

Evaluating methods for approximating stochastic differential equations

Scott D. Brown^{a,*}, Roger Ratcliff^b, Philip L. Smith^c

^a*Department of Cognitive Science, University of California Irvine, CA 92697-5100, USA*

^b*Ohio state University, USA*

^c*University of Melbourne, Australia*

Received 9 June 2005; received in revised form 26 February 2006

Available online 6 May 2006

Abstract

Models of decision making and response time (RT) are often formulated using stochastic differential equations (SDEs). Researchers often investigate these models using a simple Monte Carlo method based on Euler's method for solving ordinary differential equations. The accuracy of Euler's method is investigated and compared to the performance of more complex simulation methods. The more complex methods for solving SDEs yielded no improvement in accuracy over the Euler method. However, the matrix method proposed by Diederich and Busemeyer (2003) yielded significant improvements. The accuracy of all methods depended critically on the size of the approximating time step. The large (~10 ms) step sizes often used by psychological researchers resulted in large and systematic errors in evaluating RT distributions.

© 2006 Elsevier Inc. All rights reserved.

Over the past 40 years, models of response time (RT) for simple decision making have become very successful at capturing the details of observed data (Audley & Pike, 1965; Brown & Heathcote, 2005; Busemeyer & Townsend, 1993; Diederich, 1997; Heath, 1981; LaBerge, 1962; Lacouture and Marley, 1991; Laming, 1966; Link & Heath, 1975; Ratcliff, 1978; Ratcliff & Smith, 2004, Appendix; Ratcliff, Van Zandt, & McKoon, 1999; Smith, 1995; Vickers, 1970; Vickers & Lee, 2000). More recently, the same models have also become quite successful at explaining decision making at a neural level (Carpenter & Reddi, 2001; Cook & Maunsell, 2002; Glimcher 2003; Gold & Shadlen, 2001; Ratcliff, Cherian & Seagraves, 2003; Reddi & Carpenter, 2000; Roitman & Shadlen, 2002; Sato, Murthy, Thompson & Schall, 2001; Sato & Schall, 2003; Shadlen, Britten, Newsome & Movshon, 1996; Wang, 2002). The most successful models of decision making in both cognitive and neural domains are the sequential sampling models. These models are based on the idea that noisy stimulus information is accumulated progressively over time until sufficient information for one of the

response alternatives has been obtained. The predicted decision time in such models is obtained mathematically by solving a first-passage-time (FPT) problem, that is, the time taken for the accumulated information to reach a criterion and trigger a response. For some models, there exist explicit analytic methods or highly accurate numerical methods for solving the FPT problem (see Ratcliff & Smith, Appendix, for a survey). For other models, this problem may be complex or intractable. In such situations, researchers must resort to Monte Carlo simulation techniques to obtain predicted RT distributions and choice probabilities. We investigate the properties of such simulation techniques in this article.

One of the best-known sequential sampling models, and one that has been applied to a wide range of experimental data, is the diffusion model of Ratcliff (1978; Ratcliff & Rouder, 1998). This model assumes that evidence accumulation begins at some initial value (z) and then moves towards one of two absorbing boundaries located at a and zero. The boundaries represent the decision criteria for the two responses. The time taken to reach a boundary determines RT, except for the addition of a nondecision time, T_{ER} , which is taken as uniformly distributed for computational simplicity. The evidence accumulation process is

*Corresponding author.

E-mail address: scottb@uci.edu (S.D. Brown).

modeled mathematically as a Wiener diffusion process with constant drift¹ (I). This means that in any time interval Δt , the evidence will change from its starting value by an amount $I\Delta t + \sigma\sqrt{\Delta t}\eta$, where η is a sample from a standard normal distribution, and σ is the standard deviation of the Wiener process.

Using ideas from signal detection theory, Ratcliff (1978) improved sequential sampling models by adding other sources of variability. The drift rate, I , was assumed to vary according to a normal distribution across repeated presentations of the same decision task. In more recent incarnations of the diffusion model (e.g., Ratcliff & Rouder, 1998), the starting point, z , was also assumed to vary according to a uniform distribution across repeated decisions. Even with these extra assumptions, the FPT problem for the diffusion model can be solved analytically. That is, the probability of making each response (i.e., terminating at the boundary a or zero) and the RT distribution associated with each response, can be determined using results from the calculus of stochastic differential equations. Even these “analytic” solutions are not simple closed-form expressions giving density functions in terms of parameters. Instead, they involve approximations to infinite sums, or other such implicit quantities, and they require numerical integration over parameters that vary between decision trials (a computational tutorial for these calculations can be found in Tuerlinckx, Maris, Ratcliff, & De Boeck, 2001). The complexity of these calculations, and the need for numerical approximations, makes implementing analytic solutions a decidedly non-trivial and occasionally error-prone endeavor.

Apart from their complexity, analytic solutions are simply not available for many of the models psychologists use. We investigate just such a model below: A simplified version of the leaky competing accumulator model of Usher and McClelland (2001). In our model, a decision between two alternatives is made by allowing two evidence accumulators to “race” towards a response threshold (C). Denoting the accumulators’ levels of activation at time t by $x_1(t)$ and $x_2(t)$, we can write the equation governing the change of activation in each accumulator (Δx_i for $i = 1, 2$) over a small time interval (Δt) as:

$$\Delta x_i(t) = (I_i - kx_i(t))\Delta t + \sigma\sqrt{\Delta t}\eta.$$

The subscripted input strength (I_i) allows each accumulator to race at a different average rate, so that the accumulator corresponding to the “correct” response most often wins the race to the threshold ($I_1 > I_2$). Throughout, we assume that both accumulators begin at time $t = 0$ with zero activation (i.e., $x_1(t) = x_2(t) = 0$), although our results do not depend on this assumption.

