Behavior Research Methods, Instruments, & Computers
2004, 36 (4), 702-716

The efficient computation of the cumulative distribution and probability density functions in the diffusion model

FRANCIS TUERLINCKX
University of Leuven, Leuven, Belgium

An algorithm is described to efficiently compute the cumulative distribution and probability density functions of the diffusion process (Ratcliff, 1978) with trial-to-trial variability in mean drift rate, starting point, and residual reaction time. Some, but not all, of the integrals appearing in the model's equations have closed-form solutions, and thus we can avoid computationally expensive numerical approximations. Depending on the number of quadrature nodes used for the remaining numerical integrations, the final algorithm is at least 10 times faster than a classical algorithm using only numerical integration, and the accuracy is slightly higher. Next, we discuss some special cases with an alternative distribution for the residual reaction time or with fewer than three parameters exhibiting trial-to-trial variability.

Sequential sampling models in general and some generalizations of the Wiener process with two absorbing boundaries (henceforth, the diffusion process or model; Ratcliff, 1978; Ratcliff & Rouder, 1998; Ratcliff, Van Zandt, & McKoon, 1999) in particular, have proved to be very useful tools for fitting data from two-choice decision tasks. The basic diffusion model has some unrealistic features (e.g., equal error and correct reaction time distributions), but by allowing some key quantities of the model to vary from trial to trial, it is able to capture the most important features of the data in a two-choice decision experiment.

A major impediment in applying the diffusion model is its mathematical complexity. For simple versions of the model (an unbiased process with no trial-to-trial variability), closed-form formulas are available for the choice response probabilities, the mean response time, and the response time variance (Wagenmakers, Grasman, & Molenaar, 2004). However, more useful when applying the diffusion model are the probability density function and cumulative distribution function. Unfortunately, even for the most simple process, the latter two contain an infinite sum without an analytical solution. Moreover, by assuming trial-to-trial variability for some of the parameters, nontrivial integrals are added to the equations. Some of those integrals do not have closed-form solutions, and in past applications all integrals have been approximated numerically. However, it is clear that for reasons of precision and computing speed, it is advantageous to have easy-to-compute, closed-form solutions to at least some of the integrals.

The main focus of this article will be the computation of the cumulative distribution function (also called distribution function) and the probability density function (denoted as the density function) for the diffusion model with trial-to-trial variability in drift, starting point, and residual reaction time. Using standard results from calculus, we will present analytical solutions to some of the integrals appearing in the distribution function. We will concentrate our effort mainly on the distribution function because the latter is more commonly used in the fitting process (Ratcliff & Tuerlinckx, 2002; Ratcliff et al., 1999), and therefore it is deemed more important. However, the case of the density function is almost identical and in the second part of the article, we will sketch how the same methods can be applied to the density function.

The structure of the article is as follows: In the next section, the diffusion model with variability in three parameters (Ratcliff et al., 1999) is introduced and the classical method of calculating the distribution function is described. Subsequently, we sketch a new algorithm for computing the distribution function. Next, the new algorithm is applied to the density function, and then some special cases and alternatives are examined.

The Diffusion Model

The general mechanism behind the diffusion process is simple: When confronted with a stimulus in a two-choice decision task, the participant will sequentially sample information, which is mapped into a single signed accumulator, $Z(t)$, representing the amount of accumulated information at time point t. If the accumulator

This work was supported by National Science Foundation Grant SES00-84368 and by the Fund for Scientific Research–Flanders. I thank Andrew Gelman, Andrew Heathcote, Roger Ratcliff, Jeff Rouder, Sverker Sikström, and the editor for their helpful comments. Correspondence should be addressed to F. Tuerlinckx, Department of Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium (e-mail: francis.tuerlinckx@psy.kuleuven.ac.be).

Copyright 2004 Psychonomic Society, Inc.

crosses a fixed upper or lower boundary, the response corresponding to the crossed boundary occurs. The accrual of information occurs continuously in time, and $Z(t)$ is continuously prone to change.

The data obtained from a single trial in a two-choice decision task are a pair of observations consisting of the choice response and the reaction time. Let the random variable X represent the choice so that X takes the value 1 if absorption occurs at the upper boundary (resulting by convention in a correct response) and 0 if absorption occurs at the lower boundary (resulting in an error). The random variable T refers to the time until the (correct or incorrect) response.

First, we assume that the observed reaction time t consists of two components (see, e.g., Luce, 1986): (1) The time to encode the stimulus and execute the response, called the *residual reaction time*, denoted as t_{er}, and (2) the time devoted to the decision making, which will be modeled by the diffusion process. This decomposition can be represented as follows:

$$T = D + t_{er},$$

where T is the random variable representing the observed reaction time, D denotes the decision time, and t_{er} is the residual reaction time.

The amount of accumulated information at the onset of the decision process, $Z(t_{er})$, equals z_0. The starting point z_0 lies between the lower and upper absorbing boundaries, located at 0 and a, respectively. Thus, the information accumulation process starts at z_0 and ends as one of the absorbing boundaries is crossed, in which case $Z(t)$ equals 0 or a. The systematic component underlying the information accumulation process is called the *drift rate*, parameterized by ξ, and defined as the mean rate of information accumulation toward the upper boundary. Technically, it is the expected change in $Z(t)$ in a very small time interval.

Besides a systematic part, noise is present in the decision process, which makes the actual rate of information accumulation vary around the drift rate. The amount of variability in the actual information accumulation is represented by s (s^2dt is the variance of the change in a small time interval dt; see Cox & Miller, 1970, p. 208). If s is large, the diffusion process behaves more wildly, and if s is small, deviations of the actual information accumulation process from the drift rate are small, too. The variability parameter s is involved in a tradeoff relation with the other parameters (multiplying it with a constant and multiplying the other parameters with the same constant do not change the model's predictions), and therefore it serves only as a scaling constant; s will be set to an arbitrary value. The standard value for s in psychological applications is 0.1 (Ratcliff, 1978; Ratcliff et al., 1999), and we will follow this convention in this article. However, we retain s^2 in all equations so that one might use another value (such as $s = 1$) if wished.

Before the diffusion model can be fitted to a data set, its distribution or density function must be known. We will not discuss the derivation of the bivariate distribution function for (X,T) from the first principle here, but only the result (see Cox & Miller, 1970, for a derivation). For the diffusion process described above, the joint distribution function G for an error and the corresponding error reaction time read as (see Equation 1 at the bottom of the page) for $t > t_{er}$ and where

$$\Pr(X = 0; z_0, \xi) = \frac{\exp\left(-\frac{2a\xi}{s^2}\right) - \exp\left(-\frac{2z_0\xi}{s^2}\right)}{\exp\left(-\frac{2a\xi}{s^2}\right) - 1}. \quad (2)$$

Hence, $G_{X,T}(0,t)$ is the probability of observing an error response (therefore, $X = 0$) before time t. The corresponding probability, $G_{X,T}(1,t)$ (the probability of observing a correct response before time t) can be found by replacing z with $a-z$ and ξ with $-\xi$. This holds true because a process in which absorption at the upper boundary leads to a correct response is equivalent to a process with opposite drift rate and an appropriately relocated starting point in which absorption at the lower boundary leads to a correct response. Furthermore, it can be seen from Equation 1 that as $t \to \infty$, the joint distribution function[1] $G_{X,T}(0,t)$ approaches $\Pr(X = 0; z_0,\xi)$: $\lim_{t\to\infty} G_{X,T}(0,t) = \Pr(X = 0, T \le \infty) = \Pr(X = 0; z_0,\xi)$.

The distribution function contains an infinite alternating (due to the sine function) series with no analytical solution. Therefore, in a practical setting, the infinite series must be approximated by a finite partial sum. This is usually done by checking whether the last added term is small enough (because a necessary condition for convergence is that for large k, the added terms should go to zero). However, because of the cyclical nature of the sine function, occasionally the kth term may be zero or very close to zero (because the sine function in the kth term is zero or close to zero; e.g., if $z_0/a = 0.5$ and $k = 2$), whereas the next term $(k + 1)$ may differ again from zero. To avoid this problem, we check whether the absolute values of the last two consecutive terms are smaller than some constant (see also Ratcliff & Tuerlinckx, 2002). To

$$\begin{aligned}\Pr(X = 0, T \le t) &= G_{X,T}(0,t) \\ &= \Pr(X = 0; z_0, \xi) - \frac{\pi s^2}{a^2} \sum_{k=1}^{\infty} \frac{2k \exp\left(-z_0 \frac{\xi}{s^2}\right) \sin\left(\frac{\pi k z_0}{a}\right) \exp\left[-\frac{1}{2}\left(\frac{\xi^2}{s^2} + \frac{\pi^2 k^2 s^2}{a^2}\right)\left(t - t_{er}\right)\right]}{\left(\frac{\xi^2}{s^2} + \frac{\pi^2 k^2 s^2}{a^2}\right)} \end{aligned} \quad (1)$$

see why this works, assume $\sin(Ak) = 0$. Then $\sin[A(k + 1)] = \sin(Ak)\cos(A) + \cos(Ak)\sin(A) = \cos(Ak)\sin(A)$. Because $\cos(Ak)$ equals 1 or -1 and $0 < A < \pi$, $\sin[A(k + 1)]$ will differ from zero. To be on the safe side, we choose in our applications the constant (denoted δ) to be equal to 1×10^{-29}.

