Delta rule model in a gradually changing perceptual task.

C. A. Velázquez
National Autonomous University of Mexico

A. Bouzas
National Autonomous University of Mexico

Decisions often take place in environments that change over time. Availability of food in foraging animals may vary gradually as a function of source growth or continuous intake, so the position of objects in space moving at a certain velocity. Having accurate beliefs under such circumstances allows behavior to be better allocated and to optimize reward, e.g., by changing to a richer foraging location or predicting the correct position of an object moving towards us. In stable environments, error-driven algorithms approximate the state of the world by reducing the discrepancy of estimates and outcomes as new observations arrive. One common expression to compute this is the Delta rule:

$$\delta_t = r_t - \mu_t$$

$$\mu_{t+1} = \mu_t + \alpha\delta \tag{1}$$

where, at a given trial t, the prediction error δ_t , weighted by the learning rate α, is used to update the current estimate μ_t after outcome r_t is observed. Evidence from Experimental Psychology (Bush & Mosteller, 1951; Rescorla & Wagner, 1972; Miller et al., 1995) and Neuroscience (Schultz, 1997; Niv, 2013) provides support for this algorithm as a plausible mechanisms of learning in mammals, and it has also been implemented as an effective solution in multiple machine learning problems (Sutton & Barto, 1988). However, one of its limitations is the inability to predict behavior in non-stationary environments partly due to the fixed nature of the learning rate parameter. For example, in change-point problems, having a low α makes predictions during stable periods accurate but causes a slow adaptation after a change. A high α has the opposite effect, making inaccurate predictions during stability but having a quick adaptation to changes. Adjusting this parameter after the change-point (Nassar et al., 2010) and using multiple delta rules with their own learning rates (Wilson et al., 2013) are some of the possible solutions reported on literature.

Standard Delta rule model also has difficulties performing in environments that change gradually in time. Consider the simulation shown in figure 1. In this example, availability of food on a given day (blue dots) is given by sampling from a Gaussian distribution with variance 1 and mean X changing by $X_{t+1} = X_t + v_t$,where v_t is a velocity term following a random walk with variance 1. In 1A, predictions of Delta rule (red line) are plotted using the simulated data and $\alpha = 0.1$. In this case, previous estimations outweigh new observations, making the algorithm change so slowly that is unable to track the generative mean. Figure 1B shows the same sequence using $\alpha = 0.9$. Although it does a better job than 1A, given the high learning rate, the Delta rule predictions on a given day resemble the just-perceived outcome the day before. As a result, when the generative mean moves upwards (trial 4 to 12) or downwards (trial 18 to 23) the model adapts as though it were one step behind.

SimulationDelta.png

Figure 1. Simulated data for availability of food changing over days. (A) Delta rule model with $\alpha = 0.2$ weights previous estimations more than new outcomes, making it perform poorly. (B) Delta rule model with $\alpha = 0.9$ responds inaccurately as the mean of observations begins to drift (trial 4-12 and 18 to 23). Its high learning rate makes future predictions roughly the same as just-perceived outcomes.

One way around this problem is to estimate in a single number the rate at which the generative process changes. This quantity can be updated similarly to the delta rule algorithm as new observations arrive and added linearly to the final prediction. Such a model leads to a simple version of a Kalman Filter (Kalman, 1960; Sutton, 1992). Previous work has shown that Kalman Filter is a useful tool when modelling behavior of subjects in drifting environments, for example in Multiple Cue Probability Learning tasks (MCPL) when the cue-outcome relations vary over time (Speekenbrink, 2010), or at the acquisition and extinction of behavioral responses (Kakade & Dayan, 2000, 2002). Our goal was to build on these studies to model predictions of subjects in a noisy and gradually changing visual task. We implemented a Delta rule model that assumes participants a) estimate the rate of change in the generative process, b) modulate the impact of new observations via their learning rates depending on the level of noise and c) accompany each estimation with a certain degree of uncertainty.

We developed a novel task where participants have to predict the future location of a spaceship that orbits around the earth. Its position is sampled from a Gaussian distribution with a mean that changes over trials. We set four different values of variance that defined our conditions.

Our major findings suggest that predictions of participants change gradually in a trial-by-trial manner following the generative process and that a model that assume they estimate the rate of change is efficient in describing their performance.

