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

Yul HR Kang
Dept. of Nuerobiology, Columbia University, 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 decision is made based on a series of samples of evidence, it can be often modeled by cumulative summation of a series of noisy numbers, where the summed evidence triggers choice when it is above some threshold. The number of accurate decisions per time can be often maximized when the quality of evidence is fixed, i.e., when there is only one signal to noise ratio. When there are multiple levels of signal to noise ratio, it is optimal to decrease the threshold over time, because when the threshold is not reached after a long time, it means that the quality of evidence is poor, so it is better to move 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 forms. At least one study used a cosine basis function (Drugowitsch et al. 2012), but fitting the model is inefficient because it involves fitting many parameters using a gradient descent procedure, and it still involves 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 where there are two conditions with one drift rate, such as a two-alternative forced choice paradigm with one level of difficulty. We observe that the bounded evidence accumulation process can be modeled by a discretized diffusion process. Then, 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. Finally, we observe that the dispersion of the nondecision time allows us to determine the threshold at all times despite the fact that the data is a collection of discrete events that gives only one time point per trial which often lines the time only sparsely.

What has been previously done for the forward modeling of the RTs is to convolve the predicted decision time *r* with the nondecision time distribution $g(t,\ z)$ to obtain a prediction of the RTs, $h_{pred}(t,\ z)$. Then the likelihood of the data, $h_{\mathrm{obs}}(t,\ z)$ can be calculated. Here we apply the opposite approach to determine the likelihood of the decision time given the observed RTs, using the nondecision time backwards. That is,

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

Where i is the trial index, H_i and Z_i are the observed RT and choice of the i-th trial. To allow a flexible shape of the nondecision time, we use the gamma probability distribution with two free parameters, mean μ_g and standard deviation σ_g. Then, for each time point, we generate the prediction for the decision time for each threshold level and calculate the likelihood as a function of the threshold. For predicting the decision time, we model the evidence accumulation process with one dimensional diffusion process with two absorbing boundaries. Given a diffusion process

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

Here, σ is fixed to 1 by convention because it is redundant with the combination of $\varkappa$ and the threshold height (Palmer, Huk, and Shadlen 2005). With two absorbing boundaries $b_{z,t}$ with $z \in \{1,2\}$ for the two choices, we discretize the time and evidence space to model it numerically. That is, at the initial time $t = 0$,

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

For all later times, we use a coarse yet simple approximation (for more accurate methods, see, e.g., (Smith 2000)):

$$v_t = u_{t-\Delta t} * \varphi(\mu_c, \sigma^2)$$
$$u_{y,t} = v_{y,t} \cdot I(b_{z=1,t} < y < b_{z=2,t})$$

Importantly, we can calculate the likelihood of the threshold height efficiently in $O(TY)$ where *Y* is the number of bins on the evidence axis, because the log likelihood of *b* is obtained from the Dirichlet distribution simply as:

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

Note that the mode of $\beta_{b,t}$ over *b* can be calculated simply by matching the proportions of r and *u*, i.e., by finding *b* such that

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

Which can be done faster in $O(TY)$. 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 absolute drift rates, as is the case for an experiment with multiple levels of difficulty, we need to combine the estimates of thresholds across conditions. Fortunately, it can be done simply by summing the likelihood across conditions:

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

This is one reason why we need to compute the likelihood $\beta_{b,t,c}$. Another reason is because it allows application of priors, e.g., about smoothness.

Note that we still need to fit the parameters governing the drift and the nondecision time. It can be done efficiently using the standard gradient descent algorithm: Given an initial guess of $b_{z,t}$ and $\theta = (\varkappa, c_0, \mu_g, \sigma_g)$, compute v_t, then update $b_{z,t}$ to maximize β_t. Then using the updated $b_{z,t}$, find θ that maximizes the likelihood with gradient ascent.

Asymmetric thresholds

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

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

Then identify the maximum likelihood estimate

$$(b^+{}_t, b^-{}_t) = argmax\ \beta_{b^+{}_t, b^-{}_t}$$

The time complexity is still only $O(TY^2)$. The memory complexity is $O(Y^2)$.

Simulation and test

We simulated the data by sampling the choice and RT from the distribution predicted by the diffusion model with thresholds of various shapes. The diffusion model we used has a linear transformation between the stimulus strength c and the drift rate, controlled by two free parameters $\varkappa$ for the slope and c_0 for the bias:

$$\mu = \varkappa(c - c_0)$$

With the drift rate, we modeled the diffusion process in terms of v, u, and r as mentioned in the section "Symmetric thresholds with one drift rate." Then we computed the predicted distribution of the reaction time r by convolving the decision time h with the nondecision time g .

$$r_{forward,z,t} = \sum_i P\{R_{z,i} = t \mid H_i = t - \tau, Z_i = z\} P\{g_z = \tau\} = \sum_{i \in \{H_i < t\}} g(t - H_i, z)$$

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

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

and parameterized it with $B_{log}=\log_{10}(\beta_1\beta_2)$ which correlates with the slope of the collapse and $t_\beta=\beta_1/(\beta_1+\beta_2)$ for convenience.

