Journal of Mathematical Psychology 53 (2009) 55–68

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

On the mean and variance of response times under the diffusion model with an application to parameter estimation

Raoul P.P.P. Grasman *, Eric-Jan Wagenmakers, Han L.J. van der Maas

Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018WB Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 30 August 2007
Received in revised form
19 December 2008
Available online 18 February 2009

Keywords:
Reaction time/response time
Stochastic processes
Diffusion model
Estimation
Response time mean
Response time variance

ABSTRACT

We give closed form expressions for the mean and variance of RTs for Ratcliff's diffusion model [Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108] under the simplifying assumption that there is no variability across trials in the parameters. These expressions are more general than those currently available. As an application, we demonstrate their use in a method-of-moments estimation procedure that addresses some of the weaknesses of the EZ method [Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*, 3–22], and illustrate this with lexical decision data. We discuss further possible applications.

1. Introduction

Speeded two-alternative forced choice experiments are ubiquitous in cognitive psychology and neuroscience. Not surprisingly, the most advanced statistical models in mathematical psychology target these types of experiments. Sequential sampling models are currently the most successful in capturing the statistical features of the data observed in these experiments, and among these, one of the most prominent classes of model are diffusion models (Luce, 1986; Ratcliff, 1978). In particular, sequential sampling models can account for the speed-accuracy trade-off that has been a long-standing source of controversy in experimental psychology (Wickelgren, 1977). Interpreting speed and accuracy data in terms of the parameters that drive the underlying processes is more informative than the traditional analysis of either mean response times or percentages correct (Wagenmakers, van der Maas, & Grasman, 2007). Often, however, the mathematical complexity of these models, and their computational requirement (even with today's computers), tends to discourage researchers from using them.

To study the relationship between mean response time and response time variance predicted by this class of models, Wagenmakers, Grasman, and Molenaar (2005) (see also Palmer, Huk, and Shadlen (2005)) presented closed form expressions for the mean and variance of a simplified, yet analytically tractable, special case of Ratcliff's diffusion model.

These equations, subject to the imposed simplifying assumptions, provided the basis for a method-of-moments estimator for the diffusion parameters that only involves a direct transform of the mean response times (*MRT*), the response time variances (*VRT*), and the proportions of correct responses (P_c). We dubbed this method the "EZ method" (Wagenmakers et al., 2007).

A limitation of the equations and of the EZ method however, is that they are based on the assumption that participants are unbiased with respect to the two responses. In some experiments, participants display bias towards one or other response alternative, due to a participants' response preference, or as a consequence of an experimental manipulation (e.g., presenting 75% words and 25% nonwords in a lexical decision task). Although the equations derived in Wagenmakers et al. (2005) do cover a range of common experimental situations, they tell us little about these more general cases, as bias towards either alternative is an integral part of the decision process as conceptualized in Ratcliff's model.

Besides this limitation of the equations, their application in the EZ method has an additional weakness. Many experimental paradigms, such as, for example, the lexical decision paradigm, comprise two conditions (a 'word' condition and a 'nonword' condition) in which correct and error responses play reversed roles. These conditions are therefore logically intertwined and the diffusion processes in these conditions logically must share

* Corresponding address: Department of Psychology, University of Amsterdam, 1018KP Amsterdam, The Netherlands.
E-mail addresses: rgrasman@uva.nl, raoul.grasman@gmail.com (R.P.P.P. Grasman), ej.wagenmakers@gmail.com (E.-J. Wagenmakers), h.l.j.vandermaas@uva.nl (H.L.J. van der Maas).
URL: http://users.fmg.uva.nl/rgrasman (R.P.P.P. Grasman).

0022-2496/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2009.01.006

parameters. The EZ method does not support such constraints, and handles each experimental condition separately.

The purpose of the present article is to derive a closed form expression of the RT mean and variance for the more general case than the one considered in Wagenmakers et al. (2005). Specifically, we do not assume that the decision is unbiased with respect to the response alternatives. As a practical application of these new expressions, we consider their use in a parameter estimation procedure that is in line with the EZ method, but removes the above mentioned weaknesses. We demonstrate its usefulness with a real data example.

The EZ method is easy by virtue of the analytical invertibility of the equations obtained in Wagenmakers et al. (2005). For the new equations it is not possible to derive closed form expressions for the parameters in terms of proportion correct, RT mean, and RT variance. To use the new equations for the purpose of estimation, we therefore resort to numerical procedures. We demonstrate one such estimation procedure, and determine its effectiveness in simulations. Like the EZ method, this procedure produces method-of-moments estimates. The use of the equations is, however, not limited to method-of-moments estimators as we argue in the last section of the paper. It should be noted that the implementation of the demonstrated procedure is much easier than the statistically more optimal estimation procedures proposed in the literature (e.g., Ratcliff and Tuerlinckx (2002) and Voss and Voss (2008)). More importantly, this estimation procedure is computationally much faster than other available procedures. This can be an advantage when RT data of many participants are to be analyzed on an individual basis, and when estimates constitute the basis for online adjustments of an experiment. The use of a numerical procedure furthermore frees the algorithm from being specific to a single experimental design. With such an algorithm it becomes easy to build more extensive models that incorporate diffusion processes as building blocks for decisions in complex experimental designs, in which parameters are constrained across conditions or may be modeled as functions of covariates or design factors. This is only practically feasible when estimates are obtained quick enough, especially when various models have to be considered and compared.

The outline of this paper is as follows: In the next section, we give a general description of the diffusion model as proposed by Ratcliff (1978). In the subsequent section we derive expressions for mean and variance of a diffusion process, which are more general than those presented by Wagenmakers et al. (2005). Then we apply the derived expressions to the estimation of diffusion model parameters in a similar, but more general, vein as the EZ method (Wagenmakers et al., 2007). We demonstrate the effectiveness of this use in a simulation study, and apply it to real data obtained in a lexical decision paradigm. In the discussion we suggest other estimation methods in which the present expressions can be used. In the appendix, finally, we provide links to software we make available on the internet.

