Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Mathematical Psychology 49 (2005) 195–204
Journal of Mathematical Psychology
www.elsevier.com/locate/jmp

Notes and comments

On the relation between the mean and the variance of a diffusion model response time distribution

Eric-Jan Wagenmakers*, Raoul P.P.P. Grasman, Peter C.M. Molenaar

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

Received 23 April 2004; received in revised form 11 February 2005
Available online 31 March 2005

Abstract

Almost every empirical psychological study finds that the variance of a response time (RT) distribution increases with the mean. Here we present a theoretical analysis of the nature of the relationship between RT mean and RT variance, based on the assumption that a diffusion model (e.g., Ratcliff (1978) *Psychological Review*, *85*, 59–108; Ratcliff (2002). *Psychonomic Bulletin & Review*, *9*, 278–291), adequately captures the shape of empirical RT distributions. We first derive closed-form analytic solutions for the mean and variance of a diffusion model RT distribution. Next, we study how systematic differences in two important diffusion model parameters simultaneously affect the mean and the variance of the diffusion model RT distribution. Within the range of plausible values for the drift rate parameter, the relation between RT mean and RT standard deviation is approximately linear. Manipulation of the boundary separation parameter also leads to an approximately linear relation between RT mean and RT standard deviation, but only for low values of the drift rate parameter.

1. Introduction

Two popular dependent measures in psychological research are response accuracy (i.e., proportion of items responded to correctly) and response time (RT; i.e., time from stimulus onset until response execution). For response accuracy of a random variable X, the binomial model with success parameter p and number of observations n allows for a simple and formal description of the mean, $E(X) = np$, and its variance, $\mathrm{var}(X) = np(1-p)$. It follows that the binomial variance decreases as p (and hence $E(X)$) becomes more extreme, that is, as p gets closer to either zero or one. The fact that the binomial variance depends on the binomial mean is a violation of the "homogeneity of variance" assumption of a traditional analysis of variance (ANOVA), and this has inspired the development of variance-stabilizing transformations such as the arcsine transform (i.e., $\tilde{p} = \arcsin(\sqrt{p})$, Snedecor & Cochran, 1989, p. 289) and motivated application of alternative statistical procedures such as logistic regression (e.g., Pampel, 2000).

For RT, it has often been observed that here too the variance fluctuates with the mean. Specifically, an increase in RT mean is almost always accompanied by an increase in RT variance. The precise nature of the relationship between RT mean and RT variance is of interest for several reasons. First, knowledge of this relation may guide the search for an appropriate variance stabilizing transformation (cf. Levine & Dunlap, 1983, p. 597; Snedecor & Cochran, 1989, pp. 286–287). For instance, when the variance is proportional to the mean, such as for Poisson distributed data, the square root transformation is appropriate, whereas the logarithmic transformation is advisable when the standard deviation is proportional to the mean. A variance stabilizing transformation often also diminishes skew, further reducing the number of ANOVA violations exhibited by RT data (cf. Emerson & Stoto, 2000; Keselman, Othman, Wilcox, & Fradette, 2004).

Second, several statistical techniques assume a specific relation between mean and variance. For instance, in the

*Corresponding author. Fax: +31 20 639 0279.
E-mail address: ewagenmakers@fmg.uva.nl (E.-J. Wagenmakers).

0022-2496/$ - see front matter
doi:10.1016/j.jmp.2005.02.003

aging literature it has recently been proposed that effects of aging primarily express themselves in RT variability rather than in RT mean (e.g., Hultsch, MacDonald, & Dixon, 2002; Li, 2002; MacDonald, Hultsch, & Dixon, 2003; Shammi, Bosman, & Stuss, 1998). In order to study differences in RT variability while controlling for possible differences in RT means, researchers sometimes use a linear regression technique whose aim is to partial out effects of differences in RT mean on the observed differences in RT standard deviation. Alternatively, the coefficient of variation (i.e., $\sqrt{\text{var}(X)}/E(X)$) is sometimes used to control for baseline differences in processing speed (e.g., Segalowitz & Segalowitz, 1993). Both methods tacitly assume a linear relationship between RT mean and RT standard deviation, and their efficiency will depend on the extent to which this assumption is correct.

Third, the claim is sometimes made that RT distributions are log-normally distributed (e.g., Van Orden, Pennington, & Stone, 2001, p. 147); this is an important claim in the context of nonlinear dynamical systems theory, as the log-normal distribution can closely mimic a self-similar power law (cf. Peterson & Leckman, 1998, p. 1346; Sornette, 2000, pp. 80–82). When RTs are log-normally distributed, the standard deviation is proportional to the mean. Log-normally distributed data may require a different method of analysis than do normally distributed data (e.g., Zhou, Gao, & Hui, 1997).

Finally, we believe the relation between RT mean and RT variance is important because of general theoretical considerations. The field of mathematical psychology has invested considerable effort in the study of RT distributions (cf. Van Zandt, 2000, 2002), and the relation between RT mean and RT variance is a fundamental property of a family of RT distributions.

In order to study the precise relationship between RT mean and RT variance, we ideally need a mathematical model that is generally acknowledged to provide a close fit to a wide range of empirical RT distributions, much like the binomial model is an accepted statistical model for response accuracy. Such a mathematical model allows any conclusions to be very general and unaffected by measurement noise. The model that is the focus of our work here is the continuous random walk or diffusion model (e.g., Diederich & Busemeyer, 2003; Ratcliff, 1978, 2002; Smith, 2000).