We used stimulus strengths 0, 0.032, 0.064, 0.128, 0.256, and 0.512 for c, and generated the trials in a balanced way. For example, when we generate 1200 total trials, there are 200 trials for each stimulus strength, divided evenly between the two choice options (for strength 0 it is determined randomly). For the free parameters, we used $\varkappa$ = 5 or 10, $c_0 = 0$, $\mu_g = 0.3s$, $\sigma_g = 0.05s$, B_{log}=1 or 0.5, t_β=0.2 or 0.4s, and A_0=0.7 or 1.

Results

We show with simulated data that the method captures arbitrary shapes of the thresholds, with as few as 600 total trials.

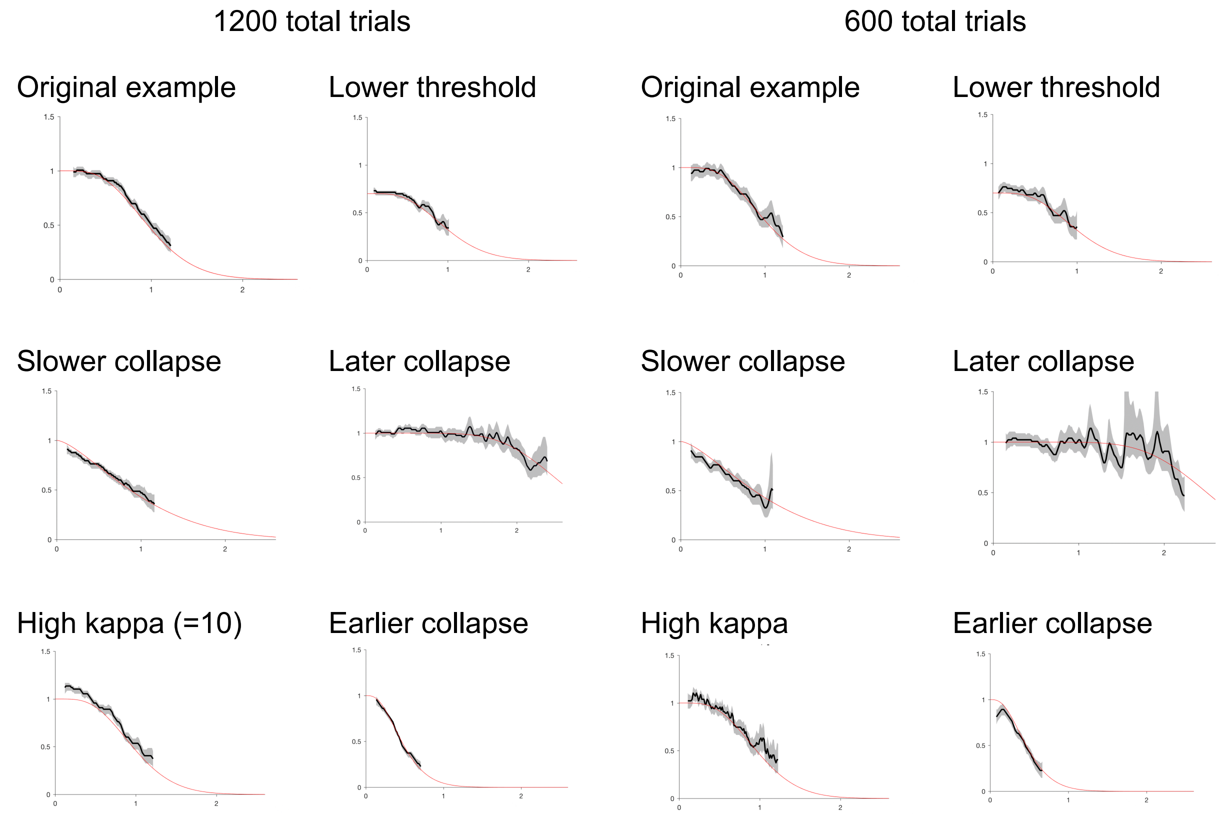

Figure 1. Comparison of the fitted and the true thresholds for symmetric case.
Examples of the fitted and the true thresholds (monotonically decreasing). Black solid line is the maximum likelihood estimate, the shade is the 95% confidence interval from the Dirichlet posterior, and the red line is the true threshold. Estimates are plotted between 5 and 95 percentile RTs.

