Fast and accurate calculations for first-passage times in Wiener diffusion models

Danielle J. Navarro[a], Ian G. Fuss[b]

[a] *School of Psychology, University of Adelaide, Adelaide SA 5005, Australia*

[b] *School of Electrical and Electronic Engineering, University of Adelaide, Adelaide SA 5005, Australia*

Abstract

We propose a new method for quickly calculating the probability density function for first passage times in simple Wiener diffusion models, extending an earlier method used by Van Zandt, Colonius and Proctor (2000). The method relies on the observation that there are two distinct infinite series expansions of this probability density, one of which converges quickly for small time values, while the other converges quickly at large time values. By deriving error bounds associated with finite truncation of either expansion, we are able to determine analytically which of the two versions should be applied in any particular context. The bounds indicate that, even for extremely stringent error tolerances, no more than 8 terms are required to calculate the probability density. By making the calculation of this distribution tractable, the goal is to allow more complex extensions of Wiener diffusion models to be developed.

Keywords: Diffusion model, Wiener processes, first passage times, response time

1 Introduction

In almost any decision-making task that humans face, the time taken to make a choice is a key dependent variable and conveys a great deal about the underlying cognitive processes. In order to relate choice response times (RTs) to some set of underlying processes, a general class of "sequential sampling models" has been developed to account for the time-course of human decision-making (e.g., Ratcliff 1978; Vickers 1979; Smith & Van Zandt 2000). Drawing on research in the statistics literature on sequential analysis (e.g., Wald 1947, see also Stone 1960), the central insight is to recognize that people seek to make good decisions quickly, and so have to solve a speed-accuracy trade-off. In general, taking the

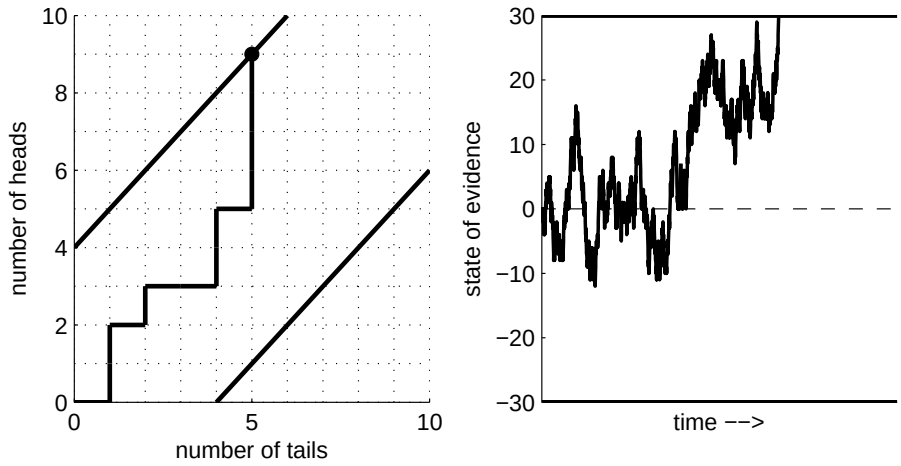

Fig. 1. Illustrations showing a discrete-time random walk model (left) and a continuous-time diffusion model (right). In the left panel, the axes used correspond to "evidence for A" and "evidence for B", so time (or more precisely, sample size) runs diagonally from the bottom left to the top right. In the right panel, the axes are rotated 45 degrees, and thus correspond to "time" and "state of evidence".

time to collect more information about the problem at hand (whatever it may be) should be expected to improve accuracy, but at the obvious cost of actually taking more time to do so. Good introductions to the area are provided by Luce (1986), Van Zandt (2000) and Ratcliff and Smith (2004).

The highly successful diffusion model proposed by Ratcliff (1978) is an example of a sequential sampling model: it assumes that, when asked to make simple perceptual decisions such as "is this line longer than that line?" the visual system samples evidence from the external stimulus environment, and continues to do so until some termination criterion is met. In both "random walk" and "diffusion" models, the process stops when the evidence for response A exceeds that for response B by some amount. If one were flipping a coin to determine if it is biased towards heads or tails, this might correspond to a "keep flipping until the number of heads exceeds the number of tails by 4 (or vice versa)", as illustrated in the left panel of Figure 1. As noted by Wald and Wolfowitz (1948), this decision process is statistically optimal when the decision-maker's loss function involves a simple kind of speed-accuracy trade-off. Although other termination rules have been proposed on the basis of psychological (Vickers 1979) and statistical (Frazier & Yu 2008) considerations, this rule remains the dominant approach in the literature.

Besides the termination rule, the most fundamental assumption made by the diffusion model is that the sampling process involved operates very quickly but with a lot of noise, with the result that it convenient to assume that evidence accrues continuously, rather than in terms of discrete samples. When constructing this model (e.g., Feller 1968, Ratcliff

1978) it is typical to derive the continuous-time "diffusion" model as a limiting version of the discrete-time "random walk" model. If we let $X(t)$ denote the "state of evidence" at time t, the result is a Wiener diffusion process: a simple dynamic system in which the state of evidence evolves via the stochastic differential equation $\frac{d}{dt}X(t) \sim \text{Normal}(v, \sigma^2)$ (see, e.g., Smith 2000 for detailed discussion). Accordingly, the marginal distribution over the state of evidence at time t is described by a normal distribution with mean $vt + z$ and variance $\sigma^2 t$. The resulting model (illustrated in the right panel of Figure 1) for human decision-making treats the choices c and corresponding decision-times t_d as random variables described by the "first passage to absorption" for this Wiener process. That is, if the initial state of evidence $X(0)$ lies in the range $0 < X(0) < a$, the model predicts that a decision is made at the first time t for which $X(t) \leq 0$ or $X(t) \geq a$, where absorption at the lower-boundary ($X(t) = 0$) corresponds to one possible choice and the absorption at the upper boundary ($X(t) = a$) results in the other choice. Statistically, if we fix the Wiener process variance[1] at $\sigma^2 = 1$ we refer to the resulting distribution as the Wiener first-passage time (WFPT) distribution, denoted $(c, t_d) \sim \text{WFPT}(v, a, z)$.

