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

* Corresponding author.
Email addresses: rgrasman@uva.nl (Raoul P. P. P. Grasman), ej.wagenmakers@gmail.com (Eric–Jan Wagenmakers), h.l.j.vandermaas@uva.nl (Han L. J. van der Maas).
URL: http://users.fmg.uva.nl/rgrasman (Raoul P. P. P. Grasman).

[1] Preparation of this article was sponsored in part by a VENI-grant from the Netherland Organisation for Scienctific research (NWO) awarded to the corresponding author. We thank the Philip Smith, Andrew Heathcote and an anonymous reviewer for helpful suggestions and comments on a previous version of this manuscript.

Abstract

We give closed form expressions for the mean and variance of response times under Ratcliff's diffusion model (Ratcliff, 1978) if the simplifying assumption is made that there is no variability across trials in the parameters. The expressions given are more general than have so far been available in the literature. As an application, we demonstrate their use in method-of-moments estimator that addresses some of the weaknesses of the EZ method (Wagenmakers, van der Maas, & Grasman, 2007), and illustrate this with lexical decision data. We discuss further possible applications.

Key words: reaction time/response time, stochastic processes, diffusion model, estimation, response time mean, response time variance

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 obtained in these experiments, and among these, one of the most prominent class of models are diffusion models (Ratcliff, 1978; Luce, 1986). In particular, sequential models are able to account for the speed-accuracy trade off that has been a major source of controversy in experimental psychology for decades (Wickelgren, 1977). Interpreting speed and accuracy data in terms of the parameters that steer the underlying processes is much more informative than the traditional analysis of either mean response times or percentages correct (Wagenmakers et al., 2007). It is therefore unfortunate

Preprint submitted to Journal of Mathematical Psychology *8 October 2008*

to see that the mathematical complexity of these models, as well as the computational load for fitting—even with todays 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, in a previous paper (Wagenmakers, Grasman, & Molenaar, 2005; see also Palmer, Huk, & Shadlen, 2005) we obtained closed form expressions for the mean and variance of a simplified, yet analytically tractable, special case of Ratcliff's diffusion model.

Subsequently, these equations suggested to us a way to alleviate the technical pain associated with fitting the model in practical data analysis. Subject to the simplifying assumptions under which they were derived, inversion of the equations provided us with a method-of-moments estimator for the 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). Appropriately enough, we dubbed this method the "EZ method" (Wagenmakers et al., 2007).

A limitation of the equations and the EZ method however, is that the special case for which the equations are valid postulates that participants are unbiased with respect to either of the two response choices. In certain experiments, participants are in fact biased towards one or another response alternative—sometimes due to a participants' response preference, sometimes due to experimental manipulation (e.g., presenting 75% words and 25% nonwords in a lexical decision experiment). 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 making 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, are comprised of 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 for each of these conditions logically must share parameters. The EZ method does not support such constraints, and handles each of these experimental conditions separately.

The purpose of the present article is to find closed form expression of response time mean and variance for the more general case than considered in Wagenmakers et al. (2005), in which it is not presumed 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 turns out not to be possible to derive closed form expressions for the parameters in terms of proportion correct, and RT-mean and RT-variance. To use the new equations for the purpose of estimation, one has to 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 & Tuerlinckx, 2002; Voss & Voss, 2008). More importantly, this estimation procedure is computationally much faster than other available procedures. This can be a major advantage, especially if response time data are to be analyzed on an individual basis; particularly when many individuals participate in a study, or 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 use 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 structure of this paper is as follows: In the next section, we first give a general description of the diffusion model as proposed by Ratcliff (1978). In section 3 we derive expressions for mean and variance similar to those in (Wagenmakers et al., 2005) for a more general diffusion process. In section 4 we apply the derived expressions to the estimation of diffusion parameters in a similar, but more general, vein as the EZ method (Wagenmakers et al., 2007). We demonstrate the effectiveness of this use of the expressions in simulations, and apply them to real data obtained in a lexical decision paradigm. In the discussion we hint at other estimation methods in which the expressions can be used. In the appendix we provide pointers to software implementations that

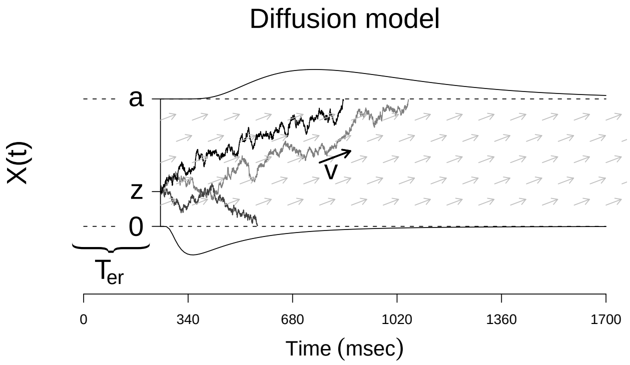

Fig. 1. Illustration of the stochastic information accumulation process underlying the decision component in the diffusion model for simple decisions.

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 represents the evidence for one versus the 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 Figure 1. The particle's movements are assumed to be governed by the stochastic differential equation

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

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 as to let the evidence $X(t)$ vary almost always without sudden jumps 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 to be conclusive for one over the 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), and this was in fact the reason for the introduction of sequential sampling models (Stone, 1960).

It is instructive to see how three parameters in the model affect the speed-accuracy tradeoff: The accumulation rate, or *drift rate*, ν controls the speed of the deterministic information accumulation. Clearly, the greater the absolute value of the drift rate, the more strongly the process is influenced by the deterministic part of (1), hence the more likely it is to exit the correct end of the interval, and the quicker the process reaches a decision. The boundary separation, controled 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 process 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, and it will do so in a shorter amount of time.

The drift rate reflects the relatedness between a probe (stimulus) presented to the subject and an item in her task related memory set—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 the quality of the memory set, and is difficult to control by the subject but strongly controlled by the experimenter. The boundary separation allows the subject to control the conservativeness of her evidence criterion (e.g., in response to task instruction). The starting point allows the decision to be biased towards one of the alternatives, which is a way for the subject to increase the response speed in the case that one alternative is to be expected more likely than another. Empirical validation for each of these interpretations was found by Voss, Rothermund, and Voss (2004).

