

ACCEPTED VERSION

Daniel J. Navarro and Ian G. Fuss

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

Journal of Mathematical Psychology, 2009; 53(4):222-230

Copyright 2009 Elsevier Inc. All rights reserved. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

PERMISSIONS

<http://www.elsevier.com/about/policies/article-posting-policy#accepted-manuscript>

Accepted Manuscript

Authors can share their accepted manuscript:

[...]

After the embargo period

- via non-commercial hosting platforms such as their institutional repository
- via commercial sites with which [Elsevier has an agreement](#)

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license – this is easy to do, click here to find out how
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our [hosting policy](#)
- not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article

19 August 2015

<http://hdl.handle.net/2440/60405>

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

Daniel 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

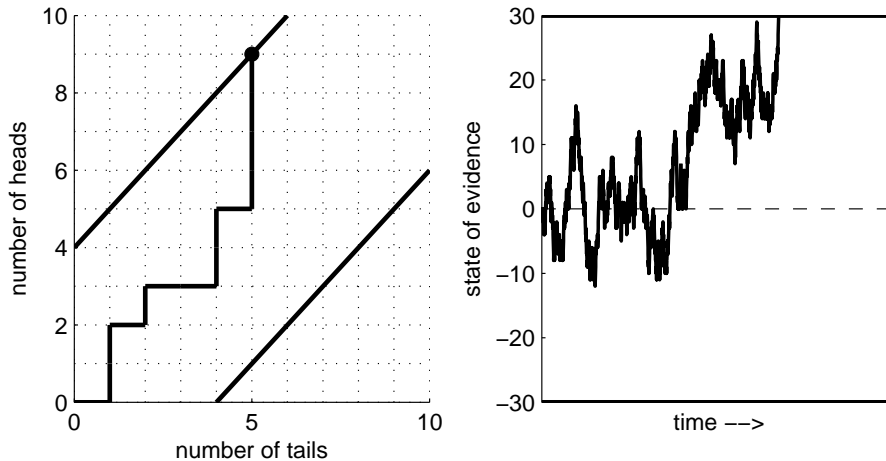


Fig. 1. Illustrations showing a discrete-time random walk model (left) and a continuous-time diffusion model (right). In the left panel, the axes used correspond to “evidence for A” and “evidence for B”, so time (or more precisely, sample size) runs diagonally from the bottom left to the top right. In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

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

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

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

31 1978) it is typical to derive the continuous-time “diffusion” model as a limiting version
 32 of the discrete-time “random walk” model. If we let $X(t)$ denote the “state of evidence”
 33 at time t , the result is a Wiener diffusion process: a simple dynamic system in which the
 34 state of evidence evolves via the stochastic differential equation $\frac{d}{dt}X(t) \sim \text{Normal}(v, \sigma^2)$
 35 (see, e.g., Smith 2000 for detailed discussion). Accordingly, the marginal distribution
 36 over the state of evidence at time t is described by a normal distribution with mean
 37 $vt + z$ and variance $\sigma^2 t$. The resulting model (illustrated in the right panel of Figure 1)
 38 for human decision-making treats the choices c and corresponding decision-times t_d as
 39 random variables described by the “first passage to absorption” for this Wiener process.
 40 That is, if the initial state of evidence $X(0)$ lies in the range $0 < X(0) < a$, the model
 41 predicts that a decision is made at the first time t for which $X(t) \leq 0$ or $X(t) \geq a$, where
 42 absorption at the lower-boundary ($X(t) = 0$) corresponds to one possible choice and the
 43 absorption at the upper boundary ($X(t) = a$) results in the other choice. Statistically, if
 44 we fix the Wiener process variance¹ at $\sigma^2 = 1$ we refer to the resulting distribution as
 45 the Wiener first-passage time (WFPT) distribution, denoted $(c, t_d) \sim \text{WFPT}(v, a, z)$.

46 The second extension to the simple random walk model that Ratcliff (1978) made when
 47 constructing the diffusion model was to assume that the start point z and the drift rates
 48 v are also random variables (generally uniformly-distributed and normally-distributed,
 49 respectively), and may differ for every observed decision. Similarly, it is assumed that
 50 observed response times consist of both the decision time t_d and time for stimulus encoding
 51 and motor response generation t_{er} , so the actual RT is given by $t_d + t_{er}$, where t_{er} is also
 52 assumed to be a random variable (usually uniformly distributed). Finally, in some versions
 53 of the model an explicit treatment is given for outlier data, where the decision-maker is
 54 assumed to be generating “junk” data with some (small) probability (Ratcliff & Tuerlinckx
 55 2002). As a result, the full diffusion model is fairly complex, and contains a number of
 56 psychologically-meaningful parameters (e.g., Voss, Rothermund & Voss, 2004).