The second extension to the simple random walk model that Ratcliff (1978) made when constructing the diffusion model was to assume that the start point z and the drift rates v are also random variables (generally uniformly-distributed and normally-distributed, respectively), and may differ for every observed decision. Similarly, it is assumed that observed response times consist of both the decision time t_d and time for stimulus encoding and motor response generation t_{er}, so the actual RT is given by $t_d + t_{er}$, where t_{er} is also assumed to be a random variable (usually uniformly distributed). Finally, in some versions of the model an explicit treatment is given for outlier data, where the decision-maker is assumed to be generating "junk" data with some (small) probability (Ratcliff & Tuerlinckx 2002). As a result, the full diffusion model is fairly complex, and contains a number of psychologically-meaningful parameters (e.g., Voss, Rothermund & Voss, 2004).

In this paper, we consider the Wiener first passage time distribution that forms the core of the diffusion model, namely $\text{WFPT}(v, a, z)$. Our approach is somewhat different to recent proposals by Tuerlinckx (2004) and Voss and Voss (2008) in that our goal is to consider the faster computation of the WFPT distribution, not the full model with the additional random variables. It also differs from the approach taken by Wagenmakers, van der Maas & Grasman (2007), in that we do not place any restrictions on the parameters (but see Grasman, Wagenmakers & van der Maas, in press, for a more general approach), and from that of Brown, Ratcliff and Smith (2006) and Diederich and Busemeyer (2003) who focus on general simulation methods. Our choice to focus on the model at this level of generality is deliberate – in the "narrow" context of modelling choice and RT in simple two-alternative decision tasks, we envision that faster WFPT calculations would

[1] Ratcliff (1978) fixes $\sigma = 0.1$, which has some practical advantages but is mathematically inconvenient for our purposes. Following Voss, Rothermund & Voss (2004) and Lee, Fuss & Navarro (2006), we use the $\sigma = 1$ version, and note that this means that estimates of v, a and z will all be 10 times the size of the corresponding estimates in Ratcliff's formalization.

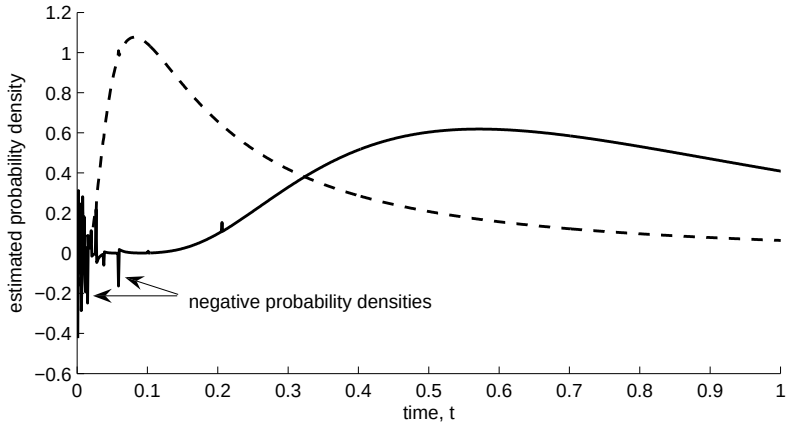

Fig. 2. A deliberately-extreme example of the kinds of pathologies that can occur when WFPT probability densities are calculated inappropriately. In this case, we produce the figure using the method described by Ratcliff & Tuerlinckx (2002) to approximate the cdf, and approximating the pdf via finite differencing. To produce errors of this magnitude, we terminate the sums far too early, using a 10^{-3} termination rule rather than the standard 10^{-29} rule. Large errors are observed at small t, even though the calculations in this case involve the evaluation of up to 60 terms.

sit naturally within the Bayesian hierarchical characterization of the "full" diffusion model (Lee, Fuss & Navarro, 2007; Vandekerckhove, Tuerlinckx & Lee, 2008). A major advantage of the Bayesian approach is that it allows *any* class of extensions to the diffusion model to be handled efficiently using modern computational statistics (e.g., Chen, Shao & Ibrahim 2000), so long as the WFPT distribution itself is tractable: in fact, Lee et al (2007) and Vandekerckhove et al (2008) both present simple extensions of this kind, in which explicit psychophysical functions are used to constrain drift rates across experimental conditions. Nevertheless, since the "standard" calculation of the WFPT distribution can sometimes require the evaluation of hundreds of terms in order to avoid pathologies at small RT values (see Figure 2 for an exaggerated example), the kind of large-scale computational methods that have become available for modelling higher-order cognition (e.g., Kemp & Tenenbaum 2008, Griffiths, Steyvers & Tenenbaum, 2007) are currently infeasible. What is required is a simple, pathology-free method for quickly computing the WFPT distribution with near-zero error. We provide such a method in this paper.

2 Computing the first-passage time densities

As the previous discussion makes clear, the most important aspect to the diffusion model is the WFPT distribution, parameterized by drift v, boundary separation a and start point z. It is convenient, however, to rewrite the start point (which varies from 0 to a)

as a relative start point $w = z/a$ (which varies from 0 to 1). Given this, the probability density function for the WFPT distribution, which describes the chance that the diffusion process is absorbed at time t at the *lower* boundary will be denoted by $f(t|v, a, w)$.

An analytic expression for this probability density was provided by Feller (1968, ch. 14, eq. 6.15). When written using the notation introduced above, the formula given for this WFPT density is, [2]

$$f(t|v, a, w) = \frac{\pi}{a^2} \exp\left(-vaw - \frac{v^2 t}{2}\right) \sum_{k=1}^{\infty} k \exp\left(-\frac{k^2 \pi^2 t}{2a^2}\right) \sin(k\pi w). \tag{1}$$

The probability density at the upper boundary is straightforward to obtain, by setting $v' = -v$ and $w' = 1-w$. Algebraically, one nice aspect to the expression is that it factorizes very simply, allowing the three-parameter density function to be written as follows:

$$f(t|v, a, w) = \frac{1}{a^2} \exp\left(-vaw - \frac{v^2 t}{2}\right) f\left(\frac{t}{a^2} \mid 0, 1, w\right). \tag{2}$$

This expression makes clear that we can, without loss of generality, consider the case where $a = 1$ and $v = 0$, and hence reduce the problem of calculating the general first passage density $f(t|v, a, w)$ to the problem of calculating a standard case, $f(t|0, 1, w)$. Accordingly, we now turn our attention to the calculations involved in this case.