In order to get an idea of the number of terms needed in the partial sum until convergence, we computed $G_{X,T}(0,t)$ for a range of values for t and for a different combinations of parameter values and recorded the number of summed terms at convergence. Reaction time t (measured in seconds) was varied from $t_{er} + 0.0001$ to $t_{er} + 0.4$ in steps of 0.00025, yielding 1,600 time points for which $G_{X,T}(0,t)$ was evaluated (it was noted that after $t_{er} + 0.4$, the number of terms needed until convergence stabilizes). The values for a were 0.08 and 0.16, for z they were $a/4$, $a/2$, and $3a/4$, and for ξ the values were 0.0, 0.1, 0.2, and 0.3 (given the choice of $s = .1$).

On average, 15 terms were included in the sum; the median number was 11 and the first and third quantiles were 8 and 15, respectively. In the most extreme case, the number of summed terms was 557, which occurred for the smallest time value $t_{er} + 0.0001$ and $a = 0.16$. The closer t is to t_{er}, the more terms are needed in the partial sum. Considering the number of terms until convergence as a function $n(t)$ of time t, it turned out that $n(t)$ decreased with increasing t as a power function (i.e., $n(t) \sim t^{-1/2}$). In 99% of the cases considered here, fewer than 45 terms were needed.

The diffusion model as presented above usually does not provide a satisfactory fit to the data from two-choice decision tasks, mainly because the model is unable to account for different reaction time distributions for error and correct responses (Ratcliff, 1978; Ratcliff & Rouder, 1998; Ratcliff et al., 1999). To remedy this shortcoming, three parameters in the original diffusion model will be assumed to vary randomly from one trial to another. First, the drift rate ξ is considered as a random variable whose distribution is normal with mean v and standard deviation η, $\xi \sim N(v,\eta^2)$. Thus, even for the same stimulus, the drift rate may be different on different trials.

The second additional source of variability is brought into the model by making the starting point of the diffusion process a random variable. For each trial, the starting point z_0 is assumed to be a random draw from a uniform density:

$$z_0 \sim U(z - s_Z/2,\ z + s_Z/2).$$

Thus, the range of the uniform density is s_Z ($s_Z > 0$) and it stretches out between $z - s_Z/2$ and $z + s_Z/2$ with density $1/s_Z$ at all intermediate points. The location of this uniform distribution, denoted by z, may take any position between 0 and a, but always with the restriction that the end points of the density do not exceed the absorbing boundaries.

As a third extension, it is assumed that t_{er} also shows random trial-to-trial variability. On each trial, t_{er} is an independent identical draw from a uniform distribution with mean T_{er}:

$$t_{er} \sim U(T_{er} - s_t/2,\ T_{er} + s_t/2).$$

The range of the uniform density is s_t ($s_t > 0$), and it stretches out between $T_{er} - s_t/2$ and $T_{er} + s_t/2$ with density $1/s_t$.

If we want to incorporate the trial-to-trial variability on these three parameters in the diffusion model, we must integrate the distribution function in Equation 1 with respect to t_{er}, z_0, and ξ and their densities (see Equation 3 at the bottom of the page). The upper limit of integration for t_{er} is not simply equal to $T_{er} + s_t/2$, because if t is smaller than $T_{er} + s_t/2$, it is impossible for t_{er} to be larger than t. For a similar reason, t_{er} is not simply equal to t, because t_{er} has as an upper boundary $T_{er} + s_t/2$. To simplify the notation in the remainder of this article, we will denote the lower and upper limits of integration with respect to z_0 and t_{er} by Z_L and Z_U and by T_L and T_U, respectively. Because the probability density functions of z_0 and t_{er} are rectangular, they can be replaced in the equation by the constant density values $1/s_Z$ and $1/s_t$, respectively.

Before discussing the computational issues involved in evaluating the distribution function of the diffusion process, we will first outline how the distribution function is used in the fitting process.

Fitting the Diffusion Model With the Chi-Square Estimation Method

Let us assume that we have collected for one person N trials in a single two-choice decision task. A particular trial i ($i = 1, \ldots, N$) results in a pair of observations (x_i,t_i), in which x_i denotes the choice response ($x_i = 1$ if the correct response was given, and zero otherwise), and t_i refers to the reaction time. Suppose that there are N_0 errors and N_1 correct choices (i.e., $N_0 + N_1 = N$). Because there are no experimental conditions, there is only one set of parameters (i.e., a, T_{er}, η, z, s_Z, s_t, and v).

One of the most frequently used methods for estimating the diffusion model parameters is the chi-square method (Ratcliff & Tuerlinckx, 2002). To apply the method, the data must be preprocessed by computing the observed quantiles for both correct and error reaction times. For each type of response, Q quantiles will be computed, and they are denoted as p_{xq}, where x denotes the type of response (i.e., 0 or 1) and q the particular quantile ($q = 1, \ldots, Q$). In previous applications, the 10%, 30%, 50%, 70%, and 90% quantiles are used successfully (such that $Q = 5$) and therefore we will continue to work with them

$$F_{X,T}(0,t) = \int_{-\infty}^{\infty} \int_{z-\frac{s_Z}{2}}^{z+\frac{s_Z}{2}} \int_{T_{er}-\frac{s_t}{2}}^{\min\left(T_{er}+\frac{s_t}{2},t\right)} G_{X,T}(0,t)\, U\left(T_{er} - \frac{s_t}{2}, T_{er} + \frac{s_t}{2}\right) U\left(z - \frac{s_Z}{2}, z + \frac{s_Z}{2}\right) N\left(v,\eta^2\right) dt_{er}\, dz_0\, d\xi \qquad (3)$$

in this article. It is also useful to define the extreme quantiles (0% and 100%) as follows: $p_{x0} = 0$ and $p_{x6} = +\infty$.

The five quantiles for the correct reaction times divide the sample of observed correct response times into six bins, and the five error quantiles do the same thing with the error response times.

Because of the definition of the quantiles, we know that the number of observations in each of the six bins equals: $0.1N_x$, $0.2N_x$, $0.2N_x$, $0.2N_x$, $0.2N_x$, $0.1N_x$ (with $x = 0$ or $x = 1$). The latter six frequencies for the response x are denoted as O_{xj} (with $j = 1, \ldots, Q + 1$). The corresponding expected frequencies can be found by plugging the observed quantiles into the distribution function. The expected frequencies for the correct response times are as follows: $E_{1j} = N[F_{X,T}(1,p_{1j}) - F_{X,T}(1,p_{1,j-1})]$. Note that for the extreme quantiles $p_{x0} = 0$ and $p_{x6} = +\infty$, the distribution function equals 0 and $\Pr(X = 1; z_0,\xi)$, respectively. The expected error frequencies are defined in an analog manner: $E_{0j} = N[F_{X,T}(0,p_{0j}) - F_{X,T}(0,p_{0,j-1})]$.

Next, the chi-square loss function that must be minimized, given the data in order to find the optimal parameter estimates, is defined as

$$\chi^2(a,T_{er},\eta,z,s_Z,s_t,\nu) = \sum_{x=0}^{1}\sum_{j=1}^{6}\frac{\left(O_{xj} - E_{xj}\right)^2}{E_{xj}}. \qquad (4)$$

If the expected frequency is zero or very close to it, numerical problems may arise during the minimization process (division by zero). To avoid these, a small quantity (e.g., 0.00001) can be added to the denominator. The chi-square loss function in Equation 4 is then minimized as a function of the parameter values using an iterative optimization algorithm (e.g., the Simplex algorithm developed by Nelder & Mead, 1965, is used for this task). Further details on the chi-square method and other estimation procedures can be found in Ratcliff and Tuerlinckx (2002).

There is good reason to focus mainly on the chi-square method for estimating the diffusion model's parameters in this article. As was already mentioned, the diffusion model with random t_{er}, z_0, and ξ is mathematically very complex. Therefore, it is important to minimize the computational effort. The chi-square method is very attractive from this point of view because for a given data set with a single experimental condition, only 10 evaluations of the distribution function are necessary within a single iteration of the optimization procedure (i.e., two × five quantiles), no matter how many trials were observed. On the other hand, the maximum likelihood estimation procedure (Myung, 2003; Van Zandt, 2000) would require the evaluation of the probability density function at each data point, which is computationally a much more expensive task.

