Some Results on Boundary Hitting Times for One-Dimensional Diffusion Processes

by Christopher Francis Shortland, BSc

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy, October, 1993

CONTENTS

Abstract

Boundary hitting times for one-dimensional diffusion processes have applications in a variety of areas of mathematics. Unfortunately, for most choices of diffusions and boundaries, the exact exit distribution is unknown, and an approximation has to be made. The primary requirements of an approximation, from a practical viewpoint, is that it is both accurate and easily computable. The main, currently used approximations are discussed, and a new method is developed for two-sided boundaries, where current methodology provides very few techniques.

In order to produce new approximations, we will make use of results about the ordering of stochastic processes, and conditioning processes not to have hit a boundary. These topics are introduced in full detail, and a number of results are proved. The ability to order conditioned processes is exploited to provide exact, analytic bounds on the exit distribution. This technique also produces a new approximation, which, for Brownian motion exiting concave or convex boundaries, is shown to be a superior approximation to the standard tangent approximation.

To illustrate the uses of these approximations, and general boundary hitting time results, we investigate a class of optimal stopping problems, motivated by a sequential analysis problem. Properties of the optimal stopping boundary are found using analytic techniques for a wide class of cost functions, and both one- and two-sided boundaries. A number of results are proved concerning the expected stopping cost in cases of "near optimality".

Numerical examples are used, throughout this thesis, to illustrate the principal results and exit distribution approximations.

Chapter 1

Introduction

The central theme to this thesis is the distribution of boundary hitting times for diffusion processes. We introduce the topic of first exit distributions and review existing techniques for finding exact, and approximate, densities. We also develop a new method for approximating exit distributions from two-sided boundaries. We shall make use of results about stochastic ordering of processes, and these are presented in a major section of the thesis. They are then applied to provide new analytic bounds on the distribution function of first exit times, which lead to a new method for approximating the exit distribution. As an application of boundary hitting time distributions, an optimal stopping problem in sequential analysis is investigated. Each chapter will contain an introduction, which will provide the most useful background references for the particular section.

1.1 Background

Diffusion processes are continuous Markov processes, which have a wide variety of applications. They are used to model continuous processes, such as stock market prices, and molecular movement. They are also employed to model other random systems, for instance discrete random walks, or to model deterministic systems, which have random perturbations. Consequently, distributional properties of one-dimensional diffusions can be used as approximations for a wide class of stochastic processes.

We shall describe a diffusion process, X, by a stochastic differential equation, see for example Øksendal (1985). In the stochastic differential equation

$$dX_t = \sigma(t, X_t)dB_t + \mu(t, X_t)dt, \tag{1.1}$$

B represents standard Brownian motion, μ is the drift coefficient and σ is the diffusion coefficient. We shall always assume, unless otherwise stated, that μ and σ satisfy the necessary smoothness and growth conditions to ensure the existence of a unique strong solution (see, for example Rogers and Williams (1987), section V.11).

The conditional expectation of a function of a diffusion can often be shown to satisfy a partial differential equation. For example, the forward and backward equations uniquely specify the transition density of a diffusion process, under appropriate boundary and initial conditions (see Williams (1979), sections I.4, I.9). The analogous forward and backward equations for a discrete time process are difference equations, which are not usually as analytically tractable as the corresponding partial differential equations. Consequently, approximating a random walk by a diffusion often allows us to obtain explicit expressions for the distribution of the process, which might not be available without such an approximation. Another advantage of approximating a random walk by a diffusion, particularly in regard to calculating first exit times, is that a diffusion has continuous sample paths. When computing the first exit time for a diffusion, the value of the process at the exit time is known and this can be exploited. However, this value is unknown for discrete random walks, resulting in a problem known as "overshoot", or "excess over the boundary", (see Siegmund (1985, p165)).

An important partial differential operator, associated with the diffusion, is the infinitesimal generator. If the diffusion, X, satisfies (1.1), then the infinitesimal generator, denoted by $\mathcal{L}$, is given by

$$\mathcal{L}\cdot = \frac{1}{2}\sigma^2(t,x)\frac{\partial^2\cdot}{\partial x^2} + \mu(t,x)\frac{\partial\cdot}{\partial x}. \tag{1.2}$$

The one-sided boundary hitting time, τ, for a functional boundary, $f(t)$, can be defined as

$$\tau = \inf_{t>0}\{t : X_t \geqslant f(t)\},$$

where it is usually assumed that $X_0 < f(0)$. If we further define

$$\phi(t,x) = \mathbf{P}[\tau > t \,|X_0 = x],$$

then ϕ satisfies

$$\mathcal{L}\phi = \frac{\partial\phi}{\partial t},$$

where $\mathcal{L}$ is defined in (1.2). By selecting suitable boundary and initial conditions, the solution to this partial differential equation is uniquely defined. However, explicit solutions are rarely available for particular choices of process and boundary, and hence, approximations are important. Most work has been concentrated on Brownian motion, the simplest diffusion process with constant drift and diffusion coefficients. The main, widely used approximation, currently available for Brownian motion is the tangent approximation (Strassen (1967), see also Lerche (1986)). This is found by approximating the boundary locally by its tangent, and using the exact exit density for the tangent, which is available. This method has the advantage of being easily, and rapidly, calculated in a compact mathematical form. The tangent approximation has also been used as a first approximation, with extensions and refinements provided by other methods (see for example Jennen (1985)). Other approximations are available, which require varying amounts of computational effort, such as a multiple integral and summation approach of Durbin (1992), which also has the tangent approximation as its leading term.

Boundary hitting time distributions are especially important in Sequential Analysis (see for example Siegmund (1985), Woodroofe (1982)) and Stochastic Control Theory (see for example Krylov (1980), Øksendal (1985)). In Sequential Analysis, plans are constructed which determine whether, or not, to continue observing a process, and the objective is to make some inference about an unknown parameter of the process. As several plans can be constructed with the same accuracy, comparisons can be made by considering the expected length of the observation period. For Markov processes, the stopping rules are often determined by comparing the process value with some, possibly time dependent, value. Thus, the length of the observation period has the same distribution as a first exit distribution across some boundary. In a subclass of stochastic control problems, the available control options are to stop or continue, and the objective is to minimise the expected stopping cost. If the costs are only dependent on the time and current value of the process, the distribution of the stopping time is sufficient to calculate the expected stopping cost.

Other applications of boundary hitting times occur in biology, where population sizes are modelled by stochastic processes. Important features of the system, such as the distribution of the time until extinction, are found using boundary hitting time methods, see for example Nobile and Ricciardi (1984a,b). In chemistry, the times of molecular collisions are also considered to be boundary hitting times for diffusion processes, see for example Balding (1988). In this paper, processes are run on the circumference of a circle, and so the techniques we develop, which will rely on the total ordering of the state-space are not applicable.

We shall exclusively consider one-dimensional diffusion processes, which run on a totally ordered state space, usually $\mathbb{R}$. This enables us to consider the ordering of processes, which will provide suitable techniques to produce some new boundary hitting time results. Daley (1968) introduced the idea of stochastic monotonicity. A process, X, is stochastically monotone if

$$\mathsf{P}[X_t \leqslant x \,|X_0 = x_1] \geqslant \mathsf{P}[X_t \leqslant x \,|X_0 = x_2],$$

for all x and $t \geqslant 0$, and all $x_1 \leqslant x_2$. We use an inequality involving probabilities to define stochastic ordering of processes. We say process X^2 is stochastically greater than process X^1 (written $[X_t^2] \overset{\text{st}}{\geqslant} [X_t^1]$) if

$$\mathsf{P}[X_t^2 \leqslant x \,|X_0^2 = x_0] \leqslant \mathsf{P}[X_t^1 \leqslant x \,|X_0^1 = x_0],$$

for all t, x and x_0.

We will make extensive use of processes conditioned not to have hit a boundary, when developing approximations to a first hitting time distribution. However, it does not follow that $\left[X_t^2 \,|\tau_2 > t\right] \overset{\text{st}}{\geqslant} \left[X_t^1 \,|\tau_1 > t\right]$ if $[X_t^2] \overset{\text{st}}{\geqslant} [X_t^1]$, where $\tau_i = \inf_{t>0}\{t : X_t^i \geqslant f^i(t)\}$, for boundary functions f^i. For a simple counterexample, see Roberts (1991a). A natural extension to investigate, would be a stronger stochastic ordering of processes, which is preserved under such conditioning.

We say a process X is strongly stochastically monotone if

$$\frac{p_t(x_2, y_2)}{p_t(x_1, y_2)} \geqslant \frac{p_t(x_2, y_1)}{p_t(x_1, y_1)},$$

for all t, $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$, where $p_t(x, y)$ is the transition density between x and y (see for example Roberts (1991a)). We define strong stochastic ordering between two processes using a non-decreasing likelihood ratio. We say process X^2 is strongly stochastically greater than process X^1 (written $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$) if

$$\frac{p_{s,t}^2(x, y_2)}{p_{s,t}^1(x, y_2)} \geqslant \frac{p_{s,t}^2(x, y_1)}{p_{s,t}^1(x, y_1)}, \tag{1.3}$$

for all x, $s \leqslant t$ and $y_1 \leqslant y_2$, where $p_{s,t}^i(x, y)$ is the transition density for process i from $X_s^i = x$ to $X_t^i = y$. It should be noted that other definitions of strong stochastic ordering are feasible. In cases where the transition density functions are unavailable, we can define strong stochastic ordering by $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$ if

$$\mathsf{P}[X_t^2 \in A_2]\mathsf{P}[X_t^1 \in A_1] \geqslant \mathsf{P}[X_t^2 \in A_1]\mathsf{P}[X_t^1 \in A_2] \qquad \text{for all } t,$$

and all A_1 and A_2 such that $a_1 \in A_1$ and $a_2 \in A_2$ implies $a_1 \leqslant a_2$. We may also define strong stochastic ordering in weaker senses by requiring (1.3) to hold only for some values of s and x. Strong stochastic ordering is preserved, in some cases, after the processes are conditioned not to have hit a boundary.

Strong stochastic monotonicity is a special case of total positivity (Karlin (1968)), which is a higher order property of transition densities. A process is totally positive of order n (TP_n) if the transition density $p_t(x, y)$ satisfies

$$M_n = \begin{vmatrix} p_t(x_1, y_1) & p_t(x_2, y_1) & \cdots & p_t(x_n, y_1) \\ p_t(x_1, y_2) & p_t(x_2, y_2) & \cdots & p_t(x_n, y_2) \\ \vdots & \vdots & & \vdots \\ p_t(x_1, y_n) & p_t(x_2, y_n) & \cdots & p_t(x_n, y_n) \end{vmatrix} \geqslant 0,$$

for all t, $x_1 < x_2 < \cdots < x_n$ and $y_1 < y_2 < \cdots < y_n$. Strong stochastic monotonicity corresponds to TP_2.

One of the aims of this thesis is to develop results on the ordering of processes and apply them to produce new results for boundary hitting time distributions. We will use processes conditioned not to have hit a boundary, and will, therefore, be particularly interested in strong stochastic ordering, which is preserved under our conditioning.

1.2 Summary of Chapters

Hitting time distributions are difficult to obtain exactly, and so a great deal of attention is focussed on the calculation of approximations to them. As a technical tool to produce such approximations, stochastic orderings will be used, necessitating an in depth study of such concepts. New results in this area are found, and then ordering results are employed to find analytic bounds on the distribution function of the hitting time, which are easily calculated — an important consideration when many distributions need to be found.

Following this introductory chapter, the second chapter reviews the current techniques for finding exact and approximate exit distributions (see Lerche (1986) for a good general introduction).

In only relatively few cases can the exact first hitting time density be found. For example, Brownian motion exiting a straight line has its exit density given by the Bachelier-Lévy formula (Lévy (1965)). Daniels (1982) introduces the method of images, which is one of the techniques discussed leading to exact exit densities for Brownian motion across more complicated, implicitly defined, functional boundaries. Partial differential equations for first exit time distributions for more general processes are investigated using Laplace transformation techniques, and also by eigenfunction expansion methods. These methods often yield infinite sums, the dominant terms defining approximations.

The main methods of approximation discussed are those which produce computationally simple distribution functions. Of these, the tangent approximation (Strassen (1967)) is applicable for the Brownian motion exit density across certain one-sided boundaries, when it can be shown to be exact asymptotically, as the boundary recedes to infinity. This formula is generalised for other processes by Durbin (1985), though derived in a different manner. For two-sided boundaries, we develop a new method which uses quasi-stationary distributions, first used in this context by Roberts (1991b), to produce an approximation for the hazard rate of the stopping time. This method is also applicable to a wider selection of processes. The chapter concludes with some numerical examples of the techniques introduced.

After this overview to the currently available methods for calculating boundary hitting distributions, we introduce, in full detail, the concepts of stochastic orderings for diffusion processes. The three main orderings introduced are almost sure ordering, stochastic ordering and strong stochastic ordering. The first two are well known for stochastic processes, and strong stochastic ordering is a property of the likelihood ratio of the two processes. Six different definitions of strong stochastic ordering are given, and relationships between the various orderings are discussed. The chapter also introduces the related topics of stochastic monotonicity (see for example Daley (1968)), strong stochastic monotonicity (see for example Roberts (1991a)) and total positivity (Karlin (1968)). A number of well known results are presented, including an almost sure ordering result for diffusions from Ikeda and Watanabe (1981), and some stochastic and strong stochastic ordering results obtained by comparing processes conditioned not to have hit a boundary (Roberts (1991a), Pollak and Siegmund (1986)).

In the next two chapters, proofs of a number of simpler ordering results are given. The first of these look at likelihood ratios for normal random variables, which will simplify the proofs of a number of results involving processes with deterministic drifts. Results indicating the conclusions which can be drawn if two processes are strongly stochastically ordered are then presented. These are followed by a number of more specific results in the cases where the drift coefficients are functions of time only. In these cases, we are able to explicitly calculate the transition densities of the processes, and can directly verify the definitions of strong stochastic ordering. This permits us to prove results concerning the ordering of process bridges and the moduli of processes.

In order to verify that processes are strongly stochastically ordered, the probability distributions, or transition densities, of the processes are required. However, these are not always explicitly available, in which case other methods to establish strong stochastic ordering are needed. In Chapter 6, this problem is tackled, first by finding necessary conditions on the drift and diffusion coefficients for strong stochastic ordering. Sufficient conditions are then found on the drift coefficients, in the case where the diffusion coefficients are identically 1, to ensure strong stochastic ordering. These new results will allow us to verify whether processes are strongly stochastically ordered by looking at the stochastic differential equations of the processes.

Our attention then returns to boundary hitting time distributions. Using results involving strong stochastic ordering of conditioned processes, and an expression for hazard rates in terms of the density derivative (Roberts (1993)), we can establish a new result producing bounds on the hazard rate of the first exit distribution for functional boundaries. The bounds are found by comparing the actual boundary with other boundaries, for which the exit distribution properties are known. In the important case of Brownian motion, explicit formulae for these bounds can be given, using straight lines as the comparison boundaries. These bounds also suggest a new approximation technique, which, for concave or convex boundaries, can be shown to be a better approximation than the tangent approximation. Numerical examples illustrating these bounds and the new approximation are also provided.

We finish by illustrating the uses of boundary hitting times with an optimal stopping problem. The optimal stopping problem discussed is motivated by a sequential analysis problem investigated in Lerche (1986), in which an inference has to be made about an unknown parameter, based on observations of the process. With error costs and observation costs, a balance has to be found between observing and getting more information about the unknown parameter, thus reducing the error costs, and the cost of this continuation. A solution to such problems can be found, at least numerically, if the exact exit density is known for all boundaries. As this is not the case, properties of the optimal stopping boundary will be investigated by analytic techniques. Throughout this chapter, a worked example will be used to illustrate the ideas developed. Some numerical approaches are discussed involving optimisation over a class of boundaries.

With the possibility of an optimal solution not being found, there follows a discussion, and a number of results, concerning ϵ−optimality (see Krylov (1980)). A solution to the optimal stopping problem which has a payoff within ϵ of the optimal payoff is termed ϵ−optimal. A number of theorems are proved, and applied in a numerical example related to the optimal stopping problem discussed earlier in the chapter.

1.3 Notation

Before proceeding with the main body of the thesis, we shall give some notational conventions which are adopted throughout. Firstly, all processes will be denoted by an upper case letter, usually X, Y, Z or B. The time index will be denoted by a subscript, such as X_t. In order to label processes, when many are required, we shall use superscripts. Thus X^i will denote process i, rather than the process X raised to the ith power. The process B will always be assumed to be a standard Brownian motion.

The distribution law of process X will be denoted by $[X]$. So, for example, we have

$$\left[B_t \,\middle|\, B_0 = 0\right] \sim N(0,t).$$

The use of o and O as orders will be as follows:

$$f(x) = o(g(x)) \Rightarrow \lim_{x \downarrow 0} \frac{f(x)}{g(x)} = 0.$$

$$f(x) = O(g(x)) \Rightarrow \lim_{x \downarrow 0} \frac{f(x)}{g(x)} = k,$$

for some k such that $0 < |k| < \infty$. The same notation will be used if the limits are as $x \to l$. Note that the definition of O is not the standard one.

Weak convergence of distributions will be denoted by $\Rightarrow$. For example, if X^n, $n = 1, 2, \ldots$, converges weakly to X, we shall write

$$[X^n] \Rightarrow [X] \qquad \text{as } n \to \infty.$$

A process $X \in L^p$ is such that

$$\mathsf{E}[|X_t^p|] < \infty, \qquad \text{for all } t.$$

For the function $f(t, x)$, we will usually denote the partial derivatives by $\frac{\partial f}{\partial x}$, *et cetera.* However, for typographical reasons, especially in Chapter 6, we will also use

$$f_x(t, x) = \frac{\partial f}{\partial x},$$

and similar notation for higher order partial derivatives.

Chapter 2

Exit Distributions

2.1 Introduction

Let X be a continuous time Markov process on $\mathbb{R}$, and suppose $A \subseteq \mathbb{R}^2$. We define

$$\tau = \inf_{t>0}\{t : (t, X_t) \notin A\}.$$

Then, the random variable τ is known as the first exit time from set A. We shall normally assume that $A = \{(t,x) : x < f(t)\}$, where $f : [0,\infty) \to \mathbb{R} \cup \{\infty\}$. In such cases, τ is called the boundary hitting time, and written

$$\tau = \inf_{t>0}\{t : X_t \geqslant f(t)\}.$$

It is generally assumed that $(0, X_0) \in A$, or alternatively $X_0 < f(0)$.

In this chapter, we shall investigate the distribution of τ, for various processes X and boundary functions f. We shall review the methods which produce exact distributions for τ, and also consider the techniques which lead to approximations to the exact distribution. We also introduce a new approximation in the case of Brownian motion exiting a two-sided boundary. We shall deal exclusively with one-dimensional diffusion processes, and usually this will be Brownian motion. A good general introduction to this subject is provided by Lerche (1986).

Boundary hitting time distributions are particularly useful in Sequential Analysis (see for example Siegmund (1985)) and in Optimal Stopping Problems (see Chapter 8). In sequential analysis, it is frequently the case that discrete time processes are used. When seeking first exit distributions, this leads to the technical complication that the value of X_τ is unknown, but greater than the boundary value. The difference is termed the "overshoot". To allow for this Siegmund discusses some methods of Woodroofe (1982), which estimate the distribution of the overshoot using renewal theory. Using these results, and a continuous approximation to the process, the exit distribution for the discrete time process can be accurately approximated. We shall only consider continuous time processes, and make no allowances for the overshoot when modelling discrete time processes by continuous ones.

2.2 Exact Distributions

2.2.1 Straight Line Boundary for Brownian Motion

We begin by considering the simplest example of a process exiting a boundary. Let B be a standard Brownian motion, with $B_0 = 0$, and define

$$\tau = \inf_{t>0}\{t : B_t \geqslant a\}$$

where a is a positive constant.

The most important result of this section is the Bachelier-Lévy formula (Lévy (1965)).

Theorem 1 (Bachelier-Lévy formula) *Let $p(t)$ denote the density of the distribution of τ. Then,*

$$p(t) = \frac{a}{t^{3/2}}\phi\left(\frac{a}{\sqrt{t}}\right),$$

where ϕ is the standard normal density function, that is $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

We shall give a simple proof of this result.

Proof

One way to derive this result is by using the reflection principle (see Karlin and Taylor (1975, p345)), which uses the symmetry of the distribution of Brownian motion, and the fact that it has independent increments. Thus if $B_{t_0} = a$, the distributions $[B_t - a]$ and $[a - B_t]$ for $t > t_0$ are identical, using the strong Markov property of Brownian motion. Hence, for each path such that $B_t \geqslant a$, it follows firstly that $\tau \leqslant t$, and secondly, there is a corresponding path, such that $B_s - a = a - B_s$ for all $s \in [\tau, t]$. This path has $B_t \leqslant a$, and $B_\tau = a$ for $\tau \leqslant t$. Thus it is easily deduced that $\mathbb{P}[\tau \leqslant t] = 2\mathbb{P}[B_t \geqslant a]$. Using the well known result that Brownian motion is normally distributed, with zero mean and variance equivalent to the time elapsed, it is clear that

$$\mathbb{P}[\tau \leqslant t] = 2\left(1 - \Phi\left(\frac{a}{\sqrt{t}}\right)\right)$$

where Φ is the standard normal distribution function.

Differentiation of this result leads to the Bachelier-Lévy formula.

★

Before proving a corollary to this result, we require the following result, the Cameron-Martin-Girsanov Theorem. This gives a simple form for the Radon-Nikodym derivative between two probability measures induced by stochastic differential equations.

Lemma 1 (Cameron-Martin-Girsanov Theorem) *Let $\mathcal{B}$ denote the Borel σ-algebra on $\mathbb{R}^+$ and C_0^2 the space of twice continuously differentiable functions. Suppose X satisfies*

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t,$$

with $X_0 = x$, $b \in \mathbb{R}$, $\sigma > 0$. Then suppose another process Y satisfies

$$dY_t = a(t,\omega)dt + b(Y_t)dt + \sigma(Y_t)dB_t,$$

with $Y_0 = x$, and where $a(t,\omega)$ is $\mathcal{B} \times \mathcal{F}$-mble, $\omega \to f_t(\omega) = a(t,\omega)$ is $\mathcal{F}_t$-mble for all t and finally $\mathbb{E}[\int_0^T (a(t,\omega))^2 dt] < \infty$ for all $T < \infty$.

Define Z by $dZ_t = \frac{1}{2}(\sigma^{-1}(Y_t)a(t,\omega))^2 dt - \sigma^{-1}(Y_t)a(t,\omega)dB_t$, $Z_0 = 0$, and let $M_t = e^{Z_t}$. Then, for all $0 \leqslant t_1 \leqslant t_2 \leqslant \cdots \leqslant t_k \leqslant t$, and all $f_i \in C_0^2$,

$$\mathbb{E}[M_t f_1(Y_{t_1}) \cdots f_k(Y_{t_k})] = \mathbb{E}[f_1(X_{t_1}) \cdots f_k(X_{t_k})] .$$

For a proof of this result, see for example Liptser and Shiryayev (1977). We now proceed to a corollary of the Bachelier-Lévy theorem.

Corollary 1 *Consider the process X satisfying*

$$dX_t = dB_t + \mu dt$$

with $X_0 = 0$. Define $\tau_\mu = \inf_{t>0}\{t : X_t \geqslant a\}$ and p_μ to be the density of the distribution of τ_μ.

Then,

$$p_\mu(t) = \frac{a}{t^{3/2}}\phi\left(\frac{a - \mu t}{\sqrt{t}}\right).$$

Proof

We apply the Cameron-Martin-Girsanov Theorem. In this case,

$$dZ_t = \frac{1}{2}\mu^2 dt - \mu dB_t,$$

so that

$$M_t = \exp\left\{\mu^2 t/2 - \mu B_t\right\},$$

assuming $B_0 = 0$. However, when considering the first exit time at t, we have $B_t = a$, so that

$$M_t = \exp\left\{\mu^2 t/2 - \mu a\right\}.$$

Thus,

$$\exp\left\{\frac{1}{2}\mu^2 t - \mu a\right\} p_\mu(t) = p_0(t),$$

where p_0 is found using the Bachelier-Lévy formula. Rearranging this expression provides the claimed result.

★

It should be noted that adding a constant drift to Brownian motion, and exiting a constant boundary is equivalent to exiting the boundary $a - \mu t$ by a standard Brownian motion.

Written in a simpler form, we have thus shown that if

$$\tau = \inf_{t>0}\{t : B_t \geqslant a + bt\},$$

then the first exit density $p(t)$ can be expressed as

$$p(t) = \frac{a}{t^{3/2}}\phi\left(\frac{a+bt}{\sqrt{t}}\right),$$

which is the more commonly stated form of the Bachelier-Lévy formula.

Giorno, Nobile and Ricciardi (1989) extended the idea of applying the reflection principle (see for example Karlin and Taylor (1975)) to find exact first exit densities for other processes. Under certain symmetry conditions on the transition density of the process, and provided that the boundary is selected to satisfy another condition, they find that

$$\mathbb{P}[\tau < t] = 2F(p(x, y, t)),$$

where p is the process' transition density, and F is a given function. In the Brownian motion case, the Bachelier-Lévy formula is attained using $F(p(x_0, y, t)) = \int_a^\infty p(x_0, y, t)dy$ in the above formulation, with constant boundary a.

2.2.2 Eigenfunction Expansions

Let X be an Itô diffusion, with $X_0 = 0$. That is, X has drift and diffusion coefficients which satisfy

$$\begin{aligned} |\mu(t,x)|^2 + |\sigma(t,x)|^2 &\leqslant K[1+|x|^2] \\ |\mu(t,x)-\mu(t,y)| + |\sigma(t,x)-\sigma(t,y)| &\leqslant K|x-y| \text{ for some } K<\infty. \end{aligned}$$

Consider the two-sided stopping time

$$\tau = \inf_{t>0}\{t : X_t \notin (-b,a)\},$$

where a and b are positive constants, and define $\phi(t,x) = \mathsf{P}[\tau > t \,|X_0 = x]$. Then it is well known that ϕ is the unique solution to the partial differential system (see Friedman (1975)):

$$\left.\begin{aligned} \mathcal{L}_t\phi &= \tfrac{\partial\phi}{\partial t} & & \\ \phi(t,a) &= 0 & &(t\neq 0) \\ \phi(t,-b) &= 0 & &(t\neq 0) \\ \phi(0,x) &= 1 & &(-b<x<a) \end{aligned}\right\} \tag{2.1}$$

where $\mathcal{L}_t\cdot$ is the infinitesimal generator of the diffusion. (In the general case where X satisfies $dX_t = \sigma(t,X_t)dB_t + \mu(t,X_t)dt$, then X has the infinitesimal generator $\mathcal{L}_t\cdot = \frac{1}{2}\sigma^2(t,x)\frac{\partial^2\cdot}{\partial x^2} + \mu(t,x)\frac{\partial\cdot}{\partial x}$.)

We now assume the process to be time homogeneous, and so we have the infinitesimal generator $\mathcal{L}_t\cdot = \mathcal{L}\cdot = \frac{1}{2}\sigma^2(x)\frac{\partial^2\cdot}{\partial x^2} + \mu(x)\frac{\partial\cdot}{\partial x}$, where μ and σ satisfy the necessary growth conditions to ensure a unique solution to the partial differential equation.

We shall seek a solution to (2.1) of the form

$$\phi(t,x) = \sum_k a_k(t)e_k(x),$$

where the functions $e_k(x)$ $(k = 1,2,\cdots)$ are eigenfunctions of the infinitesimal generator, which satisfy the boundary conditions $e_k(a) = 0$ and $e_k(-b) = 0$ for our particular choice of a and b. The $a_k(t)$'s are appropriately chosen weights. For this approach to work, we require the differential operator to have eigenfunctions leading to a countable number of distinct eigenvalues. Substituting this form of ϕ into the partial differential equation, we must solve

$$\sum_k a_k(t)\mathcal{L}e_k(x) = \sum_k \frac{da_k}{dt}e_k(x),$$

subject to the boundary conditions. Using the eigenfunction property, this reduces to

$$\sum_k a_k(t)\lambda_k e_k(x) = \sum_k \frac{da_k}{dt} e_k(x),$$

where λ_k is the eigenvalue associated with eigenfunction e_k.

Provided that the e_k's form an orthonormal basis with respect to some inner product, taking this inner product with e_j on both sides yields

$$\lambda_j a_j(t) = \frac{da_j}{dt}$$

or equivalently

$$a_j(t) = c_j e^{\lambda_j t}.$$

where c_j is a constant.

Since we are dealing with a one-dimensional diffusion in most cases, this expansion is possible, and the inner product can be written as

$$\langle f, g \rangle = \int f(x) g(x) \rho(x) dx,$$

where $\rho = \frac{dm}{ds}$, the derivative of the scale measure with respect to the speed measure of the process (see for example Karatzas and Shreve (1988, Chapt. 5)). In the most convenient cases, it is possible to find an orthonormal basis of eigenfunctions, which also has distinct, non-clustered eigenvalues. In particular this means that, as all eigenvalues are negative, the dominant term asymptotically, is the term corresponding to the least negative eigenvalue, and we have

$$\begin{aligned} \frac{\phi(t,x)}{c_1 e^{\lambda_1 t} e_1(x)} &= 1 + e^{(\lambda_2 - \lambda_1)t} g(t,x) \\ &\to 1 \text{ as } t \to \infty, \end{aligned}$$

for some function g.

The final solution is of the form

$$\phi(t,x) = \sum_k c_k e^{\lambda_k t} e_k(x).$$

The initial condition can be used to determine the values c_k, and the boundary conditions are satisfied by choosing the e_k so that $e_k(x) = 0$ at $x = a$ and $x = -b$. Thus denoting the inner product by $\langle \cdot, \cdot \rangle$, we have the final expression of

$$\phi(t,x) = \sum_k \langle 1, e_k \rangle e^{\lambda_k t} e_k(x).$$

Example: Brownian motion

As an example, consider the case of $X_t = B_t$, and again τ is defined as

$$\tau = \inf_{t>0}\{t : B_t \notin (-b, a)\}.$$

To find the eigenvalues and eigenfunctions, we solve

$$\frac{1}{2}\frac{d^2e}{dx^2} = \lambda e,$$

subject to the constraints that $e(a) = e(-b) = 0$. Clearly, if $2\lambda = -m^2$, we have the solutions

$$e(x) = A\sin(mx) + B\cos(mx),$$

to the differential equation, which we can solve for A and m by selecting $B = 1$. (Any constant multiplier of an eigenfunction is also an eigenfunction — we only seek the functional form in x.) Substituting the boundary conditions, we find

$$A = \frac{\cos(mb)}{\sin(mb)} \text{ and } \sin(ma + mb) = 0.$$

Therefore, we select

$$m = \frac{n\pi}{a+b} \text{ for } n = 1, 2, \cdots.$$

Summarising this, we see the functional form of the eigenfunction is

$$e_n(x) = \sin\left[\frac{n\pi}{a+b}(x+b)\right],$$

with corresponding eigenvalue

$$\lambda_n = -\frac{n^2\pi^2}{2(a+b)^2}.$$

Normalising the e_n, by using the inner product

$$\langle f, g\rangle = \int_{-b}^{a} f(x)g(x)dx,$$

(as $\rho(x) = 1$), we see an orthonormal basis of the space of C^2 functions with $f(a) = f(-b) = 0$, is given by

$$\left\{\sqrt{\frac{2}{a+b}}\sin\left[\frac{n\pi}{a+b}(x+b)\right]\right\}_{n=1}^{\infty}.$$

Thus, we deduce

$$\begin{aligned}\mathsf{P}[\tau > t \,|X_0 = x] &= \sum_{n=1}^{\infty} \langle 1, e_n \rangle \exp\left\{-\frac{n^2\pi^2}{2(a+b)^2}t\right\} \sin\left[\frac{n\pi}{a+b}(x+b)\right] \\ &= \sum_{k=1}^{\infty} \frac{4}{(2k-1)\pi} \exp\left\{-\frac{(2k-1)^2\pi^2}{2(a+b)^2}t\right\} \sin\left[\frac{(2k-1)\pi}{a+b}(x+b)\right].\end{aligned}$$

Note that, in the symmetrical case $a = b$, this reduces to

$$\mathsf{P}[\tau > t \,|X_0 = x] = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{4}{(2k-1)\pi}e^{-\frac{(2k-1)^2\pi^2}{8a^2}t}\cos\left(\frac{(2k-1)\pi x}{2a}\right). \quad (2.2)$$

Notice that the asymptotic behaviour is closely approximated by the leading term, and for small boundary values a, only a few terms would be necessary to calculate $\mathsf{P}[\tau > t]$ in practice, due to the fact that the probability is the sum of exponential terms. For $a = b = 1$ and $t > 0.2$, after about twenty terms of this sum, the summands are effectively zero, to computer accuracy of about $(10)^{-30}$, and so truncation of the infinite sum is justifiable at this number of terms.

Notice, also, that the non-symmetrical case can be used as a basis for a one-sided approximation, by letting b become large. In such a case, the first exit time across the two-sided boundary will be a good approximation to the case of a single boundary at a, since the probability of the first exit being across $-b$ is very small.

Example: Ornstein-Uhlenbeck process (Symmetric Boundaries)

Consider the Ornstein-Uhlenbeck process satisfying the stochastic differential equation

$$dX_t = dB_t - \frac{\alpha}{2}X_t dt.$$

Let $\tau = \inf_{t>0}\{t : |X_t| \geqslant a\}$ and $\phi(t,x) = \mathsf{P}[\tau > t\,|X_0 = x]$, which is a solution of the partial differential equation $\mathcal{L}\phi = \frac{\partial\phi}{\partial t}$. We will also define

$$\mathcal{S} = \{C^2 \text{ functions, } f \text{ such that } f(x) = 0 \text{ for } x = \pm a\}.$$

Clearly for fixed t, the function $\phi_t(\cdot) = \phi(t,\cdot) \in \mathcal{S}$. Furthermore, we define the inner product

$$\langle f, g\rangle = \int_{-a}^{a} f(x)g(x)e^{-\frac{\alpha}{2}x^2}dx,$$

under which our operator $\mathcal{L}$ is self adjoint.

We can then use spectral theory, (see for example Dunford and Schwartz (1963)) and deduce that the eigenfunctions of $\mathcal{L}$ form an orthogonal, spanning set of $\mathcal{S}$, under the inner product defined above. We may now express ϕ as

$$\phi(t,x) = \sum_i \langle \phi, e_i\rangle\, e_i(x),$$

where the e_i's are eigenfunctions satisfying

$$\frac{1}{2}\frac{d^2e}{dx^2} - \frac{\alpha x}{2}\frac{de}{dx} = \lambda_n e.$$

subject to the condition that

$$e(\pm a) = 0.$$

This system is solved by confluent hypergeometric functions, and in particular by parabolic cylinder functions (see Erdelyi (1953)).

The importance of this result is in the calculation of first exit time distributions for Brownian motion across square root boundaries. This is because the Ornstein-Uhlenbeck process may be expressed as

$$X_t = \frac{e^{-\frac{\alpha}{2}t}}{\sqrt{\alpha}} B_{e^{\alpha t}},$$

which makes it a time changed Brownian motion. A consequence of this time-scale change is that constant boundaries for the Ornstein-Uhlenbeck process correspond to square root boundaries for Brownian motion. That is

$$\{|X_t| \geqslant a\} \equiv \left\{\left|\frac{e^{-\frac{\alpha}{2}t}}{\sqrt{\alpha}} B_{e^{\alpha t}}\right| \geqslant a\right\} \equiv \{|B_{e^{\alpha t}}| \geqslant a\sqrt{\alpha} e^{\frac{\alpha}{2}t}\} \equiv \{|B_s| \geqslant a\sqrt{\alpha}\sqrt{s},\ s = e^{\alpha t}\}.$$

This connection will be discussed further in a later section.

2.3 Implicit Function Methods

2.3.1 Method of Images

Consider B to be standard Brownian motion, started from $B_0 = 0$. For this process, there exist methods to find exact exit distributions across more complex functional boundaries. One such technique is the so-called method of images (see for example Lerche (1986) Chapter 1, Section 1, or Daniels (1982)). This idea revolves around looking at the density of the Brownian motion distribution, which in addition to its point mass at zero starting measure, also has a negatively weighted starting distribution on the positive real state space (for example a point mass at (0,1) in the (t, x) plane). If we define this starting measure to be $F(d\theta)$, and

$$h(t,x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} - a^{-1} \int \frac{1}{\sqrt{2\pi t}} e^{-(x-\theta)^2/2t} F(d\theta)$$

to be the density of the resulting Brownian motion under this mixed starting measure, then it can be shown that there exists a unique solution, $x = f(t)$, to the equation $h(t, x) = 0$, for each t. We can then define

$$\begin{aligned} \tau &= \inf_{t>0}\{t : B_t \geqslant f(t)\} \\ &= \inf_{t>0}\{t : B_t = f(t)\}, \text{ by continuity of the process,} \\ &= \inf_{t>0}\{t : h(t, B_t) = 0\}. \end{aligned}$$

Several results can be proved about $h(t, x)$, and its relationship to the exit distribution and density from $f(t)$ (see Lerche (1986) for the proofs). It is known that h satisfies the following set of equations:

$$\begin{aligned} \frac{\partial h}{\partial t} &= \frac{1}{2}\frac{\partial^2 h}{\partial x^2} & \\ h(t, f(t)) &= 0 & \text{for all } t > 0 \\ h(0, \cdot) &= \delta_0 & \text{on } (-\infty, f(0)) \end{aligned}$$

where δ_0 denotes the Dirac measure at zero. Note that these are virtually the same set of equations as (2.1), satisfied by ϕ. The only difference is the initial condition, where a starting measure is given as opposed to a probability. Again, the boundary condition is that the function is zero on the boundary. However, defining $p(t, x) = \frac{\partial}{\partial x}(\mathbb{P}[\tau > t, B_t \in dx])$, it is easy to verify that p also satisfies this system, and it can be deduced that

$$\mathbb{P}[\tau > t, B_t \in dx] = h(t, x)dx,$$

using the uniqueness of the solution of such a parabolic differential system. Following from this result, by conditioning, we have

$$\mathbb{P}[\tau \leqslant t \,|B_t = x] = 1 - \frac{h(t, x)}{\frac{1}{\sqrt{t}}\phi\left(\frac{x}{\sqrt{t}}\right)}$$

where $\phi(x)$ denotes the standard normal density function. This leads to the exit distribution $\mathbb{P}[\tau \leqslant t]$ on integration. That is

$$\mathbb{P}[\tau \leqslant t] = 1 - \Phi\left(\frac{f(t)}{\sqrt{t}}\right) + a^{-1}\int_0^\infty \Phi\left(\frac{f(t) - \theta}{\sqrt{t}}\right) F(d\theta)$$

where $\Phi(x)$ is the standard normal distribution function.

The final result provides a direct link between the first exit density and $h(t, x)$. Specifically,

$$p(t) = -\left.\frac{1}{2}\frac{\partial h}{\partial x}\right|_{x=f(t)},$$

which is shown by differentiating the result that $\mathsf{P}[\tau > t] = \int_{-\infty}^{f(t)} h(t,y)dy$. Thus once we have calculated $h(t,x)$, we can quite rapidly work out $f(t)$, and its exit distribution and density. These results can easily be extended to Brownian motion with drift by the Cameron-Martin-Girsanov formula.

Examples

We can verify the Bachelier-Lévy formula. If we choose the extra starting measure to be a point mass at 2θ, then we can calculate $h(t,x)$ to be

$$h(t,x) = \frac{1}{\sqrt{t}}\phi\left(\frac{x}{\sqrt{t}}\right) - a^{-1}\frac{1}{\sqrt{t}}\phi\left(\frac{x-2\theta}{\sqrt{t}}\right).$$

This has the property that $h(t,f(t)) = 0$ for $f(t) = \theta + \frac{\log a}{2\theta}t$, which is a straight line. Now differentiation of h with respect to x, and setting $x = f(t)$ produces

$$p(t) = \frac{\theta}{t^{3/2}}\phi\left(\frac{f(t)}{\sqrt{t}}\right),$$

which is the familiar Bachelier-Lévy formula. Obviously, by varying the starting measure F, and the weighting factor a, more complex boundaries may be produced.

An extension to this method is to use a starting measure F which takes values on the whole real line. The effect of this alteration is that $h(t,x) = 0$ no longer has a unique solution, and instead, there exist two curves, $f_+(t) \geqslant 0$ and $f_-(t) \leqslant 0$ which satisfy this equation, and thus we can find the exit density in some two-sided cases.

As an example of this, and of a more complex boundary curve, we shall give the example of a family of curves discussed by Daniels (1982). In this example, a will be taken as $\frac{1}{2}k$, and $F(d\theta)$ will consist of symmetrical point masses, that is $F = \delta_\alpha + \delta_{-\alpha}$. In such a case, the two solutions to $h(t,f(t)) = 0$ can be given as

$$f_\pm(t) = \pm\frac{t}{\alpha}\cosh^{-1}\left(\frac{1}{k}e^{\alpha^2/2t}\right),$$

the shape of which has different properties according to the values of k chosen.

When $k > 1$, only values of $t < t_1 = \alpha^2/(2\log k)$ produce real values for $f(t)$, which in practice means that the two sided boundary is closed at $t = t_1$, and so, the stopping time $\tau \leqslant t_1$. Indeed its behaviour near t_1 can be shown to be approximately $\pm\sqrt{t_1 - t}$. But when $k < 1$, for large t, $f(t) \approx \gamma t + O(1)$, where $\gamma = (1/\alpha)\cosh^{-1}(1/k)$, and this case has the property that k represents the probability of eventually hitting one side of the boundary. The final case, of $k = 1$, produces the approximate square root boundary $f(t) \sim (t + \alpha^2/6)^{1/2} + O(t^{-1})$. In all cases, though, the first exit density can be expressed as

$$p(t) = \frac{1}{2t}\left\{\frac{f(t)}{\sqrt{t}}\phi\left(\frac{f(t)}{\sqrt{t}}\right) - \frac{k}{2}\frac{f(t)-\alpha}{\sqrt{t}}\phi\left(\frac{f(t)-\alpha}{\sqrt{t}}\right) - \frac{k}{2}\frac{f(t)+\alpha}{\sqrt{t}}\phi\left(\frac{f(t)+\alpha}{\sqrt{t}}\right)\right\}.$$

Remarks

The description of the method of images indicates the major drawback of this technique. The problem is that a boundary curve cannot be selected first, and its exit distribution calculated, as would be the ideal situation. Instead an image distribution is selected, from which a boundary curve can be calculated, for which the first exit properties may be deduced. This boundary itself is derived as the solution of an implicit equation, and it is often the case, in practice, that an explicit form cannot be found. Limitations do also exist on the type of boundary which may be constructed. It is proved (Lerche (1986)) that $f(t)$ must be infinitely often continuously differentiable, concave and with $f(t)/t$ monotonically decreasing. Though it does produce exact densities, its practical use may become limited.

2.3.2 Method of Weighted Likelihood Functions

A similar method for finding exact exit densities is the method of weighted likelihood functions (see Lerche (1986)), which is carried out as follows. Define

$$f(t,x) = \int \exp\left(\theta x - \frac{1}{2}\theta^2 t\right) F(d\theta)$$

where F is a positive σ-finite measure on the positive real numbers. We solve the implicit equation $f(t,x) = a$, to find $x = f(t)$. This is valid for all $t \geqslant t_0$, where t_0 is such that there exists a (t_0, x_0) with $f(t_0, x_0) < \infty$. (We assume such a t_0 exists). The solution $x = f(t)$ is unique for $t \geqslant t_0$, and has the useful property that $f(t,x) < a \Leftrightarrow x < f(t)$. Bearing in mind that $f(t)$ is only defined for $t \geqslant t_0$, we make a slight alteration to τ, and let

$$\tau = \inf_{t > t_0}\{t : f(t, B_t) \geqslant a\},$$

which clearly corresponds to the first hitting time of our boundary $f(t)$, after $t = t_0$.

It can be shown that if $0 < f(t_0, x_0) \leqslant a < \infty$, we have

$$\mathbb{P}[t_0 \leqslant \tau < \infty \,|B_{t_0} = x_0] = a^{-1} f(t_0, x_0),$$

which can be integrated to produce

$$\mathsf{P}[t_0 \leqslant \tau < \infty] = 1 - \Phi\left(\frac{f(t_0)}{\sqrt{t_0}}\right) + a^{-1}\int_0^\infty \Phi\left(\frac{f(t_0) - \theta t_0}{\sqrt{t_0}}\right) F(d\theta).$$

The proof (see Lerche (1986)) of the first result relies on the fact that a probability measure, denoted by $Q^{(t_0,x_0)}$, can be found which satisfies the property of being mutually continuous with $P^{(t_0,x_0)}$, the measure of Brownian motion started from $B_{t_0} = x_0$. Then the ratio defined by $f(t_0,x_0)/f(t,B_t)$ is a $Q^{(t_0,x_0)}$-martingale, which allows the use of the optional stopping theorem (see for example Dellacherie and Meyer (1978)) to deduce

$$\mathsf{P}[t_0 \leqslant \tau \leqslant t_1] = a^{-1} f(t_0,x_0) Q^{(t_0,x_0)}\{t_0 \leqslant \tau \leqslant t_1\}.$$

By bounding an exponential growth term, it can be shown that the term

$$Q^{(t_0,x_0)}\{t_0 \leqslant \tau \leqslant t_1\} \to 1 \text{ as } t_1 \to \infty,$$

which completes the proof.

The method of weighted likelihood functions has similarities to the method of images. In particular, a starting measure is used to deduce a boundary function by way of an implicit equation involving another function. It is this boundary for which we can derive the first exit density. In fact, Lerche (1986) has proved that the methods are the same up to an inversion of the time-scale, and the corresponding scaling of the state space. Thus the problems for the implementation of this method are similar to those of the method of images, particularly that of obtaining explicit expressions for the boundary functions.

2.4 Laplace Transformation Techniques

An example of a procedure which yields some exact densities, but also has a use in providing approximations, is the method of Laplace transformations. The technique itself is relatively simple. Recall that $\mathsf{P}[\tau > t \,|X_0 = x]$ satisfies the parabolic differential system (2.1). As an alternative approach to taking eigenfunction expansions, as in section 2.2.2, taking Laplace transforms of (2.1) with respect to t, reduces the partial differential equation into an ordinary differential equation. The only drawback to the technique, is that the inversion of the Laplace transform, back to $\mathsf{P}[\tau > t \,|X_0 = x]$, is frequently difficult.

2.4.1 Example: Brownian Motion exiting a Square Root Boundary

We shall illustrate this technique by finding the first exit density of Brownian motion across a square root boundary, which will be done using a time-scale conversion to the case of an Ornstein-Uhlenbeck process hitting a constant boundary. Define the stopping time τ as

$$\tau = \inf_{t>1}\{t : B_t \geqslant a\sqrt{t}\}.$$

Note that an infimum over $\{t > 1\}$ is taken, since the starting point $B_0 = 0$ is assumed, and the law of iterated logarithms (see Øksendal (1985)) indicates that the infimum over $\{t > 0\}$, of the same set, would be identically zero. The law of the iterated logarithm also indicates that the boundary is almost surely attained, that is $\mathsf{P}[\tau < \infty] = 1$.

