Estimation of time-varying decision thresholds
from the choice and reaction times
without assumptions on the shape

Yul HR Kang

hk2699@columbia.edu

Dept. of Neuroscience, Columbia University, New York, NY, USA.

Abstract

When a decision is made based on a series of samples of evidence, the threshold for decision may change over time. Here we propose a one-shot algorithm that gives the entire threshold from the choice and decision times. Combined with a standard gradient descent procedure, the algorithm can efficiently identify the correct shape of the threshold without any hyperparameter for the threshold.

Introduction

When a human or animal makes a binary choice based on a series of samples of scalar evidence, its speed and accuracy are typically modeled well by a diffusion process with two absorbing boundaries (Gold and Shadlen 2007; Bogacz et al. 2006). In these models, a decision is made when the cumulative sum of the evidence exceeds one of the two boundaries, or in a symmetric case, when the absolute magnitude of the cumulative sum exceeds a threshold. It has been shown that when the quality of evidence is fixed, i.e., when there is only one signal to noise ratio, the number of accurate decisions per unit time can be maximized with a threshold that is constant over time (Wald 1973; Wald and Wolfowitz 1948). But the optimal form of the threshold is different when signal to noise ratios vary between trials. There, trials with poor quality of evidence tend to take a long time to reach the threshold, and conversely, when a long time has passed and the threshold is not reached, it tends to be a trial with poor quality of evidence. Therefore, it is optimal to decrease the threshold over time, so as to move quickly onto the next decision (Drugowitsch et al. 2012; Shadlen et al. 2006).

Although it has been reported that the normative solution for the optimal threshold over time can be found using Bellman Equation (Drugowitsch et al. 2012), there has been no efficient method to estimate the shape of the threshold from the data. Many previous studies have imposed a particular functional form (linear, quadratic, exponential, Weibull, etc., see, e.g., (Hawkins et al. 2015)), and fitted its parameters, but they may have missed the form of the thresholds if it cannot be expressed by the functional form chosen. At least one study used cosine basis functions (Drugowitsch et al. 2012), which can fit thresholds of any form given a sufficient number of basis functions. However, fitting the model is inefficient because it involves fitting many parameters using a gradient descent procedure, and it requires several arbitrary hyperparameters, including the width of the cosine basis function.

Here we propose a one-shot algorithm that gives the entire threshold from the choice and decision times. Combined with a standard gradient descent procedure, the algorithm can efficiently identify the correct shape of the threshold without any hyperparameter for the threshold.

Methods

Symmetric thresholds with one drift rate

We first describe a simple case of symmetric thresholds and one drift rate. At each time step, we observe that the probability of absorption at each boundary is determined by the height of the threshold, and that it can be calculated simply by first computing the density after an unbounded diffusion then summing the density outside each proposed threshold. Then we observe that we can determine the threshold at all times thanks to the dispersion of the nondecision time, despite the fact that the data is a collection of discrete events (one time point per trial) that often populates the time bins only sparsely.

Previously, the likelihood of observing a choice $z \in \{1, 2\}$ at reaction time (RT) $h_{forward,z,t}$ has been computed by convolving the predicted decision time $r_{pred,z,t}$ with the proposed nondecision time distribution $g_{z,t}$, i.e.,

$$h_{forward,z,t} = \sum_{\tau=0}^{T-t} r_{pred,z,t-\tau} g_{z,\tau} \,,$$

where T is the maximum index of the time. Here we apply the opposite approach to determine the likelihood of the decision time given the observed RTs $h_{obs,z,t}$, by convolving the decision time with the nondecision time distribution flipped along the time axis. That is,

$$r_{back,z,t} = \sum_i P\{R_{z,i} = t \mid H_i = t + \tau, Z_i = z\} P\{g_z = \tau\} \;=\; \sum_{i \in \{H_i > t\}} g_{t-H,z}$$

