Statistical Analysis of Cell Motion

by

Edward Luke Ionides

B.A. (Cambridge University) 1994
M.A. (University of California, Berkeley) 1998

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Statistics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor David R. Brillinger, Chair
Professor David J. Aldous
Professor George F. Oster

Summer 2001

The dissertation of Edward Luke Ionides is approved:

Chair   Date

Date

Date

University of California, Berkeley

Summer 2001

Statistical Analysis of Cell Motion

Copyright 2001

by

Edward Luke Ionides

Abstract

Statistical Analysis of Cell Motion

by

Edward Luke Ionides

Doctor of Philosophy in Statistics

University of California, Berkeley

Professor David R. Brillinger, Chair

Certain biological experiments investigating cell motion result in time lapse video microscopy data which may be modeled using stochastic differential equations. These models suggest statistics for quantifying experimental results and testing relevant hypotheses, and carry implications for the qualitative behavior of cells and for underlying biophysical mechanisms. A state space model formulation is used to link models proposed for cell velocity to observed data. Sequential Monte Carlo methods enable parameter estimation and model assessment for a range of applicable models. One particular experimental situation, involving the effect of an electric field on cell behavior, is considered in detail.

There are several reasons why one might carry out parameter estimation by maximizing a smooth approximation to the likelihood in preference to the widely used method of maximum likelihood. If the likelihood is approximated by Monte Carlo simulation then a smoothed approximation may be all that can reasonably be obtained. If the likelihood function has many local maxima then a smooth approximation can lead to a more tractable and possibly more appropriate maximization problem. A theory for maximum smoothed likelihood estimation is developed using the framework of local asymptotic normality (LAN). This property of LAN is demonstrated for certain state space models whose state process is a diffusion process.

A complementary approach to direct observation of cell motion is to stain fixed cells to determine the spatial distribution throughout the cell of a molecule of interest. The data are digitized microscopy images. An algorithm is developed to quantify relevant features of data collected as part of the investigation on the effect of electric fields. Summary statistics and test statistics are proposed which are not model dependent, requiring only the symmetry of the experiment for their validity, but which can be justified in the context of a particular stochastic model for the staining process.

To my father

Here vigour failed the towering fantasy,
Yet the will rolled onward like a wheel,
In even motion, impelled by the love
That moves the sun in heaven and the stars.

Dante Alighieri

Acknowledgements

Contents

Chapter 1

Introduction

In this thesis we describe some problems in cell biology, some experimental measurements made to investigate these problems, and some statistical tools suitable for the analysis of the resulting data. The synergy from studying these two topics in some detail in one thesis is that questions of substantive scientific interest serve as a whetting stone to sharpen the statistical tools which leads in turn to improved data analyses. In particular, Chapters 2 and 4 develop statistical analyses of data collected on moving cells and fixed cells respectively. In these chapters, relevant existing statistical techniques are assembled to address particular scientific questions. Chapter 3 investigates a gap in the existing theory for statistical inference for some of the models proposed in Chapter 2. The resulting theory complements Chapter 2, while also making a contribution to many other data analyses for which similar statistical methods are applicable.

The central approach to Chapter 2 is to use stochastic differential equations to define parametric families of models whose parameters have physical interpretation. This approach was pioneered in (Kendall, 1974; Levin, 1986; Brillinger, 1997; Brillinger and Stewart, 1998). In the situation studied here, we are led toward a so-called state space model by the consideration that we model cell velocity and observe cell location at discrete times with some measurement error. A convenient way to find parameter estimates and their uncertainties in state space models turns out to be approximating the likelihood function using the method of sequential Monte

Carlo simulation developed by Kitagawa (1996). It is then tempting to estimate the likelihood function by applying a statistical smoothing technique to the Monte Carlo estimates of the likelihood at different parameter values. Chapter 3 develops some theory which suggests that using a smooth approximation to the likelihood function is a reasonable thing to do. It further demonstrates situations where it is appropriate to smooth the likelihood function deliberately. This framework is then used to find asymptotic properties of estimators for certain state space models, when the underlying state process is a diffusion.

In Chapter 4 an algorithm is constructed, using standard tools of image analysis and statistics, to quantify experimental data consisting of images of fixed cells stained for a particular protein. Summary statistics and tests of relevant hypotheses are presented, and then found to arise naturally from a statistical model. The methods are next demonstrated on some available data. The model can be seen from the data to be reasonably appropriate.

The remaining sections of this chapter give some background material to lay the foundations for the subsequent chapters. Section 1.1 gives an overview to cell motion, and Sections 1.2, 1.3 and 1.4 outline some results for stochastic differential equations, state space models and stochastic models for cell shape respectively.