We introduce a time change

$$[\alpha'(t)]^{-\frac{1}{2}} dB_{\alpha(t)} \equiv dB^\star_t \tag{2.3}$$

where $'$ denotes differentiation with respect to t, and B and $B^\star$ are both standard Brownian motions. We shall set $X_t = [\alpha'(t)]^{-\frac{1}{2}} B_{\alpha(t)}$, and note that X satisfies the stochastic differential equation

$$dX_t = dB^\star_t - \frac{1}{2}\frac{\alpha''(t)}{\alpha'(t)} X_t dt \ .$$

We select a function $\alpha(t)$ (usually $\alpha : [0,\infty) \to [0,\infty)$, one-one, monotonically increasing), so that X is a particular type of diffusion. Thus, to convert to a standard Ornstein-Uhlenbeck process (recall that this satisfies $dY_t = dB_t - \frac{1}{2}Y_t dt$), we must solve $\alpha''(t) = \alpha'(t)$. This corresponds to the time change $\alpha(t) = e^t$. Thus, $e^{-t/2}B_{e^t}$ is an Ornstein-Uhlenbeck process.

Recall our definition of τ. If we now define a new stopping time by

$$\tau_1 = \inf_{t>0}\{t : X_t \geqslant a\}$$

where X is the Ornstein-Uhlenbeck process, we can see that $\tau = e^{\tau_1}$ holds. Note

$$\{X_t \geqslant a\} \equiv \{e^{-\frac{1}{2}t} B_{e^t} \geqslant a\} \equiv \{B_{e^t} \geqslant a e^{\frac{1}{2}t}\} \equiv \{B_s \geqslant a\sqrt{s}, \quad s = e^t\}.$$

We can thus use standard distribution results to convert $\mathsf{P}[\tau_1 \leqslant t]$ into $\mathsf{P}[\tau \leqslant t]$, and so the approach of considering the Ornstein-Uhlenbeck process to a constant boundary is valid.

Many variations of this problem have been investigated in the literature. Two examples are in papers by Ricciardi and Sato (1988), and Breiman (1967). Ricciardi and Sato solve this problem to find $\psi(t,x) = \frac{\partial}{\partial t}\mathsf{P}[\tau_1 \leqslant t \,|X_0 = x]$, with the boundary at a, which also satisfies (2.1) with appropriate initial and boundary conditions. The standard formula for the Laplace transform of a derivative is $(\frac{df}{dx})^\star = yf^\star(y) - f(0)$, where * denotes Laplace transformation in x to variable

y. Noting $\psi(0,x) = 0$, but $\phi(0,x) = 1$, we see that using ψ simplifies the Laplace transform, as the constant term vanishes. Taking Laplace transforms of (2.1), (with the appropriate $\mathcal{L}$), we obtain

$$\frac{\partial^2 \psi^\star}{\partial x^2} - x\frac{\partial \psi^\star}{\partial x} - 2\lambda\psi^\star = 0,$$

(where $\psi^\star(\lambda, x) = \int_0^\infty e^{-\lambda t}\psi(t,x)dt$), which is identified to have a solution in the form of parabolic cylinder functions. Indeed, for $X_0 = x_0 < a$, the Ricciardi and Sato paper gives the solution

$$\psi^\star(\lambda, x_0) = \exp\left[\frac{x_0^2 - a^2}{4}\right]\frac{D_{-2\lambda}(-x_0)}{D_{-2\lambda}(-a)}$$

where D_λ is a parabolic cylinder function (see Erdelyi (1953)). The paper goes on to the inversion of these transforms, giving

$$\psi(t, x_0) = \frac{2|x_0|}{\sqrt{2\pi}}(e^{2t} - 1)^{-3/2}e^{2t}\exp\left[-\frac{x_0^2}{2(e^{2t}-1)}\right]$$

which is true for $x_0 < a = 0$. For other values of a, the paper also proves a more complicated expression for the density function. Denoting the expression for $\psi^\star$ above by

$$\psi^\star(\lambda, x_0) = \frac{f(2\lambda, x_0)}{f(2\lambda, a)}$$

and the zeros of $f(\lambda, z)$, for fixed z, by $\lambda_p(z)$, it can be shown that

$$\psi(t, 0) = \sum_{p=0}^{\infty} A_p(a)e^{\lambda_p a t}$$

where,

$$A_p(a) = \frac{1}{2}\sum_{n=0}^{\infty}\frac{(\sqrt{2}a)^n}{n!}\frac{\Gamma((n+\lambda_p)/2)}{\Gamma(\lambda_p/2)}\left[g\left(\frac{n+\lambda_p}{2}\right) - g\left(\frac{1+\lambda_p}{2}\right) - g\left(\frac{\lambda_p}{2}\right)\right]^{-1}$$

in which $g(x)$ is the digamma function,

$$g(x) = \frac{1}{\Gamma(x)}\frac{d\Gamma(x)}{dx}.$$

Breiman (1967) considered the two-sided case, with a view to approximation. Using the usual definition of $\phi(t,x) = \mathsf{P}[\tau_1 > t\,|X_0 = x]$, he showed

$$\phi^\star(s, 0) = e^{-a^2/4}\left[\frac{D_{-s}(0) + D_{-s}(0)}{D_{-s}(a) + D_{-s}(-a)}\right]$$

again with $D_s(x)$ as parabolic cylinder functions. Defining $-2\beta(a)$ as the largest pole of this expression, he showed that

$$\phi(t,0) = \alpha e^{-2\beta(a)t} + O(e^{-(2\beta(a)+\delta)t}) \quad (\delta > 0) \; ,$$

thus gaining an approximation to the asymptotic behaviour.

Both these results can easily be transferred to the case of Brownian motion exiting a square root boundary, by inverting the time change (2.3). It should also be noted, that although an explicit formula does exist for the first exit density, and can be found by Laplace transform techniques, the formula is not easy to work with, due to its double infinite sum form. This is the major drawback to the use of Laplace transforms in finding exit densities — the method is complicated, and neat explicit forms are hard to obtain as a result of the general difficulty in inverting Laplace transforms.

2.5 Exit Distribution Approximations

The methods for producing exact exit densities all tend to produce expressions which are difficult to work with, in most practical cases except the straight line case. Additionally, the most general two, the method of images and the method of weighted likelihood functions, do not have much direct control over the boundary for which the exit distribution can be found. Consequently, approximations are extremely useful, mainly because they may be defined in such a way that the boundary can be chosen initially, rather than arrived at through an implicit equation. Furthermore, it may also be possible to produce more compact, explicit expressions, which might be beneficial in practical work.

2.5.1 Tangent Approximation

One method of approximation is the tangent approximation (Strassen (1967), see also Lerche (1986)). This result takes families of boundary curves $\{f_a : a \in \mathbb{R}_+\}$, which are positive, increasing and continuously differentiable functions on $(0,\infty)$. A stopping time, τ_a, is defined for each curve by

$$\tau_a = \inf_{t>0}\{t : B_t \geqslant f_a(t)\},$$

where $B_0 = 0$. Let $p_a(t)$ denote the first exit density from the curve $f_a(t)$, and $\Lambda_a(t)$ denote the intercept of the tangent, at time t, with the x-axis, that is

$$\Lambda_a(t) = f_a(t) - tf_a'(t).$$

The following theorem is extracted from Lerche (1986, p60), and represents one form of the tangent approximation. Other forms do exist, which hold uniformly on sets of the form $(0, h_a)$, where $h_a \to \infty$ as $a \to \infty$.

Theorem 2 (Tangent Approximation) *Let* $0 < t_1 \leqslant \infty$ *and* $0 < \alpha < 1$. *Assume*

$$\mathsf{P}[\tau_a < t_1] \to 0 \quad \text{as } a \to \infty$$

$$\frac{f_a(t)}{t^\alpha} \quad \textit{is monotone increasing in } t \textit{ for all } a$$

$$\textit{for all } \epsilon > 0, \textit{ there exists } \delta > 0 \textit{ s.t. for all } a, \quad \left|\frac{f_a'(s)}{f_a'(t)} - 1\right| < \epsilon \quad \textit{if} \quad \left|\frac{s}{t} - 1\right| < \delta$$

for $s, t \in (0, t_1)$.

Then,

$$p_a(t) = \frac{\Lambda_a(t)}{t^{3/2}} \phi\left(\frac{f_a(t)}{\sqrt{t}}\right)(1 + o(1)) \quad \textit{uniformly on} \quad (0, t_1) \quad \textit{as} \quad a \to \infty.$$

Proof

We shall give a brief outline to the proof. For full details, see Lerche (1986). The approximation is proved using the related function $p_a^{(r,x)}(t)$, the exit density given the starting point $B_r = x < f_a(r)$. This function also satisfies the equation

$$p_a(t) = \int_{-\infty}^{f_a(r)} \mathsf{P}[\tau_a > r, B_r \in dx] p_a^{(r,x)}(t),$$

which is the key result in proving the tangent approximation, once an approximation for $p_a^{(r,x)}(t)$ has been established. By conditioning on an intermediate time point, and estimating the resulting integral, it may be shown that

$$p_a^{(r,x)}(t) = \frac{\Lambda_a(t)(1 + o(1)) - x}{(t - r)^{3/2}} \phi\left(\frac{f_a(t) - x}{\sqrt{t - r}}\right) \qquad \text{uniformly } \forall x \leqslant f_a(s)(r/s)^\beta$$

where

$$s = t\left(1 - \left(\frac{t}{(f_a(t))^2}\right)^\epsilon\right), \quad (\epsilon > 0)$$

and $\beta < 1$ is chosen so that

$$\left[(f_a(s))^2/s\right](r/s)^\beta \to 0$$

and

$$\left[(f_a(s))^2/s\right](r/s)^{2\beta - 1} \to \infty$$

uniformly on $(0, t_1]$ as $a \to \infty$.

This is sufficient to obtain the desired result.

★

The intuition behind the method is as follows. For those boundaries which recede to infinity in the required manner, the tangent to the curve lies above the curve itself. For most Brownian paths, for which $B_\tau = f(\tau)$ is the position at the first exit time of the tangent, B_t is close to $[f(\tau)/\tau]t$, the ray from the origin to the first exit point. Thus for large τ, the proportion of paths which actually exit the boundary curve, but not the tangent, prior to τ is small, since in general the curve is not close to the ray $[f(\tau)/\tau]t$. Consequently, small time events are less important (which corresponds to the times when the curve, and its tangent are far apart), so replacement of the curve by its tangent, at each time point, is sensible.

This basic result may be extended in two relatively simple ways. Firstly it can be shown that it is uniformly true on $(0, h_a)$, where $h_a \to \infty$ as $a \to \infty$, under the new assumption that $\mathsf{P}[\tau_a < h_a] \to 0$ as $a \to \infty$, instead of $\mathsf{P}[\tau_a < t_1] \to 0$, and that the boundaries are upper class functions at ∞. (That is $\mathsf{P}[B_t > f_a(t)] \to 0$ as $t \to \infty$). The other trivial extension is to the case where Brownian motion has constant drift μ, in which case we have $p_a^\mu(t)$ given by

$$p_a^\mu(t) = \frac{\Lambda_a(t)}{t^{3/2}}\phi\left(\frac{f_a(t) - \mu t}{\sqrt{t}}\right)(1 + o(1)),$$

again uniformly on $(0, t_1)$ as $a \to \infty$.

The extensions are proved using the invariance of the ratio between the exact density, and the approximation (see Lerche (1986)). The extension to uniformity on $(0, h_a)$ uses a time-scale transformation,

$$s = t/h_a, \quad y = x/\sqrt{h_a}.$$

If $\eta_a(s) = f_a(h_a s)/\sqrt{h_a}$ on $(0, 1)$, and the density of the first exit distribution over η_a is denoted by $r_a(s)$, we have

$$p_a(t)\left[\frac{f_a(t) - tf_a'(t)}{t^{3/2}}\phi\left(\frac{f_a(t)}{\sqrt{t}}\right)\right]^{-1} = r_a(t)\left[\frac{\eta_a(t) - t\eta_a'(t)}{t^{3/2}}\phi\left(\frac{\eta_a(t)}{\sqrt{t}}\right)\right]^{-1}.$$

Then the tangent approximation valid for η_a carries over to the tangent approximation for f_a on $(0, h_a)$. The second extension can be established in a similar manner.

Further extensions were carried out by Ferebee (1983) and Jennen (1985). The method employed by Jennen requires the following additional properties of the boundary:

$$\left|\frac{f_a''(s)}{f_a''(t)} - 1\right| < \epsilon \text{ if } \left|\frac{s}{t} - 1\right| < \delta$$

and

$$\text{there exists } \rho < 1,\ K < \infty \text{ s.t. } \left|t^{3/2} f_a''(t)\right| < K\left(\frac{f_a(t)}{\sqrt{t}}\right)^{1+\rho} \quad \text{for all } t \in (0, h_a).$$

We further assume that $\mathrm{P}[\tau_a < h_a] \to 0$ as $a \to \infty$. Under these further assumptions, as well as those required for the usual tangent approximation, a second order approximation can be calculated using

$$\begin{aligned} p_a(t) &= \Bigg[\Lambda_a(t) + \frac{f_a''(t)t^3}{2\Lambda_a^2(t)}(1 + o(1)) \\ &\quad - \Lambda_a(t)\mathrm{P}[\tau_a < t](1 + o(1)) + o(R_a(t))\Bigg] t^{-3/2}\phi\left(\frac{f_a(t)}{\sqrt{t}}\right), \end{aligned}$$

where $R_a(t) = \exp(-(f_a(t)/\sqrt{t})^k)$ for some $k > 0$. Again this holds uniformly on $(0, h_a)$ as $a \to \infty$.

This density consists of two approximation terms, one global and one local. The local term takes care of the approximation near to time t, and the global term takes into account the probability of hitting the curve, but not the tangent, at some earlier time. Not surprisingly, comparison of the second order approximation with the usual tangent approximation, in examples where the exact density is known from the method of images, favours this new, more complex form (see Lerche (1986, p68)).

Limitations

Consider the two-sided stopping time

$$\tau_a = \inf_{t>0}\{t : |B_t| \geqslant f_a(t)\},$$

where the boundary curves $f_a(t)$ satisfy the same conditions as before. The simplest approximation to $p_a(t)$ would be to double the estimate gained using the tangent approximation in the one-sided case. However, this would ignore the probability of a path hitting both curves prior to any time point t. So, unless the curves diverge rapidly, making this probability small, the tangent approximation cannot be simply converted to deal with the two-sided situation.

A correction factor may be obtained, to improve the estimation. The probability of a double hit can be approximated by looking at the tangent approximation of hitting one curve from the other, and integrating over the first hitting time of either curve. Subtracting this from twice the tangent approximation density will be an improvement, but in many cases will still be poor. Thus, a major deficiency with the tangent approximation is its inability to handle two-sided stopping boundaries.

The method does have other problems and drawbacks. We are still limited to boundaries which are increasing to infinity, and have an approximation which is only accurate asymptotically in t, when the boundary is large. We thus cannot accurately deal with decreasing boundaries, nor do we have any knowledge about more short term probabilities. There are also one or two other technical deficiencies within the approximation. Firstly, the tangent approximation is not a probability density, except for the case of a straight line boundary, in which case the approximation is exact. So we are not approximating a density by a density, which could be a problem. Furthermore, a result of Roberts (1993) suggests that the exit distribution is independent of the tangent. In this paper, it is shown that, for Lipschitz continuous boundaries, the hazard rate of the exit distribution is independent of the gradient of the curve. The specific result is quoted as Lemma 2 in section 2.6 of this thesis. Thus, in order to calculate the hazard rate, only the actual boundary value is required, not its derivative. Since hazard rates can be algebraically converted into densities, this fundamentally questions whether the tangent approximation has a solid base, as the tangent, and thus the gradient, form the main terms in the approximation. This result conflicts with the intuition behind the estimation, but smoothness of the boundary makes the tangent approximation work well, especially asymptotically. The approximation works best for boundaries which are approximately straight lines, as the tangent becomes a good approximation to the curve in these cases.

2.5.2 Durbin's Approximation

Another approximation technique was devised by Durbin (1985). We shall keep his notation throughout this subsection. He obtains a result for continuous Gaussian processes, X. Denoting the covariance function by $\rho(s,t)$, the following assumptions on ρ and boundary function $a(t)$ are required, in order to prove a theorem concerning the exit density:

- $a(s)$ is continuous in $0 \leqslant s < t$, and left differentiable at t.
- $\rho(s,t)$ is positive definite, and has continuous first order partial derivatives (on $\{(r,s) : 0 \leqslant r \leqslant s \leqslant t\}$), with the appropriate left and right partial derivatives only, at the boundaries.
- The variance of $X_t - X_s$ satisfies $\lim_{s\uparrow t} \frac{\mathrm{Var}[X_t - X_s]}{t-s} = \lambda_t$, where $0 < \lambda_t < \infty$.

Denoting the exact exit density by $p(t)$, these conditions lead to the result that

$$p(t) = b(t)f(t),$$

where

$$b(t) = \lim_{s \uparrow t}(t-s)^{-1}\mathsf{E}[I(s,X)(a(s) - X_s)\,|X_t = a(t)],$$

in which $I(s,X)$ is the indicator function defined to equal 1 if the sample path does not cross the boundary prior to s, and equals 0 otherwise, and $f(t)$ the density of X_t evaluated at the boundary, that is

$$f(t) = (2\pi\rho(t,t))^{-\frac{1}{2}}\exp[-a^2(t)/2\rho(t,t)].$$

The function $b(t)$ is often intractable, and thus, the approximation technique of Durbin is based on approximating this function.

The first approximation takes $I(s,X) = 1$, thus assuming that if the process hits the boundary, then it is doing so for the first time. This leads to the approximation $p_1(t) = b_1(t)f(t)$, where

$$b_1(t) = \frac{a(t)}{\rho(t,t)}\frac{\partial\rho(s,t)}{\partial s}\bigg|_{s=t} - a'(t).$$

Note that if X is standard Brownian motion, then Durbin's approximation reduces to the tangent approximation. So this technique may be thought of as an extension to the tangent approximation. Clearly, this approximation will become accurate if the boundary becomes remote, in which case, $I(s,X)$ is likely to be 1. A second approximation is also made, based on solving an integral equation for the first exit density, namely

$$p(t) = p_1(t) + \int_0^t [a'(t) - \beta_1(r,t)a(r) - \beta_2(r,t)a(t)]f(t\,|r)p(r)dr,$$

where

$$\begin{bmatrix} \beta_1(r,t) \\ \beta_2(r,t) \end{bmatrix} = \begin{bmatrix} \rho(r,r) & \rho(r,t) \\ \rho(r,t) & \rho(t,t) \end{bmatrix}^{-1} \begin{bmatrix} \rho_2(r,t) \\ \rho_1(t,t) \end{bmatrix}$$

in which $\rho_2(r,t)$ and $\rho_1(t,t)$ are partial derivatives of $\rho(r,t)$ with respect to t and r evaluated at (r,t) and (t,t), and with $f(t\,|r)$ the conditional density of X_t at $a(t)$ given that $X_r = a(r)$. This is derived by conditioning on an intermediate time, and then using the Gaussian properties of X.

The obvious approximation to make in this situation, is to replace the $p(t)$ in the integrand, by the earlier approximation $p_1(t)$. This then forms the basis of a numerical integration to find values of $p(t)$.

In later work, Durbin (1992), Durbin concentrated on Brownian motion exiting a curved boundary, and produced the following formulae for the first exit density across the boundary $a(t)$:

$$p(t) = \sum_{j=1}^{k}(-1)^{j-1}q_j(t) + (-1)^k r_k(t),$$

where

$$\begin{aligned} q_j(t) &= \int_0^t \int_0^{t_1} \cdots \int_0^{t_{j-2}} \left[\frac{a(t_{j-1})}{t_{j-1}} - a'(t_{j-1})\right] \\ &\times \prod_{i=1}^{j-1} \left[\frac{a(t_{i-1}) - a(t_i)}{t_{i-1} - t_i} - a'(t_{i-1})\right] f(t_{j-1}, t_{j-2}, \ldots, t_1, t) dt_{j-1} dt_{j-2} \ldots dt_1, \end{aligned}$$

$$\begin{aligned} r_k(t) &= \int_0^t \int_0^{t_1} \cdots \int_0^{t_{k-1}} b(t_k) \\ &\times \prod_{i=1}^{k} \left[\frac{a(t_{i-1}) - a(t_i)}{t_{i-1} - t_i} - a'(t_{i-1})\right] f(t_k, t_{k-1}, \ldots, t_1, t) dt_k dt_{k-1} \ldots dt_1, \end{aligned}$$

$t_0 = t$, b is as previously defined and $f(t_{j-1}, t_{j-2}, \ldots, t_1, t)$ is the joint density of $B_{t_{j-1}} = a(t_{j-1})$, $B_{t_{j-2}} = a(t_{j-2}), \ldots, B_t = a(t)$.

Durbin proved that $r_k \to 0$ as $k \to \infty$, and consequently the summation converges to the exact density. The benefit of this, is that the summation does not involve $b(t)$, the function which had to be approximated in the original work. The 1992 paper contains results giving bounds on the error term in the concave or convex boundary cases, and illustrates via an example, that often only a few terms of the summation are required. He has thus produced a method to numerically obtain accurate approximations to first exit densities, requiring only a multiple integration package on a computer.

2.5.3 Poisson Clumping Heuristic

A secondary way to consider first hitting times of constant boundaries, is to consider the maximum of a process until a fixed time. Let us define

$$\tau = \inf_{t>0}\{t : X_t \geqslant a\}.$$

Then

$$\{\tau < t\} = \left\{\max_{0 \leqslant s < t} X_s \geqslant a\right\}.$$

The event on the right hand side can be considered, at least for a large, to be a rare event, and in such cases the Poisson clumping heuristic can be employed to approximate its probability. For full details of this idea, see the book by Aldous (1989).

The heuristic result is that the first hitting time to boundary b is exponentially distributed. Specifically,

$$\mathsf{P}[\tau > t] \approx \exp(-\lambda_b t) \qquad \text{for } b \text{ large.}$$

The form of λ_b depends on the size of the approximating Poisson clump being used. For positive-recurrent diffusions, satisfying a stochastic differential equation of the form

$$dX_t = \sigma(X_t)dB_t + \mu(X_t)dt,$$

the parameter λ_b can be expressed as

$$\lambda_b = -\mu(b)f(b),$$

where f is the stationary density of the process.

The result of the first exit times being exponentially distributed is a common one, and is connected with the quasi-stationary distributions to be discussed in section 2.6. The papers by Nobile, Ricciardi and Sacerdote (1985a,b) also find this exponential distribution result. In the 1985a paper, they looked at the Ornstein-Uhlenbeck process, exiting constant boundary a. In the case where the boundary $a \to \infty$, they showed

$$p(t) = \frac{1}{\alpha}\exp\{-t/\alpha\} + o\left(\frac{1}{\alpha}\exp\{-t/\alpha\}\right),$$

where α is the mean hitting time for the constant boundary, height a, from a starting point of zero. This is found by expanding the Laplace transform of p in terms of moments of the first exit times, and using limiting results such as

$$\lim_{a \to \infty} \frac{t_k(a \,|x_0)}{k![t_1(a \,|0)]^k} = 1,$$

where $t_k(a \,|x_0)$ is the kth moment of the hitting time of a, when the process is started at $X_0 = x_0$. They also produce an expression for $\alpha = t_1(a \,|0)$, which enables the approximation to be used in practical applications. In a further paper, Nobile, Ricciardi and Sacerdote (1985b), this result is extended for processes with stationary distributions, letting the constant boundary recede to the extremes of the process' distribution.

2.6 Hazard Rate Approximations

In this section, we discuss methods for approximating the first exit distribution using approximations to the hazard rate of the hitting time. We review the methods of Roberts (1991b), which use quasi-stationary distributions to produce asymptotic approximations to the hazard rate, and therefore the distribution function, of the first exit time. We then extend some of these ideas to produce a new approximation technique, which is more suitable for approximating the distribution for smaller times, by taking the initial conditions into account. The method is not restricted to Brownian motion, and for some processes can produce an approximation for a one-sided hitting time distribution.

The previously discussed methods have all been approximations for either the density or distribution function of the first hitting time. We shall concentrate on approximating the hazard rate, $h(t)$, which can easily be converted to the density, p, using

$$p(t) = h(t) \exp\left\{-\int_0^t h(s)ds\right\}$$

and the distribution function, P, by

$$P(t) = 1 - \exp\left\{-\int_0^t h(s)ds\right\}.$$

Some justification for estimating the hazard rate, and obtaining the distribution function from it, is given in Chapter 7, where the tangent approximation (Strassen (1967)) is found to be inferior to a new approximation based on estimating hazard rates. Another reason for concentrating on the hazard rate is its relationship with the distribution of the process itself. Using the notation

$$\mu_t(x)dx = \mathbb{P}[X_t \in dx \,|\tau > t],$$

Roberts (1993) proves

Lemma 2 (Roberts (1993)) *For an arbitrary Lipschitz boundary $f(t)$, and an Itô diffusion X (see section 2.2.2),*

$$h(t) = \frac{1}{2}\sigma^2(t, f(t)) \lim_{x \uparrow f(t)} \frac{\mu_t(x)}{f(t) - x},$$

where σ is the diffusion coefficient of X.

Frequently, the distribution of the process, conditioned not to have hit the boundary, is accessible, at least approximately, and therefore we can use Lemma 2 to produce an approximation to the hazard rate.

If the distribution μ_t is quasi-stationary,

$$\left[X_s \,\middle|\, \tau > s\right] \sim \left[X_t \,\middle|\, \tau > t\right] \qquad \text{for all } s \geqslant t,$$

and consequently the hazard rate in constant. Therefore, we now review the methods of Roberts (1991b), which use such distributions in the approximation of hazard rates.

2.6.1 Quasi-Stationary Distributions

Quasi-stationary distributions are limiting probability distributions which exist in completely time homogeneous systems. Later, we will show how to extend these methods to approximate the boundary hitting time in the inhomogeneous case. However, we begin by assuming that X is a time homogeneous diffusion process, and define

$$\tau = \inf_{t>0}\{t : |X_t| \geqslant a\},$$

where a is a constant, and let $h(t)$ denote the hazard rate of τ. We define the quasi-stationary distribution δ_∞ as the limiting distribution

$$\delta_\infty \sim \lim_{t\uparrow\infty} \left[X_t \,\middle|\, \tau > t\right].$$

For a detailed discussion about the existence of such a distribution, see Jacka and Roberts (1987). We note that for "well-defined" diffusions on bounded domains, such a limit does exist.

We now consider the possibility of X_0 having a starting distribution, rather than a starting point. Suppose $[X_0] \sim \delta_0$, then

$$\begin{aligned}
\mu_t(y) = \mathbb{P}[X_t \in dy \,|X_0 \sim \delta_0, \tau > t] &= \int \mathbb{P}[X_t \in dy \,|X_0 = x, \tau > t]\delta_0(x)dx \\
&= \int \frac{\mathbb{P}[X_t \in dy, \tau > t \,|X_0 = x]}{\mathbb{P}[\tau > t \,|X_0 = x]}\delta_0(x)dx \\
&= \int \overline{p}_t(x,y)\delta_0(x)dx, \qquad \text{say.}
\end{aligned}$$

If we denote by $p_t^a(x,y)$ is the transition density of the process X with absorption at the boundary $\pm a$, then $\overline{p}_t$ is the normalised version of p_t^a.

It is known that δ_∞ can be expressed as a solution to the integral equation

$$e^{\alpha t}\delta_\infty(y) = \int_{-a}^{a} p_t^a(x,y)d\delta_\infty(x) \qquad \text{for all } t,\ |y| < a, \tag{2.4}$$

where $\alpha < 0$ is a left eigenvalue. If we assume $[X_0] \sim \delta_\infty$, then integrating both sides of (2.4) with respect to y yields

$$e^{\alpha t} = \mathsf{P}[\tau > t] \qquad \text{for all } t,$$

so that the hitting time has a constant hazard rate $(-\alpha)$. From the definition of δ_∞ being the limiting distribution of the conditioned process, we deduce that, asymptotically, the first exit time has hazard rate $-\alpha$, for all starting distributions.

Because the state space of the conditioned process is dependent on a, it follows that the quasi-stationary distribution is also dependent on a. We use the notation δ^a_∞, when the boundary is at $\pm a$, and h^a to denote the corresponding constant hazard rate.

2.6.2 Time Inhomogeneous Approximations

In this subsection, we assume that X is a time homogeneous diffusion process, and let

$$\tau_f = \inf_{t>0}\{t : |X_t| \geqslant f(t)\},$$

which introduces some time inhomogeneity into the system. Because the boundary is not necessarily constant, the limiting distribution

$$\lim_{t\uparrow\infty} \left[X_t \,\middle|\, \tau_f > t\right]$$

does not exist, and so we cannot directly apply the idea of the previous section. We define the *frozen* boundary $f_t(s)$ by

$$f_t(s) = \begin{cases} f(t) & s \leqslant t \\ f(s) & s > t \end{cases},$$

so that the boundary is a constant until time t. We can then use the hazard rate for $\tau_{f_t} = \inf_{s>0}\{s : |X_s| \geqslant f_t(s)\}$ to approximate the hazard rate of τ at time t. This situation is one with constant boundaries and is of the form described in the previous section. We therefore assume that

$$\left[X_t \,\middle|\, \tau_f > t\right] \sim \delta^{f(t)}_\infty. \tag{2.5}$$

Clearly this idea will work best if the boundary f is approximately constant.

Using this assumption ((2.5)), the hazard rate at time t is $h^{f(t)}$. Thus, we make the approximation

$$\mathsf{P}[\tau_f > t] = O\left(\exp\left\{-\int_0^t h^{f(s)} ds\right\}\right) \qquad \text{as } t \to \infty.$$

As an example of this, consider Brownian motion exiting a two-sided functional boundary $f(t)$. If we define

$$\tau_f = \inf_{t>0}\{t : |B_t| \geqslant f(t)\},$$

the hazard rate of τ_f at time t is approximated by

$$h^{f(t)} = \frac{\pi^2}{8f^2(t)},$$

which leads to the asymptotic approximation

$$\mathbb{P}[\tau_f > t] = O\left(\exp\left\{-\int_0^t \frac{\pi^2}{8f^2(s)}ds\right\}\right) \qquad \text{as } t \to \infty.$$

This result can be proved rigorously (see Roberts (1991b)), using monotonicity arguments which we extend in Chapters 3 – 6.

This idea is extensively investigated in Roberts (1991b), where the Ornstein-Uhlenbeck process is used, and a time change result allows the exit distribution for Brownian motion across an approximately square root boundary to be investigated. This time change was discussed in section 2.2.2.

For time inhomogeneous processes, we can make approximations by *freezing* the time components in the drift and diffusion coefficients, to their values at time t, and using the results associated with the homogeneous process with the same (*frozen*) coefficients.

The biggest problem with this approximation is that it assumes the distribution at time t of the conditioned process is $\delta_\infty^{f(t)}$. This is only a good approximation when f is asymptotically approximately constant, and t is large. The short term distribution of the process is dominated by the initial conditions, and it is unlikely that approximating this by $\delta_\infty^{f(t)}$ will be accurate.

The other obvious problem is that the method is best suited for boundaries which are asymptotically approximately constant, although this can be extended to approximately square root for the Brownian motion case, as in Roberts (1991b).

2.6.3 The Hazard Rate Approximation

In this section, we assume that X is a time homogeneous diffusion process, and will consider X with different starting distributions. We will compare the behaviour from different starting measures in order to produce a new approximation, which takes into account the initial conditions. This new approximation method may then work well for small time approximations, as well as for large times.

We seek the hazard rate of

$$\tau = \inf_{t>0}\{t : |X_t| \geqslant a\},$$

with $X_0 = 0$. We introduce the following notation

$$\begin{array}{rcl} X^i & & \text{process with starting measure } m_i \\ \tau_i & = & \inf_{t>0}\{t : |X_t^i| \geqslant a\} \\ h_i & & \text{hazard rate of } \tau_i \\ \nu_t^i(x)dx & = & \mathsf{P}[X_t^i \in dx] \\ \mu_t^i(x)dx & = & \mathsf{P}[X_t^i \in dx \,|\tau_i > t] \\ \lambda^i(t) & = & \lim_{x\uparrow a} \frac{\mu_t^i(x)+\mu_t^i(-x)}{a-x}. \end{array}$$

From Lemma 2 (adapted for the two-sided case by adding the densities at the positive and negative boundaries) we have

$$\frac{h_2(t)}{h_1(t)} = \frac{\lambda^2(t)}{\lambda^1(t)}, \tag{2.6}$$

and as we wish to find the hazard rate of τ with $X_0 = 0$, we shall select m_2 to be a point mass at zero. The discussion in section 2.6.1 suggests we choose $m_1 \sim \delta_\infty$, so that the hazard rate $h_1(t)$ is constant. We need, then, only approximate the right hand side of (2.6). We note

$$\frac{\lambda^2(t)}{\lambda^1(t)} = \lim_{x\uparrow a} \frac{\mu_t^2(x) + \mu_t^2(-x)}{\mu_t^1(x) + \mu_t^1(-x)}, \tag{2.7}$$

and make the approximation

$$\frac{\mu_t^2(x) + \mu_t^2(-x)}{\mu_t^1(x) + \mu_t^1(-x)} \approx \frac{\nu_t^2(x) + \nu_t^2(-x)}{\nu_t^1(x) + \nu_t^1(-x)}. \tag{2.8}$$

Note this approximation is based on the heuristic principle that the distributions of the processes with different starting measures are influenced more by the initial conditions than the conditioning.

Combining (2.6), (2.7) and (2.8), our approximation is

$$h_2(t) \approx h \lim_{x\uparrow a} \frac{\nu_t^2(x) + \nu_t^2(-x)}{\nu_t^1(x) + \nu_t^1(-x)}, \tag{2.9}$$

where h is the constant hazard rate associated with the δ_∞ distribution. Expressions for both unconditioned distributions, ν_t^i, can be found, providing the unconditioned process transition density is known. Thus, we can find an explicit formula for $h_2(t)$,

$$h_2(t) \approx h \frac{\nu_t^2(a) + \nu_t^2(-a)}{\nu_t^1(a) + \nu_t^1(-a)}. \tag{2.10}$$

We note that Roberts (1993) produces an inequality in (2.10) in the case where

$$\tau = \inf_{t>0}\{t : B_t \geqslant f(t)\},$$

and starting measures which are strongly stochastically ordered (see Chapter 3 for details of strong stochastic ordering). He shows

Lemma 3 (Roberts (1993, Theorem 3.5)) *Suppose the measures* m_1 *and* m_2 *satisfy* $m_2 \overset{\text{sst}}{\geqslant} m_1$ *with* $\text{supp}(m_i) \subset (-\infty, f(0)]$ *for* $i = 1, 2$. *Then*

$$\frac{h_2(t)}{h_1(t)} \leqslant \frac{\nu_t^2(f(t))}{\nu_t^1(f(t))}.$$

2.6.4 Extensions to General Boundaries

The existence of a quasi-stationary distribution for

$$\lim_{t\uparrow\infty} \left[X_t \,\middle|\, \tau > t\right],$$

where X is a time homogeneous diffusion process, is reliant on the boundary being constant. If this was not true, the state space of the conditional distribution would vary over time, and δ_∞ would not exist in general. We will require exit distribution approximations for general functional boundaries, and so we must adapt the method of section 2.6.3 to allow for this. We use the notation δ_∞^a for the quasi-stationary distribution, and h^a for the associated hazard rate, when

$$\tau = \inf_{t>0}\{t : |X_t| \geqslant a\}.$$

In order to calculate the hazard rate for the exit time

$$\tau_f = \inf_{t>0}\{t : |X_t| \geqslant f(t)\}$$

at time t, we define the boundary (as in section 2.6.2)

$$f_t(s) = \begin{cases} f(t) & s \leqslant t \\ f(s) & s > t \end{cases},$$

so that f_t is constant until time t. As such, the hazard rates of the hitting time of f_t, associated with the δ_∞ distribution (constant $a = f(t)$) will be correct until time t. We use this hazard rate to approximate the hazard rate of τ_f at time t. Therefore, the final form of the approximation for a general functional boundary is

$$h_2(t) = h^{f(t)} \frac{\nu_t^2(f(t)) + \nu_t^2(-f(t))}{\nu_t^1(f(t)) + \nu_t^1(-f(t))}. \tag{2.11}$$

Because of the approximation of f by the *frozen* boundary until time t, the method works best for boundaries which are approximately constant. We shall refer to this technique as the UDHRR method.

Example: Brownian motion

Let

$$\tau_a = \inf_{t>0}\{t : |B_t| \geqslant a\},$$

in which case, it is easy to show that

$$d\delta_\infty^a(x) = \frac{\pi}{4a}\cos\left(\frac{\pi x}{2a}\right)dx \qquad x \in [-a, a],$$

which has a corresponding hazard rate

$$h^a = \frac{\pi^2}{8a^2}.$$

We can find an expression for ν_t^1 as an integral, and ν_t^2 is a normal distribution. For a fixed boundary at $\pm a$, we can produce the approximation

$$h_2(t) \approx \frac{\pi e^{-a^2/2t}}{a\int_{-a}^{a}\left(e^{-(a-y)^2/2t} + e^{-(a+y)^2/2t}\right)\cos\left(\frac{\pi y}{2a}\right)dy}. \tag{2.12}$$

Making the necessary approximations by *freezing* the boundary prior to t, in the case

$$\tau_f = \inf_{t>0}\{t : |B_t| \geqslant f(t)\},$$

the general expression for the hazard rate of τ_f, given $B_0 = 0$, is

$$h_2(t) \approx \frac{\pi e^{-f^2(t)/2t}}{f(t)\int_{-f(t)}^{f(t)}\left(e^{-(f(t)-y)^2/2t} + e^{-(f(t)+y)^2/2t}\right)\cos\left(\frac{\pi y}{2f(t)}\right)dy}\ . \tag{2.13}$$

Some numerical examples using (2.13) are featured in section 2.7.

Example : Brownian Motion across Square Root Boundaries

The existence of quasi-stationary distributions is not restricted to Brownian motion. If X is an Ornstein-Uhlenbeck process and

$$\tau = \inf_{t>0}\{t : |X_t| \geqslant a\},$$

then the limiting distribution

$$\delta_\infty \sim \lim_{t\uparrow\infty} \left[X_t \,|\tau > t\right]$$

exists, and is in the form of a confluent hypergeometric function (see Erdelyi (1953)). Therefore, we can use the same methodology of 2.6.3 to deduce

$$h_2(t) \approx \frac{m(a)(p_t(0,a) + p_t(0,-a))}{\int_{-a}^{a}(p_t(x,a) + p_t(x,-a))\delta_\infty^a(x)dx} \tag{2.14}$$

where $p_t(x,y)$ is the transition density of the standard Ornstein-Uhlenbeck process, and $m(a)$ is the hazard rate associated with the δ_∞^a distribution. We now note that the Ornstein-Uhlenbeck process is a time changed Brownian motion:

$$X_t = e^{-t/2}B_{e^t}.$$

Therefore,

$$\{|X_t| \geqslant a\} \equiv \{|e^{-t/2}B_{e^t}| \geqslant a\} \equiv \{|B_{e^t}| \geqslant ae^{t/2}\} \equiv \{|B_s| \geqslant a\sqrt{s}\},$$

where $s = e^t$, and so knowledge of the distribution of

$$\tau = \inf_{t>0}\{t : |X_t| \geqslant a\}$$

allows us to obtain the distribution of

$$\rho = \inf_{s>1}\{s : |B_s| \geqslant a\sqrt{s}\}.$$

Clearly, if

$$\tau = \inf_{t>0}\{t : |X_t| \geqslant f(t)\},$$

we can use the approximation of a *frozen* boundary until time t to deduce an approximation. For boundaries, f, which are approximately constant, the approximation should work well. We note that if the boundary is roughly constant in the Ornstein-Uhlenbeck case, this translates to approximately square root in the time changed case, for Brownian motion. Thus, we have an approximation for Brownian motion exiting approximately square root boundaries, which ought to work well.

Remarks

Note that for some processes quasi-stationary distributions exist for one-sided boundaries. The same method will therefore produce approximations to the one-sided hitting time, in this case.

For Brownian motion, we are restricted to two-sided boundaries, and so the method complements the tangent approximation (Strassen (1967)) which is only applicable for one-sided boundaries.

Finally we note that all future references to the UDHRR method, in numerical examples, refer to the case using δ_∞ distributions for Brownian motion, producing formula (2.13). We use this formula in examples, even when the boundary is closer to square root in shape, than constant, in which case the method based on the Ornstein-Uhlenbeck process, formula (2.14), and a time change should work better.

2.7 Numerical Examples

We conclude this chapter with some numerical illustrations of the accuracy of the tangent approximation (section 2.5.1), and the UDHRR approximation (section 2.6.4). We shall use the abbreviation TA to denote the tangent approximation in the actual examples, and HRT will be used to represent the hazard rate tangent approximation, which will be introduced in section 7.6. We shall compare both exit densities and distribution functions, and in all cases the process is assumed to be Brownian motion. Note that more examples of the tangent approximation are given in Figs 7.1 – 7.6, where they are viewed in comparison with the analytic bounds and HRT approximation developed in Chapter 7.

Recall that direct comparisons between the tangent approximation (Strassen (1967)) and the UDHRR method (section 2.6.4) cannot be made, since one assumes one-sided boundaries, the other two-sided boundaries. However, in some situations, if the boundaries are sufficiently remote, we may approximate the one-sided exit density by halving the two-sided exit density. This assumes it is very unlikely that the process hits both upper and lower boundaries prior to the particular time.

For the UDHRR approximation, we will use formula (2.13), which works best if the boundary is approximately constant. We could improve these approximations for boundaries close to square root in shape, by using formula (2.14) associated with the Ornstein-Uhlenbeck process, and making the necessary time change.

For the numerical examples we considered, simple functional forms of boundaries were selected, which might be more applicable in practical situations, rather than the more complicated functions implicitly defined by the method of images (Daniels (1982)). This does cause the problem that the exact exit distributions are unknown, and will be estimated by simulating 200000 Brownian motion paths, to obtain an empirical first hitting time distribution.

To illustrate the tangent approximation, the boundary is $f(t) = \sqrt{t+4}$. This curve is increasing, with a continuous derivative and $f(t)/\sqrt{t}$ monotone decreasing, and thus the tangent approximation theoretically produces an asymptotically accurate density. Furthermore, the boundary is concave, and so the HRT approximation would be expected to be superior to the tangent approximation. The comparisons between the distribution functions and density functions are shown in Fig 2.1. As expected, the tangent approximation over-estimates the true density (as the curve is concave), and the HRT method approximately halves the error in the distribution function for large values of t.

The accuracy of the approximation for the density function is good for $t \geqslant 20$, where the theory suggests it should be. Note that for $2 \leqslant t \leqslant 10$, the simulated density is larger than the approximations — theoretically impossible. This suggests inaccuracy in the simulation, with insufficient Brownian paths being generated. However, this illustrates one advantage of the tangent approximation and the HRT method, in that they take a few seconds and a few minutes, respectively, to calculate. The simulation took about 100 times as long as the HRT method to produce the distribution, and for large values of t, the difference in the densities is negligible.

The illustration of the UDHRR method in Fig 2.2 is for the constant boundary $f(t) = 5$. In this case, the exact distribution function has been found using the eigenfunction expansion technique (formula (2.2) with $x = 0$). Notice that this is the case where we expect the approximation to work well, but in fact, it underestimates the true value of the distribution function. This is not unexpected if we recall Lemma 3.

The example in Fig 2.3 compares all the approximations, when the boundary curve is $f(t) = 2 + 0.05(t + 8\sqrt{t})$. The one-sided hitting time distribution was simulated, and so values obtained from the UDHRR method were halved, ignoring the probability of double hits. We draw three conclusions from Fig 2.3:

1. The tangent approximation is again out-performed by the HRT method, as expected, because the boundary is concave.

2. The UDHRR approximation works very well for small values of t, whereas the tangent approximation, and especially the HRT method, work well for large times. The ideal approximation might be a composite of the UDHRR and the HRT densities, using the UDHRR method for small times only, and the HRT method for intermediate and large times.