In case of asymmetric thresholds, the fits capture the qualitative difference between the two thresholds. Not surprisingly, the fit was worse compared to the symmetric case when the data was small, but it was overcome with bigger data.

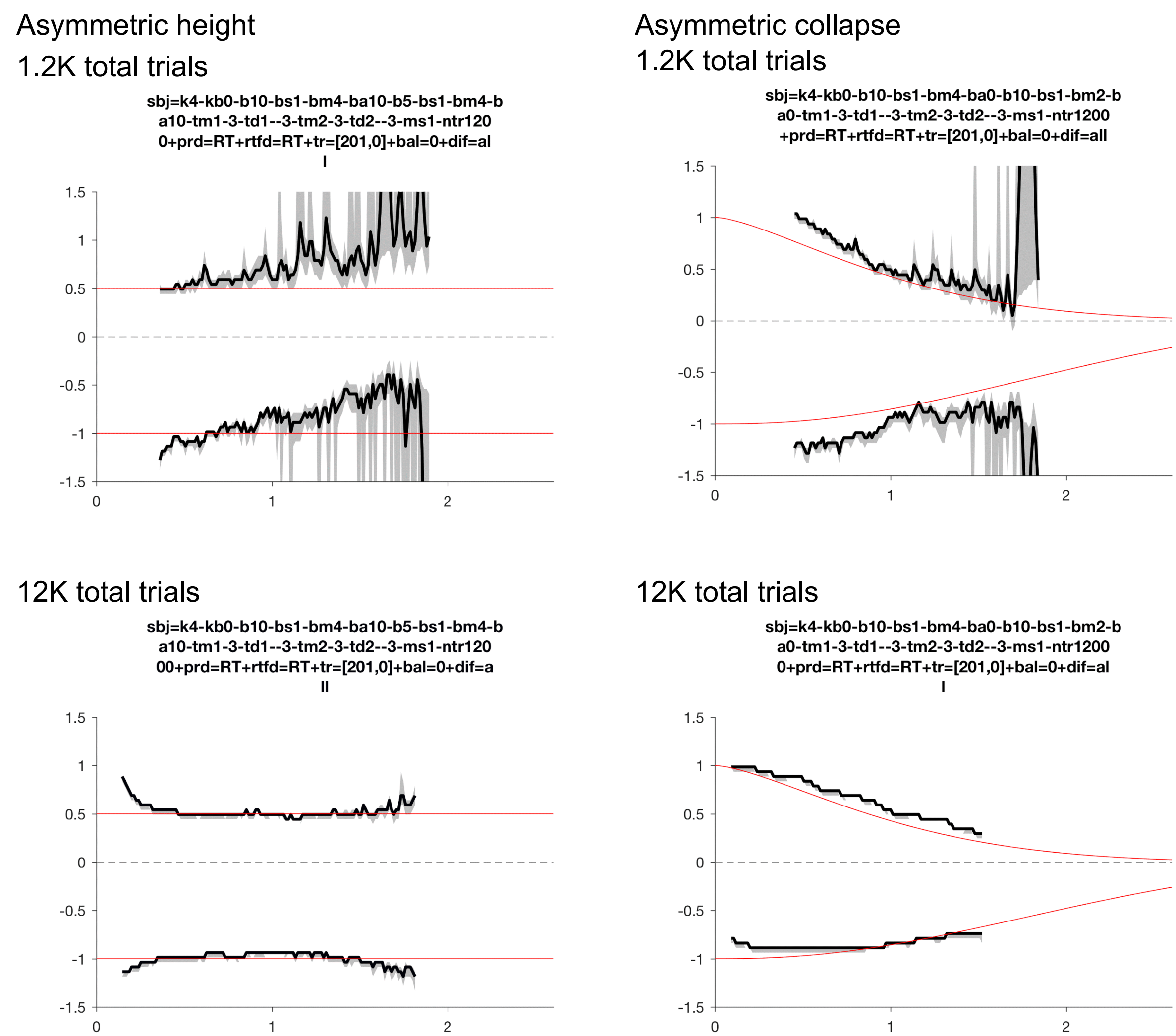

Figure 2. Comparison of the fitted and the true thresholds for asymmetric cases. Examples of the fitted and the true thresholds (flat or monotonically decreasing). Black solid line is the maximum likelihood estimate, the shade is the 95% confidence interval, and the red line is the true threshold. Estimates are drawn between 5 and 95 percentile RTs.

Discussion

We showed that we can fit time-varying decision thresholds without prior assumption about its shape with a simple greedy algorithm. The method does not require any parameter regarding 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. The last approach may resolve the discrepancy we see in our one shot estimate, especially in the asymmetric case. The method we described here may serve as a simple yet useful starting point for the more elaborate methods.

References

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.

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.