57 In this paper, we consider the Wiener first passage time distribution that forms the core
 58 of the diffusion model, namely $\text{WFPT}(v, a, z)$. Our approach is somewhat different to
 59 recent proposals by Tuerlinckx (2004) and Voss and Voss (2008) in that our goal is to
 60 consider the faster computation of the WFPT distribution, not the full model with the
 61 additional random variables. It also differs from the approach taken by Wagenmakers, van
 62 der Maas & Grasman (2007), in that we do not place any restrictions on the parameters
 63 (but see Grasman, Wagenmakers & van der Maas, in press, for a more general approach),
 64 and from that of Brown, Ratcliff and Smith (2006) and Diederich and Busemeyer (2003)
 65 who focus on general simulation methods. Our choice to focus on the model at this
 66 level of generality is deliberate – in the “narrow” context of modelling choice and RT in
 67 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.

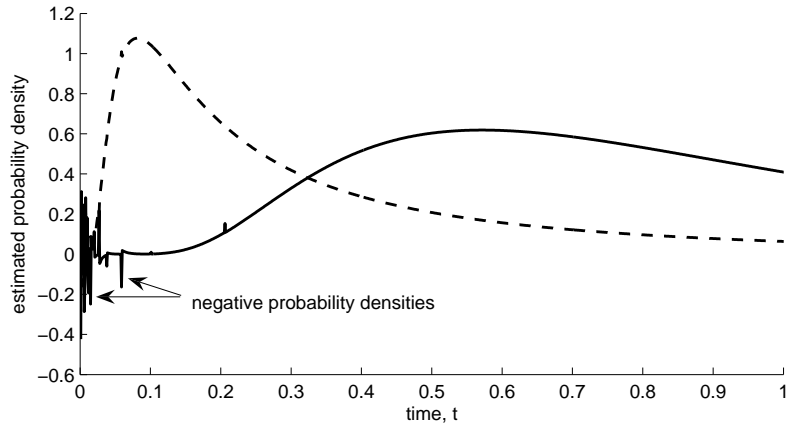


Fig. 2. A deliberately-extreme example of the kinds of pathologies that can occur when WFPT probability densities are calculated inappropriately. In this case, we produce the figure using the method described by Ratcliff & Tuerlinckx (2002) to approximate the cdf, and approximating the pdf via finite 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.

68 sit naturally within the Bayesian hierarchical characterization of the “full” diffusion model
69 (Lee, Fuss & Navarro, 2007; Vandekerckhove, Tuerlinckx & Lee, 2008). A major advantage
70 of the Bayesian approach is that it allows *any* class of extensions to the diffusion model to
71 be handled efficiently using modern computational statistics (e.g., Chen, Shao & Ibrahim
72 2000), so long as the WFPT distribution itself is tractable: in fact, Lee et al (2007) and
73 Vandekerckhove et al (2008) both present simple extensions of this kind, in which explicit
74 psychophysical functions are used to constrain drift rates across experimental conditions.
75 Nevertheless, since the “standard” calculation of the WFPT distribution can sometimes
76 require the evaluation of hundreds of terms in order to avoid pathologies at small RT
77 values (see Figure 2 for an exaggerated example), the kind of large-scale computational
78 methods that have become available for modelling higher-order cognition (e.g., Kemp &
79 Tenenbaum 2008, Griffiths, Steyvers & Tenenbaum, 2007) are currently infeasible. What is
80 required is a simple, pathology-free method for quickly computing the WFPT distribution
81 with near-zero error. We provide such a method in this paper.