2. Ratcliff's diffusion model

For a single decision, Ratcliff's diffusion model can be conceived of as an information accumulating process over a noisy channel. This process is modeled as the movements of a particle on the interval $(0, a)$. Each of the boundaries of the interval is associated with one alternative (e.g., nonwords and words in a lexical decision task). The particle's position X reflects the accumulation of evidence for one or other alternative. The initial position of the particle at time zero, denoted z, represents the bias towards either of the alternatives. The process is illustrated in Fig. 1. The particle's movements are assumed to be governed by the stochastic differential equation

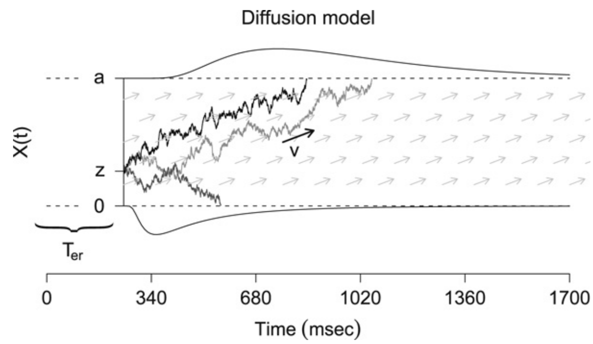

Fig. 1. Illustration of the stochastic information accumulation process underlying the decision component in the diffusion model for simple decisions.

$$dX(t) = \nu dt + s\, dW(t). \tag{1}$$

The equation expresses that the momentary change in evidence follows a constant accumulation rate ν with added random disturbances. The random disturbances, $s\, dW(t)$, are zero-mean random increments with infinitesimal variance $s^2 dt$. The infinitesimal variance ensures that the disturbances are small enough to let the evidence $X(t)$ vary almost always continuously over time, but is large enough to make the process behave erratically and ultimately unpredictably. Once the process exceeds one of the boundaries, the accumulation halts, and the evidence is taken as supporting one or other alternative. The diffusion model can be related to sequential likelihood ratio testing for optimal decision making under uncertainty (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). In fact, this provided the impetus for the introduction of sequential sampling models (Stone, 1960).

It is instructive to consider how three parameters in the model affect the speed-accuracy trade-off. The accumulation rate, or *drift rate*, ν controls the speed of the deterministic information accumulation. Clearly, the greater the value of the drift rate, the more strongly the process is influenced by the deterministic part of equation Eq. (1); hence, the more likely it is to exit the correct end of the interval and the more rapidly it will reach a decision. The boundary separation, controlled by a, not only affects the likelihood of terminating at the correct end of the interval, but also affects the amount of time the decision will take. The particle's starting position is equally important to the likelihood of leaving the correct end of the interval, and the amount of time it takes to reach it. If the process starts close to a, for instance, it will be more likely to exit through a before it can reach 0 than when it starts close to 0. Moreover, it will do so in a shorter amount of time.

The drift rate reflects the correspondence between a probe (stimulus) presented to the participant and an item in his or her task related memory set—i.e., the *goodness-of-match* (Ratcliff, 1978, 1985; Ratcliff & McKoon, 1988; Ratcliff & Smith, 2004). The drift rate is therefore determined by the properties of the probe and by the quality of the memory set, and is under the control of the experimenter rather than the participant. The boundary separation allows the participant to control the conservativeness of his or her evidence criterion (e.g., in response to task instruction). The starting point allows the decision to be biased towards one of the alternatives. Setting the starting point thus provides the participant with a means to increase the response speed in the case that one alternative is expected to be more likely than the other. Empirical validation for these interpretations was presented by Voss, Rothermund, and Voss (2004).

Because the physical properties of the stimuli, or their representation in memory, or both, may vary across trials, Ratcliff's model generalizes to repeated decisions by allowing variability in the drift rate ν. In addition, in an extended version of the model, the starting point z is also allowed to vary across trials (Ratcliff, 1978; Ratcliff & McKoon, 1988). The drift rate is usually assumed to be normally distributed around ν with variance η^2. The starting point is usually assumed to be uniformly distributed in the range $(z - s_z/2, z + s_z/2)$. These distributional assumptions are ad hoc, and should be considered as first approximations to the true underlying distributions. Furthermore, the diffusion process only models the decision process and not to the time needed to encode the stimulus and execute a response. This latter time, whose mean is denoted T_{er}, and decision time are assumed to be additive in the total RT (e.g., Luce (1986)). T_{er} is usually referred to as the (mean) non-decision component. The non-decision component is also allowed to vary randomly across trials. Its law is usually assumed to be uniform over the interval $(T_{\text{er}} - s_{T_{\text{er}}}/2, T_{\text{er}} + s_{T_{\text{er}}}/2)$ as a first approximation to the true underlying distribution. Tuerlinckx (2004) proposes a normal distribution on the basis of computational considerations.[1]

3. Decision time mean and variance

The attraction of the diffusion model for the decision process is not only its theoretical account of how information accumulates in the brain to trigger a decision, but also its ability to provide an accurate description of, and explanation for, many phenomena observed in human and animal RTs (Bogacz et al., 2006; Gold & Shadlen, 2007; Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004; Ratcliff, Thapar, & McKoon, 2001; Smith & Ratcliff, 2004). One such phenomenon concerns the well established (linear) relation between the means and standard deviations of RT distributions. The predictions made by the diffusion model about this relation were studied in Wagenmakers et al. (2005), where expressions for the first two central moments were derived for the case in which the between-trial variabilities were assumed to be absent, and the starting point z was assumed to be equidistant from the two decision boundaries. The latter assumption corresponds to unbiased decisions.

In this section, we derive closed-form expressions for the central moments in the more general case that the decision process is possibly biased. As in Wagenmakers et al. (2005), we will still assume that there is no between-trial variability in any of the diffusion parameters (i.e., $s_z = 0$ and $\eta^2 = 0$). We consider two cases. In the first case, we determine the mean and variance of the time that the process described by equation Eq. (1) exits the interval at either boundary—i.e., the cumulants of the correct and error decision times combined. In the second case we focus on the mean and variance of the time that the diffusion process exits at a particular boundary—i.e., the cumulants of the correct or error decision time only.

To put the derived expressions to some direct practical use, in the next section we apply them in a method-of-moments estimation procedure similar to the EZ method.