When calculating WFPT densities, a typical approach is to make use of this "large time" expansion (e.g., Ratcliff 1978; Luce 1986; Ratcliff & Tuerlinckx 2002; Ratcliff & Smith 2004; Tuerlinckx 2004). In terms of the standard case $f(t|0, 1, w)$, we rely on the series:

$$f(t|0, 1, w) = \pi \sum_{k=1}^{\infty} k \exp\left(-\frac{k^2 \pi^2 t}{2}\right) \sin(k\pi w). \tag{3}$$

However, Feller also provides a different "small time" representation that is less frequently used (but see Van Zandt 2000, Van Zandt, Colonius & Proctor 2000, Voss, Rothermund & Voss 2004). It is given within problem 22 in Feller (1968, ch 14), and is produced by finding limiting versions of a slightly different treatment of the discrete-time random

[2] Strictly, we should refer to this expression as describing a probability density component at time t at the lower boundary. When integrated over t, the expression yields the choice probability corresponding to the lower boundary. The lower boundary mass plus the upper boundary mass sum to 1 with probability 1. However, for the sake of simplicity we use the term "probability density function" in an unqualified fashion and assume that the more technical sense is clear from context.

walk model than the one that produces the large-time expansion. For our purposes, what matters is that this alternative version produces the series:

$$f(t|0,1,w) = \frac{1}{\sqrt{2\pi t^3}} \sum_{k=-\infty}^{\infty} (w+2k) \exp\left(-\frac{(w+2k)^2}{2t}\right). \tag{4}$$

(The reason for referring to two different representations of the WFPT densities as "large time" and "small time" expansions will be made explicit shortly). Obviously, since both expressions involve the evaluation of infinite sums, any implementation of the diffusion model must rely on a truncated version of one of these two series. In the case of the usual large-time version, the natural way to truncate the sum is to stop calculating terms once k exceeds some threshold value. Thus, in order to restrict the calculation to κ terms we obtain

$$f_\kappa^\ell(t|0,1,w) = \pi \sum_{k=1}^{\kappa} k \exp\left(-\frac{k^2\pi^2 t}{2}\right) \sin(k\pi w) \tag{5}$$

The small-time version is slightly more complicated since the series extends to infinity in both directions. In this case, a simple way to restrict the sum to κ terms is to use

$$f_\kappa^s(t|0,1,w) = \frac{1}{\sqrt{2\pi t^3}} \sum_{k=-\lceil(\kappa-1)/2\rceil}^{\lfloor(\kappa-1)/2\rfloor} (w+2k) \exp\left(-\frac{(w+2k)^2}{2t}\right), \tag{6}$$

where $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ are the floor and ceiling functions respectively.

When calculating diffusion model predictions, the difficult part is to choose a value of κ to govern the truncation. For instance, a commonly-used approach (e.g., Ratcliff & Tuerlinckx 2002) is to continue to compute the series until the value of the *cumulative* distribution function changes by less than 10^{-29} times the current value of the sum for two successive terms. While this seems to be an intuitively reasonable heuristic, a better approach would be to specify some acceptable degree of approximation error, and then compute the minimum number of terms required to ensure that the truncated sum meets the required tolerance. That is, choose κ such that the calculated density $f_\kappa(t|0,1,w)$ deviates from the true value $f(t|0,1,w)$ by no more than some target error level ϵ,

$$|f_\kappa(t|0,1,w) - f(t|0,1,w)| < \epsilon. \tag{7}$$

Of course, since the true value of the density is necessarily unknown, we cannot calculate this error (henceforth denoted $E_\kappa(t)$) exactly, but we *can* put an upper bound on it, meaning that we can instead choose the smallest value of κ for which we can *prove* that the truncation error remains below ϵ.

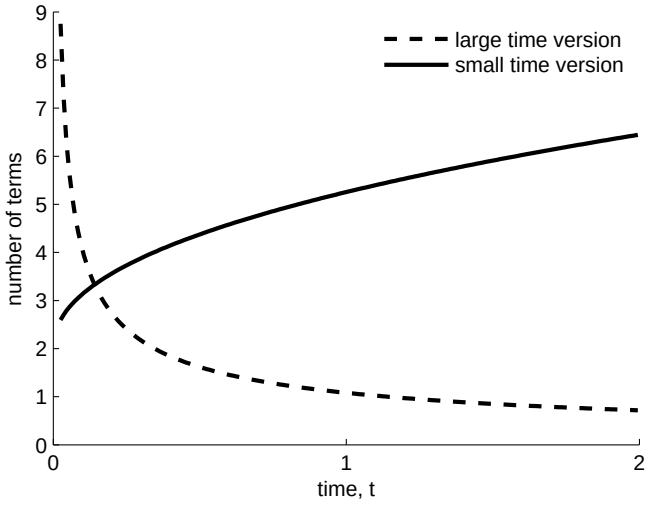

Fig. 3. Number of terms κ required to guarantee a truncation error below .001, for both the large-time (Equation 10) and small-time (Equation 11) versions. For t less than approximately 0.15 the number of terms required is smaller for the small-time expansion, after which time the large-t version is superior.

In Appendix A we show that the absolute error $E_\kappa^\ell(t)$ that results from approximating the function $f(t|0,1,w)$ by the truncated version of the large time expansion $f_\kappa^\ell(t|0,1,w)$ satisfies the inequality

$$E_\kappa^\ell(t) \leq \frac{1}{\pi t} \exp\left(-\frac{\kappa^2\pi^2}{2}t\right) \tag{8}$$

so long as $\kappa \geq 1/(\pi\sqrt{t})$. Similarly, in Appendix B we show that the error $E_\kappa^s(t)$ that results from using the truncated small-time series $f_\kappa^s(t|0,1,w)$ is bounded above as follows:

$$E_\kappa^s(t) \leq \frac{1}{2\sqrt{2\pi t}} \exp\left(-\frac{(\kappa-2)^2}{2t}\right), \tag{9}$$

a result which holds for $\kappa > 1 + \sqrt{t}$. Note that these two expressions illustrate why we refer to Equation 3 as the large-time expansion, since the error in Equation 8 tends to be largest at small t. Similarly, since the error implied by Equation 9 is largest at large t, we refer to Equation 4 it as the small-time expansion.