The Diffusion Model Equations With Random t_{er}, ξ, and z_0

In this section, we will describe a new and efficient algorithm for computing quickly and accurately the distribution function of the diffusion model with random drift, starting point, and residual reaction time. In previous applications of the diffusion model (Ratcliff & Tuerlinckx, 2002), the integration with respect to ξ, z_0, and t_{er} as shown in Equation 3 was done numerically by approximating each integral with a finite sum. An introduction to numerical integration can be found in almost any textbook on numerical analysis (Acton, 1970; Burden & Faires, 1997; Press, Flannery, Teukolsky, & Vetterling, 1989).

Approximating an integral numerically has some disadvantages compared with using a closed-form solution, if the latter exists. The first and most important disadvantage is, of course, that the numerical integral is only an approximation and is therefore subject to inaccuracies. Obviously, the accuracy can be increased by augmenting the number terms in the approximating sum, but that is at the expense of computing speed and from a certain number of terms onward, the gain in precision will become marginal. Second, integrating numerically over a single quantity is often not such a great problem and it can be done with accuracy and speed. But if several parameters are allowed to vary over trials (as in this case), the number of terms that must be summed in approximating the multidimensional integral increases exponentially. The only thing that can be done to limit the computation time is to reduce the number of quadrature nodes (i.e., the points at which the function is evaluated to approximate the integral), but again at the expense of accuracy. Therefore, it is worthwhile to try to solve the integrals analytically.

However, nothing is perfect and there also may be downsides to the use of closed-form solutions. The analytical solution may be a very complicated expression, which is hard to evaluate, or it may suffer from numerical problems. Especially problematic in our case is the presence of the infinite series. The closed-form solution of integration with respect to z_0, t_{er}, and ξ will affect the terms that are summed in the infinite series. It is not impossible that more terms are needed after the analytical integration than were needed before to achieve equal accuracy, such that the gain in speed resulting from performing the analytical integration is offset by the increased number terms in the partial sum. Hence, we need to check the speed of computation when a new algorithm is introduced. It will be shown that some of the integrals appearing in the distribution function may be solved analytically, but for the other ones we still have to rely on numerical approximation.

Let us start with Equation 3, written in full as Equation 5 at the top of the next page.

Due to the structure of the distribution function, the integration can be performed separately for the probability equation and for the part after the minus sign containing the infinite series. Therefore, we discuss the two cases separately, starting with the integration of $\Pr(X = 0; z_0,\xi)$.

Integration of $\Pr(X = 0; z_0,\xi)$. The easiest integral is the one with respect to t_{er}. Because t_{er} does not appear in the probability equation, the integral has a very sim-

$$F_{X,T}(0,t) = \frac{1}{s_t s_Z}\int_{-\infty}^{\infty}\int_{Z_L}^{Z_U}\int_{T_L}^{T_U}\left\{\Pr(X=0;z_0,\xi) - \frac{\pi s^2}{a^2}\exp\left(-z_0\frac{\xi}{s^2}\right)\right.$$
$$\left.\times\sum_{k=1}^{\infty}\frac{2k\sin\left(\frac{\pi k z_0}{a}\right)\exp\left[-\frac{1}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)\left(t-t_{er}\right)\right]}{\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)}\right\}N\left(v,\eta^2\right)dt_{er}dz_0d\xi$$
$$= \frac{T_U-T_L}{s_t s_Z}\int_{-\infty}^{\infty}\int_{Z_L}^{Z_U}\Pr\left(X=0;z_0\xi\right)N\left(v,\eta^2\right)dz_0d\xi - \frac{\pi s^2}{s_t s_z a^2}\sum_{k=1}^{\infty}2k$$
$$\times\int_{-\infty}^{\infty}\frac{\left[\int_{Z_L}^{Z_U}\exp\left(-z_0\frac{\xi}{s^2}\right)\sin\left(\frac{\pi k z_0}{a}\right)dz_0\right]\left\{\int_{T_L}^{T_U}\exp\left[-\frac{1}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)\left(t-t_{er}\right)\right]dt_{er}\right\}}{\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)}N\left(v,\eta^2\right)d\xi \quad (5)$$

ple solution, which is already given in Equation 5. In the case that $T_U = \min(t, T_{er} + s_t/2) = T_{er} + s_t/2$ (i.e., $t > T_{er} + s_t/2$), all parameters referring to the distribution of the encoding and response distribution cancel out the probability part of the equation, and the solution equals the marginal probability of absorption in the lower boundary under the diffusion model with random drift, starting point, and residual reaction time. Only in the case that $t < T_{er} + s_t/2$ do parameters related to the nondecisional time distribution affect the probability part.

Unfortunately, the integrals with respect to z_0 and ξ for $\Pr(X = 0; z_0,\xi)$ do not have analytical solutions, and therefore the intractable double integral must be approximated numerically. However, that is not too bad, because there corresponds only a single $\Pr(X = 0; z_0,\xi)$ to a given set of parameter values. Subsequently, for a set of parameter values, the numerical approximation to the double integral must be computed only once in a run of an estimation algorithm or in a program for plotting $F_{X,T}(0,t)$ as a function of t. In this article, we use common methods for numerical integration, such as Gaussian quadrature (for z_0 because it is bounded between 0 and a) or Gauss–Hermite quadrature (for ξ because the normal density appears in the integrand, and the limits of integration are $-\infty$ and $+\infty$) to approximate the integrals.

Consider first the approximation to the integral over z_0 with a Gaussian quadrature (Abramowitz & Stegun, 1974, p. 887) (see Equation 6 at the bottom of the page), where y_j and w_j are the transformed nodes and weights from a Gaussian quadrature. The standard nodes and weights of a Gaussian quadrature are denoted as y_j' and w_j', and they are suited for integrating a function between -1 and 1. Because the limits of integration in our case are Z_L and Z_U, we must transform the latter nodes and weights into

$$y_j = [(Z_U - Z_L)/2]y_j' + (Z_U + Z_L)/2 = (s_Z/)2y_j' + z$$

and

$$w_j = [(Z_U - Z_L)/2]w_j' = (s_Z/2)w_j',$$

respectively. The standard nodes and weights up to $n = 96$ can be found in Abramowitz and Stegun (pp. 916ff), and an algorithm to compute the nodes and weights for any

$$\frac{1}{s_Z}\int_{Z_L}^{Z_U}\Pr(X=0;z_0,\xi)dz_0 = \frac{1}{s_Z}\int_{Z_L}^{Z_U}\frac{\exp\left(-\frac{2a\xi}{s^2}\right)-\exp\left(-\frac{2z_0\xi}{s^2}\right)}{\exp\left(-\frac{2a\xi}{s^2}\right)-1}dz_0$$
$$\approx \frac{1}{s_Z}\sum_{j=1}^{n}\frac{\exp\left(-\frac{2a\xi}{s^2}\right)-\exp\left(-\frac{2y_j\xi}{s^2}\right)}{\exp\left(-\frac{2a\xi}{s^2}\right)-1}w_j \quad (6)$$

arbitrary n is described in Press et al. (1989). Below we will discuss how many nodes and weights are needed for the integration over z_0.

The integral with respect to ξ cannot be solved explicitly, either. Because the limits of integration with respect to ξ are $-\infty$ and ∞ and a normal density is present in the integrand, the numerical approximation can be done with a Gauss–Hermite quadrature (Abramowitz & Stegun, 1974; Naylor & Smith, 1982). The structure of the integral with respect to ξ is the following (after integrating already with respect to z_0) (see Equation 7 at the bottom of the page), where $N(\xi;v,\eta^2)$ is the normal density with mean v and variance η^2 evaluated at ξ and $q(\xi,y_j)$ equals

$$[\exp(-2a\xi/s^2) - \exp(-2y_j\xi/s^2)] / [\exp(-2a\xi/s^2) - 1].$$

Approximating the integral in the previous equation with a Gauss–Hermite quadrature gives Equation 8 (see bottom of the page), where u_i and r_i are the ith node and weight from the quadrature used to integrate numerically with respect to ξ. These nodes and weights are transformations of the standard nodes u_i' and weights r_i' ($i = 1, \ldots, m$) for a Gauss–Hermite quadrature (Naylor & Smith, 1982). The transformation of the standard nodes u_i' into u_i is as follows: $u_i = \sqrt{(2)}\eta u_i' + v$. The standard weights are transformed into r_i by means of the following rule: $r_i = r_i'/\sqrt{\pi}$. The standard Gauss–Hermite nodes and weights up to $m = 20$ can be found in Abramowitz and Stegun (1974, p. 924), and for $m > 20$ they can be found through the algorithm of Golub and Welsch (1969).