Our choice for the diffusion model as an RT counterpart to the binomial model for accuracy was motivated by mathematical tractability, by the fact that the diffusion model often performs better than competitor sequential sampling models (cf. Ratcliff & Smith, 2004), and—most important—by the fact that the diffusion model has been successfully applied to a wide range of two-choice tasks. The different paradigms to which the diffusion model has been applied include short- and long-term recognition memory tasks, same/different letter-string matching, numerosity judgments, visual-scanning tasks, brightness discrimination, letter discrimination, and lexical decision (e.g., Ratcliff, 1978, 1981, 2002; Ratcliff & Rouder, 1998, 2000; Ratcliff, Van Zandt, & McKoon, 1999; Ratcliff, Gomez, & McKoon, 2004). In all these applications, the diffusion model provided a close fit to the observed RT distributions. In addition, the above applications provide a range of plausible parameter values (i.e., the so-called practical distribution, Raftery & Zheng, 2003) that can be used in the formal study of the RT mean–variance relationship.

The outline of this article is as follows. The next section briefly describes the diffusion model used here. We then derive closed-form analytical expressions for the mean and variance of the diffusion model RT distribution. Next, these expressions are used to illustrate the mean–variance relation for plausible ranges of diffusion model parameter values.

2. Brief outline of a diffusion model for response times

The diffusion model is a continuous-time random walk sequential sampling model (for similar models see Brown & Heathcote, 2005; Link, 1992; Link & Heath, 1975; Laming, 1968). The theoretical properties of the diffusion model are well known (e.g., Luce, 1986; Ratcliff, 2002; Ratcliff & Smith, 2004; Townsend & Ashby, 1983; for a mathematical treatment see for instance Gardiner, 2004; Honerkamp, 1994) and a range of different methods for fitting the model to data is available (Diederich & Busemeyer, 2003; Ratcliff & Tuerlinckx, 2002; Smith, 2000).

In a diffusion model, illustrated in Fig. 1, noisy accumulation of information drives a decision process

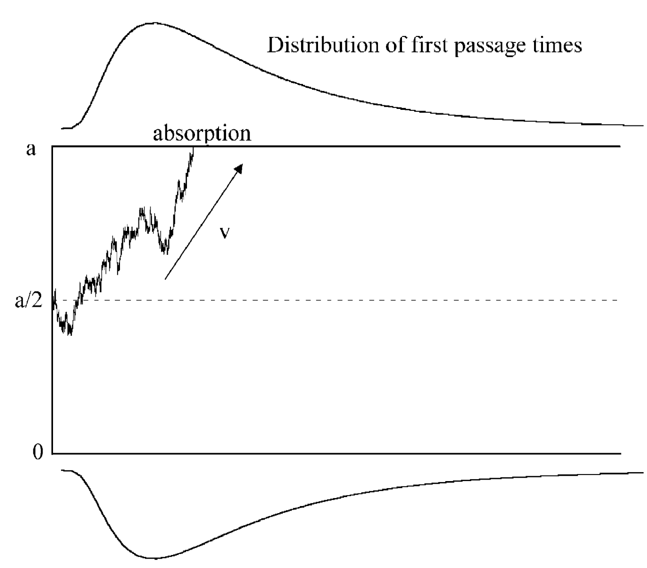

Fig. 1. A diffusion model for two-choice RT and its parameters. See text for details.

that terminates when the accumulated evidence in favor of one or the other response alternative exceeds threshold (i.e., a relative rather than absolute response criterion). The diffusion model has several key parameters: (1) drift rate v, $-\infty < v < \infty$, which quantifies the deterministic component of the continuous-time random walk process. For high absolute values of v (e.g., high-frequency words in a lexical decision task, Ratcliff et al., 2004), processing will terminate relatively quickly at one of the absorbing response boundaries;[1] in applications of the diffusion model to real data, v usually ranges from 0.1 to 0.5; (2) s^2, the variance of the diffusion function, which quantifies the random component of the continuous-time random walk process. This parameter is usually treated as a scaling parameter and set to a default value of 0.01;[2] (3) boundary separation a and starting point $z = \frac{1}{2}a$. In many applications of the diffusion model, the decision process is not very biased against one or the other response alternatives—consequently, the starting point is about equidistant from the response boundaries. As an added bonus, with the starting point in the middle there is no need to complicate the presentation of the results by conditioning on the specific boundary that was reached first: when $z = \frac{1}{2}a$, the RT distributions that terminate at the top and bottom boundaries are identical, irrespective of drift rate v (e.g., Laming, 1973, p. 192, footnote 7; Link & Heath, 1975; Smith & Vickers, 1988; Tuerlinckx, Maris, Ratcliff, & De Boeck, 2001).[3] Large values of a indicate the presence of a conservative response criterion: the system requires relatively much discriminative information before deciding on one or the other response alternative. A conservative response criterion results in long RTs, but also in highly accurate performance, since with large a it is unlikely that the incorrect boundary will be reached by chance fluctuations. Therefore, manipulation of boundary separation a provides a natural mechanism to model the speed-accuracy trade-off (e.g., Wickelgren, 1977). In practical applications, a generally ranges from 0.07 to 0.17.

Before proceeding, we would like to point out that the diffusion model outlined above is a simplified version of the model that is used to fit empirical data. When the diffusion model is fitted to empirical data, several additional free parameters come into play, such as across-trial variability in drift rate and starting point, and an additive amount of time allotted to the non-decisional component of processing (which may also vary from trial-to-trial). As the aim of this paper is not to fit empirical data, but rather to determine the general mean–variance relationship of a diffusion process, we prefer the model in its simplest form.

