Bayesian and Frequentist Multilevel Modeling

Andy Grogan-Kaylor

2020-03-12

Contents

1 *Introduction*

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + u_{0j} + e_{ij}$$

All multilevel models account for group structure, in estimating the association of x and y, by including a random intercept (u_0), and possibly one or more random slope terms ($u_1, u_2, etc....$).

Bayesian models may offer some advantages over frequentist models, but may be *substantially* slower to converge.

2 *Conceptual Appropriateness*

Following (Kruschke 2014) all Bayesian models have a *conceptual appropriateness*.

In frequentist reasoning we are estimating the probability of observing data at least as extreme as our data, while assuming a null hypothesis (H_0). Quite often, H_0, *e.g.* $\beta = 0$, or $\bar{x}_A - \bar{x}_B = 0$, is not a substantively interesting or substantively meaningful hypothesis.

3 *Accepting the Null Hypothesis (H_0)*

In Bayesian analysis, we are not rejecting a null hypothesis. Instead, we are *directly estimating the value of a parameter* such as β and are indeed estimating a *full probability distribution* for this parameter.

Such an approach means that we have the ability to accept the null hypothesis H_0 (Kruschke and Liddell 2018). This ability to accept H_0 might possibly lead to theory simplification (Morey, Homer, and Proulx 2018), as well as to a lower likelihood of the publication bias that results from frequentist methods predicated upon the rejection of H_0 (Kruschke and Liddell 2018).

4 *Prior Information*

Bayesian models allow one to incorporate prior information about a parameter of interest.

Prior information may come from the prior research literature, e.g. from systematic reviews or meta-analyses, or expert opinion or clinical wisdom.

5 *Smaller Samples*

Bayesian multilevel models may be better with small samples, especially samples with small numbers of Level 2 units (Hox et al. 2012). It is not clear to what degree this improvement in performance is dependent upon the use of informative priors.

6 *Full Distribution of Parameters*

Bayesian models of all kinds provide full distributions of the parameters (e.g. β's and random effects (u's))–both singly and jointly–rather than only point estimates.

Information about the full distribution of a parameter, such as the estimate of the probability distribution of values of a risk factor, a protective factor, or the effect of an intervention, may be substantively meaningful. Such information may be especially important when the distribution of a regression parameter is non-normal (Van de Schoot et al. 2014).

As Stata Corporation notes, "In a Bayesian multilevel model, *random effects* are model parameters just like regression coefficients and variance components" (StataCorp 2020). This ability to estimate the *distributions* of these random effects means that the *distribution* of the random effect for one group can be compared to another. For example, the *distribution* of a parameter in one country could be directly compared to the *distribution* of that same parameter in another country. One could even estimate the probability that a particular β had a higher value in one group (e.g. country), than in another. As Balov (2016) suggests, this Bayesian approach allows us to "quantify the

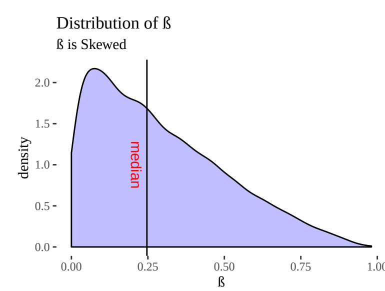

Figure 1: Distribution of a Single Parameter

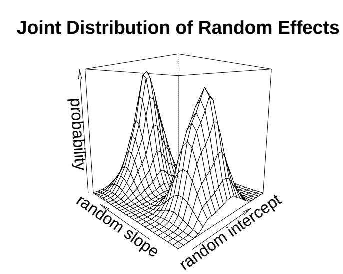

Figure 2: Joint Distribution of Parameters

credibility" of these comparisons, which would not be possible with a frequentist approach.

As an example, Stunnenberg et al. (2018) conducted a Bayesian analysis where the results of the multilevel analysis were used to inform treatment decisions. Here the data were repeated measures on patients, and thus the patients were the groups: "On completion of each treatment set, a Bayesian analysis was conducted to calculate the posterior probability of mexiletine [treatment] producing a clinically meaningful difference in the individual patient."