Fig 2.1a - Distribution Function Comparison

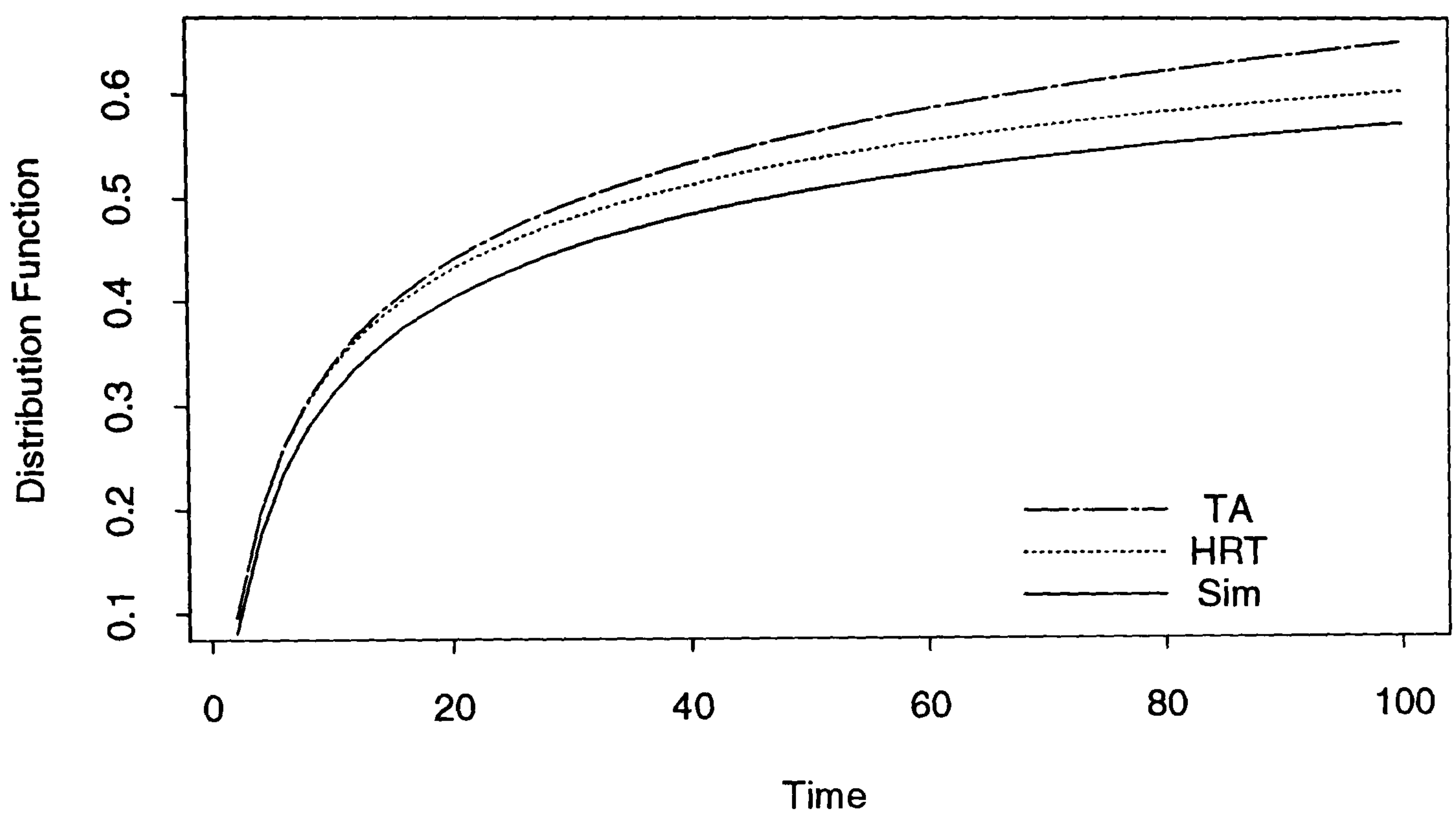
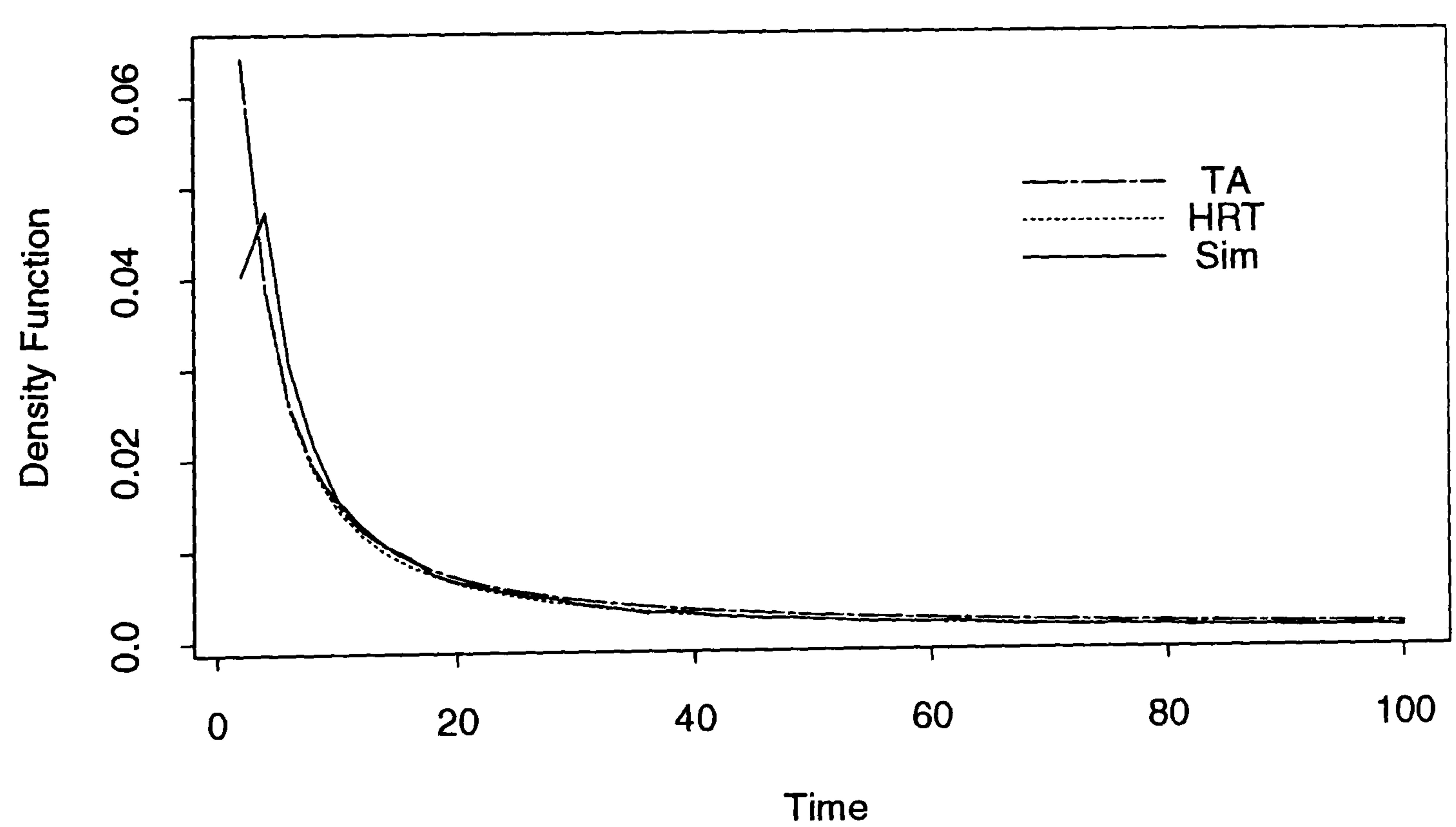


Fig 2.1b - Density Function Comparison



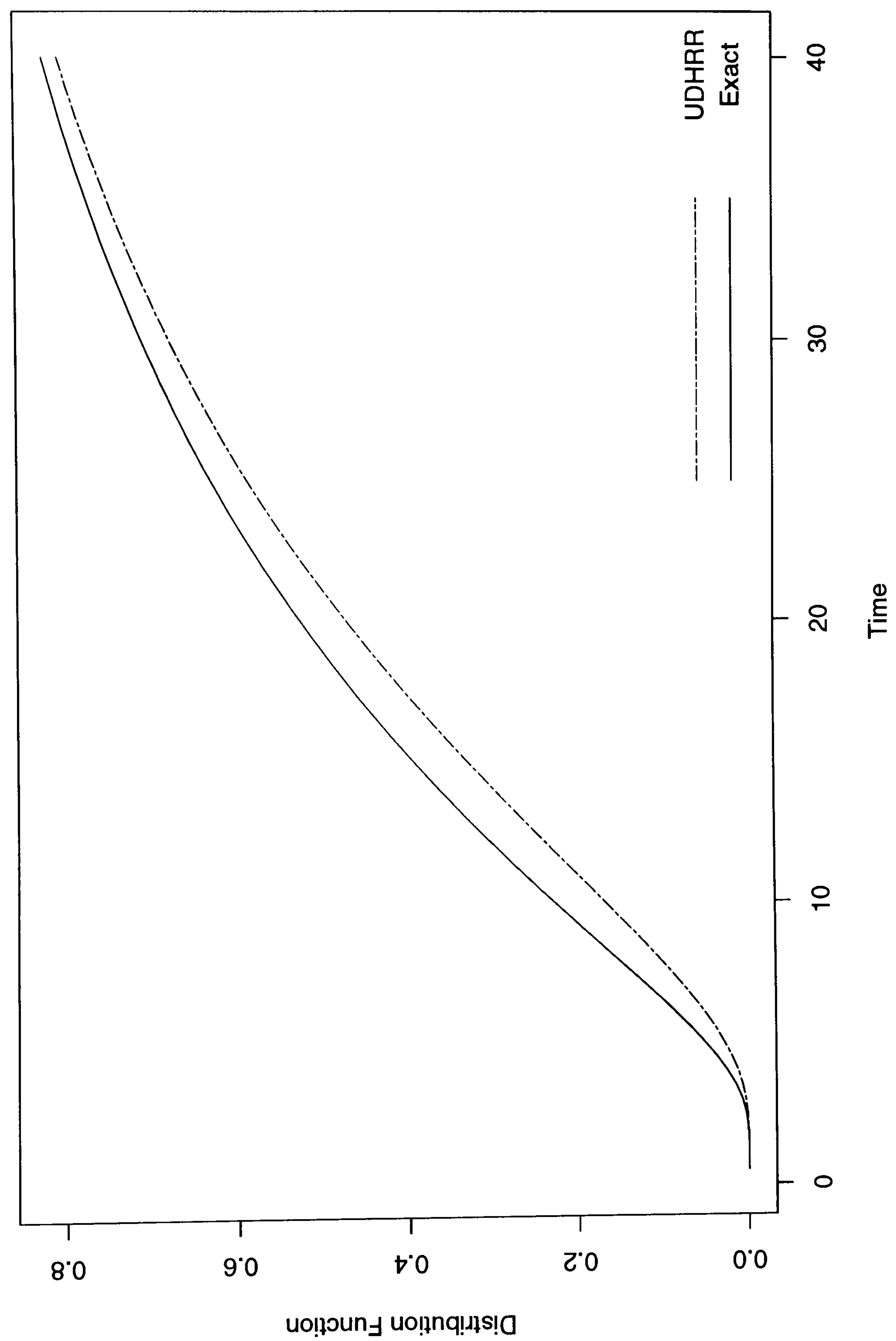


Fig 2.2 - Distribution Function Comparison

3. The UDHRR method under-estimates the density for large time values. However, for large time values, the probability of double hitting ought to become significant, and these are counted as hits by the one-sided tangent approximation, but not by the UDHRR method, because of the earlier hit. Thus, the UDHRR method would be expected to under-estimate the distribution function for large values of t. Note also, the boundary is closer to square root, than constant, and a change of time scale, and the use of the δ_∞ distributions for Ornstein-Uhlenbeck processes might be more appropriate.

The best approximation produced by the UDHRR method is illustrated in Fig 2.4 for the boundary curve $f(t) = 6 - 3e^{-t/10}$. Bearing in mind this boundary is not constant, we might expect the approximation to be worse than that of Fig 2.2, for $f(t) = 5$. However, because the boundary is increasing, the *frozen* boundary, $f_t(s)$, used to obtain (2.13) satisfies

$$f_t(s) \geqslant f(s) \qquad \text{for all } s \leqslant t.$$

Therefore, the hazard rate using this boundary over-estimates the true hazard rate (see Theorem 15, Chapter 7). Clearly, this over-estimation must cancel with the natural under-estimation, exhibited in Fig 2.2, leading to a very accurate approximation.

The "spikiness" of the simulated density is due to insufficient simulated paths. However, the simulation takes about 60 times as long as the UDHRR method to produce the distribution, and the difference is marginal. This time saving indicates why approximations are sought as an alternative to simulating.

The example illustrated in Fig 2.5 is for the boundary $f(t) = 5 - t/2$, so that

$$\tau = \inf_{t>0}\{t : |B_t| \geqslant f(t)\} \leqslant 10.$$

This boundary is decreasing, and so the hazard rate is an under-estimation, and no cancellation of errors occurs, as is the case in Fig 2.4. Thus, the distribution function is an under-estimate, though reasonably accurate considering the boundary is not close to constant.

We summarise, below, the boundaries used in the examples:

Fig 2.1	$f(t) = \sqrt{4+t}$.
Fig 2.2	$f(t) = 5$.
Fig 2.3	$f(t) = 2 + 0.05(t + 8\sqrt{t})$.
Fig 2.4	$f(t) = 6 - 3e^{-t/10}$.
Fig 2.5	$f(t) = 5 - t/2$.

Fig 2.3a - Distribution Function Comparison

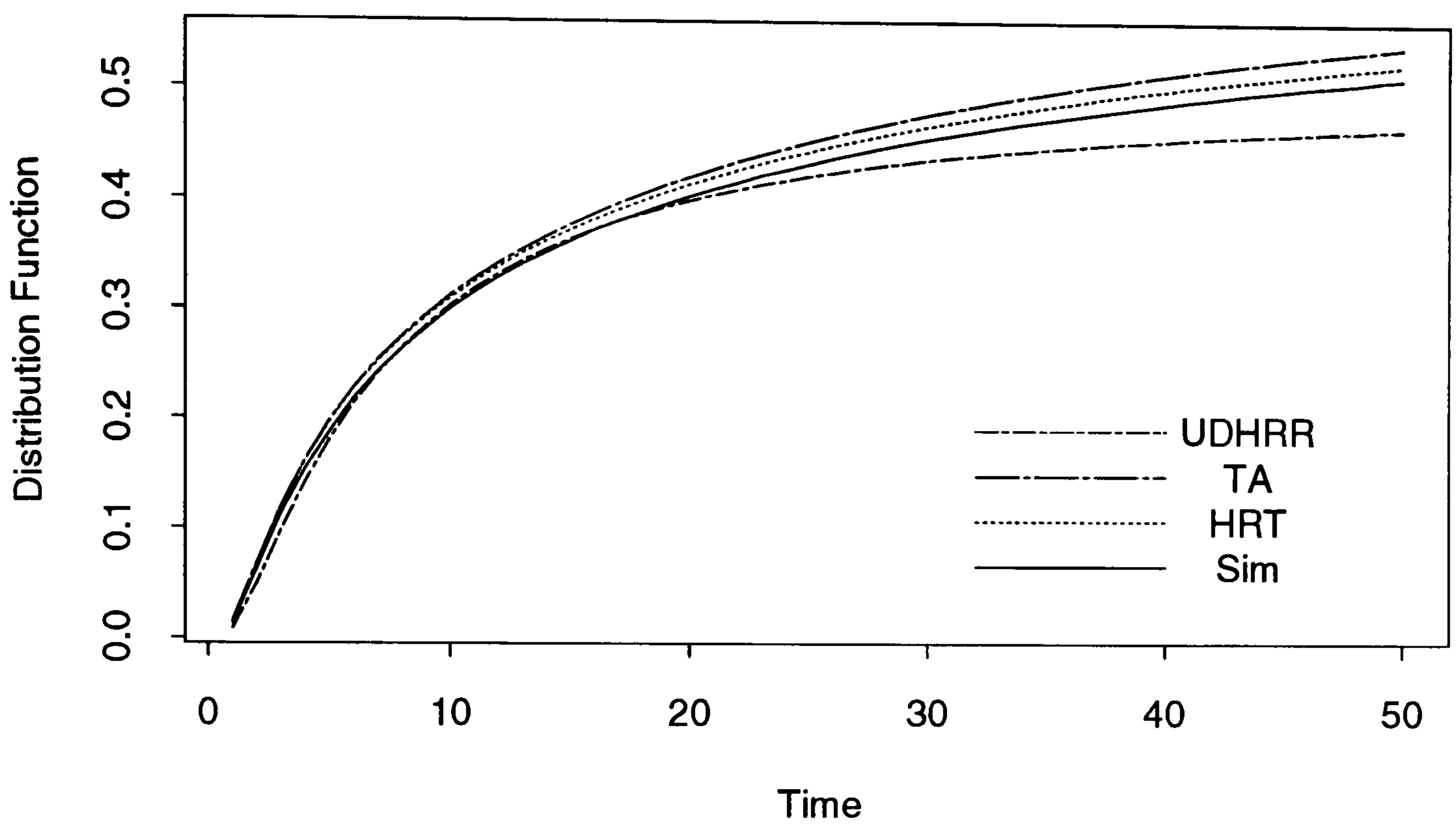


Fig 2.3b - Density Function Comparison

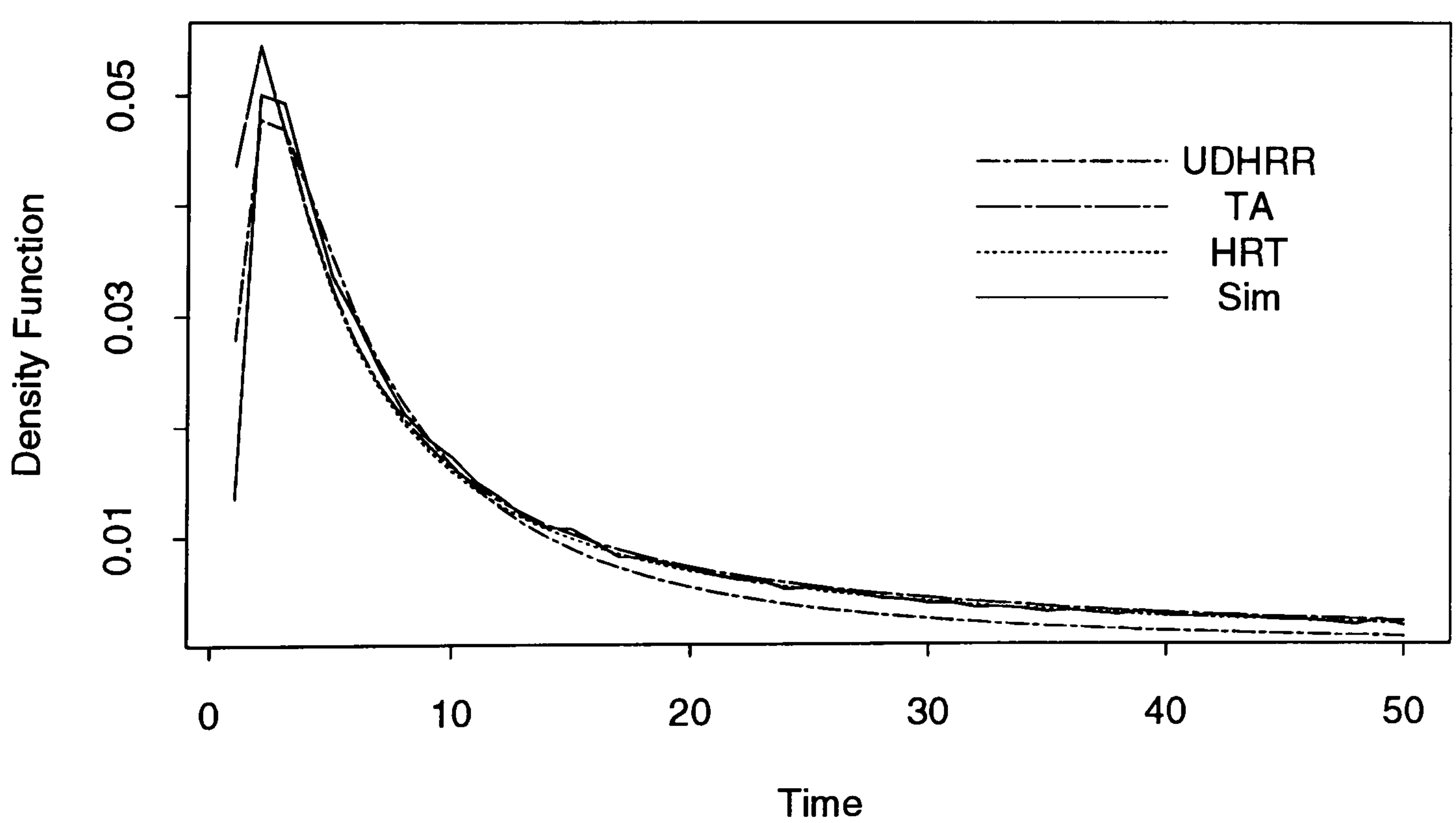


Fig 2.4a - Distribution Function Comparison

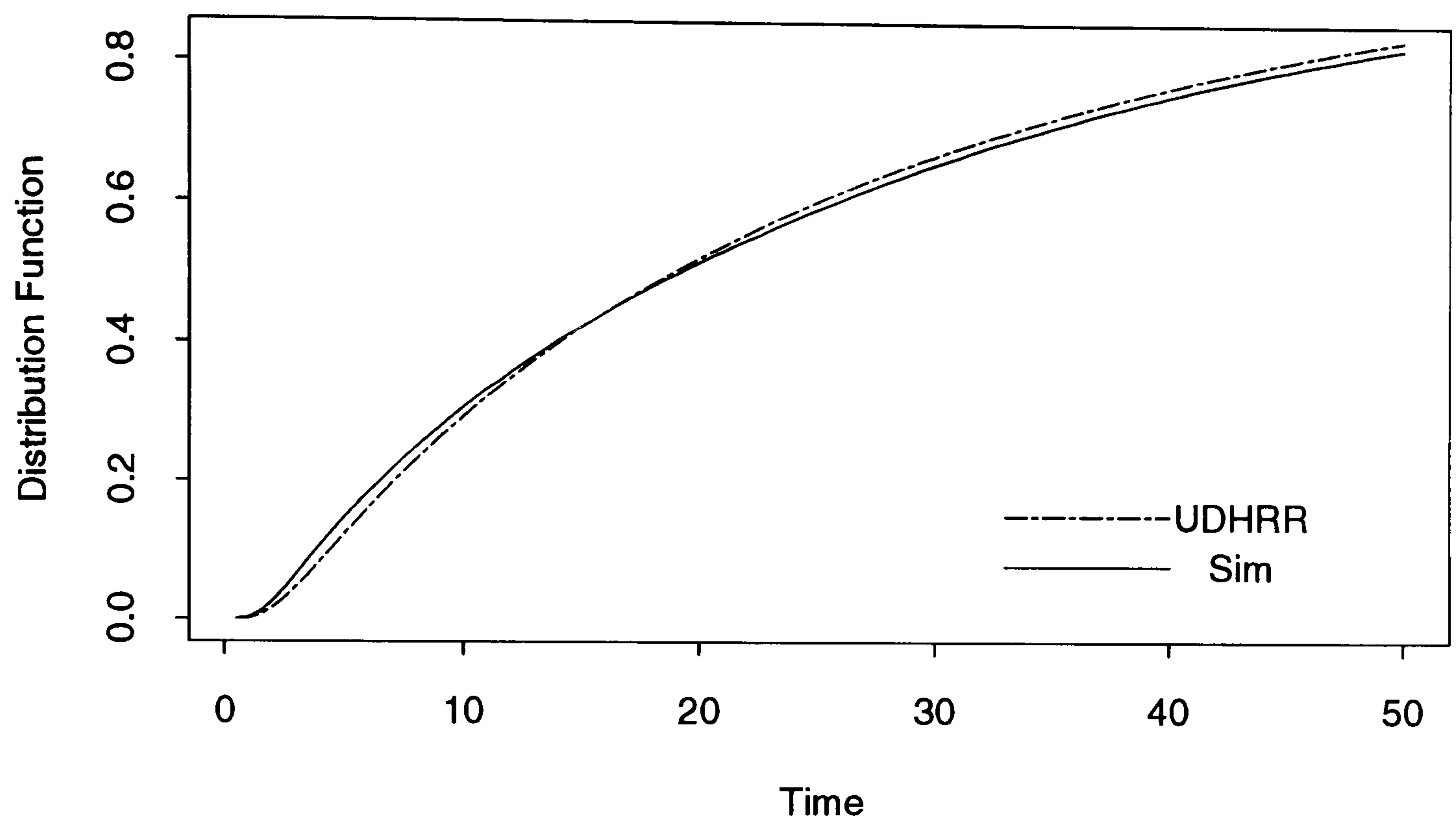


Fig 2.4b - Density Function Comparison

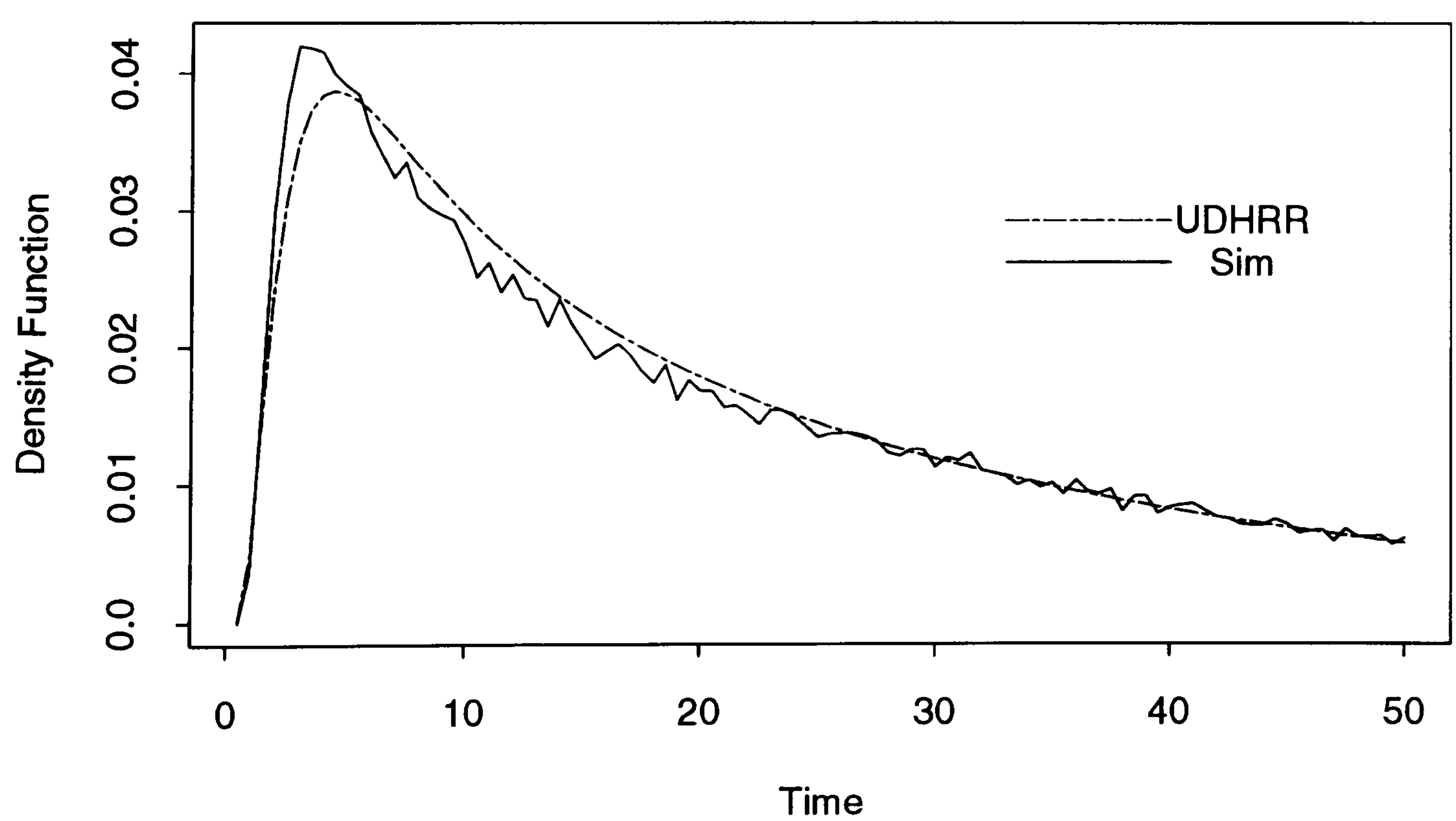


Fig 2.5a - Distribution Function Comparison

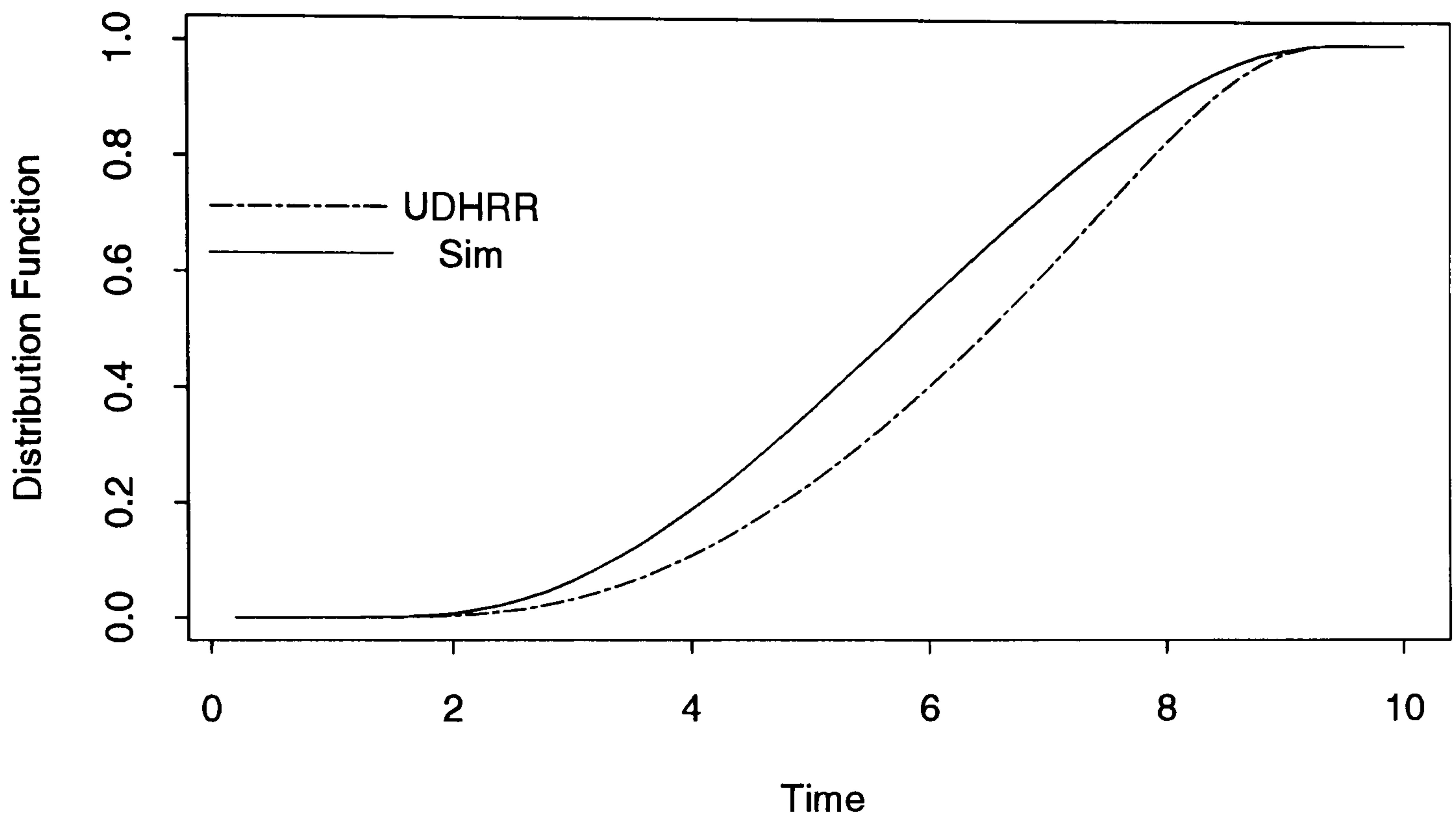
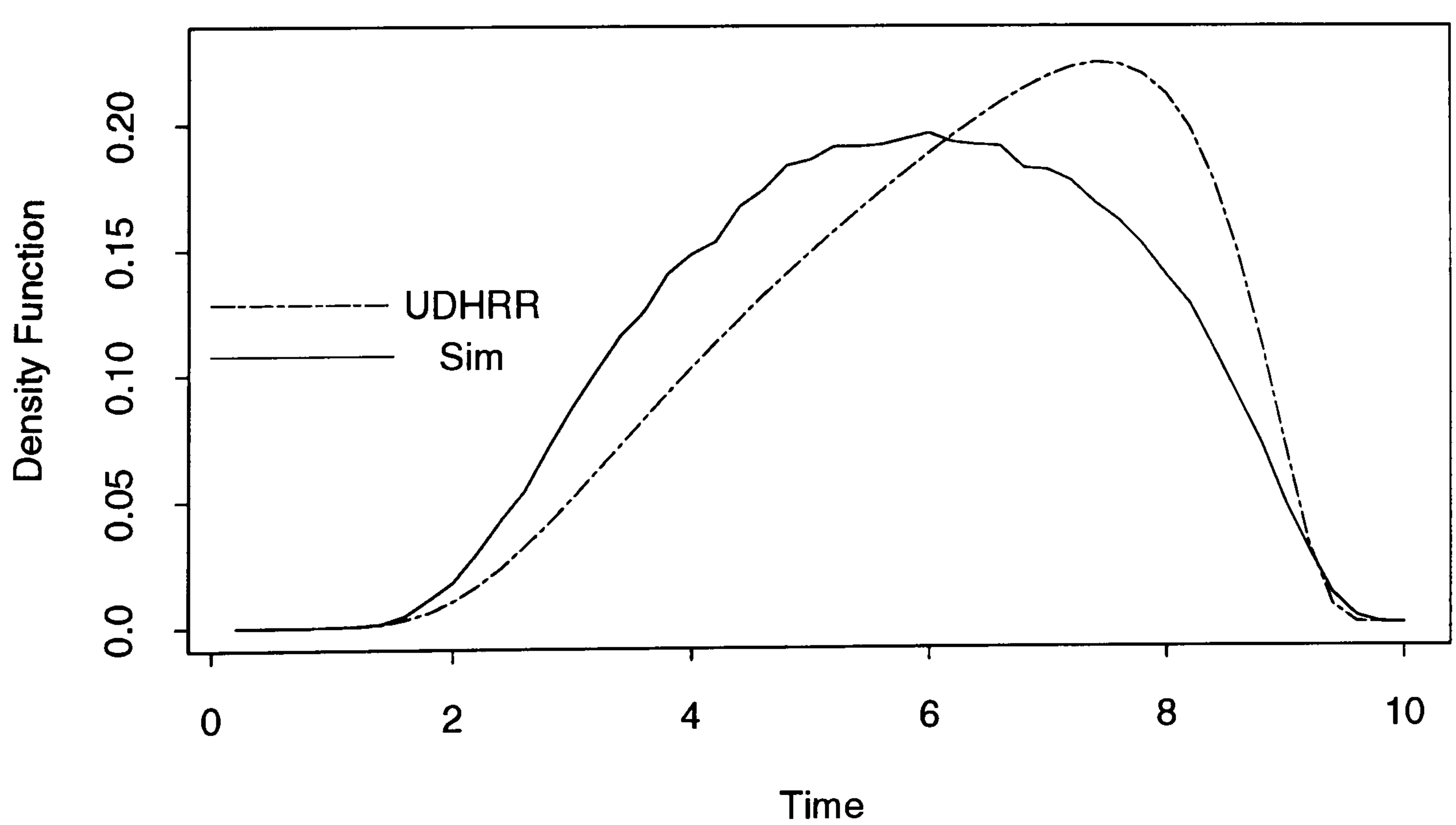


Fig 2.5b - Density Function Comparison



Chapter 3

Introduction to Ordering and Conditioning

Suppose we are given two random variables, X and Y, taking values on some totally ordered space, whose distributions we wish to compare. Strong (a.s.) comparisons are restricted to a particular probability space, $(\Omega, \mathcal{F}, \mathbb{P})$, on which it may be true that $\mathbb{P}[X(\omega) \geqslant Y(\omega)] = 1$. This is known as almost sure ordering, and written $X \overset{\text{a.s}}{\geqslant} Y$. When reference to a probability space is ignored, we can only produce weak orderings, such as stochastic ordering, when $\mathbb{P}[X \leqslant x] \leqslant \mathbb{P}[Y \leqslant x]$ for all x. Note that this is not specific to any probability space. Strong stochastic ordering is defined using the densities f_X and f_Y of random variables X and Y. We write $X \overset{\text{sst}}{\geqslant} Y$ if

$$\frac{f_X(x)}{f_Y(x)} \quad \text{is non-decreasing in } x.$$

An alternative definition, not using density functions, is also available. We also have $X \overset{\text{sst}}{\geqslant} Y$ if

$$\mathbb{P}[X \in A_2]\mathbb{P}[Y \in A_1] \geqslant \mathbb{P}[X \in A_1]\mathbb{P}[Y \in A_2]$$

for all A_1, A_2 such that $a_1 \in A_1$ and $a_2 \in A_2$ implies $a_1 \leqslant a_2$ holds. The first definition is usually quicker to verify.

In this chapter, we shall make formal definitions of these ordering concepts, when extended to Markov processes. A particular category of ordering results compare the process X with $X \,|\, \tau > t$, where $\tau = \inf_{t>0}\{t : X_t \notin A\}$. We shall give a selection of such comparison results, together with a well known almost sure ordering result for diffusion processes satisfying particular forms of stochastic differential equations.

A use of ordering techniques occurs in the field of sequential analysis, where the ordering of processes with different parameters, can be exploited when making some inferences about the unknown parameter, (see for example, Bather (1988)). Both stochastic and strong stochastic ordering may be used in such a way.

Stochastic ordering can also be defined by

$$X_2 \overset{\text{st}}{\geqslant} X_1 \text{ if } \mathsf{E}[f(X_2)] \geqslant \mathsf{E}[f(X_1)] \text{ for all non-decreasing functions } f.$$

This type of definition has been extended to produce new orderings. For example, we could define

$$X_2 \overset{\text{cx}}{\geqslant} X_1 \text{ if } \mathsf{E}[f(X_2)] \geqslant \mathsf{E}[f(X_1)] \text{ for all non-decreasing, convex functions } f.$$

For further details, and examples, of this type of ordering, see Shaked and Shanthikumar (1988) and Stoyan (1983). Levy (1992) defines stochastic orderings which are weaker than those with which we shall be interested.

3.1 Definitions

Throughout this section we assume all processes are Markov processes, possibly time inhomogeneous, and have densities with respect to the Lebesgue measure, for $t > 0$ and $x \in \mathbb{R}$. We shall then use the following notation:

X^i		$(i = 1, 2)$	- Processes for comparison.
$P_t^i(x)$	$=$	$\mathsf{P}[X_t^i \leqslant x]$	- Distribution functions of X_t^i
$p_{s,t}^i(x,y)$	$=$	$\mathsf{P}[X_t^i \in dy \,\vert X_s = x]$	- Transition densities.
$p_t^i(x,y)$	$=$	$\mathsf{P}[X_t^i \in dy \,\vert X_0 = x]$	- Transition densities.
$p_t^i(y)$	$=$	$\mathsf{P}[X_t^i \in dy \,\vert X_0 = 0]$	- Transition densities.

Unless otherwise stated, all processes X^i will also be such that $X_0^i = x_0$.

We shall start by formally defining almost sure ordering, although we shall be more concerned with weak orderings later in the thesis.

Definition 1 (Almost Sure) *Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space on which the two processes $X_t^i : \Omega \to E$, for $i = 1, 2$, where E is totally ordered, exist. Then,*

$$X_t^2 \overset{\text{a.s}}{\geqslant} X_t^1 \Leftrightarrow \mathsf{P}[X_t^2(\omega) \geqslant X_t^1(\omega) \text{ for all } t] = 1.$$

The important feature about this ordering, is that it is specific to the triple $(\Omega, \mathcal{F}, \mathsf{P})$. As a consequence of this, its usage may become limited in practical applications.

We move on to the definitions of the weak comparisons. We start with the weakest of the two, with which we shall be particularly concerned – stochastic ordering.

Definition 2 (Stochastic) *This is defined using the distribution functions of the two processes being compared.*

$$\left[X_t^2\right] \overset{\text{st}}{\geqslant} \left[X_t^1\right] \Leftrightarrow P_t^2(x) \leqslant P_t^1(x) \qquad \text{for all } t, x.$$

This type of comparison is useful in many situations, but it does have its limits. When $[X_t^2] \overset{\text{st}}{\geqslant} [X_t^1]$, it does not follow that $\left[X_t^2 \,|A\right] \overset{\text{st}}{\geqslant} \left[X_t^1 \,|A\right]$, where A is a conditioning event, such as $\tau > t$, and τ is an appropriate stopping time (see Roberts (1991a) for a counterexample). As conditioning, particularly with respect to a boundary hitting time, is of great interest in the work we are doing, this must be carefully noted.

A natural extension to make, which is preserved under conditioning (Roberts (1991a)), is the following.

Definition 3 (Strong Stochastic) *The simplest definition of strong stochastic ordering is*

$$\left[X_t^2\right] \overset{\text{sst(1)}}{\geqslant} \left[X_t^1\right] \Leftrightarrow \frac{p_t^2(x_0, y)}{p_t^1(x_0, y)} \text{ is non-decreasing in } y, \text{ for some } x_0, \text{ for all } t \geqslant 0.$$

We shall define all other forms of strong stochastic ordering in terms of likelihood ratios. However, we note that an alternative form is available, which may be more useful in showing that processes are not strongly stochastically ordered. Denoting by P_0^i the probability measure for process X_t^i given $X_0^i = x_0$, we write

$$\left[X_t^2\right] \overset{\text{sst(1)}}{\geqslant} \left[X_t^1\right] \Leftrightarrow \mathsf{P}_0^2[X_t^2 \in A_2]\mathsf{P}_0^1[X_t^1 \in A_1] \geqslant \mathsf{P}_0^2[X_t^2 \in A_1]\mathsf{P}_0^1[X_t^1 \in A_2]$$

for some x_0, all $t \geqslant 0$ and all A_1, A_2 such that $a_1 \in A_1$, $a_2 \in A_2$ implies $a_1 \leqslant a_2$.

The following Lemma shows the equivalence of these two definitions.

Lemma 4 *Let processes X^i have transition densities $p_t^i(x,y)$ (with respect to the Lebesgue measure), when $X_0^i = x$ $(i = 1, 2)$. Then,*

$$\frac{p_t^2(x,y)}{p_t^1(x,y)} \text{ is non-decreasing in } y \Leftrightarrow \tag{3.1}$$

$$\mathsf{P}[X_t^2 \in A_2]\mathsf{P}[X_t^1 \in A_1] \geqslant \mathsf{P}[X_t^2 \in A_1]\mathsf{P}[X_t^1 \in A_2] \tag{3.2}$$

for all A_1, A_2 such that $a_1 \in A_1$ and $a_2 \in A_2$ implies $a_1 \leqslant a_2$.

Proof

Let I_{A_j} denote the indicator function of A_j, for $j = 1, 2$. Then,

$$\mathsf{P}[X_t^i \in A_j] = \int I_{A_j}(y) p_t^i(x, y) dy.$$

Expressing (3.2) in this form we obtain

$$\int\int [I_{A_2}(y) I_{A_1}(z) p_t^2(x, y) p_t^1(x, z) - I_{A_1}(y) I_{A_2}(z) p_t^2(x, y) p_t^1(x, z)] dy dz \geqslant 0,$$

which may also be written as

$$\int\int_{y>z} [I_{A_2}(y) I_{A_1}(z) - I_{A_1}(y) I_{A_2}(z)][p_t^2(x, y) p_t^1(x, z) - p_t^2(x, z) p_t^1(x, y)] dy dz \geqslant 0.$$

However, under the definition of the A_i's being considered, $I_{A_1}(y) I_{A_2}(z) = 0$ for $y > z$. Hence, (3.2) reduces to

$$\int\int_{y>z} I_{A_2}(y) I_{A_1}(z) [p_t^2(x, y) p_t^1(x, z) - p_t^2(x, z) p_t^1(x, y)] dy dz \geqslant 0. \tag{3.3}$$

If (3.1) holds, the square bracketed term in (3.3) is non-negative, and hence the integrand is also non-negative. So (3.2) follows.

For the reverse implication, note that we may select $A_i = (v_i - \delta, v_i + \delta)$, $v_1 < v_2$ and when $\delta < \frac{1}{2}(v_2 - v_1)$ the relation $a_1 \in A_1$ and $a_2 \in A_2$ implies $a_1 \leqslant a_2$ holds. Letting $\delta \downarrow 0$ in such a definition, we conclude that the double integration in (3.3) reduces to a double summation, with the only significant values when $y = v_2$ and $z = v_1$. We deduce from this that

$$p_t^2(x, v_2) p_t^1(x, v_1) - p_t^2(x, v_1) p_t^1(x, v_2) \geqslant 0,$$

which is (3.1).

★

A similar result holds in the case where the processes have discrete, countable state spaces. It is proved by an almost identical argument.

Lemma 5 *Let processes X^i $(i = 1, 2)$ have discrete, countable state spaces, and let $P_t^i(x) = \mathsf{P}[X_t^i = x \,|\, X_0^i = x_0]$, for $i = 1, 2$, in some probability space $(\Omega, \mathcal{F}, \mathsf{P})$. Then,*

$$\frac{P_t^2(x)}{P_t^1(x)} \text{ is non-decreasing in } x \Leftrightarrow$$
$$\mathsf{P}[X_t^2 \in A_2] \mathsf{P}[X_t^1 \in A_1] \;\geqslant\; \mathsf{P}[X_t^2 \in A_1] \mathsf{P}[X_t^1 \in A_2],$$

for all A_1, A_2 such that $a_1 \in A_1$ and $a_2 \in A_2$ implies $a_1 \leqslant a_2$.

The second ordering of this type is applicable when the density ratio is non-decreasing from all starting points $X_0^i = x$. We define $\overset{sst(2)}{\geqslant}$ by

$$\left[X_t^2\right] \overset{sst(2)}{\geqslant} \left[X_t^1\right] \Leftrightarrow \frac{p_t^2(x,y)}{p_t^1(x,y)} \text{ is non-decreasing in } y, \text{ for all } x.$$

A final extension is that from X_0 having a starting point, to X_0 having a general starting distribution. This leads to the definition,

$$\left[X_t^2\right] \overset{sst(3)}{\geqslant} \left[X_t^1\right] \Leftrightarrow \frac{\int \mu(x) p_t^2(x,y)dx}{\int \mu(x) p_t^1(x,y)dx}$$
is non-decreasing in y, for all probability measures μ.

However, it should be noted that this definition may be too strong, as it is not clear if any pair of processes satisfy such a definition. In our work so far, a pair of processes satisfying this ordering condition, has not presented itself.

The final three definitions all make the extension to an arbitrary starting time, s. These produce:

$$\left[X_t^2\right] \overset{sst(4)}{\geqslant} \left[X_t^1\right] \Leftrightarrow \frac{p_{s,t}^2(x_0,y)}{p_{s,t}^1(x_0,y)} \text{ is non-decreasing in } y, \text{ for all } t > s,$$
and some x_0.

$$\left[X_t^2\right] \overset{sst(5)}{\geqslant} \left[X_t^1\right] \Leftrightarrow \frac{p_{s,t}^2(x,y)}{p_{s,t}^1(x,y)} \text{ is non-decreasing in } y, \text{ for all } x \text{ and } t > s.$$

$$\left[X_t^2\right] \overset{sst(6)}{\geqslant} \left[X_t^1\right] \Leftrightarrow \frac{\int \mu(x) p_{s,t}^2(x,y)dx}{\int \mu(x) p_{s,t}^1(x,y)dx} \text{ is non-decreasing in } y, \text{ for all } \mu \text{ and } t > s.$$

Again $\overset{sst(6)}{\geqslant}$ may be too strong to be useful.

The relations between the six orderings are fairly obvious. Clearly any result applicable from a general starting time s is stronger than the corresponding result from $s = 0$. Furthermore, taking μ to be a point mass means the third definition is stronger than the second, and the second is an obvious generalisation of the first. In brief we have:

$$\begin{array}{ccccc} \left(\overset{sst(6)}{\geqslant}\right) & \Rightarrow & \overset{sst(5)}{\geqslant} & \Rightarrow & \overset{sst(4)}{\geqslant} \\ \Downarrow & & \Downarrow & & \Downarrow \\ \left(\overset{sst(3)}{\geqslant}\right) & \Rightarrow & \overset{sst(2)}{\geqslant} & \Rightarrow & \overset{sst(1)}{\geqslant} \end{array}$$

It should be noted, however, that these implications are all strictly one way. Before illustrating this with counterexamples, note the following example.

Example

Let X^2 be Brownian motion with drift μ_2 and X^1 be Brownian motion with drift μ_1, with $X^1_s = X^2_s = x$. Then the transition density ratio is given by

$$\frac{p^2_{s,t}(x,y)}{p^1_{s,t}(x,y)} = \exp\left[\frac{1}{2}[2y - 2x - (t-s)(\mu_1 + \mu_2)](\mu_2 - \mu_1)\right],$$

and consequently, $[X^2_t] \overset{\text{sst}(5)}{\geqslant} [X^1_t] \Leftrightarrow \mu_2 \geqslant \mu_1$.

We now present the counterexamples.

