

## ON DURBIN'S SERIES FOR THE DENSITY OF FIRST PASSAGE TIMES

P. ZIPKIN,\* *Duke University*

### Abstract

Durbin (1992) derived a convergent series for the density of the first passage time of a Wiener process to a curved boundary. We show that the successive partial sums of this series can be expressed as the iterates of the standard substitution method for solving an integral equation. The calculation is thus simpler than it first appears. We also show that, under a certain condition, the series converges uniformly. This strengthens Durbin's result of pointwise convergence. Finally, we present a modified procedure, based on scaling, which sometimes works better. These approaches cover some cases that Durbin did not.

*Keywords:* First passage time; Wiener process

2010 Mathematics Subject Classification: Primary 60J65  
Secondary 65N21; 58J35

### 1. Introduction

The *first passage time* of a real-valued stochastic process  $W = \{W(t) : t \geq 0\}$  to a boundary  $b = b(t)$  is

$$\tau = \inf\{t > 0 : W(t) \geq b(t)\}.$$

We aim to find the density  $f$  of  $\tau$ , given  $b$ . We focus on the case where  $W$  is a standard Wiener process starting at  $W(0) = 0$ . Assume that  $b$  is continuously differentiable for  $t > 0$ , with  $b(0+) \equiv \lim_{t \rightarrow 0} \{b(t)\} > 0$ . (Additional conditions on  $b$  are discussed below.)

This problem has numerous applications in the physical and social sciences. The exact solution is known only in a few special cases. Of these, the simplest is where  $b$  is a straight line. In this case,  $f$  is an inverse Gaussian density. Consequently, there is great interest in approximations and numerical methods. The tangent approximation, proposed by Daniels [1] and Strassen [8], is perhaps the simplest approach. It approximates  $b$  locally by a straight line. It has been studied intensively by Lerche [4], among others. Several refinements have been suggested, including Lerche's second-order correction and the hazard rate tangent approximation of Roberts and Shortland [7].

Durbin [2] derived, under a certain condition, a series representation of  $f$ , the first term of which is the tangent approximation. It thus provided a useful link between exact and approximate methods. The tangent approximation is exact for linear  $b$ , and indeed the series terminates after the first term in this case. For nonlinear  $b$ , the remaining terms can be interpreted as corrections for curvature. These terms, however, appear to be increasingly difficult to compute.

We show that the successive partial sums of this series can be expressed as the iterates of the standard substitution method for solving an integral equation. (This equation too is due

---

Received 14 October 2010; revision received 23 February 2011.

\* Postal address: Fuqua School of Business, Duke University, Durham, NC, USA.

Email address: paul.zipkin@duke.edu

to Durbin, but he used it for other purposes.) The calculation is thus much simpler than it first appears. We also show that, under a certain condition, different from Durbin's, the series converges uniformly. This strengthens Durbin's result of pointwise convergence. We then present a modified procedure, based on scaling, along with a modified convergence condition, which sometimes works better. These conditions, moreover, cover some interesting cases that Durbin did not. Finally, this approach is more flexible; it can start with any estimate of the density, including the refined tangent approximations noted above.

### 2. Series

For any positive integer  $j$ , let  $\mathbf{t} = (t_i)_{i=0}^{j-1}$  denote a vector of positive times, decreasing in  $i$ , and let  $g(\mathbf{t})$  denote the joint density of  $W$  at times  $\mathbf{t}$  on the boundary, that is, the density of the event  $\{W(t_i) = b(t_i) : 0 \leq i < j \mid W(0) = 0\}$ . For  $j = 1$ , we can write  $\mathbf{t} = t_0 = t$ , so  $g(\mathbf{t})$  is just the univariate density of  $W(t)$ , evaluated at  $b(t)$ . Also, for  $0 \leq s < t$ , define  $g(t \mid s)$  as the conditional density of  $\{W(t) = b(t) \mid W(s) = b(s)\}$ . Define

$$\delta(s, t) = \frac{b(t) - b(s)}{t - s} - b'(t), \quad \delta(t) = \frac{b(t)}{t} - b'(t).$$

Now define

$$q_1(t) = q(t) = \delta(t)g(t),$$

and, for  $j > 1$ ,

$$q_j(t) = \int_0^{t_0} \int_0^{t_1} \cdots \int_0^{t_{j-2}} \delta(t_{j-1}) \left[ \prod_{i=1}^{j-1} \delta(t_i, t_{i-1}) \right] g(\mathbf{t}) dt_{j-1} \cdots dt_2 dt_1, \tag{1}$$

where  $t_0 = t$ . Also, let

$$f_k(t) = \sum_{j=1}^k (-1)^{j-1} q_j(t).$$

This is Durbin's series. The first term  $q(t)$  is precisely the tangent approximation of  $f$ .

