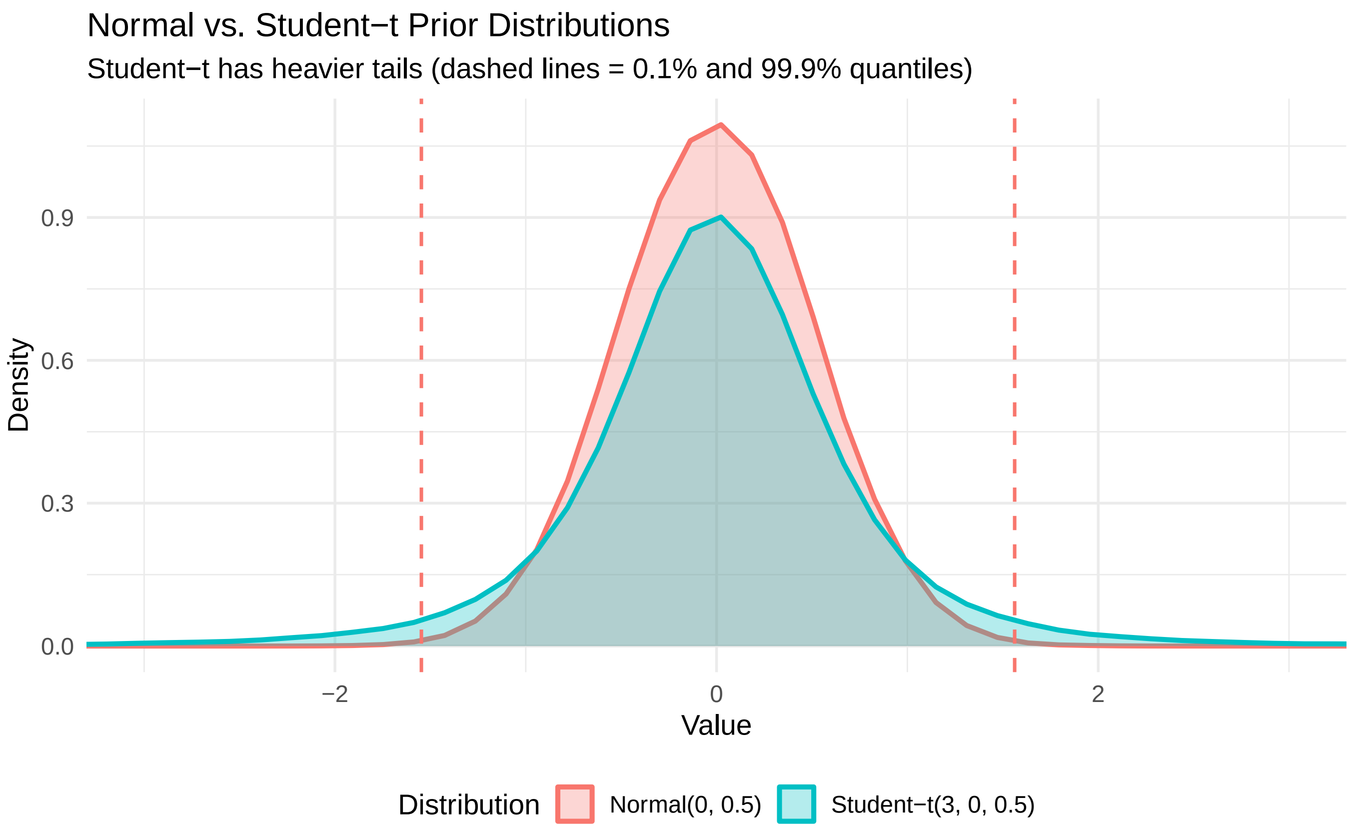
```
Tails comparison (1st and 99th percentiles):

Normal(0, 0.5):        -1.55 1.56

Student-t(3) × 0.5:    -5.26 5.16

Student-t ratio:       3.3 x wider

Student-t allows extreme values 2x more likely than normal!
```



When to use each:

- **Student-t**: Default choice (conservative, robust to outliers)
- **Normal**: When you have strong domain knowledge about plausible ranges (better for RT data in controlled experiments)

Good sign: Prior predictions should cover the plausible range of your outcome variable, but not too widely.

1.4 Summary

Key takeaways:

1. **Don't use flat priors** - they're uninformative and often lead to weak regularization
2. **Default intercept priors adapt to data** - implicit assumptions depend on your coding!
3. **Specify priors explicitly** based on domain knowledge
4. **Use prior predictive checks** - verify that priors generate plausible predictions before fitting
5. **Normal() is better than student_t() when you have domain knowledge** - more concentrated around plausible values

Next steps: - See `02_prior_predictive_checks_rt.qmd` for detailed prior validation - See `03_posterior_predictive_checks_rt.qmd` for checking if the fitted model makes sense