This decision model can be mathematically described as a pair of racing Ornstein–Uhlenbeck processes, each with a

single absorbing boundary. The model is similar to Usher and McClelland’s (2001) leaky competing accumulator model, but does not include lateral inhibition between accumulators: Usher and McClelland assumed that increased activation in one accumulator suppressed activation in the other accumulator. The Ornstein–Uhlenbeck process was first investigated as a psychological model by Busemeyer and Townsend (1993), and the racing accumulator model we investigate was investigated systematically by Smith (200) and by Ratcliff and Smith (2004). Ratcliff and Smith called this model the *leaky accumulator* to emphasize its relationship to Usher and McClelland’s *leaky competitive accumulator* model.

Numerical, integral-equation solutions to the FPT problem for the leaky accumulator are available (Smith, 2000). Although these methods can be used with quite general models, they have not yet been extended to all situations of interest to researchers. For example, Usher and McClelland (2001) assumed that the totals in the two accumulators could not become negative. This assumption was required in the model to prevent mutual inhibition between the accumulators from becoming mutual excitation. Ratcliff and Smith (2004) retained this assumption in the leaky accumulator model because of its biological plausibility, although the structure of the model does not require it because it has no mutual inhibition. Bounded accumulation processes of this form are modeled mathematically using a reflecting boundary at zero (Cox & Miller, 1965). To date, the integral-equation method has not been extended to processes with reflecting barriers, except for the simplest case. Other models, such as those in which the accumulation rate is a nonlinear function of the process’ current value, are also difficult to treat analytically. Each of these model variants is, however, easily accommodated in the matrix algebra methods of Diederich and Busemeyer (2003, discussed below).

Fortunately, there exist several methods for obtaining approximations to the RT distributions and response probabilities when analytic solutions are not available or convenient. These methods are easy to implement, even for the most complicated and intractable sequential sampling models, and the accuracy of their approximations can be made arbitrarily good, at least in theory. One disadvantage of numerical approximation methods is that they can require very long computing times, although this is less problematic with the advent of fast, cheap computers. A more serious disadvantage is that one cannot always be certain of the accuracy of the approximations. For example, in an unpublished doctoral thesis, Brown (2002) observed a difference in mean RT of over 40 ms between Usher and McClelland’s (2001) results and a more accurate implementation of exactly the same model.

Sequential sampling models can be expressed as systems of differential equations, in matrix form:

$$dX(t) = f(X)dt + \sigma dW(t). \quad (1)$$

¹We use the symbol I for drift rate instead of the traditional v . This choice ensures consistency with the corresponding quantities in other models, in particular the accumulator models upon which we focus below.

Here, $X(t)$ is a vector of activation values, with one element corresponding to activation in each accumulator. In the earlier example, $X(t)$ would be a vector of length two, namely $\{x_1(t), x_2(t)\}$. The notation $dX(t)$ represents a very small (infinitesimal) change in the value $X(t)$ during a small time period. The function f is vector-valued and specifies the average drift rates. For example, in the leaky accumulator model, each component of f is given by $f_i(t) = I_i - kx_i(t)$. In the psychological models we discuss, f always depends only on X and is independent of t , however this is not an essential constraint (see, e.g., Smith, 1995, 2000, for models with time-dependent drift functions). The term dW represents a Wiener process (see, e.g., Gard, 1988). This is a continuous random vector-valued process, such that $dW(t + \Delta t) - dW(t)$ has a multivariate normal distribution with zero mean, variance given by Δt , and zero covariance values. The coefficient of the Wiener process, σ , is always a simple scalar value in the models considered by psychologists; in general, σ can be any matrix-valued function of X and t .

There are two primary methods for obtaining approximations to arbitrary sequential sampling models². One is the matrix algebra method of Diederich and Busemeyer (2003; see also Busemeyer & Townsend, 1993), referred to hereafter as the BDT matrix method. The other solution method is actually a large class of differential equation “tracking” algorithms combined with Monte Carlo integration over sources of variability. Below, we compare the BDT matrix approximation method with three different tracking methods. Tracking algorithms have become very popular most likely because they are very simple to implement. Each decision process of a sequential sampling model can be simulated by “tracking” the system of differential equations, using computer-generated random numbers to simulate the noisy accumulation (Wiener) process. If this simulation is repeated many times, the resulting set of simulated response times and outcomes can be used to provide estimates of the RT distributions: A simple application of Monte Carlo integration. The optimal method for tracking the differential equations, however, is not always clear. This is because most differential equation tracking methods have not been developed specifically for stochastic differential equations, and their properties when applied to SDEs are not well understood. In what follows, we compare the performance of different methods for tracking SDEs. To foreshadow, we find that the simplest method is as good as the most complicated, but that the typical implementation choices made by psychological researchers can lead to large inaccuracy. We also find that the BDT matrix approximation method is faster and more accurate than any of the tracking methods, at the cost of greater implementation difficulty. Since the implementation of the matrix method

has been outlined in detail elsewhere, we refer the interested reader to Diederich and Busemeyer (2003).

1. Tracking differential equations

Consider the leaky accumulator model presented above. The RT distributions and response accuracy can be estimated by Monte Carlo integration of repeated simulated decisions. Simulating a single decision could proceed as follows:

1. Decide on values for the parameters I_1, I_2, k, σ and C .
2. Initialize time to $t = 0$ and each accumulators' activation to $x_1 = x_2 = 0$.
3. Choose a time “step size”, Δt .
4. Set $t = t + \Delta t$.
5. Sample two random numbers, η_1 and η_2 , from a standard normal distribution.
6. For $i = 1, 2$, set: $x_i := x_i + (I_i - kx_i)\Delta t + \sigma\sqrt{\Delta t}\eta_i$.
7. If $x_i > C$, choose response i and exit, with simulated RT = t .
8. Go to Step 4.

