Title
Are First Passage Time Distributions Necessary for Drift-Diffusion Modeling?

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Author
Lalvani, Shamal

Publication Date
2023

Peer reviewed

Are First Passage Time Distributions Necessary for Drift-Diffusion Modeling?

Shamal Lalvani (shamal.lalvani@northwestern.edu)
Department of Electrical and Computer Engineering, 2145 Sheridan Road
Evanston, IL 60208 USA

Aggelos Katsaggelos (a-katsaggelos@northwestern.edu)
Department of Electrical and Computer Engineering, 2145 Sheridan Road
Evanston, IL 60208 USA

Abstract

Drift diffusion models are used to model evidence accumulation in two-choice forced-tasks. The traditional approach to fitting Ratcliff's standard drift diffusion model (where the drift and diffusion are constant) usually involves explicit modeling of the first passage time distributions of the upper and lower boundaries or likelihood approximations. We present the very first technique, to the best of our knowledge, that foregoes use of explicit modeling of the first passage time distributions with a random forest regressor. A random forest regression model that takes the first five moments of the response time distribution, and the upper boundary termination proportion, is used to predict the drift and diffusion parameters from response time data. A training set of response time samples of size 2500 from 121 distinct drift-diffusion pairs is used to train the random forest regressor. On a testing set of 10,000 distinct drift-diffusion combinations with response time sample sizes of 40, we find that our model surpasses techniques that make use of some form of analytical modeling of the first passage time distributions of the boundaries for prediction of the diffusion rate, but not the drift rate. We conclude that the application of machine learning to drift-diffusion modeling of empirical data is a topic worth further investigation.

Keywords: drift-diffusion models; two-alternative forced-tasks; mathematical psychology; machine learning

Introduction

Drift-diffusion models, originally introduced by Ratcliff (see Ratcliff, 1978), typically apply to two-alternative forced-choice tasks, in which a decision maker must choose between two alternatives, typically within seconds. Drift-diffusion models have been influential in the fields of neuroscience and psychology. Drift-diffusion models have neurological correlates in brain regions including the Subthalamic Nucleus (Herz et al., 2008), correlate to Dopaminergic activity in perceptual decision making (Beste et al., 2018), and potentially segregate individuals with mental health disorders such as Schizophrenia, OCD and Social Anxiety Disorder from healthy controls (Dillon et al., 2022; Gupta et al., 2022). Drift diffusion models assume that in the decision making process, the decision maker accumulates information according to a diffusion process (Fig. 1). When the information level (to be loosely interpreted as a surrogate for the relative information needed to make a decision) first reaches either an upper or lower boundary, the process terminates, and a decision is made, with each boundary corresponding to a different decision.

Extensions to the drift diffusion model exist, such as allowing for prior distributions on the drift and diffusion parameters (Wiecki, T. V., Sofer, I., & Frank, 2013) and allowing for decisions between more than two options (Roxin, 2019).

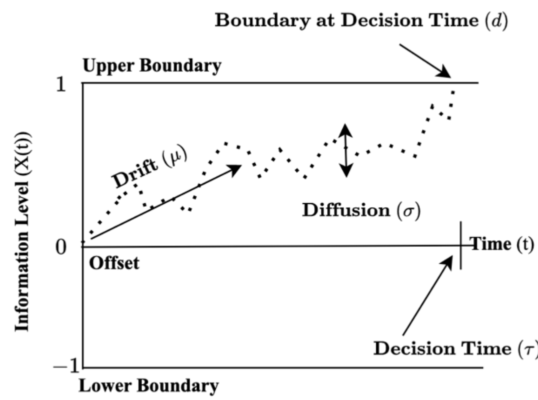

Figure 1. Illustration of a sample path (in dotted lines) from the drift- diffusion model. The sample path reaches the upper boundary (+1), after which the process terminates.

In this paper, we consider Ratcliff's model, where the drift and diffusion parameters are constant in a two-choice task, and the goal is to learn the drift and diffusion parameters from response time data (Ratcliff, 1978). The challenge here is that decision times and decisions are observed (Ratcliff & Tuerlinckx, 2002), rather than a continuous or well-represented signal of the information accumulation process. Therefore, the data for the information values cannot contain more than two values, each corresponding to the decisions at the decision boundaries. This implies that numerical techniques for solving stochastic differential equations are of less value, as they are suited to continuous signals, rather than first passage time statistics.