In addition, it should be noted that many of the additional free parameters that enter the diffusion model when it is fitted to data either do not affect the mean–variance relationship (i.e., the mean of the non-decisional processing time which simply adds to the decisional processing time), or make it more difficult to detect the underlying relationship because these free parameters add noise (i.e., across-trial variability in drift rate and starting point), somewhat comparable to overdispersion for a binomial model. As we will demonstrate below, the version of the diffusion model described here allows a closed-form mathematical solution for RT mean and RT variance, and this greatly enhances mathematical tractability and conceptual clarity.

In sum, the diffusion model is a popular sequential sampling model. Two important parameters of the model are drift rate v and boundary separation a. Fig. 2 illustrates how these parameters affect both RT mean and RT variance. Specifically, decreasing drift rate will lead to an increase in RT mean and an increase in RT variance; increasing boundary separation will also lead to an increase in RT mean and an increase in RT variance. Before studying the precise nature of the mean–variance relationship we will first discuss three procedures to obtain the mean and variance of the diffusion process outlined above.

3. Mean and variance of a diffusion model RT distribution

Several methods are available to determine the mean and variance of a diffusion model RT distribution, and three of these are described in detail below. The method of brute force simulation requires a substantial amount of iterations to converge. The method based on integrating the probability density function (pdf) will yield results that convergence in a much shorter amount of computer time. Nevertheless, integration of the diffusion pdf involves an infinite sum inside an infinite integral, which is still computationally demanding and does not provide much conceptual insight. Also, when the decision to stop the evaluation of the infinite integral and the infinite sum is premature, the solution will of course be incorrect. The third method is to analytically

[1] In most of Ratcliff's work on the diffusion model, ξ is the drift rate of an individual trial, whereas v is the mean drift rate associated with an across-trial distribution of drift rates. In this article, we ignore across-trial variability in drift rate, which implies that $v = \xi$.

[2] Several equations simplify when $s^2 = 1$ is used. We chose to use $s^2 = 0.01$ for historical reasons. Moreover, changing s^2 would also change the ranges of plausible parameter values obtained from earlier applications of the model that all use $s^2 = 0.01$.

[3] This identity no longer holds when across-trial variability in drift rate or starting point is included in the model (Ratcliff & Rouder, 1998). For the purpose of this paper (i.e., to study the relation between RT mean and RT variance in the diffusion model) these possible sources of variability have been ignored (see also below). Note that Link and Heath (1975) provide equations for mean RT in the more general case of $z \neq \frac{1}{2}a$.

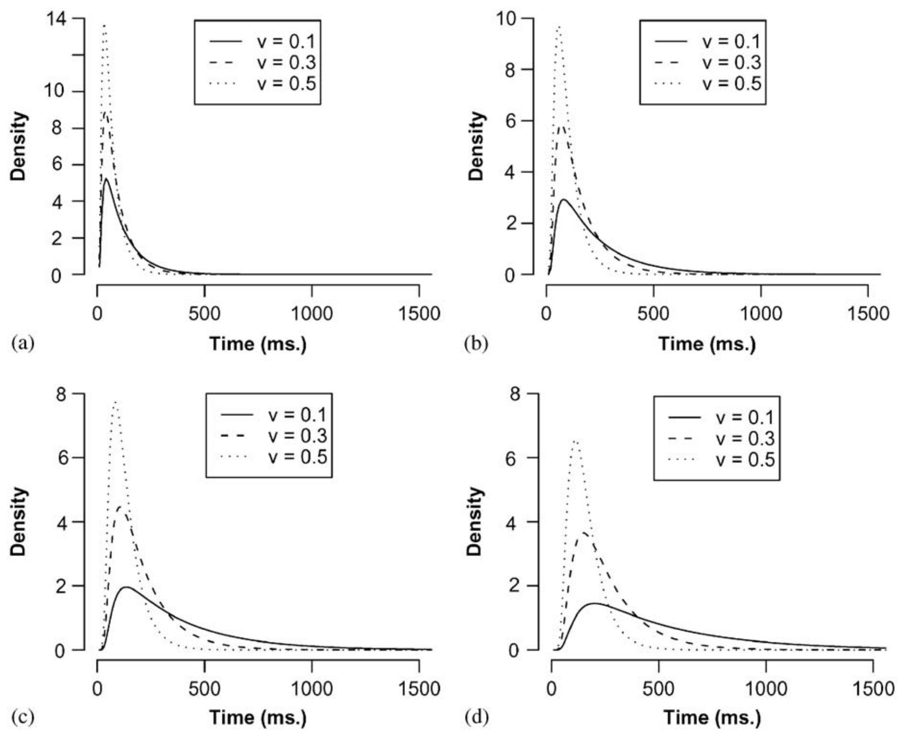

Fig. 2. Diffusion model probability density functions for several combinations of parameter values of drift rate v and boundary separation a. Note the different scaling on the y-axis.

derive closed-form solutions for the mean and variance of a diffusion model RT distribution, based on the backward Fokker–Planck equation. This derivation produces two relatively simple formulas. The reader may easily compare the properties of the various methods by using a free-ware software program written in the statistical computation environment *R*, available at the first author's internet site.[4] The reader who is mainly interested in the results may safely skip to the next section.