A problem with the numerical integration of $q(\xi,y_j)$ is that the function is not defined for $\xi = 0$ (the denominator becomes 0 at $\xi = 0$). Thus the numerical integration formula may break down if one of the quadrature nodes happens to be zero or very close to it. It is said that the integral is improper because the integrand has an integrable singularity at a known point (Press et al., 1989, p. 115). A simple solution to this problem will be described in a separate section on singularities below. For the time being, we assume that all Gauss–Hermite nodes are sufficiently different from zero.

We have treated the integration of $\Pr(X = 0; z_0,\xi)$, but the integration of $\Pr(X = 1; z_0,\xi)$ can be easily obtained from the preceding by replacing u_i with $-u_i$ and y_j with $a - y_j$ in the formulas. An easier way is to integrate $\Pr(X = 0; z_0,\xi)$ in the first place only with respect to z_0 and ξ (call this integrated quantity p_0) and then

$$(T_U - T_L)/s_t(1 - p_0)$$

gives the desired result.

Integral of the part containing the infinite series. Now we consider the integration of the second part of the distribution function in Equation 5. We see that in Equation 5, the integrals with respect to z_0 and t_{er} appear separately in the formula, and therefore they can be treated independently. First, the integration with respect to t_{er} will be discussed. Let

$$A = 1/2[(\xi^2/s^2) + (\pi^2k^2s^2/a^2)],$$

then the integral has a particularly simple solution (see Equation 9 at the top of the next page).

Also, the integral with respect to z_0 can be derived from applying basic rules of integration. Let

$$B = -\xi/s^2$$

and

$$C = \pi k/\mathrm{a},$$

then the structure of the integral appearing in Equation 5 is:

$$\int_{Z_L}^{Z_U} \exp(Bz_0)\sin(Cz_0)dz_0.$$

To obtain an analytical solution for this integral, one must apply twice the integration by parts formula ($\int udv = uv - \int vdu$) and then solve for the unknown integral. For the first application of the integration by parts formula, take $u = \exp(Bz_0)$ and $dv = \sin(Cz_0)dz_0$, so that

$$\frac{1}{s_Z}\int_{-\infty}^{\infty}\sum_{j=1}^{n}\frac{\exp\left(-\frac{2a\xi}{s^2}\right)-\exp\left(-\frac{2y_j\xi}{s^2}\right)}{\exp\left(-\frac{2a\xi}{s^2}\right)-1}w_j N(\xi;v,\eta^2)d\xi = \frac{1}{s_Z}\sum_{j=1}^{n} w_j\int_{-\infty}^{\infty} q(\xi,y_j)N(\xi;v,\eta^2)d\xi \tag{7}$$

$$\begin{aligned}\frac{1}{s_Z}\sum_{j=1}^{n} w_j\int_{-\infty}^{\infty} q(\xi,y_j)N(\xi;v,\eta^2)d\xi &\approx \frac{1}{s_Z}\sum_{j=1}^{n}\sum_{i=1}^{m} q(u_i,y_j)w_j r_i \\ &= \frac{1}{s_Z}\sum_{j=1}^{n}\sum_{i=1}^{m}\frac{\exp\left(-\frac{2au_i}{s^2}\right)-\exp\left(-\frac{2y_ju_i}{s^2}\right)}{\exp\left(-\frac{2au_i}{s^2}\right)-1}w_j r_i\end{aligned} \tag{8}$$

$$\int_{T_L}^{T_U} \exp\left[-A\left(t-t_{\text{er}}\right)\right]dt_{\text{er}} = \frac{\exp\left[-A\left(t-T_U\right)\right]-\exp\left[-A\left(t-T_L\right)\right]}{A}$$
$$= \frac{\exp\left[-\frac{1}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)\left(t-T_U\right)\right]-\exp\left[-\frac{1}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)\left(t-T_L\right)\right]}{\frac{1}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)} \quad (9)$$

$du = B\exp(Bz_0)$ and $v = (-1/C)\cos(Cz_0)$. The solution then becomes

$$\int \exp(Bz_0)\sin(Cz_0)dz_0 = -\frac{1}{C}\exp(Bz_0)\cos(Cz_0) + \frac{B}{C}\int \exp(Bz_0)\cos(Cz_0)dz_0 + K,$$

where K is the constant of integration. Next, the integration by parts formula is applied again with the same definitions of u and du as before but now with $dv = \cos(Cz_0)dz_0$ and $v = (1/C)\sin(Cz_0)$. The original integral shows up again at the right hand side so that we can solve for it. After some algebra, this gives the equation at the bottom of the page. The definite integral can then be computed as follows (see Equation 10 at the top of the next page).

Unfortunately, the integration with respect to ξ must be approximated numerically. Again, the Gauss–Hermite quadrature with the same nodes and weights as described before for integrating $\Pr(X = 0; z_0,\xi)$ with respect to ξ can be implemented here. There are no singularities for $u_i = 0$.

Collecting Equations 8, 9, and 10, and inserting them in Equation 5, the distribution function equation becomes (see Equation 11 at the bottom of the next page). The corresponding distribution function for the reaction time of a correct response can be obtained from Equation 11 by replacing z with $a - z$ (which is the same as a change of Z_L in $a - Z_U$ and a change of Z_U in $a - Z_L$) and by replacing u_i with $-u_i$.

Although it may seem that the derived solution is satisfactory, there is nevertheless a problem associated with Equation 11. As already mentioned in the introduction, the analytically solved integral may exert an influence on the rate of convergence of the infinite series. For Equation 11, the convergence of the infinite series is very good when $t > T_{\text{er}} + s_t/2$. However, for $t < T_{\text{er}} + s_t/2$,

$$\exp[-1/2(u_i^2/s^2 + \pi^2k^2s^2/a^2)(t - T_U)]$$

equals 1 because $T_U = \min(t,T_{\text{er}} + s_t/2) = t$. Thus, the last term of the last line of Equation 11 goes rapidly to zero, the last line in itself will become quickly equal to one, and consequently the infinite series will be dominated more strongly by the oscillating sine and cosine terms, leading to a poor convergence rate. In the cases we examined, we often needed more than 5,000 terms before the partial sum stabilized. Therefore, we advise using Equation 11 only for $t > T_{\text{er}} + s_t/2$. For the other case, we derive a solution in the next subsection.

The case $t \leq T_{\text{er}} + s_t/2$. Let us take Equation 5 again as the starting point. For the integration of the probability $\Pr(X = 0; z_0,\xi)$, we can proceed in the same way as described above. For the second part of the distribution function (the part containing the infinite series), we solve analytically the integral with respect to t_{er} but leave the integrals with respect to ξ and z_0 untouched for the time being. Applying Equation 9 leads to the equation for the second part of the distribution function, Equation 12 at the top of page 10.

In Equation 12, a difference between two infinite series appears. Let us consider the second term first because it is the simplest to handle. The infinite series must be approximated with a partial sum (see above) and the integrals with respect to ξ and z_0 are then numerically evaluated with a Gauss–Hermite and Gaussian quadrature, respectively. The rate of convergence of the partial sum is very good: The magnitude of consecutive terms diminishes quickly due to the presence of $-k^2$ in the exponent in the numerator and a term of the order k^4 in the denominator.

For the first infinite series in Equation 12, a closed-form formulation exists. The infinite series has essentially the following form and solution (see Prudnikov, Brychkov, & Marichev, 1986):

$$\sum_{k=1}^{\infty}\frac{k\sin(kx)}{\left(k^2+b^2\right)^2} = -\frac{\pi^2}{4b}\frac{\sinh(bx)}{\sinh^2(\pi b)} + \frac{\pi x}{4b}\frac{\cosh\left[\left(\pi-x\right)b\right]}{\sinh(\pi b)}. \quad (13)$$

$$\int \exp(Bz_0)\sin(Cz_0)dz_0 = \frac{\exp(Bz_0)}{B^2+C^2}\left[B\sin(Cz_0) - C\cos(Cz_0)\right] + K'$$

$$\begin{aligned}\int_{Z_L}^{Z_U}\exp(Bz_0)\sin(Cz_0)dz_0 &= \frac{\exp(BZ_U)}{B^2+C^2}\left[B\sin(CZ_U)-C\cos(CZ_U)\right]-\frac{\exp(BZ_L)}{B^2+C^2}\left[B\sin(CZ_L)-C\cos(CZ_L)\right]\\ &= \frac{s^2\exp\left(-\frac{\xi}{s^2}Z_U\right)}{\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)}\left[-\frac{\xi}{s^2}\sin\left(\frac{\pi k}{a}Z_U\right)-\frac{\pi k}{a}\cos\left(\frac{\pi k}{a}Z_U\right)\right]\\ &\quad-\frac{s^2\exp\left(-\frac{\xi}{s^2}Z_L\right)}{\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)}\left[-\frac{\xi}{s^2}\sin\left(\frac{\pi k}{a}Z_L\right)-\frac{\pi k}{a}\cos\left(\frac{\pi k}{a}Z_L\right)\right]\end{aligned} \tag{10}$$

Rewriting the infinite series from Equation 12 in the form of Equation 13 yields

$$\begin{aligned}\sum_{k=1}^{\infty}\frac{k\sin\left(\frac{\pi k z_0}{a}\right)}{\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)^2} &= \frac{a^4}{\pi^4s^4}\sum_{k=1}^{\infty}\frac{k\sin\left(k\frac{\pi z_0}{a}\right)}{\left(\frac{\xi^2a^2}{s^4\pi^2}+k^2\right)}\\ &= -\frac{a^3}{4s^2\pi\xi}\frac{\sinh\left(\frac{\xi z_0}{s^2}\right)}{\sinh^2\left(\frac{\xi a}{s^2}\right)}\\ &\quad+\frac{z_0a^2}{4s^2\pi\xi}\frac{\cosh\left[\frac{\xi(a-z_0)}{s^2}\right]}{\sinh\left(\frac{\xi a}{s^2}\right)}.\end{aligned} \tag{14}$$

Again, the integration with respect to z_0 and ξ is performed numerically.

