Generative Bayesian Modeling with Implicit Priors

https://arxiv.org/abs/2408.06504

Luna Fazio, Maximilian Scholz, Paul Bürkner

Simulating from Bayesian models

Sampling from the prior predictive:

$$\tilde{y} \sim p(y \mid \alpha) = \int_{\theta} p(y \mid \theta)\, p(\theta \mid \alpha)\, d\theta$$

Key challenge: How can we convince the prior predictive to generate sensible data?

Relevance (examples):

- Bayesian simulation studies
- Simulation-based calibration (SBC)

Precondition on a little bit of data

Bayesian updating of the prior using preconditionig data y_c:

$$p(\theta \mid \alpha, y_c) = \frac{p(\theta \mid \alpha)\, p(y_c \mid \theta)}{p(y_c)}$$

$p(\theta \mid \alpha, y_c)$ is implicit: we usually represent it via draws $\theta^{(s)}$

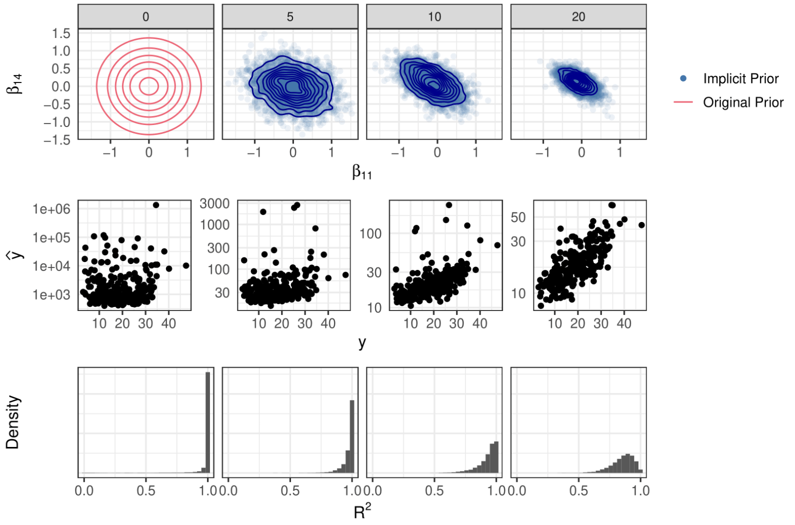
Sampling from the preconditioned prior (posterior) predictive:

$$\tilde{y} \sim p(y \mid y_c, \alpha) = \int_\theta p(y \mid \theta)\, p(\theta \mid \alpha, y_c)\, d\theta$$

Easy to approximate via sampling:

$$\tilde{y}^{(s)} \sim p(y \mid \theta^{(s)})$$

Illustrative example



Posteriors based on implicit priors

Condition on both preconditioning data y_c and "actual" data y:

$$p(\theta \mid \alpha, y_c, y) \propto p(\theta \mid \alpha, y_c)\, p(y \mid \theta) \propto p(\theta \mid \alpha)\, p(y_c \mid \theta)\, p(y \mid \theta)$$

Drawback: Increased computational requirements due to increased amounts of data

But we can usually expect y_c to be small relative to y

Sets of implicit priors

Suppose we do not have a fixed y_c but a distribution $p(y_c)$:

$$p_c(\theta \mid \alpha) := \int_{y_c} p(\theta \mid \alpha, y_c)\, p(y_c)\, d\theta\, dy_c$$

The corresponding prior predictive now looks as follows:

$$\begin{aligned}\tilde{y} \sim p_c(y \mid \alpha) &:= \int_{y_c} p(y \mid y_c, \alpha)\, p(y_c)\, dy_c \\ &= \int_{y_c} \int_{\theta} p(y \mid \theta)\, p(\theta \mid \alpha, y_c)\, p(y_c)\, d\theta\, dy_c\end{aligned}$$

Still easy to sample from $p_c(y \mid \alpha)$ by first sampling from $p(y_c)$ and then from $p(\theta \mid \alpha, y_c)$

Origin of the preconditioning data

- Historical data
- Your current data (if you are careful)
- Data simulated from the model given a sensible (fixed) parameter configuration

Simulation-based calibration (SBC)

SBC is used to verify Bayesian inference algorithms:

- Choose number J of datasets to be generated
- Sample $\theta^{(j)}$ from the prior $p(\theta \mid \alpha)$
- Sample $\tilde{y}^{(j)}$ from the likelihood $p(y \mid \theta^{(j)})$
- Sample S draws $\theta^{(j,s)}$ from $p(\theta \mid \alpha, \tilde{y}^{(j)})$
- Compute rank $R^{(j)} := \sum_{s=1}^{S} \mathbb{I}[f(\theta^{(j,s)}) < f(\theta^{(j)})]$
- If everything is correct, the distribution of ranks will be uniformly distributed

SBC with implicit priors

To run SBC with implicit priors we have to make some adjustments:

- Choose T number of preconditioning data $y_c^{(t)}$
- Choose number J of datasets to be generated per $y_c^{(t)}$
- Sample J draws $\theta_c^{(j)}$ from $p(\theta \mid \alpha, y_c^{(t)})$
- Sample $\tilde{y}^{(j)}$ from $p\left(y \mid \theta_c^{(j)}\right)$
- Sample S draws $\theta^{(j,s)}$ from $p(\theta \mid \alpha, y_c^{(t)}, \tilde{y}^{(j)})$
- Compute rank $R^{(t,j)} := \sum_{s=1}^{S} \mathbb{I}[f(\theta^{(j,s)}) < f(\theta_c^{(j)})]$
- If everything is correct, the distribution of ranks will be uniformly distributed