Counterexample 1: $\overset{\text{sst}(1)}{\geqslant} \not\Rightarrow \overset{\text{sst}(4)}{\geqslant}$.

Take X^1 to be standard Brownian motion, and let X^2 satisfy

$$dX^2_t = dB_t + b(t, X^2_t)dt,$$

where

$$b(t,x) = \begin{cases} 1 & 0 \leqslant t < 1 \\ -1 & 1 \leqslant t < 2 \\ 1 & 2 \leqslant t \end{cases} .$$

Note first that, if both processes are started from $X^i_0 = 0$ $(i = 1,2)$, then X^2 has a normal distribution with non-negative mean. Thus, $[X^2_t] \overset{\text{sst}(1)}{\geqslant} [X^1_t]$. However, if the processes are started from $X^i_1 = 0$ $(i = 1,2)$, we note $X^2_{1.5}$ has a normal $N(-0.5, 0.5)$ distribution, and hence $[X^2_t] \overset{\text{sst}(4)}{\not\geqslant} [X^1_t]$.

This counterexample also shows that $\overset{\text{sst}(2)}{\geqslant} \not\Rightarrow \overset{\text{sst}(5)}{\geqslant}$.

Counterexample 2: $\overset{\text{sst}(1)}{\geqslant} \not\Rightarrow \overset{\text{sst}(2)}{\geqslant}$.

Again take X^1 to be standard Brownian motion, and let X^2 be the modulus of a standard Brownian motion,

$$p^2_t(0,y) = \begin{cases} \frac{2}{\sqrt{2\pi t}} e^{-y^2/2t} & y \geqslant 0 \\ 0 & \text{o.w} \end{cases} .$$

Taking $x_0 = 0$ in the definition of $\overset{\text{sst}(1)}{\geqslant}$, we have $[X^2_t] \overset{\text{sst}(1)}{\geqslant} [X^1_t]$.

Now suppose $X^i_0 = 1$ $(i = 1,2)$, then

$$\begin{aligned} \frac{p^2_t(1,y)}{p^1_t(1,y)} &= \frac{e^{-(y-1)^2/2t} + e^{-(y+1)^2/2t}}{e^{-(y-1)^2/2t}} \\ &= 1 + e^{-2y/t}. \end{aligned}$$

Clearly this is non-increasing in y. Hence $[X_t^2] \overset{\text{sst(2)}}{\not\geqslant} [X_t^1]$. This counterexample also demonstrates $\overset{\text{sst(4)}}{\geqslant} \not\Rightarrow \overset{\text{sst(5)}}{\geqslant}$.

Counterexample 3: $\overset{\text{sst(2)}}{\geqslant} \not\Rightarrow \overset{\text{sst(3)}}{\geqslant}$.

Again let X^1 be standard Brownian motion, and let X^2 be Brownian motion with unit drift. Clearly $[X_t^2] \overset{\text{sst(2)}}{\geqslant} [X_t^1]$.

Now let $X_0^i \sim \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$, where δ_a represents a point mass at a. Then

$$\mathsf{P}[X_t^1 \in dy] = \frac{1}{2\sqrt{2\pi t}}\left(e^{-y^2/2t} + e^{-(y-1)^2/2t}\right)dy$$

and

$$\mathsf{P}[X_t^2 \in dy] = \frac{1}{2\sqrt{2\pi t}}\left(e^{-(y-t)^2/2t} + e^{-(y-1-t)^2/2t}\right)dy.$$

Consider $y_1 = 0.1$ and $y_2 = 0.9$. Then, using f_t^i to denote the density,

$$\begin{aligned} f_t^1(y_1) &= \frac{1}{2\sqrt{2\pi t}}\left(e^{-0.81/2t} + e^{-0.01/2t}\right) \\ &= f_t^1(y_2). \end{aligned}$$

For a counterexample to $[X_t^2] \overset{\text{sst(3)}}{\geqslant} [X_t^1]$, we need only show that there is some time t for which $f_t^2(y_2) < f_t^2(y_1)$, with $y_1 = 0.1$ and $y_2 = 0.9$.

Consider $t = 0.1$.

$$\begin{aligned} f_{0.1}^2(y_2) &= k\left(e^{-(0.8)^2/0.2} + e^{-(0.2)^2/0.2}\right) \\ &= k(e^{-3.2} + e^{-0.2}) \\ &< k, \end{aligned}$$

whereas

$$\begin{aligned} f_{0.1}^2(y_1) &= k\left(e^{-1/0.2} + 1\right) \\ &= k(1 + e^{-5}) \\ &> k. \end{aligned}$$

Hence we may deduce $[X_t^2] \overset{\text{sst(3)}}{\not\geqslant} [X_t^1]$. The same example leads to the conclusion that $\overset{\text{sst(5)}}{\geqslant} \not\Rightarrow \overset{\text{sst(6)}}{\geqslant}$.

Note that $\overset{\text{sst(3)}}{\geqslant} \not\Rightarrow \overset{\text{sst(6)}}{\geqslant}$ is not verified, due to the problems of finding processes X^1 and X^2 such that $[X_t^2] \overset{\text{sst(3)}}{\geqslant} [X_t^1]$.

Important Convention

Finally, note that when we omit the bracketed number in the strong stochastic inequality, for example $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$, we shall assume that we are using $\overset{\text{sst}(5)}{\geqslant}$. This is because $\overset{\text{sst}(5)}{\geqslant}$ is the strongest ordering for which we have found ordered processes X^1 and X^2. In any theorems involving strong stochastic ordering, it will be this definition to which we refer, unless otherwise stated.

A concept related to stochastic ordering of processes is that of stochastic monotonicity (SM). This is a comparison of the behaviour of a process, when started at different points in the state space. Let X^i denote the process X such that $X_0^i = x_i$, for $i = 1, 2$. Then we make the definition

$$X \text{ is stochastically monotone } \Leftrightarrow \left[X_t^2\right] \overset{\text{st}}{\geqslant} \left[X_t^1\right] \quad \text{for all } x_1 < x_2.$$

In a similar manner, strong stochastic monotonicity (SSM) can be defined for the same process as

$$X \text{ is strongly stochastically monotone (4) } \Leftrightarrow \left[X_t^2\right] \overset{\text{sst}(4)}{\geqslant} \left[X_t^1\right] \quad \text{for all } x_1 < x_2,$$

or

$$X \text{ is strongly stochastically monotone (1) } \Leftrightarrow \left[X_t^2\right] \overset{\text{sst}(1)}{\geqslant} \left[X_t^1\right] \quad \text{for all } x_1 < x_2.$$

Definitions involving $\overset{\text{sst}(2)}{\geqslant}$ and $\overset{\text{sst}(5)}{\geqslant}$ make little sense here, as the current definitions take care of all starting points. If we state that X is SSM, it will be assumed to be SSM(4), as SSM(1) follows trivially from this.

Strong stochastic monotonicity is a special case of total positivity (see for example Karlin (1968)). This is a higher order property of transition densities, $f(x, y)$ say. If a matrix M_n is defined as

$$M_n = \begin{bmatrix} f(x_1, y_1) & f(x_2, y_1) & \cdots & f(x_n, y_1) \\ f(x_1, y_2) & f(x_2, y_2) & \cdots & f(x_n, y_2) \\ \vdots & \vdots & & \vdots \\ f(x_1, y_n) & f(x_2, y_n) & \cdots & f(x_n, y_n) \end{bmatrix},$$

f is said to be TP_n if $|M_n| \geqslant 0$ for all $x_1 < x_2 < \cdots < x_n$, $y_1 < y_2 < \cdots < y_n$, where TP_n denotes total positivity of order n. Strong stochastic monotonicity is equivalent to the condition that the transition kernel for the process is TP_2. An important consequence of this, (see Karlin and McGregor (1959)), is that, for a process to be strongly stochastically monotone, it has to be continuous in the natural topology of its state space. That is, if a process with a continuous state space is also strongly stochastically monotone, it must be a diffusion. However, if the process has a discrete state space, it must be a birth-death process, since transitions must be made to neighbouring states. Hence, if only strongly stochastically monotone processes are being considered, we can immediately rule out jump Markov processes on $\mathbb{R}$, and all discontinuous processes except birth-death ones.

3.2 Basic Results

A well known result concerning almost sure ordering, for diffusion processes, is given, for example, in Ikeda and Watanabe (1981, p352).

Theorem 3 (Ikeda and Watanabe) *Suppose we are given the following:*

1. *A strictly increasing function ρ defined on $[0,\infty)$ such that $\rho(0) = 0$ and $\int_{0+}^{\infty} \rho(\xi)^{-2} d\xi = \infty$;*
2. *a real continuous function $\sigma(t,x)$ on $[0,\infty)\times\mathbb{R}$ such that $|\sigma(t,x)-\sigma(t,y)| \leqslant \rho(|x-y|)\ \ x,y \in \mathbb{R},\ t \geqslant 0$;*
3. *two real continuous functions $b_1(t,x)$ and $b_2(t,x)$ defined on $[0,\infty)\times\mathbb{R}$ such that $b_1(t,x) < b_2(t,x)\ \ t \geqslant 0,\ x \in \mathbb{R}$.*

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space containing a filtration $(\mathcal{F}_t)$. Suppose we have the following stochastic processes

1. *two real $(\mathcal{F}_t)$-adapted continuous processes $\hat{X}_t^1(\omega)$ and $\hat{X}_t^2(\omega)$;*
2. *an $(\mathcal{F}_t)$-Brownian motion $B_t(\omega)$ such that $B_0 = 0$ a.s;*
3. *two real $(\mathcal{F}_t)$-adapted well measurable processes $\beta_1(t,\omega)$ and $\beta_2(t,\omega)$.*

Assume

$$\hat{X}_t^i - \hat{X}_0^i = \int_0^t \sigma(s,\hat{X}_s^i)dB_s + \int_0^t \beta_i(s)ds,$$

and

$$\begin{aligned} \hat{X}_0^1 &\leqslant \hat{X}_0^2, \\ \beta_1(t) &\leqslant b_1(t,\hat{X}_t^1) \text{ for every } t \geqslant 0, \\ \beta_2(t) &\geqslant b_2(t,\hat{X}_t^2) \text{ for every } t \geqslant 0. \end{aligned}$$

Then, with probability one,

$$\hat{X}_t^1 \leqslant \hat{X}_t^2 \qquad \text{for all } t \geqslant 0.$$

(Note: If path-wise uniqueness of solutions holds for at least one of the stochastic differential equations

$$dX_t = \sigma(t, X_t)dB_t + b_i(t, X_t)dt \quad (i = 1, 2),$$

then $b_1(t,x) < b_2(t,x)$ may be replaced by $b_1(t,x) \leqslant b_2(t,x)$ for all $t \geqslant 0,\ x \in \mathbb{R}$.)

For a formal proof, see Ikeda and Watanabe (1981). Intuitively, when the processes are together, the larger drift of X^2 forces the processes to have the desired ordering immediately after this time. Continuity ensures that process 2 remains larger, until the next occasion when they are together, at which point, the same argument applies. Note that, for this to hold, the processes must have the same diffusion coefficient. This is because

$$X_t^2 \overset{\text{a.s}}{\geqslant} X_t^1 \Rightarrow \left[X_t^2\right] \overset{\text{st}}{\geqslant} \left[X_t^1\right],$$

and so if $\sigma_1(t,x) > \sigma_2(t,x)$, consideration of the tail probabilities, for small t, would contradict this statement.

Another ordering result, which has been proved by Roberts and Jacka (1994), relates strong stochastic ordering of Birth and Death processes to their transition rates. They defined these rates in the following way:

$$\alpha_j(t) = \left.\frac{d\mathbb{P}[X_{t+s} = j+1 \,|X_t = j]}{d\mu(t+s)}\right|_{s=0},$$

$$\beta_j(t) = \left.\frac{d\mathbb{P}[X_{t+s} = j-1 \,|X_t = j]}{d\mu(t+s)}\right|_{s=0},$$

where μ is a σ-finite measure, absolutely continuous with respect to the Lebesgue measure. Then the following may be proved.

Theorem 4 (Roberts and Jacka (1994)) *Let X^1 and X^2 be Birth and Death processes, such that all transition rates $\alpha_j^i(t)$ and $\beta_j^i(t)$ are bounded, for $i = 1, 2$. Assume also that the X^i are non-explosive. Then*

$$\left[X_t^2\right] \overset{\text{sst}}{\geqslant} \left[X_t^1\right] \Rightarrow \alpha_j^2(t) \geqslant \alpha_j^1(t),\ \beta_j^2(t) \leqslant \beta_j^1(t),$$

for all $j \in \mathbb{Z}^+$ and $t \in [0, R]$, for an arbitrary constant $R>0$.

The theorem formalises the intuitive result that the larger process has greater "up" rates, and smaller "down" rates, from each state and at all times.

Many ordering results link conditioned and unconditioned versions of the same process, where the conditioning is often related to a hitting time. An example of such a process is $X_. \left| \tau^f > T \right.$, where $\tau^f = \inf_{t>0}\{t : X_t \geqslant f(t)\}$ for some arbitrary functional boundary f. In many instances, there exists a quasi-stationary distribution, δ_∞, such that $\delta_\infty = \lim_{t\to\infty}[X_t \left| \tau^f > t \right.]$, (see Jacka and Roberts (1987)). One drawback, however, is that δ_∞ is often not analytically tractable, and is expressible only as the solution of an integral equation. Properties of this distribution may be investigated, and comparisons made with the distribution of the process. One such result is the following theorem.

Theorem 5 (Roberts (1991b)) *Assume X to be a diffusion process such that the distribution $\delta_\infty = \lim_{t\to\infty}[X_t \left| \tau > t \right.]$ exists, and denote $\nu_t = [X_t \left| \tau > t \right.]$. If $\nu_0 \overset{st}{\geqslant} \delta_\infty$, then $\nu_t \overset{st}{\geqslant} \delta_\infty$ for all $t > 0$. Similarly, $\nu_0 \overset{st}{\leqslant} \delta_\infty \Rightarrow \nu_t \overset{st}{\leqslant} \delta_\infty$ for all $t > 0$.*

A widely investigated topic, is the comparison between the conditioned process $X_. \left| \tau^f > T \right.$, and the unconditioned process X. The earliest result of this type is a stochastic ordering, when the boundary f is a constant.

Theorem 6 (Pollak and Siegmund (1986)) *If X is a time homogeneous process, and $f(t) = c$, a constant, then $[X_t] \overset{st}{\geqslant} \left[X_t \left| \tau^f > t \right.\right]$.*

This result has been strengthened for stochastically monotone processes and general boundaries f to

Theorem 7 (Roberts (1991a)) *If X is stochastically monotone, and $f(t)$ is such that $f(0) > X_0$, then $[X_t] \overset{st}{\geqslant} \left[X_t \left| \tau^f > t \right.\right]$.*

A further result produces a strong stochastic ordering, when the initial process X is strongly stochastically monotone, and two boundaries are used for the conditioning.

Theorem 8 (Roberts (1991a)) *If X is strongly stochastically monotone, and the cadlag boundaries f and g are such that $f(t) \leqslant g(t)$ for all t, then*

$$\left[X_t \left| \tau^g > t \right.\right] \overset{sst}{\geqslant} \left[X_t \left| \tau^f > t \right.\right].$$

These results are intuitively fairly clear. The effect of conditioning not to hit a boundary "pushes the diffusion down". Thus, the conditioned process will be "pushed down", and so the unconditioned process would be stochastically larger. The closer the process is to the boundary, the greater this effect. Consequently, the results of Roberts (1991a) also have a sound intuitive base.

Note that Theorem 8 can be applied to obtain the conclusion that the Bessel process is strongly stochastically greater than Brownian motion. For if we take $f(t) = -\infty$, $g(t) = 0$ and our process X to be standard Brownian motion (which is strongly stochastically monotone), with start point $X_s = x > 0$, we have

$$\left[X_t \,\middle|\, \tau^g > t\right] \overset{\text{sst}}{\geqslant} \left[X_t \,\middle|\, \tau^f > t\right].$$

But, $\left[X_t \,\middle|\, \tau^f > t\right] = [X_t]$, and $\left[X_t \,\middle|\, \tau^g > t\right] \sim$ Bessel process, see for example Rogers and Williams (1987). Thus we deduce the stated ordering. Knowing the Bessel process is strongly stochastically greater than Brownian motion, is useful in the Chapter 6, when strong stochastic ordering is sought by looking at the forms of the stochastic differential equations. We now have an example which any theorems cannot contradict.

Chapter 4

Preliminary Results in Ordering

4.1 Introduction

In this chapter, we shall present a selection of ordering results, which are comparatively simple. We shall give these results, together with quick proofs, and they will then be employed, without further comment, in future chapters.

4.2 Ordering Results

We start by proving two results, which although not involving ordering of processes, are useful in checking the existence of strong stochastic ordering.

Result 1 *Let $X \sim N(a, v_1)$ and $Y \sim N(b, v_2)$, where a, b, v_1 and v_2 are finite. Then*

$$\frac{\mathbb{P}[X \in dx]}{\mathbb{P}[Y \in dx]} \quad \textit{monotonic in } x \Rightarrow v_1 = v_2.$$

Proof

Let $R(x) = \frac{\mathbb{P}[X \in dx]}{\mathbb{P}[Y \in dx]}$. Then substituting in the normal densities, we have

$$R(x) = \sqrt{\frac{v_2}{v_1}} \exp\left\{\frac{1}{2v_2}(x-b)^2 - \frac{1}{2v_1}(x-a)^2\right\}.$$

Differentiating this,

$$\begin{aligned} R'(x) &= \left[\frac{x-b}{v_2} - \frac{x-a}{v_1}\right] R(x) \\ &= \frac{(v_1 - v_2)x + av_2 - bv_1}{v_1 v_2} R(x). \end{aligned}$$

Now suppose that $v_2 > v_1$. Then $R'(x) < 0$ as $x \to \infty$, and $R'(x) > 0$ as $x \to -\infty$, contradicting the monotonicity of R. The inequalities are reversed for $v_1 > v_2$, and therefore a similar contradiction is reached. Hence we must have $v_1 = v_2$.

★

Result 2 *Let $X \sim N(a,v)$ and $Y \sim N(b,v)$. Then the ratio*

$$\frac{\mathsf{P}[X \in dx]}{\mathsf{P}[Y \in dx]} \quad \text{is non-decreasing in } x \Leftrightarrow a \geqslant b.$$

Proof

Denote the density ratio by $R(x)$. We then have

$$\begin{aligned} R(x) &= \frac{\frac{1}{\sqrt{2\pi v}} \exp\left\{-\frac{1}{2v}(x-a)^2\right\}}{\frac{1}{\sqrt{2\pi v}} \exp\left\{-\frac{1}{2v}(x-b)^2\right\}} \\ &= \exp\left\{\frac{1}{2v}\left((x-b)^2 - (x-a)^2\right)\right\} \\ &= \exp\left\{\frac{1}{2v}(2x-a-b)(a-b)\right\} . \end{aligned}$$

Thus,

$$\frac{dR}{dx} = \frac{(a-b)}{v} \exp\left\{\frac{1}{2v}(2x-a-b)(a-b)\right\},$$

so that R is non-decreasing in x if and only if $a \geqslant b$ as claimed.

★

We now start to consider the ordering of processes. We begin with an observation about diffusion processes.

Result 3 (Roberts (1991a)) *Let X be a diffusion process satisfying*

$$dX_t = \sigma(t, X_t)dB_t + \mu(t, X_t)dt. \tag{4.1}$$

Then X is strongly stochastically monotone, (see Chapter 3).

Proof

Let X^i be the process satisfying the stochastic differential equation (4.1), with the initial condition that $X_0^i = x_i$, for $i = 1, 2$. We assume that $x_1 < x_2$. Consider a coupling arrangement, so that both processes are run on the same filtration, and insist that $\mathsf{P}[X_t^1 = X_t^2] = 1$ for all $t > \tau = \inf_{t>0}\{t : X_t^1 = X_t^2\}$. We are required to show that

$$\frac{p_t^2(x_2, y_2)}{p_t^2(x_2, y_1)} \geqslant \frac{p_t^1(x_1, y_2)}{p_t^1(x_1, y_1)},$$

for all $y_1 < y_2$. We introduce the notation $\mathsf{P}[X_t^i \in dy] = p_t^i(x_i, y)dy$, and adding one to both sides of the above, it is equivalent to show

$$\frac{\mathsf{P}[X_t^2 \in dy_2 \cup dy_1]}{\mathsf{P}[X_t^2 \in dy_1]} \geqslant \frac{\mathsf{P}[X_t^1 \in dy_2 \cup dy_1]}{\mathsf{P}[X_t^1 \in dy_1]},$$

or equivalently,

$$\frac{\mathsf{P}[X_t^1 \in dy_1]}{\mathsf{P}[X_t^1 \in dy_2 \cup dy_1]} \geqslant \frac{\mathsf{P}[X_t^2 \in dy_1]}{\mathsf{P}[X_t^2 \in dy_2 \cup dy_1]}. \tag{4.2}$$

Note that both sides of (4.2) represent the probabilities of the diffusions, conditioned to be at either y_1 or y_2, actually being at y_1. Thus, (4.2) is clearly true from the coupling arrangement producing almost sure ordering.

★

We shall now produce some results which follow if $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$. In the first of these, we will formally prove that strong stochastic ordering is stronger than stochastic ordering.

Result 4 *Suppose* $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$. *Then* $[X_t^2] \overset{\text{st}}{\geqslant} [X_t^1]$.

Proof

We shall assume that the processes are on the real line. If they actually have discrete state spaces, the same argument applies *mutatis mutandis*. Let $f_t^i(x)$ denote the density of X_t^i. Suppose there exists an x_0 such that

$$\int_{-\infty}^{x_0} f_t^2(x)dx > \int_{-\infty}^{x_0} f_t^1(x)dx.$$

Then, because we have probability densities, we also have

$$\int_{x_0}^{\infty} f_t^1(x)dx > \int_{x_0}^{\infty} f_t^2(x)dx.$$

Let $A_1 = (-\infty, x_0)$ and $A_2 = [x_0, \infty)$. Under these definitions, we certainly have $a_1 \in A_1$ and $a_2 \in A_2$ implies $a_1 \leqslant a_2$. But from above,

$$\mathsf{P}[X_t^2 \in A_1] > \mathsf{P}[X_t^1 \in A_1],$$

and

$$\mathsf{P}[X_t^1 \in A_2] > \mathsf{P}[X_t^2 \in A_2].$$

Combining these $\mathsf{P}[X_t^2 \in A_1]\mathsf{P}[X_t^1 \in A_2] > \mathsf{P}[X_t^1 \in A_1]\mathsf{P}[X_t^2 \in A_2]$, which contradicts the strong stochastic ordering hypothesis. Thus, we must have

$$\int_{-\infty}^{x_0} f_t^2(x)dx \leqslant \int_{-\infty}^{x_0} f_t^1(x)dx,$$

for all x_0. That is our claimed result.

★

Result 5 *Let* $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$, *and* a *and* b *be constants. Then (i) if* $a > 0$, *we have* $[aX_t^2 + b] \overset{\text{sst}}{\geqslant} [aX_t^1 + b]$, *and (ii) if* $a < 0$, *we have* $[aX_t^1 + b] \overset{\text{sst}}{\geqslant} [aX_t^2 + b]$.

Proof

We shall only prove the case $a > 0$. The other case follows by an similar method. By definition,

$$\mathsf{P}[X_t^2 \in A_2]\mathsf{P}[X_t^1 \in A_1] \geqslant \mathsf{P}[X_t^1 \in A_2]\mathsf{P}[X_t^2 \in A_1]$$

for all sets A_1 and A_2 such that $a_1 \in A_1$ and $a_2 \in A_2$ implies $a_1 \leqslant a_2$. For such a pair of sets, define

$$B_i = \left\{ x : \frac{x - b}{a} \in A_i \right\}.$$

Since a is positive, $b_1 \in B_1$ and $b_2 \in B_2$ implies $b_1 \leqslant b_2$, and consequently

$$\mathsf{P}[X_t^2 \in B_2]\mathsf{P}[X_t^1 \in B_1] \geqslant \mathsf{P}[X_t^1 \in B_2]\mathsf{P}[X_t^2 \in B_1].$$

Denoting $Y_t^i = aX_t^i + b$, we clearly have $\mathsf{P}[X_t^i \in B_j] = \mathsf{P}[Y_t^i \in A_j]$, for $i, j = 1, 2$.
Hence the result follows.

★

Result 6 *Suppose* $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$ *and* $[X_t^3] \overset{\text{sst}}{\geqslant} [X_t^2]$. *Then* $[X_t^3] \overset{\text{sst}}{\geqslant} [X_t^1]$.

Proof

From the definition of strong stochastic ordering, the first strong stochastic inequality yields

$$\frac{\mathsf{P}[X_t^2 \in A_2]}{\mathsf{P}[X_t^2 \in A_1]} \geqslant \frac{\mathsf{P}[X_t^1 \in A_2]}{\mathsf{P}[X_t^1 \in A_1]},$$

and the second gives

$$\frac{\mathsf{P}[X_t^3 \in A_2]}{\mathsf{P}[X_t^3 \in A_1]} \geqslant \frac{\mathsf{P}[X_t^2 \in A_2]}{\mathsf{P}[X_t^2 \in A_1]},$$

where the sets A_i are defined in the standard way. Combining these inequalities produces the claimed result.

★

We conclude the collection of results about general processes with some limiting results.

Result 7 *Let* $X^{1,n}$ *and* $X^{2,n}$ *be two sequences of processes, such that for each* n

$$\left[X_t^{2,n}\right] \overset{\text{sst}}{\geqslant} \left[X_t^{1,n}\right]$$

and $[X^{i,n}] \Rightarrow [X^i]$ *as* $n \to \infty$ *for* $i = 1, 2$, *where* $[X^i]$ *has a density with respect to some* σ*−finite measure. Then,*

$$\left[X_t^2\right] \overset{\text{sst}}{\geqslant} \left[X_t^1\right].$$

Proof

From the definition of strong stochastic ordering, we have

$$\mathsf{P}[X_t^{2,n} \in A_2]\mathsf{P}[X_t^{1,n} \in A_1] \geqslant \mathsf{P}[X_t^{2,n} \in A_1]\mathsf{P}[X_t^{1,n} \in A_2],$$

where $a_1 \in A_1$ and $a_2 \in A_2$ implies $a_1 \leqslant a_2$. Using the definition of weak convergence, noting P converges on all sets, from the existence of a density of $[X^i]$, we see

$$\mathsf{P}[X_t^{i,n} \in A_j] \to \mathsf{P}[X_t^i \in A_j] \qquad \text{as } n \to \infty,$$

for $i, j = 1, 2$. Taking limits, in our above inequality, produces the desired result.

★

Result 8 *Let X^1 and X^2 be two processes such that*

$$\left[X_t^2\right] \overset{\text{sst}}{\geqslant} \left[X_t^1\right]$$

and $[X_t^i] \Rightarrow [Y^i]$ as $t \downarrow s$, where $[Y^i]$ has a density with respect to some $\sigma-$finite measure. Then,

$$\left[Y^2\right] \overset{\text{sst}}{\geqslant} \left[Y^1\right].$$

Proof

From the definition of strong stochastic ordering, we have

$$\mathsf{P}[X_t^2 \in A_2]\mathsf{P}[X_t^1 \in A_1] \geqslant \mathsf{P}[X_t^2 \in A_1]\mathsf{P}[X_t^1 \in A_2],$$

where $a_1 \in A_1$ and $a_2 \in A_2$ implies $a_1 \leqslant a_2$. Using the definition of weak convergence, we see

$$\mathsf{P}[X_t^i \in A_j] \to \mathsf{P}[Y^i \in A_j] \qquad \text{as } t \downarrow s.$$

for $i, j = 1, 2$. Using this identity in the above inequality with $t \downarrow s$, produces the claimed result.

★

We now consider a result for processes with deterministic drifts.

Result 9 *Suppose X^i satisfies the stochastic differential equation*

$$dX_t^i = dB_t + \mu_i(t)dt,$$

for $i = 1, 2$. If $\mu_2(t) \geqslant \mu_1(t)$ for all t, then $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$.

Proof

Suppose the processes start from $X_s^1 = X_s^2 = x$. Then, as the drift coefficients are deterministic,

$$\left[X_t^i\right] \sim N\left(x + \int_s^t \mu_i(v)dv, t - s\right) \qquad \text{for } t \geqslant s.$$

Applying Result 2, we complete the proof.

★

Note that the converse to this result is also true, see Theorem 10 in Chapter 6.

We now produce a result about the ordering of process bridges. We define the process bridge $\bar{X}$ by

$$\bar{X}_t \sim (X_t \,|X_T = y).$$

It is known that if X satisfies the stochastic differential equation

$$dX_t = dB_t + \mu(t, X_t)dt,$$

then $\bar{X}$ satisfies

$$d\bar{X}_t = dB_t + \left[\mu(t, \bar{X}_t) + \frac{p_x(t, \bar{X}_t)}{p(t, \bar{X}_t)}\right] dt,$$

where $p(t, x)dy = \mathsf{P}[X_T \in dy \,|X_t = x]$. See, for example, Rogers and Williams (1987, IV.39-IV.40) for more details of this result.

Consider X to be Brownian motion with constant drift μ. Then the function p, defined above, can be written

$$p(t, x) = \frac{1}{\sqrt{2\pi(T-t)}} \exp\left\{-\frac{1}{2(T-t)}(y - x - (T-t)\mu)^2\right\},$$

so that

$$\frac{p_x(t, x)}{p(t, x)} = \frac{y - x - (T-t)\mu}{T-t}.$$

Therefore, the corresponding bridge process $\bar{X}$ satisfies

$$\begin{aligned} d\bar{X}_t &= dB_t + \left[\mu + \frac{y - \bar{X}_t - (T-t)\mu}{T-t}\right] dt \\ &= dB_t + \frac{y - \bar{X}_t}{T-t} dt. \end{aligned}$$

Hence, the bridge process is independent of the (constant) drift of the unbridged process. Consequently, if we seek to find conditions on the drifts of the processes, to ensure that the corresponding bridges are strongly stochastically ordered, we will need stronger conditions than the drifts being ordered.

In the case where μ is a function of t only, we can establish the following result.

Result 10 *Suppose X^1 and X^2 satisfy*

$$dX_t^2 = dB_t + \mu_2(t)dt,$$

$$dX_t^1 = dB_t + \mu_1(t)dt,$$

for $t \geqslant s$ and $X_s^i = x_0$, $(i = 1, 2)$, where $\mu_2 - \mu_1$ is a monotonic function of t. Then the process bridges $\bar{X}^1$ and $\bar{X}^2$ are strongly stochastically ordered. Specifically, if $\mu_2 - \mu_1$ is non-decreasing, then $\left[\bar{X}_t^1\right] \overset{\text{sst}}{\geqslant} \left[\bar{X}_t^2\right]$, for $t \geqslant s$, with the reverse ordering if $\mu_2 - \mu_1$ is non-increasing.

Proof

We are required to show $\mathsf{P}[\bar{X}_t^2 \in dx]/\mathsf{P}[\bar{X}_t^1 \in dx]$ is monotonic in x. However, by direct computation, we can establish

$$\begin{aligned}\frac{\mathsf{P}[\bar{X}_t^2 \in dx]}{\mathsf{P}[\bar{X}_t^1 \in dx]} &= \frac{\mathsf{P}[X_t^2 \in dx \,|X_T^2 = y]}{\mathsf{P}[X_t^1 \in dx \,|X_T^1 = y]} \\ &= \frac{\mathsf{P}[X_T^2 \in dy \,|X_t^2 = x]\mathsf{P}[X_t^2 \in dx]\mathsf{P}[X_T^1 \in dy]}{\mathsf{P}[X_T^1 \in dy \,|X_t^1 = x]\mathsf{P}[X_t^1 \in dx]\mathsf{P}[X_T^2 \in dy]}.\end{aligned}$$

Computing these probabilities, neglecting the terms not involving x, we must show the monotonicity in x of

$$\frac{\exp\left\{-\frac{1}{2(T-t)}\left(y - x - \int_t^T \mu_2(r)dr\right)^2\right\}\exp\left\{-\frac{1}{2(t-s)}\left(x - x_0 - \int_s^t \mu_2(r)dr\right)^2\right\}}{\exp\left\{-\frac{1}{2(T-t)}\left(y - x - \int_t^T \mu_1(r)dr\right)^2\right\}\exp\left\{-\frac{1}{2(t-s)}\left(x - x_0 - \int_s^t \mu_1(r)dr\right)^2\right\}}.$$

After some rearrangement, this is equivalent to the monotonicity of

$$\frac{x}{t-s}\int_s^t (\mu_2(r) - \mu_1(r))dr - \frac{x}{T-t}\int_t^T (\mu_2(r) - \mu_1(r))dr.$$

This is non-decreasing in x if $\mu_2 - \mu_1$ is non-increasing, and non-increasing in x if $\mu_2 - \mu_1$ is non-decreasing, as claimed.

★

Chapter 5

Orderings involving Process Moduli

5.1 Introduction

In this chapter, we shall examine the strong stochastic orderings between two processes, when at least one process is a modulus process. For processes with deterministic drifts, we can calculate the densities explicitly, and draw conclusions from these.

5.2 Results

Note that, in general,

$$\mathbb{P}[|X_t| \in dx] = \mathbb{P}[X_t \in dx] + \mathbb{P}[X_t \in d(-x)], \tag{5.1}$$

and so, when investigating whether $[|X_t|] \overset{\text{sst}}{\geqslant} [X_t]$, we need only consider

$$\frac{\mathbb{P}[X_t \in d(-x)]}{\mathbb{P}[X_t \in dx]}.$$

This follows from the following lemma.

Lemma 6

$$[|X_t|] \overset{\text{sst}}{\geqslant} [X_t] \Leftrightarrow \frac{\mathbb{P}[X_t \in d(-x)]}{\mathbb{P}[X_t \in dx]} \text{ is non-decreasing in } x,$$

for all $t \geqslant s$ and all choices of starting point $X_s = x_0$.

Proof

By the definition of strong stochastic ordering, we require

$$\frac{\mathsf{P}[|X_t| \in dx]}{\mathsf{P}[X_t \in dx]}$$

to be non-decreasing in x for all choices of starting point, and all $t \geqslant s$. Substituting (5.1) into the numerator, this reduces to

$$1 + \frac{\mathsf{P}[X_t \in d(-x)]}{\mathsf{P}[X_t \in dx]}$$

non-decreasing in x, which reduces to the claimed result.

★

Consider the process, X, satisfying the stochastic differential equation

$$dX_t = dB_t + \mu(t)dt, \tag{5.2}$$

such that $X_s = x_0$, and μ is integrable. Then, for $t \geqslant s$, we know

$$[X_t] \sim N\left(x_0 + \int_s^t \mu(r)dr, t-s\right).$$

In the sequel, we shall denote $D = \int_s^t \mu(r)dr$. Then we have,

$$\begin{aligned} \frac{\mathsf{P}[X_t \in d(-x)]}{\mathsf{P}[X_t \in dx]} &= \frac{\exp\left\{-\frac{1}{2(t-s)}(-x-x_0-D)^2\right\}}{\exp\left\{-\frac{1}{2(t-s)}(x-x_0-D)^2\right\}} \\ &= \exp\left\{\frac{1}{2(t-s)}[2x][-2(x_0+D)]\right\} \\ &= \exp\left\{-\frac{2(x_0+D)x}{t-s}\right\}. \end{aligned} \tag{5.3}$$

We shall use this identity in the following two results.

Result 11 *Let X satisfy* (5.2). *Then,*

$$[|X_t|] \overset{\text{sst}(5)}{\not\geqslant} [X_t].$$

Proof

From (5.3), we note that

$$\frac{\mathsf{P}[X_t \in d(-x)]}{\mathsf{P}[X_t \in dx]} = \exp\left\{-\frac{2(x_0+D)x}{t-s}\right\}.$$

Clearly, we may select x_0 large enough, so that this expression is not non-decreasing in x for all $t \geqslant s$. Thus, from Lemma 6, we have our claimed result.

★

However, it is certainly possible to obtain $\overset{\text{sst}(4)}{\geqslant}$ ordering between the modulus and original process.

Result 12 *Let X satisfy* (5.2). *If $\mu(t) \leqslant 0$ for all t,*

$$[|X_t|] \overset{\text{sst}(4)}{\geqslant} [X_t].$$

Proof

When considering $\overset{\text{sst}(4)}{\geqslant}$ ordering, we shall select $x_0 = 0$. In this case, we see from (5.3) that

$$\frac{\mathsf{P}[X_t \in d(-x)]}{\mathsf{P}[X_t \in dx]} = \exp\left\{-\frac{2Dx}{t-s}\right\}.$$

Clearly, this is non-decreasing in x if $D \leqslant 0$, which follows under the imposed condition on μ.

★

The final result of this chapter compares the modulus of a process with the modulus of Brownian motion.

Result 13 *Suppose X^1 and X^2 satisfy*

$$\begin{aligned} dX_t^2 &= dB_t + \mu(t)dt \\ dX_t^1 &= dB_t. \end{aligned}$$

Then, for any integrable function μ,

$$\left[|X_t^2|\right] \overset{\text{sst}(4)}{\geqslant} \left[|X_t^1|\right].$$

Proof

When considering $\overset{\text{sst}(4)}{\geqslant}$ ordering, we will take $x_0 = 0$. Let $R(x)$ denote the density ratio,

$$R(x) = \frac{\mathsf{P}[|X_t^2| \in dx]}{\mathsf{P}[|X_t^1| \in dx]}$$

for $x \geqslant 0$, where both processes start from $X_s^i = 0$. We let $D = \int_s^t \mu(r)dr$. Then,

$$\begin{aligned} R(x) &= \frac{\exp\left\{-\frac{1}{2(t-s)}(x-D)^2\right\} + \exp\left\{-\frac{1}{2(t-s)}(x+D)^2\right\}}{2\exp\left\{-\frac{1}{2(t-s)}x^2\right\}} \\ &= \frac{1}{2}\left[\exp\left\{\frac{1}{2(t-s)}[x^2-(x-D)^2]\right\} + \exp\left\{\frac{1}{2(t-s)}[x^2-(x+D)^2]\right\}\right] \\ &= \frac{1}{2}\exp\left\{-\frac{D^2}{2(t-s)}\right\}\left[\exp\left\{\frac{Dx}{t-s}\right\} + \exp\left\{-\frac{Dx}{t-s}\right\}\right]. \end{aligned}$$

Differentiating, we see

$$\frac{dR}{dx} = \frac{1}{2}\exp\left\{-\frac{D^2}{2(t-s)}\right\}\frac{D}{t-s}\left[\exp\left\{\frac{Dx}{t-s}\right\} - \exp\left\{-\frac{Dx}{t-s}\right\}\right].$$

Remembering that $x \geqslant 0$, we note that $\frac{dR}{dx} \geqslant 0$ for all values of D. Therefore, we deduce that

$$\left[|X_t^2|\right] \overset{\text{sst}(4)}{\geqslant} \left[|X_t^1|\right].$$

★

This result can also be used to show $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$ does not imply $[|X_t^2|] \overset{\text{sst}}{\geqslant} [|X_t^1|]$. For if we use Result 13 with $\mu(t) = -1$, we have $[X_t^1] \overset{\text{sst}}{\geqslant} [X_t^2]$, but $[|X_t^1|] \overset{\text{sst}}{\ngeqslant} [|X_t^2|]$.

Chapter 6

Drift Ordering

6.1 Introduction

Suppose we have two diffusion processes X^1 and X^2 satisfying the following stochastic differential equations:

$$\begin{aligned} dX_t^1 &= \sigma_1(t, X_t^1)dB_t + \mu_1(t, X_t^1)dt \\ dX_t^2 &= \sigma_2(t, X_t^2)dB_t + \mu_2(t, X_t^2)dt \end{aligned}$$

with $X_0^1 = X_0^2 = x$. Throughout this chapter, we will assume, unless otherwise stated, that all drift coefficients, μ, and diffusion coefficients, σ, satisfy:

$$|\mu(t,x)|^2 + |\sigma(t,x)|^2 \leqslant K[1+|x|^2] \tag{6.1}$$
$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leqslant K|x-y| \text{ for some } K < \infty. \tag{6.2}$$

We often need these processes to be strongly stochastically ordered, but checking the definitions of strong stochastic ordering (see Chapter 3) is not always convenient, because the distributions of the processes need not be known. Consequently, it would be advantageous if the existence of strong stochastic ordering could be determined directly from the stochastic differential equations. We seek conditions on the drift and diffusion coefficients in order that this may be done.

6.2 Necessary Conditions for Strong Stochastic Ordering

Before finding some conditions which are necessary for $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$, we shall establish a lemma concerning the weak convergence of the distribution of a diffusion process satisfying a stochastic differential equation.

Lemma 7 *Suppose X satisfies the stochastic differential equation*

$$dX_t = \sigma(t, X_t)dB_t + \mu(t, X_t)dt,$$

with $X_s = x$. If $X \in L^2$, and μ and σ are Lipschitz in t and x, then

$$\frac{X_t - x - (t-s)\mu(s,x)}{\sqrt{t-s}} \Rightarrow N(0, \sigma^2(s,x)) \text{ as } t \downarrow s.$$

Proof

Consider a second process, Y, which satisfies a similar stochastic differential equation to X, except that the coefficients are *frozen* to their initial values. That is,

$$dY_t = \sigma(s, x)dB_t + \mu(s, x)dt,$$

with $Y_s = x$. Then, Y is a Brownian motion with constant drift and diffusion coefficients, and thus

$$[Y_t] \sim N(x + (t-s)\mu(s,x), (t-s)\sigma^2(s,x)) \qquad t \geqslant s.$$

We now compare processes X and Y. Let

$$Z_t = \frac{X_t - Y_t}{\sqrt{t-s}} \qquad t \geqslant s.$$

From the forms of the stochastic differential equations, we can express Z_t as

$$Z_t = \frac{1}{\sqrt{t-s}} \int_s^t [\sigma(r, X_r) - \sigma(s,x)]dB_r + \frac{1}{\sqrt{t-s}} \int_s^t [\mu(r, X_r) - \mu(s,x)]dr.$$

We would like to show that Z is small. Consider $\mathbb{E}[|Z_t|]$.

$$\begin{aligned} \mathbb{E}[|Z_t|] &\leqslant \mathbb{E}\left[\frac{1}{\sqrt{t-s}}\left|\int_s^t [\sigma(r, X_r) - \sigma(s,x)]dB_r\right|\right] \\ &+ \mathbb{E}\left[\frac{1}{\sqrt{t-s}}\left|\int_s^t [\mu(r, X_r) - \mu(s,x)]dr\right|\right] \\ &= E_1 + E_2, \text{ say.} \end{aligned}$$

Using the property that variances are positive, $(\mathbb{E}[X])^2 \leqslant \mathbb{E}[X^2]$, and denoting by k_1 and k_2 the Lipschitz constants for σ, in t and x respectively, we have

$$
\begin{aligned}
E_1^2 &\leqslant \mathsf{E}\left[\frac{1}{t-s}\left(\int_s^t [\sigma(r,X_r)-\sigma(s,x)]dB_r\right)^2\right] \\
&= \mathsf{E}\left[\frac{1}{t-s}\int_s^t [\sigma(r,X_r)-\sigma(s,x)]^2 dr\right] \\
&\leqslant \frac{1}{t-s}\mathsf{E}\left[\int_s^t [k_1^2(r-s)^2+2k_1k_2(r-s)|X_r-x|+k_2^2(X_r-x)^2\right] dr \\
&= \frac{1}{t-s}\int_s^t k_1^2(r-s)^2 dr \\
&+ \frac{1}{t-s}\int_s^t 2k_1k_2(r-s)\mathsf{E}[|X_r-x|]dr \\
&+ \frac{1}{t-s}\int_s^t k_2^2\mathsf{E}[(X_r-x)^2]dr.
\end{aligned}
$$

The first equality follows from the isometry for stochastic integrals, see Øksendal (1985). The next inequality is derived from the Lipschitz properties of σ and the final equality uses $X \in L^2$. We now note that by selecting $t-s$ arbitrarily small, the integrands and hence E_1 can be made arbitrarily small. Specifically,

$$\text{For each } \epsilon_1 > 0 \text{ there exists a } t_1 \text{ such that } s < t < t_1 \Rightarrow E_1 < \epsilon_1. \tag{6.3}$$

We now consider E_2, denoting by k_3 and k_4 the Lipschitz constants for μ,

$$
\begin{aligned}
E_2 &\leqslant \mathsf{E}\left[\frac{1}{\sqrt{t-s}}\int_s^t |\mu(r,X_r)-\mu(s,x)|dr\right] \\
&\leqslant \mathsf{E}\left[\frac{1}{\sqrt{t-s}}\int_s^t [k_3(r-s)+k_4|X_r-x|dr\right] \\
&= \frac{1}{t-s}\int_s^t \left(\sqrt{t-s}k_3(r-s)+\sqrt{t-s}k_4\mathsf{E}[|X_r-x|]\right) dr.
\end{aligned}
$$

The first inequality comes from taking the modulus inside the integrand, the second from the Lipschitz properties of μ. The equality is obtained using $X \in L^1$. Again, we are averaging the integrand between s and t, and so making the integrand arbitrarily small makes E_2 arbitrarily small. Thus,

$$\text{For each } \epsilon_2 > 0 \text{ there exists a } t_2 \text{ such that } s < t < t_2 \Rightarrow E_2 < \epsilon_2. \tag{6.4}$$

Combining (6.3) and (6.4), from Markov's inequality (see for example Ross (1984)) we deduce that

$$\mathsf{P}[|Z_t| > c] \to 0 \text{ as } t \downarrow s,$$

for any $c > 0$.

To complete the proof, we note that

$$\frac{X_t - x - (t-s)\mu(s,x)}{\sqrt{t-s}} = \frac{Y_t - (t-s)\mu(s,x) - x}{\sqrt{t-s}} + Z_t$$

where the right hand side is the sum of a normal random variable and a random variable which converges to zero in probability. Standard weak convergence results (see Billingsley (1968)) enable us to conclude the claimed result.

★

The first of the necessary conditions for $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$ involves the diffusion coefficients.

Theorem 9 *Suppose that*

$$dX_t^1 = \sigma_1(t, X_t^1)dB_t + \mu_1(t, X_t^1)dt$$

$$dX_t^2 = \sigma_2(t, X_t^2)dB_t + \mu_2(t, X_t^2)dt,$$

with $X_s^2 = X_s^1 = x$. *If* $X^1, X^2 \in L^2$, μ_i *and* σ_i *are Lipschitz in* t *and* x, $i = 1, 2$, *and* $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$, *then* $\sigma_2(t,x) = \sigma_1(t,x)$ *for all* t, x.

Proof

Assume, without loss of generality, that $\sigma_1(s,x) > \sigma_2(s,x)$ for some s, x. Denoting the transition density of process i by $p_{s,t}^i$, we have

$$\frac{p_{s,t}^2(x, x + \delta\sqrt{t-s})}{p_{s,t}^1(x, x + \delta\sqrt{t-s})} \geqslant \frac{p_{s,t}^2(x,x)}{p_{s,t}^1(x,x)} \qquad \text{for all } t \geqslant s,\ \delta > 0, \tag{6.5}$$

using the definition of $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$ (see Chapter 3). We note that

$$\phi\left(\frac{\delta}{\sigma_2(s,x)}\right) < \phi\left(\frac{\delta}{\sigma_1(s,x)}\right),$$

and so from Lemma 7, the left hand side of (6.5) is strictly less than 1, in the limit as $t \downarrow s$, whereas the right hand side is 1. This contradicts Result 8 of Chapter 4, and so we deduce $\sigma_2(s,x) \geqslant \sigma_1(s,x)$ for all s, x. The reverse inequality is found by taking $\delta < 0$, and so we conclude that $\sigma_2(s,x) = \sigma_1(s,x)$ for all s, x.