Here we switch to the terminology that is common in the literature on stochastic processes, and refer to a particle's position and exit time rather than accumulated evidence and decision time.

[1] Assuming a normal instead of a uniform distribution allows to reduce the computational complexity of evaluating the density function that is implied by the diffusion model because one integral can be carried out analytically.

3.1. Moments of exit times irrespective of exit boundary

Most steps for this case were already discussed in Wagenmakers et al. (2005); hence we only briefly summarize the derivation here and quickly turn to the more general expressions.

The process in Eq. (1) is associated with a partial differential equation (PDE) that governs the evolution of the probability distribution of $X(t)$ across time, given that the process started out at the point z:

$$\partial_t p(x, t|z, 0) = \nu\, \partial_z p(x, t|z, 0) + \frac{s^2}{2}\, \partial_z^2 p(x, t|z, 0). \tag{2}$$

This equation is known as the *Kolmogorov backward equation*, or to be more precise, this is one form of the Kolmogorov backward equation of a time homogenous system. The Kolmogorov backward equation, as opposed to the associated *Kolmogorov forward* or *Fokker–Planck equation*, is the usual starting point for considerations about the exit times of a diffusion process.

We consider the exit time T of the process. Let $G(t, z) = P(T > t)$ denote the probability that a process that started at z exits the interval after time t. Recall that the process terminates as soon as it hits one of the boundaries; i.e., the boundaries are absorbing. Now, suppose the process exits the interval after time t, i.e., $T > t$. Then, because of the absorbing boundaries, the process must still be in the interval at time t (otherwise the process would have stopped earlier than, or at, time t; that is, $T \le t$, which would contradict our assumption that $T > t$). Hence, for $p(x, t|z, 0)$ to be a valid function for the density of this process that is subject to absorbing boundaries, it must satisfy the equality

$$G(t, z) = \int_0^a p(x, t|z, 0)dx,$$

in addition to satisfying the backward equation (2). The backward equation then implies that G satisfies

$$\partial_t G(t, z) = \nu\, \partial_z G(t, z) + \frac{s^2}{2}\, \partial_z^2 G(t, z), \tag{3}$$

with boundary conditions $G(t, 0) = G(t, a) = 0$, as both boundaries are absorbing and hence, $P(T > t) = 0$ for any $t > 0$ (cf., Gardiner (2004) and Wagenmakers et al. (2005)). Note that the assumption that the process starts in z can be stated symbolically as $\lim_{t\downarrow 0} p(x, t|z, 0) = \delta(x - z)$, where $\delta(\cdot)$ is Dirac's delta. Furthermore, as indicated, absorbing boundaries mean that the probability that the particle reenters the interval after it has visited a boundary is zero, which, because the process is (time) homogenous, can be stated formally as $p(x, t|a, 0) = p(x, t|0, 0) = 0$.

The moments of the exit times are given by

$$T_n(z) \equiv \mathrm{E}\{T^n\} = \int_0^\infty t^n [\partial_{t'} P(T \le t')]_t dt$$
$$= -\int_0^\infty t^n [\partial_{t'} G(t', z)]_t dt = n \int_0^\infty t^{n-1} G(t, z) dt,$$

where the latter equality results from integration by parts. This equation can be applied in (3) to obtain the equation for the moments of the exit times:

$$\nu\, \partial_z T_n(z) + \frac{s^2}{2}\, \partial_z^2 T_n(z) = -n T_{n-1}(z). \tag{4}$$

Note that the equation is recursive in the moment order n. Busemeyer and Townsend (1992) provide an alternative derivation of the analogous equation for the more general Ornstein–Uhlenbeck process.

A general solution can be obtained by direct integration of (4) (see Gardiner (2004), but we shall not do so here—the result is analogous to the derivation of the mean an variance of the correct responses that is outlined in the next section. For the first and second order moments the equations turn out to be analytically solvable, which allows us to obtain expressions for the mean and variance of the exit times:

$$\mathrm{E}\{T\} = -\frac{z}{\nu} + \frac{a}{\nu} Z/A, \tag{5}$$

and

$$\mathrm{Var}(T) = \frac{-\nu a^2(Z+4)Z/A^2 + ((-3\nu a^2 + 4\nu za + s^2 a)Z + 4\nu za)/A - s^2 z}{\nu^3}, \tag{6}$$

where, $A = \exp\{-2\nu a/s^2\} - 1$, and $Z = \exp\{-2\nu z/s^2\} - 1$.

As indicated, these equations are the moments of the exit times conditioned on the starting point, but *irrespective of their point of exit*. In RT terms: These are the first two cumulants of the RTs of the aggregated correct and incorrect responses. We next consider the cumulants of the exit times given that the process exits through a particular end of the interval—i.e. the cumulants of the responses conditioned on the correctness of the response. Some exact and approximate results for the discrete time random walk counterpart of the diffusion model in this case were derived by Schwartz (1991).

3.2. Mean and variance of exit times through the lower bound

Before we proceed, consider again the Kolmogorov backward equation in (2) associated with the decision process. As indicated above, this equation is associated with the *Kolmogorov forward* or *Fokker–Planck equation*, which reads

$$\partial_t p(x, t|z, 0) = -\nu\, \partial_x p(x, t|z, 0) + \frac{s^2}{2}\, \partial_x^2 p(x, t|z, 0). \tag{7}$$

This equation is in fact a completely equivalent, but slightly alternative, specification of the probability density $p(x, t|z, t')$. Both equations give rise to the same probability density function (Gardiner, 2004).