Where i is the trial index, and H_i and Z_i are the observed RT and choice of the i-th trial. Then, for each time point, we first generate the prediction for the decision time for each threshold level and calculate the likelihood as a function of the threshold. To predict the decision time, we model the the accumulated evidence y at time t with a diffusion process

$$y(t) = \mu t + \sigma W_t$$
$$y(0) = 0$$

where μ is the drift rate, σ is the diffusion coefficient, and W_t is the Wiener process. The process has two symmetric absorbing boundaries $b_{z,t}$ with $-b_{1,t} = b_{2,t} > 0$. We fix σ to 1 without loss of generality, since for any values of μ, σ, and $b_{z,t}$, we can always use $\mu' = \mu/\sigma$, $b'_{z,t} = b_{z,t}/\sigma$, and $y' = y/\sigma$ instead of μ, b, and y and get the same results (Palmer, Huk, and Shadlen 2005). At the initial time $t = 0$, the accumulated evidence is defined to be zero, so the probability $u_{y,t}$ of having the accumulated evidence y is

$$u_{y,t=0} = \delta(y = 0) \,.$$

For all later times, we simulate the diffusion process in discretized time and evidence space (for more accurate methods, see, e.g., (Smith 2000)). We first compute the probability $v_{y,t}$ of having the accumulated evidence y assuming there is no absorbing boundary at time t as

$$v_{y,t} = u_{y,t-\Delta t} \varphi(y - \mu_c \Delta t, \sigma^2 \Delta t)$$

where φ is the Gaussian probability density function. This step can be done using the fast Fourier transform and the convolution theorem in time $O(Y \log Y)$ where Y is the number of bins on the

evidence axis. Then we compute the probability $u_{y,t}$ of having accumulated evidence y when there are two absorbing boundaries $b_{z,t}$:

$$u_{y,t} = v_{y,t} \cdot I(b_{z=1,t} < y < b_{z=2,t})$$

Importantly, we can calculate the likelihood of the threshold height efficiently in $O(Y)$, because the log likelihood of b can be obtained from the Dirichlet distribution as follows:

$$\beta_{b,t} := log\, P\{B_t = b \mid v_{.,t}\} \propto r_{z=-,t} log \sum_{y=-\infty}^{-b} u_{y,t} + r_{z=+,t} log \sum_{y=b}^{\infty} u_{y,t} + (\sum_{y=-b}^{b} u_{y,t})\, log \sum_{y=-b}^{b} u_{y,t}$$

We can then simply find the mode of $\beta_{b,t}$ over b to obtain the maximum likelihood estimate

$$\hat{b}_{2,t} = argmax_b\, \beta_{b,t}$$

and

$$\hat{b}_{1,t} = -\hat{b}_{2,t}$$

since we assumed that the thresholds are symmetric. Note that the mode can also be calculated faster without calculating β (hence without using the log function) by matching the proportions of r and u, i.e., by finding $b = \hat{b}_{2,t} = -\hat{b}_{1,t}$ such that

$$r_{z=1,t} + r_{z=2,t} = 1 - \frac{\sum_{y=-b}^{b} u_{y,t}}{\sum_{y=-\infty}^{\infty} u_{y,t}} \,.$$

However, to combine the results across multiple drift rates, we need the values of $\beta_{b,t}$, as explained in the following section.

Symmetric thresholds with multiple drift rates

When there are multiple drift rates, as is the case for an experiment with multiple conditions with different levels of difficulty, we need to combine the estimates of thresholds across conditions. Fortunately, it can be done simply by summing the log likelihood across conditions:

$$\beta_{b,t,\cdot} = \sum_c \beta_{b,t,c}$$

This is one reason why we need to compute the log likelihood $\beta_{b,t,c}$. Another reason is because it allows the application of priors, for example those relating to smoothness.

Asymmetric thresholds

The method can be easily generalized to two asymmetric thresholds. Here, we calculate the joint log likelihood

$$\beta_{b_1,b_2,t} := log\, P\{B^+_t = b^+, B^-_t = b^- \mid v_{.,t}\} \propto r_{z=-,t} log \sum_{y=b^+_t}^{\infty} u_{y,t} + r_{z=-,t} log \sum_{y=-\infty}^{b^-_t} u_{y,t} + (\sum_{y=b^-_t}^{b^+_t} u_{y,t}) log \sum_{y=b^-_t}^{b^+_t} u_{y,t}$$

then identify the maximum likelihood estimate

$$(\hat{b}_{1,t}, \hat{b}_{2,t}) = argmax_{(b_1,b_2)}\, \beta_{b_1,b_2,t}$$

The time complexity is still only $O(CTY^2 logY)$ where C is the number of unique drift rates. The memory complexity is $O(Y^2)$, since we need to store β for one time bin only.