In the literature, drift-diffusion models are typically fit to data by techniques that involve analytical modeling of the first passage-time distributions of each boundary. Specifically, the drift and diffusion parameters are varied to minimize a distance metric between the empirical distribution and the analytical distribution. Some

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

implementations in the literature include the weighted least squares method, maximum likelihood estimation and the chi-squared method (Ratcliff & Tuerlinckx, 2002). Likelihood approximation algorithms have also been established in cognitive science and applied to drift diffusion models (Turner & Sederberg, 2014; Turner & Van Zandt, 2012; Turner & Van Zandt, 2018), as well as machine learning techniques such as the use of deep neural networks in implementation, where more generally, the drift and diffusion rate are not constant (Wiecki, T. V., Sofer, I., & Frank, 2013).

In this paper, we make use of random forest regression to explore if it is possible to recover drift and diffusion parameters from empirical data. To the best of our knowledge, this is the first paper to do so.

Methods

Table 1: List of symbols.

Symbol	Meaning	Unit
μ	Drift Rate	Seconds^{-1}
σ^2	Diffusion Rate	Seconds^{-2}
t	Time	Seconds
dt	Time-step for Diffusion Model Simulation	Seconds
τ	Time at Process Termination	Seconds
m	Mean Time of Process Termination	Seconds
d	Boundary at Process Termination (1 or -1)	NA
X(t)	Information Level at Time t	NA
W(t)	Standard Wiener Process at Time t	NA
μ_i	The i'th Drift Rate in the Training Set	Seconds^{-1}
σ_i	The i'th Drift Rate in the Training Set	Seconds^{-2}
T_i	The i'th set of response times in the Training Set	Seconds
d_i	The i'th Set of Decisions in the Training Set	NA
μ'_i	The i'th Drift Rate in the Testing Set	Seconds^{-1}
σ'_i	The i'th Drift Rate in the Testing Set	Seconds^{-2}
T'_i	The i'th Set of Response Times in the Testing Set	Seconds
d'_i	The i'th Set of Decisions in the Testing Set	NA

Simulating the Drift-Diffusion Process

The diffusion process was simulated in Python as follows. The process initially started at the offset, which was fixed to zero for all simulations. The process was updated according to the standard diffusion equation (1), where dt = 0.01 seconds and W(t) is a standard Wiener process (i.e., dW(t) is normally distributed with mean zero and variance dt)

$$X(t+dt) = X(t) + \mu\,dt + \sigma\,dW(t) \qquad (1)$$

The run-time of the process was fixed to be a maximum of 20 seconds, after which the process terminated if it had not reached the upper or lower boundary (+1 and -1, respectively). If the process terminated before a decision boundary was reached, the decision variable (d) was chosen to be the boundary that is closest to the final information level (i.e., X(20), see Table 1 for a summary of symbols).

Training and Testing Datasets

The training dataset was generated by selecting drift and diffusion values from a grid, where the drift varied from -0.5 to +0.5 with spacing 0.1, and the diffusion varied from 0 to +1 with spacing 0.1. This resulted in 11 unique values for the drift parameter and 11 unique values for the diffusion parameter, yielding a total of 121 unique drift and diffusion combinations. For each of these drift diffusion combinations, a sample size of 2500 was drawn for reaction times and terminal boundaries of the process through simulations in python.

The testing dataset was generated similarly, where the drift varied from -0.5 to +0.5 with a spacing of 0.01, and the diffusion varied from 0 to +1 with a spacing of 0.01, yielding 10,000 unique drift diffusion pairs. For each pair, a sample of size 40 was drawn for reaction times and the corresponding termination boundaries, which falls in the smaller end of the range of the typical sample size for behavioral experiments in the literature (Mueller et al., 2017; Mulder et al., 2013; Ratcliff, Gomez & McKoon, 2004, Van Der Groen et al., 2019; Winkel et al., 2012).

Random Forest Regressor

The inputs to the random forest regressor were the first five moments of the reaction time data, as well as the proportion of times the decision corresponded to the upper boundary in the data (Fig. 2). The outputs were the drift and diffusion parameters. The random forest was trained in Python using the package *Sklearn* (Pedregosa et al., 2018), with a forest depth of 1,000,000.

We note that the decision to use the first five moments of the reaction time data was arbitrarily chosen with the intention of having a small feature set for input to the random forest regressor, while still capturing key components of the sampling distributions. We justify using a small number of moments by noting that the inverse gaussian distribution (which may be used to model first passage time distributions of drift diffusion models with a single boundary), only has two parameters, which could be solved for by knowledge of the first two moments of the distribution.

