Drift Diffusion Modeling Approaches to Tone Discrimination Tasks

A Report Presented in Fulfillment of
PSYC 396

Megan Kairiss

Department of Psychology
Under the supervision of Dr. Rosemary Bagot
McGill University
April 2019

Contents

List of Figures

List of Tables

1 Introduction

Two alternative-force decision processes, otherwise known as diffusion processes, are heavily applied in the fields of psychology and cognitive neuroscience to model the noisy evolution of thought of an individual participating in a binary-choice task. We discuss the application of a specific diffusion process, Drift Diffusion Models (DDM), to model a Tone Discrimination (TD) task in mice. The use of this computational approach on decision making tasks has been shown to help gauge a better understanding of the underlying decision-making processes in such tasks. Modeling drift rates allow us to examine differences and disruptions in how animals accumulate information from their environment. The focus of this paper is to explore this modeling approach of choice response time in decision making, as well as examine the usability of a particular DDM: the hierarchical drift diffusion model (HDDM).

Drift Diffusion models fall under the category of sequential-sampling models, and more specifically, "random walk" models [1]; however DDMs operate under the assumption that the information which motivates a decision process is accumulated continuously across time, instead of in discrete steps (thus, utilizing continuous time and continuous evidence criteria). Response Time (RT) is used as a metric of information accumulation, under the assumption that individuals make rapid-decisions (typically under 200ms) when prompted by a stimulus. Diffusion models explain the decision making process using several free parameters which are formalized later in this paper.

Drift Diffusion models allow for an experimental design with only two alternative options (multi-alternative choice models do exist, however we will not discuss these). Subjects noisily fluctuate between two options until a certain decision threshold has been reached, akin to the vibrating frequencies of a tuning fork (Ratcliff 1978). These decision thresholds represent positive and negative response boundaries, and we paramaterize the speed at which the process approaches a boundary (measured as the mean amount of information accumulated per second) as the *drift rate*.

The application of diffusion processes to study choice behavior has been available for more than 40 years, however the use of these diffusion models on the choice reaction time of animal behavior has yet to gain popularity. There are currently two model approaches commonly applied to DDM modeling efforts, namely Fast-DM and HDDM. Both toolkits estimate the same parameters, but differ in their methods of parameter estimation (Bayesian vs non-Bayesian, hierarchical vs non-hierarchical). In this paper, we present a demonstration of the Hierarchical Drift Diffusion Model (HDDM). Recent advances in computational methods has enabled the computation of a hierarchical Bayesian posterior distribution of DDM parameters within a group, with the aid of Markov Chain Monte Carlo (MCMC) sampling to estimate the joint distribution of free parameters via HDDM. This model utilizes the powerful assumption that computing enough parameter estimates across a population can strengthen the recovery of parameters of the individual; while other approaches (such as Fast-DM) often struggle to provide a robust framework for parameter identification using alternative methods, such as Maximum Likelihood Estimation.

We explore the implementation of a hierarchical drift diffusion model on a Tone Discrimination (TD) task. All code, statistical analysis, and data used in this paper can be found here: https://github.com/MeganKai/Drift-Diffusion-Modeling.

[1] Random walk models are discrete-time precursors of diffusion models (Ratcliff et. al 2016), and do not involve a decision starting point bias - hence we need a more specific implementation to represent cognitive biases of an individual.