Optimization of free parameters other than the thresholds

In the preceding sections, we described methods that estimate $\hat{b}_{z,t}$ for both choices and all times given the predicted decision time $r_{pred z,t}$ and the proposed nondecision time distribution $g_{z,t}$. Although those methods do not need any free parameter, we do need free parameters to compute the predicted decision and nondecision times, $r_{pred z,t}$ and $g_{z,t}$. The free parameters can be fit using the standard gradient ascent procedure.

In this paper, we use $\mu = \varkappa(c - c_0)$ where c is the signed stimulus strength, with $c > 0$ supporting the choice $z = 2$ and $c < 0$ supporting the choice $z = 1$. $\varkappa$ and c_0 are free parameters for sensitivity and bias. To allow a flexible shape of the nondecision time, we use the gamma distribution with two free parameters, mean μ_g and standard deviation σ_g. Then, given an initial guess of $b_{z,t}$ and $\theta = (\varkappa, c_0, \mu_g, \sigma_g)$, we compute v_t, update $b_{z,t}$ to maximize β_t, and using the updated $b_{z,t}$, find θ that maximizes the likelihood

$$L = \sum_t \beta_{\hat{b}_{1,t}, \hat{b}_{2,t}}$$

with gradient ascent.

Note that the methods of the preceding section do not require a particular functional forms of μ or g. We only need them for the gradient descent procedure.

Simulation and test

We simulated the data by sampling the choice and RT from $h_{pred z,t,c}$, the distribution predicted by the diffusion model with thresholds of various shapes. We modeled the diffusion process in terms of v, u, and r, then computed the predicted distribution of the reaction time r by convolving the decision time r with the nondecision time g, as described in the section "Symmetric thresholds with one drift rate."

To parametrically generate collapsing thresholds, we used the incomplete beta function

$$A_t = A_0(1 - I_{t'}(\beta_1, \beta_2))$$

where $t' = t/t_{max}$ is the relative time. When we fit the parameters, we transformed β_1 and β_2 into $t_\beta = \beta_1/(\beta_1 + \beta_2)$ and $B_{log} = log_{10}(\beta_1, \beta_2)$ because they were more interpretable. Here, t_β is approximately the relative time of the steepest collapse (i.e., $t_\beta \approx argmax_t |dA_t/dt|$), and B_{log} correlates with the slope of the collapse.

We used absolute drift rates (i.e., stimulus strengths) 0, 0.032, 0.064, 0.128, 0.256, and 0.512 for c, and generated the drift rates in a balanced way. For example, when we generate 1200 total trials, there are 200 trials for each absolute drift rate, divided evenly between positive and negative drift rates (for drift rate 0 the two are identical). Then we calculated $h_{pred z,t,c}$ and sampled the choice and the reaction time (H_i, Z_i) following

$$P\{(H_i, Z_i) = (h, z) \mid c\} = h_{pred z,t,c}.$$

For the free parameters, we used $\varkappa = 5$ or 10, $c_0 = 0$, $\mu_g = 0.3s$, $\sigma_g = 0.02s$, $B_{log} = 1$ or 0.5, $t_\beta = 0.2$ or $0.4s$, and $A_0 = 0.7$ or 1.

