Toolbox for the Reinforcement Learning Drift Diffusion Model

Mads L. Pedersen (m.l.pedersen@psykologi.uio.no)
Department of Psychology, University of Oslo, Forskningsveien 3A
Oslo, Oslo 0370 Norway

Michael J. Frank (michael_frank@brown.edu)
Cognitive, Linguistic & Psychological Sciences, Brown University, 190 Thayer Street
Providence, RI 02912 USA

Abstract:

The continuous development of computational models drives understanding of cognitive mechanisms and their neurobiological underpinnings. Here we extend HDDM, an open source python toolbox for Bayesian hierarchical parameter estimation of the drift diffusion model, to also support reinforcement learning (RL). Moreover, our extension affords the ability to model instrumental learning paradigms in which the choice rule is replaced with the DDM (RLDDM), thus account for evolution of both choices and RT distributions with learning. RLDDM simultaneously estimates parameters of learning and dynamic decision processes by assuming decisions are made by accumulating evidence of the difference in expected rewards between choice options until reaching a decision threshold. Here we validate the model with a parameter recovery test and illustrate the usability of the toolbox, with posterior predictive checks, by fitting pre-collected data on an instrumental learning task.

Keywords: computational modeling; decision making; reinforcement learning; drift diffusion model

Background

Traditional reinforcement learning (RL) models (Rescorla & Wagner, 1972) typically assume static decision processes, e.g. softmax (Luce, 1959), that do not capture the dynamics of choice processes. The drift diffusion model (DDM; Ratcliff, 1978), on the other hand, typically assumes static decision variables, i.e. stimuli are modeled with the same drift rate across trials. The reinforcement learning drift diffusion model (RLDDM; Pedersen, Frank & Biele, 2017) combines dynamic decision variables from RL and dynamic choice process from DDM by assuming trial-by-trial drift rate that depends on the difference in expected rewards, which are updated on each trial by a rate of the prediction error dependent on the learning rate. The potential benefit of the RLDDM is thus to gain a better insight into decision processes in instrumental learning by also accounting for speed of decision making. Indeed, recent studies (Ballard & McClure, 2019; Shahar et al., 2019) have shown that modeling reaction time in reinforcement learning improves identifiability of learning parameters.

We have extended the HDDM toolbox (Wiecki, Sofer & Frank, 2013) with a module that allows users to run hierarchical Bayesian RLDDM models on their dataset. An online tutorial illustrates how to use the toolbox. Here we show that the model can recover parameters and illustrate its potential benefits by analyzing pre-collected data from a two-alternative instrumental learning task.

Methods

The RLDDM assumes expected rewards (Q) for an option *i* on trial *t* is updated according to a delta learning rule:

$$Q_t = Q_{t-1} + \alpha(Reward_t - Q_{t-1}),$$

where the learning rate α weights the rate of learning from the prediction error (reward - expected reward). Further, the model assumes trial-by-trial drift rates (*v*) can be modeled as the scaled difference in expected rewards for the two options (a and b):

$$v_t = (Qa_t - Qb_t) * scaling,$$

where scaling captures sensitivity to rewards. Lastly, combined reaction time and choice on trial t is modeled with the wiener first passage time distribution (wfpt) with parameters for decision threshold (A), non-decision time (T) and drift rate (v_t):

$$wfpt(choice_t+rt_t, A, T, v_t)$$

Parameter Recovery

To validate that the RLDDM can recover parameter values we generated 81 synthetic datasets with different combinations of values for each decision and learning parameter. Each dataset contained 50 subjects performing 70 trials in each of three conditions with varying levels of probability of reward for the best and worst option.

Figure 1 illustrates that the model successfully recovered the values used to generate the synthetic datasets.