Evaluating Model Performance

The mean squared errors of the drift and diffusion coefficients (denoted MSE(μ) and MSE(σ) respectively) were used to evaluate model performance on the testing data (Fig. 3) for the trained random forest regressor.

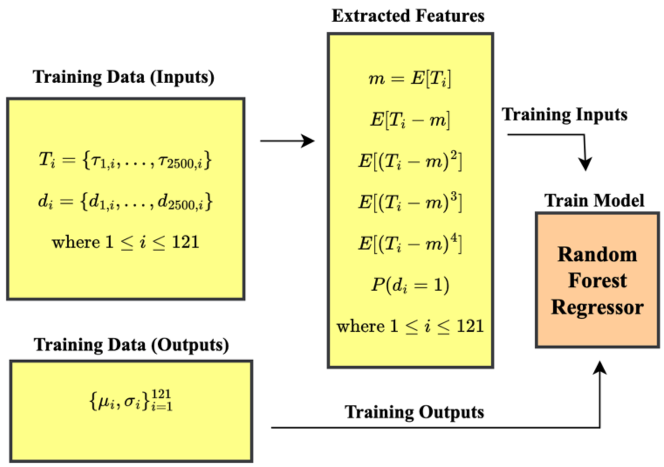

Figure 2. Training of the random forest classifier.

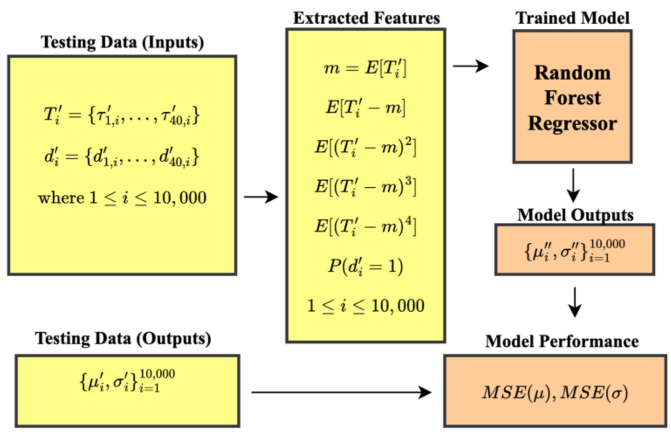

Fig. 3 Evaluating model performance of the random forest classifier.

Additionally, to compare performance of the random forest classifier to standard techniques, MSE(μ) and MSE(σ) were also obtained by using the following methods: maximum likelihood estimation and MSE minimization using the package *Pyddm* (Shinn et al., 2020), a popular python toolbox for drift-diffusion modeling, with the recommended optimizer (differential evolution). Finally, the backward Euler algorithm was also implemented in *Pyddm* to fit the drift-diffusion data, although this technique is typically not used in the literature, because it designed for fitting continuous signals.

Results and Discussion

The random forest regression resulted in a larger mean squared error for estimation of the drift parameter (table 2) when compared to maximum likelihood estimation and MSE minimization, whose errors are roughly 15% and 66% of the random forest regression error respectively.

Table 2: Performance on testing dataset.

Method	MSE(μ)	MSE(σ)
Random Forest Regression	0.00194	0.0611
Maximum Likelihood Estimation	0.000290	0.267
MSE Minimization	0.00128	0.273
Backward Euler	0.038	0.275

On the other hand, the random forest regression produced the smallest MSE for estimation of the diffusion coefficient. The MSE of the diffusion parameter for the random forest regression are roughly 23% the size of the other approaches.

Although the error of the drift parameter for the random forest regression is considerably larger than errors produced by the MSE minimization and maximum likelihood estimation, evaluation of the drift error is application dependent. For example, the error may be acceptable in specific aging effects, as a meta-analysis reports a wide range of effect sizes of the diffusion parameter (Theisen et al., 2021). In general, if the measurement error is small compared to the sampling variation of the drift and diffusion errors, the measurement error may be acceptable.

We also note that the MSE of the backward Euler method uncured the largest errors for the drift and diffusion parameters. This is a relatively expected result because numerical techniques for solving stochastic differential equations are not designed for fitting first passage time data.