★

We shall now assume, without loss of generality, that the diffusion coefficient is identically 1. This condition can be satisfied by an appropriate time change. We now find a necessary condition for the drift coefficient.

Theorem 10 *Suppose X^1 and X^2 satisfy the stochastic differential equations*

$$dX_t^i = dB_t + \mu_i(t, X_t^i)dt,$$

with $X_s^i = x$ and μ_i satisfies (6.1) *and* (6.2) *with $\sigma = 1$ ($i = 1, 2$).*

If $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$ then $\mu_2(s,x) \geqslant \mu_1(s,x)$ for all s, x.

Proof

Note that

$$\lim_{\delta \downarrow 0} \frac{\mathsf{E}[X_{s+\delta}^i - X_s^i \,|X_s^i = x]}{\delta} = \mu_i(s,x),$$

or alternatively,

$$\lim_{\delta \downarrow 0} \frac{\mathsf{E}[X_{s+\delta}^i \,|X_s^i = x]}{\delta} - \frac{x}{\delta} = \mu_i(s,x).$$

However, $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$ implies $[X_t^2] \overset{\text{st}}{\geqslant} [X_t^1]$ (see Result 4, Chapter 4), and hence

$$\mathsf{E}[X_{s+\delta}^2 \,|X_s^2 = x] \geqslant \mathsf{E}[X_{s+\delta}^1 \,|X_s^1 = x] \qquad \text{for all } \delta \geqslant 0.$$

Consequently,

$$\lim_{\delta \downarrow 0} \frac{\mathsf{E}[X_{s+\delta}^2 \,|X_s^2 = x] - \mathsf{E}[X_{s+\delta}^1 \,|X_s^1 = x]}{\delta} \geqslant 0,$$

which is equivalent to

$$\mu_2(s,x) \geqslant \quad \mu_1(s,x),$$

which leads to the claimed result, since the choice of s and x was arbitrary.

★

We can use the same argument to prove the following theorem.

Theorem 11 *Suppose X^1 and X^2 satisfy the stochastic differential equations*

$$dX_t^i = dB_t + \mu_i(t, X_t^i)dt,$$

with $X_s^i = x$, $i = 1, 2$.

If $[X_t^2] \overset{\text{st}}{\geqslant} [X_t^1]$ then $\mu_2(s,x) \geqslant \mu_1(s,x)$ for all s, x.

6.3 Sufficient Conditions for Strong Stochastic Ordering

In the case where $\mu_i(t,x) = \mu_i(t)$, we have seen that the condition $\mu_2(t) \geqslant \mu_1(t)$ for all t is also sufficient (see Result 9, Chapter 4). But, in the general case, when the drift coefficient is state dependent, this no longer holds. Consider the following counterexample.

Counterexample : $\mu_2(x) \geqslant \mu_1(x),\ \forall x,\ \not\Rightarrow [X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$.

Consider the processes X^1 and X^m for $m = 2, 3, \ldots$, with $X_s^i = x_0$, satisfying

$$dX_t^1 = dB_t$$

$$dX_t^m = dB_t + \mu_m(X_t^m)dt,$$

where

$$\mu_m(x) = \begin{cases} 0 & x > 0 \\ m & x \leqslant 0 \end{cases}.$$

Clearly $\mu_m(x) \geqslant 0$ for all x. We now recall Counterexample 2 of Chapter 3, that $[|B_t|] \overset{\text{sst}}{\not\geqslant} [B_t]$. For m sufficiently large, we can make the distribution of X^m sufficiently close to that of $|B|$, and so the inequality in the definition of strong stochastic ordering will no longer hold for all choices of x_0 and s. This provides a counterexample in the present case.

It is therefore apparent that some additional smoothness conditions need to be made about $\mu(t,x)$. We shall seek to establish such conditions.

Suppose we have two diffusion processes, X^1 and X^2, satisfying the stochastic differential equations

$$dX_t^1 = dB_t \tag{6.6}$$

$$dX_t^2 = dB_t + \mu(t, X_t^2)dt, \tag{6.7}$$

with $X_s^1 = X_s^2 = x$. We shall compare likelihood ratios by considering the expected values of indicator functions. We will assume process 2 has its drift coefficient replaced by $\delta\mu(t,x)$, where $\delta > 0$ is assumed to be small. Thus, we will start by establishing conditions on μ so that $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$, where X^1 and X^2 satisfy

$$\begin{aligned} dX_t^1 &= dB_t \\ dX_t^2 &= dB_t + \delta\mu(t, X_t^2)dt, \end{aligned} \tag{6.8}$$

with $X_s^1 = X_s^2 = x$. The rôle of the Cameron-Martin-Girsanov Theorem will enable a relatively straightforward extension from this case to the case comparing processes satisfying (6.6) and (6.7).

6.3.1 Cameron-Martin-Girsanov Theorem

Let P_t^i denote the measure, at time t, associated with process i, defined in (6.6) and (6.8), for $i = 1, 2$. Then, as was described in Chapter 2, the Radon-Nikodym derivative can be found using the Cameron-Martin-Girsanov Theorem, under certain conditions on the drift coefficient. Specifically, we need to impose the condition that

$$\mathsf{E}\left[\int_0^T \mu^2(t, X_t)dt\right] < \infty \qquad \text{for all } T > 0.$$

Then, for any path X, we have the Girsanov transformation

$$G(X) = \exp\left\{\delta \int_s^T \mu(t, X_t)dX_t - \frac{\delta^2}{2}\int_s^T \mu^2(t, X_t)dt\right\}. \tag{6.9}$$

6.3.2 The Conditions on $\mu(t, x)$ - SEE ERRATA.

Recall the definition of strong stochastic ordering:

$$\left[X_T^2\right] \overset{\text{sst}}{\geqslant} \left[X_T^1\right] \Leftrightarrow \frac{p_{s,T}^2(x, y_2)}{p_{s,T}^1(x, y_2)} \geqslant \frac{p_{s,T}^2(x, y_1)}{p_{s,T}^1(x, y_1)},$$

for all $T \geqslant s$ and $y_1 \leqslant y_2$, where the $p_{s,T}^i$ are the transition densities of the two processes. We may express the right hand side in terms of expectations:

$$\frac{\mathsf{E}_2[f_2(X)]}{\mathsf{E}_1[f_2(X)]} \geqslant \frac{\mathsf{E}_2[f_1(X)]}{\mathsf{E}_1[f_1(X)]} \tag{6.10}$$

where E_i denotes expectation with respect to the measure associated with process i, and $f_i(X) = I(X_T \in dy_i \,|X_s = x)$ are indicator functions.

We use (6.9) in (6.10), and therefore require to show

$$\frac{\mathsf{E}[f_2(X)G(X)]}{\mathsf{E}[f_2(X)]} \geqslant \frac{\mathsf{E}[f_1(X)G(X)]}{\mathsf{E}[f_1(X)]}.$$

Relabelling the right hand side and rearranging, we must prove

$$\mathsf{E}[f_2(X)f_1(Z)(G(X) - G(Z))] \geqslant 0, \tag{6.11}$$

where all expectations are now with respect to process 1, (standard Brownian motion in this case).

Denoting $b(X, Z) = G(X) - G(Z)$, we have $b(X, Z) \geqslant 0$ (for those paths which contribute to the expectation) if

$$\begin{aligned} b_2(X, Z) &= \delta\int_s^T \mu(t, X_t)dX_t - \delta\int_s^T \mu(t, Z_t)dZ_t - \frac{\delta^2}{2}\int_s^T (\mu^2(t, X_t) - \mu^2(t, Z_t))dt \\ &\geqslant 0, \end{aligned}$$

for all Brownian bridges with $X_T = y_2$, $Z_T = y_1$.

Theorem 12 *Let X^1 and X^2 satisfy* (6.6) *and* (6.8), *with $\mu(t,x) \geqslant 0$ such that*

$$\mathsf{E}\left[\int_0^T \mu^2(t, X_t^1)dt\right] < \infty \qquad \textit{for all } T > 0.$$

If $b_2(X, Z) \geqslant 0$ for almost all Brownian bridges with $X_T = y_2 \geqslant Z_T = y_1$, then

$$\left[X_t^2\right] \overset{\text{sst}}{\geqslant} \left[X_t^1\right].$$

Proof

The proof follows from the preceding discussion.

★

We shall seek conditions on μ such that $b_2(X, Z) \geqslant 0$. Note that for $\delta = 0$, $b_2(X, Z) = 0$. We shall study the behaviour of b_2 for small δ by differentiating. Let

$$b_3(X, Z) = \left.\frac{\partial b_2}{\partial \delta}\right|_{\delta=0} = \int_s^T \mu(t, X_t)dX_t - \int_s^T \mu(t, Z_t)dZ_t.$$

If we can show $b_3(X, Z) \geqslant 0$, then the required result would hold, at least for sufficiently small δ. We make use of the coupling (see for example Roberts (1993))

$$Z_t = X_t - \frac{(y_2 - y_1)(t - s)}{T - s}.$$

Then,

$$\begin{aligned} b_3(X, Z) &= \int_s^T \mu(t, X_t)dX_t - \int_s^T \mu\left(t, X_t - \frac{t-s}{T-s}(y_2 - y_1)\right) dX_t \\ &+ \frac{y_2 - y_1}{T - s}\int_s^T \mu\left(t, X_t - \frac{t-s}{T-s}(y_2 - y_1)\right) dt. \end{aligned}$$

Let $\epsilon = y_2 - y_1$, and note that $b_3(X, Z) = 0$ for $\epsilon = 0$. Consider

$$\left.\frac{\partial b_3}{\partial \epsilon}\right|_{\epsilon=0} = \int_s^T \frac{t-s}{T-s}\mu_x(t, X_t)dX_t + \frac{1}{T-s}\int_s^T \mu(t, X_t)dt.$$

By consideration of Itô's formula for $f(t, X_t) = \frac{t-s}{T-s}\mu(t, X_t)$, we may write

$$\left.\frac{\partial b_3}{\partial \epsilon}\right|_{\epsilon=0} = \mu(T, X_T) - \int_s^T \frac{t-s}{T-s}\left[\mu_t(t, X_t) + \frac{1}{2}\mu_{xx}(t, X_t)\right] dt.$$

We define the following condition:

(C1) For each y_2, and almost all Brownian Bridges, X, from (s,x) to (T,y_2), for all s, T, x,

$$\mu(T,y_2) - \int_s^T \frac{t-s}{T-s}\left[\mu_t(t,X_t) + \frac{1}{2}\mu_{xx}(t,X_t)\right] dt \geqslant 0 \text{ a.s.}$$

Under condition (C1), we have

$$\left.\frac{\partial b_3}{\partial \epsilon}\right|_{\epsilon=0} \geqslant 0,$$

which holds for all values of y_2. Note that evaluating at $\epsilon = 0$ is equivalent to evaluating at $y_1 = y_2$. Since the above expression is valid for all y_2, it follows that

$$\frac{\partial b_3}{\partial \epsilon} \geqslant 0,$$

and so $b_3(X,Z) \geqslant 0$, or alternatively,

$$\left.\frac{\partial b_2}{\partial \delta}\right|_{\delta=0} \geqslant 0.$$

We now consider the comparison between

$$\begin{aligned} dX_t^2 &= dB_t + \gamma(t,X_t^2)dt + \delta\mu(t,X_t^2)dt \\ dX_t^1 &= dB_t + \gamma(t,X_t^1)dt. \end{aligned}$$

If we denote the Girsanov transformation between process X^1 and standard Brownian motion by G_1,

$$G_1(X) = \exp\left\{\int_s^T \gamma(t,X_t)dX_t - \frac{1}{2}\int_s^T \gamma^2(t,X_t)dt\right\}. \tag{6.12}$$

We require to prove

$$\mathsf{E}\left[f_2(X)f_1(Z)G_1(X)G_1(Z)(G(X)-G(Z))\right] \geqslant 0.$$

We note the similarity between this expression and (6.11), in that we have our result provided $b_2(X,Z) \geqslant 0$, where b_2 is exactly as before. However, we can again establish

$$\left.\frac{\partial b_2}{\partial \delta}\right|_{\delta=0} \geqslant 0.$$

Importantly, we can choose to take $\gamma(t,x) = k\mu(t,x)$ for any value of k, and so, noting the idea of the proof of Result 6 in Chapter 4, we have

$$\left.\frac{\partial b_2}{\partial \delta}\right|_{\delta=k} \geqslant 0 \qquad \text{for all } k.$$

Thus, if we now integrate this expression, we have $b_2(X,Z) \geqslant 0$ for almost all Brownian bridges with $X_T = y_2,\ Z_T = y_1$, with $y_1 \leqslant y_2$.

We have the following theorem.

Theorem 13 *If* (C1) *holds,* X^1 *and* X^2 *satisfy* (6.6) *and* (6.8)*, where* $\mu(t,x) \geqslant 0$ *is such that*

$$\mathbb{E}\left[\int_0^T \mu^2(t, X_t^1)dt\right] < \infty \qquad \textit{for all } T > 0,$$

then $[X_t^2] \overset{\text{sst}}{\geqslant} [X_t^1]$ *for all* δ.

Proof

The preceding discussion cites condition (C1) as sufficient for $b_2(X,Z) \geqslant 0$. We can then apply Theorem 12.

★

We can produce the following generalisation:

Theorem 14 *If* X^1 *and* X^2 *satisfy*

$$\begin{aligned} dX_t^2 &= dB_t + \mu_2(t, X_t^2)dt \\ dX_t^1 &= dB_t + \mu_1(t, X_t^1)dt \end{aligned}$$

where $\mu_1(t,x)$ *satisfies*

$$\mathbb{E}\left[\int_0^T \mu_1^2(t, X_t^1)dt\right] < \infty \qquad \textit{for all } T > 0,$$

and $\mu(t,x) = \mu_2(t,x) - \mu_1(t,x)$ *satisfies the conditions of Theorem* 13*, then*

$$\left[X_t^2\right] \overset{\text{sst}}{\geqslant} \left[X_t^1\right].$$

Proof

We may express the stochastic differential equation for X^2 as

$$dX_t^2 = dB_t + \mu_1(t, X_t^2)dt + (\mu_2(t, X_t^2) - \mu_1(t, X_t^2))dt,$$

and note that the Girsanov transformation G_1, defined in (6.12), exists in this case, and so we require to prove

$$\mathsf{E}\left[f_2(X)f_1(Z)G_1(X)G_1(Z)(G(X) - G(Z))\right] \geqslant 0.$$

We can invoke Theorem 13, with $\delta = 1$ to establish this.

★

Condition (C1) is strong enough to ensure processes conditioned not to hit the same boundary are also strongly stochastically ordered.

Corollary 2 *Under the conditions of Theorem* 14,

$$\left[X_t^2 \,|\tau_f > t\right] \overset{\text{sst}}{\geqslant} \left[X_t^1 \,|\tau_f > t\right].$$

Proof

Consider the processes X^{fi}, which are modified versions of X^i, with absorption at a moving boundary, so that

$$X_s^{fi} = f(s) \Rightarrow X_t^{fi} = f(t) \qquad \text{for all } t \geqslant s.$$

Then, if $f(t) \notin A_j$,

$$\mathsf{P}[X_t^{fi} \in A_j] = \mathsf{P}[X_t^{fi} \in A_j \cap \tau_f > t]$$

for all t. Thus,

$$\left[X_t^{f2}\right] \overset{\text{sst}}{\geqslant} \left[X_t^{f1}\right] \Rightarrow \left[X_t^{f2} \,|\tau_f > t\right] \overset{\text{sst}}{\geqslant} \left[X_t^{f1} \,|\tau_f > t\right].$$

We also have $\left[X_t^{fi} \,|\tau_f > t\right] \sim \left[X_t^i \,|\tau_f > t\right]$. Thus, if we can show our modified processes are strongly stochastically ordered, we will have also shown that our original processes, conditioned not to have hit the boundary, are also strongly stochastically ordered.

If we use the previous definitions of Girsanov transformations G and G_1, and indicator functions f_1 and f_2, ((6.9), (6.10) and (6.12)), we have

$$\begin{aligned} \mathsf{P}[X_t^{fi} \in dy_j] &= \mathsf{E}_i^f[f_j(X)] \\ &= \mathsf{E}_i[f_j(X)b(X)], \end{aligned}$$

where $b(X) = I(X_s < f(s)$ for all $s < t)$, and E_i^f denotes expectation with respect to process X^{fi}. Therefore, $\left[X_t^{f2}\right] \overset{\text{sst}}{\geqslant} \left[X_t^{f1}\right]$ follows if

$$\frac{\mathsf{E}_1[G(X)f_2(X)b(X)]}{\mathsf{E}_1[f_2(X)b(X)]} \geqslant \frac{\mathsf{E}_1[G(X)f_1(X)b(X)]}{\mathsf{E}_1[f_1(X)b(X)]}$$

or equivalently,

$$\mathsf{E}\left[f_2(X)f_1(Z)b(X)b(Z)G_1(Z)G_1(X)(G(X) - G(Z))\right] \geqslant 0,$$

with expectation taken with respect to standard Brownian motion. Recall condition (C1) is sufficient to enforce $G(X) - G(Z) \geqslant 0$ for the relevant paths, and thus

$$\left[X_t^2 \left| \tau_f > t\right.\right] \overset{\text{sst}}{\geqslant} \left[X_t^1 \left| \tau_f > t\right.\right].$$

★

Corollary 3 *Under the conditions of Theorem* 14,

$$h_2(t) \geqslant h_1(t) \qquad \textit{for all } t,$$

where h_i is the hazard rate of the stopping time $\tau^i = \inf_{t>0}\{t : X_t^i \geqslant f(t)\}$.

Proof

The result follows from Corollary 2, Lemma 2 of Chapter 2 and an argument similar to that which will be used to prove Theorem 15.

★

Remarks

We shall look for specific examples where (C1) holds. We start by considering deterministic drifts, where $\mu(t, x) = \mu(t) \geqslant 0$. In such a case, $\mu_{xx} = 0$, and condition (C1) reduces to checking

$$I(T) := \mu(T) - \frac{1}{T - s}\int_s^T (t - s)\mu_t(t)dt \geqslant 0.$$

Integrating by parts,

$$\int_s^T (t - s)\mu_t(t)dt = (T - s)\mu(T) - \int_s^T \mu(t)dt,$$

and therefore

$$I(t) = \frac{1}{T - s}\int_s^T \mu(t)dt.$$

Since $\mu(t) \geqslant 0$ for all t, $I(T) \geqslant 0$ for all T, and we may apply our theorems. Note that this manner of verification is far quicker than looking at the ratio of transition densities, which were used to prove this result in Result 9.

Thus, if we have two processes

$$\begin{aligned} dX_t^2 &= dB_t + \mu_2(t, X_t^2)dt \\ dX_t^1 &= dB_t + \mu_1(t, X_t^1)dt, \end{aligned}$$

with

$$\mathsf{E}\left[\int_0^T \mu_1^2(t, X_t)dt\right] < \infty \qquad \text{for all } T < \infty,$$

and $\mu_2(t, x) = \mu_1(t, x) + \mu(t)$, where $\mu(t) \geqslant 0$ for all t, then

$$\left[X_t^2\right] \overset{\text{sst}}{\geqslant} \left[X_t^1\right].$$

Chapter 7

Hazard Rate Bounds

7.1 Introduction

As was discussed in the Exit Distributions Chapter, there are only a few cases in which a boundary f can be selected, for which an Itô diffusion X has a known first exit density. As an alternative to directly approximating the distribution function of

$$\tau = \inf_{t>0}\{t : X_t \geqslant f(t)\},$$

we shall find analytic upper and lower bounds upon it, which can then be used as approximations, with the maximum error determined by the other bound. This can be done by comparing the hazard rate of τ with the hazard rates of the first exit time across other boundaries, for which the exact exit distribution is known.

In the case of Brownian motion, we shall use straight lines for comparisons, since the Bachelier-Lévy formula (see Lerche (1986)) furnishes us with the exact first exit distribution, and therefore hazard rate. We shall choose the straight lines for comparison so that they envelope the boundary curve f, which will then allow the ordering of the boundaries to be exploited.

This idea of enclosing the boundary curve between straight lines also appears in Roberts (1993). He uses the method to produce inequalities for the ratio of first hitting time densities, for a process started at two different points.

In the remainder of this chapter, we shall restrict attention to the subclass of processes which are strongly stochastically monotone, and to boundaries f which are Lipschitz continuous.

We will also introduce a new approximation technique, for Brownian motion first exit distributions, based on approximating the hazard rate. The tangent approximation (Strassen (1967)) is shown to be inferior for concave and convex boundaries.

7.2 Notation

The following notation will be taken as standard:

$$\begin{array}{lcl} \tau_\lambda & = & \inf_{t>0}\{t : X_t \geqslant \lambda(t)\} \text{ for an arbitrary function } \lambda \\ \mu_t^\lambda(x)dx & = & \mathsf{P}[X_t \in dx \,|\tau_\lambda > t] \\ r_\lambda(t) & - & \text{hazard rate of } \tau_\lambda. \end{array}$$

Also for each fixed value of $t > 0$, and variable s, we shall define the enveloping curves $g_t(s)$ and $h_t(s)$ such that

$$g_t(s) \geqslant f(s) \geqslant h_t(s) \qquad \text{for all } s \leqslant t,$$

with

$$g_t(t) = f(t) = h_t(t),$$

and g_t and h_t are Lipschitz. Finally we assume that X is a strongly stochastically monotone Itô diffusion, satisfying

$$dX_t = \sigma(t, X_t)dB_t + \eta(t, X_t)dt.$$

7.3 Preliminary Lemmas

We shall employ two helpful results from Roberts (1993) and Roberts (1991a). The first provides a characterisation of the hazard rate of a hitting time, and was previously quoted as Lemma 2 in Chapter 2. We quote it again here for easy reference.

Lemma 2 (Roberts (1993)) *For an arbitrary Lipschitz boundary $\lambda(t)$ and an Itô diffusion X,*

$$r_\lambda(t) = \frac{1}{2}\sigma^2(t, \lambda(t)) \lim_{x\uparrow\lambda(t)} \frac{\mu_t^\lambda(x)}{\lambda(t) - x}.$$

Intuitively, this result states that the hazard rate is proportional to the derivative of the density of the conditioned process, evaluated close to the boundary.

The second lemma was previously quoted as Theorem 8 in Chapter 3. Again we quote it here for convenience.

Theorem 8 (Roberts (1991a)) *If X is strongly stochastically monotone, and the cadlag boundaries f and g are such that $f(t) \leqslant g(t)$ for all t, then*

$$\left[X_t \,|\tau_g > t\right] \overset{\text{sst}}{\geqslant} \left[X_t \,|\tau_f > t\right].$$

The intuition behind this result is that the conditioning is more severe from boundary f, and this "pushes down" process X more than the other conditioning does. As a result, the above ordering presents itself.

7.4 The Hazard Rate Bounds Theorem

The two lemmas may be combined to provide a bound on the hazard rate $r_f(t)$ by using $r_{g_t}(t)$ and $r_{h_t}(t)$. Note that the curves g_t and h_t are dependent on the value of t at which the hazard rates are being compared. We shall drop the subscripts t in the remainder, for notational convenience.

Theorem 15 *Let X be a strongly stochastically monotone Itô diffusion process, f be a Lipschitz continuous boundary, and $g(s)$ and $h(s)$ be as defined in section 7.2. Then*

$$r_h(t) \leqslant r_f(t) \leqslant r_g(t) \text{ for all } t. \tag{7.1}$$

Proof

Fix $t > 0$. First note that since f is Lipschitz, we can define curves g and h with the desired ordering. Therefore, applying Theorem 8, we have

$$[X_t \,|\tau_g > t] \overset{\text{sst}}{\geqslant} [X_t \,|\tau_f > t] \overset{\text{sst}}{\geqslant} [X_t \,|\tau_h > t].$$

We shall consider only the first of these strong stochastic inequalities, and prove the second hazard rate inequality. The first can be proved in the same manner, using the second strong stochastic inequality.

Note Lemma 2 yields

$$\begin{aligned} r_g(t) - r_f(t) &= \frac{1}{2}\sigma^2(t, g(t)) \lim_{x \uparrow g(t)} \frac{\mu_t^g(x)}{g(t) - x} \\ &- \frac{1}{2}\sigma^2(t, f(t)) \lim_{x \uparrow f(t)} \frac{\mu_t^f(x)}{f(t) - x} \end{aligned}$$

and since $f(t) = g(t)$, this reduces to

$$r_g(t) - r_f(t) = \frac{1}{2}\sigma^2(t, f(t)) \lim_{x \uparrow f(t)} \left[\frac{\mu_t^g(x)}{f(t) - x} - \frac{\mu_t^f(x)}{f(t) - x}\right].$$

Suppose that $\lim_{x \uparrow f(t)} \frac{\mu_t^g(x)}{x - f(t)} < \lim_{x \uparrow f(t)} \frac{\mu_t^f(x)}{x - f(t)}$, so that

$$\lim_{x \uparrow f(t)} \frac{\mu_t^g(x)}{\mu_t^f(x)} < 1.$$

Then, since our strong stochastic inequality is equivalent to (see Definition 3, Chapter 3)

$$\frac{\mu_t^g(x)}{\mu_t^f(x)} \leqslant \frac{\mu_t^g(y)}{\mu_t^f(y)} \qquad \text{for all } x \leqslant y,$$

we have

$$\frac{\mu_t^g(x)}{\mu_t^f(x)} < 1 \qquad \text{for all } x \leqslant f(t).$$

That is,

$$\mu_t^g(x) < \mu_t^f(x),$$

which is impossible since both are densities, and must integrate to 1.

Thus we conclude $r_g(t) \geqslant r_f(t)$.

★

7.5 Remarks and Corollaries

It is often more convenient, from an intuitive perspective, to use distribution functions rather than hazard functions. We can also produce bounds on the distribution function of our hitting time τ_f, by exploiting the algebraic relationship between hazard rates and distribution functions.

Corollary 4 *Let f be a Lipschitz continuous boundary, P_λ denote the distribution function for the first exit time across boundary $\lambda(t)$, and $g(s)$ and $h(s)$ be the curves defined in section 7.2. Then,*

$$P_h(t) \leqslant P_f(t) \leqslant P_g(t)$$

for all t.

The proof notes that $P_\lambda(t) = 1 - \exp\{-\int_0^t r_\lambda(s)ds\}$. Combined with Theorem 15, this immediately provides the claimed ordering.

Clearly, it does not make sense to seek ordering for the density functions in a similar way, as this would violate the density property of $\int p(t)dt = 1$. Thus, any ordering could only exist on a particular set of intervals, with the reverse ordering holding elsewhere.

For the Brownian motion case, we can explicitly calculate the form of the hazard rate across any straight line, using the Bachelier-Lévy formula. Therefore, we select our curves g_t and h_t to be straight lines. We use the following definitions:

$$\begin{aligned} m_t^2 &= \sup_{s<t} \tfrac{f(t)-f(s)}{t-s}. \\ m_t^1 &= \inf_{s<t} \tfrac{f(t)-f(s)}{t-s}. \\ c_g &= f(t) - m_t^1 t. \\ c_h &= f(t) - m_t^2 t. \\ g_t(s) &= m_t^1 s + c_g. \\ h_t(s) &= m_t^2 s + c_h. \end{aligned}$$

Then we have $g_t(s) \geqslant f(s) \geqslant h_t(s)$ for all $s \leqslant t$, with equality at time t. From the Bachelier-Lévy formula, if $\lambda(t) = a + bt$, and

$$\tau_\lambda = \inf_{t>0}\{t : B_t \geqslant \lambda(t)\},$$

we have the density of τ_λ given by

$$p_\lambda(t) = \frac{a}{t^{3/2}}\phi\left(\frac{a+bt}{\sqrt{t}}\right)$$

and distribution function

$$P_\lambda(t) = 1 - \Phi\left(\frac{a+bt}{\sqrt{t}}\right) + e^{-2ab}\Phi\left(\frac{bt-a}{\sqrt{t}}\right),$$

where ϕ and Φ denote the standard normal density and distribution function, respectively. Consequently, we can obtain the following corollary to Theorem 15.

Corollary 5 *If f is a Lipschitz continuous boundary, and r_f denotes the hazard rate of the first exit time of Brownian motion across it, then*

$$\frac{c_h\phi\left(\frac{f(t)}{\sqrt{t}}\right)}{t^{3/2}\left[\Phi\left(\frac{f(t)}{\sqrt{t}}\right) - e^{-2c_h m_t^2}\Phi\left(\frac{f(t)-2c_h}{\sqrt{t}}\right)\right]} \leqslant r_f(t) \leqslant \frac{c_g\phi\left(\frac{f(t)}{\sqrt{t}}\right)}{t^{3/2}\left[\Phi\left(\frac{f(t)}{\sqrt{t}}\right) - e^{-2c_g m_t^1}\Phi\left(\frac{f(t)-2c_g}{\sqrt{t}}\right)\right]},$$

where c_g, c_h and m_t^i are defined above.

7.6 Hazard Rate Tangent Approximation

For the Brownian motion case, we introduce a new approximation for the first exit distribution. This is based on estimating the hazard rate of the first exit time by the hazard rate of the tangent at the same time point, which can be found exactly by the Bachelier-Lévy formula. We shall denote this approximation by HRT. In the case where

$$\tau_f = \inf_{t>0}\{t : B_t \geqslant f(t)\},$$

is the first exit time from a concave, Lipschitz boundary, the tangent to the curve at each time point is the same as our upper enveloping straight line. Consequently, the upper analytic bound and the HRT method produce identical distributions. If f is a convex, Lipschitz boundary, the tangent to the curve at each time point is the same as the lower enveloping straight line. Therefore, the lower analytic bound and the HRT method produce the same distributions. In either of these cases, we can prove that the HRT technique produces more accurate approximations, to the distribution function, than the tangent approximation (Strassen (1967)) does.

In the remainder of this section, we shall use the following notation:

$f(s)$	boundary function.
$u_t(s)$	tangent to f at time t.
τ	$\inf_{t>0}\{t : B_t \geqslant f(t)\}$.
p, P	density, distribution function of τ.
p_T, P_T	tangent approximation to density, distribution function.
p_H, P_H	HRT approximation to density, distribution function.
r	hazard rate of τ.
r_T	hazard rate of the first hitting time of u_t.

We first show that the densities produced by the HRT method and the tangent approximation are ordered if the boundary is either concave or concave.

Theorem 16 *If f is a concave, Lipschitz boundary, then*

$$p_H(t) \leqslant p_T(t) \qquad \textit{for all } t.$$

Proof

Since f is concave, we have $u_t(s) \geqslant f(s)$ for all $s \leqslant t$. Therefore,

$$\mathsf{P}[\tau_{u_t} > t] \geqslant \mathsf{P}[\tau > t] = 1 - P(t), \tag{7.2}$$

and from Corollary 4

$$P(t) \leqslant P_H(t). \tag{7.3}$$

Then, by definition,

$$\begin{aligned} p_T(t) &= r_T(t)\mathsf{P}[\tau_{u_t} > t] \\ &\geqslant r_T(t)(1 - P(t)), && \text{by (7.2)}, \\ &\geqslant r_T(t)(1 - P_H(t)), && \text{by (7.3)}, \\ &= p_H(t). \end{aligned}$$

★

For convex boundaries, we can prove the following theorem, by reversing all the inequalities.

Theorem 17 *If f is a convex, Lipschitz boundary, then*

$$p_H(t) \geqslant p_T(t) \qquad \textit{for all } t.$$

The most important ordering results are for the distribution functions produced by the two approximation methods, which imply that the HRT method is superior to the tangent approximation.

Theorem 18 *If f is a concave, Lipschitz boundary, then*

$$P(t) \leqslant P_H(t) \leqslant P_T(t) \qquad \textit{for all } t.$$

Proof

This follows trivially from Corollary 4 and integrating the result of Theorem 16.

★

Similarly, we have

Theorem 19 *If f is a convex, Lipschitz boundary, then*

$$P(t) \geqslant P_H(t) \geqslant P_T(t) \qquad \textit{for all } t.$$

Another advantage of the HRT method over the tangent approximation, is that it does integrate to 1, if the process is such that $\tau < \infty$ a.s. This is not true of the tangent approximation, which always over-estimates the density for concave boundary curves, and so integrates to greater than 1. For convex boundaries, the tangent approximation always under-estimates the density and integrates to less than 1.

For other shapes of boundary curves, one possible advantage of the HRT method over the tangent approximation, is that the density

$$p(t) = r(t) \exp\left\{-\int_0^t r(s)ds\right\}$$

takes the previous estimates into account. If the previous density estimates have all been over-estimates, this will be reflected in over-estimates of r. Consequently $\exp\left\{-\int_0^t r(s)ds\right\}$ is smaller than it ought to be, and this reduces the estimate to $p(t)$. Conversely, if the previous densities have all been under-estimates, $\exp\left\{-\int_0^t r(s)ds\right\}$ is liable to be larger than it should be, thus increasing the current density estimate. This feedback effect makes fluctuations between over- and under-estimation less drastic, and intuitively, this may lead to a better approximation.

7.7 Numerical Examples

We conclude this chapter with some numerical examples illustrating the analytic bounds and our new HRT method. We also compare the distributions produced by these methods with the tangent approximation (see section 2.5.1). We shall use L to denote the lower bound, U to denote the upper bound and TA represents the tangent approximation.

The first two examples (Figs 7.1 and 7.2) use the same form of boundary function, but with different parameters. We use

$$f(t) = \frac{\theta}{2} - \frac{t}{\theta}\log\left(\frac{1}{2}\frac{\alpha}{a} + \left[\frac{1}{4}\left(\frac{\alpha}{a}\right)^2 + \frac{1-\alpha}{a}\exp\{-\theta^2/t\}\right]^{1/2}\right),$$

with the choices of parameters as follows:

Fig 7.1:- $\theta = 20$ $\alpha = 0.3$ $a = 0.3$.
Fig 7.2:- $\theta = 10$ $\alpha = 0.25$ $a = 1.05$.

Both of these curves are concave, and as a consequence of this, the upper analytic bound is equivalent to the HRT approximation. We also include the tangent approximation for comparison, and, as expected, the distribution it produces lies outside of our analytic bounds. The exact exit densities are found using the method of images, and the appropriate formulae are given in Lerche (1986). Note that the upper bound (HRT approximation) is tighter than the lower bound, because the tangent is a closer approximation to the curve than the lower enveloping straight line.

For boundaries which are neither concave or convex, the HRT method is no longer theoretically superior to the tangent approximation. However, we can still use hazard rates to compute the analytic bounds. As an example, the function $f(t) = 2+0.1t+0.25\sin(t)$ was used. Fig 7.3 gives a simulated "exact" distribution and density function, together with the analytic bounds. We also include the HRT and tangent approximations, which remain within the analytic bounds. The HRT method is again better than the tangent approximation, which has successive periods of over- and under-estimation partially cancelling the errors.

The next example is a convex boundary curve. In this case we choose a parabola, $f(t) = 3 + \frac{1}{12}(t-6)^2$. Again a simulation is used to produce the "exact" exit distribution. For purely convex curves, there exists a time, t_1, such that the tangent's intercept with the x axis is negative, for all $t \geqslant t_1$. In this example, $t_1 = \sqrt{72}$, beyond which the distribution function of the tangent approximation, and the analytic lower bound (equivalent to the HRT approximation in this convex boundary example), will remain constant. In this case, (Fig 7.4), the analytic upper and lower bounds will diverge, and thus the accuracy of any approximation using either, will become poorer. To counteract this, for larger time values, we

