Hierarchical Bayesian estimation of the combined-controller model in Gillan et al. (2016)

Joaquín Rapela

October 21, 2021

Abstract

Here I outline the implementation of a hierarchical Bayesian algorithm to estimate the parameters of the combined-controller model described in pages 19 and 20 of Gillan et al. (2016), and I compare these estimates to those from a maximum-likelihood algorithm.

1 Introduction

In Section 2 I describe the estimation method, in Section 3 I mention computer implementation details, in Section 4 I compare results of the application of this and a maximum-likelihood estimation methods, and I conclude in Section 5.

2 Estimation method

I estimated the parameters of the combined-controlled model described in pages 19 and 20 of Gillan et al. (2016) using the method in Huys et al. (2011).

Let

$$
\begin{aligned}
Y_i &= \left\{r_t^i, c_{2t}^i, s_t^i, c_{1t}^i\right\}_{t=1}^T \\
Y &= \{Y_i\}_{i=1}^N \\
\theta^i &= \left\{\alpha^i, \beta^{stage2_i}, \beta^{MB_i}, \beta^{MF0_i}, \beta^{MF1_i}, \beta^{stick_i}\right\} \\
Z &= \{\theta_i\}_{i=1}^N
\end{aligned}
$$

where Y_i and θ_i are the experimental observations and model parameters for the ith subject, respectively, and Y and Z are the experimental observations and model parameters of all subjects.

Briefly, the model in Huys et al. (2011) estimates the parameter m_i^{MAP} of the ith subject model that maximize the posterior distribution, as in Eq. 1.

$$
m_i^{\mathrm{MAP}} = \arg\max_{\theta_i} P(Y_i|\theta_i)\ P(\theta_i|\theta) \tag{1}
$$

using a Normal prior $P(\theta_i|\theta) = \mathcal{N}(\theta_\rangle|\mu, \mathrm{diag}(\nu^2))$.

The prior parameter θ in Eq. 1 were estimated by maximum likelihood, as in Eq. 2.

$$
\begin{aligned}
\theta^{\text{ML}} &= \arg\max_{\theta} P(Y|\theta) \\
&= \arg\max_{\theta} \int P(Y|Z)P(Z|\theta)dZ \\
&= \prod_{i=1}^{N} \arg\max_{\theta_i} \int P(Y_i|\theta_i)P(\theta_i|\theta)d\theta_i
\end{aligned} \tag{2}
$$

We solved the maximization problem in Eq. 2 by expectation maximization. For the E-step we used the Laplace approximation, as in Eq. 3,

$$
P(\theta_i|Y_i, \theta^{\text{old}}) \simeq N(\theta^i|m_i, \text{diag}(\sigma_i^2)) \tag{3}
$$

$$
\begin{aligned}
m_i &= \arg\max_{\theta_i} \; \log P(\theta_i|Y_i, \theta^{\text{old}}) \\
&= \arg\max_{\theta_i} \; \left[\log P(Y_i|\theta^i) + \log P(\theta^i|\theta^{\text{old}})\right]
\end{aligned} \tag{4}
$$

$$
\sigma_i^2 = -\frac{1}{\frac{\partial^2}{\partial \theta^{i2}} \log P(\theta^i|Y^i, \theta^{\text{old}})}
$$

and for the M-step we used the update equations in Eq. 5.

$$
\begin{aligned}
\mu &= \frac{1}{N}\sum_{i=1}^{N} m_i \\
\nu^2 &= \frac{1}{N}\sum_{i=1}^{N}[m_i^2 + \sigma_i^2] - \mu^2
\end{aligned} \tag{5}
$$

Once the prior parameter was estimated, we then maximized the right-hand side of Eq. 1 to calculate MAP estimates of the parameters of every subject model.

3 Computer implementation details

I maximized the parameters in Eq. 4 by constrained gradient ascent using Python and the *minimize* function of the *scipy* library with method *L-BFGS-B*. Gradients were automatically computed using *PyTorch's Autograd*. All parameters were constrained to the interval $[-5.0, 5.0]$ with the exception of α that was constrained to the interval $[0.1, 1.0]$.

In the E-step, parameters of different subjects were estimated concurrently using Python's `multiprocessing` library. In a server with 128 CPUs, each EM iteration took on 86.0 ± 7.64 seconds. The final MAP estimation of each subject model parameters was also done in parallel. The final estimation of the 253 subjects model parameters took 82 seconds in the previous server.

The Python script doHBIEstimations.py[1] performs the hierarchical Bayesian estimation described above.

A problem in the current implementation is that sometimes the estimated variance in Eq. 5 becomes negative. This could be due to roundoff problems. To temporarily solve this problem, when a variance value becomes negative, I set it to zero. However, I want to find a more principled solution.

[1] https://github.com/joacorapela/mlGillanEtAl16/blob/master/code/scripts/doHBIEstimations.py

4 Estimation results

I estimated the parameters of the above model using behavioral data from 253 subjects, each performing 200 trials of the two-stage decision task described in Gillan et al. (2016). Figure 1 top shows box plots and values of all estimated parameters of all subjects models using the hierarchical Bayesian model described above. For comparison, Figure 1 bottom shows estimates from the maximum-likelihood method described previously[2]. Both types of estimates methods look similar, but the hierarchical model seems to achieve a little less variability and less extreme estimates.