82 2 Computing the first-passage time densities

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

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

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

$$92 \quad f(t|v, a, w) = \frac{\pi}{a^2} \exp\left(-vaw - \frac{v^2t}{2}\right) \sum_{k=1}^{\infty} k \exp\left(-\frac{k^2\pi^2t}{2a^2}\right) \sin(k\pi w). \quad (1)$$

93 The probability density at the upper boundary is straightforward to obtain, by setting
 94 $v' = -v$ and $w' = 1 - w$. Algebraically, one nice aspect to the expression is that it factorizes
 95 very simply, allowing the three-parameter density function to be written as follows:

$$96 \quad f(t|v, a, w) = \frac{1}{a^2} \exp\left(-vaw - \frac{v^2t}{2}\right) f\left(\frac{t}{a^2} \mid 0, 1, w\right). \quad (2)$$

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

101 When calculating WFPT densities, a typical approach is to make use of this “large time”
 102 expansion (e.g., Ratcliff 1978; Luce 1986; Ratcliff & Tuerlinckx 2002; Ratcliff & Smith
 103 2004; Tuerlinckx 2004). In terms of the standard case $f(t|0, 1, w)$, we rely on the series:

$$104 \quad f(t|0, 1, w) = \pi \sum_{k=1}^{\infty} k \exp\left(-\frac{k^2\pi^2t}{2}\right) \sin(k\pi w). \quad (3)$$

105 However, Feller also provides a different “small time” representation that is less frequently
 106 used (but see Van Zandt 2000, Van Zandt, Colonius & Proctor 2000, Voss, Rothermund
 107 & Voss 2004). It is given within problem 22 in Feller (1968, ch 14), and is produced
 108 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.

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

$$111 \quad f(t|0, 1, w) = \frac{1}{\sqrt{2\pi t^3}} \sum_{k=-\infty}^{\infty} (w + 2k) \exp\left(-\frac{(w + 2k)^2}{2t}\right). \quad (4)$$

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

$$119 \quad f_{\kappa}^{\ell}(t|0, 1, w) = \pi \sum_{k=1}^{\kappa} k \exp\left(-\frac{k^2\pi^2 t}{2}\right) \sin(k\pi w) \quad (5)$$

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

$$122 \quad f_{\kappa}^s(t|0, 1, w) = \frac{1}{\sqrt{2\pi t^3}} \sum_{k=-\lfloor(\kappa-1)/2\rfloor}^{\lfloor(\kappa-1)/2\rfloor} (w + 2k) \exp\left(-\frac{(w + 2k)^2}{2t}\right), \quad (6)$$

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

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

$$133 \quad |f_{\kappa}(t|0, 1, w) - f(t|0, 1, w)| < \epsilon. \quad (7)$$

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

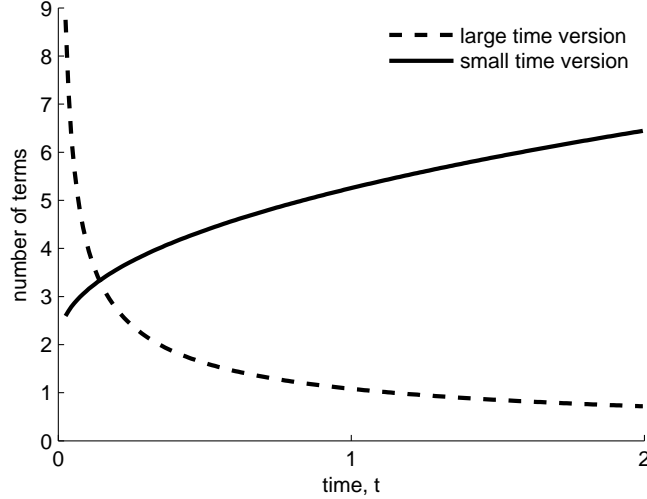


Fig. 3. Number of terms κ required to guarantee a truncation error below .001, for both the large-time (Equation 10) and small-time (Equation 11) versions. For t less than approximately 0.15 the number of terms required is smaller for the small-time expansion, after which time the large- t version is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

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

$$141 \quad E_{\kappa}^{\ell}(t) \leq \frac{1}{\pi t} \exp\left(-\frac{\kappa^2 \pi^2}{2} t\right) \quad (8)$$

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

$$144 \quad E_{\kappa}^s(t) \leq \frac{1}{2\sqrt{2\pi t}} \exp\left(-\frac{(\kappa - 2)^2}{2t}\right), \quad (9)$$

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

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

150 below ϵ the number of terms needed is

$$151 \quad \kappa \geq \sqrt{\frac{-2 \log(\pi t \epsilon)}{\pi^2 t}} \quad (10)$$