7 Distributional Models

Bayesian estimators allow one to directly model σ_{u_0}, the variance of the Level 2 units, as a function of covariates (Burkner 2018).

8 Non-Linear Terms

Bayesian estimators allow for the incorporation of non-linear terms (Burkner 2018).[1]

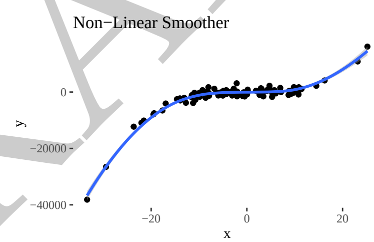

Figure 3: Non-Linear Terms

[1] Such non-linear terms offer ways of non-parametrically fitting curvature. Do they represent over-fitting? Do they provide substantively interpretable results?

9 Maximal Models

Bayesian estimators allow for the estimation of so called *maximal models* (Barr et al. 2013; Frank 2018), which allow for the inclusion of a large number of random slopes, e.g. $u_1, u_2, u_3, ..., etc.$ even when some of those estimated slopes are close to 0.

In contrast, (Matuschek et al. 2017) argue that such a *maximal* approach may lead to a loss of statistical power and further argue that one should adhere to "a random effect structure that is supported by the data."

References

Balov, Nikolay. 2016. "Bayesian Hierarchical Models in Stata." In *2016 Stata Conference*. https://ideas.repec.org/p/boc/scon16/30.html.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. "Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal." *Journal of Memory and Language* 68 (3): 255–78. https://doi.org/10.1016/j.jml.2012.11.001.

Burkner, Paul-Christian. 2018. "Advanced Bayesian Multilevel Modeling with the R Package Brms." *The R Journal* 10 (1): 395–411.

Frank, Michael. 2018. "Mixed Effects Models: Is It Time to Go Bayesian by Default?" http://babieslearninglanguage.blogspot.com/2018/02/mixed-effects-models-is-it-time-to-go.html.

Hox, Joop, Joop J. C.M. Hox, Rens van de Schoot, and Suzette Matthijsse. 2012. "How Few Countries Will Do? Comparative Survey Analysis from a Bayesian Perspective." *Survey Research Methods*. https://doi.org/10.18148/srm/2012.v6i2.5033.

Kruschke, John K. 2014. *Doing Bayesian Data Analysis: A Tutorial with R, Jags, and Stan, Second Edition*. Elsevier Science. https://doi.org/10.1016/B978-0-12-405888-0.09999-2.

Kruschke, John K, and Torrin M Liddell. 2018. "The Bayesian New Statistics: Hypothesis Testing, Estimation, Meta-Analysis, and Power Analysis from a Bayesian Perspective." *Psychonomic Bulletin & Review* 25 (1): 178–206. https://doi.org/10.3758/s13423-016-1221-4.

Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, Harald Baayen, and Douglas Bates. 2017. "Balancing Type I Error and Power in Linear Mixed Models." *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2017.01.001.

Morey, Richard D., Saskia Homer, and Travis Proulx. 2018. "Beyond Statistics: Accepting the Null Hypothesis in Mature Sciences." *Advances in Methods and Practices in Psychological Science*. https://doi.org/10.1177/2515245918776023.

StataCorp. 2020. "Bayesian Multilevel Models." https://www.stata.com/features/overview/bayesian-multilevel-models/.

Stunnenberg, Bas C., Joost Raaphorst, Hans M. Groenewoud, Jeffrey M. Statland, Robert C. Griggs, Willem Woertman, Dick F. Stegeman, et al. 2018. "Effect of Mexiletine on Muscle Stiffness in Patients with Nondystrophic Myotonia Evaluated Using Aggregated N-of-1 Trials." *JAMA - Journal of the American Medical Association*. https://doi.org/10.1001/jama.2018.18020.

Van de Schoot, Rens, David Kaplan, Jaap Denissen, Jens B. Asendorpf, Franz J. Neyer, and Marcel A. G. van Aken. 2014. "A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research." *Child Development* 85 (3): 842–60. https://doi.org/10.1111/cdev.12169.