Modeling individual differences in sentence comprehension

Himanshu Yadav*,[1], Garrett Smith*, Dario Paape*, Shravan Vasishth* (*University of Potsdam)
[1]hyadav@uni-potsdam.de

Computational models of sentence processing generally focus on modeling average effects, even though systematic differences in individual-level performance have important implications for theory development [5]. Systematic variation in model parameters can yield new insights into the underlying causes for observed individual differences in behavior [7]. In this work, we present a principled approach to compute individual-level parameter estimates in a particular model of sentence processing, the cue-based retrieval model of Lewis and Vasishth [1, 6] (LV05 model). A common practice in modeling has been to use grid search to estimate model parameters, but this approach has several important limitations [4]: grid search is inefficient and the uncertainty associated with a parameter estimate cannot be taken into account. Approximate Bayesian Computation (ABC) [8] solves these problems. In ABC, simulated data is generated from the model using different parameter values. Parameter values that generate simulated data that is close to the observed data become part of the posterior distribution. Since ABC is rooted in Bayesian statistics, it provides the uncertainty associated with parameter estimates. Here, we use ABC for hierarchically modeling individual differences in the LV05 model. The data come from a relatively large-sample eyetracking study, (181 participants) [3] that investigated (inter alia) facilitatory interference effects in subject-verb agreement dependencies.

The facilitatory interference effect in subject-verb agreement: Cue-based retrieval predicts a facilitatory interference effect when multiple items in memory are candidates for retrieval but none of the items perfectly match the retrieval cues. For example, in (a), retrieval cues (subject, plural) at the verb, *were*, partly match the items *the bodybuilder* (+subject) and *the trainers* (+plural). In this situation, either of the candidates may be retrieved, resulting in a speedup in reading times compared to (b), where there is only one partially-matching retrieval candidate [1].

$(a)\ {}^{*}The\ bodybuilder^{+Subj}_{-PL}\ who\ worked\ with\ the\ trainers^{-Subj}_{+PL}\ amazingly\ were^{Subj}_{PL}\ ...$

$(b)\ {}^{*}The\ bodybuilder^{+Subj}_{-PL}\ who\ worked\ with\ the\ trainer^{-Subj}_{-PL}\ amazingly\ were^{Subj}_{PL}\ ...$

The data used for modeling individual-level effects were the shrunken estimates for each participant, extracted from a hierarchical Bayesian model fit. Then, using ABC (see Note 1), we estimated the latency factor (LF) parameter for each participant. LF is a free parameter in the LV05 model, with larger values leading to longer processing times. In the LV05 model, the difference in reading times between condition (a) and condition (b) above increases with an increase in LF. Therefore, the magnitude of the facilitatory interference effect is directly affected by the LF in LV05. For a participant who is a relatively slow reader, LF would be higher, and as a result the interference effect is predicted to be larger.

Results: From the shrunken estimates of the participant-level effects, we computed participant-level as well as population-level latency factor estimates for LV05 using hierarchical ABC. The estimated latency factor had mean 0.24 (95% credible interval [0.22, 0.27]). The participant-level latency factor estimates were used to generate posterior predictions from LV05. Figure 1 shows individual-level predictions of LV05 and the mean shrunken estimates of by-participant interference effects. Note that the model underestimates the effect size for the biggest effects from some of the participants. This is a positive feature of the model; the unrealistically large estimates from the data, along with their uncertainty, strongly suggests that these are probably Type M errors [2]. The correlation between the participant-level (mean) model predictions and the participant-level (mean) shrunken estimates from the data was 0.84 (95% CrI [0.81,0.87]).

In summary, ABC allows us to model individual-level variability. This is the first attempt to compute the individual-level parameter estimates in the LV05 model, where the uncertainty associated with the parameter estimates has been taken into account. In future work, we plan to use ABC to simultaneously estimate other parameters of the LV05 model such as cue weighting and activation noise, as these also lead to systematic individual-level variability [10].