152 for the large-time approximation, whereas for the small-time approximation the corre-
153 sponding number of terms is given by

$$154 \quad \kappa \geq 2 + \sqrt{-2t \log(2\epsilon\sqrt{2\pi t})}. \quad (11)$$

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

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

$$166 \quad \lambda(t) = 2 + \sqrt{-2t \log(2\epsilon\sqrt{2\pi t})} - \sqrt{\frac{-2 \log(\pi t \epsilon)}{\pi^2 t}} \quad (12)$$

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

$$173 \quad f(t|0, 1, w) \approx \begin{cases} \frac{1}{\sqrt{2\pi t^3}} \sum_{k=-\lfloor(\kappa-1)/2\rfloor}^{\lceil(\kappa-1)/2\rceil} (w + 2k) \exp\left(\frac{(w + 2k)^2}{2t}\right) & \text{if } \lambda(t) < 0 \\ \pi \sum_{k=1}^{\kappa} k \exp\left(-\frac{k^2\pi^2 t}{2}\right) \sin(k\pi w) & \text{if } \lambda(t) \geq 0 \end{cases}, \quad (13)$$

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

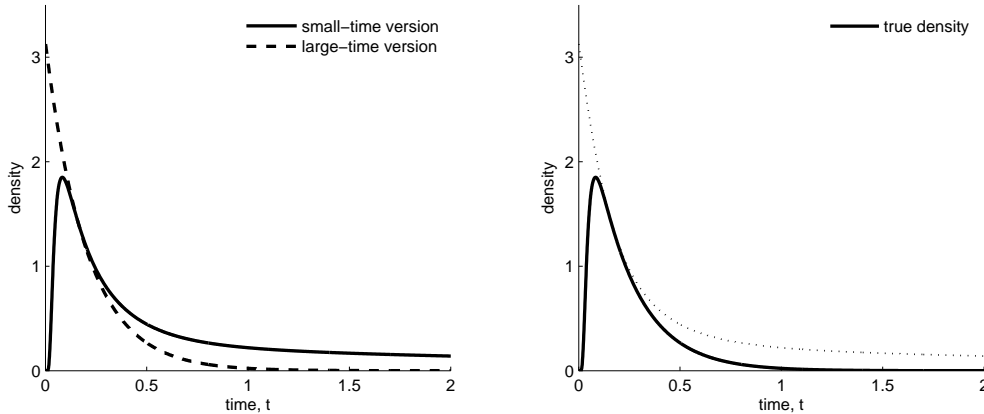


Fig. 4. The $\kappa = 2$ truncations for the small-time and large-time expansions (left), as compared to the true density (right). As one might expect, the large-time version (dashed line) performs very poorly for small values of t when only two terms are used, while the small-time version (solid line)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

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

182 3 Effectiveness of the Method

183 We now turn to some simple tests of the effectiveness of the proposed method for com-
 184 puting the WFPT distribution. As a first test, Figure 5 plots the Wiener first passage
 185 time predictions for a process with $v = 1$, $a = 2$ and $z = .5$. On the left, predictions are
 186 made using the slower “classical” method discussed in Ratcliff & Tuerlinckx’s (2002) pa-
 187 per: namely, to terminate when two successive terms in the cdf calculation remain below
 188 10^{-29} times the current value of the sum. In the middle, predictions are made using the
 189 “fast truncations” presented in this paper (i.e., using Equation 13), with a stringent error
 190 tolerance $\epsilon = 10^{-29}$ so as to roughly mimic the standard applied in the classical version.
 191 The panel on the right shows the difference between the two versions, which is very small.
 192 Although the two methods make very similar predictions, they differ dramatically in the
 193 amount of computation required to do so, as shown in Figure 6, which shows the number
 194 of terms required to compute both versions, as a function of t . At large t the two versions
 195 are comparable, which is what one would expect since for these t values both approaches
 196 rely on Equation 4. At small t , the classical method needs to calculate a very large number