Because stimuli, their activation in memory, or both may vary across trials, Ratcliff's model generalizes to multiple 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 on the other hand is usually assumed to be uniformly distributed in the range $(z - s_z/2, z + s_z/2)$. Both these distributional assumptions are ad hoc and should be considered as first approximations to the true underlying distributions. Furthermore, the diffusion process only pertains to 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 response time (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 instead to use a normal distribution, which is motivated by computational considerations. [2]

3 Decision time mean and variance

The important merit of the diffusion model for the decision process is not only its credible account of how information accumulates in the brain to trigger a decision, but is also its ability to provide an accurate description of, and explanation for, many phenomena observed in human and animal response times (Ratcliff & Rouder, 1998; Ratcliff, Thapar, & McKoon, 2001; Ratcliff & Smith, 2004; Smith & Ratcliff, 2004; Bogacz et al., 2006; Gold & Shadlen, 2007). One of these phenomena concerns the consistently found (linear) relation between the means of samples of response time observations and their standard deviation. 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 across trial variabilities were assumed to be absent, and the starting point z was assumed to lie equidistant from the two decision boundaries. The latter assumption corresponds to unbiased decisions.

In this section, we find closed form expressions for the central moments in the more general case in which it is not a priori assumed that the decision process is unbiased. As in Wagenmakers et al. (2005), we will still assume however that there is no across trial variability in any of the diffusion parameters (i.e., $s_z = 0$ and $\eta^2 = 0$). We discuss two different cases: In the first case, we

[2] 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.

determine the mean and variance of the time that the process described by equation (1) exits the interval on either side—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 through a particular interval—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 in line with the EZ method.

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

3.1 Moments of exit times irrespective of exit boundary

As most of this case was already discussed in (Wagenmakers et al., 2005), we only briefly summarize the derivation here and quickly turn to the resulting expressions.

The process in equation (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 from 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*. 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 for the process. Let $G(t, z) = Prob(T > t)$ denote the probability that a process that started at z exits the interval after time t. Recall that the process is terminated 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 $T \leq t$, which would contradict our assumption that $T > t$). Hence, if $p(x, t|z, 0)$ is 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 $G(t, 0) = 0 = G(t, a) = 0$, as both boundaries are absorbing (cf., Gardiner, 2004; Wagenmakers et al., 2005). 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 response time terms: These are the first two cumulants of the response times 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. of the responses conditioned on the correctness of the response. Approximate and some exact results for the discrete time random walk counter part 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 proces. As indicated before, 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, here the direction of flow is 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 of a particle that started at z to leave the interval at the lower boundary after time t is

$$g_0(z,t) = -\int_t^\infty j(0,t'|z,0)dt' = \int_t^\infty \Big(-\nu + \frac{s^2}{2}\partial_x\Big)\, p(x,t' \mid z,0)\Big|_{x=0} dt'$$

(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 depends on the starting point z of the particle. The change of total probability that the particle is 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

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

where the second equality result from integration by parts.

On the other hand, using the PDE for g_0 above

$$-\; g_0(z,0) T_n(z,0) \;=\; \nu \;\; \partial_z \int_0^\infty t^n g_0(z,t) dt \;+\; \frac{s^2}{2} \;\; \partial_z^2 \int_0^\infty t^n g_0(z,t) dt$$

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

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

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

$$\pi_0(a)T(a,0) = \pi_0(0)T(0,0) = 0, \quad (11)$$

which result directly from the boundary conditions of the backward Fokker-Planck equation in case of absorbing boundaries (the decision process terminates as soon as it hits one of the boundaries). Following Gardiner (2004, p. 143), clearly $T(0,0) = 0$, as a process starting at the boundary immediately terminates, and $\pi_0(a) = 0$, as the chance 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 $g_0(z,0) = \pi_0(z)$,

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

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 response time 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 $(\nu, z) \mapsto (-\nu, a - z)$.

Unconditional versus conditional cumulants We note a couple of differences between the conditional and unconditional mean and variance. First, both mean and variance of the exit time conditioned on the point of exit converge to an asymptotic value as the starting point approaches the other end. The unconditional mean and variance on the other hand, both become zero when the starting point approaches either end of the interval, which is of course to be expected. A second, perhaps 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 $\nu = 0$. 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 stick to the method-of-moments which is the approach of the EZ method.

But first we give a quick overview other approaches that obtain estimators with statistically more desirable properties, but with the drawback of long computation times. As indicated earlier, there are several situations in which computation time becomes an issue, which include situations in which estimates per subject 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 put forward (Ratcliff & Tuerlinckx, 2002; Voss, Rothermund, & Voss, 2004; Vandekerckhove & Tuerlinckx, 2007; 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 squares 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 necessary for full maximum likelihood estimation. They also have an option to use the method of Brown and Heathcote (2003). Voss et al. (2004), Voss and Voss (2008) propose to minimize the maximum of the Kolmogorov-Smirnov statistics of correct and error response time distributions.

Although ML estimators are in many cases preferred (save for cases that undermine the usual assumptions—some of which that are relevant to response times, are discussed by Cheng & Iles, 1987 and Heathcote & Brown, 2004), 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, and more precise than WLS estimators.

4.2 EZ estimation method

Despite the substantial payoff of the use of the diffusion model in terms of interpretability of the speed and accuracy data, the methodology 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 time these methods require – even on modern computers. The latter becomes especially problematic when a researcher wishes to try different models for complex experimental designs or fit the model on a large group of subjects on an individual bases. For online estimation as required in adaptive experimental paradigms (e.g., if stimulus discriminability is to be equalized across subjects) these methods are impractical.

The EZ method (Wagenmakers et al., 2007) bridges the gap by providing easily computable estimators for the parameters of the diffusion model. These could be obtained by virtue of the analytical invertability 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. The EZ method furthermore ignores variability in parameters across trials. Thus the EZ method sacrifices some aspects of the full diffusion model and consequently has a more modest range of applicability. The simulations presented by Wagenmakers et al. (2007) show however that these method-of-moments estimators perform quite well, even when either of the simplifying assumptions were slightly violated. The method has recently been criticized however (see Ratcliff, in press and Wagenmakers, van der Maas, Dolan, & 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 section 3 can be used to address the starting point problem of EZ. 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, response time means and response time variances) to the expression of their theoretical population values of section 3 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 still simple and fast enough to be computed in a web page script and is straightforwardly implemented in a spreadsheet program. To ease the discussion we appropriately refer to this method as "EZ2".