The forward equation can be written

$$\partial_t p(x, t|z, 0) + \partial_x j(x, t|z, 0) = 0,$$

where $j(x, t|z, 0) = \nu p(x, t|z, 0) - \frac{s^2}{2}\, \partial_x p(x, t|z, 0)$. The function $j(x, t; z, 0)$ is termed the *probability current* because mathematically, it behaves as a physical current or flux (see Gardiner (2004), sect. 5.2). The probability current describes how much of the probability per unit time flows through a particular point x at time t, as the probability density $p(x, t|z, 0)$ evolves over time. By convention, the direction of flow is here assumed to be pointing to the right. In particular, for the type of processes under consideration, $-j(0, t|z, 0)$ and $j(a, t|z, 0)$ measure the amount of probability that leaks away per unit time at the end points of the interval. Clearly then, the probability that a particle which started at z leaves the interval at the lower boundary after time t is

$$\begin{aligned} g_0(z, t) &= -\int_t^\infty j(0, t'|z, 0)dt' \\ &= \int_t^\infty \left(-\nu + \frac{s^2}{2}\partial_x\right)p(x, t'|z, 0)\bigg|_{x=0} dt' \end{aligned}$$

(cf., Gardiner (2004)), where the first equality expresses the total amount of probability that leaks through 0 after time t. Therefore, the probability that the exit time, $T(0, z)$, of the particle is larger than t given that it exits through 0 is

$$P(T(0, z) > t) = g_0(z, t)/g_0(z, 0), \tag{8}$$

Here, the notation $T(0, z)$ emphasizes that the exit is through the lower boundary 0 and that it depends on the starting point z of the particle. The change in the total probability of the particle being inside the interval at time t is the total probability current that flows out of the interval at the boundaries

$$\frac{\partial P(X(t) \in (0, a))}{\partial t} = j(a, t) - j(0, t)$$

where the minus sign arises because the current is taken to point to the right.

By calculating the partials $\partial_t g_0$, $\partial_z g_0$, and $\partial_z^2 g_0$, and using the backward equation (2), it may be verified that $g_0(z, t)$ therefore satisfies the equation

$$\partial_t g_0(z, t) = j(0, t|z, 0) = \nu\, \partial_z g_0(z, t) + \frac{s^2}{2}\, \partial_z^2 g_0(z, t). \tag{9}$$

As was the case for $G(z, t)$ in the previous section, $g_0(z, t)$ gives rise to an equation for the moments of the exit times, given that the exit is at 0. The n-th order moment of $T(z, 0)$, $T_n(z, 0)$, is defined by

$$\begin{aligned} T_n(z, 0) &= -\int_0^\infty t^n \partial_{t'} P(T(z, 0) > t')\bigg|_t dt \\ &= n\int_0^\infty t^{n-1} g_0(z, t)/g_0(z, 0)dt, \end{aligned}$$

where the second equality result from integration by parts.

On the other hand, using the PDE for g_0 above

$$\begin{aligned} -g_0(z, 0)T_n(z, 0) &= \nu\, \partial_z \int_0^\infty t^n g_0(z, t)dt \\ &\quad + \frac{s^2}{2}\, \partial_z^2 \int_0^\infty t^n g_0(z, t)dt. \end{aligned}$$

Combining these equations, and defining $\pi_0(z) = g_0(z, 0)$, we obtain

$$\begin{aligned} &\nu\, \partial_z(\pi_0(z)T_n(z, 0)) + \frac{s^2}{2}\, \partial_z^2(\pi_0(z)T_n(z, 0)) \\ &\quad = -n\pi_0(z)T_{n-1}(z, 0). \end{aligned} \tag{10}$$

This equation recursively relates the moments of the exit times to each other, conditioned on the exit point 0. Note that the zero-th moment $T_0(z, 0) \equiv 1$. It is clear that the boundary conditions for the solution $\pi_0(z)T(z, 0)$ are

$$\pi_0(a)T(a, 0) = \pi_0(0)T(0, 0) = 0, \tag{11}$$

which result directly from the boundary conditions of the backward Kolmogorov equation in case of absorbing boundaries (the decision process terminates as soon as it hits one of the boundaries). Following (Gardiner (2004), p. 143) $T(0, 0) = 0$, as a process starting at the boundary immediately terminates, and $\pi_0(a) = 0$, as the probability that the process terminates at a if it started at the boundary 0 is zero.

If t in (9) approaches 0, the equation reduces to an equation for $g_0(z, 0) = \pi_0(z)$,

$$\nu\, \partial_z \pi_0(z) + \frac{s^2}{2}\, \partial_z^2 \pi_0(z) = 0, \tag{12}$$

which, together with the obvious boundary conditions $\pi_0(0) = 1$ and $\pi_0(a) = 0$, gives rise to the equation for the probability of an error response given in Ratcliff (1978).

We obtain the mean RT of the error responses by solving (10) for $T_1(z, 0)$, subject to the indicated boundary conditions. An alternative expression was obtained in Palmer et al. (2005)

using different methods. Note that $T_0(z, 0) \equiv 1$. Introducing $\varphi(x, y) = \exp\{2\,\nu\,y/s^2\} - \exp\{2\,\nu\,x/s^2\}$, the solution is found by straightforward integration:

$$T_1(z, 0) = \frac{z\,(\varphi(z-a, a) + \varphi(0, z)) + 2\,a\,\varphi(z, 0)}{\nu\,\varphi(z, a)\varphi(-a, 0)}. \tag{13}$$

The derivation of the expression for the second moment of the decision times is outlined in Appendix A. The variance is obtained by subtracting the square of the mean. Tedious simplifications yield

$$\begin{aligned}\mathrm{Var}(T(z, 0)) = {}& \\ & \frac{-2\,a\,\varphi(0, z)(2\,\nu\,a\,\varphi(z, 2\,a) + s^2\,\varphi(0, a)\,\varphi(z, a))\,e^{2\,\nu a/s^2}}{\nu^3\,\varphi^2(0, a)\,\varphi^2(z, a)} \\ & + \frac{4\nu\,z\,(2\,a - z)\,e^{2\nu\,(z+a)/s^2} + z\,s^2\,\varphi(2\,z, 2\,a)}{\nu^3\,\varphi^2(z, a)}.\end{aligned} \tag{14}$$

To obtain the corresponding equations for the correct responses, use $(\nu, z) \mapsto (-\nu, a - z)$.