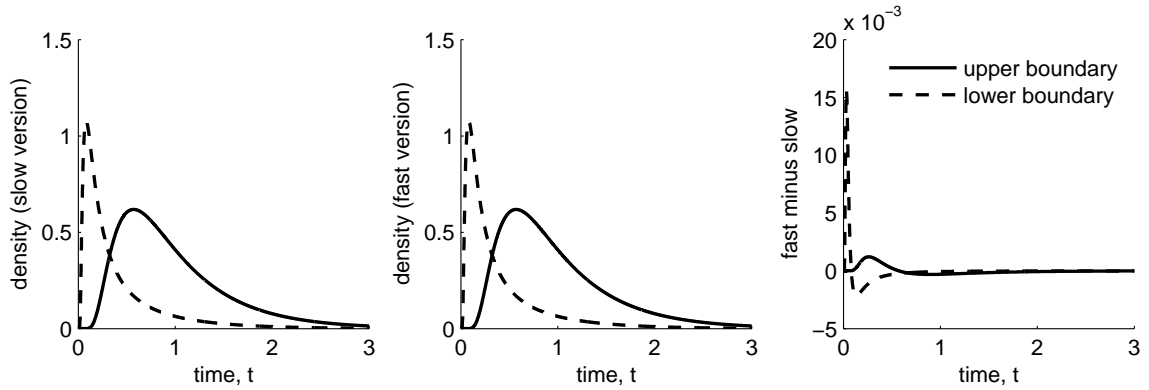


Fig. 5. Illustration of the performance of the approximation, for $v = 1$, $a = 2$ and $z = .5$. The left panel shows model predictions calculated using the “slow” classical method, while the middle panel shows the predictions calculated using the fast truncation method suggested here. As is illustrated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

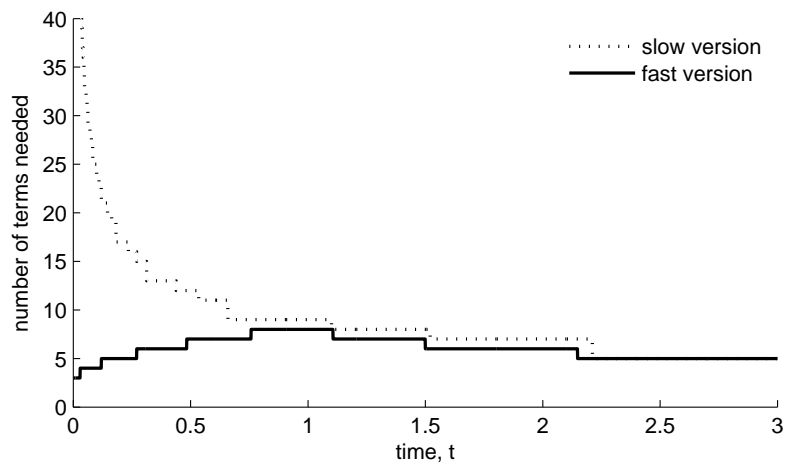


Fig. 6. Illustration of the extent of the speed up: the number of terms required to calculate the slower classical heuristic (dotted) and fast truncation (solid) predictions from Figure 5, in which $\epsilon = 10^{-29}$. Also note that the full extent of the speed up for very small t is masked since 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.

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

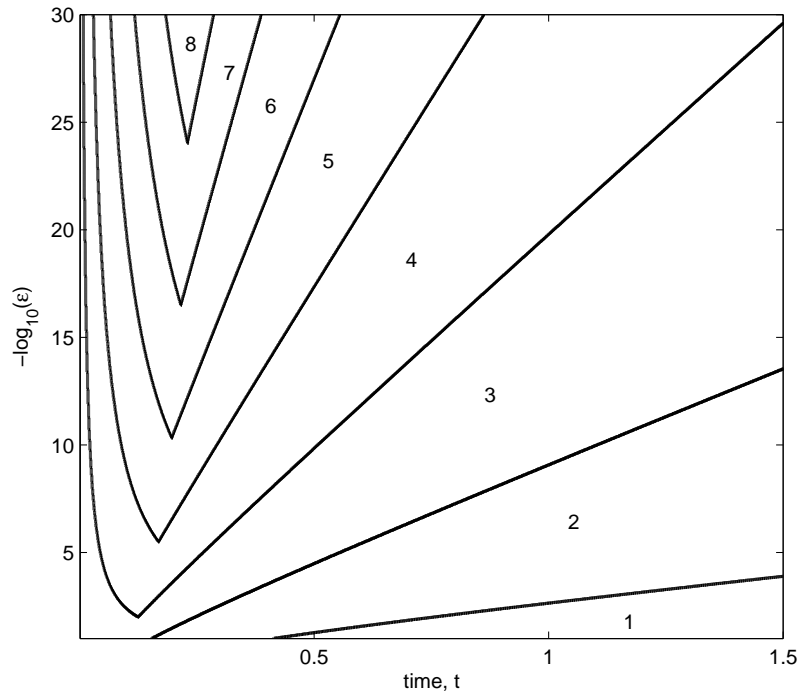


Fig. 7. Number of terms required to compute approximate the standard WFPT densities $f(t|0, 1, w)$, as a function of the time t and the error tolerance ϵ . Since the bounds are derived independently of the parameters v , a and w these results hold quite generally.