Models

Hyperboloid model
Tradeoff model

Bayesian Estimation

Two models were evaluated: one from the alternative-based choice family (Hyperboloid model) and the other from the attributed-based choice family (Trade-off model). The latter model can account for intransitive patters, which the alternative-based models cannot accommodate. The models' comparison was done with Bayesian Modeling in order to infer individual parameters.

The used notation for describing the bayesian analysis was adopted from Lee & Wagenmakers (2014). In this representation, shaded nodes correspond to observed variables, whereas unshaded nodes stand for latent variables. Double-bordered nodes are deterministic, while single-bordered nodes are stochastic. Circles represent continuous variables, and squares portray discrete variables.

Hyperboloid Model

The evaluated models were analysed for time and probability, therefore there was a total of four bayesian models. The analysis was based on the statistical work of Scholten et al. (2014) were they made model comparison with bayesian analysis. However in the present work, the parametrical estimation was done at an individual level.

$$Q(w(ts), w(tl)) = \frac{\kappa}{\alpha} log\left(1 + \alpha\left(\frac{w(tl) - w(ts)}{\vartheta}\right)^{\vartheta}\right) \qquad (2)$$

The four models share the following structure: All of them have three rectangles that enclose independent replications of 1) number of participants i; 2) number of questions j; 3) number of repetitions for the same questions r. The individual responses, $C_i jr$, were modeled with a Bernoulli process with parameter , that represents the probability of choosing the larger reward. The node $x_i^s j$ is the small outcome, and $x_i^l j$ is the larger outcome.

The difference in between models of time and probability are the following: For the time models, the nodes $t_i^s j$ y $t_i^l j$ are the sooner delay and the later delay, respectively. For the probability task, the probabilities were transformated into odds against the receipt of a reward: $\Omega = \frac{p-1}{p}$. Therefore, the node $\Omega_i^s j$ are the odds against receipt of the safer smaller reward, and $\Omega_i^l j$ are the odds against the receipt of the riskier larger reward.

Figures tal and tal, are the bayesian graphical models for time and probability of the hyperboloid function. Both have the same structure where the probability of choosing the larger reward, $\theta_i j$ is obtained with the discounted values of both alternatives; where the discounted value of the larger later/riskier is divided between the sum of the discounted value of the larger later/riskier and the discounted value of the smaller sooner/safer. Each discounted value is gathered with the discount factor and the absolute outcome.

The discount factor $d_i j^s l$ has the hyperboloid structure with two free parameters: the parameter κ_i which reflects the degree of discounting (with higher values of it associated with steeper discounting); and the parameter τ_i, which determines the shape of the discounting function (Green, Cita). These parameters were given a lognormal distribution with a mean of zero and a standard deviation of one. This distribution does not allow negative values, it is heavy tailed, and it puts little weight on values extremely close to zero.

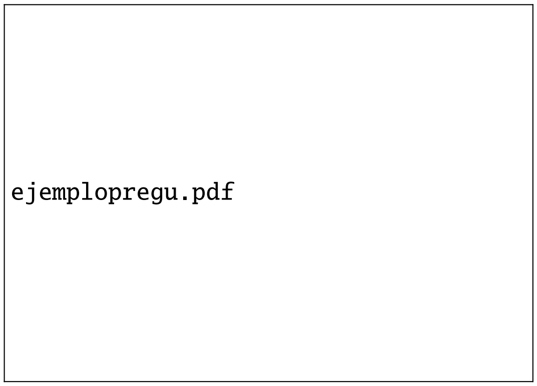

Figure 2. hola

Trade-off Model

The Trade-off model has four main functions. The first one is the time-weighing function which gives a weight for time/probability. en la figura **??**

The second one is the value function which gives a subjective value for outcomes and has the same mathematical structure of the weighing function. The third one, is the tradeoff function that evaluates the differences among the time/probability intervals. The last function, is one that gives the probability of choosing the larger reward according to Luces's Choice axiom.

The time weighing function is a concave over the time/probability. It has the parameter τ which is diminishing absolute sensitivity to time/probability. Increments in τ make the absolute time/probability have less weight. The value function has the same structure of the weighing function but with parameter γ a goes over outcomes. The differences between the values of the outcomes are going to be call effective compensations.

Method

Participants. 25 students from the School of Psychology. For their participation they enteren in a raffle were they could win a card of Netflix, iTunes or Spotify, according their preference.