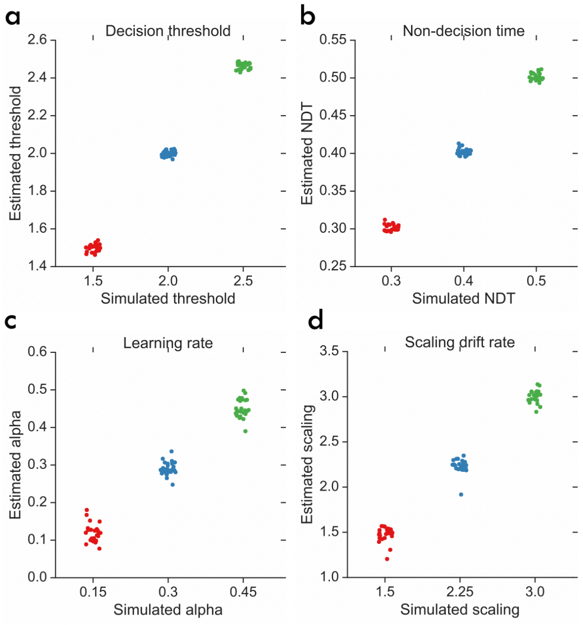

Figure 1: Parameter recovery across decision and learning parameters. Simulated group values on x-axis and estimated group estimates on y-axis. Colors identify the different values used to generate data for each parameter.

Results

The toolbox is created to allow users to easily run and validate models. Here we show an example by applying the model to a pre-collected dataset (Frank et al., 2007) on the probabilistic selection task (PST). The PST includes three conditions with varying levels of reward probability for the best and worst option.

We performed posterior predictive checks, which assess the validity of a model by generating data with estimated parameter values. One hundred datasets were generated by sampling parameter values from the posterior distribution. Figure 2 shows the ability of the model to recreate observed choice and reaction time patterns across learning, separately for the difficulty conditions.

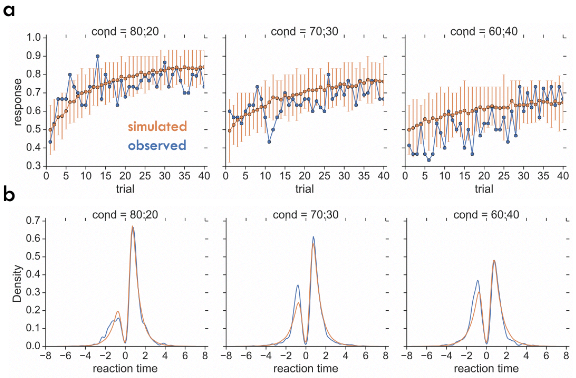

Figure 2: Posterior predictive check of choice and reaction across learning and difficulty conditions. a) Mean observed (blue) and simulated (orange) response in favor of best option across trials. Error bars for generated data represent 90% highest density interval of 100 generated datasets with samples extracted from the posterior distribution of parameters. b) Observed (blue) and simulated (orange) reaction time distributions separated for best and worst option responses as positive and negative RTs, respectively. Cond represents the probability of reward (in percentage) for the best and worst option, respectively.

Conclusion

The RLDDM-toolbox could be a helpful tool for analyzing instrumental learning data and has the potential to be useful given the recent interest in accounting for reaction time in reinforcement learning models.

Acknowledgments

We thank Thomas Wiecki for help with implementing the code. M.L. Pedersen is funded by the Research Council of Norway (262327).

References

Ballard, I. C., & McClure, S. M. (2019). Joint Modeling of Reaction Times and Choice Improves Parameter Identifiability in Reinforcement Learning Models. Journal of Neuroscience Methods. https://doi.org/10.1016/j.jneumeth.2019.01.006

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proceedings of the National

Academy of Sciences, 104(41), 16311–16316. https://doi.org/10.1073/pnas.0706111104

Luce, R. D. (1959). Individual choice behavior. Oxford, England: John Wiley.

Pedersen, M.L, Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. Psychonomic Bulletin & Review, 24(4), 1234–1251. https://doi.org/10.3758/s13423-016-1199-y

Ratcliff, R. (1978). A theory of memory retrieval. Psychological Review, 85(2), 59. https://doi.org/10.1037/0033-295x.85.2.59

Rescorla, R., & Wagner, A. (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement (Appleton-Century-Crofts, New York).

Shahar, N., Hauser, T. U., Moutoussis, M., Moran, R., Keramati, M., & Dolan, R. J. (2019). Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. PLOS Computational Biology, 15(2), e1006803. https://doi.org/10.1371/journal.pcbi.1006803

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. Frontiers in Neuroinformatics, 7, 14. https://doi.org/10.3389/fninf.2013.00014