3.3. Unconditional versus conditional cumulants

We note several differences between the conditional and unconditional mean and variance. First, conditioned on the point of exit, both the mean and the variance of the exit times converge to an asymptotic value as the starting point approaches the opposite end. That is, for the exits of the process through the lower boundary 0, the mean exit time tends to a finite limit as $z \to a$, and vice versa. The same holds for the variance of the exit time. The unconditional mean and variance on the other hand, both become zero when the starting point approaches either boundary, which is to be expected. A more noteworthy difference is that, while the unconditional mean and variance are reflected in the point $z = a/2$ as the sign of ν is changed, the conditional mean and variance are even functions of ν—i.e., they are symmetrical in the point $\nu = 0$. That is, given a drift rate ν and starting point z, the unconditional mean exit time equals the unconditional mean exit time with drift $-\nu$ and starting point $a - z$. The same is true for the variance of the exit times. The conditional mean and variance on the other hand, are the same for drift rates ν and $-\nu$. The latter implies that the conditional mean and variance do not provide information about the sign of ν, whereas the unconditional mean and variance do. If $z = a/2$ then both unconditional and conditional mean and variance are even functions of ν, and neither contains information about the sign of ν. Only the proportion of correct responses provides information about the sign of ν in that case.

4. Application to parameter estimation

In this section we use the derived equations in a estimation procedure similar to the EZ method. Although the use of the equations and the technique presented in this section can be easily extended to more general use, for simplicity here we concentrate on the method-of-moments, which is the approach of the EZ method.

First, however, we give a brief overview of other approaches that produce estimators with statistically more desirable properties, at the cost of greater computational burden. As indicated earlier, there are several situations in which computation time may be an issue. These include situations in which estimates per participant are desired, and situations in which different complex models, possibly including covariates, need to be compared. A further situation in which computational speed is important is for instance an experimental procedure in which stimulus properties are adaptively changed in response to a participants' performance. In such a situation (near) real-time estimation is necessary.

4.1. Chi-square, WLS, and ML estimation methods

Several methods for estimating the parameters have been proposed (Ratcliff & Tuerlinckx, 2002; Vandekerckhove & Tuerlinckx, 2007; Voss et al., 2004; Wagenmakers, in press). Ratcliff and Tuerlinckx (2002) have extensively reviewed and evaluated three of these methods, namely, minimum chi-square, a weighted least squares method, and maximum likelihood. In the minimum chi-square method the distribution is binned by computing a number of quantiles from the cumulative distribution of both correct and error responses, and fits the model by minimizing the (χ^2-) discrepancy between observed bin frequencies and bin sizes. The weighted least squares (WLS) method on the other hand, directly minimizes the squared differences between computed quantiles and observed quantiles, weighted by their asymptotic accuracy. The maximum likelihood (ML) method used by Ratcliff and Tuerlinckx (2002) evaluates the likelihood by numerically differentiating the cumulative distribution function.

Vandekerckhove and Tuerlinckx (2007) proposed a grouped data maximum likelihood approach to reduce the computation time of full maximum likelihood estimation. In their software, they also provide an option to use the method of Brown and Heathcote (2003). This method, called QMLE, quantizes the observed correct and error RT distributions into several bins, to which a multinomial distribution is fitted by maximum likelihood. The multinomial probabilities are predicted from the diffusion model. Voss et al. (2004) and Voss and Voss (2008) propose to minimize the maximum of the Kolmogorov–Smirnov statistics of correct and error RT distributions.

ML estimators are preferred in many cases, in view of the associated optimal asymptotic properties. (Cases that undermine the usual assumptions of ML theory, some of which are relevant to RTs, are discussed by Cheng and Iles (1987) and Heathcote and Brown (2004).) Even so, Ratcliff and Tuerlinckx (2002) recommend the use of the chi-square estimator, because in their simulations these were more robust than the outlier sensitive ML estimators. Moreover, they were more precise than WLS estimators.

4.2. EZ estimation method

Despite the evident utility of the diffusion model in interpreting the speed and accuracy data, it has failed to catch on in a wider audience of researchers. This may have several causes, the most prominent of which are probably the amount of effort a researcher needs to invest in devising an implementation of one of the estimation methods, and the computational burden of these methods — even on modern computers. The latter is especially problematic when one wishes to investigate different models for complex experimental designs, or to fit the model at the individual level in a large group of participants. For online estimation, as required in adaptive experimental paradigms (e.g., if stimulus discriminability is to be equalized across participants), these methods are impractical.

The EZ method (Wagenmakers et al., 2007) provides easily computable estimates of the parameters of the diffusion model. These are obtained by virtue of the analytical invertibility of the expressions for the moments derived in the previous section for the special case that $z = a/2$—i.e., for the case that the decision is unbiased with respect to either response categories. Furthermore, the EZ method ignores variability in parameters across trials. Thus the EZ method sacrifices some aspects of the full diffusion model for computational ease, and consequently has a more modest range of applicability. The simulations presented by Wagenmakers et al. (2007) showed however that these method-of-moments estimators perform quite well, even when the simplifying assumptions were slightly violated. The method has recently been criticized however (see Ratcliff (in press) and Wagenmakers, van der Maas, Dolan, and Grasman (in press)).

A second disadvantage alluded to earlier is that the EZ method handles a single experimental condition at a time. Random intermixing of trials from different conditions however necessitates that boundary separation must be the same in different types of trials. The EZ method gives separate estimates for each condition however. This constitutes a somewhat inefficient use of the data.

4.3. *Easy estimation method for biased decisions*

In this section we discuss how the equations of the previous section can be used to address the starting point problem of the EZ method. Note that the problem of parameter constraints across conditions becomes more prominent in the biased response case. We therefore will have to address this problem too.

To obtain method-of-moment estimators, we have to equate as many observed moments (i.e., proportions of errors, RT means and variances) to the expression of the corresponding theoretical population values derived in the previous section as there are unknown parameters, and then solve for the unknown parameters.

Unlike the EZ case, analytical inversion of the method-of-moment equations is not possible, and therefore, closed form expressions for the estimators cannot be found. Hence we resort to a numerical algorithm. The resulting estimation procedure turns out to be sufficiently fast to be computed in a web page script, and is simple to implement in a spreadsheet program. To ease the discussion we refer to this method as "EZ2".