2 Theoretical Formalization

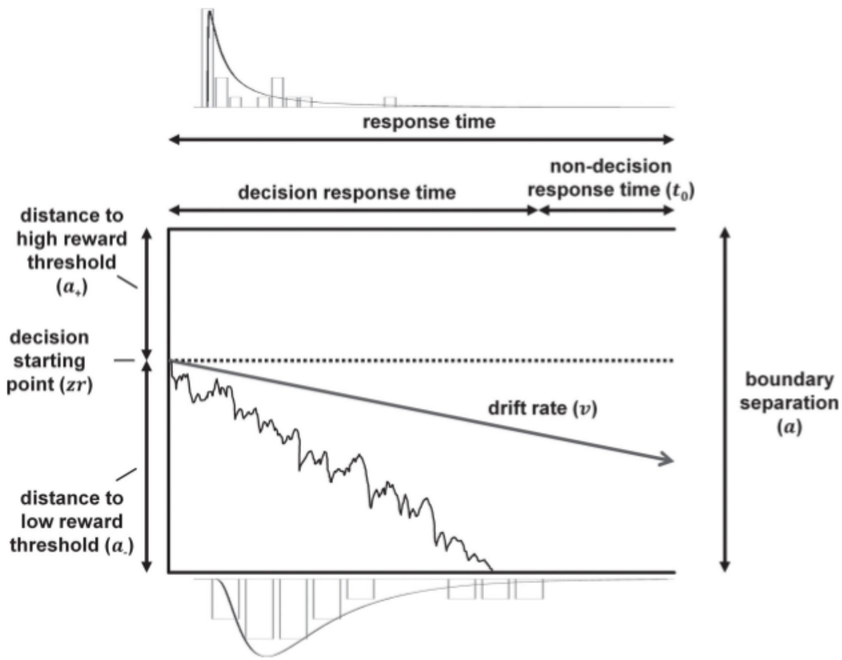
We rely on empirical measures of response time to implement a drift diffusion model. The decision-making process is initiated once prompted by a stimulus, and we accumulate information from the environment until a decision boundary is reached. The initial starting point of our decision-making process has the potential to be biased towards either threshold, or this starting point can be halfway between either boundary, indicating no bias towards either a positive decision boundary or negative decision boundary. The relative distance between these two boundaries explains how much information must accumulated before there is enough evidence to reach a boundary. When a threshold is close to the starting point, this process happens quite fast, but is indicative of more impulsive decisions (Wiecki et. al 2013). The opposite case represents more thoughtful responses, with more time taken before a threshold is reached. It is important to note that a measure of response time actually includes time taken for both decision-making and time for action execution. Figure 1 shows the relationship of all drift diffusion model parameters, and illustrates the division of response time into both decision response time ($\boldsymbol{t}$) and time taken for non-decision processes ($\boldsymbol{t_0}$).

Drift rate, $\boldsymbol{v}$, corresponds to the strength and direction of the influence of information accumulation on the diffusion process as we approach a threshold. The width of the interval between two decision thresholds is parameterized by $\boldsymbol{\alpha}$. The bias parameter $\boldsymbol{z}$ represents the decision starting point between the two thresholds (assumed to be at 0.5 if no bias is present). Values from individual thresholds, both positive and negative, can be derived from α and z. We can decompose this single threshold parameter α into two separate threshold values, corresponding to α_+ and α_- boundaries as seen in Figure 1. Mathematically, these values are computed as in (1):

$$\begin{aligned} \alpha_+ &= z \cdot \alpha \\ \alpha_- &= \alpha - \alpha_+ \end{aligned} \tag{1}$$

It is also possible to include additional parameters to account for bias and inter-trial variability: $\boldsymbol{sv}$ for inter-trial variability in the drift rate, and $\boldsymbol{szr}$ for inter-trial variability in the decision starting point. However, these parameters cannot be used for individual participants since they would lack any statistical strength. Rather, their intent is to help explain variability across participants.

Figure 1: Diffusion Model Parameters



Source: C.A. Hales, E.S.J. Robinson, and C.J. Houghton 2016.
α_+ denotes the distance to the high reward decision threshold, α_- denotes the distance to the low reward decision threshold, and the decision starting point zr is considered to include bias when $\alpha_+ \neq \alpha_-$ (i.e. the individual has either a negative or positive cognitive bias towards either outcome). The average drift rate v follows the trajectory of the information accumulation. The horizontal axis depicts overall response time, split into decision response time t and non-decision response time t_0. Note that in the model, we parameterize the decision threshold with respect to a single parameter: α

3 Experimental Approach

3.1 Experimental Design

Seven mice were trained on a tone discrimination task, consisting of 50 trials per mouse per day across a period of 18 days. Individual trials involved a lever extension (two levers, one correct lever, and one incorrect depending on the trial type). Trials are divided into two types: positive trials, and negative trials. Both trials involve a tone stimulus. During positive trials, a tone stimulus is emitted at 9000Hz and 63db, levers are extended, and the animal must choose the correct lever to receive a reward. During negative trials, a lower-frequency tone is emitted at 2000Hz and 75db, and animals must select the other lever to prevent the onset of a half-second shock of 30mA. Note that during this experimental set-up, only the first response is recorded; e.g. if a mouse's first response is incorrect, any further lever presses within that individual trial, correct or incorrect, will be ignored.

Figure 2 shows the progression of each animal learning the task, in terms of overall percent correct lever presses. It is clear that the Tone Discrimination task requires a heavy learning curve, thus the data actually used to perform model parameter estimation is taken from the final three days of training, at which point the number of correct lever presses reaches a plateau.