We consider as an example the common situation where there are two types of trials in which a correct response for one type of trial is an error response for the other and vice versa—a lexical decision task, say. 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 determine in advance what the condition of the next trial is. 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, hypothesize that response times 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 have the following considerations. Firstly, from section 3 we know that in order to be able to estimate the sign of ν we have to include at least one proportion of errors or an unconditional moment. Secondly, to be able to estimate T_{er} we have to use at least one mean response time. In fact, the mean response time is not only the sole moment that provides information about T_{er}, it essentially *only* provides information about T_{er} and *scarcely* provides information about any of the other parameters. This can be seen if one considers the partition $MRT = MDT + T_{\mathrm{er}}$, 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 response time, which is clearly required, T_{er} will absorb any discrepancy between observed and predicted mean response time. Hence the observed mean response time 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 response times for words, a variance for the correct response times for nonwords, a percentage of errors for the words and a percentage of errors for the nonwords, The non-linear 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 non-negative potential function, whose gradient involves the system (e.g., a least squares function) 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.[3] The next section demonstrates the ability of this procedure to produce valid parameter estimates in a number of numerical simulations.

[3] 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.

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.

Setup

We simulated response times under a lexical decision task like setup. The values of the drift rates, boundary separation and starting point, as 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 response times, the percentages of errors, and the response time variances corresponding to these parameter values.

For each combination of parameters, we simulated 100 data sets, with $N = 50, 250$, or 1000 trials, with $N/2$ for each condition.

A problem with few trials is the occurrence of perfect performance. 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

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

Table 1

Parameter values used in the simulation and the corresponding mean response times (MRT), percentages of errors (Pe), and response time variances (VRT) for the correct responses. Units of MRT and VRT in this table were rescaled and rounded to milliseconds. $Ter = 0.25$ in all cases.

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.

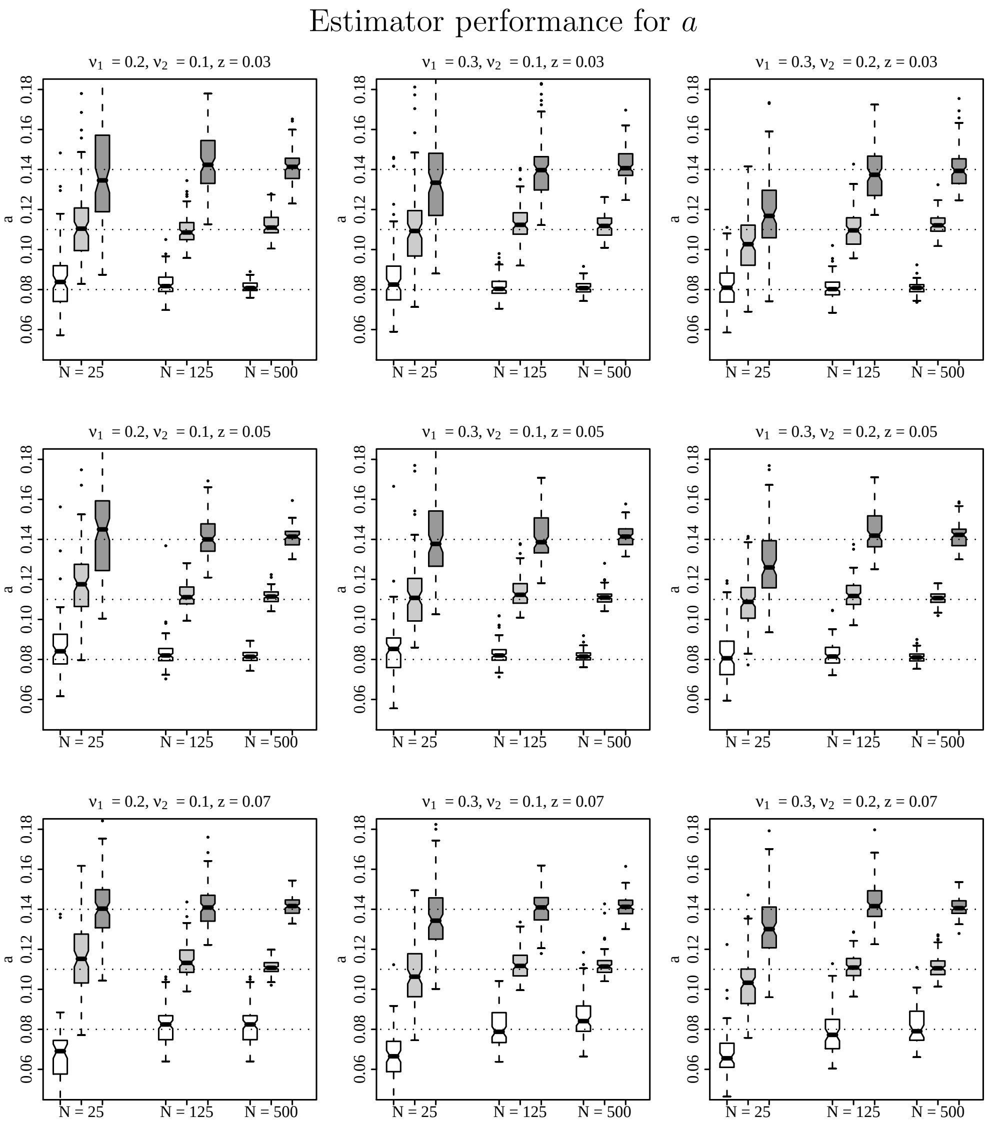

Fig. 2. Box-and-whisker plots for the EZ2 estimates of the boundary seperation a. The dotted line indicate the true values $a = 0.08$ (white boxes), $a = 0.11$ (light gray boxes), and $a = 0.14$ (dark gray boxes).