3.1. The method of brute force simulation

The method of brute force simulation (for details see Feller, 1968, Chap. 14; Ratcliff & Tuerlinckx, 2002, pp. 441–442; for a review of four different brute force methods see Tuerlinckx et al., 2001) is usually employed whenever it is necessary to obtain quantities from the diffusion model that are not available from more efficient analytical procedures or whenever diffusion model data are needed to check the adequacy of specific model fitting procedures. The brute force procedure is based on the approximation of the continuous-time random walk process by a discrete-time, small-step process. Let the step size in time be given by h, and let the step size in space be given by $\delta = s\sqrt{h}$. It can be shown that if the probability of taking a step of size δ down is given by $P(down) = \frac{1}{2}[1 - (v\sqrt{h}/s)]$, the diffusion process can be simulated by using a random number generator and taking a small step down when the randomly drawn variable $X \sim Uniform(0, 1)$ is smaller than $P(down)$, and taking a step up otherwise.[5] After the process has reached a boundary, the RT for a single trial is calculated as nh, where n is the number of steps taken. Obviously, this is a time consuming process. Moreover, we found that even 10,000 trials are not always sufficient to get a stable estimate for the variance.

3.2. The method of integrating the probability density function

One of the standard methods to obtain the mean and the variance for a given model is by integrating over the pdf and calculating expected values. The pdf of first-passage times for a diffusion process in which the starting point in equidistant from the response boundaries is given by (e.g., Feller, 1968;

[4]The open source statistical computation environment R (R Development Core Team, 2004) is available free of charge from http://cran.r-project.org/. The software program that performs the calculations is available at http://www.psych.nwu.edu/~ej/diffvar.R, and the accompanying help file is available at http://www.psych.nwu.edu/~ej/help.txt

[5]Ratcliff and Tuerlinckx (2002, p. 442) give the probability of a step down as $P(down) = \frac{1}{2}[1 - (v\sqrt{h/s})]$—a typographical error.

Ratcliff & Smith, 2004)

$$g(t) = \frac{\pi s^2}{a^2} \exp\left(\frac{va}{2s^2} - \frac{v^2 t}{2s^2}\right) \times \sum_{k=1}^{\infty}\left[k \exp\left(\frac{-k^2\pi^2 s^2 t}{2a^2}\right)\sin\left(\tfrac{1}{2}k\pi\right)\right], \quad (1)$$

which can be further simplified to (Tuerlinckx et al., 2001):

$$g(t) = \frac{\pi s^2}{a^2} \exp\left(\frac{va}{2s^2} - \frac{v^2 t}{2s^2}\right) \times \sum_{n=0}^{\infty}\left[(2n+1)(-1)^n \exp\left(\frac{-(2n+1)^2\pi^2 s^2 t}{2a^2}\right)\right]. \quad (2)$$

The mean, $E(T)$, can then be calculated as $E(T) = \int_{t=0}^{\infty} t \times g(t)\, dt$, and the variance, $\mathrm{var}(T)$, can be calculated as

$$\begin{aligned}\mathrm{var}(T) &= E([T - E(T)]^2) = E(T^2) - [E(T)]^2 \\ &= \int_{t=0}^{\infty} t^2 \times g(t)\, dt - \left[\int_{t=0}^{\infty} t \times g(t)\, dt\right]^2, \quad (3)\end{aligned}$$

where E denotes statistical expectation. The application of this procedure presents two challenges. First, the pdf contains an infinite sum over k. Ratcliff and Tuerlinckx (2002, p. 478) recommend to truncate this sum at the point where the series contains two consecutive values that are both less than some tolerance value, say 10^{-29} times the current sum of the series. Second, the integral over t also needs to be truncated at some point. One possibility is to again terminate the integral (or the sum, since we approximate the integral by discrete small steps) at the point where the values add less than some tolerance value. Other solutions to this problem exist (i.e., explicitly solving the integral over t or transforming the interval of integration) but these will not be explored here.

3.3. The closed-form solutions

The most satisfying solution is to derive closed form expressions for RT mean and RT variance.[6] In mathematical terms, our aim is to derive the moments of the first passage time distribution for a homogeneous (i.e., drift rate and diffusion variance are independent) diffusion process with two absorbing boundaries. These moments may be obtained using the adjoint or "backward" Fokker–Planck equation, and the general method of solution is described for instance in Gardiner (2004, pp. 136–138) and Honerkamp (1994, pp. 279–285); see also (Karlin & Taylor, 1981, p. 197). For the Ornstein–Uhlenbeck process, Busemeyer and Townsend (1992, p. 271) present equations for the raw moments that are derived on the same basis.

To briefly reiterate the notation, let the drift rate be v, the diffusion variance be s^2, let the bottom boundary be located at 0 and the top boundary be located at a, and let the starting point be z, $z \in [0, a]$. Let $p(x, t|z, 0)$ denote the time-dependent pdf of the continuous random walk occupying position x at time t, given that the process started at z. The walk ends as soon as it hits one of the absorbing boundaries. Thus, the probability that the position x of the walker is in the interval [0, a] after absorption is zero. Consequently, the probability that the process never left the [0, a] interval before a certain time t is given by $G(t|z) = \int_0^a p(x, t|z, 0)\, dx$. At the same time, if T is the time of absorption, that is, the time at which the random walk reaches one of the two boundaries, then also $\Pr(T \geqslant t) = \int_0^a p(x, t|z, 0)\, dx$. Consequently, $G(t|z) = \Pr(T \geqslant t)$, and hence $F_T(t) = 1 - G(t|z) = \Pr(T < t)$ gives the distribution of T. The moments of T as a function of the starting point z are given by