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

216 4 Conclusion

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

234 5 Acknowledgements

235 Correspondence concerning this article should be addressed to Daniel Navarro, School of
236 Psychology, University of Adelaide, SA 5005, Australia. Tel +61 8 8303 5265, Fax +61
237 8 8303 3770, E-mail: daniel.navarro@adelaide.edu.au. DJN was supported by an Aus-
238 tralian Research Fellowship (ARC grant DP-0773794). We thank Roger Ratcliff, Francis
239 Tuerlinckx, Trish Van Zandt, E.J. Wagenmakers and three anonymous reviewers for their
240 comments and suggestions. Code associated with this article is available on the first au-
241 thor’s website.

242 6 References

- 243 Brown, S.D., Ratcliff, R., & Smith, P.L. (2006). Evaluating methods for approximating
244 stochastic differential equations. *Journal of Mathematical Psychology*, *50*, 402-410.
- 245 Chen, M., Shao, Q. & Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Compu-*
246 *tation*. New York, NY: Springer.
- 247 Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion
248 models of choice probability, choice response time, and simple response time. *Journal*
249 *of Mathematical Psychology*, *47*(3), 304322.

- 250 Feller, W. (1968). *An Introduction to Probability Theory and its Applications* (Vol. 1, 3rd
251 Ed). New York: Wiley.
- 252 Frazier, P. I. & Yu, A. J. (2008). Sequential hypothesis testing under stochastic deadlines.
253 In J. C. Platt, D. Koller, Y. Singer & S. Roweis (Eds.) *Advances in Neural Information*
254 *Processing Systems, 20* (pp. 465-472). Cambridge, MA: MIT Press
- 255 Grasman, R. P. P. P., Wagenmakers, E.-J., & van der Maas, H. L. J. (in press). On the
256 mean and variance of response times under the diffusion model with an application to
257 parameter estimation. *Journal of Mathematical Psychology*.
- 258 Griffiths, T. L., Steyvers, M. & Tenenbaum, J. B. (2007). Topics in semantic representa-
259 tion. *Psychological Review, 114*, 211-244.
- 260 Kemp, C. & Tenenbaum, J. B. (2008). Learning structural forms. *Proceedings of the*
261 *National Academy of Sciences of the USA, 105*, 10687-10692
- 262 Lee, M. D., Fuss, I. & Navarro, D. J. (2007). A Bayesian approach to diffusion models
263 of decision-making and response time. In B. Schölkopf, J. Platt & T. Hoffman (Eds.)
264 *Advances in Neural Information Processing Systems, 19* (pp. 809-815). Cambridge, MA:
265 MIT Press
- 266 Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organi-*
267 *zation*. Oxford, UK: Oxford University Press.
- 268 Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59-108.
- 269 Ratcliff, R. & Smith, P. L. (2004). A comparison of sequential sampling models for two-
270 choice reaction time. *Psychological Review, 111*, 333-367.
- 271 Ratcliff, R. & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Ap-
272 proaches to dealing with contaminant reaction times and parameter variability. *Psy-*
273 *chonomic Bulletin & Review, 9*, 438-481.
- 274 Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foun-
275 dational primer. *Journal of Mathematical Psychology, 44*, 408-463.
- 276 Smith, P. L. & Van Zandt, T. (2000). Time-dependent Poisson counter models of response
277 latency in simple judgment. *British Journal of Mathematical and Statistical Psychology,*
278 *53*, 293-315.
- 279 Stone, M. (1960). Models for choice-reaction time. *Psychometrika, 25*, 251-260.
- 280 Tuerlinckx, F. (2004). The efficient computation of the cumulative distribution and proba-
281 bility density functions in the diffusion model. *Behavior Research Methods, Instruments,*
282 *& Computers, 36*, 702-716.
- 283 Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2008). A Bayesian approach to diffusion
284 process models of decision-making. In V. Sloutsky, B. Love & K. McRae (Eds.), *Proceed-*
285 *ings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1429-1434).
286 Austin, TX: Cognitive Science Society.
- 287 Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin &*
288 *Review, 7*, 424-465.
- 289 Van Zandt, T., Colonius, H. & Proctor, R. W. (2000). A comparison of two response time
290 models applied to perceptual matching. *Psychonomic Bulletin & Review, 7*, 208-256.
- 291 Vickers, D. (1979). *Decision Processes in Visual Perception*. New York, NY: Academic
292 Press