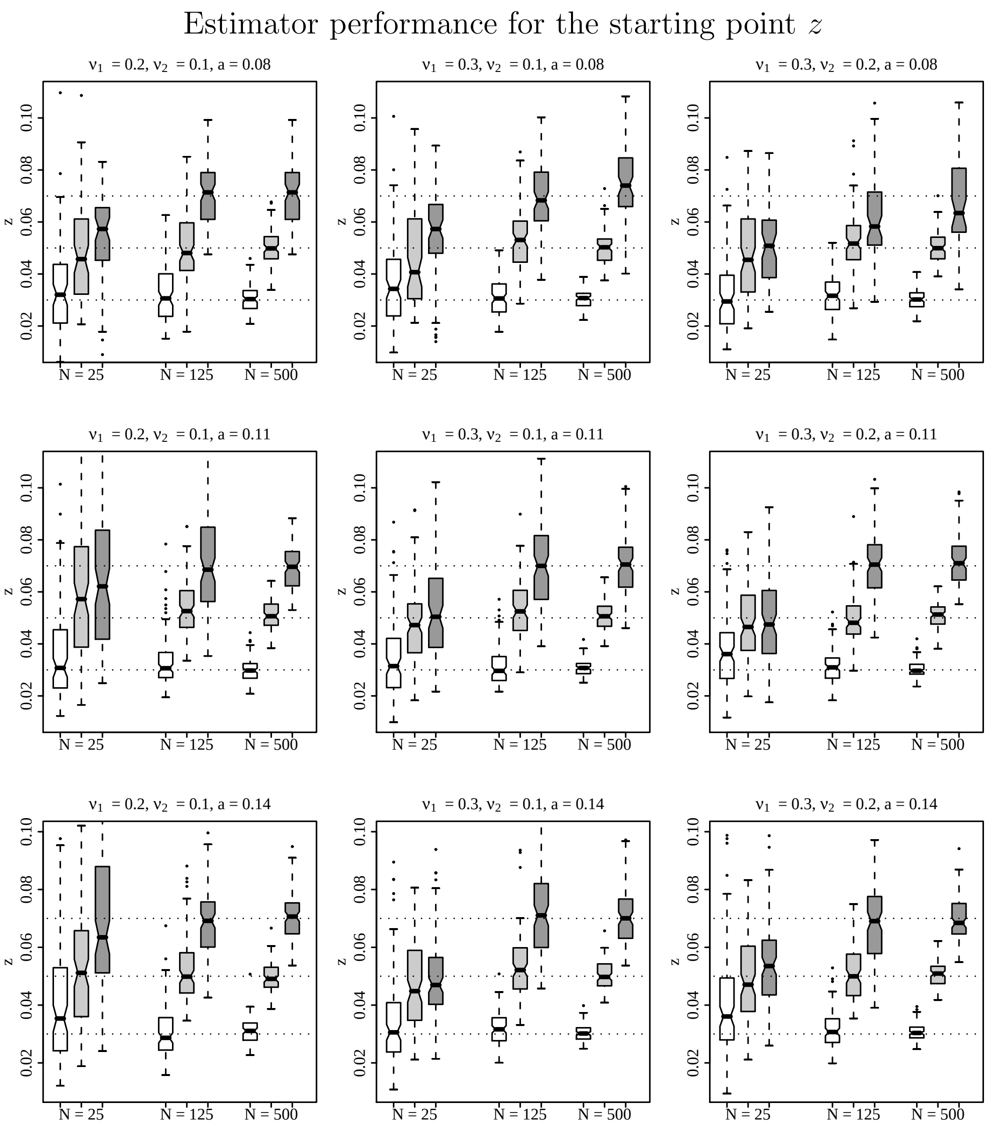

Fig. 3. Box-and-whisker plots for the EZ2 estimates of the parameter z. The dotted line indicate the true values $z = 0.03$ (white boxes), $z = 0.05$ (light gray boxes), and $z = 0.07$ (dark gray boxes).

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 (Hooke & Jeeves, 1961; Kaupe Jr., 1963; Gill, Wright, & Murray, 1986; Seber & Wild, 1989; Press et al., 1993). The algorithms did not differ very much, although the Hooke and Jeeves al-