$$\begin{aligned}M_n(z) = E(T^n|z) &= \int_{-\infty}^{\infty} t^n f_T(t)\, dt \\ &= -\int_0^{\infty} t^n \partial_t G(t|z)\, dt, \quad (4)\end{aligned}$$

where $f_T(t) = \frac{d}{dt}F_T(t) = -\partial_t G(t|z)$ is the density function. Integrating the right-hand side of Eq. (4) by parts yields $M_n(z) = \left[n \int_0^{\infty} t^{n-1} G(t|z)\, dt\right] - [t^n G(t|z)|_0^{\infty}]$. If $-t^n G(t|z) \to 0$ as $t \to \infty$, it follows that the second term vanishes. It can be seen as follows that this is in fact the case. Let $u = 1/t$, and write $\lim_{t\to\infty} t^n G(t|z) = \lim_{u\to 0} G(1/u|z)/u^n$. Because as $t \to \infty$ the process will terminate with probability one, $G(t|z)$ as well as all of its first n derivatives approach zero, and hence one may apply l'Hôpital's rule n times to show that this limit is in fact zero.[7] Therefore,

$$M_n(z) = n \int_0^{\infty} t^{n-1} G(t|z)\, dt. \quad (5)$$

The pdf of the diffusion process under investigation, $p(x, t|z, 0)$, is governed by the backward Fokker–Planck equation (Gardiner, 2004):

$$\partial_t p(x, t|z, 0) = v\partial_z p(x, t|z, 0) + \tfrac{1}{2}s^2\partial_z^2 p(x, t|z, 0). \quad (6)$$

Integrating both sides: $\partial_t \int_0^a p(x, t|z, 0)\, dx = v\partial_z \int_0^a p(x, t|z, 0)\, dx + \frac{1}{2}s^2\partial_z^2 \int_0^a p(x, t|z, 0)\, dx$.[8]

[6] A Maple spreadsheet that shows the derivation is available at http://www.psych.nwu.edu/~ej/meanvarderivation.mws

[7] We thank Richard Chechile for providing this argument.

[8] Exchanging the order of integration and differentiation is permitted because the following three conditions hold (e.g., Amemiya, 1985, Theorem 1.3.2): First, p is continuous in both t and x, as their derivatives exist by Eq. (6). Second, the integral $\int_0^a p(x, t|z, 0)\, dx = P(T \geqslant t)$ is bounded because it represents a probability. Third, $\int_0^a |\partial_t p(x, t|z, 0)|\, dx$ is bounded because otherwise Eq. (6) has no solution in the interval from 0 to a.

Recalling that $G(t|z) = \int_0^a p(x, t|z, 0)\, dx$ then yields the governing equation for G:

$$\partial_t G(t|z) = v\partial_z G(t|z) + \tfrac{1}{2}s^2 \partial_z^2 G(t|z). \tag{7}$$

Note that G satisfies the boundary conditions:

$$G(0|z) = 1 \quad 0 \leqslant z \leqslant a,$$
$$G(0|z) = 0 \quad z \text{ elsewhere;}$$

which state that the time until absorption is greater than zero when the starting point is located in between the two boundaries. Another boundary condition is $G(t|0) = G(t|a) = 0$, which states that if the process starts at one of the absorbing boundaries, it is absorbed immediately.

After multiplying both sides of Eq. (7) by t^{n-1}, and integrating t from 0 to ∞, we are left with $\int_0^\infty t^{n-1}\partial_t G(t|z)\, dt = v\partial_z \int_0^\infty t^{n-1} G(t|z)\, dt + \frac{1}{2}s^2\partial_z^2 \int_0^\infty t^{n-1} G(t|z)\, dt$. From the definition of moments it follows that $\int_0^\infty t^{n-1}\partial_t G(t|z)\, dt = -M_{n-1}(z)$, and from Eq. (5) it follows that $\int_0^\infty t^{n-1} G(t|z)\, dt = M_n(z)/n$. Substituting these identities, and multiplying both sides of the equation by n gives the recursion equation

$$-nM_{n-1}(z) = v\partial_z M_n(z) + \tfrac{1}{2}s^2\partial_z^2 M_n(z), \tag{8}$$

which holds for all existing moments of T (e.g., Gardiner, 2004, p. 138). In particular, for $n = 1$ this yields

$$v\partial_z M_1(z) + \tfrac{1}{2}s^2\partial_z^2 M_1(z) = -1, \tag{9}$$

as $M_0 = 1$. All moments are subject to the boundary conditions mentioned above, that is, $M_n(0) = M_n(a) = 0$, $n = 1, 2, \ldots$. Solving Eq. (9) for $M_1(z)$ and evaluation in the symmetric starting point $z = \frac{1}{2}a$ results in the mean absorption time

$$E(T) = M_1\left(\frac{1}{2}a\right) = \left[\frac{a}{2v}\right]\frac{1-\exp(y)}{1+\exp(y)}, \tag{10}$$

where $y = -va/s^2$. In the limit when drift rate v goes to zero, $v \to 0$, the mean absorption time is given by $E(T) = \frac{a^2}{4s^2}$. Solving the second order moment $M_2(\frac{1}{2}a)$ from the recursion equation and using the equality $\text{var}(T) = E(T^2) - [E(T)]^2$ one obtains

$$\begin{aligned}\text{var}(T) &= M_2\left(\frac{1}{2}a\right) - \left[M_1\left(\frac{1}{2}a\right)\right]^2 \\ &= \left[\frac{a}{2v}\right]\left[\frac{s^2}{v^2}\right]\frac{2y\,\exp(y) - \exp(2y) + 1}{(\exp(y)+1)^2}.\end{aligned} \tag{11}$$

In the limit of $v \to 0$, $\text{var}(T) = \frac{a^4}{24s^4}$.