5 Final comments

Here I used a simple type of parallelization by taking advantage of multiple processes in a server with a large number of CPUs. More optimal parallelization might be achieved by distributing computing across multiple computers, as in Shattuck et al. (2002) or by using newer Python solutions[3].

References

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., and Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *Elife*, 5:e11305.

Huys, Q. J., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., and Dayan, P. (2011). Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS computational biology*, 7(4):e1002028.

Shattuck, D., Rapela, J., Asma, E., Chatzioannou, A., Qi, J., and Leahy, R. (2002). Internet2-based 3d pet image reconstruction using a pc cluster. *Physics in Medicine & Biology*, 47(15):2785.

[2] http://www.gatsby.ucl.ac.uk/ rapela/mlGillanEtAl16/dawRLmodel.pdf

[3] e.g., https://dask.org/

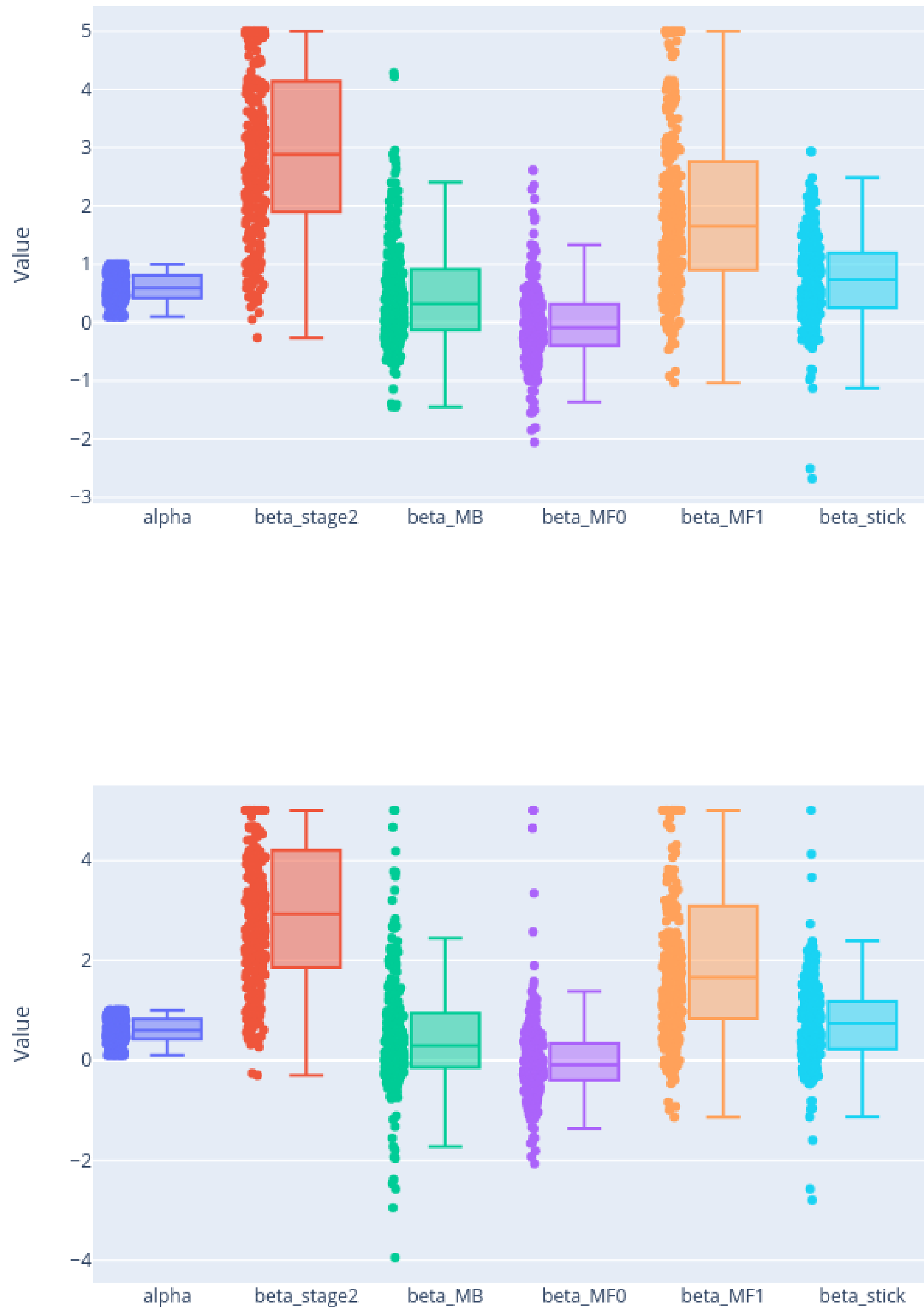

Figure 1: Box plots and values of all parameters of subject models estimated with the Hierarchical Bayesian model (top) and the maximum-likelihood method (bottom). Click on the figure to see its interactive version.