Case Study 1: Gamma regression

Predicting bodyfat percentage from simple body measurements

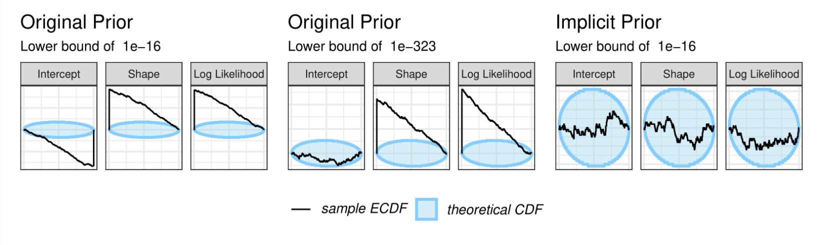
Model description with weakly informative prior:

$$
\begin{aligned}
y_i &\sim \text{Gamma}\left(\alpha, \frac{\alpha}{\mu_i}\right) \\
\log(\mu_i) &\sim \beta_0 + \sum_{k=1}^{13} \beta_k x_{ki} \\
\beta_0 &\sim \text{Normal}(2, 5) \\
\beta_k &\sim \text{Normal}(0, 1) \\
\alpha &\sim \text{Gamma}(0.1, 0.1)
\end{aligned}
$$

Without a lower censoring (or truncation) bound, numerical underflow to zero prevents the model from fitting at all on the prior predictive data

Gamma regression: SBC results

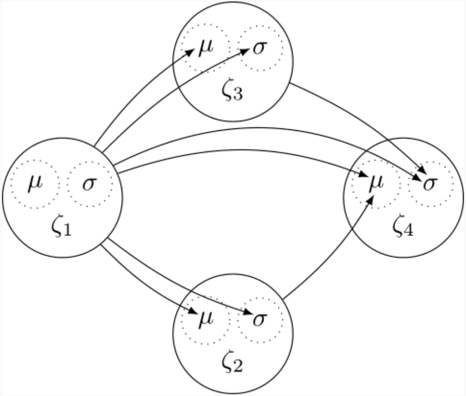
What happens with SBC as we employ lower censoring bounds?



Even very small truncation bounds break SBC

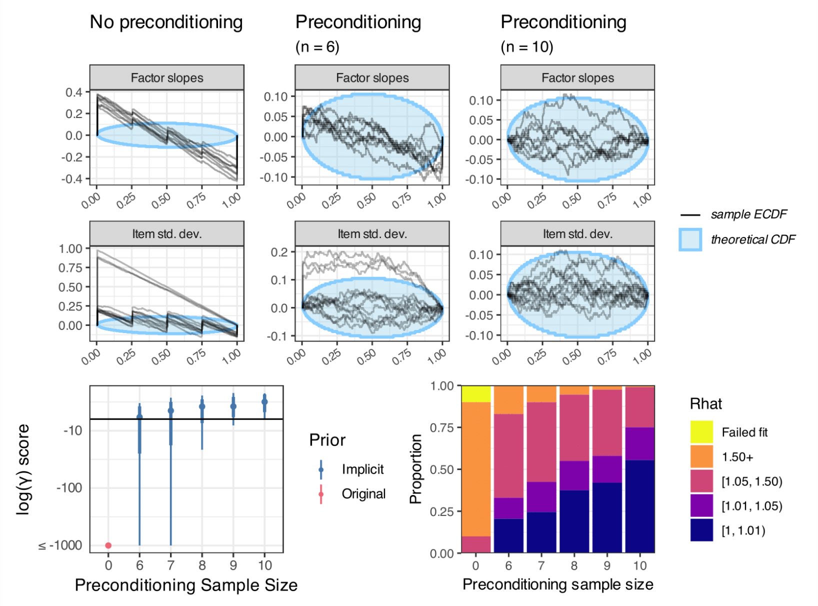
With some preconditioning data ($N = 15$) SBC works perfectly

Case Study 2: Distributional SEM



More details in: https://arxiv.org/abs/2404.14124

Distributional SEM: Results



Summary

- Preconditioning the prior on a little bit of data greatly helps with any kind of prior predictive simulations
- In particular, we can prevent SBC from failing due to priors implying non-sensible data
- The computational overhead of using the resulting implicit priors is small in most cases
- Thanks to Luna and Maximilian for their work on this paper!
- Our preprint is available at: https://arxiv.org/abs/2408.06504
- Also check out my lab's website: https://paulbuerkner.com

Representative implicit priors

Power-scale the preconditioning data likelihood to a weight of M observations:

$$p_M(\theta \mid \alpha, y_{c,N}) := \frac{p(\theta \mid \alpha)\, p(y_{c,N} \mid \theta)^{M/N}}{\int_\theta p(\theta \mid \alpha)\, p(y_{c,N} \mid \theta)^{M/N}\, d\theta}$$

Under mild assumptions the limit of $N \to \infty$ exists:

$$p_M(\theta \mid \alpha, y_{c,\infty}) := \lim_{N \to \infty} p_M(\theta \mid \alpha, y_{c,N})$$