We take as an example the common situation where there are two types of trials in which a correct response for one type is an error response for the other and vice versa—a lexical decision task, for example. Assume that the decision processes associated with the two conditions (i.e., words and nonwords) share the starting value z and the boundary separation a, which is appropriate if a participant cannot know in advance the nature of the next trial. Assume further that the decision process associated with each type of condition has its own drift parameter—ν_0 for nonwords and ν_1 for words, say. In addition, assume that RTs modeled with both types of processes have the same non-decision time T_{er}. Then there are five unknown parameters and we need five moment equations.

In both the 'word' and the 'nonword' condition, the proportion of errors, conditional and unconditional means, and conditional and unconditional variances can be calculated. This constitutes a total of ten observed moments. In order to choose an appropriate subset of moments, we consider the following. First, from the previous section we know that in order to be able to estimate the sign of ν we have to include at least one error proportion or an unconditional moment. Second, to be able to estimate T_{er} we have to use at least one mean RT. In fact, the mean RT is not only the sole moment that provides information about T_{er}, it essentially *only* provides information about T_{er}. This can be seen if one considers the partition $MRT = MDT + T_{er}$, where MDT is the mean decision time (or mean exit time in diffusion terms) determined by the diffusion parameters. As long as MDT is smaller than the observed mean RT, which is clearly required, T_{er} will absorb any discrepancy between observed and predicted mean RT. Hence the observed mean RT only bounds the region in which the diffusion parameters must lie, and does not provide information about the specific values within that region. Often, furthermore, T_{er} is not of primary interest and the equations involving means then can safely be ignored (except of course for checking the condition $MDT < MRT$). Finally, it sometimes seems reasonable to assume that error responses have a higher proportion of contamination and, therefore, to restrict the attention to correct responses. We are then left with 4 observed moments and 4 unknown parameters: a variance for the correct RTs for words, a variance for the correct RTs for nonwords, a percentage of errors for the words and a percentage of errors for the nonwords, The nonlinear system that needs to be solved then consists of 4 equations. The simulations presented below focus on this setting.

Numerical methods to solve such nonlinear systems of equations are discussed in Press, Flannery, Teukolsky, and Vetterling (1993). These generally involve defining a nonnegative potential function (e.g., a least-squares function), whose gradient involves the system in a way that the gradient is zero if and only if the system is solved. The system is then solved by finding the minimum of the potential function using an optimization scheme.[2] The next section demonstrates the ability of this procedure to produce valid parameter estimates in a number of numerical simulations.

4.4. *Simulations*

The simulations follow essentially the same setup as those in Wagenmakers et al. (2007). Overall the simulations show that when the starting point is not too close to the boundary separation parameter, the EZ2 estimators perform well when the number of trials per condition exceeds about 250, or when the number of trials per condition exceeds 125 and drift rates are not very high. Overall it appears to be more difficult to estimate parameters when the drift rates are very high and when the proportions of errors are very low.

4.4.1. *Setup*

We simulated RTs under conditions like those in a lexical decision task. The values of the drift rates, boundary separation and starting point, are listed in Table 1. Drift rates ν_1 and ν_2 (for 'word' and 'nonword' conditions) were chosen such that ν_1 was always strictly larger than ν_2. The table also shows the theoretical mean RTs, the percentages of errors, and the RT variances corresponding to these parameter values. For each combination of parameters, we simulated 100 data sets, with $N = 50$, 250, or 1000 trials, of which the word- and nonword conditions each received $N/2$ trials.

A problem with using only few simulated trials is the possible occurrence of perfect performance (absence of errors). Because the method only works if the proportion of errors is nonzero, we discarded data sets without error responses. The results below are therefore conditioned on the presence of error responses. Perfect performance can be dealt with as suggested in Wagenmakers et al. (2007). Here we did not do so, in order to be able to separate pure estimator bias from bias due to bias in the estimated moments.

We found the EZ estimates of ν, a, and T_{er}, together with z equal to half the estimate of a, to be effective starting values. We obtained two sets of EZ estimates – one based on the statistics from one condition and one based on the statistics from the other – and used both in a separate round of fitting. We retained those estimates where the gradient of the potential had the smallest L_2-norm.

We have explored several standard optimization algorithms, including the Nelder–Mead (or 'simplex') algorithm, the Hooke and Jeeves algorithm, and quasi-Newton and Newton–Raphson algorithms (Gill, Wright, & Murray, 1986; Hooke & Jeeves, 1961;

[2] Note that although this may seem very similar to a least squares fit, it is in fact not—the difference being that in order to solve the system, the minimum of the objective function must be identically zero.

Table 1
Parameter values as used in the simulations, and corresponding moments of the correct responses.

Parameters			Moments		
ν	z	a	% Error	MRT	VRT
0.1	0.03	0.08	43.5	424.9	15827.4
0.2	0.03	0.08	27.1	404.7	11514.4
0.3	0.03	0.08	15.8	381.5	7499.4
0.1	0.05	0.08	20.8	372.8	13531.0
0.2	0.05	0.08	9.9	355.7	9532.4
0.3	0.05	0.08	4.2	337.0	5915.4
0.1	0.07	0.08	5.6	296.5	6437.5
0.2	0.07	0.08	2.1	288.7	4239.8
0.3	0.07	0.08	0.7	280.7	2420.1
0.1	0.03	0.11	49.3	594.3	52057.5
0.2	0.03	0.11	29.3	534.4	30808.8
0.3	0.03	0.11	16.4	478.1	16583.4
0.1	0.05	0.11	28.9	542.2	49761.1
0.2	0.05	0.11	12.5	485.4	28826.9
0.3	0.05	0.11	4.8	433.5	14999.4
0.1	0.07	0.11	15.3	465.9	42667.6
0.2	0.07	0.11	4.9	418.3	23534.2
0.3	0.07	0.11	1.4	377.2	11504.1
0.1	0.03	0.14	52.0	801.5	121424.9
0.2	0.03	0.14	29.9	675.9	58264.2
0.3	0.03	0.14	16.5	577.3	27069.2
0.1	0.05	0.14	32.7	749.3	119128.5
0.2	0.05	0.14	13.2	626.9	56282.2
0.3	0.05	0.14	5.0	532.7	25485.3
0.1	0.07	0.14	19.8	673.1	112035.0
0.2	0.07	0.14	5.7	559.9	50989.6
0.3	0.07	0.14	1.5	476.4	21989.9

Note. Units of MRT and VRT in this table were rescaled and rounded to milliseconds. $T_{er} = 250$ ms in all cases.

