Testing Order Constraints: Qualitative Differences Between Bayes Factors and Normalized Maximum Likelihood

Supplementary Material

Daniel W. Heck, Eric-Jan Wagenmakers, and Richard D. Morey

June 10, 2015

Using Standard NML and the Bayes Factor with Jeffreys' Prior

In the following, we show that the two qualitative differences also emerge in a comparison of standard NML with the Bayes factor using Jeffreys' prior, which are known to result in asymptotically similar results. Note that Jeffreys' prior should in general not be used for model selection (Jeffreys, 1961, p. 251). However, in order to enable a comparison to standard NML, we will use it for the purpose of illustration.

For the Bayes factor, the uniform prior on the binomial rate θ is replaced by Jeffreys' prior,

$$p(\theta \mid \mathcal{M}) \propto |\mathcal{I}(\theta)\,|^{-1} = \theta^{-1/2}(1-\theta)^{-1/2}, \qquad (1)$$

which gives the beta distribution with parameters $a = 1/2$ and $b = 1/2$ for the full model $\mathcal{M}_1$. For the order-constrained model $\mathcal{M}_0$ with $\theta \leq z$, the density has to be normalized accordingly,

$$p(\theta \mid \mathcal{M}_0) = \frac{\theta^{-1/2}(1-\theta)^{-1/2}}{\int_0^z \theta^{-1/2}(1-\theta)^{-1/2}\,\mathrm{d}\theta}. \tag{2}$$

Since the prior distributions are proportional on the interval $\theta \leq z$, the Bayes factor can be computed as the ratio of posterior-to-prior mass as shown in the main text.

The luckiness NML, defined for a model $\mathcal{M}_i$ as

$$\mathrm{LNML}_i = \frac{p(y \mid \hat{\theta}^L_{y,i}) \exp(-a(\hat{\theta}^L_{y,i}))}{\int_{\mathcal{X}} p(x \mid \hat{\theta}^L_{x,i}) \exp(-a(\hat{\theta}^L_{x,i}))\,\mathrm{d}x}, \tag{3}$$

simplifies to the standard NML by using a constant luckiness function $a(\theta) = 1$ (Grünwald, 2007, p. 313). Similarly as in the main text, the normalizing integral of the NML distribution can simply be computed by summing the maximum likelihoods across the whole data space. The following figures show that both of the two qualitative differences also emerge in a comparison between standard NML and the Bayes factor based on Jeffreys' prior.

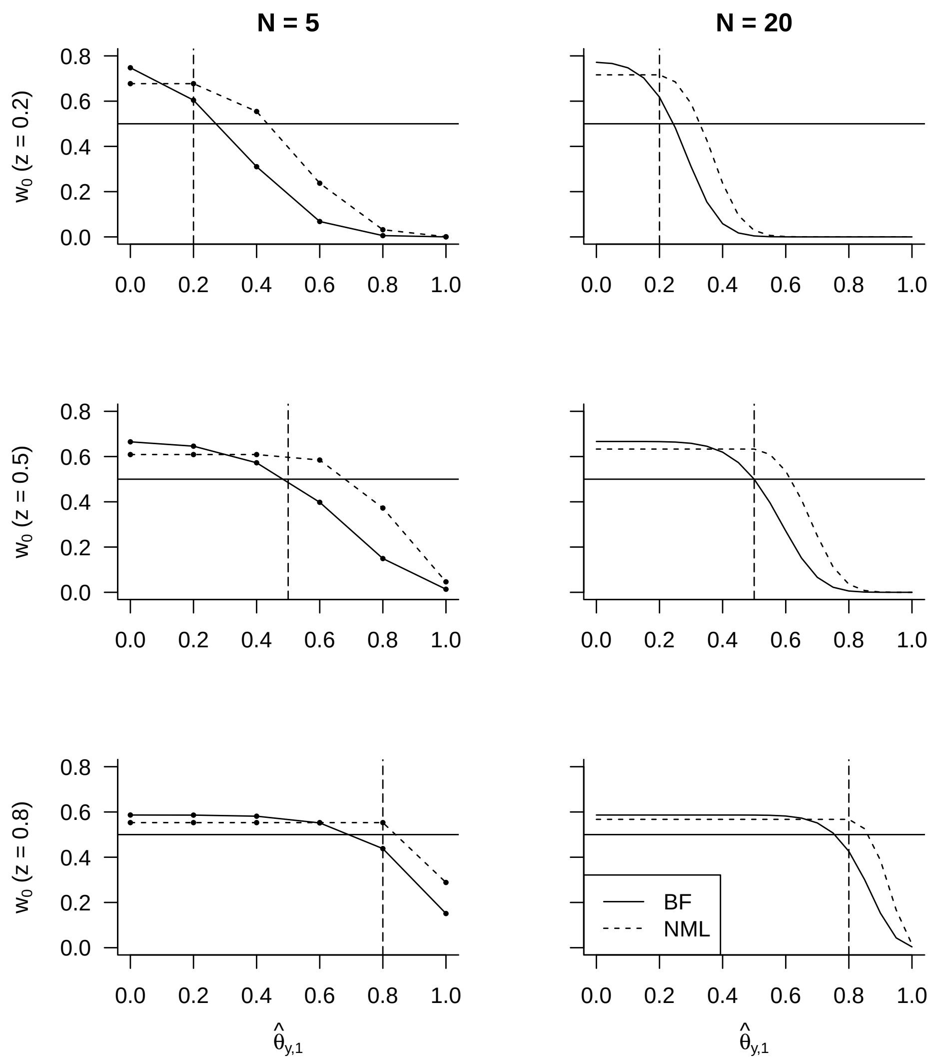

Figure 1: In contrast to the Bayes factor, NML model selection is independent of the observed data if the order constraint is satisfied, as shown by constant model weights w_0 in the range of $\hat{\theta}_{y,1} \leq z$. The boundary z is shown as a vertical, dashed line. For $N = 20$, dots for discrete observations are omitted. Weights that exceed the horizontal line ($w_0 = 0.5$) indicate a preference for the constrained model.