We mention a few limitations. Firstly, our approach compares point estimates of the models, but does not compare error estimates of our model to other approaches. Such a process would be necessary to completely understand the relationships between the two approaches. Secondly, note that recovery of drift diffusion parameters is implemented through a fixed sample size of 40 reaction times and decisions, but it remains to compare performance of the approaches as the sample size increases. Thirdly, a systematic means of feature selection (for example, evaluating the number of moments in the input to the random forest regressor or evaluating the use of quantiles as features) is not performed in this paper. Fourthly, the offset and bias are fixed to zero in this study. Finally, we consider the case where the drift and diffusion are constants and non-variable.

Conclusion

The findings of this paper suggest that random forest regression may be a viable technique for fitting response time data to drift-diffusion models. We highlight that our model, trained only on 121 unique drift-diffusion pairs, produced more accurate results for estimation of the diffusion parameter than standard approaches that use some degree of analytical modeling of the first passage time distributions of the upper and lower boundary. Our results suggest that random forest regression may provide better estimations of the diffusion parameter when compared to other techniques in the literature. It is therefore of interest to investigate various machine learning techniques and larger training datasets for drift-diffusion modeling, as well as feature selection (such as the number of quantiles or moments to be inputted into the model).

References

Beste, C., Adelhöfer, N., Gohil, K., Passow, S., Roessner V., & Li, S. C. (2018). Dopamine Modulates the Efficiency of Sensory Evidence Accumulation During Perceptual Decision Making. The international journal of neuropsychopharmacology, 21(7), 649–655.

Dillon, D. G., Lazarov, A., Dolan, S., Bar-Haim, Y., Pizzagalli, D. A., & Schneier, F. R. (2022). Fast evidence accumulation in social anxiety disorder enhances decision making in a probabilistic reward task. Emotion (Washington, D.C.), 22(1), 1–18.

Gupta, A., Bansal, R., Alashwal, H., Kacar, A. S., Balci, F., & Moustafa, A. A. (2022). Neural Substrates of the Drift-Diffusion Model in Brain Disorders. Frontiers in computational neuroscience, 15, 678232.

Herz, D. M., Zavala, B. A., Bogacz, R., & Brown, P. (2016). Neural Correlates of Decision Thresholds in the Human Subthalamic Nucleus. Current biology: CB, 26(7), 916–920.

Mueller, C. J., White, C. N., & Kuchinke, L. (2017). Electrophysiological correlates of the drift diffusion model in visual word recognition. Human brain mapping, 38(11), 5616–5627.

Mulder, M. J., Keuken, M. C., van Maanen, L., Boekel, W., Forstmann, B. U., & Wagenmakers, E. J. (2013). The speed and accuracy of perceptual decisions in a random-tone pitch task. Attention, perception & psychophysics, 75(5), 1048–1058.

Pedregosa, F., Varoquaux, G. Gramfort, V., Thirion, Grisel, O. et al. (2018). “Scikit-learn: Machine Learning in Python” in arxiv, 2018.

Ratcliff, R. (1978). A theory of memory retrieval. Psychological Review, 85(2), 59–108.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. Psychological review, 111(1), 159–182.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. Psychonomic bulletin & review, 9(3), 438–481.

Roxin A. (2019). Drift-diffusion models for multiple-alternative forced-choice decision making. Journal of mathematical neuroscience, 9(1), 5.

Shinn, M., Lam, N. H., & Murray, J. D. (2020). A flexible framework for simulating and fitting generalized drift-diffusion models. eLife, 9, e56938. https://doi.org/10.7554/eLife.56938

Theisen, M., Lerche, V., von Krause, M., & Voss, A. (2021). Age differences in diffusion model parameters: a meta-analysis. Psychological research, 85(5), 2012–2021.

Turner, B. M., and Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. Psychonomic Bulletin and Review. 21, 227-250.

Turner, B. M., and Van Zandt, T. (2012). A tutorial on approximate Bayesian computation. Journal of Mathematical Psychology. 56, 69-85.

Turner, B. M. and Van Zandt, T. (2018). Approximating Bayesian inference through model simulation. Trends in Cognitive Science. 22, 826-840.

Van der Groen, O., Tang, M. F., Wenderoth, N., & Mattingley, J. B. (2018). Stochastic resonance enhances the rate of evidence accumulation during combined brain stimulation and perceptual decision-making. PLoS computational biology, 14(7), e1006301.

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.

Winkel, J., van Maanen, L., Ratcliff, R., van der Schaaf, M.E., van Schouwenburg, M. R., Cools, R., & Forstmann, B. U. (2012). Bromocriptine does not alter speed-accuracy tradeoff. Frontiers in neuroscience, 6, 126.