This method of tracking stochastic differential equations is called Euler's method. It is the simplest method available and has been almost the only method employed in psychological research. However, researchers in applied mathematics have found Euler's method to be inefficient and error prone (e.g., Kloeden & Platen, 1992). Errors are introduced due to the finite number of steps taken in approximating the Wiener process by a sequence of random samples. Larger values of Δt lead to fewer samples, and hence faster, but less accurate approximations.

Burrage and Burrage (1996; see also Burrage, Burrage, & Tian, 2004) examined this problem, and introduced a series of more complicated methods, by analogy with traditional higher-order Runge–Kutta methods for tracking deterministic differential equations. They examined the expected “best case” error level for each of these methods and identified two methods that provided considerable improvements in accuracy and efficiency over Euler's method. Below, we implement Euler's method and Burrage and Burrage's two more complicated methods, and examine their performance in evaluating the leaky accumulator model.

The methods described by Burrage and Burrage (1996) are suitable for very general stochastic differential equations, but the equations used in psychological models are almost always taken from a simplified subset. In Eq. (1), f is independent of t , and σ is a constant scalar. These constraints greatly simplify Burrage and Burrage's methods: The three methods we examine here are:

1. Euler: This is the method presented above in an example for the leaky accumulator model. For the system of differential equations given in (1), the algorithm for finding $X(t + \Delta t)$ is:

$$X(t + \Delta t) = X(t) + \Delta t \cdot f(X(t)) + \sigma\sqrt{\Delta t} \cdot \eta, \quad (2)$$

²Smith (1995, 2000) and Heath (1992) describe a third method based on integral equations. This method is quite general, but has not yet been applied to accumulator models with reflecting boundaries.

where η is a vector of independent samples from a standard normal distribution.

2. *Explicit two-step (E2)*: Let $Y_1 = f(X(t))$ and $Y_2 = f(X(t) + \frac{2}{3}\Delta t \cdot Y_1 + \frac{2}{3}\sigma\sqrt{\Delta t} \cdot \eta)$. Again, η is a vector of independent samples from a standard normal distribution. Then the updated accumulator activations are given by: $X(t + \Delta t) = X(t) + h/4(Y_1 + 3 \cdot Y_2) + \sigma\sqrt{\Delta t} \cdot \eta$. Note that the vector of samples (η) is used twice in this method.

3. *Explicit four-step (E4)*: Let

$$Y_1 = f(X(t)),$$

$$Y_2 = f(X(t) + \frac{2}{3}\Delta t \cdot Y_1 + \frac{2}{3}\sigma \cdot J_1),$$

$$Y_3 = f(X(t) + \Delta t(\frac{2}{3}Y_1 - \frac{1}{3}Y_2) + \frac{2}{3}\sigma \cdot J_1 - \frac{2}{3}\sigma \cdot J_2),$$

$$Y_4 = f(X(t) + \frac{7}{6}\Delta t \cdot Y_1 + \frac{2}{3}\sigma \cdot J_2).$$

Values for J_1 and J_2 are calculated on each cycle, as follows. Let u and v be vectors of independent standard normal deviates, and then $J_1 = \sqrt{\Delta t} \cdot u$ and $J_2 = \frac{1}{2}\sqrt{\Delta t}(u + \frac{1}{\sqrt{3}}v)$. Then the updated accumulator activations are given by $X(t + \Delta t) = X(t) + \frac{h}{4}(Y_1 + 3 \cdot Y_2 - 3 \cdot Y_3 + 3 \cdot Y_4) + \sigma J_1$.

2. Evaluation methods

To match the kind of parameter settings and tasks faced during real RT modeling, we used a fixed set of parameters for the model, along with factorial combinations of four different input strengths (drift rates, I) and two different response criteria (C). This simulated four experimental conditions of different difficulty, each of which is given in both speed- and accuracy-emphasis conditions (see, e.g., Ratcliff & Rouder, 1998). The particular model we used was $dX(t) = (I - kX(t))dt + \sigma dW(t)$. The accumulator activations, X were initialized at zero, and the time counter, t was initialized at a random sample from a uniform distribution on [0.204 s, 0.281 s]. The leakage parameter was $k = 2.6$ and the noise parameter was $\sigma = 0.241$. The response threshold used to model the speed-emphasis conditions was $C = 0.169$, for accuracy conditions it was $C = 0.240$. Four values were used for the input strength to the first response accumulator: $I_1 = 0.552, 0.675, 0.764$ and 0.987 . Input strength to the second accumulator (I_2) was set at $1 - I_1$. We forced all activation values to be bounded at zero: If on any time step the first accumulator was negative ($x_1 < 0$) we set it zero ($x_1 = 0$), and similarly for the second accumulator (x_2). This approximated a reflecting barrier at $x = 0$.

Each of the three tracking methods outlined above was used with six different step sizes: $\Delta t = 50, 20, 10, 1, 0.1$ and 0.01 ms. For each step size and for each set of parameter values, 20,000 decisions were simulated. To allow for the uncertainty in threshold crossing time introduced by finite step-sizes, $\Delta t/2$ was subtracted from each simulated response time. Response accuracy was calculated as the

proportion of times x_1 reached threshold before x_2 . We also calculated sample estimates of the 10%, 30%, 50%, 70% and 90% quantiles for distributions of correct and error RTs.