Kaupe, 1963; Press et al., 1993; Seber & Wild, 1989). The algorithms did not differ very much, although the Hooke and Jeeves algorithm seemed to be slightly more accurate than the simplex algorithm, and is far simpler to implement than the other algorithms.

Although possible (e.g., Gill et al. (1986)), we did not put any effort into imposing any of the natural constraints on parameters (e.g., $0 < z < a$). We never encountered estimates that violated these constraints,[3] thus keeping the method simple.

4.4.2. Results

Figs. 2–5 display the EZ2 results for the parameters a, z, ν_1 and ν_2, respectively in box-and-whisker plots. These estimates were based on the correct responses only. The results based on the pooled correct and error responses were very similar, and the conclusions that can be drawn from these simulations are essentially the same. We therefore limit the discussion to the results displayed in Figs. 2–5. We discuss the performance of the parameter estimators in terms of bias below.

Consecutive columns in the three-by-three panel array in Fig. 2 indicate that the boundary separation a is well recovered. The performance deteriorates, however, as the drift rate increases, unless the number of trials is increased. The distance between z and a also influences the recovery of a, but the adverse effects of the distance on the estimate disappear when the number of trials is high.

Similar conclusions hold for the starting point z. Higher drift rates also worsen the recovery of z, as do smaller distances between starting point and boundary separation. The latter is especially noticeable from the top row of panels in Fig. 3. The distribution of z estimates is also more symmetrical and narrower as z is more equidistant from the boundaries.

The recovery of the drift rates is also affected by the values of the drift rates themselves (compare middle row panels in Figs. 4 and 5), as well as by the distance of starting point from the boundaries (see bottom row panels of Figs. 4 and 5). However as trial numbers increase, the bias quickly vanishes in all cases.

In conclusion, the parameter recovery performance of these method-of-moment estimators seems satisfactory, provided numbers of trials are sufficiently large when drift rates are large or the decision bias is strong. The key factor in the performance of this estimator seems to be the proportion of errors: the fewer the errors, the poorer the recovery. Incidentally, Ratcliff and Tuerlinckx (2002) draw the same conclusion for the chi-square, WLS, and ML methods. Bearing these results in mind, we apply this method to data from an actual experiment in the next section.

4.5. Application to lexical decision data

For illustrative purposes, we apply the EZ2 method to empirical data. The complete task is described in Wagenmakers, Ratcliff, Gomez, and McKoon (2008); here we only summarize the relevant features. The RT data were collected in 19 university students, who participated in a lexical decision task with 75% nonwords and 25% words. Word frequency was varied from 'very low' to 'low' to 'high'. The preponderance of nonwords presumably biases the starting point towards the nonword boundary, whereas the word frequency should affect drift rate for words but not for nonwords—that is, higher frequency words are presumably more strongly represented in memory, and hence their drift rate should be higher. The nonwords consisted of pseudo-words that were generated by changing the vowels of existing high frequency, low frequency, and very low frequency words. Because 'very low', 'low', and 'high' frequency words were randomly intermixed, the bias should not be affected by word frequency, and neither should boundary separation and nonword drift rate. Two of the participants showed perfect performance in one of the conditions. Although this can be dealt with using the method suggested in Wagenmakers et al. (2008), we discarded these two cases from the present, illustrative analysis. Variances of correct responses and percentages of errors of 17 participants were fitted individually to a model in which the lower boundary and upper boundary corresponded to a word response and a nonword response, respectively. The word and nonword responses from different word frequencies were fitted separately, so that for each word frequency condition, we obtained a boundary separation (a), a starting point (z), a drift rate for words (ν_1) and a drift rate for nonwords (ν_0). The means of the parameter estimates across participants are given in Table 2, along with their standard errors in parentheses. A Hotelling's T^2 test revealed significant differences in parameter vectors for the different word frequencies ($F(8, 9) = 5.144$, $p = .0122$). Post hoc analysis revealed that these could only be attributed to differences between very low and high frequency words ($F(4, 13) = 12.51$, $p = .0008$), to differences between low and high frequency words ($F(4, 13) = 7.509$, $p = .0023$), but not between very low and low frequency words ($F(4, 13) = .404$, $p = .316$). Subsequent t-tests revealed significant differences *only* for the word drift rates (ν_1) between low and high word frequencies ($t(16) = 3.259$, $p = .005$) and between very low and high word frequencies ($t(16) = 5.731$, $p = .00003$).

Note that these results are consistent with our expectations, except perhaps for the lack of the anticipated difference between the word drift rates in the very low word frequency and the low frequency conditions. The latter however may be due to a lack of statistical power. Note furthermore that the drift rate for nonwords

[3] This should not be surprising because both the variance formulas as well as the error proportion formula become negative when z is outside of $(0, a)$, and the observed values of course never are.