Fig 7.1a - Distribution Function Comparison

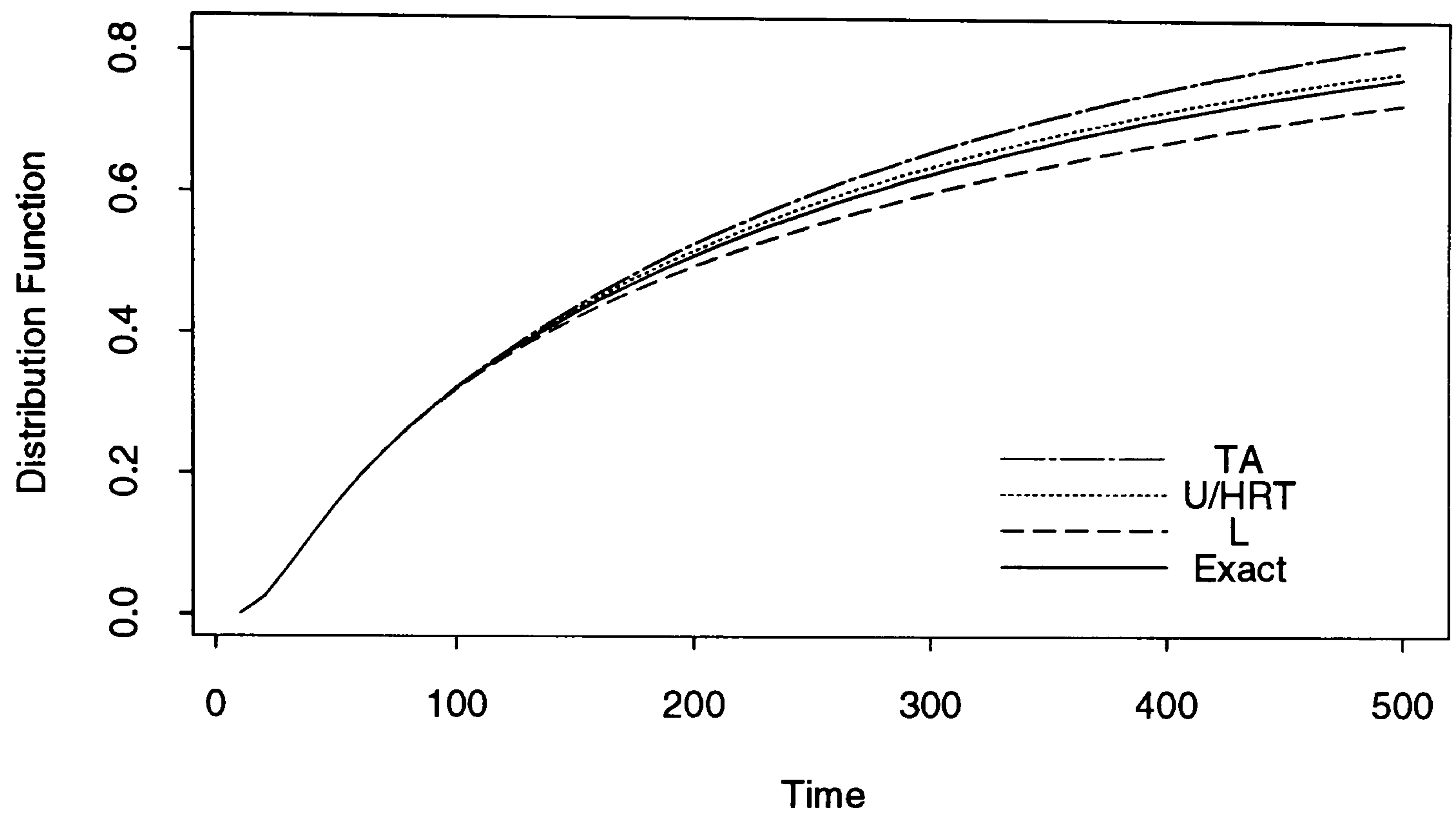


Fig 7.1b - Density Function Comparison

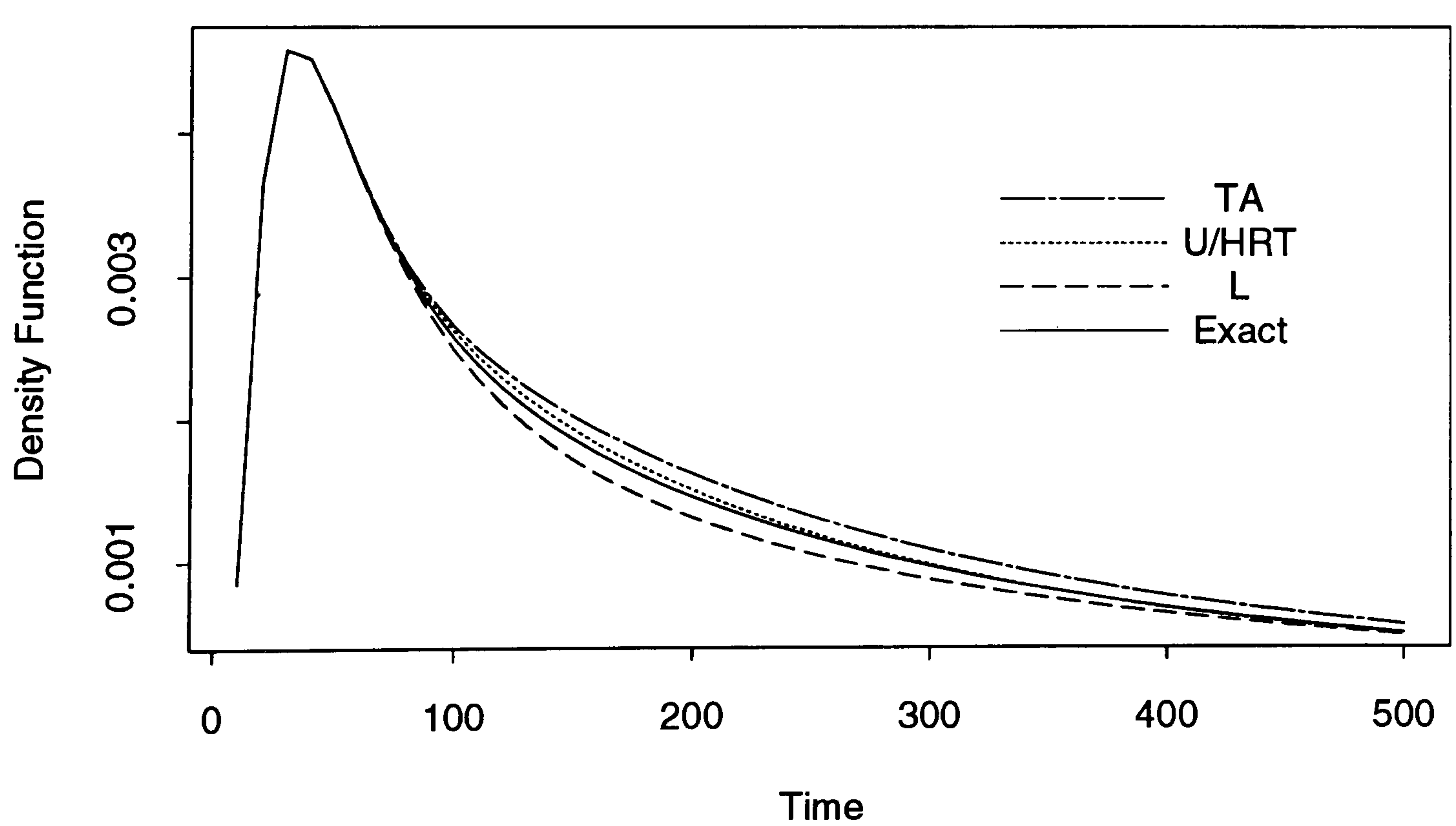


Fig 7.2a - Distribution Function Comparison

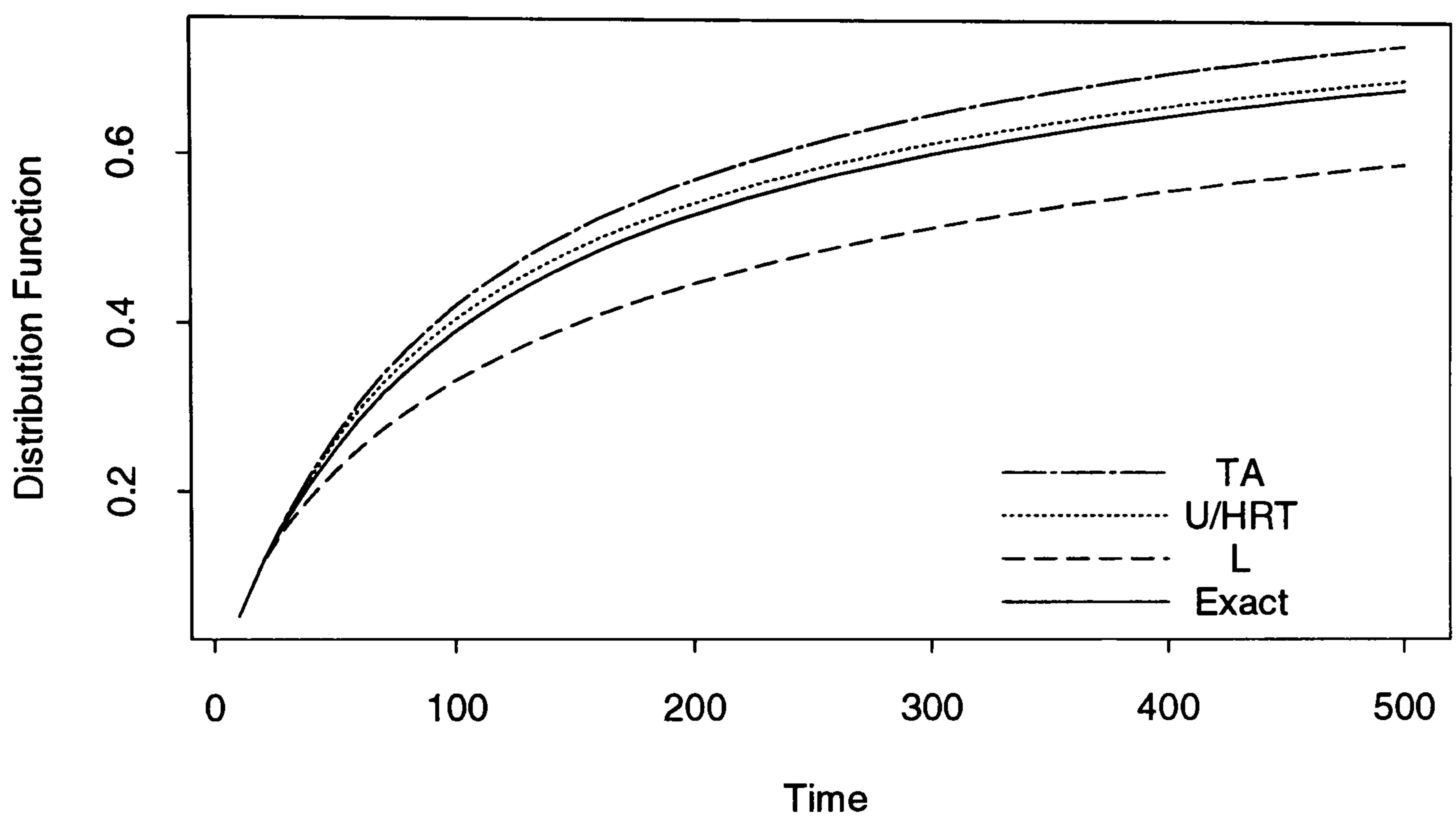


Fig 7.2b - Density Function Comparison

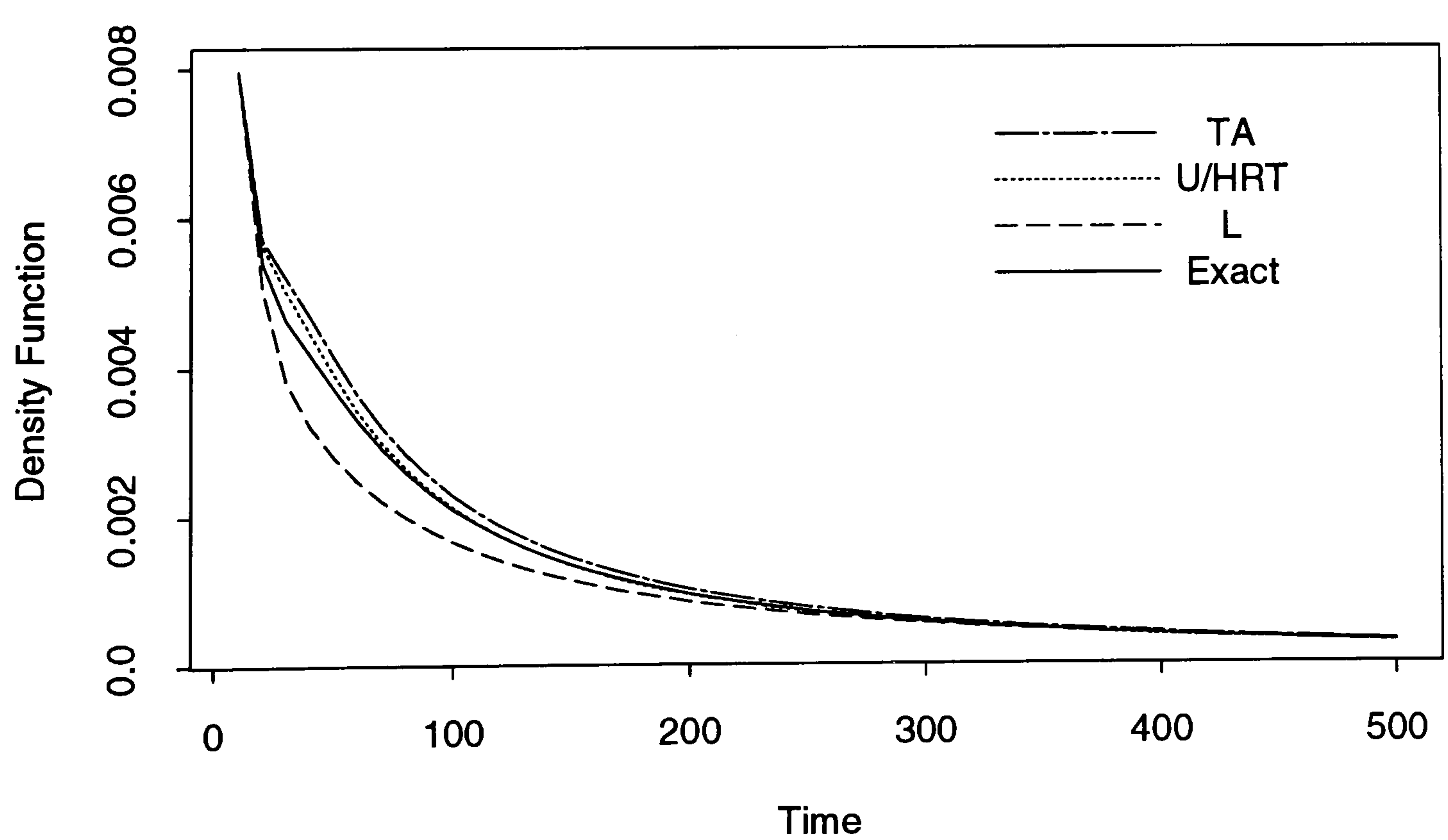


Fig 7.3a - Distribution Function Comparison

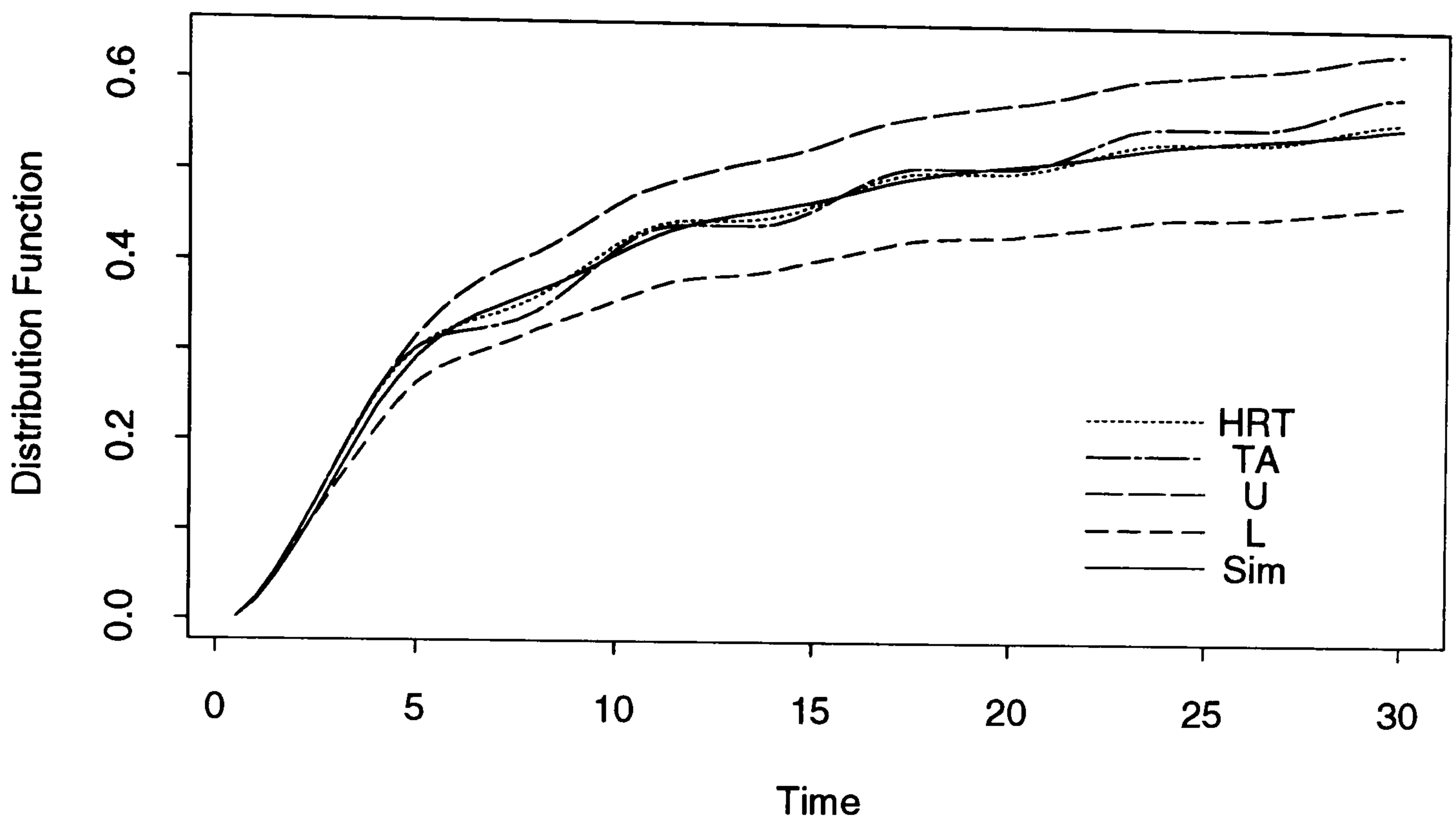
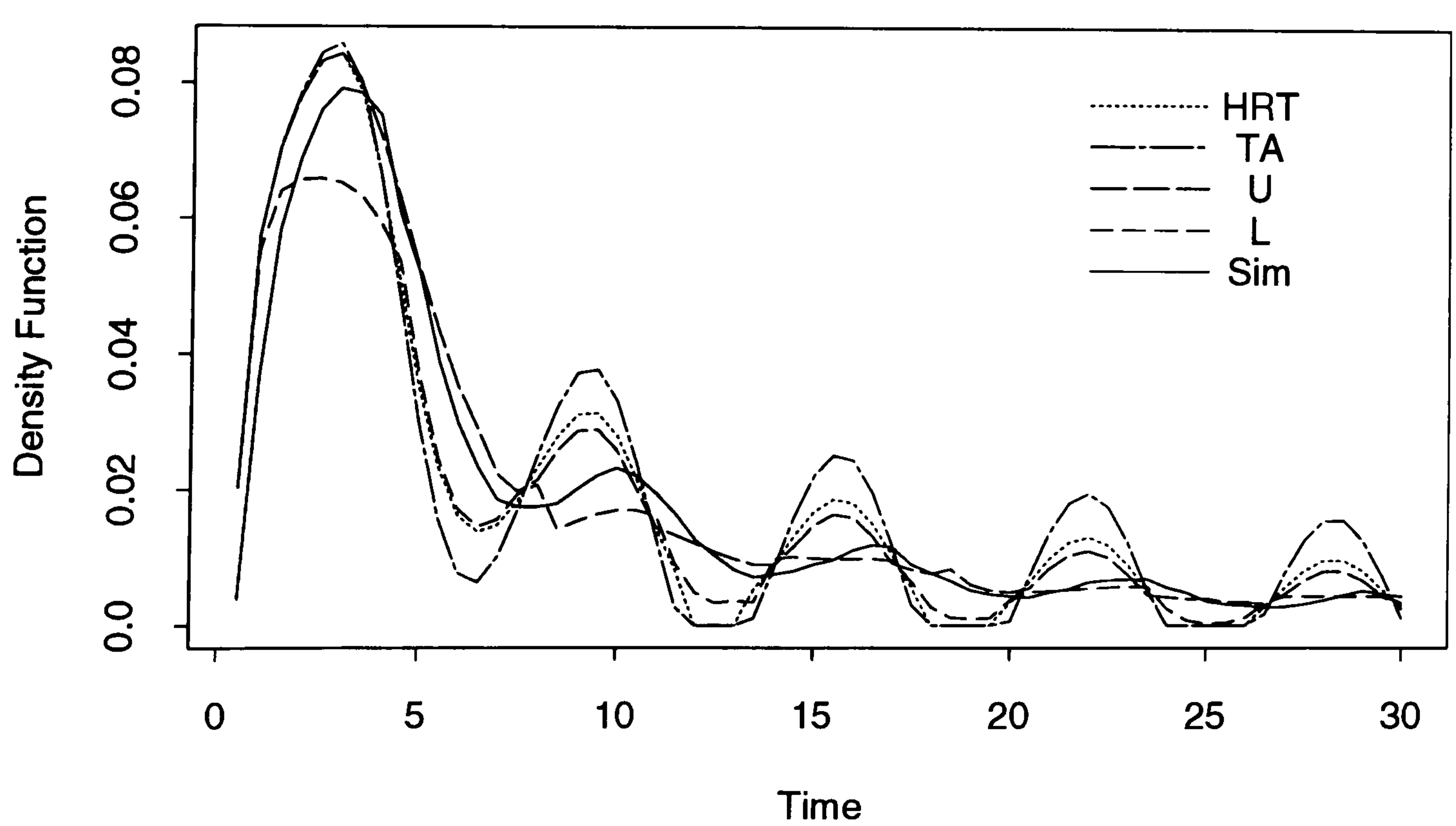


Fig 7.3b - Density Function Comparison



might consider using other approximation methods, for these times, perhaps based on quasi-stationary distribution functions.

The boundary curve $f(t) = 4 - e^{-t/10}$ is used to produce Fig 7.5, and the results are predictable, in that the HRT method is superior to the tangent approximation in this concave boundary case. The analytic bounds are tight, mainly because the curve can be enveloped between straight lines of similar gradient. Note that the simulation is slightly inaccurate, as the density is not over-estimated by the tangent approximation for all values of t. This is probably due to the discrete nature of the approximation technique.

The final example (Fig 7.6) is for $f(t) = 6e^{-t/4} + 2e^{-4/t}$. This curve is neither concave, nor convex, and the most noticeable feature is that the tangent approximation does not remain within the analytic bounds, whereas the HRT method and lower bound are virtually indistinguishable, and very accurate. Note that the curve is convex for $t < 24$, and so the lower bound and HRT methods produce identical results over this period. The lower bound produces a tight bound because the curve is well approximated by the lower enveloping straight line, except for small t, when few exits occur.

Note that in all cases where the exact distribution has been simulated, we used 200000 sample paths to obtain the empirical distribution.

We summarise, below, the boundaries used for the examples:

Fig 7.1 $\quad f(t) = 10 - (t/20)\log\left(\frac{1}{2} + \left[\frac{1}{4} + \frac{7}{3}e^{-400/t}\right]^{1/2}\right).$

Fig 7.2 $\quad f(t) = 5 - (t/10)\log\left(\frac{5}{42} + \left[\frac{25}{1764} + \frac{15}{21}e^{-100/t}\right]^{1/2}\right).$

Fig 7.3 $\quad f(t) = 2 + 0.1t + 0.25\sin(t).$

Fig 7.4 $\quad f(t) = t^2/12 - t + 6.$

Fig 7.5 $\quad f(t) = 4 - e^{-t/10}.$

Fig 7.6 $\quad f(t) = 6e^{-t/4} + 2e^{-4/t}.$

Fig 7.4a - Distribution Function Comparison

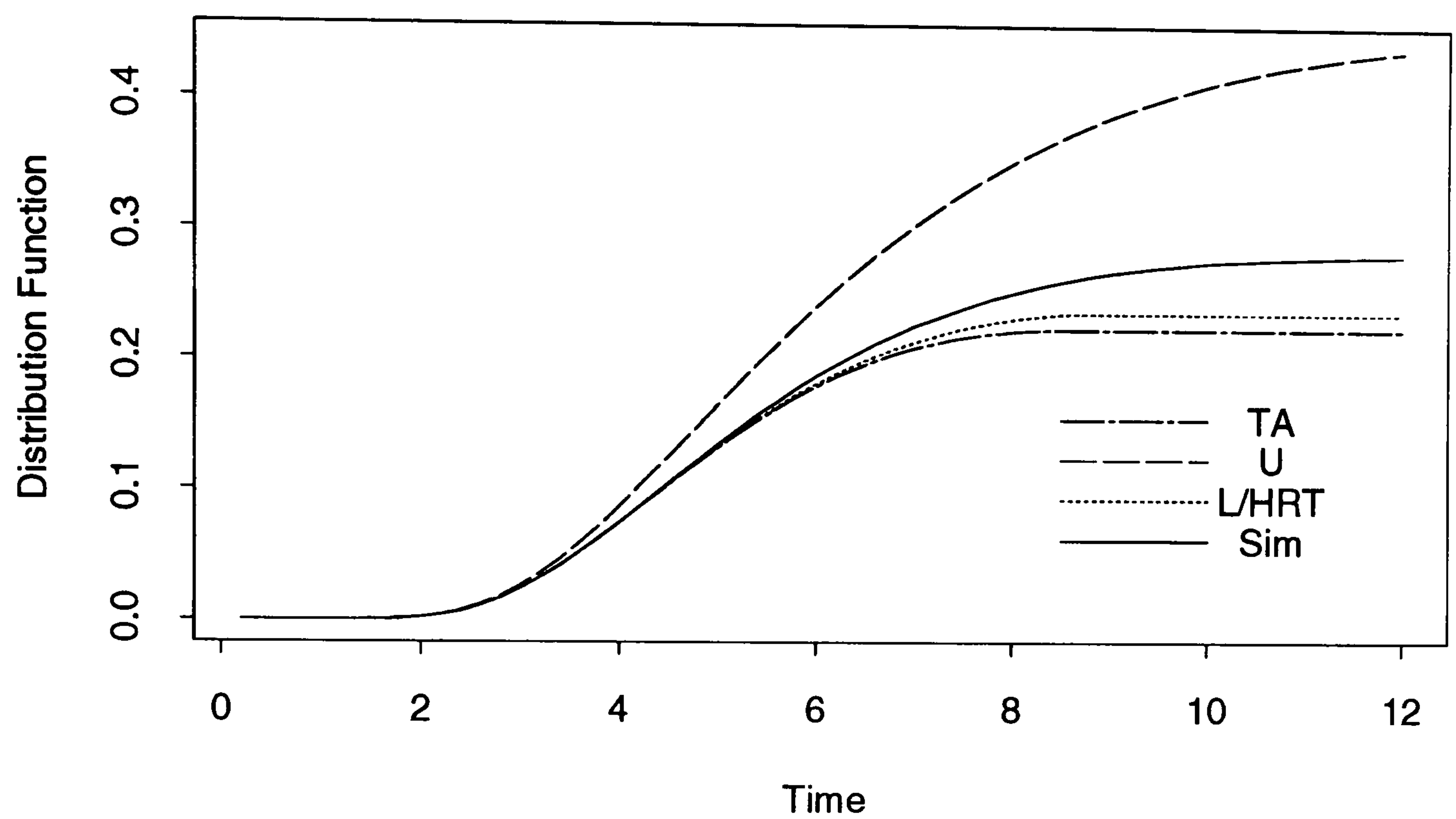


Fig 7.4b - Density Function Comparison

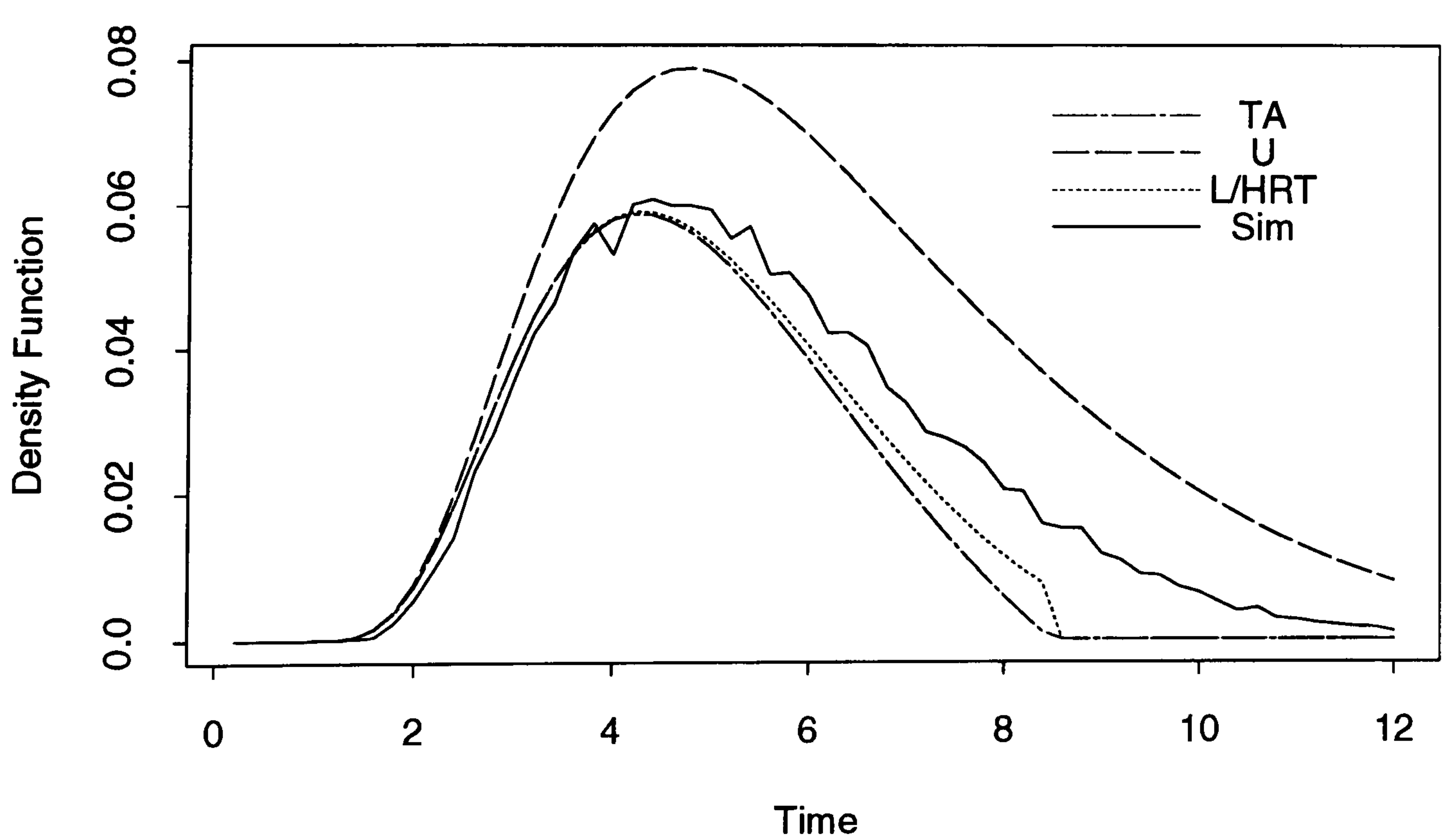


Fig 7.5a - Distribution Function Comparison

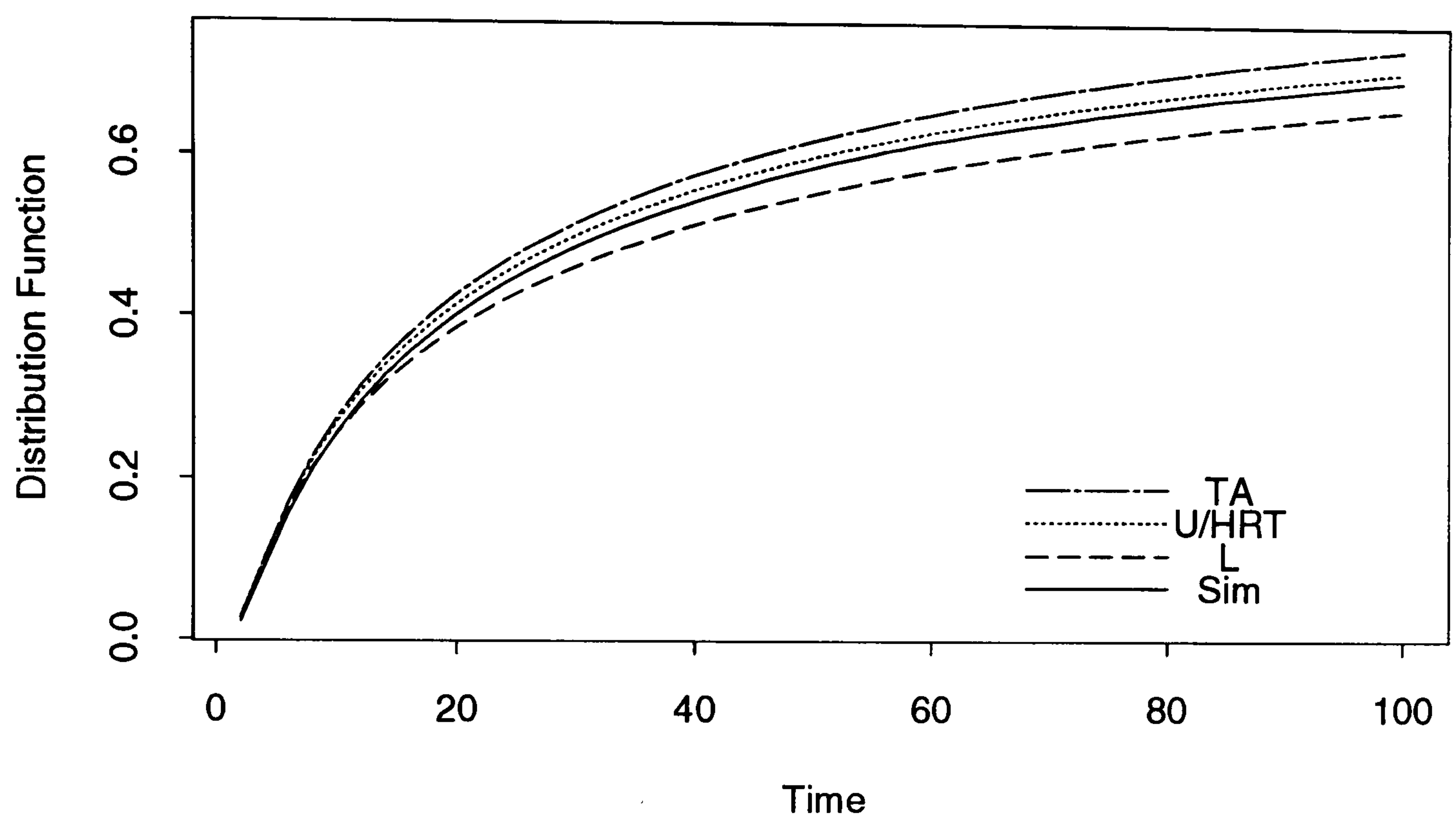


Fig 7.5b - Density Function Comparison

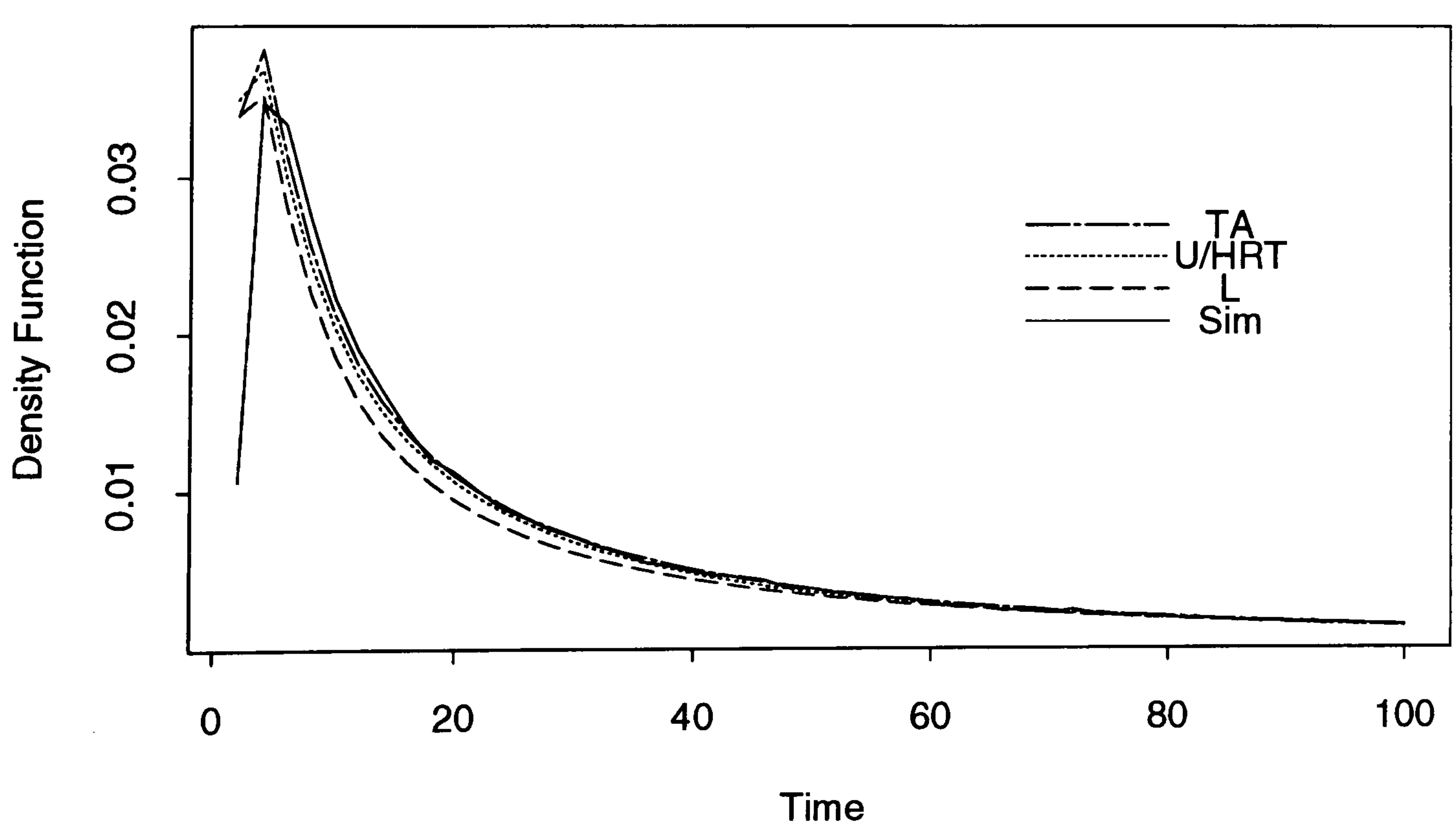


Fig 7.6a - Distribution Function Comparison

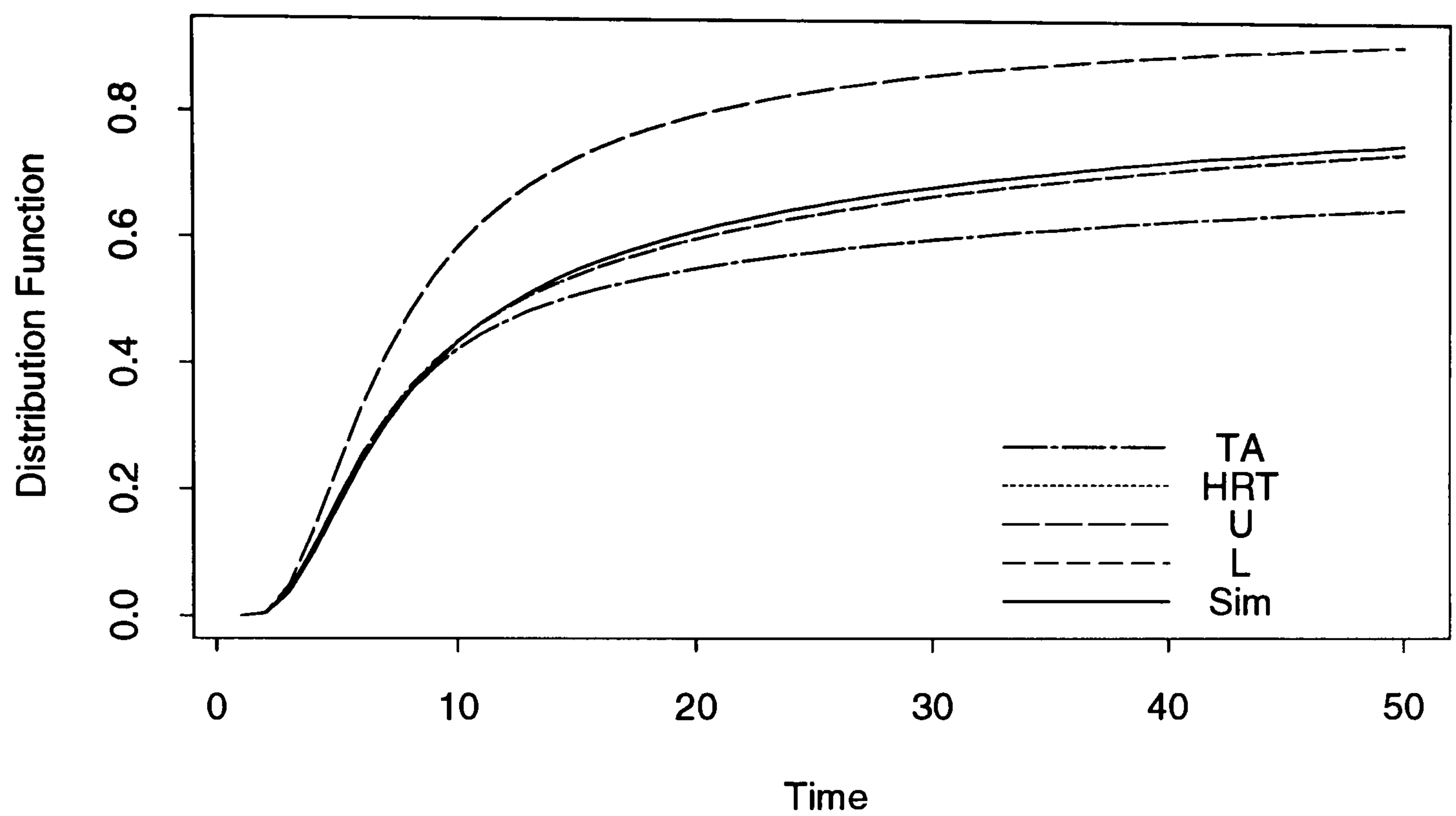
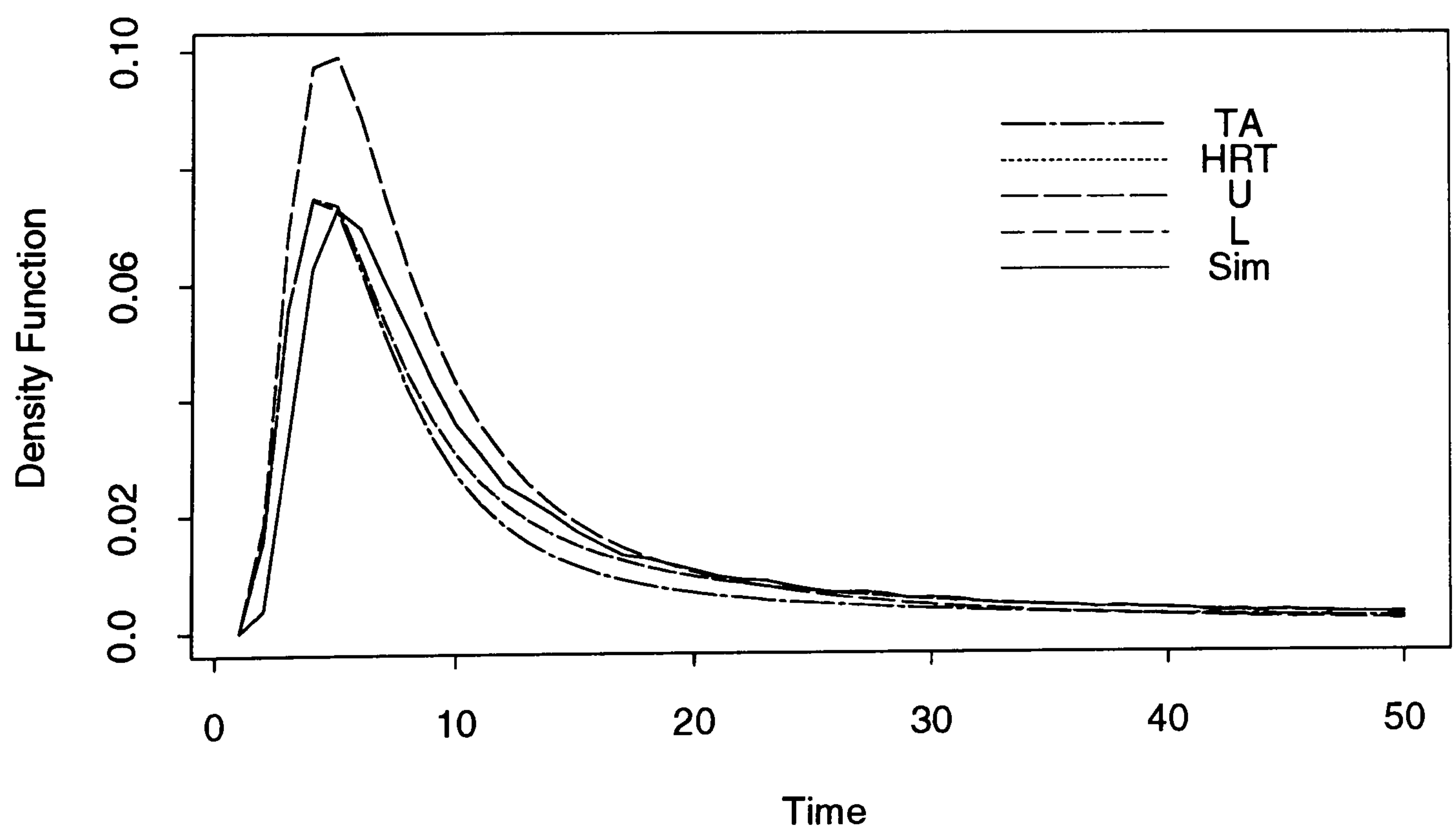


Fig 7.6b - Density Function Comparison



Chapter 8

An Optimal Stopping Problem

8.1 Introduction

To conclude this work, we illustrate the ideas of the previous chapters by considering an optimal stopping problem. The solution to this problem will be investigated using boundary hitting time techniques, and stochastic calculus. As an example of the use of approximations to the first exit distribution, applications in numerical optimisation and $K\epsilon$–optimality results will be highlighted.

In a class of optimal stopping problems, an inference has to be made about an unknown parameter of an observable process. However, observations have an associated cost, and a further cost is incurred for an inaccurate inference. Thus, the objective is to make the inference, whilst minimising the expected value of a pre-defined cost function, C. The functional dependence of C usually reflects the length of the observation period, and some feature of the path the process followed, such as the terminal value, X_τ.

In applications, the actual minimum value of the expected cost, is less important than the stopping rule, τ, which generates it. In many cases, the (t, x) plane can be split into two regions, C and $S = C^C$, for which the stopping rule can be expressed as

$$\tau = \inf_{t>0}\{t : (t, X_t) \notin C\}.$$

For obvious reasons, C is known as the continuation region, and S the stopping region. The boundary of C, ∂C, is called the optimal stopping boundary, because τ can be written as

$$\tau = \inf_{t>0}\{t : (t, X_t) \in \partial C\}.$$

To solve such problems analytically, the standard technique uses Bellman's equations, a system of a partial differential equation and boundary conditions. Since the optimal boundary is unknown, these are referred to as free boundary problems (see Krylov (1980)). Solving this partial differential system is usually difficult. Often the solution, and boundary, are guessed by heuristic arguments, and then proved to be correct (see Bĕnes, Shepp and Witsenhausen (1980) for examples). Another difficulty is that the condition of continuity of the solution across the boundary is insufficient to produce a unique solution. In such a case, the heuristic of smooth fit is often invoked, which says that the first derivative of the solution is also continuous across the boundary. Because this technique is non-trivial, we shall seek properties of the optimal stopping boundary by more direct methods in this chapter.

Bather, Chernoff and Petkau (1989) discuss methods for finding approximate, and exact results about the shape of the optimal continuation region. These techniques include the use of analytic methods to find bounds on the regions. They also discuss the use of known solutions to the heat equation to find asymptotic series expansions for the optimal payoff, and then find the boundary. The final method discussed is the use of simple random walks as approximations, in order to calculate the boundary numerically.

Throughout this chapter, we shall assume that we are given a cost function $C(t, X_t)$, which is dependent only on the time and position of the process. Clearly, in some problems, the entire history of the process X may be relevant in determining the cost. We will not look at such problems.

This chapter can be split into two parts. The first part looks at a stochastic control problem, and the second considers ϵ−optimal results (see Krylov (1980)). The specific stochastic control problem considered is that of optimally stopping Brownian motion, with respect to a cost function which depends only on the final position and the observation time. Properties of the optimal stopping boundary are then established by stochastic calculus techniques. As an example of this, throughout the exposition, we consider a worked example, motivated by a sequential analysis problem. In the following subsection, we will show how this problem can be transformed into a stochastic control problem of the form being investigated. Thus results for this problem will also be given.

In many cases, whilst the optimal solution cannot be found, it is possible to find a solution which produces an expected payoff within ϵ of the optimal expected cost. Such a solution is termed ϵ−optimal (see Krylov (1980)). In the final section of this chapter, we shall prove a number of ϵ−optimality results, in cases where the boundary is only approximated, and also when the payoff is calculated using an approximate exit distribution. The results with density approximations are illustrated by a numerical example, related to the worked example in the first section, making use of the UDHRR approximation developed in section 2.6.4.

8.1.1 A Worked Example

Sequential analysis problems may be categorised as a particular type of optimal stopping problem. In these problems, a process is observed and then stopped according to some pre-determined rule. Frequently a choice is then made between the null and alternative hypotheses. Many policies (a stopping rule and a decision rule) exist with the same operational characteristics, and comparisons between them are often made by considering the expected length of the observation period.

Consider the process

$$dX_t = dB_t + \theta dt,$$

that is, Brownian motion with constant drift θ, and suppose that θ is an unknown parameter. Lerche (1986) discusses tests of power one to decide $H_0 : \theta = 0$ or $H_1 : \theta \neq 0$. He selects a prior distribution for θ, F, given by

$$F = \gamma\delta_0 + (1-\gamma)N(0, 1/r),$$

where $0 < \gamma < 1$ and $r > 0$ are constants, $N(a, b)$ denotes a normal distribution with mean a and variance b, and δ_0 is a point mass at zero. To calculate the posterior distribution of θ, based on the path of the process until time t, we only require the final value of the path, X_t, as this is sufficient for θ. That is, the posterior distribution of θ, $P(\theta \,|X_t) = P(\theta \,|X_s,\ 0 \leqslant s \leqslant t)$. The objective is to find a stopping rule to minimise the Bayes risk given by

$$\rho(\tau) = \gamma \mathsf{P}_0[\tau < \infty] + (1-\gamma)c\int_{-\infty}^{\infty} \theta^2 \mathsf{E}_\theta[\tau]\phi(\sqrt{r}\theta)\sqrt{r}d\theta,$$

in which P_λ and E_λ denote the probability measure and expectation when the drift is λ. We use ϕ to denote the standard normal density, and $c\theta^2$ represents the cost of observation for one time unit. The decision to stop observation is equivalent to agreeing with the hypothesis that $\theta \neq 0$, since if $\theta = 0$, observations are free and therefore it would not be sensible to stop.

One unsatisfactory feature of this model, is that observation costs are chosen to be proportional to θ^2. In many examples, proportionality to $|\theta|$ is more natural, but this would lead to infinite Bayes risk, for all stopping times of tests of power one, if substituted into Lerche's formulation. This follows from a lemma of Darling and Robbins (1968), which states that if $\mathsf{P}_0[\tau < \infty] < 1$, then

$$\mathsf{E}_\theta[\tau] \geqslant \frac{-2\ln(\mathsf{P}_0[\tau < \infty])}{\theta^2}.$$

Consequently, if the observation costs are $c|\theta|$, the Bayes risk contains the integral

$$\int_{-\infty}^{\infty} |\theta|\mathsf{E}_\theta[\tau]\phi(\sqrt{r}\theta)\sqrt{r}d\theta \geqslant \int_{-\infty}^{\infty} \frac{|\theta|}{\theta^2}(-2\ln(\mathsf{P}_0[\tau < \infty])\phi(\sqrt{r}\theta)\sqrt{r}d\theta,$$

which is infinite, due to the singularity of the integrand at $\theta = 0$.

We shall, therefore, reformulate the problem in such a way that observation costs of $c|\theta|$ per unit time are permitted. Because of the above result, we cannot consider this as a sequential analysis type of hypothesis testing problem, and instead we shall switch to a stochastic control framework. We shall continue to observe the process, and thus improve our estimate for θ, until cost considerations make this no longer viable. Rather than choosing to reject or accept $H_0 : \theta = 0$, we shall give a posterior distribution for θ.

Instead of Lerche's mixed prior, we shall make the prior assumption that $[\theta] \sim N(0, 1/r)$, which leads to the posterior distribution

$$\left[\theta \,\middle|\, X_t\right] \sim N\left(\frac{X_t}{t+r}, \frac{1}{t+r}\right). \tag{8.1}$$

We will minimise the expectation of a cost function of the form

$$\text{cost} = t|\theta| + (\text{error costs}).$$

However, since θ is unknown, we shall use the estimate $\hat{\theta}_t = \mathsf{E}[\theta \,|X_t]$ in its place, when evaluating the cost at time t, and as an error cost, we shall use a constant multiple of $\mathsf{E}[(\theta - \hat{\theta}_t)^2 \,|X_t]$. Both these are readily available from (8.1). Note that the drift and diffusion coefficient for a diffusion process satisfy

$$\begin{aligned}
\mu(t, X_t) &= \lim_{h \downarrow 0} \frac{\mathsf{E}[X_{t+h} - X_t \,|X_t]}{h} \\
\sigma^2(t, X_t) &= \lim_{h \downarrow 0} \frac{\mathsf{E}[(X_{t+h} - X_t)^2 \,|X_t]}{h}.
\end{aligned}$$

We can evaluate these expectations by conditioning on θ, to obtain

$$\begin{aligned}
\mu(t, X_t) &= \frac{X_t}{t+r} \\
\sigma^2(t, X_t) &= 1.
\end{aligned}$$

These lead to the "posterior" stochastic differential equation

$$dX_t = dB_t + \frac{X_t}{t+r}dt, \tag{8.2}$$

which we will use.

Thus, we have now formulated a problem with θ no longer explicitly involved, but in which $\hat{\theta}$ is present. We shall, therefore, observe process X satisfying (8.2), and attempt to minimise the expected value of

$$C^\star(t, X_t) := t\left|\frac{X_t}{t+r}\right| + \frac{c}{t+r}. \tag{8.3}$$

On stopping we shall then report the posterior distribution, $\left[\theta \,|X_\tau\right]$.

Strictly speaking, in estimating the cost function, we should use $\mathsf{E}[|\theta| \,|X_t]$ to correctly model the structure. However, a simple explicit expression for such an expectation is not available, as is the case for $\left|\mathsf{E}[\theta \,|X_t]\right|$, which we use. Since

$$\left|\int f(x)dx\right| \leqslant \int |f(x)|dx,$$

we under-estimate $|\theta|$ in our cost function.

We will make use of the fact that a decision to stop at $|X_t| = x$ is made on the basis of our future expected cost on stopping, that is

$$\text{stop at } |X_t| = x \Leftrightarrow C^\star(t,x) \leqslant \inf_{\tau > t} \mathsf{E}[C^\star(\tau, X_\tau) \,\big| |X_t| = x], \tag{8.4}$$

where τ represents a stopping time. Note that by continuity of $C^\star(t,x)$, and the process X itself, the inequality in (8.4) can be replaced by an equality. Furthermore, (8.4) can be used to split the (t,x) plane into two regions, the stop region and the go or continuation region, where as the names suggest, you either stop observing the process, or continue observing the process, respectively.

Changing the time-scale

Although the process is observed in the (t,x) co-ordinate system, this is not necessarily the most useful time scale in which to make progress. Bather (1983) gives the example of observations comprising of a sequence of independent normal random variables, with unknown mean, θ, and known variance, σ. If a normal prior is placed on θ, it is well known that a $N(u,v)$ prior leads to a $N(u',v')$ posterior distribution, after the observation, x, where

$$u' - u = \frac{v}{v + \sigma^2}(x - u)$$

and

$$v' - v = -\frac{v^2}{v + \sigma^2}.$$

So, if the process is observed in (u,v) space, the transitions of v are deterministic. Thus, if the continuous analogue is used, and the stopping cost is $C(u_\tau, v_\tau)$, it is possible to separate out a deterministic part of the cost function. That is, we can write $C(u,v) = C_1(u,v) + C_2(v)$. In many applications, having a deterministic part to the cost will be advantageous. Notice that, if the variance is used as a "time" scale, we run the process backwards to zero, which corresponds to full information about θ.