Another minor problem is associated with Equation 14, because the integration with respect to ξ contains a singularity at $\xi = 0$, since $\sinh(0) = 0$. Again this singularity causes a problem only if one of the Gauss–Hermite quadrature nodes is zero or very close to zero. In the next subsection, we will deal with the treatment of this and the already mentioned singularity (in the integration of the probability).

Improper integrals: Singularities at $\xi = 0$. As we have described, there are two situations in which integrable singularities at $\xi = 0$ occur, and in such cases the integral is called improper. Consequently, if one of the nodes in the numerical approximation is zero or very close to zero, the program may break down. Both singularities can be removed, but the degree of complexity of the two solutions differ.

For the integration of the probability $q(\xi,y_j)$ over $N(\xi;v,\eta^2)$ in Equation 7, a simple solution exists. The limiting value of $q(\xi,y_j)$ as $\xi \to 0$ equals

$$q(0,y_j)\to\frac{a-y_j}{a}, \tag{15}$$

and this result is found by taking the limit to zero and then applying l'Hôpital's rule. This limiting value is the probability of absorption at the lower boundary if the

$$\begin{aligned}F_{X,T}(0,t) &= \\ &\approx \frac{T_U-T_L}{s_ts_Z}\sum_{i=1}^{m}\frac{1}{\exp\left(-\frac{2au_i}{s^2}\right)-1}\sum_{j=1}^{n}\left[\exp\left(-\frac{2au_i}{s^2}\right)-\exp\left(-\frac{2y_ju_i}{s^2}\right)\right]r_iw_j\\ &\quad-\frac{2\pi s^4}{s_ts_za^2}\sum_{k=1}^{\infty}2k\sum_{i=1}^{m}\left(\frac{u_i^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)^{-3}\\ &\quad\times\left\{\exp\left(-\frac{u_i}{s^2}Z_U\right)\left[-\frac{u_i}{s^2}\sin\left(\frac{\pi k}{a}Z_U\right)-\frac{\pi k}{a}\cos\left(\frac{\pi k}{a}Z_U\right)\right]\right.\\ &\quad\left.-\exp\left(-\frac{u_i}{s^2}Z_L\right)\left[-\frac{u_i}{s^2}\sin\left(\frac{\pi k}{a}Z_L\right)-\frac{\pi k}{a}\cos\left(\frac{\pi k}{a}Z_L\right)\right]\right\}\\ &\quad\times\left\{\exp\left[-\frac{1}{2}\left(\frac{u_i^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)(t-T_U)\right]-\exp\left[-\frac{1}{2}\left(\frac{u_i^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)(t-T_L)\right]\right\}r_i\end{aligned} \tag{11}$$

$$\frac{4\pi s^2}{s_t s_z a^2}\int_{-\infty}^{\infty} N\left(v,\eta^2\right)\int_{Z_L}^{Z_U}\exp\left(-z_0\frac{\xi}{s^2}\right)\sum_{k=1}^{\infty}\frac{k\sin\left(\frac{\pi k z_0}{a}\right)\left\{1-\exp\left[-\frac{1}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)\left(t-T_L\right)\right]\right\}}{\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)^2}dz_0d\xi$$

$$=\frac{4\pi s^2}{s_t s_z a^2}\int_{-\infty}^{\infty} N\left(v,\eta^2\right)\int_{Z_L}^{Z_U}\exp\left(-z_0\frac{\xi}{s^2}\right)$$

$$\left\{\sum_{k=1}^{\infty}\frac{k\sin\left(\frac{\pi k z_0}{a}\right)}{\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)^2}-\sum_{k=1}^{\infty}\frac{k\sin\left(\frac{\pi k z_0}{a}\right)\exp\left[-\frac{1}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)\left(t-T_L\right)\right]}{\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)^2}\right\}dz_0d\xi \qquad (12)$$

drift rate equals zero and the starting point is y_j (see Cox & Miller, 1970). Thus to remove the singularity, one may replace the formula in Equation 7 by $(a - y_j)/a$ if the node u_i equals 0 or lies in a small neighborhood of 0. To avoid any problems, we check in practice whether the absolute value of u_i is smaller than 1×10^{-7} (note that this is a heuristic and somewhat arbitrary rule).

A singularity also occurs at $\xi = 0$ if $t \le T_{er} + s_t/2$, and the closed-form expression in Equation 14 of the infinite series has been used. Unfortunately, the solution is more complicated. Instead of continuing to work with the equation that resulted in the singularity, we start again from Equation 1 and integrate analytically with respect to z_0 and t_{er} and approximate numerically the integral with respect to ξ; the result is again Equation 11. Now suppose $u_i = 0$ (here again, we resort to this procedure if $|u_i| < 1\times 10^{-7}$), then some of the terms in the second part of Equation 11 (containing the infinite series) can be dropped, yielding Equation 16 (see the bottom of the page). This equation contains three separate infinite series. For the first two similar ones, a closed-form solution is available; but the last one must be approximated numerically again.

The two similar infinite series appearing in the last equation have the same structure, and a closed-form solution (Prudnikov et al., 1986) exists:

$$\sum_{k=1}^{\infty}k^{-4}\cos(kx)=\frac{\pi^4}{90}-\frac{\pi^2x^2}{12}+\frac{\pi x^3}{12}-\frac{x^4}{48}. \qquad (17)$$

Applying the latter identity to the first infinite series from Equation 16 is straightforward and results in

$$\sum_{k=1}^{\infty}k^{-4}\cos\left(\frac{\pi k}{a}Z_L\right)=\frac{\pi^4}{90}-\frac{\pi^4Z_L^2}{12a^2}-\frac{\pi^4Z_L^3}{12a^3}-\frac{\pi^4Z_L^4}{48a^4}. \qquad (18)$$

Replacing Z_L with Z_U gives the solution for the second infinite series. The third infinite series in Equation 16 must be approximated numerically with a partial sum, but again, this series converges very quickly.

Summary of the computational strategy. The algorithm to compute the distribution function $F_{X,T}(x,t)$ of the diffusion model with random drift, starting point, and residual time component has been programmed in Fortran and Matlab (the Matlab code is a literal translation of the Fortran code). The Fortran and Matlab codes can be found in the Web-based archive of *Behavior Research Methods, Instruments, & Computers* (see the Archived Materials section at the end of this article).

The global structure of the program is as follows. First, it is tested whether t is larger than the lowest theoretical possible value $T_{er} - s_t/2$. To avoid possible numerical problems, t should be 0.001 s larger than $T_{er} - s_t/2$ (given that $s = 0.1$ and the reaction time is measured in seconds). If that is not the case, the distribution function value is set to zero and the program ends. If $t > T_{er} - s_t/2 + 0.001$, the probability of an error response will be computed. In most situations, all Gauss–Hermite quadrature nodes will be sufficiently different from zero so

$$-\frac{\pi s^4}{s_t s_z a^2}\sum_{k=1}^{\infty}2k\left(\frac{\pi^2k^2s^2}{a^2}\right)^{-3}\frac{\pi k}{a}\left[\cos\left(\frac{\pi k}{a}Z_L\right)-\cos\left(\frac{\pi k}{a}Z_U\right)\right]\left\{1-\exp\left[-\frac{1}{2}\left(\frac{\pi^2k^2s^2}{a^2}\right)\left(t-T_L\right)\right]\right\}r_i$$

$$=-\frac{2a^4}{s_t s_z\pi^4s^2}\left\{\sum_{k=1}^{\infty}k^{-4}\cos\left(\frac{\pi k}{a}Z_L\right)-\sum_{k=1}^{\infty}k^{-4}\cos\left(\frac{\pi k}{a}\right)Z_U\right.$$

$$\left.-\sum_{k=1}^{\infty}k^{-4}\left[\cos\left(\frac{\pi k}{a}Z_L\right)-\cos\left(\frac{\pi k}{a}Z_U\right)\right]\exp\left[-\frac{1}{2}\left(\frac{\pi^2k^2s^2}{a^2}\right)\left(t-T_L\right)\right]\right\}r_i. \qquad (16)$$

that Equation 8 can be used. As said above, the latter equation is not valid for $|u_i| < \epsilon$ (where $\epsilon = 1 \times 10^{-7}$), in which case we resort to the simple formula described in Equation 15. Then, we compute the second part of the distribution function, containing the infinite series. If t is larger than $T_{er} + s_t/2$, we rely on Equation 11 for our computations. However, if the opposite is true ($t \leq T_{er} + s_t/2$), then we use Equations 12 and 14 for the Gauss–Hermite quadratures that are larger than ϵ. If $t \leq T_{er} + s_t/2$ and a Gauss–Hermite quadrature node is less than ϵ, we apply Equations 16 and 18. Both programs use 10 nodes and weights for the Gaussian and Gauss–Hermite quadratures (the values for the nodes and weights are supplied with the program code).