Rearrangement of these bounds implies that, in order to guarantee a truncation error

below ϵ the number of terms needed is

$$\kappa \geq \sqrt{\frac{-2\log(\pi t\epsilon)}{\pi^2 t}} \tag{10}$$

for the large-time approximation, whereas for the small-time approximation the corresponding number of terms is given by

$$\kappa \geq 2 + \sqrt{-2t\log(2\epsilon\sqrt{2\pi t})}. \tag{11}$$

In both cases it is assumed that the expressions are real valued: when the small-t version is used, the error tolerance should be set such that $\epsilon \leq 1/(2\sqrt{2\pi t})$. However, in such cases it is straightforward to lower the value of ϵ to $1/(2\sqrt{2\pi t})$, which yields the requirement that $\kappa \geq 2$. A similar constraint applies to the large-t version, namely that $\epsilon \leq 1/(\pi t)$. Lowering ϵ to the smallest allowable value would lead to $\kappa \geq 0$, but since the derivation of the error bound only holds for $\kappa \geq 1/(\pi\sqrt{t})$, this sets the value of κ in this case.

These functions are shown in Figure 3, for $\epsilon = .001$. Note that across all values of t, it would take no more than four terms to keep the truncation error within this tolerance for at least one of the two versions. More generally, for a fixed error level ϵ, in order to minimize the number of terms evaluated, the bounds imply that we should use the small t version when the function

$$\lambda(t) = 2 + \sqrt{-2t\log(2\epsilon\sqrt{2\pi t})} - \sqrt{\frac{-2\log(\pi t\epsilon)}{\pi^2 t}} \tag{12}$$

is less than zero. To illustrate why this changeover occurs, Figure 4 plots the $\kappa = 3$ approximations for both the small-time version and the large-time version, for $w = .5$. As one would expect given the nature of the bounds described earlier, the two expansions show pathologies in different parts of the density function: the large-t approximation (dashed lines) works poorly at small time values, while the small-t approximation (solid lines) works poorly at larger time values. This suggests the natural approximation:

$$f(t|0,1,w) \approx \begin{cases} \dfrac{1}{\sqrt{2\pi t^3}} \displaystyle\sum_{k=-\lfloor(\kappa-1)/2\rfloor}^{\lceil(\kappa-1)/2\rceil} (w+2k)\exp\left(\frac{(w+2k)^2}{2t}\right) & \text{if } \lambda(t) < 0 \\ \pi \displaystyle\sum_{k=1}^{\kappa} k\exp\left(-\frac{k^2\pi^2 t}{2}\right)\sin(k\pi w) & \text{if } \lambda(t) \geq 0 \end{cases}, \tag{13}$$

where, as indicated earlier, we construct the more general pdf via the relationship in Equation 2. It should be noted that the basic idea is not new – for instance, Van Zandt et al. (2000) compute WPFT predictions by switching between the small-time and large-time

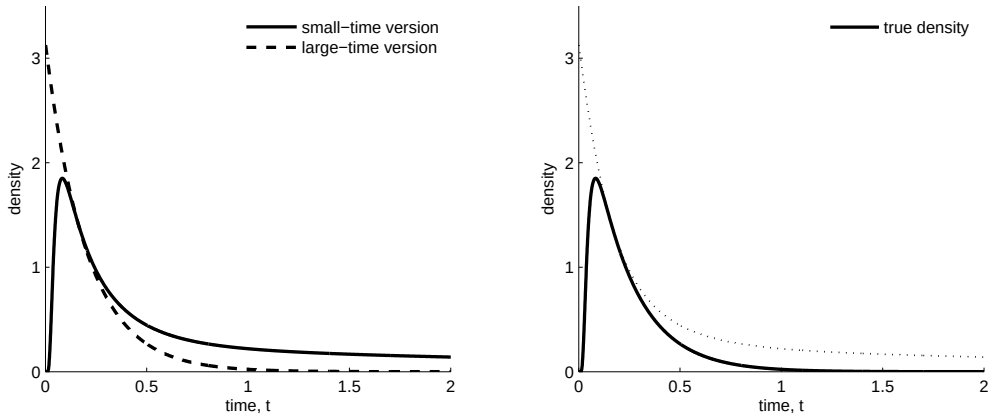

Fig. 4. The $\kappa = 2$ truncations for the small-time and large-time expansions (left), as compared to the true density (right). As one might expect, the large-time version (dashed line) performs very poorly for small values of t when only two terms are used, while the small-time version (solid line) performs poorly in the tails.

versions, using a heuristic method (Van Zandt, personal communication) to govern the changeover. The novel aspect to our proposal is that the function $\lambda(t)$ allows the switch to be made in an analytically-derived fashion, and the total number of terms computed is governed by the explicit derivation of the error bounds introduced above. MATLAB code implementing this method is attached in Appendix C.

3 Effectiveness of the Method

We now turn to some simple tests of the effectiveness of the proposed method for computing the WFPT distribution. As a first test, Figure 5 plots the Wiener first passage time predictions for a process with $v = 1$, $a = 2$ and $z = .5$. On the left, predictions are made using the slower "classical" method discussed in Ratcliff & Tuerlinckx's (2002) paper: namely, to terminate when two successive terms in the cdf calculation remain below 10^{-29} times the current value of the sum. In the middle, predictions are made using the "fast truncations" presented in this paper (i.e., using Equation 13), with a stringent error tolerance $\epsilon = 10^{-29}$ so as to roughly mimic the standard applied in the classical version. The panel on the right shows the difference between the two versions, which is very small. Although the two methods make very similar predictions, they differ dramatically in the amount of computation required to do so, as shown in Figure 6, which shows the number of terms required to compute both versions, as a function of t. At large t the two versions are comparable, which is what one would expect since for these t values both approaches rely on Equation 4. At small t, the classical method needs to calculate a very large number