- 293 Voss, A., Rothermund, K. & Voss J. (2004) Interpreting the parameters of the diffusion
 294 model: An empirical validation. *Memory & Cognition*, 32, 1206-1220
- 295 Voss, A. & Voss, J. (2008). A fast numerical algorithm for the estimation of diffusion
 296 model parameters. *Journal of Mathematical Psychology*, 52, 1-9
- 297 Wagenmakers, E. J., van der Maas, H. L. J. & Grasman, R. P. P. P. (2007). An EZ-
 298 diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, 14,
 299 3-22.
- 300 Wald, A. (1947). *Sequential Analysis*. New York, NY: Wiley.
- 301 Wald, A. & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio
 302 test. *Annals of Mathematical Statistics*, 19, 326-339.

303 A Large-time error bound

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

$$306 \quad g(k, t, w) = k \exp\left(-\frac{k^2 \pi^2 t}{2}\right) \sin(k\pi w), \quad (\text{A.1})$$

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

$$310 \quad E_{\kappa}^{\ell}(t) = \pi \left| \sum_{k=\kappa+1}^{\infty} g(k, t, w) \right|. \quad (\text{A.2})$$

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

$$312 \quad E_{\kappa}^{\ell}(t) \leq \pi \sum_{k=\kappa+1}^{\infty} |g(k, t, w)|. \quad (\text{A.3})$$

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

$$|g(k, t, w)| = k \exp\left(-\frac{k^2 \pi^2 t}{2}\right) |\sin(k\pi w)| \quad (\text{A.4})$$

$$\leq k \exp\left(-\frac{k^2 \pi^2 t}{2}\right). \quad (\text{A.5})$$

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

$$317 \quad E_{\kappa}^{\ell}(t) \leq \pi \sum_{k=\kappa+1}^{\infty} k \exp\left(-\frac{k^2 \pi^2 t}{2}\right). \quad (\text{A.6})$$

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

$$329 \quad \sum_{x=a+1}^b h(x) \leq \int_a^b h(x) dx \leq \sum_{x=a}^{b-1} h(x). \quad (\text{A.7})$$

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

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

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

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

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

336 **B Small-time error bound**

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

$$342 \quad f(t|0, 1, w) = \frac{1}{\sqrt{2\pi t^3}} \sum_{k=-\infty}^{\infty} (w + 2k) \exp\left(-\frac{(w + 2k)^2}{2t}\right). \quad (\text{B.1})$$

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

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

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

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

$$346 \quad S^- = - \sum_{k=1}^{\infty} (-w + 2k) \exp\left(-\frac{(-w + 2k)^2}{2t}\right). \quad (\text{B.4})$$

347 This allows us to rewrite the first passage time density as

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

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

$$350 \quad f_1(t|0, 1, w) = \frac{1}{\sqrt{2\pi t^3}} w \exp\left(-\frac{w^2}{2t}\right), \quad (\text{B.6})$$

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

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

353 and if κ is an odd integer larger than 1 then

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

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

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

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

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

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

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

$$368 \quad S = \sum_{k=\nu/2}^{\infty} (c + 2k) \exp\left(-\frac{(c + 2k)^2}{2t}\right). \quad (\text{B.11})$$

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

$$371 \quad g(k) = (c + 2k) \exp\left(-\frac{(c + 2k)^2}{2t}\right). \quad (\text{B.12})$$

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

$$378 \quad \int_{\nu/2}^{\infty} (c + 2k) \exp\left(-\frac{(c + 2k)^2}{2t}\right) dk \leq \sum_{k=\nu/2}^{\infty} (c + 2k) \exp\left(-\frac{(c + 2k)^2}{2t}\right) \quad (\text{B.13})$$