We now aim to simplify this calculation. Let  $\mathcal{B}$  denote the integral operator defined by

$$(\mathcal{B}h)(t) = \int_0^t h(s)\delta(s, t)g(t \mid s) ds,$$

and let  $\mathcal{A} = q - \mathcal{B}$ , that is,

$$(\mathcal{A}h)(t) = q(t) - \int_0^t h(s)\delta(s, t)g(t \mid s) ds.$$

(For now, the domain of both operators is the set  $\mathcal{C}$  of continuous functions  $h$  on  $[0, \infty)$ .)

**Lemma 1.** For  $j > 0$ ,  $q_{j+1} = \mathcal{B}q_j$ .

*Proof.* By the Markov property,

$$g(\mathbf{t}) = g(t_{j-1}) \prod_{i=1}^{j-1} g(t_{i-1} \mid t_i).$$

Consequently,

$$\begin{aligned}
 q_{j+1}(t_0) &= \int_0^{t_0} \int_0^{t_1} \cdots \int_0^{t_{j-1}} \delta(t_j) \left[ \prod_{i=1}^j \delta(t_i, t_{i-1}) g(t_{i-1} | t_i) \right] g(t_j) dt_j \cdots dt_2 dt_1 \\
 &= \int_0^{t_0} \delta(t_1, t_0) g(t_0 | t_1) \int_0^{t_1} \delta(t_2, t_1) g(t_1 | t_2) \cdots \int_0^{t_{j-1}} \delta(t_j) g(t_j) dt_j \cdots dt_2 dt_1 \\
 &= \int_0^{t_0} \delta(t_1, t_0) g(t_0 | t_1) q_j(t_1) dt_1 \\
 &= (\mathcal{B}q_j)(t_0).
 \end{aligned}$$

**Proposition 1.** For  $k \geq 0$ ,  $f_{k+1} = \mathcal{A}f_k$ , with  $f_0 = 0$ .

*Proof.* We have

$$\begin{aligned}
 f_{k+1} &= \sum_{j=1}^{k+1} (-1)^{j-1} q_j \\
 &= q - \sum_{j=2}^{k+1} (-1)^j q_j \\
 &= q - \sum_{i=1}^k (-1)^{i-1} q_{i+1} \\
 &= q - \sum_{i=1}^k (-1)^{i-1} \mathcal{B}q_i \\
 &= q - \mathcal{B} \left( \sum_{i=1}^k (-1)^{i-1} q_i \right) \\
 &= q - \mathcal{B}f_k \\
 &= \mathcal{A}f_k.
 \end{aligned}$$

The calculation of  $q_{j+1}$  appears from (1) to require a  $j$ -dimensional integral for each  $t$ . Lemma 1 shows, however, that it entails just a one-dimensional integral, using the previous term  $q_j$ . Likewise, according to the proposition, the entire partial sum  $f_{k+1}$  can be obtained from the previous one  $f_k$  by a one-dimensional integral.

Consider the integral equation

$$h = \mathcal{A}h. \tag{2}$$

Durbin showed that the true  $f$  solves this equation. (This is his Equation (10), although it looks different. We have stated it in a form closer to Peskir’s [5] equivalent Equation (5.4).) The sequence  $\{f_k\}$  can thus be viewed as the outcome of the substitution algorithm applied to (2), starting with  $f_0 = 0$ .

### 3. Convergence

Durbin assumed that  $\delta(t) > 0$ . (It appears, however, that this condition can be dropped. He used it in a proof of (2), but as he mentioned, there are other proofs that do not require it. We will revive it, for a different reason, in the next section.)