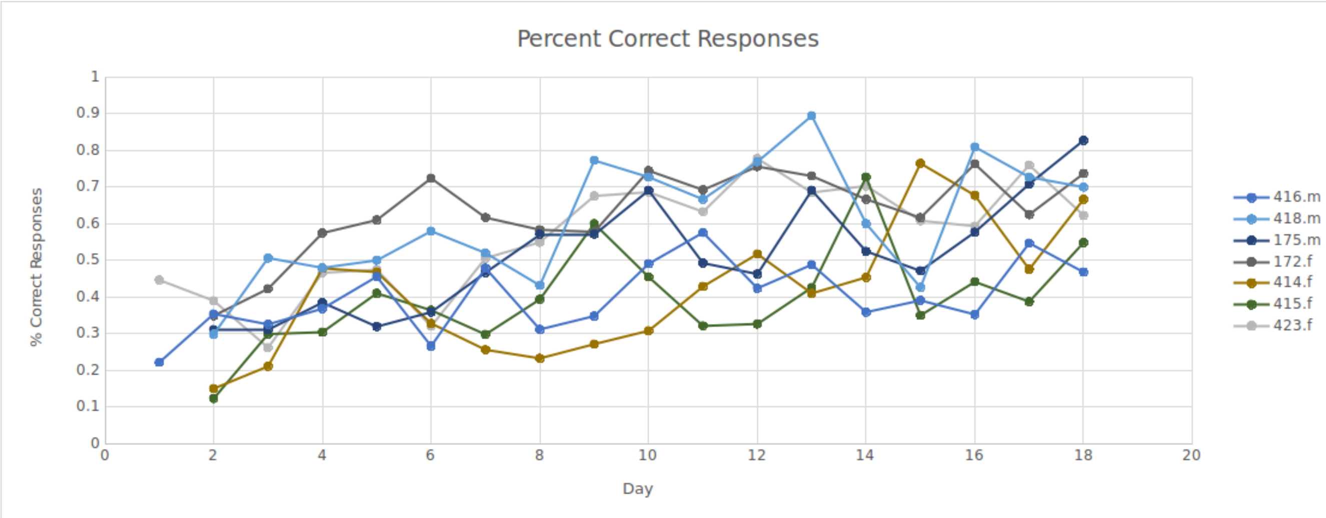


Figure 2: Percent Correct Responses by Participant
Across seven animals, four females and three males performed the Tone Discrimination task. Percent correct responses are a function of trials completed overall, and trials with which the animal selected the correct lever in response to the tone stimulus.

3.2 Data Collection

Inputs to both modeling tools, HDDM and Fast-DM, are identical. In order to estimate parameter values, a DDM model requires the following: response time (in terms of latency to first lever press), stimulus (tone), and response (in terms of a correct or incorrect lever press). A neurological measure (e.g. EEG, photometry) can also be provided as appropriate. A sample of the first few lines of the input table to both models is provided below in Table 1:

Subj	Stim	Response	RT
278	2	1	3.46
278	2	0	56.16
278	1	1	26.38
280	1	1	54.62
280	1	1	108.21

Table 1: Sample input to a DDM

In practice, we might remove the 'Subj' column depending on whether or not we are interested in hierarchical or non-hierarchical parameter estimates. 'Stim' represents the stimulus; in our Tone-Discrimination task, animals are presented with one of two tones per trial. The stimulus input is not limited to two options; depending on the experimental design you can have several. The response column is necessarily always binary.

4 Modeling

4.1 Hierarchical Drift Diffusion Model: HDDM

HDDM is an open-source drift diffusion modeling toolkit for Python (Wiecki et. al 2013) that allows for a hierarchical Bayesian parameter estimation of drift diffusion models. HDDM has been demonstrated to successfully recover model parameters through simulation efforts, and is perhaps most useful in instances with a minimum number of trials. This is particularly advantageous in animal behaviour research, which is constrained by the nature and noise of animal behaviour.

How does HDDM model the DDM parameters? Since we are using a Bayesian inferential method to compute the full posterior distribution of DDM parameters, we are able to provide much more information than a simple expected value, or point-estimate. In accordance with the model parameters explained above in the theoretical formalization, we compute the conditional probability of a forward-model of DDM: the probability of observing the observed response times, given the joint probability of the model parameters υ, α, and z (the decision starting point) is derived in (2):

$$
\begin{aligned}
f(x|v,a,z) =& \frac{\pi}{a^2} \exp\left(-vaz - \frac{v^2 x}{2}\right) \\
& \times \sum_{k=1}^{\infty} k \exp\left(-\frac{k^2 \pi^2 x}{2a^2}\right) \sin(k\pi z)
\end{aligned} \tag{2}
$$

The model computes the optimal parameters for a given RT distribution even if variability between parameters is present, via analytic integration over the infinite sum. The joint probability distribution is approximated via Markov Chain Monte Carlo (MCMC) sampling.