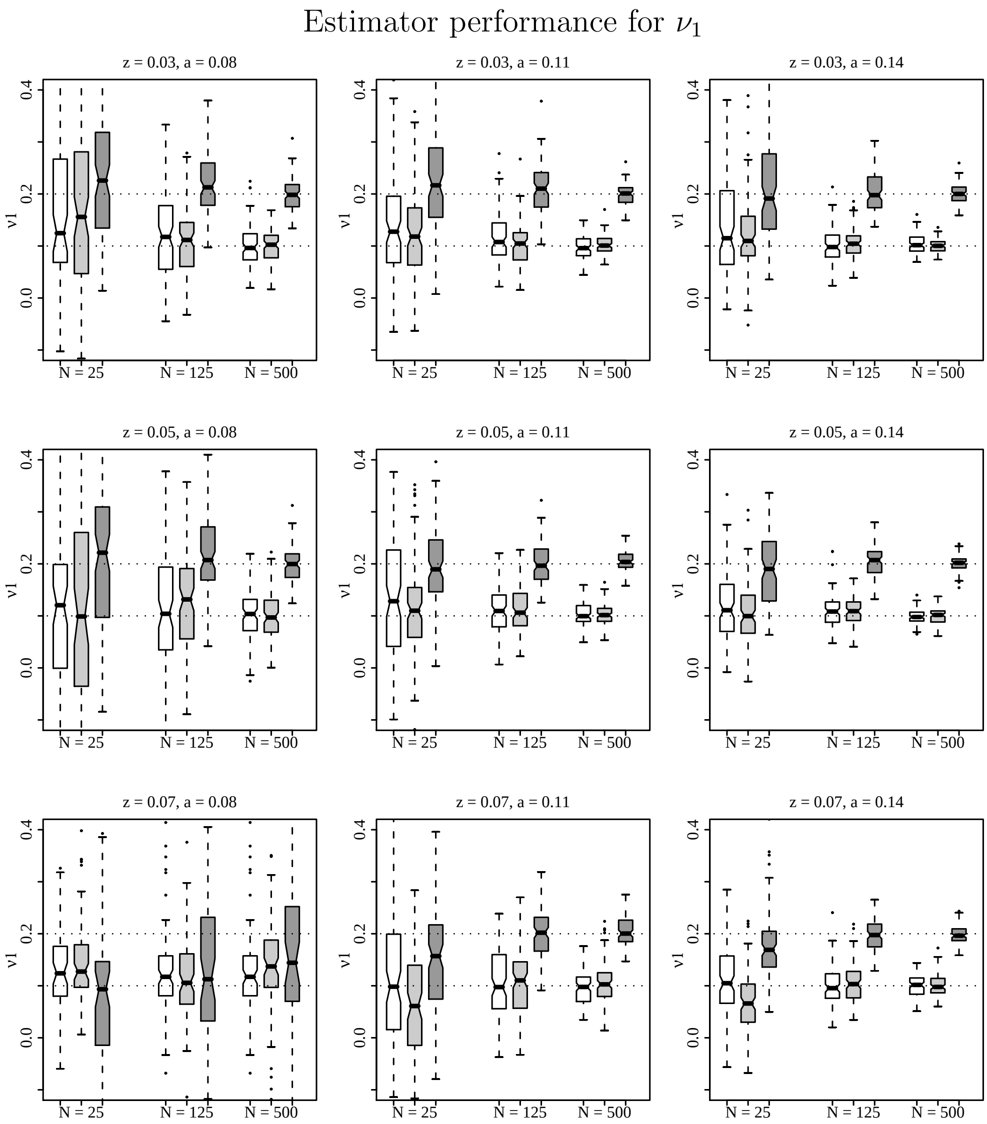

Fig. 4. Box-and-whisker plots for the EZ2 estimates of drift rate ν_1. Dotted horizontal lines indicate true values of ν: $\nu_1 = 0.1$ with $\nu_2 = 0.2$ (white boxes), $\nu_1 = 0.1$ with $\nu_2 = 0.3$ (light gray boxes), and $\nu_1 = 0.2$ with $\nu_2 = 0.3$ (dark gray boxes).

gorithm 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

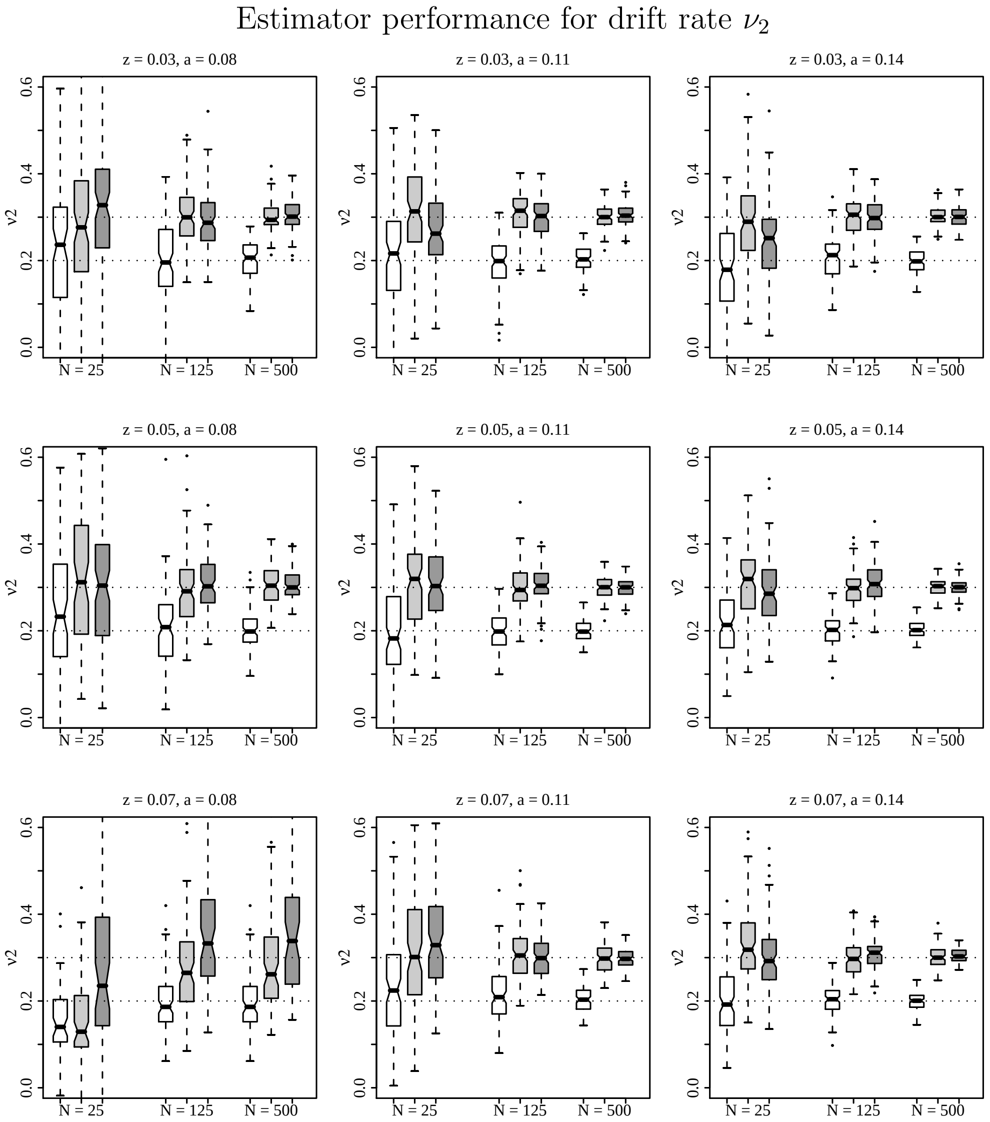

Fig. 5. Box-and-whisker plots for the EZ2 estimates of the parameter ν_2. Dotted horizontal lines indicate true values of ν: $\nu_2 = 0.2$ with $\nu_1 = 0.1$ (white boxes), $\nu_2 = 0.3$ with $\nu_1 = 0.1$ (light gray boxes), and $\nu_2 = 0.3$ with $\nu_1 = 0.2$ (dark gray boxes).

never encountered estimates that violated these constraints [4], thus keeping

[4] 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.

the method simple.

Results

Figures 2 – 5 display the EZ2 results for the parameters a, z, ν_1 and ν_2 respectively in box-and-whisker plots. These estimates where 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 Figures 2–5. We discuss the performance of the parameter estimators in terms of bias below.