Define

$$r_k(t) = f(t) - f_k(t), \quad \gamma(t) = \sup\{|\delta(r, s)| : 0 \leq r < s \leq t\}.$$

Assume that  $\gamma(t)$  is finite for all  $t$ . Durbin proved that

$$|r_k(t)| \leq \bar{r}_k(t) \equiv \sqrt{\frac{2\pi}{t} \frac{[\eta(t)]^k}{\Gamma(k/2)}},$$

where  $\eta(t) = \gamma(t)\sqrt{t/2}$ . For each  $t$ ,  $\bar{r}_k(t)$  converges to 0 as  $k \rightarrow \infty$ , and, therefore,  $f_k(t) \rightarrow f(t)$ .

This result is remarkable in several ways. First, it is quite general; it places no restrictions on  $b$ , besides those above. Second, the error bound  $\bar{r}_k(t)$  need not decrease monotonically in  $k$ , because  $\eta(t)$  may be greater than 1. However,  $\bar{r}_k(t)$  does ultimately decrease for large  $k$ , due to the factor  $\Gamma(k/2)$ . In fact, it decreases faster than geometrically. The form of  $\bar{r}_k(t)$  is like that of the terms in the series expansion of  $e^x$ , and it decays similarly. This point can be seen even more clearly by examining the even terms:

$$\bar{r}_{2k}(t) = \sqrt{\frac{2\pi}{t} \frac{[\eta^2(t)]^k}{\Gamma(k)}}.$$

Third, as Durbin pointed out,  $|\delta(r, s)|$  can be viewed as a measure of the curvature of  $b$  between  $r$  and  $s$ . So,  $\gamma(t)$  is a summary measure of its nonlinearity up to  $t$ . Specifically, if  $b$  is linear over  $[0, t]$  then  $\gamma(t) = 0$ . Thus, the result indicates that the convergence should be fast when the slope of  $b$  changes slowly.

The assumption that  $\gamma(t)$  be finite is of course necessary for this approach. Consider  $b(t) = b_0 + ct^\alpha$ , where  $b_0, c$ , and  $\alpha$  are constants, with  $b_0 > 0, c \neq 0$ , and  $0 < \alpha < 1$ . Then,

$$|\delta(r, s)| = |c| \left( \frac{s^\alpha - r^\alpha}{s - r} - \alpha s^{\alpha-1} \right) = |c| \left( \frac{1 - (r/s)^\alpha}{1 - r/s} - \alpha \right) s^{\alpha-1}.$$

For any fixed  $s, |\delta(r, s)|$  is maximized at  $r = 0$ , and

$$|\delta(0, s)| = |c|(1 - \alpha)s^{\alpha-1}.$$

As  $s \rightarrow 0, |\delta(0, s)| \rightarrow \infty$ . Thus,  $\gamma(t) = \infty$  for all  $t > 0$ . The convergence argument thus does not work here. We will see below that, nevertheless, the algorithm does converge for this  $b$ , at least for most values of  $\alpha$ . (For  $\alpha > 1, \gamma(t) = |c|(\alpha - 1)t^{\alpha-1}$ , which is indeed finite.)

Even for finite  $\gamma(t)$ , the convergence here is pointwise, not uniform over  $t$ . The quantity  $\gamma(t)$  is increasing in  $t$ , and, therefore,  $\eta(t)$  increases at least as fast as  $\sqrt{t/2}$ . So, according to this analysis, the calculation may take a while to ‘settle down’ for large  $t$ . We now show that, when  $b$  is nearly linear, in a sense different from that measured by  $\gamma(t)$ , the convergence is indeed uniform.

Consider the subspace  $\mathfrak{B}$  of  $\mathfrak{C}$  comprising *bounded*, continuous functions  $h = h(t)$ , that is, such that

$$\|h\| \equiv \sup\{|h(t)| : t > 0\} < \infty.$$

Assume that  $g \in \mathfrak{B}$ . We aim to show that, under a certain condition,  $\mathcal{A}$  is a *contraction* on  $\mathfrak{B}$ . That is, for some constant  $0 \leq \rho < 1$ , and any two functions  $h^1$  and  $h^2 \in \mathfrak{B}$ ,

$$\|\mathcal{A}(h^2 - h^1)\| \leq \rho \|h^2 - h^1\|.$$

(Peskir [5] verified this property for a different integral equation, under an entirely different condition.) This property guarantees uniform, geometric convergence to  $f$ . Furthermore,  $f$  is the unique solution in  $\mathfrak{B}$  to the integral equation (2). See, e.g. [3, Chapter 5].