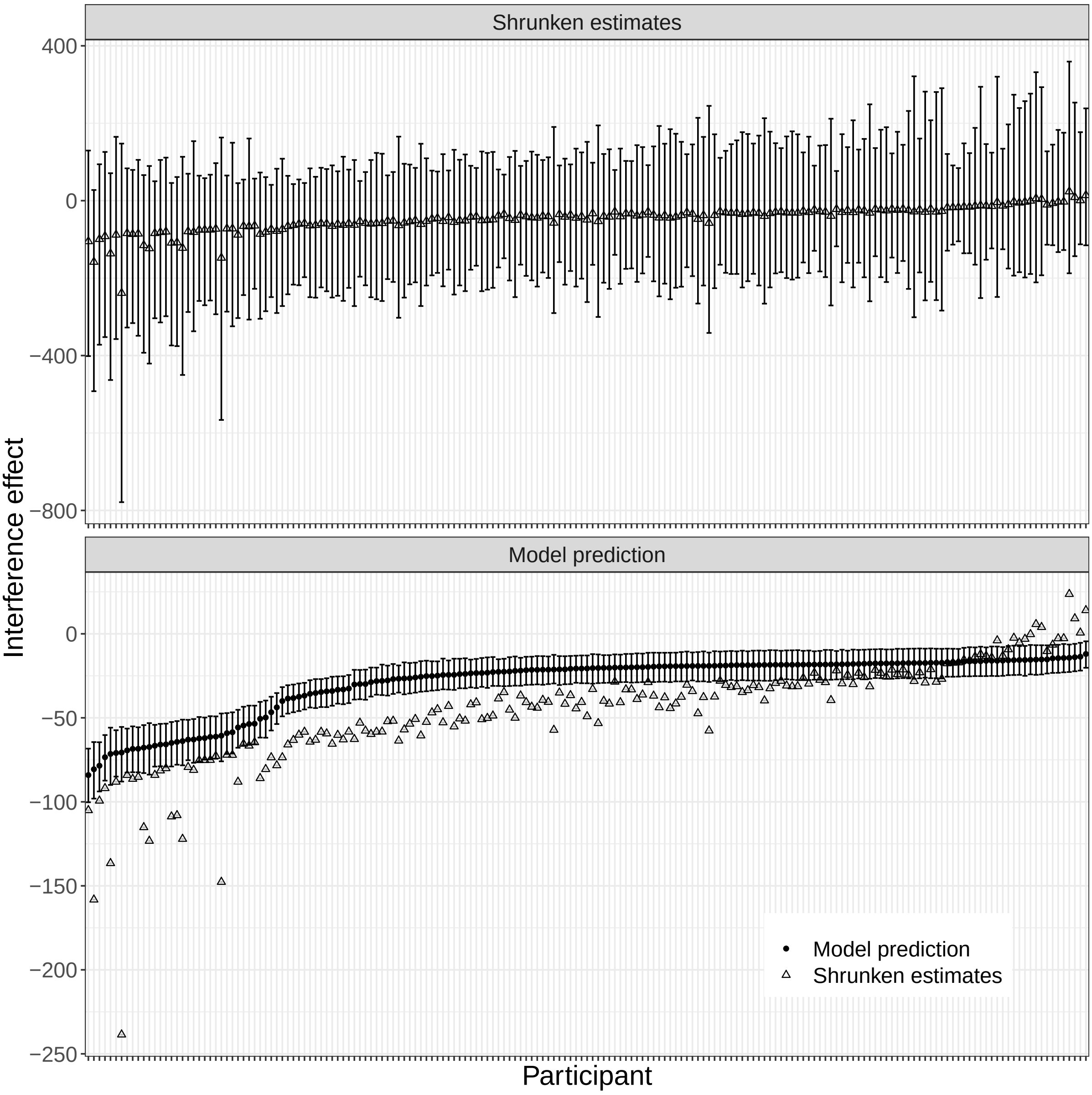

Figure 1: *Shrunken estimates* of participant-level interference effect for agreement dependencies from [3], and predicted interference effect by the model conditional on latency factor estimated by hierarchical ABC method. Note that the (mean) shrunken estimates from the top panel are also shown in the bottom panel.

Note 1: Hierarchical ABC method for individual-level parameter estimation
Consider a simple model, say the output X comes from a function of parameter θ i.e. $X \sim f(\theta)$. Now, say, we collect some data, Y from human participants, and assume that Y comes from $f(\theta)$. A key problem in modeling is to infer what value(s) of the parameter θ would have generated the observed data Y. Bayesian modeling allows us to estimate the distribution of posterior probabilities of the parameter values given the observed data i.e., $\pi(\theta|Y)$. The estimation of posterior distribution requires knowledge about the likelihood, $\pi(Y|\theta)$ (i.e., probability of seeing the data for given parameter value) and the prior knowledge about the parameter, $\pi(\theta)$. For some models, the likelihood function i.e., $\pi(Y|\theta)$ is unknown or difficult to express mathematically. ABC provides a principled approach to approximate the likelihood function using a simulation step. In ABC method, the model is simulated to generate output, say X, for given parameter values and the discrepancy between simulated and observed data i.e., $\rho(X, Y)$, is evaluated to approximate