Procedure. There were two experimental sessions, one for the time task, and another for the probability task. Each session last about 35 minutes. Both sessions were differed about 24 hours. Both tasks were developed in PsychoPy v1.83.04 (Pierce, 2007). For the time task, the two alternatives were the smaller sooner reward and the larger later reward. For probability, there were the smaller safer and the larger riskier. Both tasks had the same alternative selection structure where participants had to click in the letter which indicated the prefered alternative; the letter was illuminated when the mouse was arround it (Figure 1), see Appendix A for instructions. The positions of the smaller and larger options were randomized across trials.

Experimental Desing. The used task, was created to evaluate interval effects in intertermporal choice, therefore we made a analogous task for risky choice. The design consisted in 12 fixed alternatives that have within them linear increments; for outcome the increment was $150 Mexican pesos, for time was one week, and for probability was .10 of probability. The combination of alternative pairs created 22 questions which were classified in 4 sets (6 questions each): 1) Short delays, 2) Long delays, 3) Long intervals, and 4) Small outcomes, long intervals. The sets for probability were the following: 1) Large probabilities, 2) Small probabilities, 3) Long intervals, and 4) Small outcomes and long intervals. Each set had 6 questions, two questions were used y 2 sets. Therefore in total there were 22 formal questions for the time task and 22 questions for the probability task. The tasks were based on the 2nd study of Scholten et al. (2014). The outcomes were presented in Mexican pesos, with an exchange rate of $10 pesos per dollar. In order to consider choice variability, each question was presented 10 times. In such a manner, there were 220 trials for the time task and 220 for the probability task.

Results

For each participant (one per row) there are two graphics, time and probability. In each graphic, there are 24 bars (6 by set/color). Each bar represents a question, the leftmost end indicates the smaller reward, and the rightmost end the larger reward; alternatives are indicated in the superior axis. The darkness of the bar represents the participant's choice proportion of the larger reward. For example, if the participant opted for the larger reward 8 times, the proportion of the bar is darker 8/10. Furthermore, in each bar there are two vertical lines, the pink one represents the prediction from the Hyperboloid model while the purple one represents the prediction from the Trade-off model. The right panel shows the posterior densities from each model.

Choice proportions from pooled data, following the structure of Figure 2. Each rectangle has 25 lines, representing the choice proportions for the large reward (darker color) from all participants. The lines are organized from lowest to highest choice proportion. The left side of the graphic shows the choice proportions for the time task, while the right side shows the choice proportions for the probability task.

Predictions of choice proportions for all participants fo reach model, ordered by sets/colors. The X-axis shows the choice proportions of the larger reward, while the Y-axis presents the prediction of choice proportions from each model. The squares show the number of choice proportions and predictions.

Discussion

The main finding was that most of the sample chose the larger reward in the shortest intervals, and the smaller reward in the longest interval (superadditivity); there was more variability in medium-size intervals. These patterns were found mostly in the time task.

Results were better accounted for by the Trade-off (attributed-based) model than by the Hyperboloid (alternative-based) model. However, the Trade-off Model overpredicted interval effects for Small Outcomes.

Furthermore, participants did not show the same intransitive pattern across both tasks.

References

Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The perception of probability. *Psychological review, 121(1),* 96.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering,* 82(1), 35-45.

Miller R.R., Barnet R.C. & Grahame NJ (1995) Assessment of the Rescolra-Wagner model. *Psychological Bulletin* 117: 363âĂŞ386.

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience,* 30(37), 12366-12378.

Rescorla R.A. & Wagner A.R. (1972) A Theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF, editors, Classical conditioning II: current research and theory. New York: Appleton Century Crofts. chapter 3. pp. 64âĂŞ99.

Ricci, M. & Gallistel, R. (2017). Accurate step-hold tracking of smoothly varying periodic and aperiodic probability.*Attention, Perception, & Psychophysics,* 1-15.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science,* 275(5306), 1593-1599.

Speekenbrink, M., & Shanks, D. R. (2010). Learning in a changing environment. *Journal of Experimental Psychology: General,* 139(2), 266.

Sutton, R. S. (1992). Gain adaptation beats least squares. In *Proceedings of the 7th Yale workshop on adaptive and learning systems* (Vol. 161168).

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning (Vol. 135)* . Cambridge: MIT Press.

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2013). A mixture of delta-rules approximation to Bayesian inference in change-point problems. *PLoS computational biology,* 9(7), e1003150.