Define

$$\beta(t) = \int_0^t |\delta(s, t)|g(t | s) ds, \quad \beta = \sup\{\beta(t) : t > 0\}.$$

Evidently,  $\beta$  is an alternative summary measure of nonlinearity.

**Proposition 2.** *If  $\beta < 1$  then  $\mathcal{A}$  is a contraction with  $\rho = \beta$ .*

*Proof.* We have

$$\begin{aligned} |(\mathcal{A}(h^2 - h^1))(t)| &= \left| \int_0^t [h^2(s) - h^1(s)]\delta(s, t)g(t | s) ds \right| \\ &\leq \int_0^t |h^2(s) - h^1(s)||\delta(s, t)|g(t | s) ds \\ &\leq \beta(t)\|h^2 - h^1\| \\ &\leq \beta\|h^2 - h^1\|. \end{aligned}$$

It appears difficult in general to compute  $\beta$ , but it can be simplified in some cases. Reconsider the example above:  $b(t) = b_0 + ct^\alpha$  with  $0 < \alpha < 1$ . It is not difficult to check that indeed  $q \in \mathfrak{B}$ . Let  $\phi$  denote the standard normal density. Then

$$\begin{aligned} \beta(t) &= \int_0^t |c| \left( \frac{t^\alpha - s^\alpha}{t - s} - \alpha t^{\alpha-1} \right) \frac{1}{\sqrt{t-s}} \phi \left( |c| \frac{t^\alpha - s^\alpha}{\sqrt{t-s}} \right) ds \\ &= \int_0^t |c| \frac{t^\alpha}{t} \left( \frac{1 - (s/t)^\alpha}{1 - s/t} - \alpha \right) \frac{1}{t^{1/2}} \frac{1}{\sqrt{1-s/t}} \phi \left( |c| \frac{t^\alpha}{t^{1/2}} \frac{1 - (s/t)^\alpha}{\sqrt{1-s/t}} \right) ds \\ &= \int_0^1 \left( \frac{1 - x^\alpha}{1 - x} - \alpha \right) \frac{1}{\sqrt{1-x}} |c| t^{\alpha-1/2} \phi \left( |c| t^{\alpha-1/2} \frac{1 - x^\alpha}{\sqrt{1-x}} \right) dx. \end{aligned}$$

Evidently, this depends on  $t$  and  $c$  only through the parameter  $\theta = |c|t^{\alpha-1/2}$ . Let us write

$$\begin{aligned} \bar{\beta}(\theta) &= \int_0^1 \left( \frac{1 - x^\alpha}{1 - x} - \alpha \right) \frac{1}{\sqrt{1-x}} \theta \phi \left( \theta \frac{1 - x^\alpha}{\sqrt{1-x}} \right) dx, \tag{3} \\ \bar{\beta} &= \sup\{\bar{\beta}(\theta) : \theta > 0\}. \end{aligned}$$

For  $\alpha \neq \frac{1}{2}$ , clearly,  $\beta = \bar{\beta}$ , and this is the same for all  $c$ . For  $\alpha = \frac{1}{2}$ ,  $\theta = |c|$ , and  $\beta(t)$  is independent of  $t$ , so  $\bar{\beta}$  gives the worst case of  $\beta$  over all values of  $c$ . In both cases, for fixed  $\alpha$ , if  $\bar{\beta} < 1$  then  $\mathcal{A}$  is a contraction for all  $c$ , with  $\rho \leq \bar{\beta}$ .

Also, we can show that  $\bar{\beta}$  is decreasing as a function of  $\alpha$ . (For any fixed  $x$  and  $\theta$ , the integrand in (3) is decreasing in  $\alpha$ .) So, if we can verify that  $\bar{\beta} < 1$  for some particular  $\alpha$  then the same is true for all larger  $\alpha < 1$ . (It would be nice to determine the limit as  $\alpha \rightarrow 0$ , but we have not yet been able to do so.)

It appears that the integral in (3) cannot be further simplified. It is not difficult to compute it numerically, however, and we have done so;  $\bar{\beta}(\theta)$  appears to be unimodal in all cases. (This is not a surprise. It is clearly true of the integrand for each fixed  $v$ .) For  $\alpha = \frac{1}{32}$ , the mode occurs at about  $\theta^* \approx 22.5$ , with  $\bar{\beta} \approx 0.10$ . This is well below 1, and so  $\mathcal{A}$  is a contraction for

TABLE 1: Contraction test:  $b(t) = b_0 + ct^\alpha$ .