Subsequent columns in the three-by-three panel array in Figure 2 indicate that while the boundary separation a is well recovered, as drift rate increases, performance deteriorates unless the number of trials is increased. The distance between z and a also influences the recovery of a, but any of 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 deteriorate 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 Figure 3. The distribution of z estimates is also more symmetrical and narrower if 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 Figures 4 and 5), as well as by the distance of starting point from the boundaries (see bottom row panels of Figures 4 and 5). However as trial numbers increase, the bias quickly vanishes in all cases.

In conclusion, the recovery performance of these method-of-moment estimators seems to be fine, as long as sufficient numbers of trials are collected when drift rates are expected to be large or decision bias is strong. The key factor in parameter recovery performance of this estimator seems to be the proportion of errors that is made: the fewer errors the worse 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 illustration purposes, we apply the EZ2 methods to empirical data. The complete task is described in Wagenmakers, Ratcliff, Gomez, and McKoon (2008); here we only summarize the important features. The response time data were collected from 19 university students who participated in a lexical decision task with 75% nonwords and 25% words, and word frequency was varied from 'very low' to 'low' to 'high'. The word-nonword imbalance presumably biases participants 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 stronger represented in memory and hence their drift rate should be higher. The nonwords consisted of pseudowords 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), since we only mean to illustrate the use of the method, we simply discarded these two cases from the analysis. In Wagenmakers et al. (2008), the data were judged to conform the diffusion model characteristics so that application of the EZ2 method is warranted, although not entirely correct, as across trial variation is ignored. Individual variances (of correct responses only) and percentages of errors of 17 participants were fitted to a model in which the lower boundary corresponded to a word response and the upper boundary to a nonword response. 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 multivariate repeated measures omnibus 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 these could only be attributed to differences between very low and high frequency words ($F(4, 13) = 12.51$, $p = .0008$) and 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 conform the expectations, except perhaps for the lack of the anticipated difference between the word drift rates in the very low

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)

Table 2

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.

word frequency and the low frequency conditions. The latter however may to be due to a lack of power rather than due to an absence of the expected difference. Note furthermore that the drift rate for nonwords is close to the drift rate for very low frequency words [5], which seems quite reasonable theoretically for the pseudo-words used if drift rate is indicative of the quality of the memory representation for the item. In addition, the starting point z is closer to a, the nonword boundary, indicating a clear bias towards nonword responses as 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

[5] 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.

word frequency	ν_0	ν_1	z	a
very low	.172 (.014)	.188 (.025)	.1049 (.0065)	.150 (.0077)
low	.166 (.010)	.259 (.023)	.1068 (.0069)	.148 (.0084)
high	.183 (.011)	.352 (.027)	.0945 (.0054)	.143 (.0077)

Table 3

Parameters estimates from fits to variances of the pooled correct and error responses and error percentages in the lexical decision task. Standard errors as determined from across participant variance are indicated between parentheses. The differences between the values of ν_1 are all significant. Differences between conditions for other parameters are all non-significant.

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 response times variances of pooled error and correct responses instead of using only correct responses hardly affects the estimates and their standard errors [6], except for a slightly diminished 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 previously, except that in this case an additional marginal difference was detected in ν_1 between low frequency words and very low frequency words which is 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 par-

[6] correlations between parameter estimates all $> .9$; for z and a all $> .96$

ticipants. In the fit of the model in that paper, parameters were constraint to fit an additional condition with 75% words and 25% nonwords (the opposite of the data analyzed here). The estimates are tabulated in Table 4 for comparison.

Qualitatively, the difference between estimates is not very large. The estimated drift rates for very low frequency and low frequency are very close, but the full diffusion model drift rate for high frequency words is a bit more sizable than the EZ2 estimate. Also, the full diffusion model drift rate for the nonwords is more sizable than the EZ2 estimate, and the EZ2 estimates of z and a are larger than the full diffusion model estimates. However, the EZ2 nonword drift rates estimates do not seem unreasonable from a theoretical point of view when compared to the very low frequency drift rate estimate. Also, while the EZ2 estimates of z and a are both larger than their full diffusion model counterparts, the ratio between the starting point and the boundary separation estimates, z/a, are similar for both methods: $z/a = .669$ for the full diffusion model estimates, while this ratio is $.699, .722$ and $.661$ for the EZ2 estimates in respectively, the high, low, and very low word frequencies conditions. Furthermore, the EZ2 estimates of a are close to the full diffusion model estimate in the 75% word condition for which it was .13 (not given in the Table). In understanding the differences, it should be kept in mind that Tables 2 and 3 were produced by averaging parameter estimates across participants, whereas Table 4 was produced by deriving the estimates of the full diffusion model from quantiles averages, which may partly explain the observed differences. Other major causes for the differences are the inclusion of data of a 75% words condition in the full diffusion model fit, and the equality restrictions on starting points, boundary separations and the nonwords drift

word frequency	ν_0	ν_1	z	a
very low	.252	.169	.079	.118
low	-	.260	-	-
high	-	.476	-	-

Table 4

Parameter estimates from chi-square fits of the full diffusion model to participant averaged quantiles in two conditions (75% nonwords vs. 25% words and 25% nonwords vs. 75% words). Parameters were constraint across these two conditions. Only the estimates for the 75% nonword condition are reproduced here. A hyphen indicates that the parameter value was constraint to be identical to the one in the row above. Because participant averaged quantiles were used, no sample standard errors were given. No estimate standard errors were calculated.

rates in the latter fit. Note that, unlike the method-of-moments estimation paradigm adopted here and throughout this paper, a least squares estimation framework would be able to address all of these differences. Exploring these possibilities is beyond the scope of the this paper however (see the discussion section for more detailed remarks on this issue).

All in all, the results show that for as far as the EZ2 parameters are concerned, conclusions that may drawn from the averaged EZ2 estimates pretty much confer to the conclusions that may be drawn from a full diffusion model fit to participant averaged data in this example.

4.5.1 Computational Speed

As indicated before the algorithm above is faster than any of the currently available estimation algorithms. For comparison we computed estimates for the data described above again with fast-dm (Voss et al., 2004; Voss & Voss, 2008) under the same model as the EZ2 method estimates in Tables 2 and 3, and registered the computation time. Although it is difficult to compare computation times taken by fast-dm and EZ2 because fast-dm is implemented in C while the implementation of EZ2 we used to estimate this time is in a web page using javascript, the speed difference is quite substantial: To compute Table 2 fast-dm took a total of about 6 minutes while the EZ2 web page implementation took a total of about 6 seconds.[7] It should be mentioned however that fast-dm always fits the model three times from different starting values, while this is not the case for the EZ2 method implementation we used for timing.