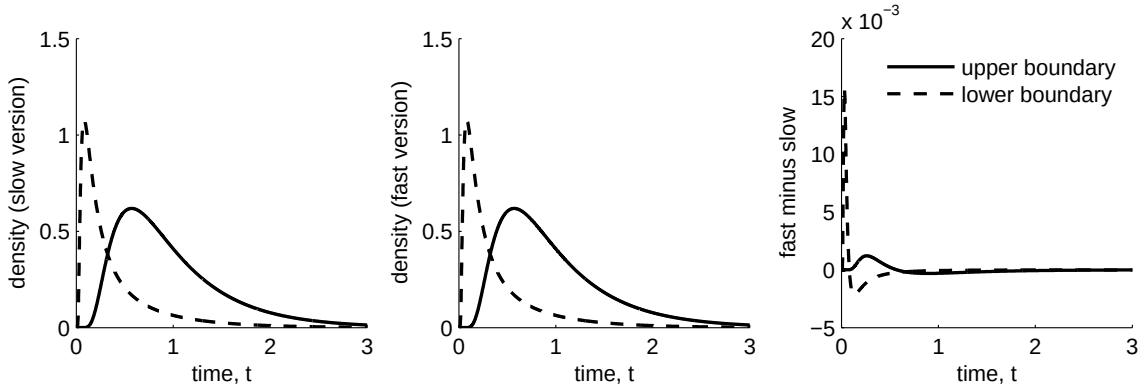

Fig. 5. Illustration of the performance of the approximation, for $v = 1$, $a = 2$ and $z = .5$. The left panel shows model predictions calculated using the "slow" classical method, while the middle panel shows the predictions calculated using the fast truncation method suggested here. As is illustrated in the panel on the right, the differences between the two predictions (fast minus slow) are minimal.

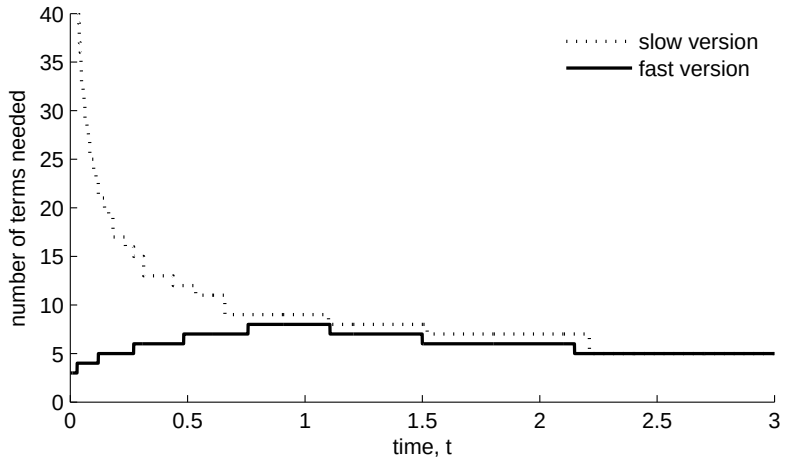

Fig. 6. Illustration of the extent of the speed up: the number of terms required to calculate the slower classical heuristic (dotted) and fast truncation (solid) predictions from Figure 5, in which $\epsilon = 10^{-29}$. Also note that the full extent of the speed up for very small t is masked since the dotted line accelerates very rapidly beyond the bounds of the plot. For instance, at $t = .001$, the slow version required 227 terms, compared to a mere 8 terms required at worst for the fast version.

of terms in order to avoid the pathologies observed in Figure 2: at $t = .001$, the number required was 227 terms. In comparison, the fast truncation method never required more than 8 terms to be evaluated.

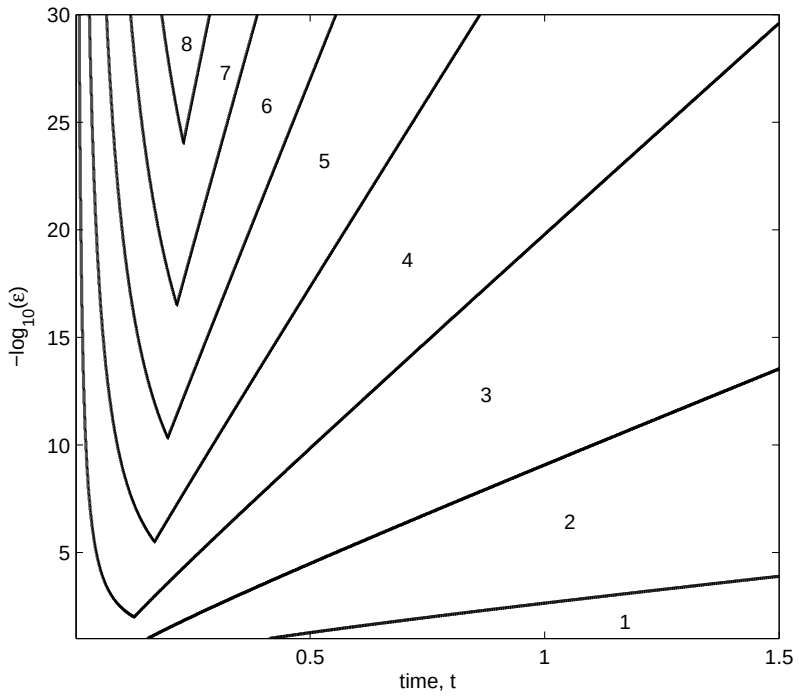

Fig. 7. Number of terms required to compute approximate the standard WFPT densities $f(t|0, 1, w)$, as a function of the time t and the error tolerance ϵ. Since the bounds are derived independently of the parameters v, a and w these results hold quite generally.