$\alpha$	$\theta^*$	$\bar{\beta}$
2	1.00	0.16
4	0.83	0.54
6	0.76	0.93
8	0.72	1.33

any  $c$  and all  $\alpha \geq \frac{1}{32}$ . (We also computed  $\bar{\beta}$  for a couple of larger values of  $\alpha$ , to check for consistency. For  $\alpha = \frac{1}{8}$ ,  $\theta^* \approx 6.0$  and  $\bar{\beta} \approx 0.09$ . For  $\alpha = \frac{1}{2}$ ,  $\theta^* \approx 1.9$  and  $\bar{\beta} \approx 0.06$ .)

It is worth noting that the results will be even better, in some cases, when  $b$  has nonzero asymptotic slope. Consider the more general function  $b(t) = b_0 + c(b_1t + t^\alpha)$ , where  $b_1 \geq 0$ . What is the effect of  $b_1$  on  $\beta(t)$ ? Observe that  $\delta(s, t)$  is independent of  $b_1$ , and the  $\phi$  term in  $g(t | s)$  becomes

$$\phi\left(|c| \frac{t^\alpha - s^\alpha + b_1(t - s)}{\sqrt{t - s}}\right).$$

This is decreasing in  $b_1$  for all  $s < t$ . Consequently, the same is true of  $\beta(t)$  itself and of  $\beta$ . So, if  $\beta < 1$  for  $b_1 = 0$  (as we have already shown for  $\frac{1}{32} \leq \alpha < 1$ ) then the same is true for all  $b_1 \geq 0$ . This logic applies to any function of the form  $\tilde{b}(t) = b_0 + c[b_1t + \tilde{b}(t)]$ , where  $\tilde{b}(t)$  is nondecreasing with  $\tilde{b}(0+) = 0$ .

Returning to  $b_1 = 0$ , for  $\alpha > 1$ , the derivation and interpretation of (3) remain valid, with a change of sign to account for the fact that  $\delta(s, t) < 0$ . Table 1 records some results. Evidently,  $\bar{\beta} < 1$  for some cases, but not all.

### 4. Scaling

These mixed results lead us to explore a modification of the approach above. Suppose that we know or suspect that  $f$  behaves badly, in a way that may cause numerical difficulties. Suppose that we know that a function  $v = v(t)$ , positive for  $t > 0$ , behaves in this same way. One possible remedy is to work with a scaled version of  $f$ , namely  $f^\sim = f/v$ . This is a standard technique of analysis.

In terms of scaled functions, the integral equation (2) now reads

$$\begin{aligned} h^\sim(t) &= \frac{h(t)}{v(t)} \\ &= \frac{q(t)}{v(t)} - \frac{1}{v(t)} \int_0^t h(s)\delta(s, t)g(t | s) ds \\ &= \frac{q(t)}{v(t)} - \frac{1}{v(t)} \int_0^t v(s)h^\sim(s)\delta(s, t)g(t | s) ds. \end{aligned}$$

Defining

$$\begin{aligned} q^\sim(t) &= \frac{q(t)}{v(t)}, & \delta^\sim(s, t) &= \frac{v(s)\delta(s, t)}{v(t)}, & (\mathcal{B}^\sim h^\sim)(t) &= \int_0^t h^\sim(s)\delta^\sim(s, t)g(t | s) ds, \\ & & & & \text{and } \mathcal{A}^\sim &= q^\sim - \mathcal{B}^\sim, \end{aligned}$$

the equation can be written as

$$h^\sim = \mathcal{A}^\sim h^\sim.$$

The substitution algorithm now becomes  $f_0^\sim = 0$ ,  $f_{k+1}^\sim = \mathcal{A}^\sim f_k^\sim$ .

This is a different calculation. We can verify that it works, in the sense above, by showing that  $q^\sim \in \mathfrak{B}$  and  $\mathcal{A}^\sim$  is a contraction. Define

$$\beta^\sim(t) = \int_0^t |\delta^\sim(s, t)|g(t | s) ds, \quad \beta^\sim = \sup\{\beta^\sim(t) : t > 0\}.$$

Then, by a minor adjustment to its proof, Proposition 2 holds with  $\beta^\sim$  replacing  $\beta$  and  $\mathcal{A}^\sim$  replacing  $\mathcal{A}$ . Again, the key condition is  $\beta^\sim < 1$ .