From Eqs. (10) and (11) it follows that a closed-form expression for the relation between mean and variance is given by

$$\text{var}(T) = \begin{cases} E(T)\left[\dfrac{s^2}{v^2}\right]\dfrac{\exp(2y) - 2y\exp(y) - 1}{\exp(2y) - 1} & \text{if } v \neq 0, \\ E(T)\dfrac{a^2}{6s^2} & \text{if } v = 0. \end{cases} \tag{12}$$

To the best of our knowledge, Eqs. (11) and (12) have not yet been reported in the psychological literature on RT modeling. The standard literature on stochastic differential equations (e.g., Gardiner, 2004; Honerkamp, 1994) also does not mention these equations, although they may be easily derived.

Note that Eqs. (11) and (12) are not conditional on whether responses are correct or in error. For a diffusion model with starting point equidistant from the response boundaries and with no variability across trials in drift-rate, the distribution of T is the same for correct and incorrect responses (e.g., Laming, 1973, p. 192, footnote 7).

4. Relation between diffusion model mean and variance as a function of drift rate

As mentioned earlier, drift rate v, $-\infty < v < \infty$, quantifies the deterministic component of the decision process, and in practical applications v usually ranges from absolute values of 0.1 to about 0.5 (e.g., Ratcliff, Thapar, & McKoon, 2003; Ratcliff et al., 2004). Setting the diffusion variance s^2 to its default value of 0.01, the left panel of Fig. 3 plots the relation between mean and variance when drift rate is systematically decreased from 0.5 to 0.1. Each of the six lines corresponds to a different plausible value of the boundary separation parameter a. The left panel of Fig. 3 clearly shows that the relation between RT mean and RT variance resulting from a change in drift rate is highly nonlinear, as the variance increases faster than the mean. In contrast, the relation between RT mean and RT standard deviation, shown in the right panel of Fig. 3, is very close to linear across the entire parameter space of plausible drift rate values. Moreover, this approximate linearity holds for the entire range of plausible boundary separation values. Thus, a decrease in drift rate makes the standard deviation increase approximately linearly with the mean.

The above result implies that effects of processing speed on response variability may perhaps best be discounted by using the coefficient of variability, $\sqrt{\text{var}(X)}/E(X)$. That is, if participants or experimental conditions only differ in processing speed (i.e., drift rate), their CVs should be almost identical. A further consequence of the almost perfect linearity between RT mean and RT standard deviation as a function of drift

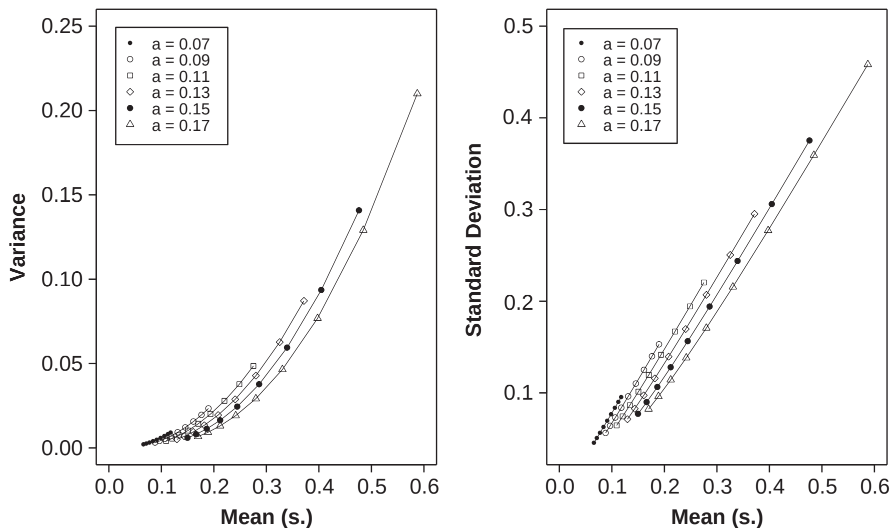

Fig. 3. The relation between RT mean and RT variance (left panel) and RT mean and RT standard deviation (right panel) for 6 levels of boundary separation a, and 9 values for drift rate v. The parameter ranges for a and v (i.e., $a \in [0.07, 0.17]$; $v \in [0.1, 0.5]$) were motivated by previous applications of the model.

rate differences is that the log transform gains credence as a suitable variance-stabilizing transformation.

Finally, it is interesting that the literature on automaticity and learning has shown that practice decreases the mean and the standard deviation at approximately the same rate (e.g., Logan, 1988, 1992; Kramer, Strayer, & Buckley, 1990)—a phenomenon predicted by Logan's instance theory (e.g., Logan, 1988, 1992). Cohen, Dunbar, and McClelland (1990, pp. 345–346) developed a neural network for which simulations showed that practice decreases RT mean and RT standard deviation at approximately the same rate. In their model, output from a neural net drives a random walk decision process. The right panel from Fig. 3 confirms that if automaticity or learning selectively affects the drift rate parameter of a diffusion process, an approximate linear relation between RT mean and RT standard deviation will result.