This illustration highlights a few key points. Firstly, since explicit bounds are used to control the truncation, the accuracy of the approach is guaranteed (as per Figure 5): the main issue at hand is how fast this accuracy can be obtained (as in Figure 6). Secondly, as Figure 6 makes clear, the classical method can be made to perform as poorly as desired, simply by taking t as close to zero as we like. Thirdly, since the fast method (by definition) uses whichever of the small t and large t versions is superior, it is never worse than the existing method. Taken together, these observations make clear that there is little to be learned by making extensive comparisons between the two approaches. Rather, the interesting question relates mainly to the number of terms required for different values of t and ϵ. For the standard case, $f(t|0, 1, w)$, the number of terms required is illustrated in Figure 7: since the top of the figure corresponds to the case where $\epsilon = 10^{-30}$ it is clear that one would have to set an extraordinarily low tolerance for error before requiring more than 10 terms for *any* value of t. Indeed, given the imprecision associated with real data, one would rarely expect to be calculating more than 5 terms when using this method. The general case is a straightforward extension: all that is needed is to take account of the additional multiplicative term in Equation 2, namely $(1/a^2) \exp(-vaw - v^2t/2)$.

4 Conclusion

The main goal of this paper was to find a method for computing finite approximations to the WFPT distribution with as little effort as possible. By deriving upper bounds on the truncation errors associated with Feller's small-time and large-time expansions of the true density, we are able to propose a hybrid method that uses whichever of these two methods is most appropriate. The number of terms required in our simulations never exceeded 8, even when the error tolerance ϵ was set to unnecessarily low levels, and in no case did the method produce negative probability densities. We hope that this method will assist in the application of models reliant on this distribution, such as the full diffusion model (Ratcliff 1978) and its extensions (e.g., Lee et al., 2007; Vandekerckhove et al. 2008). In particular, most Bayesian estimation methods (e.g., Chen et al. 2000) rely heavily on the use of the exact probability density functions rather than χ^2 statistics, for instance, and so can benefit from this approach. Moreover, the Bayesian framework lends itself naturally to the development of rich, hierarchically structured stimulus representations (e.g., Kemp & Tenenbaum, 2008; Griffiths et al. 2007). In our view, tractable WFPT distributions allow psychologically-plausible models for decision-times to be integrated with psychologically-interesting approaches to stimulus representation, allowing the construction of time-sensitive models for complex, real-world decisions.

5 Acknowledgements

Correspondence concerning this article should be addressed to Daniel Navarro, School of Psychology, University of Adelaide, SA 5005, Australia. Tel +61 8 8303 5265, Fax +61 8 8303 3770, E-mail: daniel.navarro@adelaide.edu.au. DJN was supported by an Australian Research Fellowship (ARC grant DP-0773794). We thank Roger Ratcliff, Francis Tuerlinckx, Trish Van Zandt, E.J. Wagenmakers and three anonymous reviewers for their comments and suggestions. Code associated with this article is available on the first author's website.

6 References

Brown, S.D., Ratcliff, R., & Smith, P.L. (2006). Evaluating methods for approximating stochastic differential equations. *Journal of Mathematical Psychology, 50*, 402-410.

Chen, M., Shao, Q. & Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation.* New York, NY: Springer.

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*(3), 304322.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications* (Vol. 1, 3rd Ed). New York: Wiley.

Frazier, P. I. & Yu, A. J. (2008). Sequential hypothesis testing under stochastic deadlines. In J. C. Platt, D. Koller, Y. Singer & S. Roweis (Eds.) *Advances in Neural Information Processing Systems, 20* (pp. 465-472). Cambridge, MA: MIT Press

Grasman, R. P. P. P., Wagenmakers, E.-J., & van der Maas, H. L. J. (in press). On the mean and variance of response times under the diffusion model with an application to parameter estimation. *Journal of Mathematical Psychology.*

Griffiths, T. L., Steyvers, M. & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114*, 211-244.

Kemp, C. & Tenenbaum, J. B. (2008). Learning structural forms. *Proceedings of the National Academy of Sciences of the USA, 105,* 10687-10692

Lee, M. D., Fuss, I. & Navarro, D. J. (2007). A Bayesian approach to diffusion models of decision-making and response time. In B. Schölkopf, J. Platt & T. Hoffman (Eds.) *Advances in Neural Information Processing Systems, 19* (pp. 809-815). Cambridge, MA: MIT Press

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization.* Oxford, UK: Oxford University Press.

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

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

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.

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. & Van Zandt, T. (2000). Time-dependent Poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology, 53*, 293-315.

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

Tuerlinckx, F. (2004). The efficient computation of the cumulative distribution and probability density functions in the diffusion model. *Behavior Research Methods, Instruments, & Computers, 36*, 702-716.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2008). A Bayesian approach to diffusion process models of decision-making. In V. Sloutsky, B. Love & K. McRae (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1429-1434). Austin, TX: Cognitive Science Society.

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

Van Zandt, T., Colonius, H. & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review, 7*, 208-256.

Vickers, D. (1979). *Decision Processes in Visual Perception.* New York, NY: Academic Press

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. 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.

Wald, A. (1947). *Sequential Analysis*. New York, NY: Wiley.

Wald, A. & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics, 19*, 326-339.

A Large-time error bound

In this section we derive the upper bound referred to in Equation 8, since this is the easier of the two bounds to construct. To do so, we begin by defining the function

$$g(k,t,w) = k\exp\left(-\frac{k^2\pi^2 t}{2}\right)\sin(k\pi w), \tag{A.1}$$

corresponding to the kth term in the series in Equation 3. Accordingly, the truncation error that results when one uses only the first κ terms in the large-time expansion is given by the magnitude of the sum of the remaining terms:

$$E_\kappa^\ell(t) = \pi\left|\sum_{k=\kappa+1}^{\infty} g(k,t,w)\right|. \tag{A.2}$$

Using the convexity of the absolute value function, we can state that

$$E_\kappa^\ell(t) \leq \pi\sum_{k=\kappa+1}^{\infty}\left|g(k,t,w)\right|. \tag{A.3}$$

This is equivalent to making the "worst case" assumption that all of the omitted terms are working in concert and do not cancel out at all. Moreover, noting that $-1 \leq \sin(k\pi w) \leq 1$ and thus $|\sin(k\pi w)| \leq 1$ we can place a very simple bound on the error, since

$$\begin{aligned}|g(k,t,w)| &= k\exp\left(-\frac{k^2\pi^2 t}{2}\right)|\sin(k\pi w)| &\text{(A.4)}\\ &\leq k\exp\left(-\frac{k^2\pi^2 t}{2}\right). &\text{(A.5)}\end{aligned}$$