All this holds for any positive function  $v$ . We do know one that behaves something like  $f$ , namely, the tangent approximation  $q$  itself. Strassen [8] demonstrated that  $q(t) \sim f(t)$  for small  $t$ . (See also [6].) For larger  $t$ , the situation is less clear; Lerche [4] included some partial results. Of course, to set  $v = q$  requires  $q(t) > 0$  for  $t > 0$  (that is,  $\delta(t) > 0$ , as Durbin assumed). This is true in many cases. For example, it holds when  $b$  is concave.

When we can and do set  $v = q$ , we immediately get  $q^\sim = 1 \in \mathfrak{B}$ . Also, suppose that the sign of  $\delta(s, t)$  does not fluctuate for  $s < t$  near 0. (This holds when  $b(t)$  is either consistently concave or consistently convex for small  $t$ .) We have  $\lim_{t \rightarrow 0}\{f^\sim(t)\} = \lim_{t \rightarrow 0}\{q^\sim(t)\} = 1$ , and so

$$\begin{aligned} 0 &= \lim_{t \rightarrow 0}\{|f^\sim(t) - (\mathcal{A}^\sim f^\sim)(t)|\} \\ &= \lim_{t \rightarrow 0}\left\{\left|\int_0^t f^\sim(s)\delta^\sim(s, t)g(t | s) ds\right|\right\} \\ &= \lim_{t \rightarrow 0}\left\{\int_0^t f^\sim(s)|\delta^\sim(s, t)|g(t | s) ds\right\} \\ &\geq \lim_{t \rightarrow 0}\left\{\inf\{f^\sim(s) : 0 < s < t\} \int_0^t |\delta^\sim(s, t)|g(t | s) ds\right\} \\ &= \lim_{t \rightarrow 0}\{\inf\{f^\sim(s) : 0 < s < t\}\} \lim_{t \rightarrow 0}\left\{\int_0^t |\delta^\sim(s, t)|g(t | s) ds\right\} \\ &= 1 \cdot \lim_{t \rightarrow 0}\{\beta^\sim(t)\}. \end{aligned}$$

Since  $\beta^\sim(t) \geq 0$ , we conclude that  $\lim_{t \rightarrow 0}\{\beta^\sim(t)\} = 0$ . So,  $\beta^\sim(t)$  at least starts small.

It is not always true, however, that  $q > 0$ . In particular, it is not true for the example above,  $b(t) = b_0 + ct^\alpha$ , when  $c > 0$  and  $\alpha > 1$ . Here,

$$\delta(t) = \frac{b_0}{t} - c(\alpha - 1)t^{\alpha-1},$$

which is positive for small  $t$ , but crosses 0 at a certain positive value of  $t$  and remains negative thereafter. For such cases, we propose that  $v(t) = |\delta(0, t)|g(t | 0)$ . This is like  $|q(t)|$  but without the terms involving  $b_0$ . Finally, we make one small adjustment, explained below. Redefine  $v(t) = \sqrt{t}|\delta(0, t)|g(t | 0)$ . (For the example, it is still true that  $q^\sim \in \mathfrak{B}$ , due to the term  $b_0/\sqrt{t}$  in the exponent of  $g(t)$ .)

For the example,  $\beta^\sim(t)$  is the same as  $\beta(t)$ , except that the integral includes the additional factor

$$\frac{v(s)}{v(t)} = \frac{s^{\alpha-1/2}\phi(|c|s^{\alpha-1/2})/\sqrt{s}}{t^{\alpha-1/2}\phi(|c|t^{\alpha-1/2})/\sqrt{t}}.$$

Letting  $x = s/t$ , this becomes

$$\frac{v(tx)}{v(t)} = \frac{(tx)^{\alpha-1}\phi(|c|(tx)^{\alpha-1/2})}{t^{\alpha-1}\phi(|c|t^{\alpha-1/2})} = \frac{x^{\alpha-1}\phi(|c|t^{\alpha-1/2}x^{\alpha-1/2})}{\phi(|c|t^{\alpha-1/2})}.$$

TABLE 2: Contraction test with scaling:  $b(t) = b_0 + ct^\alpha$ .