5. Relation between diffusion model mean and variance as a function of boundary separation

In practical applications, the parameter for boundary separation usually varies from 0.07, a very risky progressive response criterion, to 0.17, which is a very safe conservative response criterion (e.g., Ratcliff et al., 2003, 2004). Similar to our illustration of the effect of drift rate differences, we now examine the effect of an increase in boundary separation on the mean–variance relationship.

The left panel of Fig. 4 plots five lines, one for each of five plausible values of the drift rate parameter. Each line separately is constructed by calculating both mean and variance for a range of different plausible values for boundary separation. Fig. 4, left panel, shows that the relation between mean and variance is approximately linear for high values of drift rate, but becomes strongly nonlinear for low values of drift rate, such that the variance increases as a faster rate than the mean.

The right panel of Fig. 4 plots the mean against the standard deviation. For low values of the drift rate parameter, the mean varies approximately linearly with the standard deviation. For high values of the drift rate parameter, however, the relation is more curvilinear. Thus, it turns out that the relation between mean and standard deviation is more complex for a difference in boundary separation than it is for a difference in drift rate: the relation between mean and standard deviation that results from a difference in boundary separation is conditional on the value of the drift rate parameter, whereas the effects of differences in drift rate are qualitatively unaffected by the specific values of the boundary separation parameter (cf. Fig. 3, right panel).

6. Summary and conclusion

In this article, we studied the relation between the mean and the variance of a diffusion model RT distribution.[9] We used a simple diffusion model (i.e., starting point always equidistant from the response boundaries, no across-trial variability in drift rate or starting point) to obtain closed-form expressions for RT

[9]The focus of this article has been on response time, and not on response accuracy. Equations for diffusion model response accuracy can be found, for instance, in Ratcliff (1978) and in Link (1992).

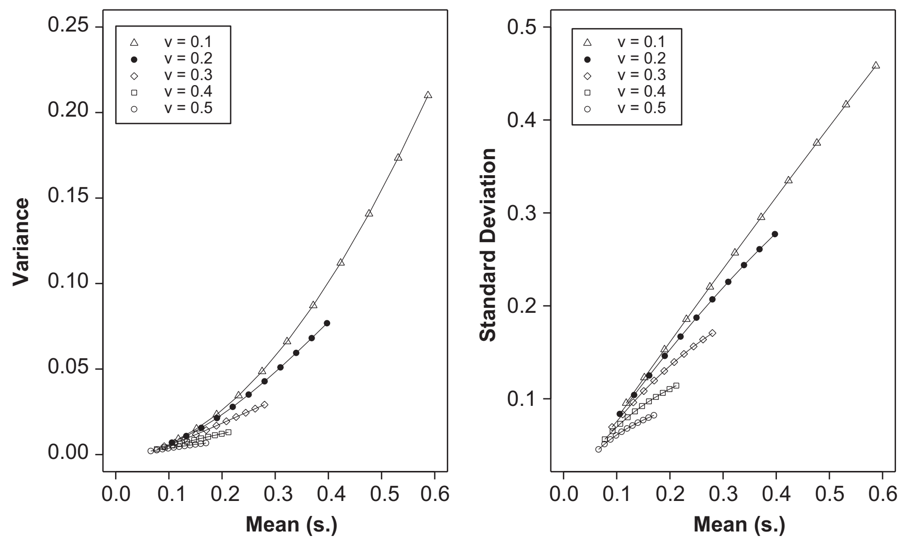

Fig. 4. The relation between RT mean and RT variance (left panel) and RT mean and RT standard deviation (right panel) for 5 levels of drift rate v, and 11 values for boundary separation a. The parameter ranges for a and v (i.e., $a \in [0.07, 0.17]$; $v \in [0.1, 0.5]$) were motivated by previous applications of the model.

mean and RT variance as a function of drift rate, diffusion variance, and boundary separation. Next, we studied how the variance goes with the mean when either drift rate or boundary separation is gradually increased along a range of plausible parameter values. The results showed that RT mean increases in an approximately linear fashion with RT standard deviation as drift rate is decreased. When boundary separation is gradually increased, the relation between RT mean and RT standard deviation depends on the value of the drift rate parameter: only when drift rate is relatively low will the relation between mean and standard deviation be approximately linear.

In this theoretical note, the focus has been entirely on the diffusion model for choice RT. It is certainly possible, and potentially informative, to study the mean–variance relationship for alternative models of choice RT, such as accumulator models (e.g., Smith & Vickers, 1988), Poisson counter models (e.g., LaBerge, 1994; Pike, 1966, 1973; Townsend & Ashby, 1983), Ornstein–Uhlenbeck models with non-negligible decay in drift-rate, and the recently proposed ballistic model of choice RT (Brown & Heathcote, 2005). These alternative models may or may not produce approximate linearity between RT mean and RT standard deviation as the efficiency of processing is manipulated. As mentioned above, at least one alternative model (i.e., Logan's instance model) yields results that are similar to those derived from the diffusion model.