The BDT matrix approximation method also requires a step size parameter, analogous to Δt in the tracking methods. The matrix approximation method works by breaking the continuous state and time spaces into a matrix of discrete values. The step size parameter determines the fineness of this discrete approximation. We used two different values for the time step size parameter for the matrix method: $\Delta t = 1$ and 0.1 ms (the state-space step size was set at $\sqrt{\Delta t}$). We used fewer values of Δt for the BDT matrix method (two) than the tracking methods (six) because initial simulations showed that these values of Δt allowed the matrix method to be both faster and more accurate than the tracking methods (hence, there was no use investigating less accurate values of Δt).

3. Test results

3.1. Implementation and computation times

The computation time for each of the methods we investigated is inversely proportional to the step size parameters, at least approximately. Using a standard desktop computer (32 bit CPU, about 2 GHz clock speed, 1 GB memory), the matrix approximation method required about 2 s to evaluate distributions associated with a single set of parameters when $\Delta t = 0.1$ ms, and about 0.2 s when $\Delta t = 1$ ms. Euler's method, using 20,000 Monte Carlo repetitions, required about 20 s when $\Delta t = 0.1$ ms and about 2 s when $\Delta t = 1$ ms. The E2 and E4 methods required about 50% more time than the Euler method.

For the particular leaky accumulator model we have investigated, the matrix method was much more computationally efficient: Producing results of a fixed accuracy takes about 20 times longer with the Euler method than the matrix method. However, this advantage may be balanced by implementation costs, depending on the particular user's costs for computer time and programming time. Implementing the tracking methods is very simple, requiring minimal coding time even for the most complex models and least competent programmer. Implementing the BDT matrix method may be more difficult and time consuming for less proficient programmers, although Diederich and Busemeyer (2003) provide Matlab code to facilitate implementation of the BDT matrix method for some models.

The situation is quite different for more complex models, such as those that include between-trial variability in several parameters (e.g., Ratcliff's, 1978, diffusion model) or those that include lateral inhibition terms (e.g., Usher & McClelland, 2001). For those models, implementing the BDT matrix method can become quite complex. The numerical efficiency of the BDT method may also be reduced if memory storage considerations limit the size of the matrices considered (e.g., adding lateral inhibition

would increase memory usage by squaring, which can be very costly). It is possible to greatly reduce memory usage, obviating these problems, by employing sparse matrix methods. These methods make efficiency gains by considering only nonzero elements of large matrices, and the gains can be quite considerable if a great many zero elements are present. Once again, however, the gains in efficiency come at the cost of increased implementation complexity—sparse matrix operations are always more complex to consider than standard matrix operations. For the tracking methods, models that are more complex are no more difficult to analyze than simple models. More complex models may increase computation time, as extra parameter variability can require more Monte Carlo iterations for a fixed level of accuracy.

In summary, for simple models such as the one investigated here, the BDT method is much faster to compute, and only a little more difficult to implement than the tracking methods. For more complex models the implementation difficulty of the BDT method will increase, and its computational advantage may reduce.

3.2. Response accuracy

Response accuracy did not change dramatically for any integration method or time step size: The largest difference occurred for $I_1 = 0.764$ under a speed-emphasis threshold ($C = 0.169$). In that condition, the “true” accuracy rate, given by the very small step-size implementations of each method was 75.1%. The largest step size gave an accuracy estimate of 76.9% (using the E2 method), a difference of only 1.8%. All other accuracy differences were far smaller than this.

3.3. RT distributions

RT distributions, on the other hand, varied systematically with step size. In general, as expected, smaller step sizes lead to greater accuracy. There was essentially no difference between the results from the tracking methods using step sizes 0.01 and 0.1 ms and from the BDT matrix method with step size 0.1 ms or 1 ms. From here on, we use one of these distributions to be the “true” values against which others are plotted (arbitrarily, we chose the distributions generated by the Euler method with step size 0.01 ms). Fig. 1 demonstrates that the BDT matrix method produced RT distributions that were almost identical to those from the Euler method with the smallest step size. Fig. 1 plots 10%, 30%, 50%, 70% and 90% quantiles calculated from the small-step Euler distributions against those from the matrix method distributions (separately for matrix method step sizes of 0.1 and 1 ms). If the methods produced exactly the same RT distributions, all data would lie on the $y = x$ line (dotted in Fig. 1). The left column of plots show data from the speed-emphasis conditions, the right column from the accuracy-emphasis conditions. The four rows of plots correspond to the four input strength