5 Discussion

The aim of the present paper was to derive closed form expressions for the first two central moments of response time distributions predicted from Ratcliff's diffusion model under less restrictive assumptions than the ones made in Wagenmakers et al. (2005), and to consider their use for estimation purposes. In particular, we demonstrated how they can be used in a vein similar to the EZ

[7] Timing was done on a 2.33 GHz Intel Mac running Mac OS X Tiger. Fast-dm-29 sources were downloaded from `http://seehuhn.de/pages/fast-dm`. The web page implementation of EZ2 was run in the Safari 3.1 web browser. Note that this is the fastest browser we have tested.

method (Wagenmakers et al., 2007) to obtain method-of-moment estimators.

Although we demonstrated the effectiveness of using the expression in estimation, we do not wish to suggest that this procedure can be considered as a complete substitute for an analysis with the full diffusion model. The assumptions made for deriving the expressions constitute a drastic simplification of the full model. One of these assumptions can be easily repaired (it should be straightforward to include a non-decision time range component s_T), but others are not so easily removed. The method may however be considered valid for a somewhat more coarse level of analysis.

As is true for the EZ method, since both methods provide method-of-moments estimators and the first two central moments are not sufficient statistics for response time distributions, these estimators should be expected to be less precise than for instance maximum likelihood estimators. This is a disadvantage that is to be weighted against the advantage of a substantially reduced computation times. The simulations demonstrate that their sampling errors do not overshadow their usefulness.

Note that the use of the expressions for estimation purposes is not limited to method-of-moment estimation. They can be straightforwardly used in a least squares procedure that fits diffusions used as building blocks to model decisions in different experimental conditions to the observed moments. This is similar to covariance structure modeling as used, e.g., in linear structural relations modeling (e.g., LISREL, see Jöreskog, 1981; Bollen, 1989). The straightforward method is ordinary least squares (OLS) estimation with more equations than unknown parameters. OLS is however, generally dominated by its cousin generalized least squares (GLS) estimation (e.g., Browne, 1974, 1984) or generalized minimum chi-square estimation (Ferguson, 1996, chap. 23), in which squared differences between modeled and observed moments are weighted in accordance with their precision. GLS may result in asymptotically efficient (i.e., maximum likelihood equivalent or best asymptotic normal (BAN)) estimators (Browne, 1974, 1984). Such an approach could be viable in the current case: An estimate of the covariance matrix from which the precision can be calculated can be obtained by bootstrapping the mean and the variance of the response times if only correct responses are used, and the error rate, and mean and variance of the response times if pooled error and correct responses are used.

Recently, Ratcliff and Tuerlinckx (2002) point out the importance of contaminant response times. They showed in their simulations that outliers and contaminant responses in general can have important effects on parameter estimates, and therefore propose to fit a mixture model in which the proportion of contaminants is estimated in addition to the other model parameters. It is not entirely straightforward perhaps to include a 'proportion of contaminants' parameter in the estimation procedure, although not entirely impossible.[8]

[8] One could modify the equations for the variance to $(1-\rho)\,VRT + \rho\,\sigma_c + \rho\,(1-\rho)\,(MRT-\mu_c)^2$, where ρ would indicate the proportion of contaminants, and μ_c and σ_c their mean and variance. This introduces 3 extra parameters and can only be estimated if 3 more equations are available. This can be realized if multiple conditions are analyzed in which these parameters are assumed to be constant. If μ_c and σ_c are functionally dependent, as for instance is the case if a chi-square distribution is assumed for the contaminants for instance, then the number of extra parameters can be reduced by one parameter. Alternatively, as an anonymous reviewer pointed out, if the approach of (Ratcliff & Tuerlinckx, 2002) is used, one could assume that the contaminants have a uniform distribution across the range from the lowest observed

Alternatively, one can try to find more robust estimators of the mean and variance. Such estimators for skewed distribution are available (e.g., Wang & Raftery, 2002). Likewise, the expressions for the decision time variance can be augmented with a variance of the non-decision component that we have ignored all along. It remains to be evaluated if such additions are advantageous.

The currently explored estimation use of the expressions for response time mean and variance thus leaves room for future improvement—both in terms of (relatively straightforward) generalizations to handling contaminants and non-decision time variability, as well as in terms of more complex generalizations to handling a complicated experimental designs with multiple factors, fit assessment and model selection.

References

Apostol, T. M. (1969). *Calculus* (Vol. 2). New York: Wiley.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*, 700–65.

Bollen, K. A. (1989). *Structural equations with latent variables* (1st ed.). New York, USA: Wiley.

Brown, S., & Heathcote, A. (2003). QMLE: fast, robust, and efficient estimation of distribution functions based on quantiles. *Behavior Research Methods, Instruments, & Computers*, *35*, 485–92.

to the highest observed response time. Then, only the proportion of contaminants ρ has to be estimated.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, *8*, 1-24.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83.

Busemeyer, J. R., & Townsend, J. T. (1992). Fundamental derivations from decision field theory. *Mathematical Social Sciences*, *23*, 255–282.

Cheng, R. C. H., & Iles, T. C. (1987). Corrected maximum likelihood in non-regular problems. *Journal of the Royal Statistical Society. Series B (Methodological)*, *1*, 95–101.

Ferguson, T. S. (1996). *A course in large sample theory* (1 ed.). Boca Raton: Chapman & Hall/CRC.

Gardiner, C. W. (2004). *Handbook of stochastic methods* (3 ed.). Berlin: Springer Verlag.

Gill, P. E., Wright, M. H., & Murray, W. (1986). *Nonlinear programming.* Stanford: Stanford University Press.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–74.

Heathcote, A., & Brown, S. (2004). Reply to Speckman and Rouder: A theoretical basis for QML. *Psychonomic Bulletin & Review.*

Hooke, R., & Jeeves, T. A. (1961). Direct search solution of numerical and statistical problems. *Journal of the ACM*, *8*, 212–229.

Jöreskog, K. G. (1981). Analysis of covariance structures. *Scandinavian Journal of Statistics*, *8*, 65-92.