$\alpha$	$\theta^*$	$\bar{\beta}^-$
2	2.4	0.15
4	1.16	0.14
6	0.86	0.14
8	0.72	0.14

Again, this depends on  $t$  and  $c$  only through  $\theta = |c|t^{\alpha-1/2}$ , and

$$\frac{v(tx)}{v(t)} = \frac{x^{\alpha-1}\phi(\theta x^{\alpha-1/2})}{\phi(\theta)}$$

Thus, we can define

$$\bar{\beta}^-(\theta) = \int_0^1 \frac{x^{\alpha-1}\phi(\theta x^{\alpha-1/2})}{\phi(\theta)} \left| \frac{1-x^\alpha}{1-x} - \alpha \left| \frac{1}{\sqrt{1-x}} \theta \phi \left( \theta \frac{1-x^\alpha}{\sqrt{1-x}} \right) \right| \right| dx,$$

$$\bar{\beta}^- = \sup\{\bar{\beta}^-(\theta) : \theta > 0\},$$

and use them as above.

This formula suggests why the scaling approach might work. The largest values of the second factor in the integrand (derived from  $\delta(tx, t)$ ) are those for small  $x$ . The term  $x^{\alpha-1}$  in the new factor helps to suppress those values. (This is why we guessed that the  $\sqrt{t}$  adjustment above might help. Otherwise, the new term would be  $x^{\alpha-3/2}$ , which would have a weaker effect. It would even magnify the large values for  $1 < \alpha < \frac{3}{2}$ . Anyway, we tried the calculation with and without the adjustment, and it does help a bit.)

We present the results in Table 2. As can be seen, we obtain significant improvements, especially for larger  $\alpha$ .

This approach also allows us to relax the assumption that  $b(0+) > 0$  made at the outset. This assumption does simplify the discussion. But, as a glance at the literature reveals, there are interesting cases with  $b(0+) = 0$ . A broader condition is that  $b$  be an *upper function*, namely, that  $\tau$  has a nondegenerate distribution, specifically,  $\Pr\{\tau > 0\} = 1$ . It is sufficient that  $b(0+) > 0$ , but it is also possible that  $b(0+) = 0$ , provided that  $b(t)$  grows very fast for small  $t$ . (The precise condition is rather technical. See [6].) Examples include  $b(t) = ct^\alpha$  for  $c > 0$  and  $0 < \alpha < \frac{1}{2}$ , and  $b(t) = c\sqrt{-t \ln(t)}$  for  $c > 0$ . (Of course, the latter formula is valid only for  $t \leq 1$ . We mean a function that behaves in this way for small  $t$ . One such function is given below.) In some of these cases,  $q \in \mathfrak{B}$ . This is so for the first example for all  $c > 0$ , and also the second provided that  $c > \sqrt{2}$ . When this condition holds, the analysis of the previous section remains valid. Specifically, the calculation of  $\bar{\beta}$  there for power functions does not depend at all on  $b_0$ . Those results therefore apply even to the case of  $b_0 = 0$ .

In other cases,  $q$  is *not* bounded, specifically  $\lim_{t \rightarrow 0}\{q(t)\} = \infty$ . This is so for  $b(t) = c\sqrt{-t \ln(t)}$  when  $0 < c \leq \sqrt{2}$ . As indicated above, this implies that  $\lim_{t \rightarrow 0}\{f(t)\} = \infty$ . A scaling approach is essential in such cases.

Consider specifically

$$b(t) = c \begin{cases} \sqrt{-t \ln(t)}, & 0 \leq t \leq 1/e, \\ \sqrt{\frac{1}{e}}, & t > 1/e, \end{cases}$$

TABLE 3: Contraction test with scaling:  $b(t) = c\sqrt{-t \ln(t)}$ .