the likelihood. For example, likelihood can be approximated using a kernel function of $\rho(X, Y)$, such that $\pi(Y_j|\theta_j) \approx \Psi(\rho(X, Y)|\delta)$. ABC algorithms employ the approximate likelihood function to sample from an approximate posterior distribution, e.g., $\pi(Y|\theta) \approx \Psi(\rho(X, Y)|\delta).\pi(\theta)$. In this work, we use hierarchical Gibbs ABC algorithm [9] to estimate participant-level latency factor in the LV05 model.
Let Y be the observed data, Y_j be the observed data for j^{th} subject, θ_j be the individual-level parameter for j^{th} subject, X_j be the simulated output from the model conditional on θ_j, μ be the population-level mean, and σ be the population-level variance.
Consider posterior distribution for an individual-level parameter θ_j conditional on data Y_j and population-level parameters μ and σ. As the conditional distribution of each of the θ_j depends on the data exclusive to the j^{th} subject,

$$\pi(\theta_j|Y_j, \mu, \sigma) \propto \pi(Y_j|\theta_j)\pi(\theta_j|\mu, \sigma) \tag{1}$$

ABC method allows us to approximate the likelihood, $\pi(Y_j|\theta_j)$, as $\pi(Y_j|\theta_j) \approx \Psi(\rho(Y_j, X_j)|\delta)$, where $\Psi(.|\delta)$ is a Kernel function, and X_j is simulated data from the model for θ_j. The Kernel function could be normal distribution e.g., $Normal(0, \delta)$, where δ is the tolerance parameter. Lower the value of δ, better is the approximation of the likelihood.
The approximate conditional posterior for individual-level parameters θ_j becomes,

$$\pi(\theta_j|Y_j, \mu, \sigma) \approx \Psi(\rho(Y_j, X_j)|\delta)\pi(\theta_j|\mu, \sigma) \tag{2}$$

The posterior distribution for population-level parameters conditional on individual-level parameters will be,

$$\pi(\mu|\theta_{1:n}, \sigma) \propto \pi(\theta_{1:n}|\mu, \sigma)\pi(\mu) \tag{3}$$

$$\pi(\sigma|\theta_{1:n}, \mu) \propto \pi(\theta_{1:n}|\mu, \sigma)\pi(\sigma) \tag{4}$$

The conditional distribution of a population-level parameter can be derived as a full conditional distribution in some cases. For example, in the above case, if $\theta \sim Normal(\mu, \sigma^2)$ and conjugate priors are chosen for μ and σ, then $\pi(\mu|\theta_{1:n}, \sigma^2)$ will be a normal distribution and $\pi(\sigma^2|\theta_{1:n}, \mu)$ will be an Inverse-Gamma distribution.

We implement the above method as an ABC sampler. The sampler works as follows. In each iteration: (a) Sample each participant-level parameter from their approximate posterior distribution given in equation 2, (b) Sample each population-level parameter from their conditional posterior (conditional on all other parameters) as shown in equations 3 and 4.

References

[1] Engelmann, Jäger, and Vasishth, 2020. *Cognitive Science*. URL `10.1111/cogs.12800`.
[2] Gelman and Carlin, 2014. *Perspectives on Psychological Science*, 9(6):641–651.
[3] Jäger, Mertzen, Van Dyke, and Vasishth, 2020. *Journal of Memory and Language*, 111. URL `https://psyarxiv.com/7c4gu/`.
[4] Kangasrääsiö, Jokinen, Oulasvirta, Howes, and Kaski, 2019. *Cognitive Science*, 43(6): e12738.
[5] Kidd, Donnelly, and Christiansen, 2018. *Trends in Cognitive Sciences*, 22(2):154–169.
[6] Lewis and Vasishth, 2005. *Cognitive Science*, 29(3):375–419.
[7] Mätzig, Vasishth, Engelmann, Caplan, and Burchert, 2018. *Topics in Cognitive Science*, 10 (1):161–174. Allen Newell Best Student-Led Paper Award at MathPsych/ICCM 2017.
[8] Sisson, Fan, and Beaumont. *Handbook of approximate Bayesian computation*. Chapman and Hall/CRC, 2018.
[9] Turner and Van Zandt, 2014. *Psychometrika*, 79(2):185–209.
[10] Vasishth, Nicenboim, Engelmann, and Burchert, 2019. *Trends in Cognitive Sciences*, 23: 968–982.