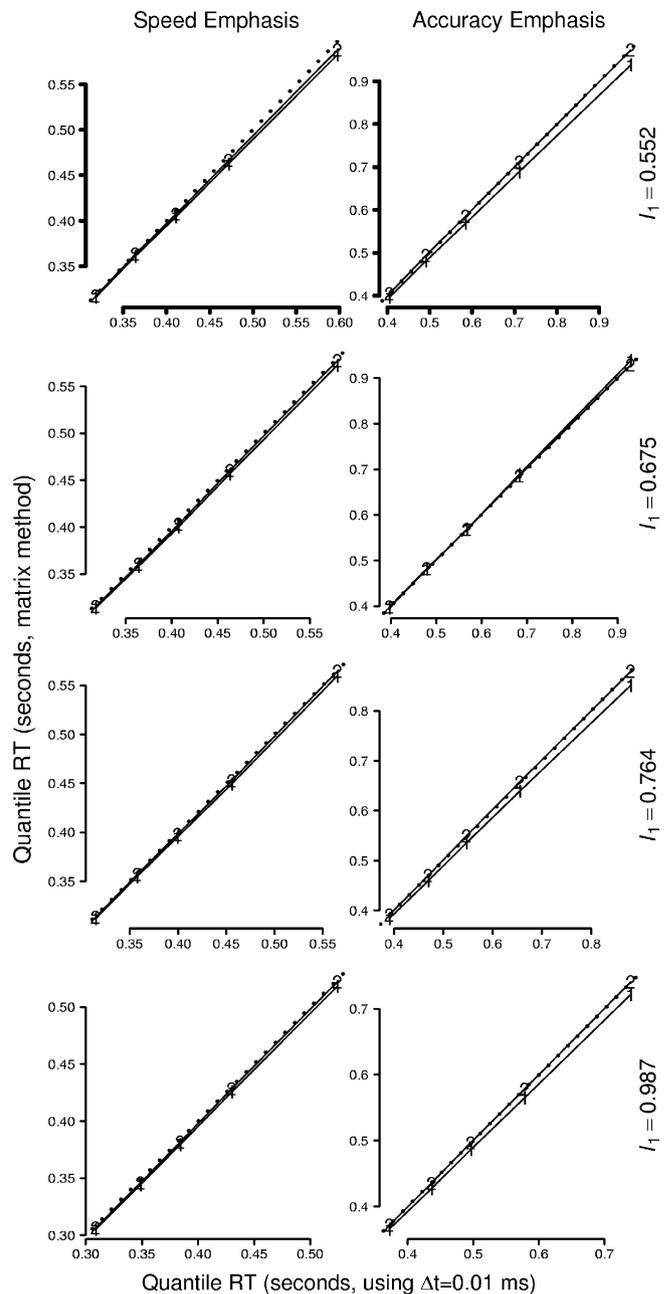


Fig. 1. Quantiles from the BDT matrix method distributions (y-axis) plotted against quantiles from the Euler method with step size $\Delta t = 0.01$ ms (x-axis). The two different BDT matrix method step sizes are represented by lines #1 ($\Delta t = 0.1$ ms) and #2 ($\Delta t = 1$ ms). Dotted lines show $y = x$. All units in seconds.

drift rates. There are some very small deviations for the 1 ms step size (line #2 in each plot), but none for the 0.1 ms step size. We conclude that the matrix approximation method produces RT distributions that are indistinguishable from the best Euler method distributions when using a step size of 0.1 ms, and are nearly as good with a step size of 1 ms. Note that errors enter into BDT’s matrix method because of a finite approximation to the response criterion. The state space is discretized: the only allowed values for x are $\Delta x, 2\Delta x, 3\Delta x, \dots$. This means that the actual response

criteria ($C = 0.169$ and 0.240 , for speed and accuracy conditions) must be approximated by the nearest integral multiple of Δx . In the $\Delta t = 1$ ms the actual values we used were $C = 0.1677$ and 0.2363 . For $\Delta t = 0.1$ ms the actual values were $C = 0.1687$ and 0.2410 .

The results for different step sizes using the Euler method are shown in Fig. 2, for the correct RT distributions (error RT distributions were similar, and are discussed later).

Each plot contains five lines: the sample quantile estimates for the five larger step sizes ($\Delta t = 0.1, 1, 10, 20$ and 50 ms) plotted against the “true” values obtained with

$\Delta t = 0.01$ ms. Deviations from the $y = x$ line (dotted) indicate prediction differences.

From Fig. 2, it is clear that $\Delta t = 0.1$ ms produced almost identical results to $\Delta t = 0.01$ ms, as the #1 lines are almost always coincident with the $y = x$ line. The 1–50 ms step sizes (#2–#5 lines) produced longer RT distributions. The quantile–quantile plots are very close to linear, suggesting that larger step sizes simply “scaled up” the RT distributions by some factor. Simple linear regression for the lines shown in Fig. 2, with intercepts fixed at zero, showed that the amount of RT inflation compared to the $\Delta t = 0.01$ ms standard was less than 1% for $\Delta t = 0.1$ ms, around 3% for $\Delta t = 1$ ms, 8% for $\Delta t = 10$ ms, 11% for $\Delta t = 20$ ms, and 16% for $\Delta t = 50$ ms. Comparison of Figs. 1 and 2 illustrates that the tracking methods required step sizes (Δt) around 10–100 times larger than the BDT matrix method to achieve a comparable accuracy level. This is the chief benefit of the BDT method—its results are both faster to calculate and much more accurate for any fixed step size.

Fig. 3 shows the inflation percentages for correct RT distributions (x-axis) plotted against those for the corresponding error RT distributions for the tracking methods. The points are clustered tightly around the $y = x$ line, indicating that the inflation for any single combination of parameters was almost equal for the correct and error RT distributions that were generated. Also, points corresponding to different parameter settings within each step size (represented by different plot symbols in Fig. 3) are also closely clustered together. These two results are important, as they mean that researchers can use a large, efficient step size for model evaluations and be confident that all their predicted RT distributions have been inflated by the same constant amount. Thus, when a more accurate approximation is required, a simple scaling calculation