Results

When we estimated the symmetric thresholds from the simulated data, the method captured the shapes with as few as 600 total trials.
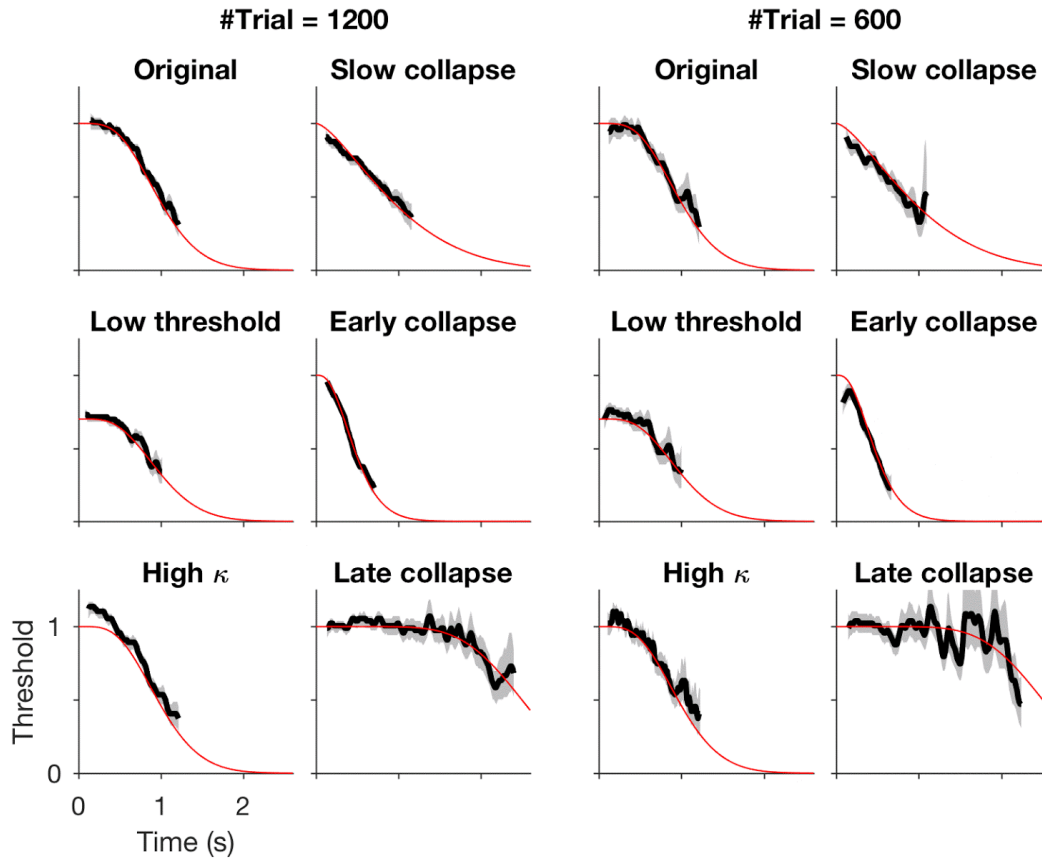

Figure 1. Comparison of the fitted and true thresholds for symmetric case.

Examples of the fitted and the true thresholds. The black solid line is the maximum likelihood estimate, the shade is the 95% confidence interval from the Dirichlet distribution, and the red line is the true threshold. Estimates are plotted between 5 and 95 percentile RTs.

In case of asymmetric thresholds, the fits captured the qualitative difference between the two thresholds. Not surprisingly, the fit was worse compared to the symmetric case with fewer trials, but the fit improved with more trials.
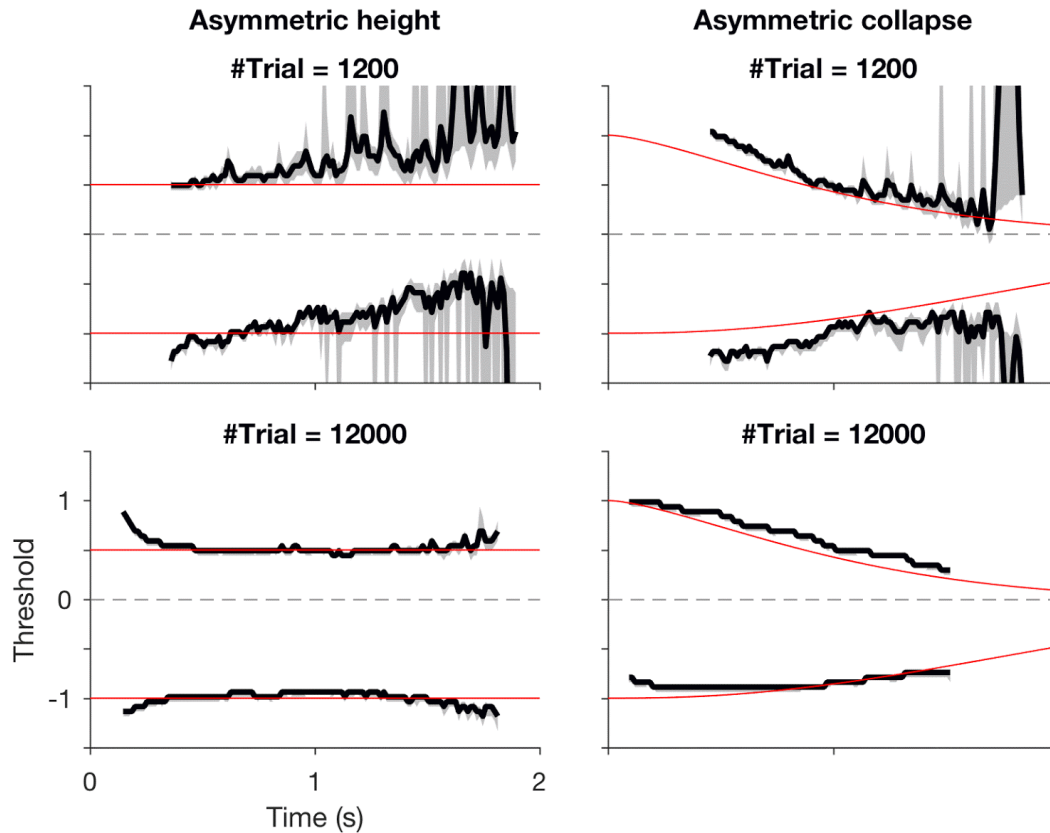