The theoretical work presented here also outlines a qualitative prediction for the diffusion model that could be subjected to empirical tests. That is, two-choice experiments that manipulate task-difficulty (e.g., word frequency in a lexical decision task) across many different levels should find an approximate linear relationship between RT mean and RT standard deviation. Unfortunately, most experiments to date have manipulated task difficulty across only two or three levels, and this is clearly an insufficient number to empirically assess the mean–variance relationship. A few experiments, however, did systematically manipulate task difficulty across many levels (Chocholle, 1940; Green & Luce, 1971). The results, summarized in Luce (1986, p. 64), support the theoretical analysis reported here, as both studies found a strong linear relationship between RT mean and RT standard deviation. Also, the extensive data sets presented by Logan (1988, 1992) provide evidence that the result of practice is to decrease RT mean and RT standard deviation at the same rate. This result is consistent with a diffusion model account in which the effect of practice is to increase drift rate.

In sum, Eq. (12) gives the relation between the mean and variance of a diffusion model RT distribution. In general, the variance will always increase with the mean, but the specific form of this relation depends on the nature of the model parameters that differ between experimental conditions or participants.

Acknowledgments

Preparation of this article was supported by a VENI grant from the Netherlands Organisation for Scientific Research (NWO). We would like to thank Scott Brown, Richard Chechile, Michael Lee, and Jim Townsend for helpful comments.

References

Amemiya, T. (1985). *Advanced econometrics*. Cambridge, MA: Harvard University Press.

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, *112*, 117–128.

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

Chocholle, R. (1940). Variations des temps de réaction auditifs en fonction de l'intensité a diverses fréquences. *L'Année Psychologique*, *41*, 65–124.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332–361.

Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, *47*, 304–322.

Emerson, J. D., & Stoto, M. A. (2000). Transforming data. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Understanding robust and exploratory data analysis* (pp. 97–128). New York: Wiley.

Feller, W. (1968). *An introduction to probability theory and its applications*. New York: Wiley.

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

Green, D. M., & Luce, R. D. (1971). Detection of auditory signals presented at random times: III. *Perception & Psychophysics*, *9*, 257–268.

Honerkamp, J. (1994). *Stochastic dynamical systems*. New York: VCH Publishers.

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journal of Gerontology: Psychological Sciences*, *57B*, 101–115.

Karlin, S., & Taylor, H. M. (1981). *A second course in stochastic processes*. New York: Academic Press.

Keselman, H. J., Othman, A. R., Wilcox, R. R., & Fradette, K. (2004). The new and improved two-sample *t* test. *Psychological Science*, *15*, 47–51.

Kramer, A. F., Strayer, D. L., & Buckley, J. (1990). Development and transfer of automatic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 505–522.

LaBerge, D. A. (1994). Quantitative models of attention and response processes in shape identification tasks. *Journal of Mathematical Psychology*, *38*, 198–243.

Laming, D. R. J. (1968). *Information theory of choice-reaction times*. London: Academic Press.

Laming, D. R. J. (1973). *Mathematical psychology*. New York: Academic Press.

Levine, D. W., & Dunlap, W. P. (1983). Data transformation, power, and skew: A rejoinder to games. *Psychological Bulletin*, *93*, 596–599.

Li, S.-C. (2002). Connecting the many levels and facets of cognitive aging. *Current Directions in Psychological Science*, *11*, 38–43.

Link, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, *40*, 77–105.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 883–914.

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

MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria longitudinal study. *Psychology and Aging*, *18*, 510–523.

Pampel, F. C. (2000). *Logistic regression: A primer*. Thousand Oaks, CA: Sage.

Peterson, B. S., & Leckman, J. F. (1998). The temporal dynamics of tics in Gilles de la Tourette syndrome. *Biological Psychiatry*, *44*, 1337–1348.

Pike, A. R. (1966). Stochastic models of choice behaviour: Response probabilities and latencies of finite Markov chain systems. *British Journal of Mathematical and Statistical Psychology*, *21*, 161–182.

Pike, A. R. (1973). Response latency models for signal detection. *Psychological Review*, *80*, 53–68.

R Development Core Team (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available: http://www.Rproject.org

Raftery, A. E., & Zheng, Y. (2003). Discussion: Performance of Bayesian model averaging. *Journal of the American Statistical Association*, *98*, 931–938.

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

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, *88*, 552–572.

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review*, *9*, 278–291.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). Diffusion model account of lexical decision. *Psychological Review*, *111*, 159–182.

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

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 127–140.

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. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics*, *65*, 523–535.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*, 438–481.

Ratcliff, R., Van Zandt, T., & McKoon, R. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *102*, 261–300.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*, 369–385.

Shammi, P., Bosman, E., & Stuss, D. T. (1998). Aging and variability in performance. *Aging, Neuropsychology, and Cognition*, *5*, 1–13.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, *44*, 408–463.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*, 135–168.

Snedecor, G. W., & Cochran, W. G. (1989). *Statistical methods* (8th ed). Ames (IA): Iowa State University Press.

Sornette, D. (2000). *Critical phenomena in natural sciences*. New York: Springer.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, UK: Cambridge University Press.

Tuerlinckx, F., Maris, E., Ratcliff, R., & De Boeck, P. (2001). A comparison of four methods for simulating the diffusion process. *Behavior Research Methods, Instruments, & Computers*, *33*, 443–456.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science*, *25*, 111–172.

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, *7*, 424–465.

Van Zandt, T. (2002). Analysis of response time distributions. In J. T. Wixted (Ed.), *Stevens' handbook of experimental psychology* (vol. 4. 3rd ed., pp. 461–516).

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

Zhou, X.-H., Gao, S., & Hui, S. L. (1997). Methods for comparing the means of two independent log-normal samples. *Biometrics*, *53*, 1129–1135.