379 and

$$380 \quad \sum_{k=\nu/2}^{\infty} (c + 2k) \exp\left(-\frac{(c + 2k)^2}{2t}\right) \leq \int_{(\nu-2)/2}^{\infty} (c + 2k) \exp\left(-\frac{(c + 2k)^2}{2t}\right) dk. \quad (\text{B.14})$$

381 Evaluation of the integrals gives

$$382 \quad \frac{t}{2} \exp\left(-\frac{(c + \nu)^2}{2t}\right) \leq \sum_{k=\nu/2}^{\infty} (c + 2k) \exp\left(-\frac{(c + 2k)^2}{2t}\right) \leq \frac{t}{2} \exp\left(-\frac{(c + \nu - 2)^2}{2t}\right). \quad (\text{B.15})$$

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

$$385 \quad E_{\kappa}^s(t) \leq \frac{1}{2\sqrt{2\pi t}} \left[\exp\left(-\frac{(w + \kappa)^2}{2t}\right) - \exp\left(-\frac{(w + \kappa + 2)^2}{2t}\right) \right]. \quad (\text{B.16})$$

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

388 is always the larger one). This gives

$$389 \quad E_{\kappa}^s(t) \leq \frac{1}{2\sqrt{2\pi t}} \left[\exp\left(-\frac{(-w + \kappa - 1)^2}{2t}\right) - \exp\left(-\frac{(w + \kappa + 1)^2}{2t}\right) \right]. \quad (\text{B.17})$$

390 To simplify matters, we note that since the exponential function is positive valued, we
 391 can set 0 as an upper bound on the $-\exp(x)$ terms in both equations. Similarly, since
 392 $0 \leq w \leq 1$ we can set it to the worst possible value ($w = 0$ for even κ and $w = 1$ for odd
 393 κ). By doing so we observe that, irrespective of whether κ is odd or even,

$$394 \quad E_{\kappa}^s(t) \leq \frac{1}{2\sqrt{2\pi t}} \exp\left(-\frac{(\kappa - 2)^2}{2t}\right), \quad (\text{B.18})$$

395 which is the error bound in Equation 9. As observed above, this bound only holds for
 396 sufficiently large κ , which in this case corresponds to $\kappa > \sqrt{t} - c$. Noting that the smallest
 397 value of c used in any of the expressions is -1 , we can state that this bound holds for
 398 $\kappa > \sqrt{t} + 1$. In short, this bound holds for small t , and in those cases the truncation error
 399 is provably small.

400 C MATLAB code

```

401 function p=wfpt(t,v,a,z,err)
402
403 tt=t/(a^2); % use normalized time
404 w=z/a; % convert to relative start point
405
406 % calculate number of terms needed for large t
407 if pi*tt*err<1 % if error threshold is set low enough
408     kl=sqrt(-2*log(pi*tt*err)./(pi^2*tt)); % bound
409     kl=max(kl,1/(pi*sqrt(tt))); % ensure boundary conditions met
410 else % if error threshold set too high
411     kl=1/(pi*sqrt(tt)); % set to boundary condition
412 end
413
414 % calculate number of terms needed for small t
415 if 2*sqrt(2*pi*tt)*err<1 % if error threshold is set low enough
416     ks=2+sqrt(-2*tt.*log(2*sqrt(2*pi*tt)*err)); % bound
417     ks=max(ks,sqrt(tt)+1); % ensure boundary conditions are met
418 else % if error threshold was set too high
419     ks=2; % minimal kappa for that case

```

```

420 end
421
422 % compute f(tt|0,1,w)
423 p=0; %initialize density
424 if ks<k1 % if small t is better...
425 K=ceil(ks); % round to smallest integer meeting error
426     for k=-floor((K-1)/2):ceil((K-1)/2) % loop over k
427         p=p+(w+2*k)*exp(-((w+2*k)^2)/2/tt); % increment sum
428     end
429     p=p/sqrt(2*pi*tt^3); % add constant term
430
431 else % if large t is better...
432 K=ceil(k1); % round to smallest integer meeting error
433 for k=1:K
434 p=p+k*exp(-(k^2)*(pi^2)*tt/2)*sin(k*pi*w); % increment sum
435 end
436 p=p*pi; % add constant term
437 end
438
439 % convert to f(t|v,a,w)
440 p=p*exp(-v*a*w -(v^2)*t/2)/(a^2);

```