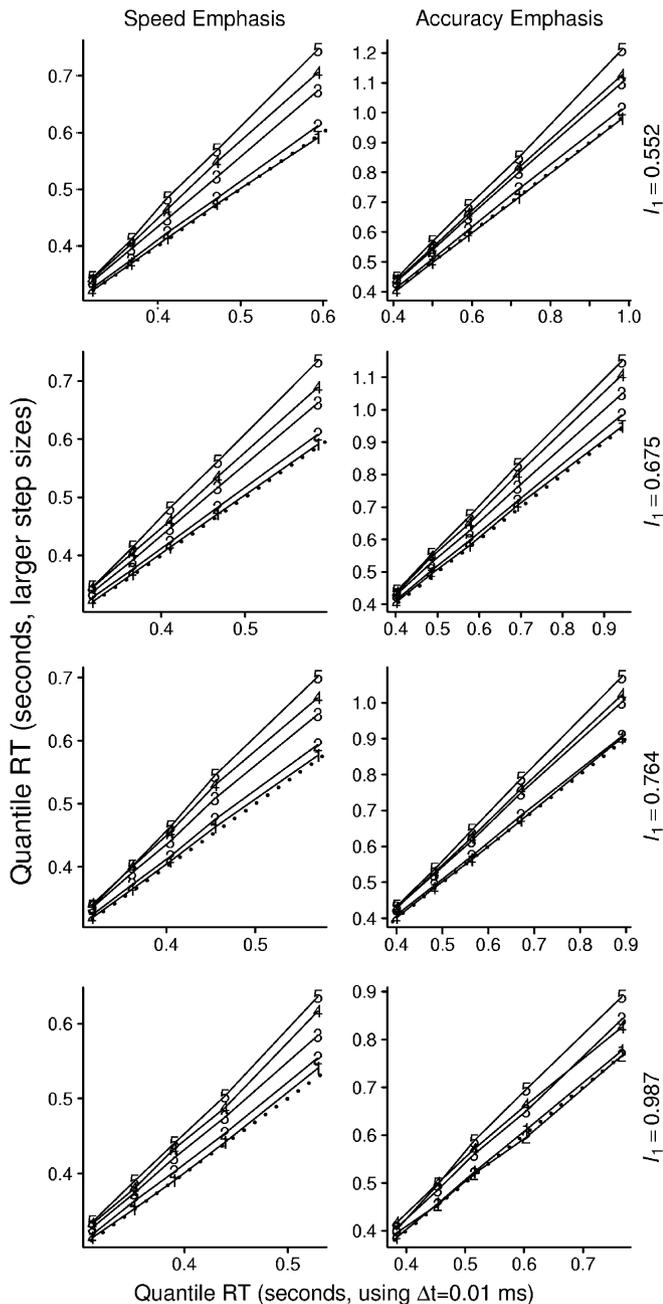


Fig. 2. Plot of estimated quantiles (y-axis) for step sizes $\Delta t = 0.1, 1, 10, 20$ and 50 ms (lines numbered #1–#5, respectively) against quantiles for $\Delta t = 0.01$ ms (x-axis). Dotted line shows $y = x$. All units in seconds.

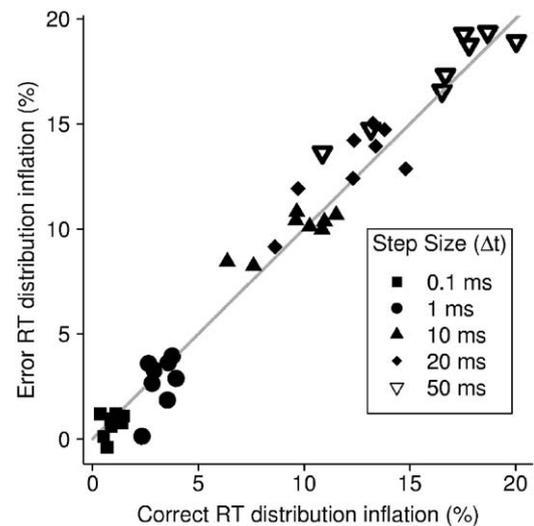


Fig. 3. Inflation coefficients (percentages) for correct and error RT distributions. Each symbol represents a different step size: squares for $\Delta t = 0.1$ ms; circles for $\Delta t = 1$ ms; filled upright triangles for $\Delta t = 10$ ms; diamonds for $\Delta t = 20$ ms; and un-filled inverted triangles for $\Delta t = 50$ ms. Solid gray line shows $y = x$.

will recover the true RT distributions (more on this point later).

The linearity of the quantile–quantile plots in Fig. 2 was also observed for both other integration methods (E2 and E4). Thus, subsequently we report inflation percentages rather than repeating plots like Fig. 2. The performance of the two other integration methods was, disappointingly, almost identical to that of the Euler method. For the E2 method, correct RT distributions were inflated by 0.91%, 3.2%, 9.9%, 14% and 20% for step sizes $\Delta t = 0.1, 1, 10, 20$ and 50 ms, respectively (error RT distributions were inflated by similar amounts). For the E4 method, correct RT distributions were inflated by 0.7%, 3.0%, 10%, 14% and 21% (similar for error RT distributions).

As well as comparing across step sizes within each method, we compared across methods to check that each was converging on the same (true) result with decreasing step size. For all four step sizes, the two more complicated methods (E2 and E4) agreed very closely with the Euler method: inflation of either method compared to Euler was always less than 0.6% for correct RT distributions, and less than 0.9% for error RT distributions (and most often much less than this).

4. General discussion

4.1. A note on parameter scaling

The stochastic differential equation models used in psychology are sufficiently complex that tradeoffs between various model parameters are not always well understood. As an example, the model we have investigated has two different parameter scaling properties that must *both* be understood to allow accurate parameter estimation. Firstly, there is a scaling that adjusts RT distributions. If the following parameters are all scaled by some factor, say w , the predicted RT distributions will all become faster by a factor of w :

$$t_0 \rightarrow t_0/w, \quad k \rightarrow wk, \quad I_{1,2} \rightarrow wI_{1,2}, \quad \sigma \rightarrow \sigma\sqrt{w}.$$

Note that the standard deviation of the Wiener process, σ , is scaled by the square root of w . This scaling solution is useful in adjusting for inflation caused by too-large time steps in numerical integration (see below for an example).