Accuracy and the choice of number of quadrature nodes. The quality of numerical methods can be compared with respect to speed and accuracy, and both aspects are very often involved in a tradeoff relation with each other. The importance of a fast method is obvious: Given two equally accurate methods, we prefer the method that results in the fastest evaluation of the distribution function. Concerning accuracy and holding speed constant, we prefer the method for which the computed solution has the smallest deviations from the truth. Because we give priority to the accuracy of the method, we will assess that aspect first, and in the next subsection the computational speed of our method is evaluated.

When evaluating the accuracy, in an ideal situation, we should be able to compare the results from numerical methods with the true values. However, in our case (and many others) the true distribution function is unknown. First, there are common roundoff and cancelation errors, and second, two other kinds of approximations occur in both the classical algorithm and our new algorithm for computing the distribution function that may affect the numerical accuracy: The infinite series is replaced by a partial sum, and some or all integrals are replaced by finite sums. However, we expect the new algorithm to have a higher degree of accuracy than the classical method because in the latter method all integrals must be approximated numerically, while in the former at least some closed-form solutions are available.

If we want to increase the numerical accuracy, there is little we can do on the side of the partial sum because the maximal accuracy is already achieved by requiring the two last terms added to the sum to be smaller than 1×10^{-29}. On the other hand, the number of quadrature points of the finite sums that approximate the integrals can be increased and in the following, the effect of such an operation will be examined.

Using the Matlab program, we computed for a variety of parameter values and reaction times the distribution function with the classical algorithm and with our new algorithm. The number of quadrature nodes varied from 5 to 10 to 20 for all numerically approximated integrals. The range of parameter values was the following: $a = 0.08$ or 0.16; $\eta = 0.08$; $T_{er} = 0.3$; $z = a/4$ or $a/2$; $s_Z = 0.01$ if $a = 0.08$ and $z = 0.02$, 0.03 if $a = 0.16$ and $z = 0.04$, 0.05 if $a = 0.08$ and $z = 0.05$, 0.10 if $a = 0.16$ and $z = 0.08$; $s_t = 0.2$; $v = 0.0$, 0.15, or 0.30. The reaction time was varied from 0.2 to 2.5 sec in steps of 0.05 sec, and the distribution function was always computed for a correct response (hence, $X = 1$). Next, we calculated for all possible parameter and reaction time values the maximum absolute difference between the distribution function values for each method and the number of quadrature nodes. The results can be found in Table 1 (the last column can be ignored for the moment).

A first observation is that the maximum absolute differences between the distribution functions evaluated with different methods and a different number of nodes are very small to moderately small (ranging from 2×10^{-7} to 3×10^{-4}). This shows that both methods give almost identical results under a variety of circumstances (different parameter values, different reaction times, different number of nodes). A second observation is that for both methods, with an increasing number of nodes, the maximum absolute differences between them become smaller; hence, they seem to converge to the same value. If we use 20 nodes for the classical and for the new method, the maximum absolute difference attains its minimum values (2×10^{-7}). Third, it seems that 10 nodes for approximating the integrals with respect to the three random parameters is sufficient when using the classical methods because the maximum absolute difference with the distribution function when using 20 nodes is only 2×10^{-5}. Also, the agreement is good with the classical method with 20 nodes for each dimension. Doubling the nodes in the new algorithm from 10 to 20 does not add much to the precision. Therefore, we conclude that the accuracy

Table 1
Maximum Absolute Difference Between the Computed Distribution Function Using the Classical Method and New Method With a Different Number of Quadrature Points

		Method						
		New			Classical			Avg. Time/ Function Call
Method	Number of Nodes	5 Nodes	10 Nodes	20 Nodes	5 Nodes	10 Nodes	20 Nodes	(sec)
New	5	0	2×10^{-4}	2×10^{-4}	3×10^{-4}	2×10^{-4}	2×10^{-4}	0.002
	10		0	1×10^{-6}	3×10^{-4}	2×10^{-5}	1×10^{-6}	0.004
	20			0	3×10^{-4}	2×10^{-5}	2×10^{-7}	0.012
Classical	5				0	3×10^{-4}	3×10^{-4}	0.023
	10					0	2×10^{-5}	0.184
	2						0	1.500

$$g_{X,T}(0,t)=\frac{\pi s^2}{a^2}\exp\left(-z_0\frac{\xi}{s^2}\right)\sum_{k=1}^{\infty}k\sin\left(\frac{\pi k z_0}{a}\right)\exp\left[-\frac{1}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)\left(t-t_{\text{er}}\right)\right] \tag{19}$$

of the new method is satisfactory and comparable to using an algorithm with only numerical integrations.

Computational speed. After having concluded that the accuracy of our new method is excellent, we should examine its speed of computation and compare it with the classical method. Here, the main advantage of our new algorithm is revealed because it is very fast compared with the classical approach in which all integrals are approximated numerically. To support this conclusion, we carried out the same computations as described above to evaluate the accuracy of the methods and then we used Matlab's profile function to assess how much CPU time is spent on computing the distribution functions with a different number of quadrature nodes.

The last column of Table 1 shows the average time per function call for the two algorithms and different number of quadrature nodes. First, the classical algorithm always takes much more computing time than the new algorithm (10 times more for 5 nodes, 46 times more for 10 nodes, and 125 times more for 20 nodes). Second, doubling the number of quadrature nodes (for all three dimensions) multiplies the average time per function call for the new algorithm with a factor of 2 to 3, while for the classical algorithm with a factor of 8 ($= 20^3/10^3$). (Note that the average times per function call are less in absolute value when running Fortran.)

In conclusion, it seems that using 10 quadrature nodes for all three random parameters in the new algorithm leads to an accurate result within a very reasonable computing time.

The Density Function and Some Special Cases

The final section of this article outlines how the approach taken for the distribution function can be carried over to the density function with only a few changes. Moreover, an alternative distribution function for the residual response time is proposed. Next, we elaborate the analytical solution of the integral of the density function with respect to the drift rate. Finally, a simplification of the infinite series in the case of an unbiased diffusion process is given.

The density function. Calculating the distribution function is necessary when one wants to use the chi-square method for estimating the model parameters. An alternative for the chi-square method is the maximum likelihood estimation method (Myung, 2003; Van Zandt, 2000), and then the probability density function is needed. The joint density function of (X,T) (evaluated at $X = 0$ and $T = t$) is obtained from Equation 1 by differentiating $G_{X,T}(0,t)$ with respect to t (see Equation 19 at the top of the page). This equation does not integrate to 1 (over t) but to $\Pr(X = 0; z_0,\xi)$ [i.e., $\int_0^\infty g_{X,T}(0,t)\,dt = \Pr(X = 0; z_0,\xi)$], the probability of an error. Again, the corresponding formulas for $X = 1$ can be found by changing z to $a - z$ and ξ to $-\xi$. Likewise, $g_{X,T}(1,t)$ integrates to $\Pr(X = 1; z_0,\xi)$: $\int_0^\infty g_{X,T}(1,t)\,dt = \Pr(X = 1; z_0,\xi)$.

