

Turtles all the way down

Richard Torkar

April 2021

UNDERSTANDING WHY MULTILEVEL MODELS should be the default is important and worth repeating:

1. Improved estimates for repeat sampling.
2. Improved estimates for imbalance in sampling.
3. Estimates of variation.
4. Avoid averaging, retain variation.

But it is also important to understand the three main challenges,

- New assumptions (maximum entropy helps us here).
- Estimation is harder (MCMC helps us here).
- Harder to understand (we need to decide what to focus on).

Several concepts are important to understand when working with MLMs, i.e., varying effects (intercepts and slopes), hyperparameters, hyperpriors, regularizing priors, shrinkage, pooling.

One needs to understand why we now must use MCMC.¹

The concept of pooling is key to MLMs. Three types of pooling is commonly used, in this case we use partial pooling. What are the three types, and how do they work? Why do we want partial pooling?

When sampling from MLMs it is not uncommon that you face difficulties. Being careful with prior predictive checks is important, but even that doesn't always help. No, something else needs to be done when you face *divergent transitions*. We need to reparameterize the model. In short, we need to change the model specification so that we still estimate what we want, but the landscape the sampler needs to explore will change. The example in the book, *The Devil's Funnel*, is commonly used to explain the ideas of *centered* and *non-centered* parameterization.

Finally, once we have an MLM that we can put some confidence in, we need to start using it for posterior predictions. Often we divide these predictions into *average distribution*, *marginal distribution*, and *simulations*.

Figure 1: The world supported on the backs of four elephants, themselves resting on the back of a turtle. When asked upon what the turtle stands on, a guru is said to reply, "it's turtles all the way down." Please open this PDF in a reader that accepts GIFs (e.g., Acrobat Reader) and click the picture to enjoy the multimedia.

¹ hint: When a prior is itself a function of parameters, there are two levels of uncertainty.

Changing `adapt_delta` doesn't always help either.

Please see pp. 428–30.