The second scaling property for the model we investigated allows a set of parameters to be altered with *no change* to model predictions. The parameters for the start and end points of the evidence accumulation process (x_0 and C), the input strength parameters (I_1 and I_2) and the standard deviation of the Wiener noise (σ) can all be multiplied by any common value without changing any model predictions. This second scaling property has an important consequence for model selection: One of these parameters (x_0, C, I_1, I_2, σ) can be fixed to an arbitrary value without loss of generality. Typically, researchers have either fixed $\sigma = 1$ or $I_1 + I_2 = 1$.

The two scaling properties we have identified for the leaky accumulator model we have investigated hold more generally for most any stochastic differential equation model. Such models used in psychology can almost always be written in vector form as:

$$d\mathbf{x}(t) = (\mathbf{A} - \mathbf{B}\mathbf{x}(t))dt + \sigma d\mathbf{W}.$$

Here, \mathbf{x} is a vector of accumulator activations (one for each response), \mathbf{B} is a matrix of co-efficients representing leakage, self-excitation, lateral inhibition and perhaps other terms, and $d\mathbf{W}$ is a vector-valued Wiener process. Let the starting points of the accumulators be given by the vector $\mathbf{x}(t_0) = \mathbf{x}_0$, and the response thresholds by vector \mathbf{C} . To give a concrete example of the specification, the model we investigated above had $\mathbf{C} = C$, $\mathbf{A} = [I_1, I_2]$, $\mathbf{x}_0 = 0$, and \mathbf{B} proportional to the identity matrix.

With this model specification, the first scaling property that reduces RT distributions by a factor of w can be written as: parameters \mathbf{A} and \mathbf{B} are multiplied by w ; parameter σ is multiplied by \sqrt{w} ; and t_0 is divided by w . The second scaling property, that does not change any model predictions allows for the parameters: \mathbf{A} , \mathbf{x}_0 , \mathbf{C} and σ to be multiplied by any common factor without changing model predictions.

5. Conclusions

Our results are important for working with sequential sampling models. For those researchers using Euler's method who are interested only in the goodness-of-fit of various models, a large and computationally fast step size (e.g., $\Delta t = 20$ ms) is quite adequate. However, if the goal is to compare or interpret estimated parameter *values*, any step size larger than $\Delta t = 1$ ms may lead to significant bias. There are two possible solutions to this problem. The simplest solution is to use a small step size (e.g., $\Delta t = 1$ ms) and absorb the resulting computational costs. A more efficient solution is to exploit the RT scaling properties of psychological models described above. To illustrate this concretely, suppose one used a computationally efficient, but inaccurate step size, such as $\Delta t = 20$ ms when investigating the model we have used above. After best-fitting parameters were identified by search, the inflation factor could be determined as above, by comparing results from the $\Delta t = 20$ ms simulation with a more accurate $\Delta t = 1$ ms simulation. Suppose this inflation factor was found to be 10%, corresponding to multiplication by 1.1, then the parameters of the model can simply be scaled to reduce all RT distributions using the first scaling solution identified above:

$$t_0 \rightarrow t_0/1.1, \quad k \rightarrow 1.1k, \quad I_{1,2} \rightarrow 1.1 \cdot I_{1,2}, \quad \sigma \rightarrow \sigma\sqrt{1.1}.$$

Similar results hold for all other psychological models. This scaling will reduce both correct and error RT distributions by exactly 10%. The scaled parameters will not perfectly match the previous goodness-of-fit, however, as it is

possible that the predicted response accuracy changes by a few percent when moving from $\Delta t = 20$ ms to $\Delta t = 1$ ms. The solution is to “fine-tune” the scaled parameter values by running a new, small-scale parameter estimation starting from the scaled parameters. This optimization should be carried out with the smaller step size, of course. This two-stage method of parameter estimation allows large, fast steps to be employed for the bulk of the parameter optimization search, using the expensive, small steps only at the final stage.

Aside from implementation and computational costs, the BDT and tracking methods agreed closely on the model predictions, for small step sizes. For larger step sizes, each method becomes less accurate, as seen in Figs. 1 and 2. It is important to note that the particular type of inaccuracies introduced by larger step sizes will differ between the methods. For example, larger step sizes will introduce coarser discretization in both time and state spaces for the BDT method, but only in the time domain for the tracking methods. This does not necessarily mean that the tracking methods will be more robust with larger step sizes: errors in the state-space domain increase approximately as the square of time step sizes for those methods.

The similarity of performances of the three tracking methods we investigated is surprising. Burrage et al. (2004) show that the E2 and E4 methods can provide considerable improvements over the Euler method (up to two orders of magnitude reduction in error). The question to answer is: Why doesn't this possible improvement materialize in evaluating our models? The answer lies in the simplicity of the stochastic models investigated by psychological researchers. All accumulator models used in psychology are governed by systems of stochastic differential equations like (1), with f a linear function and σ constant. These models simplify the methods E2 and E4 considerably. For example, for the accumulator model we examined, both E2 and E4 methods reduce to a modified version of the Euler method, with each new step given by:

$$X(t + \Delta t) = X(t) + \Delta t \cdot f(X(t))[1 - R] + \sigma\sqrt{\Delta t}[1 + R] \cdot \eta, \quad (3)$$

where $R = \frac{1}{2}hk$ for method E2 and $R = \frac{1}{2}hk + \frac{1}{6}(hk)^2$ for method E4. Thus, the methods E2 and E4 are identical to the Euler method except for terms of the order of hk or smaller. For the model we investigated, $k = 2.6$ and h (when expressed in seconds, the units of analysis) varied from 0.05 s (50 ms) down to 0.00001 s (0.01 ms). These very small values mean that the adjustment factor R was quite small, and the E2 and E4 method agreed with the Euler method to many significant places. Note that this same argument holds for accumulator models that include lateral inhibition, and for re-parameterization in terms of seconds (or hours, or days, etc).