$c$	$t^*$	$\beta^\sim$
$(\sqrt{2})^{-4} = \frac{1}{4}$	0.000 000 09	0.75
$(\sqrt{2})^{-3}$	0.004	0.74
$(\sqrt{2})^{-2} = \frac{1}{2}$	1/e	0.74
$(\sqrt{2})^{-1}$	1/e	0.58
$(\sqrt{2})^0 = 1$	1/e	0.32
$(\sqrt{2})^1$	1/e	0.15

with  $0 < c \leq \sqrt{2}$ . We have

$$b'(t) = c \begin{cases} -\frac{\ln(t) + 1}{2\sqrt{-t \ln(t)}}, & 0 < t \leq 1/e, \\ 0, & t > 1/e. \end{cases}$$

Note that  $b'(1/e-) = 0$ . So,  $b$  is continuously differentiable. Also,

$$\delta(s, t) = c \begin{cases} \frac{\sqrt{-t \ln(t)} - \sqrt{-s \ln(s)}}{t - s} + \frac{\ln(t) + 1}{2\sqrt{-t \ln(t)}}, & 0 < t \leq 1/e, \\ \frac{\sqrt{1/e} - \sqrt{-s \ln(s)}}{t - s}, & 0 \leq s \leq 1/e < t, \\ 0, & s > 1/e. \end{cases}$$

For  $0 < t \leq 1/e$ ,  $\delta(s, t)$  is maximized at  $s = 0$ . Some algebra reveals that

$$\delta(0, t) = \frac{-\ln(t) + 1}{2\sqrt{-t \ln(t)}} > 0,$$

and  $\lim_{t \rightarrow 0} \{\delta(0, t)\} = \infty$ . Therefore,  $\gamma(t) = \infty$  for all  $t > 0$ , as in the example in the previous section. Again, the pointwise convergence argument for the unscaled calculation breaks down.

Also,  $b$  is concave, so we can and do set  $v = q$ . The integral defining  $\beta^\sim(t)$  is complex and not especially interesting, so we omit it. It seems there is no simplifying parameter like  $\theta$  above, so we must treat each case separately. Table 3 displays some results. Observe that  $\beta^\sim < 1$  in all cases.

(The details are rather peculiar. In all but the first two cases,  $\beta^\sim(t)$  is unimodal, with a mode at  $t^* = 1/e \approx 0.37$ . In the second case ( $c = 1/2\sqrt{2}$ )  $\beta^\sim(t)$  is bimodal, with one mode at  $1/e$  and the other at the much smaller value 0.004. In the first case ( $c = \frac{1}{4}$ ) again there is only one mode, this time at the tiny value shown in the table. Also, it appears that  $\beta^\sim$  is decreasing in  $c$ , but this is not true of  $\beta^\sim(t)$  for each  $t$ . For instance, for  $c = \frac{1}{4}$ ,  $\beta^\sim(1/e) \approx 0.60$ , which is *smaller* than the corresponding value for  $c = \frac{1}{2}$ . We are unable to discern any meaningful pattern here.)

### 5. Concluding remark

The contraction property implies that the substitution algorithm can start with *any*  $f_0 \in \mathfrak{B}$  (or  $f_0^\sim \in \mathfrak{B}$ , if we use scaling), not just 0. As mentioned in the introduction, several authors have suggested refinements of the tangent approximation. Any of these can be taken as the starting

point. Durbin's series is sometimes regarded as an exclusive alternative to these methods, but in fact they can be combined in this manner.

### References

- [1] DANIELS, H. E. (1974). The maximum size of a closed epidemic. *Adv. Appl. Prob.* **6**, 607–621.
- [2] DURBIN, J. (1992). The first-passage density of the Brownian motion process to a curved boundary (with an appendix by D. Williams). *J. Appl. Prob.* **29**, 291–304.
- [3] GRIFFEL, D. H. (2002). *Applied Functional Analysis*. Dover, Mineola, NY.
- [4] LERCHE, H. R. (1986). *Boundary Crossing of Brownian Motion*. Springer, Berlin.
- [5] PESKIR, G. (2002). On integral equations arising in the first-passage problem for Brownian motion. *J. Integral Equations Appl.* **14**, 397–423.
- [6] PESKIR, G. (2002). Limit at zero of the Brownian first-passage density. *Prob. Theory Relat. Fields* **124**, 100–111.
- [7] ROBERTS, G. O. AND SHORTLAND, C. F. (1995). The hazard rate tangent approximation for boundary hitting times. *Ann. Appl. Prob.* **5**, 446–460.
- [8] STRASSEN, V. (1967). Almost sure behavior of sums of independent random variables and martingales. In *Proc. 5th Berkeley Symp. Math. Stat. Prob.* (Berkeley, CA, 1965/66), Vol. II, University of California Press, Berkeley, pp. 315–343.