HDDM was used to fit data from the Tone Discrimination task outlined in the previous section. Starting values for the model were obtained via gradient ascent optimization, and 10000 samples were taken to estimate the posterior values. The model was fit to include the potential probability of outliers occuring with p = 0.05 in the data. In order to test for MCMC model convergence, we used the same method outlined in (Wiecki et al 2013); the Gelman Rubin diagnostic. This test essentially compares the variance between several models, to make sure that sampling and burn-in rates are adequate, under the premise that multiple chains sampled from the same data should return indistinguishable results. The Gelman-Rubin diagnostic, applied on five instances of our model, returned ranges all within a tight bound of 1, indicative of a properly converged markov chain (values of < 1.001 are considered good).

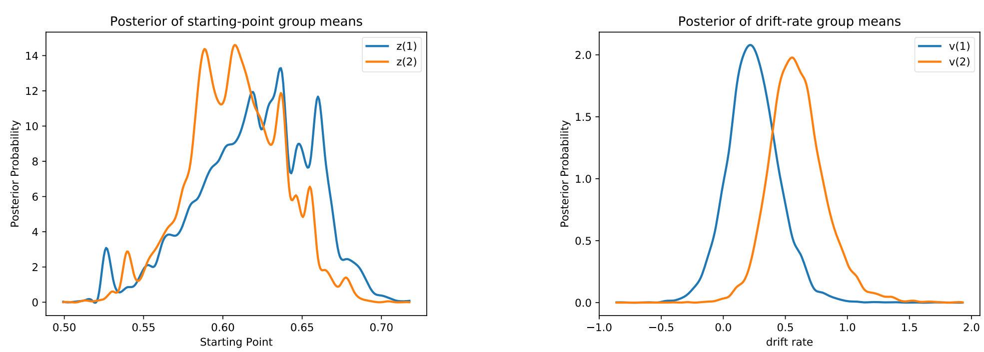
The computed parameter estimates are outlined below:

Parameter	Mean	Std
a_1	4.743	0.65
a_2	6.495	0.91
v_1	0.249	0.20
v_2	0.598	0.22
z_1	0.619	0.035
z_2	0.607	0.023

Table 2: HDDM Parameter Estimates
Mean and standard deviation of parameter estimates for a given model which models effects of two tones: a subscript of (1) represents positive trial types, with the high-tone stimulus, and a subscript of (2) represents negative trial types with the low-tone stimulus.

We see a clear difference in drift rates for the two separate tones, indicating that the animals were indeed able to discriminate between the two stimuli. The negative trial type corresponding to the lower-frequency tone has a larger drift rate, suggesting that the diffusion process occurs within a slightly shorter duration with a smaller percent error. The negative trial type also has significantly large decision threshold width, indicating that this condition requires much more information accumulation before decisions can be made. This larger value also corresponds to fewer error responses, and altogether a large response time distribution.

Perhaps the most interesting difference across trial types is that of the drift rate. HDDM allows us to easily visualize the posterior probabilities of the individual drift rates. Figure 3 shows the difference in means of both stimuli.



(a) Starting Point Posterior

(b) Drift Rate Posterior

Figure 3: Posterior Probabilities of Drift Rate, Starting Point
(a) HDDM provides a visualization of drift rates with respect to different conditions/stimuli. Here we can see a clear difference in the drift rates dependent on the negative or positive trial type: z(1) denotes trials containing the positive tone stimulus and z(2) denotes trials with the negative tone stimulus. (b) No clear difference in the values of the starting point parameter across different trial types is present: with v(1) denoting trials with the positive tone stimulus, and v(2) denoting trials with the negative tone stimulus.