If we apply this inequality to our existing bound, we obtain the new bound:

$$E_\kappa^\ell(t) \leq \pi \sum_{k=\kappa+1}^{\infty} k \exp\left(-\frac{k^2\pi^2 t}{2}\right). \tag{A.6}$$

Now consider the function $h(k,t) = k\exp(-\frac{k^2\pi^2 t}{2})$ that describes the summands involved in this new upper bound. From inspection it is clear that for small k the linear term will dominate and the function will be increasing. For larger k, however, the exponential term dominates and the function is decreasing. The stationary point at which this occurs is found by setting $\frac{d}{dk}h(k,t) = 0$, which occurs when $k = 1/(\pi\sqrt{t})$ and trivially as $k \to \infty$. When k is below this critical value, the amplitude of the sinusoidal terms is increasing, and it would probably be unwise to truncate the sum at any such value. So the interesting cases occur when $\kappa \geq 1/(\pi\sqrt{t})$, and in these cases we may treat $h(k,t)$ as a monotonic decreasing function of k. Given this, note that the sum in question is in effect a rectangle approximation to the corresponding integral, and so elementary integration theory gives us the following inequalities:

$$\sum_{x=a+1}^{b} h(x) \leq \int_a^b h(x)dx \leq \sum_{x=a}^{b-1} h(x). \tag{A.7}$$

The left inequality allows us to use the integral as an upper bound on the sum. Applying this inequality and solving the integral gives us the error bound referred to in the main text:

$$E_\kappa^\ell(t) \leq \pi \int_\kappa^\infty k \exp\left(-\frac{k^2\pi^2 t}{2}\right) dk \tag{A.8}$$

$$= \frac{1}{\pi t} \int_{\kappa\pi\sqrt{t/2}}^{\infty} 2u\exp(-u^2)du \tag{A.9}$$

$$= \frac{1}{\pi t}\exp\left(-\frac{\kappa^2\pi^2 t}{2}\right). \tag{A.10}$$

As noted previously, this bound holds for all interesting cases (i.e., when $\kappa > 1/(\pi\sqrt{t})$). In short, the bound derived above holds for large t, and in those cases the truncation error is provably small.

B Small-time error bound

We now turn to the derivation of the upper bound on the error associated with truncating the small-time expansion of the first-passage time density. Since the expansion now involves a sum from $-\infty$ to $+\infty$, the derivation is a little more complex, and so we need to be somewhat more careful. In this instance, the function we are interested in truncating is the one from Equation 4

$$f(t|0,1,w) = \frac{1}{\sqrt{2\pi t^3}} \sum_{k=-\infty}^{\infty} (w+2k) \exp\left(-\frac{(w+2k)^2}{2t}\right). \tag{B.1}$$

It is convenient to partition the sum into two parts such that $f(t|0,1,w) = \frac{1}{\sqrt{2\pi t^3}}(S^+ + S^-)$, where

$$S^+ = \sum_{k=0}^{\infty} (w+2k) \exp\left(-\frac{(w+2k)^2}{2t}\right) \text{ and} \tag{B.2}$$

$$S^- = \sum_{k=-\infty}^{-1} (w+2k) \exp\left(-\frac{(w+2k)^2}{2t}\right). \tag{B.3}$$

Having done so, we can rewrite S^- as follows:

$$S^- = -\sum_{k=1}^{\infty} (-w+2k) \exp\left(-\frac{(-w+2k)^2}{2t}\right). \tag{B.4}$$

This allows us to rewrite the first passage time density as

$$\begin{aligned} f(t|0,1,w) = \frac{1}{\sqrt{2\pi t^3}} \Bigg[&\sum_{k=0}^{\infty} (w+2k) \exp\left(-\frac{(w+2k)^2}{2t}\right) \\ &- \sum_{k=1}^{\infty} (-w+2k) \exp\left(-\frac{(-w+2k)^2}{2t}\right) \Bigg] \end{aligned} \tag{B.5}$$

Using this expression, we define our truncated series by allowing the indexing variable k to stop at some finite value in both of the two sums. So, if we set $\kappa = 1$, this yields

$$f_1(t|0,1,w) = \frac{1}{\sqrt{2\pi t^3}} w \exp\left(-\frac{w^2}{2t}\right), \tag{B.6}$$

which is similar to the small-time approximation used by Lee, Fuss & Navarro (2007). More generally, however, if κ is an even positive integer then

$$f_\kappa(t|0,1,w) = \frac{1}{\sqrt{2\pi t^3}} \left[\sum_{k=0}^{(\kappa-2)/2} (w+2k) \exp\left(-\frac{(w+2k)^2}{2t}\right) - \sum_{k=1}^{\kappa/2} (-w+2k) \exp\left(-\frac{(-w+2k)^2}{2t}\right) \right] \tag{B.7}$$

and if κ is an odd integer larger than 1 then

$$f_\kappa(t|0,1,w) = \frac{1}{\sqrt{2\pi t^3}} \left[\sum_{k=0}^{(\kappa-1)/2} (w+2k) \exp\left(-\frac{(w+2k)^2}{2t}\right) - \sum_{k=1}^{(\kappa-1)/2} (-w+2k) \exp\left(-\frac{(-w+2k)^2}{2t}\right) \right]. \tag{B.8}$$

Taken together, Equations B.6, B.7 and B.8 correspond to a rewritten version of the finite-term truncation described in Equation 6. With that in mind, the truncation error associated with Equation 6 can be described in the following way when κ is an even positive integer:

$$E^s_\kappa(t) = \frac{1}{\sqrt{2\pi t^3}} \left| \sum_{k=\kappa/2}^{\infty} (w+2k) \exp\left(-\frac{(w+2k)^2}{2t}\right) - \sum_{k=(\kappa+2)/2}^{\infty} (-w+2k) \exp\left(-\frac{(-w+2k)^2}{2t}\right) \right|. \tag{B.9}$$

Similarly, when κ is odd, then the error is given by

$$E^s_\kappa(t) = \frac{1}{\sqrt{2\pi t^3}} \left| \sum_{k=(\kappa+1)/2}^{\infty} (w+2k) \exp\left(-\frac{(w+2k)^2}{2t}\right) - \sum_{k=(\kappa+1)/2}^{\infty} (-w+2k) \exp\left(-\frac{(-w+2k)^2}{2t}\right) \right|. \tag{B.10}$$

In the even case (B.9) the first series is larger than the second series if $t < \kappa^2$. This can be seen to be true by noting that if $t < \kappa^2$ then the leading term of the first series is larger than the leading term of the second series and similarly for each successive pair of terms: hence, the first series is larger than the second. Given this observation, in order to obtain a simple bound on $E^s_\kappa(t)$, we derive an upper bound for the first sum, and subtract from it a lower bound for the second sum. In the odd case (B.10) the opposite applies, and so our bound is constructed in the opposite fashion, by finding an upper bound on the second sum and subtracting that from a lower bound on the first sum.