The results of our investigation have several important implications for researchers:

1. The BDT matrix approximation method provides accurate solutions with less computation time than the tracking methods, but it is more complex to implement. However, if a more complex model is used (such as one that includes lateral inhibition), the BDT method may be more difficult to implement.
2. The more complicated integration methods we investigated above provide no benefit over the simplest method (Euler) for the model and parameter values we investigated.
3. For the tracking methods, larger step sizes produce large errors in RT distribution methods, but these errors are mostly simply linear scalings.
4. Researchers can use one of two approaches, depending on their computational resources, when using the Euler method to accurately identify parameter values:
 - (a) Use a small step size ($\Delta t = 1$ ms) always.
 - (b) Use a large, fast step size (e.g., $\Delta t = 20$ ms) for initial parameter search. Then use a small step size ($\Delta t = 1$ ms) to determine the inflation factor, and scale the estimated parameters to adjust for this inflation. Finally, fine-tune the estimated parameters with a small-scale optimization starting from the scaled parameters, using $\Delta t = 1$ ms.

References

- Audley, R. J., & Pike, A. R. (1965). Some stochastic models of choice. *British Journal of Mathematical and Statistical Psychology*, 18, 207–225.
- Brown, S. (2002). *Quantitative approaches to skill acquisition in choice RT*. Unpublished doctoral dissertation, University of Newcastle, Australia. Retrieved June 6, 2004, from University of Newcastle, School of Behavioural Sciences Web site: <http://www.newcastle.edu.auchool/behav-sci/ncl/publications.html>.
- Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, 112(1), 117–128.
- Burrage, K., & Burrage, R. M. (1996). High strong order explicit Runge–Kutta methods for stochastic ordinary differential equations. *Applied Numerical Mathematics*, 22, 81–101.
- Burrage, K., Burrage, P. M., & Tian, T. (2004). Numerical methods for strong solutions of SDES. *Proceeding of the Royal Society London*, 460 (2041), 373–402.
- Busmeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100, 432–459.
- Carpenter, R. H. S., & Reddi, B. A. J. (2001). Letters to the editor (reply). *Nature Neuroscience*, 4, 337.
- Cook, E. P., & Maunsell, J. H. R. (2002). Attentional modulation of behavioral performance and neuronal responses in middle temporal and ventral intraparietal areas of Macaque monkey. *The Journal of Neuroscience*, 22(5), 1994–2004.
- Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, 41, 260–274.
- Diederich, A., & Busmeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, 47(3), 304–322.
- Gard, T. C. (1988). *Introduction to stochastic differential equations*. New York: Marcel Dekker.

- Glimcher, P. W. (2003). The neurobiology of visual-saccadic decision-making. *Annual Review of Neuroscience*, *26*, 133–179.
- Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, *5*, 10–16.
- Heath, R. A. (1981). A tandem random-walk model for psychological discrimination. *British Journal of Mathematical and Statistical Psychology*, *34*, 76–92.
- Heath, R. A. (1992). A general nonstationary diffusion model for two choice decision making. *Mathematical Social Sciences*, *23*, 283–309.
- Kloeden, P. E., & Platen, E. (1992). *The Numerical Solution of Stochastic Differential Equations*. New York: Springer.
- LaBerge, D. A. (1962). A recruitment theory of simple behavior. *Psychometrika*, *27*, 375–396.
- Lacouture, Y., & Marley, A. A. J. (1991). A connectionist model of choice and reaction time in absolute identification. *Connection Science*, *3*, 401–433.
- Laming, D. R. J. (1966). A new interpretation of the relation between choice reaction time and the number of equiprobable alternatives. *British Journal of Mathematical and Statistical Psychology*, *19*, 139–149.
- Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, *40*, 77–105.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367.
- Ratcliff, R., Cherian, A., & Seagraves, M. (2003). A comparison of Macaque behavior and superior colliculus neuronal activity to predictions from models of simple two-choice decisions. *Journal of Neurophysiology*, *10*.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Comparing connectionists and diffusion models of reaction time. *Psychological Review*, *106*, 261–300.
- Reddi, B. A. J., & Carpenter, R. H. S. (2000). The influence of urgency on decision time. *Nature Neuroscience*, *3*, 827–830.
- Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, *22*, 9475–9489.
- Sato, T. R., Murthy, A., Thompson, K. G., & Schall, J. D. (2001). Search efficiency but not response interference affects visual selection in frontal eye field. *Neuron*, *30*, 583–591.
- Sato, T. R., & Schall, J. D. (2003). Effects of stimulus–response compatibility on neural selection in frontal eye field. *Neuron*, *38*, 637–648.
- Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *The Journal of Neuroscience*, *16*, 1486–1510.
- Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, *102*, 567–593.
- Smith, P. L. (2000). Stochastic, dynamic models of response times and accuracy: A foundational primer. *Journal of Mathematical Psychology*, *44*, 408–463.
- Tuerlinckx, F., Maris, E., Ratcliff, R., & De Boeck, P. (2001). A comparison of four methods for simulating the diffusion process. *Behavior Research Methods, Instruments, and Computers*, *33*, 443–456.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.
- Vickers, D. (1970). Evidence for an accumulator of psychophysical discrimination. *Ergonomics*, *13*, 37–58.
- Vickers, D., & Lee, M. D. (2000). Dynamic models of simple judgments: II. Properties of self-organizing PAGAN (parallel, adaptive, generalized accumulator network) model for multi-choice tasks. *Non-Linear Dynamics, Psychology and Life Sciences*, *4*, 1–31.
- Wang, Xiao-Jing. (2002). Probabilistic decision-making by slow reverberation in cortical circuits. *Neuron*, *36*, 955–968.