Kaupe Jr., A. F. (1963). Algorithm 178: Direct search. *Communications of the ACM*, *6*, 313.

Kreyszig, E. (1993). *Advanced engineering mathematics.* New York: Wiley.

Luce, R. D. (1986). *Response times.* New York: Oxford University Press.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(5), 376–404.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1993). *Numerical Recipes in C: the art of scientific computing* (2nd ed.). Cambridge: Cambridge University Press.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review*, *92*, 212–225.

Ratcliff, R. (in press). The EZ diffusion method: Too EZ? *Psychonomic Bulletin & Review.*

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95.*

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two–choice reaction time. *Psychological Review*, *111*, 333–367.

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and aging*, *16*, 323–41.

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.

Schwartz, W. (1991). Variance results for random walk models of choice reaction time. *British Journal of Mathematical and Statistical Psychology*, *44*, 251–264.

Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression.* New York: Wiley.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*, 161–8.

Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, *25*, 251–260.

Tuerlinckx, F. (2004). The efficient computation of the distribution function of the diffusion process. *Behavior Research Methods, Instruments, & Computers*, *34*.

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, *14*, 1011–26.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*, 1206–1220.

Voss, A., & Voss, J. (2008). A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology*, *52*, 1–9.

Wagenmakers, E., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, *58*.

Wagenmakers, E.-J. (in press). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology.*

Wagenmakers, E.-J., Grasman, R. P. P. P., & Molenaar, P. C. M. (2005). On the relation between the mean and the variance of a diffusion model response time distribution. *Journal of Mathematical Psychology*, *49*, 195–204.

Wagenmakers, E.-J., van der Maas, H. L. J., Dolan, C. V., & Grasman, R. P. P. P. (in press). EZ does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review*.

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.

Wang, N., & Raftery, A. E. (2002). Nearest-neighbor variance estimation (NNVE): Robust covariance estimation via nearest-neighbor cleaning. *Journal of the American Statistical Association*, *97*, 994–1005.

Wickelgren, W. A. (1977). Speed–accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*, 67–85.

A Second moment of decision time of error responses

In this appendix we derive the second order moment of the exit time of the error responses. We solve equation (10) by variation of parameters (e.g., Apostol, 1969; Kreyszig, 1993). In (10), write $y(z) = \pi_a(z)T_2(z,0)$, the general solution to the homogeneous equation associated with (10) is

$$C_1 + C_2\, e^{-2\nu z/s^2},$$

so that $\{y_1, y_2\} = \{1, e^{-2\nu z/s^2}\}$ is a linear basis for the solutions of the homogeneous problem. The Wronskian of this basis is

$$W = y_1\, \partial_z y_2 - y_2\, \partial_z y_1 = -\frac{2\nu}{s^2}\, e^{-2\,\nu\, z/s^2}$$

A particular solution, y_p, is given by

$$y_p = -y_1 \int \frac{-2\, y_2\, \pi_0\, T_1}{W}\, dz + y_2 \int \frac{-2\, y_1\, \pi_0\, T_1}{W}\, dz$$

Maple evaluates the integrals to unpleasant lengthy expressions involving higher transcendental functions, but tedious derivations show that they can be brought down to expressions involving only exponentials. Denote $\phi(x) = e^{2\nu\, x/s^2}$, the particular solution can be written

$$y_p = \frac{1}{2}\frac{\phi(-z)}{\nu^4\,(\phi(a)-1)^2}\Big((2\,z\,\nu\,s^2 - s^4 - 8\,\phi(a)\,a\,\nu^2\,z - 2\,z^2\,\nu^2)\;\phi(z)$$

$$+ (2\,\nu^2 z^2 + 2\nu z s^2 - 8a\nu^2 z - 4\nu a s^2 + s^4)\;\phi(a) + (-2\,\nu z s^2 - 2\nu^2 z^2 - s^4)\;\phi(2a)$$

$$+ (2\nu^2 z^2 + 4\nu a s^2 + s^4 - 2\nu z s^2)\;\phi(x+a)\Big).$$

The general solution to (10) then is

$$y = y_p + C_1 + C_2\, e^{-2\nu z/s^2}.$$

The coefficients C_1 and C_2 are solved for by imposing the side conditions (11), and are substituted back into the solution. Dividing y by $\pi_a(z)$ yields an expression for the second order moment $T_2(z,0)$ which we do not give here. Instead, we gave the variance in (14), which results from subtracting the square of equation (13).

B Estimation Software

We provide several pointers to software implementations of the estimation procedure of the EZ2 diffusion model:

B.1 Web Application

A web application that can be used directly, can be found at `http://purl.oclc.org/net/rgrasman/jscript/ez2`. The application allows users to specify a model for complex experimental designs involving several EZ2-diffusions with separate and or shared parameters. To this end, the user i) creates a set of parameters, ii) chooses a multiple of the non-linear equations in (5), (6), (13), and (14) (i.e., one equation corresponding to each observed percentage of errors, response time variance and/or mean response time), iii) specifies the corresponding observed values, and iv) indicates on which parameters each equation depends. The user can then specify starting values, or use EZ estimators, and presses the 'solve' button to find the estimates. The application can be used both for finding method-of-moment estimators (the number of non-linear equations equals the number of unknown parameters) or for (ordinary) least squares estimation (the number of nonlinear equations exceeds the number of unknowns). The application also provides batch estimation functionality.

Note that the application is written in Javascript (ECMAScript) and DHTML. Javascript is clearly not intended for heavy numerical computations, yet the application is sufficiently fast to gain some first hands-on experience with modeling this way. Performance speed varies considerably across browsers. Notably Firefox (versions 1.5, 2.0 and 3.0, tested on a Windows XP machine) seems to be a bit slow. Microsoft's Internet Explorer (IE6 & IE7 on Windows platform) is appreciably faster, as are Opera 9.0 and Safari 3.0/3.1 (both for Mac OS X & Windows).

B.2 Estimation in an Excel Sheet

An example Excel sheet, including a tutorial can be found at `http://purl.oclc.org/net/rgrasman/excel/ez2`.

B.3 R Routines

An R packages with non user-friendly R routines, including documentation, can be downloaded from `http://purl.oclc.org/net/rgrasman/r/ez2` .