In all cases, we are seeking upper and lower bounds for functions of the form

$$S = \sum_{k=\nu/2}^{\infty} (c+2k) \exp\left(-\frac{(c+2k)^2}{2t}\right). \tag{B.11}$$

Applying the logic used in Appendix A, we define a function of k that corresponds to the summands,

$$g(k) = (c+2k) \exp\left(-\frac{(c+2k)^2}{2t}\right). \tag{B.12}$$

As before, the function is initially increasing since the linear term dominates, but for larger k it becomes a decreasing function since the exponential term comes to dominate. Again, the stationary point is found by setting $\frac{d}{dk}g(k) = 0$, which occurs when $k = (\sqrt{t} - c)/2$ and also as $k \to \infty$. Thus, $g(k)$ may be treated as a monotonic decreasing function so long as $\kappa > \sqrt{t} - c - 1$. Repeating the observation made in Appendix A, we can use elementary integration theory to show that

$$\int_{\nu/2}^{\infty} (c+2k) \exp\left(-\frac{(c+2k)^2}{2t}\right) dk \leq \sum_{k=\nu/2}^{\infty} (c+2k) \exp\left(-\frac{(c+2k)^2}{2t}\right) \tag{B.13}$$

and

$$\sum_{k=\nu/2}^{\infty} (c+2k) \exp\left(-\frac{(c+2k)^2}{2t}\right) \leq \int_{(\nu-2)/2}^{\infty} (c+2k) \exp\left(-\frac{(c+2k)^2}{2t}\right) dk. \tag{B.14}$$

Evaluation of the integrals gives

$$\frac{t}{2} \exp\left(-\frac{(c+\nu)^2}{2t}\right) \leq \sum_{k=\nu/2}^{\infty} (c+2k) \exp\left(-\frac{(c+2k)^2}{2t}\right) \leq \frac{t}{2} \exp\left(-\frac{(c+\nu-2)^2}{2t}\right). \tag{B.15}$$

Hence, to construct the bound for even-valued κ, we apply the upper bound to the first sum and the lower bound to the second sum, which gives us the expression:

$$E^s_\kappa(t) \leq \frac{1}{2\sqrt{2\pi t}} \left[\exp\left(-\frac{(w+\kappa)^2}{2t}\right) - \exp\left(-\frac{(w+\kappa+2)^2}{2t}\right)\right]. \tag{B.16}$$

In contrast, if κ is odd-valued, then we apply the lower bound to the first sum, and subtract this off the upper bound for the second sum (since in this case the second sum

is always the larger one). This gives

$$E_\kappa^s(t) \leq \frac{1}{2\sqrt{2\pi t}} \left[\exp\left(-\frac{(-w+\kappa-1)^2}{2t} \right) - \exp\left(-\frac{(w+\kappa+1)^2}{2t} \right) \right]. \tag{B.17}$$

To simplify matters, we note that since the exponential function is positive valued, we can set 0 as an upper bound on the $-\exp(x)$ terms in both equations. Similarly, since $0 \leq w \leq 1$ we can set it to the worst possible value ($w = 0$ for even κ and $w = 1$ for odd κ). By doing so we observe that, irrespective of whether κ is odd or even,

$$E_\kappa^s(t) \leq \frac{1}{2\sqrt{2\pi t}} \exp\left(-\frac{(\kappa-2)^2}{2t} \right), \tag{B.18}$$

which is the error bound in Equation 9. As observed above, this bound only holds for sufficiently large κ, which in this case corresponds to $\kappa > \sqrt{t} - c$. Noting that the smallest value of c used in any of the expressions is -1, we can state that this bound holds for $\kappa > \sqrt{t} + 1$. In short, this bound holds for small t, and in those cases the truncation error is provably small.

C MATLAB code

```
function p=wfpt(t,v,a,z,err)

tt=t/(a^2); % use normalized time
w=z/a; % convert to relative start point

% calculate number of terms needed for large t
if pi*tt*err<1 % if error threshold is set low enough
    kl=sqrt(-2*log(pi*tt*err)./(pi^2*tt)); % bound
    kl=max(kl,1/(pi*sqrt(tt))); % ensure boundary conditions met
else % if error threshold set too high
    kl=1/(pi*sqrt(tt)); % set to boundary condition
end

% calculate number of terms needed for small t
if 2*sqrt(2*pi*tt)*err<1 % if error threshold is set low enough
    ks=2+sqrt(-2*tt.*log(2*sqrt(2*pi*tt)*err)); % bound
    ks=max(ks,sqrt(tt)+1); % ensure boundary conditions are met
else % if error threshold was set too high
    ks=2; % minimal kappa for that case
```

```
end

% compute f(tt|0,1,w)
p=0; %initialize density
if ks<kl % if small t is better...
K=ceil(ks); % round to smallest integer meeting error
    for k=-floor((K-1)/2):ceil((K-1)/2) % loop over k
        p=p+(w+2*k)*exp(-((w+2*k)^2)/2/tt); % increment sum
    end
    p=p/sqrt(2*pi*tt^3); % add constant term

else % if large t is better...
K=ceil(kl); % round to smallest integer meeting error
for k=1:K
p=p+k*exp(-(k^2)*(pi^2)*tt/2)*sin(k*pi*w); % increment sum
end
p=p*pi; % add constant term
end

% convert to f(t|v,a,w)
p=p*exp(-v*a*w -(v^2)*t/2)/(a^2);
```