We seek a time-scale transformation which will simplify our problem. Consider the standard time change result for Brownian motion,

$$[\alpha'(t)]^{-\frac{1}{2}} dB_{\alpha(t)} = dB_t^\star,$$

where $'$ denotes differentiation with respect to t; B and $B^\star$ are standard Brownian motions. We see that putting $X_t = [\alpha'(t)]^{-\frac{1}{2}} B_{\alpha(t)}$ yields

$$dX_t = dB_t^\star - \frac{1}{2}\frac{\alpha''(t)}{\alpha'(t)} X_t dt. \tag{8.5}$$

Comparison of (8.5) with (8.2) suggests that we should equate

$$-\frac{1}{2}\frac{\alpha''(t)}{\alpha'(t)} = \frac{1}{t+r},$$

which can be solved to yield $\alpha(t) = k - \frac{1}{t+r}$, and we choose the constant k so that $\alpha(0) = 0$, which means $k = \frac{1}{r} = T$, say. Substituting

$$X_t = (t+r) B_{T-\frac{1}{t+r}}$$

into (8.3), we obtain the cost function

$$t\left|B_{T-\frac{1}{t+r}}\right| + \frac{c}{t+r},$$

which, after substituting $T - s = \frac{1}{t+r}$, gives

$$C_s := C(s, B_s) = \frac{s}{T(T-s)}|B_s| + c(T-s), \tag{8.6}$$

and our objective now is to optimally stop B_s in order to minimise our expected value for C_s. For convenience, we will define $C(T,0) = 0$, since $\mathsf{P}[B_T = 0] = 0$.

We have now simplified the problem in two ways. Firstly, we are stopping a Brownian motion process, the properties of which are more accessible than those of our previous diffusion X. Secondly, we have a cost function in which the deterministic part is linear in time. Note our transformation is similar to Bather's, except we run the process index forwards to the time horizon, T, rather than backwards to "time" zero.

8.2 The Optimal Stopping Boundary

We now introduce the optimal stopping problem that we will investigate, of which the worked example, previously introduced, is a specific case. The objective will be to optimally stop a process subject to the cost function $C(t, x)$.

We make the following assumptions about the cost function $C(t,x)$:

1. $\lim_{t\uparrow T} C(t,x) = \infty$, $C(t,x) = \infty$ for all $t > T$ and $x \neq 0$.
2. $C(T,0) = 0$.
3. $C(t,x) = C(t,-x)$.
4. C has a continuous partial derivative in t, and a continuous second partial derivative in x for $x \neq 0$.

The process will be assumed to be Brownian motion, and we will use the notation $C_t = C(t, B_t)$. The first of these conditions is used to enforce an upper bound, T, on the optimal stopping time, τ. For suppose $p = \mathsf{P}[\tau \geqslant T]$, then

$$\begin{aligned} \mathsf{E}[C_\tau] &= p\mathsf{E}[C_\tau \,|\tau \geqslant T] + (1-p)\mathsf{E}[C_\tau \,|\tau < T] \\ &< \infty \qquad \text{if and only if } p = 0. \end{aligned}$$

Clearly the minimum expected stopping cost should be finite, as it is assumed the initial cost is finite.

We define the optimal expected payoff by

$$J(t,x) = \inf_{\tau > t} \mathsf{E}[C_\tau \,||B_t| = x] \tag{8.7}$$

and so a stopping rule of the form of (8.4) may be written

$$\text{stop at } |B_t| = x \Leftrightarrow J(t,x) \geqslant C(t,x). \tag{8.8}$$

We now consider the cost function as a diffusion. By Itô differentiation,

$$dC_t = \frac{\partial C}{\partial x} dB_t + \left(\frac{\partial C}{\partial t} + \frac{1}{2}\frac{\partial^2 C}{\partial x^2} \right) dt. \tag{8.9}$$

We shall denote the drift coefficient of this process by $D(t,x)$. That is,

$$D(t,x) := \frac{\partial C}{\partial t} + \frac{1}{2}\frac{\partial^2 C}{\partial x^2} \;. \tag{8.10}$$

We make the following assumptions about D, which will be used in the theorems of this section:

(D1) $D(t,0) < 0$.
(D2) (a) D is non-decreasing in t, for all t, x,
(b) D is non-decreasing in x, for all t, $x > 0$.
(D3) There exists $g : [0,\infty) \to \mathbb{R} \cup \{\infty\}$ such that $D(t, g(t)) = 0$ for all t.

Worked Example

Finally note that in our worked example, $C(t,x) = \frac{t}{T(T-t)}|x| + c(T-t)$, which has $\lim_{t\uparrow T} C(t,x) = \infty$ for $(x \neq 0)$ and $C(t,x) = C(t,-x)$. Itô differentiation gives

$$dC_t = \left(\frac{|B_t|}{(T-t)^2} - c\right) dt + \frac{t}{T(T-t)} d|B_t|.$$

This poses some technical problems in that we have $d|B_t|$ rather than dB_t. However, noting $|B|$ can be expressed as $M + l$, where M is a martingale and l is a local time at zero, these problems are only significant when B is close to zero. Such cases will be carefully considered. We will take $D(t,x) = \frac{|x|}{(T-t)^2} - c$, as if we do have dB_t. It is clear that the properties (D1), (D2)(a), (D2)(b) and (D3) hold.

8.2.1 Existence

In practice, the stopping rule in the form of (8.8) is not convenient, since the calculation of J takes time, whereas any decision to stop should be made instantly, especially for continuous processes. For this reason, we often seek a functional boundary, such that the stopping rule may be expressed as

$$\text{stop at } |B_t| = x \Leftrightarrow x \geqslant f(t). \tag{8.11}$$

Alternatively, we allow rules of the form

$$\text{stop at } |B_t| = x \Leftrightarrow t \geqslant \bar{f}(x) \tag{8.12}$$

to be termed an optimal stopping boundary. In such cases, the stopping rule is quickly verified, and in fact, if the process is continuous, we need only to monitor it closely when $(t, |B_t|)$ is close to $(t, f(t))$ or $(\bar{f}(x), x)$ in the (t,x) plane.

A variation in these results is possible in the two sided case. It is a common feature of optimal stopping problems that large values of the process correspond to significant effects, in which case the costs are high. These high costs encourage the process to be stopped quickly, when the feature is more clearly seen. However, it is possible that high values of the process lead to low costs, and so stopping rules of the form

$$\text{stop at } |B_t| = x \Leftrightarrow x \leqslant f(t)$$

and

$$\text{stop at } |B_t| = x \Leftrightarrow t \leqslant \bar{f}(x)$$

may be relevant.

To show the existence of an optimal stopping boundary, we merely have to show the existence of one of the rules (8.11) and (8.12) as optimal stopping rules. In the two-sided case, which we are considering, the easiest to show is (8.12), since this creates no technical difficulties when using the modulus of Brownian motion.

We will need some properties of the function

$$C^+(t,x) = C^+(t_1,t_2,t,x) = C(t,x) - C(t-t_2+t_1,x) \qquad (t_2 \leqslant t \leqslant T,\ x \geqslant 0),$$

to hold for all choices of $t_1 < t_2$:

(P1) For each $t_1 < t_2$ there exists $\epsilon = \epsilon(t_1,t_2)$ such that

$$\frac{\partial C^+}{\partial t} \geqslant 0 \qquad \text{for all } x \geqslant 0,\ t \geqslant T-\epsilon > t_2$$

$\dfrac{\partial^2 C^+(T-\epsilon,x)}{\partial x^2}$ is Lipschitz continuous in x and non-negative for all $x \geqslant 0$.

For the worked example, we have

$$C^+(t,x) = \frac{t_2-t_1}{(T-t)(T-t+t_2-t_1)}|x| + c(t_2-t_1),$$

which satisfies (P1) for $\epsilon < (t_2-t_1)/2$.

We have the following theorem:

Theorem 20 *If* (D2)(a) *and* (P1) *hold, then there exists a function* $\bar{f}$ *such that* (8.12) *holds.*

Proof

We shall prove this result by comparing the cost process when starting from two different times, at one of which it is assumed optimal to stop. Let $t_1 < t_2$ and assume that it is optimal to stop at $|B_{t_1}| = x$, and compare two processes, $|B^1|$ and $|B^2|$, which are coupled so that $|B^2_t| = |B^1_{t-t_2+t_1}|$ for $t_2 \leqslant t \leqslant T$. Let $C^i_t = C(t,B^i_t)$, $(i=1,2)$, so that $C^+(t,B^2_t) = C^2_t - C^1_{t-t_2+t_1}$ under our coupling.

Then, by definition,

$$\begin{aligned} J(t_2,x) &= \inf_{\tau>t_2} \mathsf{E}[C^2_\tau \,\big|\, |B^2_{t_2}| = x] \\ &= \inf_{\tau>t_2} \mathsf{E}[C^2_\tau - C^1_{\tau-t_2+t_1} + C^1_{\tau-t_2+t_1} \,\big|\, |B^2_{t_2}| = x] \\ &\geqslant \inf_{\tau>t_2} \mathsf{E}[C^2_\tau - C^1_{\tau-t_2+t_1} \,\big|\, |B^2_{t_2}| = x] + \inf_{\tau>t_2} \mathsf{E}[C^1_{\tau-t_2+t_1} \,\big|\, |B^2_{t_2}| = x] \\ &\geqslant \inf_{\tau>t_2} \mathsf{E}[C^+(\tau,B^2_\tau) \,\big|\, |B^2_{t_2}| = x] + C(t_1,x). \end{aligned} \tag{8.13}$$

The first inequality uses $\inf(A+B) \geqslant \inf(A) + \inf(B)$, and the second uses the assumption that it is optimal to stop at $|B_{t_1}| = x$.

For ϵ defined in (P1), consider the function

$$C^\epsilon(t,x) = \begin{cases} C^+(t,x) & t < T-\epsilon \\ C^+(T-\epsilon, x) & t \geqslant T-\epsilon \end{cases}$$

which, from (P1) satisfies $C^+(t,x) \geqslant C^\epsilon(t,x)$ for all t, x. Therefore,

$$\inf_{\tau > t_2} \mathsf{E}[C^+(\tau, B_\tau^2) \,||B_{t_2}^2| = x] \geqslant \inf_{\tau > t_2} \mathsf{E}[C^\epsilon(\tau, B_\tau^2) \,||B_{t_2}^2| = x]. \tag{8.14}$$

But,

$$dC_t^\epsilon = \begin{cases} [D(t,B_t) - D(t-t_2+t_1, B_t)]dt + a(t,B_t)dB_t & t < T-\epsilon \\ \frac{1}{2}\frac{\partial^2 C^+(T-\epsilon, B_t)}{\partial x^2} dt + b(B_t)dB_t & t \geqslant T-\epsilon \end{cases}, \tag{8.15}$$

where $a(t,x) = \frac{\partial C^+(t,x)}{\partial x}$, $b(x) = \frac{\partial C^+(T-\epsilon,x)}{\partial x}$ and D is defined in (8.10). From the existence and Lipschitz continuity of $\frac{\partial^2 C^+}{\partial x^2}$, we deduce the solution of (8.15) is well defined, and furthermore it is a submartingale.
Hence, for any stopping time τ, conditioning on $B_{T-\epsilon} = X$,

$$\begin{aligned} \mathsf{E}[C_\tau^\epsilon \,||B_{t_2}^2| = x, \tau > T-\epsilon] &= \mathsf{E}_X\left[\mathsf{E}[C_\tau^\epsilon \,||B_{t_2}^2| = x, \tau > T-\epsilon, B_{T-\epsilon} = X]\right] \\ &\geqslant \mathsf{E}_X[C^\epsilon(T-\epsilon, X) \,||B_{t_2}^2| = x] \\ &\geqslant C^\epsilon(t_2, x) \end{aligned}$$

and

$$\mathsf{E}[C_\tau^\epsilon \,||B_{t_2}^2| = x, \tau \leqslant T-\epsilon] \geqslant C^\epsilon(t_2, x),$$

using Doob's optional stopping theorem (see for example Dellacherie and Meyer (1978)), since both cases have a submartingale running until a bounded stopping time. Therefore,

$$\begin{aligned} \mathsf{E}[C_\tau^\epsilon \,||B_{t_2}^2| = x] &\geqslant C^\epsilon(t_2, x) \\ &= C(t_2,x) - C(t_1,x), \end{aligned}$$

as $t_2 < T-\epsilon$. Using (8.13) and (8.14), we conclude

$$J(t_2, x) \geqslant C(t_2, x),$$

so that it is optimal to stop at $|B_{t_2}| = x$.

We formally define

$$\bar{f}(x) = \inf_t \{t : J(t,x) \geqslant C(t,x)\},$$

and $\bar{f}$ is our optimal stopping boundary. This set cannot be empty, as

$$T \in \{t : J(t,x) \geqslant C(t,x)\},$$

for all x.

★

Worked Example

We can almost apply the theorem to show the existence of an optimal stopping bound, since $\frac{\partial D}{\partial t} = \frac{|x|}{(T-t)^3} \geqslant 0$. However, we do have to verify that

$$\left(\frac{\partial C(t,B_t)}{\partial x} - \frac{\partial C(t-t_2+t_1,B_t)}{\partial x}\right) d|B_t| \tag{8.16}$$

does not interfere with the submartingale behaviour of $C_t^2 - C_{t-t_2+t_1}^1$. The term (8.16) does not present a problem in this case, since $\frac{\partial C}{\partial x} = \frac{1}{(T-t)^2}$ is increasing in t, and $|B|$ is a submartingale. Hence we obtain our desired result.

8.2.2 Lower Bound

In order to prove the existence of a lower bound, we require an extra condition on the function g defined in (D3), and a further condition on the cost function.

(D4) $0 \leqslant g(T-\delta) \leqslant k\delta^p$ for some $p > 0$ and $k > 0$ as $\delta \downarrow 0$.
(D5) $C(t,g(t)) \leqslant \bar{C}$ on $[T-\delta, T]$ for small δ, for some constant $\bar{C}$.

Theorem 21 *If* (D1), (D2)(a) *or* (b), (D3), (D4) *and* (D5) *hold, then the function g, defined in* (D3), *acts as a lower bound to the optimal stopping boundary.*

Proof

First note that (D1) and (D3) yield $g(t) > 0$. Secondly, as the function $D(t,x)$ is non-decreasing in at least one of its variables, we deduce $D(t,x) \leqslant 0$ on A, where

$$A := \{(t,x) : |x| < g(t)\}.$$

We define the stopping time τ to be the first exit time from this set. That is,

$$\tau := \inf_{t>0}\{t : (t,B_t) \notin A\}.$$

We note the cost diffusion is a supermartingale on $A \cap \{t < T - \epsilon\}$, for any fixed $\epsilon > 0$. Thus,

$$\mathsf{E}[C_{\tau\wedge(T-\epsilon)}] \leqslant \mathsf{E}[C_0].$$

We wish to let $\epsilon \downarrow 0$.

Fix $0 < \delta < \frac{3}{4}T$. Since our diffusion process is Brownian motion, we have

$$\begin{aligned}
\mathsf{P}[\tau > T-\delta] &< \mathsf{P}[B_{T-\delta} \in (-g(T-\delta), g(T-\delta))] \\
&= \int_{-g(T-\delta)}^{g(T-\delta)} \frac{1}{\sqrt{2\pi(T-\delta)}} e^{-x^2/[2(T-\delta)]} dx \\
&< \frac{2g(T-\delta)}{\sqrt{T-\delta}} \\
&< \frac{4}{\sqrt{T}} g(T-\delta).
\end{aligned}$$

We shall now consider

$$
\begin{aligned}
|C_{\tau\wedge(T-\epsilon)} - C_0| &= |C_{\tau\wedge(T-\epsilon)} - C_{\tau\wedge(T-\delta)} + C_{\tau\wedge(T-\delta)} - C_0| \\
&\leqslant |C_{\tau\wedge(T-\epsilon)} - C_{\tau\wedge(T-\delta)}| \qquad (8.17)\\
&+ |C_{\tau\wedge(T-\delta)} - C_0|, \qquad (8.18)
\end{aligned}
$$

for $\epsilon < \delta$. But $\mathsf{E}[|C_{\tau\wedge(T-\epsilon)} - C_{\tau\wedge(T-\delta)}|]$ can be expressed as

$$
\begin{aligned}
\mathsf{E}[|C_{\tau\wedge(T-\epsilon)} - C_{\tau\wedge(T-\delta)}|] &= \mathsf{E}[|C_{\tau\wedge(T-\epsilon)} - C_{\tau\wedge(T-\delta)}|I(\tau \leqslant T-\delta)] \\
&+ \mathsf{E}[|C_{\tau\wedge(T-\epsilon)} - C_{\tau\wedge(T-\delta)}|I(\tau > T-\delta)] \\
&= 0 + \mathsf{E}[|C_{\tau\wedge(T-\epsilon)} - C_{\tau\wedge(T-\delta)}|\,|\tau > T-\delta]\mathsf{P}[\tau > T-\delta] \\
&\leqslant \bar{C} \times \frac{4}{\sqrt{T}} g(T-\delta).
\end{aligned}
$$

By Doob's optional stopping theorem (see for example Dellacherie and Meyer (1978)), (8.17) is integrable, as $\epsilon \downarrow 0$. So we deduce $|C_{\tau\wedge(T-\epsilon)} - C_0|$ is also integrable as $\epsilon \downarrow 0$, producing

$$\mathsf{E}[C_\tau] \leqslant \mathsf{E}[C_0].$$

Hence we conclude that it is never optimal to stop within A. So, if f satisfying (8.11) exists, we have $f(t) \geqslant g(t)$ for all t.

★

Worked Example

We apply the theorem noting that $g(t) = c(T-t)^2$, so $g(T-\delta) = c\delta^2$ satisfies the requirements of (D4), and that $C(T-\delta, g(T-\delta)) = 2c\delta - 2c\delta^2/T$, is bounded. Thus we conclude, if the optimal stopping bound, f, exists, then $f(t) \geqslant c(T-t)^2$.

8.2.3 Non-increasing

Having established the existence of the optimal stopping boundary by proving a rule of the form (8.12) can be found, we can deduce that this boundary is non-increasing if we can additionally show a rule of the form (8.11) exists. Note, however, that the use of $|B|$ in the cost function prevents the use of the coupling $|B_t^2| = |B_t^1| + k$, for all t, since this would force $|B^2|$ to be at least k, thus it will not behave as $|B|$. So we restrict the set of paths for which we apply such a coupling. On this restricted set of paths, we shall use the function C^- in a similar way to that in which C^+ was used for Theorem 20, where

$$C^-(t,x) = C^-(t,x,x_1,x_2) = C(t,x) - C(t, x - x_2 + x_1) \text{ for } x \geqslant x_1,$$

with $x_1 < x_2 < (3/2)x_1$.

We assume properties of this function:

(P2) For each $x_1 < x_2 < (3/2)x_1$ assume there exists $\epsilon = \epsilon(x_1, x_2)$ such that

$$\frac{\partial C^-}{\partial t} \geqslant 0 \text{ for all } t > T - \epsilon, \ x \geqslant x_1$$

$$\frac{\partial C^-(T-\epsilon, x)}{\partial x} \geqslant 0 \text{ for all } x > x_1$$

$$C^-(T-\epsilon, x) \text{ is convex in } x \text{ for } x > x_1.$$

The following theorem can now be proved.

Theorem 22 *If* (D1), (D2)(a), (D2)(b), (D3), (D4), (D5), (P1) *and* (P2) *hold, then there exists a non-increasing function* f, *which satisfies* (8.11).

Proof

From Theorem 21, we know that it is never optimal to stop at zero. So, by selecting $0 < x_1 < x_2 < (3/2)x_1$, the first inequality places no additional constraints on the problem, when we make the assumption that it is optimal to stop at $|B_s| = x_1$. Define

$$A := \{\tau > s : |B_t^2| \geqslant x_1 \ , \text{ for all } t \in [s, \tau]\}.$$

On A

We use the coupling $|B_t^2| = |B_t^1| + (x_2 - x_1)$, noting the choices of x_1 and x_2 allow this to be done without forcing $|B_t^1| < 0$. Denoting by C_t^i the cost function associated with $|B_t^i|$, from (8.9) we have

$$\begin{aligned} dC_t^-(t, |B_t^2|) &= [D(t, |B_t^2|) - D(t, |B_t^2| - x_2 + x_1)]dt \\ &+ [E(t, |B_t^2|) - E(t, |B_t^2| - x_2 + x_1)]d|B_t^2|, \end{aligned}$$

where D is defined in (8.10) and $E(t,x) = \frac{\partial C}{\partial x}$. Note that either $B_t^2 \geqslant x_1$ or $B_t^2 \leqslant -x_1$ on A, so that $|B_t^2|$ is indistinguishable from B_t^2. Hence

$$dC_t^- = [D(t, B_t^2) - D(t, B_t^2 - x_2 + x_1)]dt + a(t, B_t^2)dB_t^2,$$

where $a(t,x) = E(t,x) - E(t, x - x_2 + x_1)$.

Let ϵ be as defined in (P2) and Q be the conditioning event

$$Q = \{|B_s^2| = x_2, \tau \in A, \tau > T - \epsilon, B_{T-\epsilon} = X\},$$

then

$$\begin{aligned}
\mathsf{E}[C_\tau^- \,|Q] &= \mathsf{E}\left[\int_{T-\epsilon}^{\tau} \frac{\partial C^-(t, B_\tau^2)}{\partial t} + C^-(T-\epsilon, B_\tau^2)\,|Q\right] \\
&\geqslant 0 + C^-(T-\epsilon, \mathsf{E}[B_\tau^2\,|Q]) \\
&\geqslant C^-(T-\epsilon, X).
\end{aligned}$$

The first inequality uses $C^-(t,x)$ non-decreasing in t, the convexity property in (P2) and Jensen's inequality (see for example Ross (1984)). Noting that we have Brownian motion conditioned not to hit a lower boundary, we apply Theorem 7 of Chapter 3 to deduce $\mathsf{E}[B_\tau^2\,|Q] \geqslant X$. Then, the second inequality follows from the non-decreasing property of $C^-(T-\epsilon, x)$ in x.

Thus, for any $\tau \in A$,

$$\begin{aligned}
\mathsf{E}[C_\tau^-\,||B_s^2| = x_2, \tau > T-\epsilon] &\geqslant \mathsf{E}[C_{T-\epsilon}^-\,||B_s^2| = x_2] \\
&\geqslant C^-(s, x_2),
\end{aligned}$$

and

$$\mathsf{E}[C_\tau^-\,||B_s^2| = x_2, \tau \leqslant T-\epsilon] \geqslant C^-(s, x_2)$$

using Doob's optional stopping theorem (see for example Dellacherie and Meyer (1978)). Therefore,

$$\inf_{\tau\in A} \mathsf{E}[C_\tau^-\,||B_s^2| = x_2] \geqslant C^-(s, x_2). \tag{8.19}$$

Hence,

$$\begin{aligned}
\inf_{\tau\in A} \mathsf{E}[C_\tau^2\,||B_s^2| = x_2] &= \inf_{\tau\in A} \mathsf{E}[C_\tau^- + C_\tau^1\,||B_s^2| = x_2] \\
&\geqslant \inf_{\tau\in A} \mathsf{E}[C_\tau^-\,||B_s^2| = x_2] + \inf_{\tau\in A} \mathsf{E}[C_\tau^1\,||B_s^2| = x_2] \\
&\geqslant C^-(s, x_2) + C(s, x_1) = C(s, x_2),
\end{aligned} \tag{8.20}$$

from (8.19) and the assumption that it is optimal to stop at $|B_s| = x_1$.

<u>On A^C</u>

For each path with stopping time $\tau \in A^C$, there is a stopping time

$$t' = \inf_{t>s}\{t : |B_t^2| = x_1\} < \tau.$$

Since it is optimal to stop at $|B_{t'}^2| = x_1$ (by Theorem 20), we have

$$\begin{aligned}
\inf_{\tau\in A^C} \mathsf{E}[C_\tau^2\,||B_s^2| = x_2] &\geqslant \mathsf{E}[C_{t'}^2\,||B_s^2| = x_2] \\
&\geqslant \inf_{\tau\in A} \mathsf{E}[C_\tau^2\,||B_s^2| = x_2] \text{ as } t' \in A. \\
&\geqslant C(s, x_2).
\end{aligned} \tag{8.21}$$

Therefore, combining (8.20) and (8.21),

$$J(s, x_2) \geqslant C(s, x_2).$$

That is, it is optimal to stop at $|B_s| = x_2$. We can extend this argument as necessary to show it is optimal to stop at $|B_s| = x_2$, for all $x_2 > x_1$.

We require to show that there exists an x_1 such that it is optimal to stop at $|B_s| = x_1$. If we define

$$f(s) = \inf_x \{x : \ J(s, x) \geqslant C(s, x)\},$$

taking $f(s) = \infty$ if this set is empty, then we may set $x_1 = f(s)$ in our result. Combining this result with Theorem 20, which also holds under the conditions of this theorem, the discussion prior to the statement of this theorem allows us to conclude the optimal stopping bound in non-increasing. It should be noted that there may exist some times at which it is always optimal to continue ($f(t) = \infty$), but this does not detract from the non-increasing nature of the optimal stopping boundary.

★

Worked Example

In this case, $C^-(t, x) = t(x_2 - x_1)/(T(T - t))$ for $x \geqslant x_1$, $x_1 < x_2 < (3/2)x_1$, and so (P2) holds. Furthermore, we can simplify the proof by exploiting the linearity of the cost function in $|B|$. We note that we can express

$$C_t^2 = C_t^1 + \frac{(x_2 - x_1)t}{T(T - t)} \tag{8.22}$$

and the increasing nature of the deterministic function part of (8.22) makes the submartingale behaviour of $C_t^2 - C_t^1$ clearer.

8.2.4 Continuity

For practical applications, continuity of f is a desirable property. We shall adopt the following notation:

$$\begin{aligned} f(t-) &= \lim_{s \uparrow t} f(s) \\ f(t+) &= \lim_{s \downarrow t} f(s). \end{aligned}$$

We begin by proving f is right continuous.

Theorem 23 *If f is non-increasing, then f is right continuous.*

Proof

Since f is non-increasing, $f(t) \geqslant f(t+)$. Assume that $f(t) > f(t+)$, define $x = \frac{1}{2}(f(t)+f(t+))$ and consider the optimal policy from $|B_t| = x$. Since $x < f(t)$, it is optimal to continue from $|B_t| = x$. We define

$$\tau = \inf_{s>t}\{s : |B_s| \geqslant f(s)\}.$$

Since the sample paths of Brownian motion are continuous, we deduce $\tau = t$ almost surely, and thus have a contradiction.

We conclude it is optimal to stop at $|B_t| = x$, and have established right continuity.

★

Before looking at left continuity, we shall prove a useful lemma.

Lemma 8 *Let $S = \{(t,x) : c_1 < x < c_2\}$, $x_0 = \frac{1}{2}(c_1 + c_2)$ and*

$$\tau_S = \inf_{t>0}\{t : (t, B_t) \notin S \,|\, B_0 = x_0\}.$$

Then, for any $p > 0$

$$\frac{1 - \mathsf{P}[\tau_S > t]}{t^p} \to 0 \text{ as } t \downarrow 0.$$

Proof

Note that $\{\tau_S \leqslant t\} \subset (\{\tau_1 \leqslant t\} \cup \{\tau_2 \leqslant t\})$, where

$$\tau_i = \inf_{t>0}\{t : B_t = c_i \,|\, B_0 = x_0\}.$$

Thus,

$$\begin{aligned} \mathsf{P}[\tau_S \leqslant t] &\leqslant (\mathsf{P}[\tau_1 \leqslant t] + \mathsf{P}[\tau_2 \leqslant t]) \\ &= 4\left(1 - \Phi\left(\frac{c_2 - x_0}{\sqrt{t}}\right)\right) \end{aligned}$$

by the Bachelier-Lévy formula. But, for $k > 0$,

$$\begin{aligned} \lim_{t\downarrow 0} \frac{1 - \Phi(k/\sqrt{t})}{t^p} &= \lim_{t\downarrow 0} \frac{(k/2)t^{-3/2}\phi(k/\sqrt{t})}{pt^{p-1}} \\ &= 0. \end{aligned}$$

The first equality is obtained using L'Hôpital's rule, and the second uses the fact that $\phi(x) = ce^{-x^2/2}$. As the limit is non-negative, we apply the Sandwich theorem, noting $\phi(k/\sqrt{t}) \leqslant \alpha t^v$ for some $\alpha > 0$ and any $v > 0$.

Since this holds for all $p > 0$, we deduce our result.

★

To prove left continuity, we require additional conditions on the cost function, and the lower bound g, defined in (D3). Let $P := \{(t,x) : \frac{\partial C}{\partial t} + \frac{1}{2}\frac{\partial^2 C}{\partial x^2} > 0\}$ and let the following two properties hold for $(t,x) \in P$:

(D6) Let $C(t,\cdot)$ be convex for each t, and C be bounded for $t < T$ and x bounded.
(D7) $C(t,x) - C(t-\epsilon,x) \geqslant k\epsilon^p,\ k > 0,$ for some $p > 0$ and ϵ sufficiently small.

We finally assume that

(D8) g is Lipschitz in t, $|g(t) - g(s)| \leqslant k_2|t-s|$.

Theorem 24 *Let* (D6)–(D8) *hold and assume f is non-increasing. Then f is left continuous.*

Proof

We shall assume that $f(t-) > f(t)$, and seek a contradiction. Let

$$b_1 = f(t) + k_2\epsilon$$

$$b_2 = f(t-) - k_2\epsilon$$

and

$$x = \frac{1}{2}(b_1 + b_2),$$

where ϵ is sufficiently small to ensure that $b_1 < b_2$. We shall consider optimal policy from $|B_{t-\epsilon}| = x$, and define

$$A := \{(t,z) : b_1 < z < b_2\}$$

and the first exit time from this set

$$\tau_A := \inf_{s>t-\epsilon}\{s : (s,B_s) \notin A\}.$$

Note that $A \subseteq P$. Our optimal payoff J satisfies

$$\begin{aligned} J(t-\epsilon,x) &\geqslant \mathsf{P}[\tau_A > t]\mathsf{E}[C(t,|B_t|)\,|\tau_A > t,\ |B_{t-\epsilon}| = x] \\ &\geqslant \mathsf{P}[\tau_A > t]C(t,\mathsf{E}[|B_t|\,|\tau_A > t,\ |B_{t-\epsilon}| = x]) \\ &= \mathsf{P}[\tau_A > t]C(t,x). \end{aligned}$$

In this expression, the first inequality is due to the possibility of paths not having their optimal stopping time at time t, which leads to an additional positive expectation. The second follows from property (D6) and an application of Jensen's inequality (see for example Ross (1984)). To establish the final equality, we use the fact that, on A, $|B|$ and B are indistinguishable. Furthermore, as the distribution of Brownian motion is symmetric about its starting point and we also have symmetric conditioning, it follows that

$$\mathsf{E}[B_t \,|\tau_A > t,\; B_{t-\epsilon} = x] = x.$$

Therefore,

$$\begin{aligned} J(t-\epsilon, x) - C(t-\epsilon, x) &\geqslant C(t,x) - C(t-\epsilon,x) - (1 - \mathsf{P}[\tau_A > t])C(t,x) \\ &\geqslant k\epsilon^p - (1 - \mathsf{P}[\tau_A > t])C(t,x) \qquad (8.23) \\ &= \epsilon^p \left(k - \frac{(1 - \mathsf{P}[\tau_A > t])C(t,x)}{\epsilon^p} \right) \\ &> 0, \end{aligned}$$

for ϵ sufficiently small, using (D7), then (D6) and Lemma 8. (Note (8.23) also assumes ϵ is sufficiently small. We therefore take ϵ small enough to allow (8.23) to follow, and the final inequality to hold.)

Thus we have a contradiction to the optimality of continuing from $|B_{t-\epsilon}| = x$ and have established left continuity.

★

Worked Example

We have previously shown that the optimal stopping boundary in our worked example is non-increasing, and so right continuity follows trivially from Theorem 23. Note that the lower bound, $g(t) = c(T-t)^2$, is such that $|g(t) - g(t-\epsilon)| \leqslant 2cT|\epsilon|$ and for $x > g(t)$ we have $C(t,x) - C(t-\epsilon,x) > c\epsilon^2/T$. Thus we may apply Theorem 24 with $k = c/T$, $k_2 = 2cT$ and $p = 2$. Note also that the linearity of the cost function in $|x|$, circumvents the need of convexity and Jensen's inequality, so the second inequality in the proof may be replaced by equality.

8.2.5 Upper Bound

In most cases, the optimal stopping boundary is bounded above by a sufficiently large constant. However, from a practical point of view, this is not useful. In some cases, it is possible to do considerably better, and establish an upper bound on f, such that the upper bound also decreases to zero. For the worked example, we shall prove the existence of such a bound, using another coupling idea, incorporating the self-scaling property of Brownian motion. We shall find a function h such that $h(t) \geqslant f(t)$ for all t. However, before we obtain this result, where $h(t) = k\sqrt{T-t}$ (k constant), we shall give a result about the behaviour of f for t close to T.

An Asymptotic Result for f

Consider starting the process from $B_s = 0$, for some s. Then, the cost if the process is stopped immediately is $c(T-s)$. We define the curve $k_s(t)$ such that if the process is stopped on this curve, the cost will be $c(T-s)$. Thus, k_s is the solution to

$$\frac{tk_s(t)}{T(T-t)} + c(T-t) = c(T-s)$$

or

$$k_s(t) = cT(T-t)(1-s/t).$$

If $f(t) > k_s(t)$ for all $t \in (s,T)$, then

$$\mathsf{E}[C(\tau, f(\tau))\,|B_s = 0] > \mathsf{E}[C(\tau, k_s(\tau))\,|B_s = 0],$$

where τ is the optimal stopping time, using the non-decreasing behaviour of C in x. The right hand expectation is $c(T-s)$, by definition of k_s, and so it is optimal to stop at $B_s = 0$ — a clear contradiction to Theorem 21. Thus, we must have

$$\inf_{t>s}\{t : f(t) = k_s(t)\} < T.$$

This result holds for all s.

Now consider the behaviour of $k_s(t)$. We have $k_s'(t) = cT\left((Ts/t^2) - 1\right)$, so that $k_s'(T) \to 0$ as $s \to T$. Thus, we deduce

$$\limsup_{\epsilon \downarrow 0} \frac{-f(T-\epsilon) + f(T)}{\epsilon} = 0.$$

Using this statement, we further deduce an asymptotic bound of f. Specifically, there must exist a constant $\kappa > 0$ and a $\delta_1 > 0$ such that

$$f(t) \leqslant \kappa(T-t) \qquad \text{for } t > T - \delta_1.$$

Global Upper Bound

We intend to show $f(t) \leqslant k\sqrt{T-t}$, for some constant k, and so we select the initial values in the coupling to reflect this. We select t_2 and x_2 such that

$$t_2 = pt_1 + (1-p)T \text{ and } x_2 = \sqrt{p}x_1 \ (0 < p < 1),$$

so that both points are on the same $\sqrt{T-t}$ curve. We define B^i to be the Brownian motion process started from $|B^i_{t_i}| = x_i$, $(i = 1,2)$. $C(t,x)$ is defined in (8.6). We use the self-scaling properties of Brownian motion, so that both processes have indexes running to T, which is reached simultaneously, in the coupling

$$|B^2_{tp+(1-p)T}| = \sqrt{p}|B^1_t|. \tag{8.24}$$

Assume it is optimal to stop at $|B^1_{t_1}| = x_1$. So, by definition,

$$\inf_{\tau > t_1} \mathsf{E}[C(\tau, B^1_\tau) \,\big|\, |B^1_{t_1}| = x_1] \geqslant C(t_1, x_1).$$

We now consider optimal policy from $|B^2_{t_2}| = x_2$. By definition, the stopping time $\tau = \inf_{s>t_2}\{s : |B^2_s| \geqslant f(s)\}$ is such that $J(t_2,x_2) = \mathsf{E}[C(\tau, B^2_\tau) \,\big|\, |B^2_{t_2}| = x_2]$. Therefore, using the coupling arrangement, we have

$$J(t_2,x_2) = \mathsf{E}\left[\frac{\tau|B^1_{(\tau-(1-p)T)/p}|\sqrt{p}}{T(T-\tau)} + c(T-\tau) \,\big|\, |B^1_{t_1}| = x_1\right].$$

Making the substitution $s = (\tau - (1-p)T)/p$, so that $s > t_1$, we obtain

$$\begin{aligned}
J(t_2,x_2) &= \mathsf{E}\left[\frac{ps+(1-p)T}{Tp(T-s)}\sqrt{p}|B^1_s| + cp(T-s) \,\big|\, |B^1_{t_1}| = x_1\right] \\
&= \mathsf{E}\left[\sqrt{p}\left(\frac{s|B^1_s|}{T(T-s)} + c(T-s)\right)\right. \\
&+ \left.\frac{1-p}{\sqrt{p}(T-s)}|B^1_s| + (p-\sqrt{p})c(T-s) \,\big|\, |B^1_{t_1}| = x_1\right] \\
&= \sqrt{p}\mathsf{E}[C(s,B^1_s) \,\big|\, |B^1_{t_1}| = x_1] + (\sqrt{p}-p)\mathsf{E}[c(s-T) \,\big|\, |B^1_{t_1}| = x_1] \\
&+ \frac{1-p}{\sqrt{p}}\mathsf{E}\left[\frac{|B^1_s|}{T-s} \,\big|\, |B^1_{t_1}| = x_1\right],
\end{aligned}$$

where all the expectations are multiplied by positive constants. The first expectation is at least $\inf_{\sigma > t_1} \mathsf{E}[C(\sigma, B_\sigma) \,\big|\, |B^1_{t_1}| = x_1] \geqslant C(t_1,x_1)$, since it is optimal to stop at $|B^1_{t_1}| = x_1$. The second expectation is at least $c(t_1 - T)$, as $s - T$ in non-decreasing in s. For the third expectation, we note that $\frac{1}{T-s} \geqslant \frac{1}{T-t_1}$, so that

$$\mathsf{E}\left[\frac{|B^1_s|}{T-s} \,\big|\, |B^1_{t_1}| = x_1\right] \geqslant \mathsf{E}\left[\frac{|B^1_s|}{T-t_1} \,\big|\, |B^1_{t_1}| = x_1\right].$$

We now use the fact that $|B|$ is a submartingale, and hence

$$\mathsf{E}\left[\frac{|B^1_s|}{T-t_1} \,\big|\, |B^1_{t_1}| = x_1\right] \geqslant \frac{x_1}{T-t_1}.$$

Combining the lower bounds on the three expectations, and rearranging, we have

$$J(t_2, x_2) \geqslant C(t_2, x_2).$$

That is, it is optimal to stop at $|B_{t_2}| = x_2$. This result will hold for any choice of t_2, x_2 such that the point lies on the correct $\sqrt{T-t}$ curve. So, to show that there is an upper bound of this form, we need show that $f(0)$ is finite, that is, there exists an $x < \infty$ such that it is optimal to stop at $|B_0| = x$.

Let $t_0 = \inf_{t \geqslant 0}\{t : f(t) < \infty\}$. There are three possibilities. Either $t_0 = 0$, in which case we have nothing to prove, or $t_0 = T$, which would yield infinite expected stopping costs as $\mathsf{P}[|B_T| > 0] > 0$ and is, therefore, discounted, or lastly $0 < t_0 < T$. We shall assume this case. Consider the process beginning from $|B_0| = x > f(t_0)$. Then we can find a lower bound on the probability that $|B_{t_0}|$ is in the range $(f(t_0), 2x - f(t_0))$ in which case the process should be stopped at precisely this time. By the Bachelier-Lévy formula, we have

$$\mathsf{P}[|B_{t_0}| \in (f(t_0), 2x - f(t_0)) \,\big|\, |B_0| = x] \geqslant 4\Phi\left(\frac{x - f(t_0)}{\sqrt{t_0}}\right) - 3 = P_d \text{ say.}$$

Furthermore, by a similar argument to that which was used in Theorem 24 for left continuity, we can establish the fact that

$$\mathsf{E}[C_\tau \,\big|\, |B_0| = x] \geqslant \left(\frac{t_0 x}{T(T - t_0)} + c(T - t_0)\right) P_d.$$

For it to be optimal to stop at $|B_0| = x$, we require that

$$\mathsf{E}[C_\tau \,\big|\, |B_0| = x] > cT,$$

which clearly holds if

$$\left(\frac{t_0 x}{T(T - t_0)} + c(T - t_0)\right) P_d > cT.$$

Rearranging this, we see that we require

$$\begin{aligned} x &> \frac{T(T-t_0)}{P_d t_0}\left[cT - c(T - t_0)P_d\right] \\ &= \frac{cT^2(T-t_0)}{P_d}\frac{1-P_d}{t_0} + cT(T-t_0). \end{aligned} \tag{8.25}$$

Since we require this result to hold for all t_0, we need to take limits as $t_0 \downarrow 0$ in the expression. Recalling Lemma 8, $(1 - P_d)/t_0 \to 0$ as $t_0 \downarrow 0$, and thus the right hand side of (8.25) has the finite limit cT^2 as $t_0 \downarrow 0$. Thus we can clearly choose a finite value of x such that this equation holds. Denoting such a choice by $\hat{X}^0$, we have established $f(0) \leqslant \hat{X}^0$, where $\hat{X}^0 > cT^2$, and so, we can now obtain the upper bound

$$f_u(t) \leqslant \frac{\hat{X}^0}{\sqrt{T}}\sqrt{T-t},$$

which holds for all $t \geqslant 0$, and hence $f(T) = 0$.

8.2.6 One-Sided Cost Functions

Another category of cost functions are the one-sided cost functions, $C_1(t,x)$, such that $\lim_{t\uparrow T} C_1(t,x) = \infty$, but in which symmetry in x is not a requirement. A similar set of results may be established about an optimal stopping boundary in such a case, where stopping rules of the form

$$\text{stop at } B_t = x \quad \Leftrightarrow \quad x \geqslant f(t) \tag{8.26}$$

$$\text{stop at } B_t = x \quad \Leftrightarrow \quad t \geqslant \bar{f}(x) \tag{8.27}$$

are sought.

In a similar manner to the two-sided case, we shall assume that $C_1 \in C^{1,2}$, and apply Itô's formula to obtain

$$dC_{1t} = \frac{\partial C_1}{\partial x} dB_t + D_1(t, B_t)dt,$$

where

$$D_1(t,x) = \frac{1}{2}\frac{\partial^2 C_1}{\partial x^2} + \frac{\partial C_1}{\partial t}.$$

We can then use properties of D_1 in a similar way to those of D earlier, to produce equivalent results.

We also need the one-sided analogues of (P1) and (P2). Define

$$C_1^+(t,x) = C_1(t,x) - C_1(t - t_2 + t_1, x) \qquad (t_2 \leqslant t \leqslant T),$$

and let (P1b) be the following condition:

(P1b) For each $t_1 < t_2$ there exists $\epsilon = \epsilon(t_1, t_2)$ such that

$$\frac{\partial C_1^+}{\partial t} \geqslant 0 \qquad \text{for all } x,\ t > T - \epsilon > t_2\ ,$$

$$\frac{\partial^2 C_1^+(T-\epsilon, x)}{\partial x^2} \text{ is Lipschitz continuous and non-negative, for all } x.$$

Similarly, let

$$C_1^-(t,x) = C_1(t,x) - C_1(t, x - x_2 + x_1) \qquad (x \geqslant x_1),$$

and (P2b) be the condition

(P2b) For each $x_1 < x_2$ there exists $\epsilon = \epsilon(x_1, x_2)$ such that

$$\frac{\partial C_1^-}{\partial t} \geqslant 0 \qquad \text{for all } t > T - \epsilon,\ x \geqslant x_1,$$

$$\frac{\partial C_1^-(T-\epsilon, x)}{\partial x} \geqslant 0 \qquad \text{for all } x \geqslant x_1,$$

$$C_1^-(T-\epsilon, x) \text{ is convex for } x \geqslant x_1.$$

Theorem 25 *If $D_1(t,x)$ is non-decreasing in t and x,* (P1b) *and* (P2b) *hold, then there exists an optimal stopping bound f, which is non-increasing.*

Proof

We may use the same coupling techniques as in the proofs of Theorems 20 and 22 to establish the existence of rules of the form (8.26) and (8.27) above. Note that we have no problems, with the coupling, in the proof of a rule of the form (8.26), and need not make any further restrictions on x_1 and x_2 other than $x_1 < x_2$. We may, therefore, use the obvious coupling for all paths.

★

Let $g_1(t)$ satisfy $D_1(t, g_1(t)) = 0$, for all t. We have the following theorem about g_1.

Theorem 26 *If for large $N > 0$, some $\epsilon > 0$ we have $g_1(t) < -N$ for all $t > T - \epsilon$, and the value of $C(t, g_1(t))$ is bounded for such t, then $f(t) \geqslant g_1(t)$ for all t.*

Proof

The proof follows in virtually the same fashion as that of Theorem 21. Our condition involving N ensures a small probability of not exiting g_1 prior to $T - \epsilon$ replaces the condition on the way g approaches zero in the earlier theorem. The cost bound performs the same task.

★

Finally, the results on continuity follow in a similar manner.

Theorem 27 *If f is non-increasing, then f is right continuous.*

Theorem 28 *If f is non-increasing, g_1 is Lipschitz continuous, and for $x > g_1(t)$ we have: $C(t,x)$ is convex in x for each t, bounded on bounded (t,x) and*

$$C(t,x) - C(t-\epsilon, x) \geqslant k\epsilon^p,\ k = k(t,x) > 0, \text{ for some } p > 0,\ \epsilon \text{ suff. small,}$$

then f is left continuous.

The proofs of these results mirror those of Theorems 23 and 24.

8.2.7 Remarks

Cost Functions with Other Processes

As an alternative to Brownian motion, assume another diffusion is used, which satisfies

$$dX_t = \sigma(t, X_t)dB_t + \mu(t, X_t)dt.$$

If a time-scale change to a Brownian motion is not possible, then we can use the following approach. Letting $C_2(t,x)$ be a cost function of the form previously used, and $C_{2t} = C_2(t, X_t)$ be the corresponding cost diffusion process. Itô's formula yields

$$\begin{aligned} dC_{2t} &= \frac{\partial C_2}{\partial x}dX_t + \left(\frac{1}{2}\frac{\partial^2 C_2}{\partial x^2} + \frac{\partial C_2}{\partial t}\right)dt \\ &= \sigma(t, X_t)\frac{\partial C_2}{\partial x}dB_t + \left(\frac{1}{2}\frac{\partial^2 C_2}{\partial x^2} + \frac{\partial C_2}{\partial t} + \mu(t, X_t)\frac{\partial C_2}{\partial x}\right)dt. \end{aligned}$$

Thus, we can repeat the earlier results by considering

$$D_2(t,x) = \frac{1}{2}\frac{\partial^2 C_2}{\partial x^2} + \frac{\partial C_2}{\partial t} + \mu(t,x)\frac{\partial C_2}{\partial x},$$

instead of the usual $D(t,x) = \frac{1}{2}\frac{\partial^2 C}{\partial x^2} + \frac{\partial C}{\partial t}$, and assuming properties of this function.