If we assume again that a single two-choice decision task with N trials has been carried out, the data of a particular trial i ($i = 1, \ldots, N$) are a pair of observations (x_i,t_i), with x_i being the choice response ($x_i = 1$ if the correct response was given and zero otherwise) and t_i being the reaction time. The likelihood L is then defined as the product of the individual density functions:

$$L\left(a,T_{\text{er}},\eta,z,s_Z,s_t,v\right)=\prod_{i=1}^{N}g_{X,T}\left(x_i,t_i\right). \tag{20}$$

Before maximizing L as a function of the unknown parameters, the logarithm is usually taken (which does not affect the location of the maximum), and the log likelihood is maximized. It is obvious that a likelihood analysis is computationally more burdensome than the chi-square method, because for the former method the density must

$$\begin{aligned}
f_{X,T}(0,t) = &\\
&\approx \frac{\pi s^4}{s_t s_z a^2}\sum_{k=1}^{\infty}k\sum_{i=1}^{m}\left(\frac{u_i^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)^{-2}\\
&\times\left\{\exp\left(-\frac{u_i}{s^2}Z_U\right)\left[-\frac{u_i}{s^2}\sin\left(\frac{\pi k}{a}Z_U\right)-\frac{\pi k}{a}\cos\left(\frac{\pi k}{a}Z_U\right)\right]\right.\\
&\qquad\left.-\exp\left(-\frac{u_i}{s^2}Z_L\right)\left[-\frac{u_i}{s^2}\sin\left(\frac{\pi k}{a}Z_L\right)-\frac{\pi k}{a}\cos\left(\frac{\pi k}{a}Z_L\right)\right]\right\}\\
&\times\left\{\exp\left[-\frac{1}{2}\left(\frac{u_i^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)\left(t-T_U\right)\right]-\exp\left[-\frac{1}{2}\left(\frac{u_i^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)\left(t-T_L\right)\right]\right\}r_i
\end{aligned} \tag{21}$$

$$h_{X,T}(0,t) = \frac{\pi s^2}{a^2\sqrt{2\pi}\, s_t}\exp\left(-z_0\frac{\xi}{s^2}\right)\sum_{k=1}^{\infty} k\sin\left(\frac{\pi k z_0}{a}\right)$$
$$\times\int_{-\infty}^{t}\exp\left[-\frac{1}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)\left(t-t_{\text{er}}\right)\right]\exp\left[-\frac{\left(t_{\text{er}}-T_{\text{er}}\right)^2}{2s_t^2}\right]dt_{\text{er}} \quad (24)$$

be evaluated at every data point, while for the latter method the distribution function only has to be evaluated for the selected quantiles.

Returning to the computation of the density function, it appears that the formulas of the density function are simpler than for the distribution function because $\Pr(X = 1; z_0,\xi)$ is not present in the equation. The same basic integrals as the ones applied to the distribution function can be used in this case, and the density function becomes as shown in Equation 21 (at the bottom of the previous page), where the same definitions apply for the limits of integration and the nodes and weights of the quadrature as in Equation 11. The latter formula works fine if $t > T_{\text{er}} + s_t/2$, but the convergence of the infinite series is again very slow for the opposite case. However, calculations similar to those of the distribution function case lead to a numerically stable algorithm. The main difference is that an infinite series of the form

$$\sum_{k=1}^{\infty}\frac{k\sin(kx)}{\left(k^2+b^2\right)} = \frac{\pi}{b}\frac{\sinh[(\pi-x)b]}{\sinh(\pi b)} \quad (22)$$

appears instead of Equation 13, and

$$\sum_{k=1}^{\infty}k^{-2}\cos(kx) = \frac{\pi^2}{6}-\frac{\pi x}{2}+\frac{x^2}{4} \quad (23)$$

replaces the infinite series in Equation 17. The two sets of equations (one set for the distribution function and the other for the density function) differ slightly because in the distribution function case, the terms in the infinite series are divided by an additional factor containing k^2.

An alternative distribution for t_{er}. Next, we consider the normal distribution for t_{er} instead of a rectangular one. No closed-form solution will be available for the convolution integral but it will be shown that the numerical approximation comes down to finding a way to calculate cumulative standard normal probabilities. Moreover, it is still possible to integrate analytically with respect to z_0, although we do not describe the solution in detail because it does not add anything new. Also because of simplicity, we present in the following only the results concerning the density function. The results for the distribution function are analog and can be derived readily from the following material.

At first, it may seem unrealistic to assume a normal distribution because positive mass is assigned to values for t_{er} that are smaller than zero, which is physically impossible. However, it is quite likely that after fitting the normal distribution, the estimated values for the mean and variance will lead to a distribution that predicts negative values only with a very small probability (note also that the same problem occurs with the ex-Gaussian distribution).

The convolution of the diffusion density with the normal density for t_{er} is (where the mean of t_{er} is T_{er} and the standard deviation is s_t) is as shown in Equation 24 (see the top of the page). Completing the squares in the exponent and writing the standard normal cumulative distribution function as $\Phi(\cdot)$ yields the following expression (see Equation 25 at the bottom of the page). In this case, there is no analytical solution for the convolution integral, but most software packages have built-in functions to evaluate the cumulative standard normal integral (or the related error function[2]) quickly and with very little error. Therefore, for most cases one can rely on the built-in routines to evaluate $\Phi(\cdot)$.

A problem with Equation 25 arises if the counter in the infinite sum, k, becomes large, so that the value in the $\exp(\cdot)$ becomes very large and the value in the standard normal cumulative distribution function $\Phi(\cdot)$ approaches zero. In this case, the standard numerical approximations may fail. However, as noted by Schwarz (2001), one can use an approximation formula given by Derenzo (1977) to circumvent the computational problem. Derenzo (1977) shows that if $y \geq 5.5$,

$$\Phi(-y) = 1-\Phi(y) \approx \frac{1}{y\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}-\frac{0.94}{y^2}\right). \quad (26)$$

$$h_{X,T}(0,t) = \frac{\pi s^2}{a^2\sqrt{2\pi}\, s_t}\exp\left(-z_0\frac{\xi}{s^2}\right)\sum_{k=1}^{\infty} k\sin\left(\frac{\pi k z_0}{a}\right)$$
$$\times\exp\left[\frac{s_t^2T_{er}-t}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)+\frac{s_t^2}{8}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)^2\right]$$
$$\times\Phi\left[\frac{t-T_{er}}{s_t}-\frac{s_t}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)\right] \quad (25)$$

$$h_{X,T}(0,t) \approx \frac{\pi s^2}{a^2\sqrt{2\pi}s_t}\exp\left(-z_0\frac{\xi}{s^2}\right)\sum_{k=1}^{\infty}k\sin\left(\frac{\pi k z_0}{a}\right)\frac{1}{\sqrt{2\pi}\left[\frac{s_t}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)-\frac{t-T_{\text{er}}}{s_t}\right]}$$

$$\times\exp\left\{\frac{s_t^2T_{\text{er}}-t}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)+\frac{s_t^2}{8}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)^2\right.$$

$$\left.-\frac{1}{2}\left[\frac{s_t}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)-\frac{t-T_{\text{er}}}{s_t}\right]^2-\frac{0.94}{\left[\frac{s_t}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2k^2s^2}{a^2}\right)-\frac{t-T_{\text{er}}}{s_t}\right]^2}\right\} \tag{27}$$

The relative error of this approximation is smaller than 0.04% for $y \geq 5.5$. Thus, given that

$$(t - T_{\text{er}})/s_t - s_t/2(\xi^2/s^2 + \pi^2k^2s^2/a^2) \leq -5.5,$$

Equation 25 can be approximated as follows (see Equation 27 at the top of the page). The main advantage of this approximation is that the standard normal cumulative distribution is approximated with an exponential function, and therefore the large negative terms previously in the $\Phi(\cdot)$ function and the large positive terms in the $\exp(\cdot)$ function can now be taken together in a single $\exp(\cdot)$ function. This will lead to a less extreme value in the exponent.

One advantage of assuming that the distribution of t_{er} is Gaussian can be seen if one looks at Equation 11. The formula for the distribution function contains the upper limit of integration over t_{er}, T_U, which was equal to $\min(t, T_{\text{er}} + s_t/2)$. The latter upper limit is a discontinuous function of t, but also of T_{er} and s_t. This may result in discontinuity points in the objective functions that are minimized in order to find parameter estimates (chi-square distances, log likelihoods, etc.). Such discontinuities may cause problems if the minimization routine makes use of (numerical) derivatives. In contrast, when the distribution of t_{er} is assumed to be Gaussian, this discontinuity does not arise.

Trial-to-trial variability for ξ, but not for z_0 and t_{er}. In some situations, a researcher may want to make ξ random but leave z_0 and t_{er} fixed over trials. Of course, whether it is feasible to restrict some parameters to zero is in the first place an empirical issue (i.e., if the estimates for s_Z and s_t happen to be close to zero, there is no harm in restricting them effectively to zero). However, it is possible to give some general guidelines. For instance, if there are no fast errors, that may be an indication of the lack of variability in z_0. Also, small variability over conditions in the leading edge of the reaction time distribution points to the nonexistence of variability in t_{er}. Further information can be found in Ratcliff and Rouder (1998) and Ratcliff et al. (1999).