4.2 Alternative Modeling Methods

Fast-DM is another computational tool for diffusion modeling and analysis, introduced by Voss and Voss 2007. Fast-DM estimates the same parameters as HDDM, however uses a non-Bayesian and non-hierarchical approach to parameter estimation. Fast-DM calculates the Cumulative Distribution Function (CDF) for the parameters of the Response-Time distribution from the diffusion model. This process involves solving the Kolmogorov backward equation. Recent developments of Fast-DM have allowed users to select an alternative method of parameter estimation (Voss et. al 2015), selecting between Kolmogorov-Smirnov, Chi-Squared, or Maximum Likelihood. Fast-DM is however much more user-friendly, and does not require significant programming knowledge or familiarity with Python. This comes with the caveat that we cannot perform simultaneous estimation of individual and group parameters, and thus we cannot provide measures of uncertainty in these parameters. The main disadvantage of Fast-DM is in the number of trials required for proper parameter estimation, with the minimum number of trials recommended per participant being approximately 100.

5 Discussion

This paper has shown a proof of applicability of hierarchical Drift Diffusion models to choice response behaviour. Drift Diffusion Models offer a powerful explanation of underlying decision-making processes, and HDDM provides an easily reproducible framework for parameter estimation. Specifically, hierarchical methods of DDM parameter estimation allow for full Bayesian analysis of parameters, conditional modelling of the effects of individual stimuli, and optional integration of neural mechanisms.

HDDM is perhaps superior to Fast-DM as it allows for population estimates of the parameters, which can in turn positively constrain the individual participant. Instead of assuming an uninformed prior, one can use the power of group distributions to influence and constrain parameter estimates of an individual. Additionally, Fast-DM generally requires a higher total number of trials-to-participant ratio, and uses an alternative means of computing a probabilistic distribution over the free parameters.

6 Next Steps

The data used in the model above was collected from a Tone Discrimination task. In order to investigate the effects of cognitive biases on decision making, we will further use HDDM to model parameters on an Ambiguous Cue Interpretation Task. This task is the same as the Tone Discrimination, however we include an additional ambiguous tone stimulus that lies between the high-frequency tone and the low-frequency tone. The expectation is that the model will have a biased decision starting point for this ambiguous tone, allowing the use of these metrics to explain positive and negative cognitive biases among animal participants. Due to the significant amount of time required to gather data for an ACI task, this paper was confined to describing the modeling efforts of a TD task.

Another benefit of using a hierarchical drift diffusion model is that we are capable of including a measure of neural activity into the model. The next step in this study will involve the use of optogenetics and *in vivo* calcium imaging techniques to bidirectionally manipulate and observe neural activity in striatal afferent pathways and jointly estimate posterior distributions.

We are also interested in exploring the idea of combining reinforcement learning models with measures of response time. An interesting recent study by Pedersen et al (2017) described efforts to use drift diffusion model parameters as a choice updating rule in a reinforcement learning model, instead of using the typical expected value choice process via a softmax function. The drawback of regular reinforcement learning models is their inability to capture within-trial decision making processes. Rather, they model the decision process as trial-by-trial accumulation of information. Pedersen et al. aim to combine drift diffusion models (which are advantageous when investigating an individuals' decision process with respect to choice boundary thresholds) with reinforcement learning models (which are useful in describing an individual's learning rate and choice sensitivity within their environment). Response latency is usually not encoded into reinforcement learning models, thus the integration of response time to RL modeling can provide a more complex explanation of decision processes.

References

[1] C. A. Hales, E. S. J. Robinson, and C. J. Houghton, "Diffusion modelling reveals the decision making processes underlying negative judgement bias in rats," *PLOS ONE*, vol. 11, pp. 1–25, 03 2016.

[2] M. L. Pedersen, M. J. Frank, and G. Biele, "The drift diffusion model as the choice rule in reinforcement learning," *Psychonomic Bulletin & Review*, vol. 24, pp. 1234–1251, Aug 2017.

[3] R. Ratcliff, "A theory of memory retrieval," *Psychological Review*, vol. 85, no. 2, pp. 59–108, 1978.

[4] A. Voss, J. Voss, and V. Lerche, "Assessing cognitive processes with diffusion model analyses: a tutorial based on fast-dm-30," *Frontiers in Psychology*, vol. 6, p. 336, 2015.

[5] A. Voss and J. Voss, "Fast-dm: A free program for efficient diffusion model analysis," *Behavior research methods*, vol. 39, pp. 767–75, 12 2007.

[6] T. Wiecki, I. Sofer, and M. Frank, "Hddm: Hierarchical bayesian estimation of the drift-diffusion model in python," *Frontiers in Neuroinformatics*, vol. 7, p. 14, 2013.

[7] R. Ratcliff, P. L. Smith, S. Brown, and G. McKoon, "Diffusion decision model: Current issues and history," *Trends in Cognitive Sciences*, vol. 20, 03 2016.