No Time Horizon

In cases where $\lim_{t\to\infty} C(t,x) = \infty$, so that no time horizon naturally exists, there are two options worth considering. The first of these is the idea used in the worked example, of performing a time-scale change to produce a new, but related, problem which does have a time horizon. It should be noted, however, that properties, such as a non-increasing boundary, may not be true for both the original and transformed problems, although results such as existence, and lower bounds, will transform between frameworks.

The other alternative approach is the modification of the proofs of the Theorems, by allowing $T \to \infty$. This will require some restatements of assumptions and required properties, notably the alteration of "$t > T - \epsilon$ for some ϵ (small)" to "$t > T$ for some T (large)". One technique for this would be to artificially add a time horizon T, by altering the cost function, to $C^\star$ say, and letting $T \to \infty$, so that $C^\star \to C$.

Numerical Approaches

Direct attempts to numerically calculate the optimal stopping bound include the method explained in Bather, Chernoff and Petkau (1989). This method approximates the process by a random walk, which is run on an appropriate lattice.

Then, working backwards from a time horizon, when stopping is optimal, it can be determined, at each point of the lattice, whether it is optimal to stop or continue, and an approximate boundary can be placed to separate the two regions. Thus, if the lattice is made fine enough, the accuracy of this boundary to optimality can be made arbitrarily small. Clearly, for finer lattices, the computational effort becomes larger, especially as the time interval between each step becomes smaller, and more points are required for determining the boundary over a set time interval prior to the time horizon. Consequently, this method is probably best suited to finding the asymptotic behaviour of the boundary, close to the time horizon.

One method, which can be employed to help determine the form of the optimal stopping boundary, is numerical optimisation. Although this method cannot guarantee to find the optimal boundary, it allows us to optimise over classes of boundaries. We look at boundaries of the form $f_k(t) = kb(t)$, for some fixed function b, and seek to optimise over the choice of k. In order to accomplish this, we need to calculate

$$\mathsf{E}_{\tau_k}[C(\tau_k, f_k(\tau_k)) \,|X_0 = 0],$$

which invariably requires the use of exit distribution approximations. However, if it is assumed that the accuracy of the various distribution approximations is similar, for values of k close together, we might still be able to optimise over k. The choice of the function b is largely dependent on the problem in question. If we have a one-sided stopping boundary, we may consider classes of b so that f_k is linear. In such a case, the Bachelier-Lévy formula provides the exact exit density, and then our optimisation over this class of boundaries would be precise. Perhaps a more suitable choice of function b is that of the known lower bound to the optimal stopping boundary.

In the worked example, we search for the optimal stopping boundary of the form $f_k(t) = k(T-t)^2$, recalling the lower bound to the optimal stopping boundary is $c(T-t)^2$. To produce an approximate first exit density, the tangent approximation (Strassen (1967)) is completely unsuitable, as we have a two-sided boundary, which is convex decreasing in the upper half plane. Consequently, we shall use the UDHRR method of section 2.6.4. Using this approximation, for a variety of choices of c and T, the optimal choice of k was found to be as follows:

c \ T	4	2	1	0.5
2	3.59	2.79	2.27	1.64
4	6.88	6.67	5.48	4.19
8	15.95	13.36	12.00	9.78

Analysing these results, we notice that the optimal value of k increases in T and c. Notice that in the case $c = 2$, $T = 0.5$, the optimal choice is $k = 1.64$. That is, the optimal curve of the form $k(T-t)^2$ lies below the known lower bound, and is contained within the supermartingale region of the cost diffusion. The initial response to this would be that the error is caused by an inaccuracy in the exit distribution approximation.

We could verify this by simulation. It should be noted, however, that simulation is very computer expensive, particularly for optimisation. To obtain simulated results, to the same accuracy, takes about fifty times as long as it takes using the UDHRR approximation method. This is generally not practical when many examples need to be considered.

To save some time, three pairs of c and T were investigated, and a selection of values of k were chosen, in order that an approximate value for the optimal choice could be found. The same level of accuracy as before was not sought, and the results from simulating 200000 paths, in each case, are as follows:

Case $c = 1, T = 1$ k	Expected cost
0.6	0.904245
0.7	0.898620
0.75	0.897035
0.8	0.896590
0.85	0.896655
0.9	0.897565
1.0	0.901430

Case $c = 2, T = 2$ k	Expected cost
1.7	2.219855
1.75	2.217765
1.8	2.214525
1.85	2.214510
1.9	2.215520
1.95	2.216130

Case $c = 2, T = 0.5$ k	Expected cost
1.4	0.938750
1.5	0.938190
1.6	0.938150
1.7	0.938670
1.8	0.940220
1.9	0.942255
2.0	0.944700

One noticeable feature in all three cases is that the optimal choice of k, in each case, is less than the value of c, which corresponds to the theoretical lower bound to the optimal stopping boundary. Therefore, these optimal values of k cannot be correct and there are two possible explanations. Either insufficient simulations have been produced, leading to statistical errors, or a fundamental error is incurred when using the simulation approach.

We consider the standard errors on the expected costs found by simulation. In each case, the magnitude of the standard error is approximately the same for each value of k, and the approximate values of the standard errors are given below:

Case	Standard Error
$c = 1, T = 1$	± 0.015
$c = 2, T = 2$	± 0.1
$c = 2, T = 0.5$	± 0.01

In all cases, the standard errors are sufficiently large for us to be unable to conclude that the optimal choice of k is definitely less than c. The sizes of the errors would suggest about 20 million paths need to be simulated. This is completely impractical.

There are also systematic errors in the simulation approach. Consider the following. Let

$$A(k,t) = C(t, f_k(t)) = \left(c + \frac{kt}{T}\right)(T-t),$$

so that

$$\frac{\partial A}{\partial t} = (k-c) - \frac{2kt}{T} \text{ and } \frac{\partial^2 A}{\partial k \partial t} = 1 - \frac{2t}{T}.$$

So, for $t < T/2$, $\frac{\partial A}{\partial t}$ is more negative for smaller values of k. Consequently, the inaccuracy due to stopping too late is accentuated for smaller k. Because a discrete time approximation to Brownian motion is simulated, the stopping times are slightly greater than they ought to be. Combining these two ideas, the expected stopping costs for these examples are too small, and this error is increased as k gets smaller. The exit distributions for the curves with $k = 1$, $T = 1$ and $k = 2$, $T = 0.5$ have been found by simulation (see Fig 8.1). In these cases, most exits occur prior to $T/2$. It is reasonable to assume that this is true for all choices of (k, T) investigated in the cases $c = 1$, $T = 1$ and $c = 2$, $T = 0.5$. This possibly contributes to the phenomena seen, of the optimal choice of $f_k(t)$ being inside the supermartingale region of the cost diffusion.

It should also be noted that increasing the gradient of the boundary (or equivalently reducing k, in this case), increases the probability of the Brownian motion path exiting the boundary between consecutive observation times, but returning prior to observation. In such a scenario, the exit time is not recorded until at least one time interval later than its actual occurrence. This effect will make the

Fig 8.1a - Distribution Function (k=1,T=1)

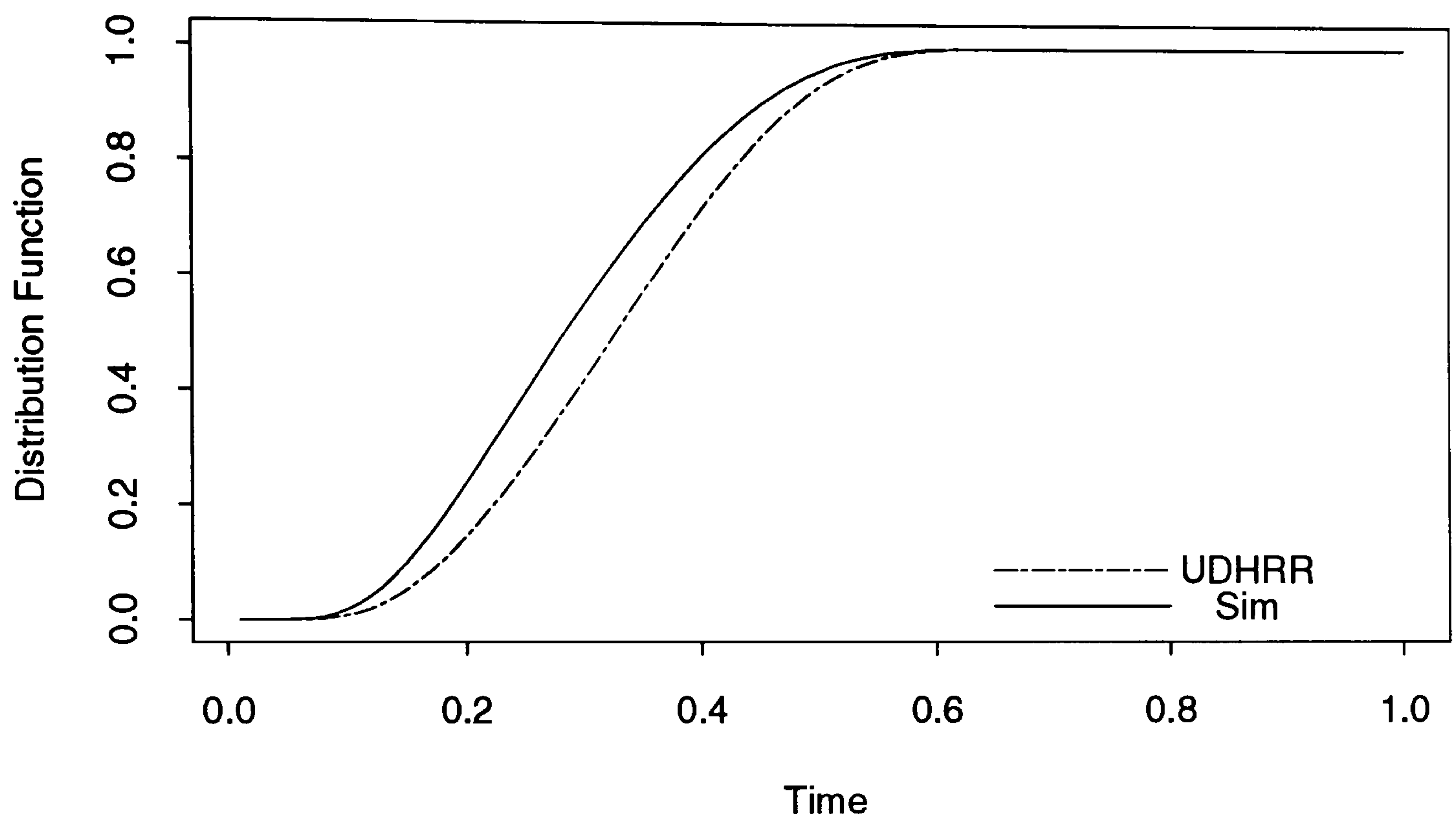
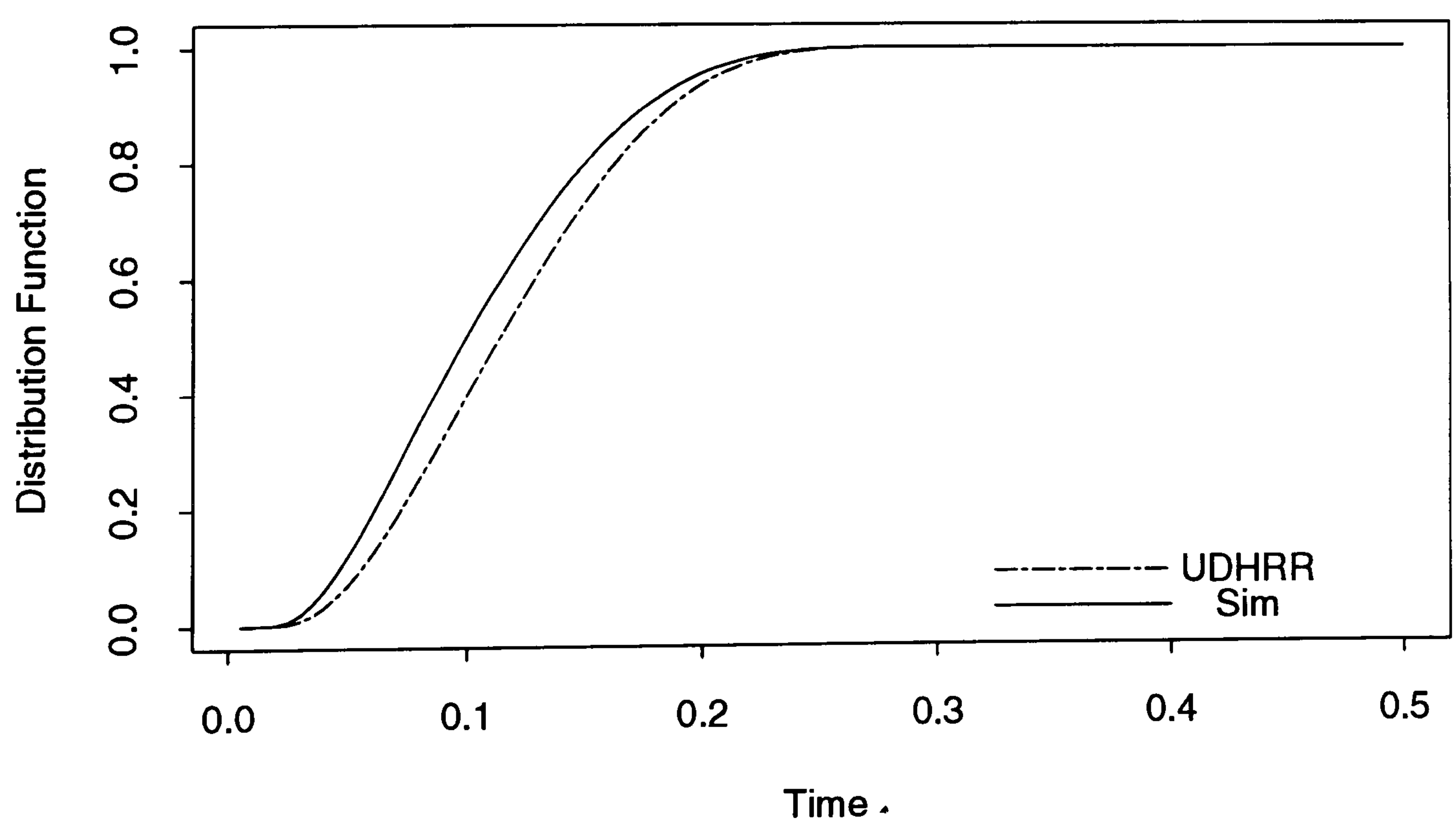


Fig 8.1b - Distribution Function (k=2,T=1/2)



expected stopping cost smaller than it ought to be in all cases, more notably for smaller values of k. This will, to some degree, negate the increasing error for larger values of k, when exits occur after $t = T/2$.

8.3 Approximations to Optimal Expected Stopping Costs

8.3.1 Introduction

In optimal stopping problems, the objective is to find a boundary f to minimise

$$\mathsf{E}_\tau[C(\tau, f(\tau)) \,|X_0 = x_0],$$

where C is the cost function and τ is the first exit time across the boundary. So far, we have concentrated on the practical side of these problems, finding the optimal stopping boundary, so that the objective is achieved. However, we often also need to know the value of the optimal stopping cost.

In many problems, the minimal expected cost cannot be calculated precisely, since either the optimal stopping boundary or the first exit distribution are not known exactly, but as approximations. In these cases, it is necessary to find an error bound on the calculated minimal expected cost, $e(\epsilon)$ say, where ϵ is the error in the approximated term. It is desirable to have $e(\epsilon) \to 0$ as $\epsilon \downarrow 0$.

When the source of the error is in the exit distribution approximation, we place no restriction on the process for which the results hold. However, when the error is in the boundary itself, we require some assumptions on the drift and diffusion coefficients of the process. If

$$dX_t = \sigma(t, X_t)dB_t + \mu(t, X_t)dt, \tag{8.28}$$

we insist that there exist α and ν such that

$$\left.\begin{array}{lr} 0 < \alpha < \sigma(t,x) & \text{for all } t,x \\ -\infty < \nu < \mu(t,x) & \text{for all } t,x. \end{array}\right\} \tag{8.29}$$

Property (8.29) is usually stated as the infinitesimal generator is strongly elliptic, see for example Friedman (1975).

In many problems, a stopping rule can be found, such that the expected stopping cost is within ϵ of the optimal stopping cost. Such a rule is said to be $\epsilon-$optimal, (see for example Krylov (1980)). Krylov proves the existence of such a stopping rule in a wide class of problems. We use a similar concept when using approximate boundaries, as this creates an approximation to the optimal stopping rule. When using density approximations, though, we are using the optimal stopping rule, but only approximating the payoff. In the two subsections to follow, we shall give a few results of this type. Note that many other results are possible, but the forms we have chosen relate most closely to the situation encountered in the worked example earlier in this chapter.

8.3.2 Boundary Approximations

Notation

We shall adopt the following notation:

$C(t,x)$	-	The cost incurred when stopping at $X_t = x$.
$f(t),\ g(t)$	-	Stopping boundaries .
τ_λ	-	$\tau_\lambda = \inf_{t>0}\{t : X_t \geqslant \lambda(t)\}$.
$\overline{T},\ \overline{x}$	-	$\sup_{(t,x)}\left\lvert\frac{\partial C}{\partial t}\right\rvert$, $\sup_{(t,x)}\left\lvert\frac{\partial C}{\partial x}\right\rvert$.
$\overline{C}_\lambda$	-	$\sup_t \lvert C(t,\lambda(t))\rvert$ for boundary function λ .
J_λ^V	-	Expected cost when stopping at boundary $\lambda(t)$, with cost $V(t,x)$.

The first result is in the infinite time horizon problem, when we have two boundaries for which we can calculate the expected payoff. No assumption is made that either of these boundaries is the optimal stopping boundary, and they could represent an upper and lower boundary to the optimal stopping boundary. We make the following assumptions:

(A1) $f(t) \leqslant g(t)$ for all t, $\sup_t(g(t) - f(t)) \leqslant \epsilon$ for some $\epsilon > 0$.
(A2) $f,\ g$ are uniformly continuous functions, with $|f(t) - f(t+\gamma)| \leqslant k\gamma$, $|g(t) - g(t+\gamma)| \leqslant k\gamma$ for all t and all $\gamma > 0$.
(A3) $C(t,x)$ is bounded and has positive, bounded first partial derivatives in t and x.

We then arrive at,

Theorem 29 *If* (A1)–(A3) *hold and* $\nu \geqslant 0$ (ν *defined in* (8.29)), *then*

$$|J_g^C - J_f^C| \leqslant (\overline{C}_g + \overline{C}_f)\frac{k+1}{\sqrt{2\pi\alpha}}\sqrt{\epsilon} + (\overline{T} + (k+1)\overline{x})\epsilon + O(\epsilon^{3/2}).$$

Proof

In addition to X satisfying (8.28), we require the three additional diffusions:

$$\begin{aligned} Y \text{ satisfying } dY_t &= \nu dt + \sigma(t, Y_t)dB_t \\ Y^2 \text{ satisfying } dY_t^2 &= \sigma(t, Y_t^2)dB_t \\ Z \text{ satisfying } dZ_t &= \alpha dB_t, \end{aligned}$$

where ν and α are as in (8.29). Since we are going to use several diffusions, we shall use superscripts to distinguish the hitting times of each process. For example,

$$\tau_g^Y = \inf_{t>0}\{t : Y_t \geqslant g(t)\}.$$

Consider a path for which $\tau_f^X = t$. By (A1), we must have $\tau_g^X \geqslant t$. We are interested in those paths which satisfy $\tau_f^X = t$ and $\tau_g^X < t+\delta$, for some appropriately selected δ. Denote by $p = \mathsf{P}[\tau_g^X < t+\delta \,|\tau_f^X = t]$, $\bar{g} = \max_{s\in(t,t+\delta)} g(s)$ and let $h(s) = \bar{g} - \nu(s-t)$. Clearly we have

$$p \geqslant \mathsf{P}[\tau_{\bar{g}}^X < t+\delta \,|\tau_f^X = t],$$

since prior to $t+\delta$, $\bar{g} \geqslant g(s)$. Now note that X always has greater drift than Y, which yields

$$\mathsf{P}[\tau_{\bar{g}}^X < t+\delta \,|\tau_f^X = t] \geqslant \mathsf{P}[\tau_{\bar{g}}^Y < t+\delta \,|\tau_f^Y = t].$$

Allowing for the constant drift difference between Y and Y^2,

$$\mathsf{P}[\tau_{\bar{g}}^Y < t+\delta \,|\tau_f^Y = t] = \mathsf{P}[\tau_h^{Y^2} < t+\delta \,|\tau_f^{Y^2} = t].$$

Note next that,

$$\mathsf{P}[\tau_h^{Y^2} < t+\delta \,|\tau_f^{Y^2} = t] \geqslant \mathsf{P}[\tau_{\bar{g}}^{Y^2} < t+\delta \,|\tau_f^{Y^2} = t], \tag{8.30}$$

as $h(s) \leqslant \bar{g}$ for $s \in (t, t+\delta)$. Finally, consider the time change

$$\beta(t) = \int_0^t \frac{\sigma^2(s, Y_s^2)}{\alpha^2} ds,$$

which provides

$$\sigma(t, Y_t^2)dB_t = \alpha dB_{\beta(t)},$$

so that $dY^2_{\beta^{-1}(t)} = dZ_t$. Hence, noting we are looking at exiting a constant boundary,

$$\begin{aligned}\mathsf{P}[\tau_{\bar{g}}^Z < t+\delta \,|\tau_f^Z = t] &= \mathsf{P}[\tau_{\bar{g}}^{Y^2} < \beta^{-1}(t)+\delta \,|\tau_f^{Y^2} = t] \\ &\leqslant \mathsf{P}[\tau_{\bar{g}}^{Y^2} < t+\delta \,|\tau_f^{Y^2} = t],\end{aligned}$$

since $\beta^{-1}(t) \leqslant t$. Combining these, and using the Bachelier-Lévy formula, we deduce that

$$p \geqslant 2\left(1 - \Phi\left(\frac{k\delta+\epsilon}{\sqrt{\alpha\delta}}\right)\right)$$

as $\bar{g} - f \leqslant k\delta + \epsilon$.

If we select $\delta = \epsilon$, we have $p \geqslant 2\left(1 - \Phi\left(\frac{(k+1)}{\sqrt{\alpha}}\sqrt{\epsilon}\right)\right)$. Also for paths with $\tau_g \leqslant \tau_f + \epsilon$, we have $g(\tau_g) \leqslant f(\tau_f) + (k+1)\epsilon$. Using (A3), we substitute this information into the cost function to obtain

$$\begin{aligned} C(\tau_g, g(\tau_g))I(\tau_g \leqslant \tau_f + \epsilon) &\leqslant C(\tau_f + \epsilon, f(\tau_f) + (k+1)\epsilon) \\ &\leqslant C(\tau_f, f(\tau_f)) + \overline{T}\epsilon + (k+1)\overline{x}\epsilon. \end{aligned}$$

Denoting by P_{τ_λ} the distribution of the hitting time of $\lambda(t)$, we obtain

$$\begin{aligned} |J_g^C - J_f^C| &= \left|\int\int [pC(\tau_g, g(\tau_g))I(\tau_g \leqslant \tau_f + \epsilon)\right. \\ &+ \left.(1-p)C(\tau_g, g(\tau_g))I(\tau_g > \tau_f + \epsilon) - C(\tau_f, f(\tau_f))]dP_{\tau_f}dP_{\tau_g}\right| \\ &= \left|\int\int p[C(\tau_g, g(\tau_g))I(\tau_g \leqslant \tau_f + \epsilon) - C(\tau_f, f(\tau_f))]dP_{\tau_f}dP_{\tau_g}\right. \\ &+ \left.(1-p)\int\int [C(\tau_g, g(\tau_g))I(\tau_g > \tau_f + \epsilon) - C(\tau_f, f(\tau_f))]dP_{\tau_f}dP_{\tau_g}\right| \\ &\leqslant (\overline{T} + (k+1)\overline{x})p\epsilon + (1-p)(\overline{C}_f + \overline{C}_g) \\ &= (\overline{T} + (k+1)\overline{x})\epsilon + (1-p)(\overline{C}_f + \overline{C}_g - (\overline{T} + (k+1)\overline{x})\epsilon) \\ &\leqslant (\overline{T} + (k+1)\overline{x})\epsilon \\ &+ \left(2\Phi\left(\frac{k+1}{\sqrt{\alpha}}\sqrt{\epsilon}\right) - 1\right)(\overline{C}_f + \overline{C}_g - (\overline{T} + (k+1)\overline{x})\epsilon). \end{aligned}$$

Noting that $2\Phi\left(\frac{k+1}{\sqrt{\alpha}}\sqrt{\epsilon}\right) - 1 = \frac{k+1}{\sqrt{\alpha}}\sqrt{\epsilon}\phi(0) + O(\epsilon^{3/2})$, we obtain our result.

★

Note that, as $C \geqslant 0$, we may replace $\overline{C}_g + \overline{C}_f$ by twice $\max(C(t,x))$, which exists by assumption (A3).

We may produce a similar result for $\nu < 0$.

Theorem 30 *If* (A1)-(A3) *hold and* $\nu < 0$ (ν *defined in* (8.29)), *then*

$$|J_g^C - J_f^C| \leqslant (\overline{C}_g + \overline{C}_f)\frac{k - \nu + 1}{\sqrt{2\pi\alpha}}\sqrt{\epsilon} + (\overline{T} + (k+1)\overline{x})\epsilon + O(\epsilon^{3/2}).$$

Proof

The proof of this result follows in a similar manner to Theorem 29. The only change is in the construction of a lower bound for p, when we produce $h(s) \leqslant \bar{g} - \nu\delta$, which alters (8.30) to

$$\mathsf{P}[\tau_h^{Y^2} < t + \delta \,|\tau_f^{Y^2} = t] \geqslant \mathsf{P}[\tau_{\bar{g}-\nu\delta}^{Y^2} < t + \delta \,|\tau_f^{Y^2} = t].$$

We continue the proof in the same manner as before, and obtain the bound

$$p \geqslant 2\left(1 - \Phi\left(\frac{(k-\nu)\delta + \epsilon}{\sqrt{\alpha\delta}}\right)\right).$$

The remainder of the proof follows as before.

★

Remark

Theorem 29 and Theorem 30 are the same if we use $\nu_2 = \min(0, \nu)$ in the statements and proofs of the results. We shall, therefore, use this notation in the following theorem.

In many cases, there exists a time horizon, T, by which time the process must have stopped. The result of this, in the two-sided case, is that all sensible stopping boundaries (not giving infinite expected stopping cost), are forced to have $f(T) = 0$. Consequently, it may be possible to have the following assumption to take this into account.

(A4) $f(t) \leqslant g(t)$, f, g non-increasing , $\sup_t(g(t) - f(t)) \leqslant \epsilon(T - t)$.

We then have the following theorem.

Theorem 31 *If* (A3) *and* (A4) *hold, then*

$$|J_g^C - J_f^C| \leqslant \frac{T - \nu_2}{\sqrt{2\pi\alpha}}(\overline{C}_g + \overline{C}_f)\sqrt{\epsilon} + (\overline{T} + T\overline{x})\epsilon + O(\epsilon^{3/2}).$$

Proof

This proof is similar to that of Theorem 29. In order to establish a lower bound on p, we use the same technique, of comparing the stopping times associated with the other diffusions, to yield

$$p \geqslant 2\left(1 - \Phi\left(\frac{-\nu_2\delta + T\epsilon}{\sqrt{\alpha\delta}}\right)\right),$$

as the non-increasing boundaries make the maximum value of

$$\overline{g} - \nu_2\delta - f(t) \leqslant \epsilon T - \nu_2\delta.$$

We shall again use $\delta = \epsilon$. The non-increasing nature of the boundaries lead to the statement

$$\begin{aligned} C(\tau_g, g(\tau_g))I(\tau_g \leqslant \tau_f + \epsilon) &\leqslant C(\tau_f, f(\tau_f)) + \epsilon(\overline{T} + (T - \tau_f)\overline{x}) \\ &\leqslant C(\tau_f, f(\tau_f)) + \epsilon(\overline{T} + T\overline{x}). \end{aligned}$$

The proof proceeds as before.

★

Remark

We will not look at any numerical examples of these results because of the requirement of the exact exit density across the two different functional boundaries. As these are unavailable in the majority of cases, we do not have any natural illustration to take from our worked example. This will not be the case for the density inaccuracies described in the following section.

8.3.3 Density Inaccuracies

We shall now look at the cases when the inaccuracy is as a result of exit distribution approximations, rather than boundary approximations. We shall assume that we know the optimal stopping boundary, although this is not strictly necessary. For, if the optimal stopping boundary is also an approximation, this will merely lead to a different error in the exit distribution approximation. No specific properties of the boundary will be directly imposed, though properties of the cost function evaluated along these boundaries will be assumed. We shall use the following notational conventions:

$f(t)$	-	The optimal stopping boundary .
$p_1,\ P_1$	-	Exact density, distribution function of τ_f .
$p_2,\ P_2$	-	Approximate density, distribution function of τ_f.
J_1	-	Optimal expected payoff (associated with $p_1,\ P_1$) .
J_2	-	Approximate expected payoff (associated with $p_2,\ P_2$).

We start by looking at the effect of inaccuracy of exit time densities, and shall make the following assumptions:

(A5) $C(t,x)$ is such that $C(t,f(t))$ is integrable.
(A6) $p_1(t),\ p_2(t)$ satisfy $\max_t |p_1(t)-p_2(t)| < \epsilon$.

We can now prove,

Theorem 32 *If* (A5) *and* (A6) *hold, then* $|J_1 - J_2| < \epsilon \int_0^\infty C(t,f(t))dt$.

Proof

By definition of the expected payoff, we clearly have,

$$\begin{aligned} |J_1 - J_2| &= \left| \int_0^\infty C(t, f(t))(p_1(t) - p_2(t))dt \right| \\ &\leqslant \int_0^\infty C(t, f(t))|p_1(t) - p_2(t)|dt \\ &< \epsilon \int_0^\infty C(t, f(t))dt, \end{aligned} \tag{8.31}$$

which is the stated result.

★

The next theorem is similar to Theorem 32, except that it assumes that the density ratio remains close to one. So we need,

(A7) $p_1(t)$, $p_2(t)$ are such that $\left|\frac{p_1(t)}{p_2(t)} - 1\right| < \epsilon$.

Theorem 33 *If* (A5) *and* (A7) *hold, then* $\left|\frac{J_1}{J_2} - 1\right| < \epsilon$.

Proof

Not surprisingly, we can virtually copy the proof of Theorem 32. We need only note that

$$\left|\frac{p_1(t)}{p_2(t)} - 1\right| < \epsilon \Rightarrow |p_1(t) - p_2(t)| < |p_2(t)|\epsilon,$$

and so we just have to replace this in (8.31) to see

$$|J_1 - J_2| < \epsilon \int_0^\infty C(t, f(t))p_2(t)dt.$$

Hence we produce our result.

★

The final result involves distribution functions for the exit times across the optimal boundary. This requires slightly more restrictions on the cost function evaluated along the boundary. We now assume

(A8) $C(t, f(t))$ is bounded and differentiable.
(A9) $|P_1(t) - P_2(t)| < \epsilon$.
(A10) $C(t, f(t))$ is monotonic.

Theorem 34 *If* (A8)–(A10) *hold, then*

$$|J_1 - J_2| < 2\overline{C}_f\epsilon.$$

Proof

Again starting from the definition of expected cost, we have

$$J_i = \int_0^\infty C(t, f(t)) p_i(t) dt.$$

This can be integrated by parts:

$$\begin{aligned} J_i &= \int_0^\infty C(t, f(t)) p_i(t) dt \\ &= [C(t, f(t)) P_i(t)]_0^\infty - \int_0^\infty \frac{d}{dt}(C(t, f(t))) P_i(t) dt \\ &= C_\infty - \int_0^\infty \frac{d}{dt}(C(t, f(t))) P_i(t) dt, \end{aligned}$$

where $C_\infty = \lim_{t\to\infty} C(t, f(t))$ exists by (A8) and (A10).

Therefore,

$$\begin{aligned} |J_1 - J_2| &= \left| \int_0^\infty \frac{d}{dt}(C(t, f(t)))(P_2(t) - P_1(t)) dt \right| \\ &< \epsilon \left| \int_0^\infty \frac{d}{dt} C(t, f(t)) dt \right| \\ &= \epsilon |C_\infty - C(0, f(0))| \\ &\leqslant 2\overline{C}_f \epsilon, \end{aligned}$$

which is the required result.

★

Remarks

Note that assumption (A10) can be relaxed to $C(t, f(t))$ is piecewise monotonic. If this is the case, the integral will need to be broken up into the monotonic sections, and the result will be $|J_1 - J_2| < 2k\overline{C}_f \epsilon$, where k is the number of regions into which the piecewise monotonicity splits the time-scale up.

Finally, we remark that any combination of boundary approximation, and density approximation can be handled in the same case, by adding the corresponding upper bounds on the errors.

Numerical Example

Consider the stopping boundary $f(t) = \pm 2\left(\frac{1}{2} - t\right)^2$, with the associated cost function

$$C(t,x) = \frac{t|x|}{\frac{1}{2}\left(\frac{1}{2}-t\right)} + 2\left(\frac{1}{2} - t\right).$$

The example has been chosen with the worked example of the optimal stopping problem, discussed earlier, in mind and f is the lower bound to the optimal stopping boundary. For the purpose of this illustration, we will assume that f is actually the optimal stopping boundary. One consequence of this choice of f, is that the exact exit distribution to the curve is unknown, and so the exit distribution has been simulated. At this point, we should note that, in practice, we would simulate the expected stopping cost rather than the exit distribution from which bounds on the expected cost are found. However, the current approach is taken to illustrate Theorems 32–34. The approximate exit distribution has been found using the UDHRR method of section 2.6.4. For one-sided stopping boundaries, we could use the methods of Chapter 7 to find the necessary maximum values of the difference between exact and approximate distribution functions, and so the actual exact values are not strictly needed, making the theorems more justifiable.

We first check the required conditions on C for the theorems to hold. Note that we have a finite horizon problem, which requires alterations to some of the conditions and definitions to be made. Specifically, we replace

$$\int_0^\infty C(t,f(t))dt \text{ by } \int_0^{1/2} C(t,f(t))dt,$$

and write

$$J_i = \int_0^{1/2} C(t,f(t))p_i(t)dt.$$

The final necessary notational change is that C_∞ in the proof of Theorem 34 should be defined as $C_\infty = \lim_{t\uparrow 1/2} C(t,f(t))$. All properties, such as differentiability, are also assumed to hold on $[0, 0.5]$, rather than $\mathbb{R}^+$.

Checking (A5), (A8) and (A10), we see that

$$C(t,f(t)) = (4t+2)\left(\frac{1}{2} - t\right) = 1 - 4t^2,$$

which is bounded, monotone and differentiable on $[0, 0.5]$. Furthermore, we have

$$\int_0^{1/2} C(t,f(t))dt = \frac{1}{3}$$

and $\overline{C}_f = 1$. Thus, all of the conditions required for Theorems 32–34 hold, and we need only find the appropriate bounds on the distribution functions and densities. Comparing the simulated values with the approximations, we obtain the following:

$$\begin{aligned} |p_1(t) - p_2(t)| &< 1.842 \\ \left|\frac{p_1(t)}{p_2(t)} - 1\right| &< 0.982 \\ |P_1(t) - P_2(t)| &< 0.1032. \end{aligned}$$

The final information required is that the approximate expected stopping cost is $J_2 = 0.93396$.

We start by applying Theorem 32, although we would not expect a particularly accurate result, as in this case $\epsilon = 1.842$. Nevertheless, the theorem provides us with the equation

$$|J_1 - 0.93396| < \frac{1}{3}(1.842),$$

which rearranges to become

$$0.31996 < J_1 < 1.54796.$$

Similarly, we would not expect Theorem 33 to produce a tight bound on J_1, as we have $\epsilon = 0.982$. We obtain

$$\left|\frac{J_1}{0.93396} - 1\right| < 0.982,$$

or

$$0.017 < J_1 < 1.851,$$

which is extremely poor. Finally we apply Theorem 34, with $\epsilon = 0.1032$. Substituting the appropriate values into the theorem, results in the inequality

$$|J_1 - 0.93396| < 2 \times 1 \times 0.1032 = 0.2064,$$

which reduces to

$$0.72756 < J_1 < 1.14036.$$

Note that, since $\lim_{t \uparrow 0.5} C(t, f(t)) = 0$, the proof of Theorem 34 could be amended, and the factor 2 dropped. Thus, we could actually establish

$$0.83076 < J_1 < 1.03716.$$

Of the bounds found for J_1, the final one is the tightest, and would be used in applications. As a check, the actual expected stopping cost was also found by simulation. After 200000 simulated paths, the expected stopping cost was found to be

$$J_1 = 0.9447 \qquad (\text{standard error } \pm 0.01),$$

which is well inside all the bounds.

Despite the crudity of the calculated bounds, it should be stressed that the theorems were proved by worst case estimates, and in many cases, they can be tightened, just as Theorem 34 was in our example. The other important fact to note is that they can be produced extremely rapidly, especially in comparison to the time taken to simulate the actual value. The one drawback is the need of the exact distribution of the stopping time, although the methods of Chapter 7 suggest a way to find the necessary bounds between exact and approximate values, without actually finding the exact value. Thus, particularly in one-sided cases, the theorems suggested provide fast methods to establish reasonably tight bounds on the expected stopping cost.

Chapter 9

Concluding Remarks

We finish this work with some general comments about the results we have established. The main achievements of this thesis have been made in finding results about the first exit distribution of diffusion processes across functional boundaries. The most thorough exposition has been for the case of Brownian motion, because this process is the simplest, widely used diffusion. For Brownian motion, we can compute analytic bounds on the exit distribution across one-sided boundaries (Corollary 4, Chapter7). We have established the necessary theory to extend such results to other strongly stochastically monotone processes, although explicit forms of the bounds are not currently available.

Methods to find exact and approximate exit distributions have also been discussed. For one-sided boundaries, and the Brownian motion process, we have developed the HRT method (section 7.6), which can be shown to be superior to the tangent approximation, in the case of concave and convex boundaries. For other types of boundaries, the numerical examples we have investigated, suggest it is also more accurate. Thus, in the case of one-sided boundaries, the Brownian motion exit distribution is best approximated by the HRT method. For two-sided boundaries, the previous methodology was limited, and the UDHRR method developed, (section 2.6.4), appears to be accurate. The basic method, using only Brownian motion quasi-stationary distributions, is most applicable for boundaries which are approximately constant, though cancellation of errors makes it work well for slowly increasing boundaries. However, the use of time-scale changes, and exploitation of the same result for the Ornstein-Uhlenbeck process, enable us to produce a good approximation for Brownian motion exiting approximately square root boundaries. The UDHRR method is also applicable to a wide class of processes, for which a limiting distribution of $\lfloor X \,|\tau > t \rfloor$ is available, and may provide exit distribution approximations for one- or two-sided boundaries.

A section of the thesis was devoted to results on stochastic orderings. A number of previously known results have been collated, and grouped with newly established results, to produce a compendium of ordering results (Chapters 3-5). The main new results in this section (Chapter 6) enable us to determine the existence of strong stochastic ordering, by comparing the forms of the stochastic differential equations of the diffusions. This is particularly helpful in situations where the distributional properties of either process cannot be found explicitly, and so checking the appropriate definitions is not directly possible. As an example of the usage of strong stochastic ordering, the application to exit distribution theory is illustrated in Chapter 7, where analytic bounds on the exit distribution are found.

The optimal stopping problem (Chapter 8) is primarily included as an example of the application of boundary hitting time results. We have now established properties of the drift coefficient of the cost diffusion, which enable us to quickly conclude results about the optimal stopping boundary. The properties, of existence, monotonicity, continuity and upper and lower bounds, are found using a combination of boundary hitting results, and stochastic calculus.

More direct applications of boundary hitting time distributions are found in section 8.3. The error in the calculation of the optimal expected stopping cost is computed, when there are errors in either the optimal stopping boundary, or the exit distribution across it. These results also provide a natural application of the analytic bounds on the exact exit distribution.

Acknowledgements

Many thanks to Gareth Roberts for steering me in the right directions and suffering the many early drafts of this thesis. I am also grateful to the University of Cambridge for allowing me to complete this work in their Statistical Laboratory. Thanks also to the Science and Engineering Research Council for providing a grant, and to Ladbrokes plc for supplying a further source of income.

Chapter 10

Bibliography

ALDOUS (1989). Probability Approximations via the Poisson clumping heuristic. *Springer-Verlag, New York, London.*

BALDING (1988). Diffusion-Reaction in one-dimension. *J. Appl. Prob. 25, 733-743.*

BATHER (1983). Optimal stopping of Brownian motion: A comparison technique. *Recent Advances in Statistics, 19-48.*

BATHER (1988). Stopping rules and ordered families of distributions. *Seq. An. 7, 111-126.*

BATHER, CHERNOFF AND PETKAU (1989). The effect of truncation on a sequential test for the drift of Brownian motion. *Seq. An. 8, 169-190.*

BĚNES, SHEPP AND WITSENHAUSEN (1980). Some solvable stochastic control problems. *Stochastics 4, 39-83.*

BILLINGSLEY (1968). Convergence of Probability Measures. *Wiley, New York, Chichester, Brisbane, Toronto.*

BREIMAN (1967). First exit times from a square root boundary. *Proc. Fifth Berk. Symp. Math. Stat. Probab., vol II part II, 9-16.*

DALEY (1968). Stochastically monotone Markov chains. *Z. Wahr. Verw. Geb. 10, 305-317.*

DANIELS (1982). Sequential tests constructed from images. *Ann. Stats. 10, 394-400.*

DARLING AND ROBBINS (1968). Some further remarks on inequalities for partial sums. *Proc. Nat. Acad. Sci. U.S.A. 60, 1175-1182.*

DELLACHERIE AND MEYER (1978). Probabilities and Potential. *North-Holland, Paris, Hermann, Amsterdam, New York.*

DUNFORD AND SCHWARTZ (1963). Linear Operators II. Spectral Theory. *Interscience publishers, New York, London.*

DURBIN (1985). The first-passage density of a continuous Gaussian process to a general boundary. *J. Appl. Prob. 22, 99-122.*

DURBIN (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.*

ERDELYI (1953). Higher Transcendental Functions. *McGraw-Hill.*

FEREBEE (1983). An asymptotic expansion for one-sided Brownian exit densities. *Z. Wahr. Verw. Geb. 61, 309-326.*

FRIEDMAN (1975). Stochastic Differential Equations and Applications. *Acad. Press, New York.*

GIORNO, NOBILE AND RICCIARDI (1989). A symmetry-based constructive approach to probability densities for one dimensional diffusion processes. *J. Appl. Prob 27, 707-721.*

IKEDA AND WATANABE (1981). Stochastic Differential Equations and Diffusion Processes. *North-Holland, Amsterdam.*

JACKA AND ROBERTS (1987). Conditional diffusions: their infinitesimal generators and limit laws. *Warwick Univ., Dept. of Statistics, Research Report No. 127.*

JENNEN (1985). Second-order approximations for Brownian first exit distributions. *Ann. Probab. 13, 126-144.*

KARATZAS AND SHREVE (1988). Brownian Motion and Stochastic Calculus. *Springer-Verlag, New York, Heidelberg, Berlin, London.*

KARLIN (1968). Total Positivity, Vol. 1. *Stanford, Cal.*

KARLIN AND MCGREGOR (1959). Coincidence probabilities. *Proc. J. Maths. 9, 1141-1164.*

KARLIN AND TAYLOR (1975). A First Course in Stochastic Processes. *Academic Press, New York.*

KRYLOV (1980). Controlled Diffusion Processes. *Springer-Verlag, New York.*

LERCHE (1986). Boundary Crossing of Brownian Motion. *Lecture Notes in Statistics 40, Springer-Verlag, Berlin, Heidelberg, London, Paris, Tokyo.*

LÉVY (1965). Processus Stochastiques et Mouvement Brownien. *Gauthier-Villars, Paris.*

LEVY (1992). Stochastic dominance and expected utility: Survey and analysis. *Man. Sci. 38, 555-593.*

LIPTSER AND SHIRYAYEV (1977). Statistics of Random Processes I: General Theory. *Springer-Verlag, New York.*

NOBILE AND RICCIARDI (1984a). Growth with regulation in fluctuating environments. I. Alternative logistic-like diffusion models. *Biol. Cybern. 49, 179-188.*

NOBILE AND RICCIARDI (1984b). Growth with regulation in fluctuating environments. II. Intrinsic lower bounds to population size. *Biol. Cybern. 50, 285-299.*

NOBILE, RICCIARDI AND SACERDOTE (1985a). Exponential trends of Ornstein-Uhlenbeck first passage time densities. *J. Appl. Prob. 22, 360-369.*

NOBILE, RICCIARDI AND SACERDOTE (1985b). Exponential trends of first passage time densities for a class of diffusion processes with steady state distribution. *J. Appl. Prob. 22, 611-618.*

ØKSENDAL (1985). Stochastic Differential Equations. *Springer-Verlag, Berlin, Heidelberg, New York.*

POLLAK AND SIEGMUND (1986). Convergence of quasi-stationary distribution for stochastically monotone Markov processes. *J. Appl. Prob. 23, 215-220.*

RICCIARDI AND SATO (1988). First-passage time density and moments of the Ornstein-Uhlenbeck process. *J. Appl. Prob. 25, 43-57.*

ROBERTS (1991a). A Comparison theorem for conditioned Markov processes. *J. Appl. Prob. 28, 74-83.*

ROBERTS (1991b). Asymptotic approximations for Brownian motion boundary hitting times. *Ann. Prob. 19, 1689-1731.*

ROBERTS (1993). Some inequalities for one-dimensional conditioned diffusions and boundary hitting times. *Submitted to J. Appl. Prob.*

ROBERTS AND JACKA (1994). Weak convergence of conditioned Birth and Death processes. *To appear J. Appl. Prob.*

ROGERS AND WILLIAMS (1987). Diffusions, Markov Processes and Martingales. Vol. 2. *Wiley, New York.*

ROSS (1984). A First Course in Probability, 2^{nd} edition. *MacMillan, New York.*

SHAKED AND SHANTHIKUMAR (1988). Stochastic convexity and its applications. *Adv. Appl. Prob. 20, 427-446.*

SIEGMUND (1985). Sequential Analysis *Springer-Verlag, New York, Berlin, Heidelberg, Tokyo.*

STOYAN (1983). Comparison methods for queues and other stochastic models. *Wiley, Chichester.*

STRASSEN (1967). Almost sure behaviour of sums of independent random variables and martingales. *Proc. Fifth Berk. Symp. Math. Stat. Probab., vol II part I, 315-343.*

WILLIAMS (1979). Diffusions, Markov Processes and Martingales. Vol. 1: Foundations. *Wiley, Chichester, New York.*

WOODROOFE (1982). Nonlinear Renewal Theory in Sequential Analysis. *Society for Industrial and Applied Mathematics, Philadelphia.*