If there is no variability in z_0 and t_{er}, the integration with respect to ξ has a closed-form solution for the density function (but not for the distribution function). To explain this solution, we start with the integral with respect to ξ (see Equation 28 at the bottom of the page). The squares in the exponent can be completed, such that a function proportional to the normal density appears (called the kernel of a normal density). After some algebraic manipulations, one can find the normalizing factor of the normal density kernel, and the integral can be replaced by the following quantity:

$$\exp\left[\frac{1}{2}\frac{\left(\frac{v}{\eta^2}-\frac{z_0}{s^2}\right)^2}{\left(\frac{t-t_{\text{er}}}{s^2}+\frac{1}{\eta^2}\right)}\right]\sqrt{\frac{2\pi}{\left(\frac{t-t_{\text{er}}}{s^2}+\frac{1}{\eta^2}\right)}}.$$

$$h_{X,T}(0,t) = \frac{\pi s^2}{a^2}\exp\left(-\frac{v^2}{2\eta^2}\right)\frac{1}{\sqrt{2\pi}\eta}\int_{-\infty}^{\infty}\exp\left[-\xi^2\left(\frac{t-t_{\text{er}}}{2s^2}+\frac{1}{2\eta^2}\right)+\xi\left(\frac{v}{\eta^2}-\frac{z_0}{s^2}\right)\right]d\xi$$

$$\times\sum_{k=1}^{\infty}k\sin\left(\frac{\pi k z_0}{a}\right)\exp\left[-\frac{1}{2}\frac{\pi^2k^2s^2}{a^2}(t-t_{\text{er}})\right] \tag{28}$$

$$h_{X,T}(0,t) = \frac{\pi s^2}{a^2}\exp\left(-\frac{v^2}{2\eta^2}\right)\frac{1}{\sqrt{\frac{(t-t_{er})\eta^2}{s^2}+1}}\exp\left[\frac{1}{2}\frac{\left(\frac{v}{\eta^2}-\frac{z_0}{s^2}\right)^2}{\left(\frac{t-t_{er}}{s^2}+\frac{1}{\eta^2}\right)}\right]$$
$$\times\sum_{k=1}^{\infty} k\sin\left(\frac{\pi k z_0}{a}\right)\exp\left[-\frac{1}{2}\frac{\pi^2 k^2 s^2}{a^2}(t-t_{er})\right] \quad (29)$$

$$p_{X,T}(0,t) = \frac{\pi s^2}{a^2}\exp\left[-\frac{1}{2}\frac{v^2(t-t_{er})}{(t-t_{er})\eta^2+s^2}\right]\frac{1}{\sqrt{\frac{(t-t_{er})\eta^2}{s^2}+1}}\exp\left\{-\frac{1}{2}\frac{\left(2vz_0-\frac{z_0^2\eta^2}{s^2}\right)}{\left[(t-t_{er})\eta^2\right]+s^2}\right\}$$
$$\times\sum_{k=1}^{\infty} k\sin\left(\frac{\pi k z_0}{a}\right)\exp\left[-\frac{1}{2}\frac{\pi^2 k^2 s^2}{a^2}(t-t_{er})\right] \quad (30)$$

$$g_{X,T}(0,t) = \frac{\pi s^2}{a^2}\exp\left(-\frac{a\xi}{2s^2}\right)\sum_{n=0}^{\infty}(2n+1)(-1)^n\exp\left[-\frac{1}{2}\left(\frac{\xi^2}{s^2}+\frac{\pi^2(2n+1)^2 s^2}{a^2}\right)(t-t_{er})\right] \quad (31)$$

The formula for $h_{X,T}(0,t)$ now becomes as follows (see Equation 29 at the top of the page). Again the density $h_{X,T}(1,t)$ follows by replacing z_0 by $a - z_0$ and v by $-v$.

There are some numerical problems associated with Equation 29 because an overflow may occur for small values of η (typically for η smaller than 0.03 given that $s = 0.1$). This is due to the fact that in one of the exponents in Equation 29, there is a division by a factor of order η^4, which may result in a very large value that must be exponentiated if η is small. However, by expanding the squares in the exponent, a numerically much more stable formula is obtained (see Equation 30 at the top of the page).

Unfortunately, this derivation is only valid for the density function, and there is no analog closed-form solution for the integral over ξ in the distribution function. Moreover, if one wants to integrate afterward with respect to t_{er} or z_0, this must be done numerically because there are no closed-form solutions available.

The unbiased diffusion model. If there is no integration with respect to z_0, a useful simplification in the density and distribution function formulas can be obtained for the unbiased diffusion model (with $z_0 = a/2$). In that case, the sine term in the infinite series from Equation 1 reduces to $\sin(\pi k/2)$. Next, we may use the following equalities:

$$\sin\left(\frac{\pi k}{2}\right) = \begin{cases} 1 & \text{if } k = 1+4d \\ 0 & \text{if } k = 2d \\ -1 & \text{if } k = 3+4d \end{cases}$$

where d is a nonnegative integer. Then, the original density formula in Equation 19 reduces to (see Equation 31 above). With this simplification, fewer terms have to be calculated than when the sine is left in the formula. For an unbiased diffusion process, the infinite sum in the distribution function can also be simplified in this way.

CONCLUSION

In this article, we have presented several explicit solutions for the integration of the distribution and density function of the diffusion process with respect to parameters that are allowed to vary from trial to trial. These explicit solutions will be helpful if one wants to fit the diffusion model to data. It is shown for the most important case that the new algorithm is sufficiently accurate and fast. It is also illustrated that closed-form solutions for integrals are in general better than numerical approximations but that it is important to check for new computational problems when closed-form solutions are used.

REFERENCES

Abramowitz, M., & Stegun, I. (1974). *Handbook of mathematical functions*. New York: Dover.

Acton, F. S. (1970). *Numerical methods that work*. New York: Harper & Row.

Burden, R. L., & Faires, J. D. (1997). *Numerical analysis* (6th ed.). Pacific Grove, CA: Brooks/Cole.

Cox, D. R., & Miller, H. D. (1970). *The theory of stochastic processes*. London: Methuen.

Derenzo, S. E. (1977). Approximations for hand calculators using small integer coefficients. *Mathematics of Computation*, **31**, 214-225.

Golub, G. H., & Welsch, J. H. (1969). Calculation of Gauss quadrature rules. *Mathematics of Computation*, **23**, 221-230.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, **47**, 90-100.
Naylor, J. C., & Smith, A. F. M. (1982). Applications of a method for the efficient computation of posterior distributions. *Applied Statistics*, **31**, 214-225.
Nelder, J. A., & Mead, R. (1965). A simple method for function minimization. *Computer Journal*, **7**, 308-313.
Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1989). *Numerical recipes: The art of scientific computing*. Cambridge: Cambridge University Press.
Prudnikov, A. P., Brychkov, Y. A., & Marichev, O. I. (1986). *Integrals and series. Vol. 1: Elementary functions*. New York: Gordon & Breach.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.
Ratcliff, R., & Rouder, J. (1998). Modeling response times for two-choice decisions. *Psychological Science*, **9**, 347-356.
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**, 438-481.
Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, **106**, 261-300.
Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, & Computers*, **33**, 457-469.
Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, **7**, 424-465.
Wagenmakers, E.-J., Grasman, R., & Molenaar, P. C. M. (2004). *On the relation between the mean and the variance of a diffusion model response time distribution*. Manuscript submitted for publication.

NOTES

1. We use the term *distribution function*, although that is, strictly speaking, not entirely correct. The function is only a distribution function with respect to the reaction time; it gives the probability mass for the choice response.

2. For instance, in Matlab (without the statistics toolbox) only the error function is available. The error function is defined as $\text{erf}(z) = 2/\sqrt{\pi} \int_0^z e^{-t^2} dt$. From the error function, the cumulative normal probability can be computed as follows: $\Phi(a) = 0.5 + 0.5 \times \text{erf}(a/\sqrt{2})$.

ARCHIVED MATERIALS

The following materials associated with this article may be accessed through the Psychonomic Society's Norms, Stimuli, and Data archive, http://www.psychonomic.org/archive/.

To access these files or links, search the archive for this article using the journal (*Behavior Research Methods, Instruments, & Computers*), the first author's name (Tuerlinckx), and the publication year (2004).

File: Tuerlinckx-BRMIC-2004.zip.

Description: The compressed archive file comprises four files:

CDFDif.f90, containing as an 11K text file the Fortran 90 code for the subroutine CDFDif, which computes the cumulative distribution function from the diffusion model with random residual reaction time, starting point, and drift rate.

ExCDFDif.f90, containing as a 1K text file the Fortran 90 code for the program that illustrates the use of the subroutine CDFDif.

ReadMe.txt, containing as a 2K text file the instructions on how to compile the files CDFDif.f90 and ExCDFDif.f90.

CDFDif.m, containing as an 8K text file the Matlab code for the function CDFDif, which computes the cumulative distribution function from the diffusion model with random residual reaction time, starting point, and drift rate.

Author's e-mail address: francis.tuerlinckx@psy.kuleuven.ac.be.

(Manuscript received June 30, 2003;
revision accepted for publication May 19, 2004.)