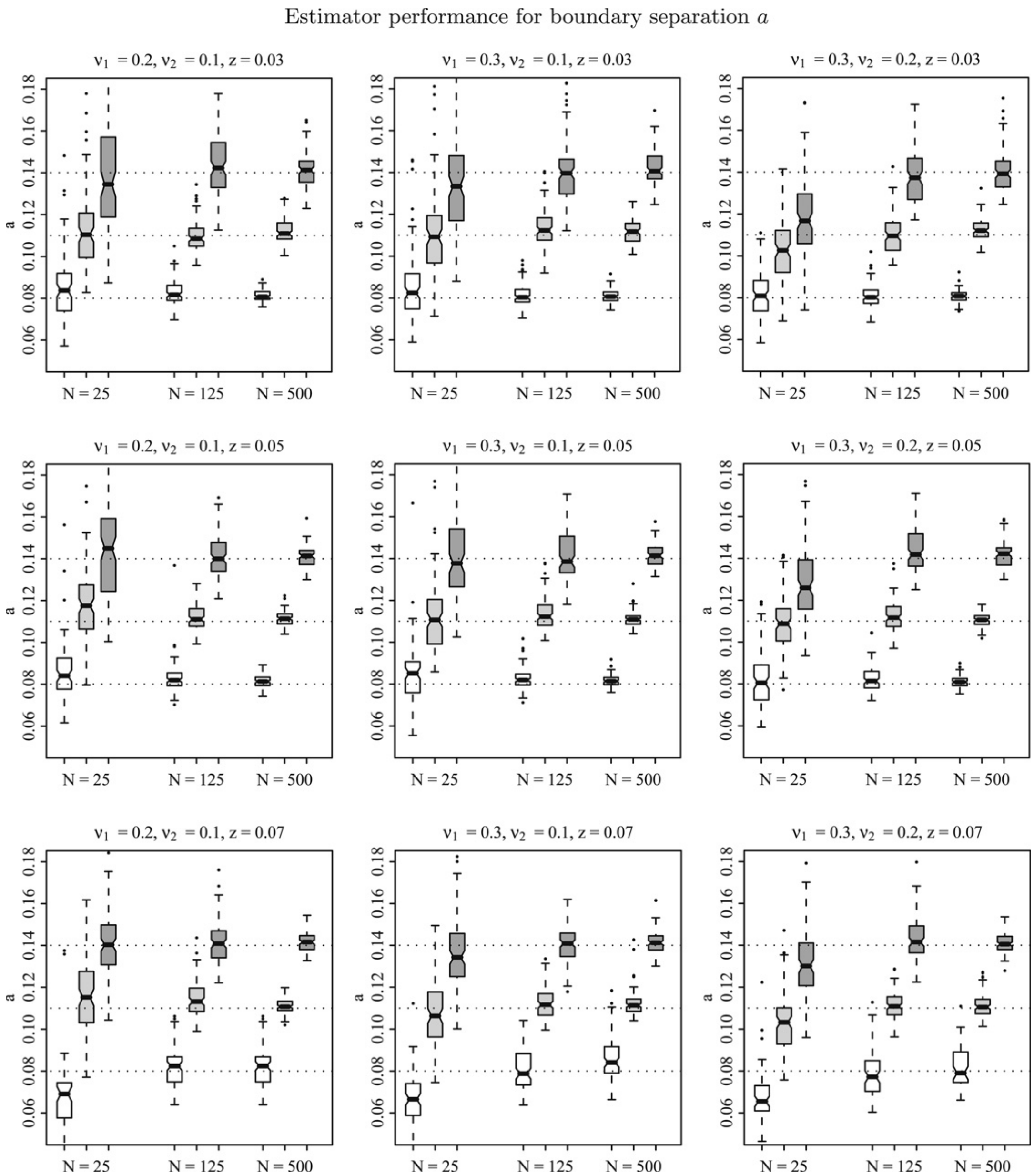

Fig. 2. Box-and-whisker plots for the EZ2 estimates of the boundary separation a. The dotted line indicate the true values $a = 0.08$ (white boxes), $a = 0.11$ (light gray boxes), and $a = 0.14$ (dark gray boxes).

Table 2
EZ2 Parameter estimates for correct responses in the lexical decision task.

Word frequency	ν_0	ν_1	z	a
Very low	.177 (.018)	−.195 (.028)	.1013 (.0069)	.149 (.0083)
Low	.168 (.012)	−.252 (.022)	.1034 (.0064)	.143 (.0073)
High	.186 (.013)	−.362 (.028)	.0939 (.0054)	.141 (.0075)

Note. Parameter estimates from fits to variances of correct responses and error percentages in the lexical decision task. Standard errors as determined from across participant variance are indicated between parentheses. Only the differences in words drift rate ν_1 between low frequency words condition and the high frequency words condition, and between very low frequency words condition and the high frequency words condition are statistically significant.
Note that the drift rates for words are here signed negatively by the convention. In the EZ2-method, word and nonword conditions are conceptualized as separate diffusions of which the starting point is constraint by the starting point in the word conditions (i.e., if z is the starting point in the nonword condition, $a - z$ is taken to be the starting point in the word condition). In this conceptualization, the correct alternative is always assigned to the upper boundary and the incorrect alternative is always assigned to the lower boundary, so that a positive drift rate will always indicate a drift towards the correct decision boundary. Hence both word conditions and nonword conditions will be associated with a positive drift rate in normal circumstances. In the more conventional conceptualization the boundaries are always associated with response alternatives. Hence, the drift rates for the alternative assigned to the lower boundary (e.g., words) will be associated with a negative drift rate.

Fig. 3. Box-and-whisker plots for the EZ2 estimates of the parameter z. The dotted line indicate the true values $z = 0.03$ (white boxes), $z = 0.05$ (light gray boxes), and $z = 0.07$ (dark gray boxes).

is close to the drift rate for very low frequency words.[4] Given that drift rate is indicative of the quality of the memory representation for the item, this seems quite reasonable theoretically for the pseudo-words used. In addition, the starting point z is closer to a, the nonword boundary, which indicates a clear bias towards nonword responses, as is to be expected from the nonword/word ratios.

Because we only used correct responses for the parameter estimation we may have lost information that will enable us to detect the word drift rate difference between the very low and low word frequencies conditions. We repeated the analysis on parameter estimates that were obtained from fitting the percentages of errors and variances computed over the pooled error and correct responses. The means of the estimates are tabulated in Table 3. Using RTs variances of pooled error and correct responses instead of using only correct responses hardly affects the estimates and their standard errors,[5] except for a slightly lower mean estimated value of ν_1 in the very low frequency words condition (i.e., .188 vs. .195). The statistical analysis of these estimates led to the same results as above, except that here an additional marginal difference was detected in ν_1 between low frequency words and very low frequency words. This is presumably caused by a somewhat more pronounced difference between the low word frequencies condition and the very low word frequencies condition.

In Wagenmakers et al. (2008) the chi-square method was used to fit the full diffusion model to the .1, .3, .5, .7, .9 quantiles that were averaged across participants. In the fit of the model

4 A pairwise comparison did not detect a significant difference between ν_1 and ν_0 for the very low word frequencies whereas it did for the low and high frequency words.

5 Correlations between parameter estimates all >.9; for z and a all >.96.