Figure 2. Comparison of the fitted and the true thresholds for asymmetric cases.
Examples of the fitted and the true thresholds. The black solid line is the maximum likelihood estimate, the shade is the 95% confidence interval, and the red line is the true threshold. Estimates are drawn between 5 and 95 percentile RTs.

Discussion

We showed that we can fit time-varying decision thresholds without prior assumption about their shape with a simple greedy algorithm. The method does not require any hyperparameter for the threshold, yet it is highly accurate and efficient, in that it captures the threshold shape with as few as 600 total trials.

There are a number of directions that the method can be extended. The simplest would be to introduce a smoothness parameter. The appropriate smoothness may be found using cross validation. Another direction is to use a full posterior for the density v by considering the dispersion of the estimate of the threshold at each time point. Yet another approach is to derive the gradient of the likelihood with regards to the thresholds and parameters. Such elaboration may resolve the discrepancy we see in our estimates, especially in the asymmetric case. The method we described here may serve as a simple yet useful starting point for more elaborate methods.

References

Bogacz, Rafal, Eric Brown, Jeff Moehlis, Philip Holmes, and Jonathan D. Cohen. 2006. "The Physics of Optimal Decision Making: A Formal Analysis of Models of Performance in Two-Alternative Forced-Choice Tasks." *Psychological Review* 113 (4): 700–765.

Drugowitsch, Jan, Rubén Moreno-Bote, Anne K. Churchland, Michael N. Shadlen, and Alexandre Pouget. 2012. "The Cost of Accumulating Evidence in Perceptual Decision Making." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 32 (11): 3612–28.

Gold, Joshua I., and Michael N. Shadlen. 2007. "The Neural Basis of Decision Making." *Annual Review of Neuroscience* 30: 535–74.

Hawkins, Guy E., Birte U. Forstmann, Eric-Jan Wagenmakers, Roger Ratcliff, and Scott D. Brown. 2015. "Revisiting the Evidence for Collapsing Boundaries and Urgency Signals in Perceptual Decision-Making." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 35 (6): 2476–84.

Palmer, John, Alexander C. Huk, and Michael N. Shadlen. 2005. "The Effect of Stimulus Strength on the Speed and Accuracy of a Perceptual Decision." *Journal of Vision* 5 (5): 376–404.

Shadlen, Michael N., Timothy D. Hanks, Anne K. Churchland, Roozbeh Kiani, and Tianming Yang. 2006. "The Speed and Accuracy of a Simple Perceptual Decision: A Mathematical Primer." In *Probabilistic Approaches to Neural Coding*, 208–37.

Smith, P. L. 2000. "Stochastic Dynamic Models of Response Time and Accuracy: A Foundational Primer." *Journal of Mathematical Psychology* 44 (3): 408–63.

Wald, A. 1973. *Sequential Analysis*. Dover Books on Advanced Mathematics. Dover Publications.

Wald, A., and J. Wolfowitz. 1948. "Optimum Character of the Sequential Probability Ratio Test." *Annals of Mathematical Statistics* 19 (3). Institute of Mathematical Statistics: 326–39.