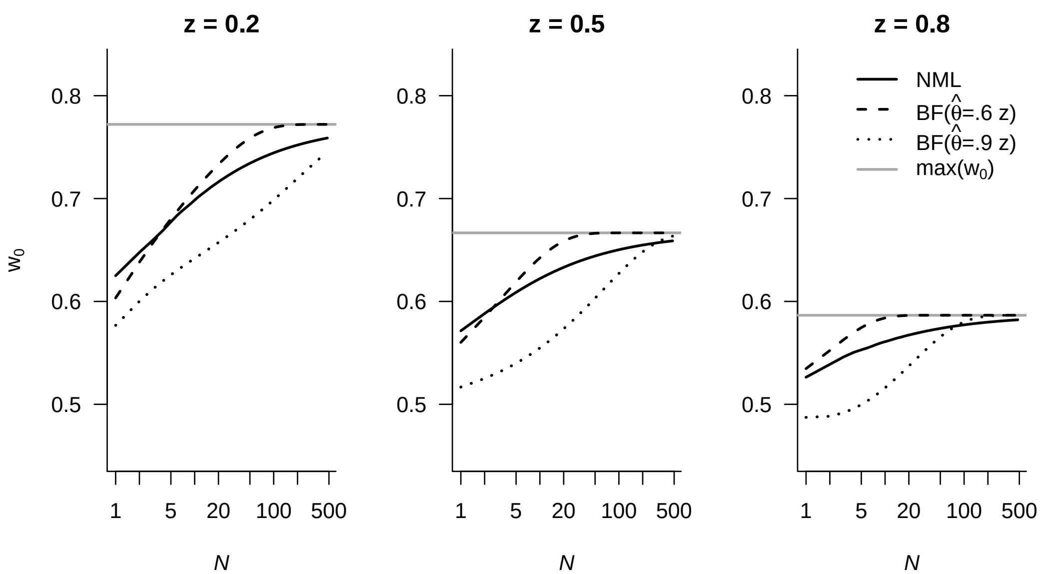

Figure 2: Convergence of standard NML and the Bayes factor with Jeffreys' prior to the maximum model weight in favor of the order constraint (i.e., max(BF)). The depicted Bayes factors are based on data resulting in different ML estimates (e.g., $\hat{\theta}_{y,1} = 0.9z$ for all N).

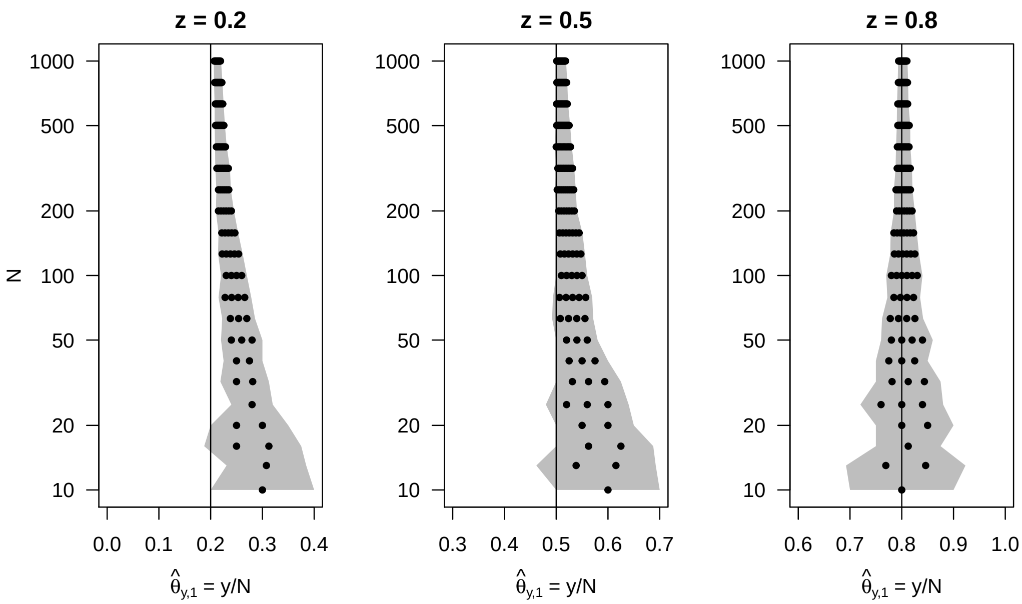

Figure 3: Black dots represent possible data for which NML prefers the order-constrained model, whereas the matching Bayes factor with Jeffreys' prior favors the full binomial model. The gray area shows divergence in model preference and is bounded to the left and the right by data for which the Bayes factor and NML agree which model to prefer.

References

Grünwald, P., 2007. The Minimum Description Length Principle. MIT Press, Cambridge, MA.

Jeffreys, H., 1961. Theory of Probability, 3rd Edition. Oxford University Press, Oxford, UK.