1.1   Some Background to Cell Motion

Active migration of blood and tissue cells is essential to a number of physiological processes such as inflammation, wound healing, embryogenesis and tumor cell metastasis (Bray, 1992). It also plays an important role in the functioning of many bioartificial tissues and organs (Langer and Vacanti, 1999), such as skin equivalents (Parenteau, 1999) and cartilage repair (Mooney and Mikos, 1999). Modern techniques in microscopy, genetics and pharmacology helped to make some progress in unraveling the complex biophysical processes involved in cell motion (Maheshwari and Lauffenburger, 1998). Although different cell types show diverse methods of locomotion, there are general principles that are widely applicable for cells moving along a substrate. First a cell extends a protrusion by actin filament polymerization

(Mogilner and Oster, 1996), which then attaches to the substrate using integrin adhesion receptors (Huttenlocher et al., 1995). A contractile force is next generated which moves the cell body. Finally the cell must attach from the substrate at its trailing end.

Various mathematical models incorporating the above principles of cell motion have been proposed. The most ambitious of them attempt to represent all the physical and chemical processes involved in the motion of an entire cell (Tranquillo and Alt, 1996; Dickinson and Tranquillo, 1993; Dembo, 1989). Others concentrate on a specific process such as extension of a protrusion (Mogilner and Oster, 1996) or receptor dynamics (Lauffenburger and Linderman, 1993). The primary purpose of these biophysical models is to demonstrate that the proposed mechanisms can in fact produce the forces and behaviors observed experimentally.

Another approach to modeling cell motion is phenomenological in nature. The so-called correlated random walks of Alt (1980), Dunn and Brown (1987) and Shenderov and Sheetz (1997) have been proposed to describe observations of isolated cells locomoting on a substrate. For applications, the behavior of cell populations may be of more direct interest, and here diffusion approximations to population behavior are widely used, for example in Ford et al. (1991). The theoretical relationships between single cell models and population models are studied in Alt (1980), Dickinson and Tranquillo (1995), Ford and Lauffenburger (1991). An empirical comparison between single cell and cell population models is given in Farrell et al. (1990). Phenomenological models are used for quantifying experimentally observed cell behavior, and do not require justification in terms of a proposed mechanism. Nevertheless the line dividing biophysical from phenomenological models is in fact only a difference in complexity, and can become blurred as even the simpler phenomenological models can have implications concerning underlying biophysical mechanisms (Dunn and Brown, 1987).

The questions of scientific and engineering interest about cell motion can be broadly summarized into the following: What biophysical processes are involved in cell motion? How can the speed and direction of the motion be modeled? One approach toward answering these questions is to collect temporal sequences of images of

moving cells. This is the data type that will be considered in later chapters. Various experimental protocols for studying cell motion are discussed in Alt, Deutsch and Dunn (1997) and Alt and Hoffmann (1990). Cells may be observed moving separately on a microscope slide, or in a three-dimensional collagen gel. The cells under investigation may be connected to form a tissue, with a stain used to identify and track groups of cells.

Computer-assisted microscopy can be used to build three-dimensional images of moving cells (Murray et al., 1992; Wessels et al., 1994). Traction forces can be measured by observing the wrinkles produced by a cell moving on an elastic substrate (Oliver et al., 1995). The particular experimental procedure that will recur repeatedly as an example in this thesis is described in more detail in Example 1 below. It is an investigation of the reaction to a stimulus of single cells locomoting on a microscope slide using time-lapse observations. Analysis of this relatively simple cell motion experiment can be taken as motivation for the theory developed in Chapters 2, 3 and 4.

Example 1. *Investigation of Galvanotaxis.* Human keratinocytes (skin cells) migrate toward the negative pole in direct current electric fields of physiological strength. This phenomenon is termed galvanotaxis and is of particular interest in wound healing (Nishimura et al., 1996). It has also recently become a tool for more general study of directional cell motility, as in Fang et al. (1999). One of the challenges in practice of investigating directional response to a stimulus is to set up experimentally the controlled, uniform gradients required for clear and reproducible results. However, such gradients are relatively easy to attain for DC electric fields, making galvanotaxis a convenient model system for investigating basic aspects of directional cell motility.

The data analyzed in Section 2.2 were collected by Dr. Kathy Fang at University of California, Davis, to investigate the effect of calcium ion (Ca^{2+}) concentration on galvanotaxis. The experimental method is similar to that used in Fang et al. (1999) to demonstrate the role of the epidermal growth factor receptor (EGFR) in galvanotaxis. Specifically, normal human keratinocytes from neonatal foreskin epidermis were cultured and plated onto a glass coverslip coated with extra-cellular matrix (col-

lagen). The treated coverslip was placed in a galvanotaxis chamber as described in Nishimura et al. (1996). Cells were then observed using phase contrast or differential interference contrast optics with video images being digitally captured. The images were typically captured at intervals of ten minutes, during a one hour observation period, resulting in seven images per experiment. Further details of the experimental method are given in Fang et al. (1999).

1.2   Diffusion Processes and Stochastic Calculus

A stochastic process $\{X(t), t \in \mathbb{R}\}$ is said to be an m-dimensional *diffusion process* if it has continuous sample paths in $\mathbb{R}^m$ and possesses the strong Markov property, that for a stopping time τ, $\{X(t), t \geq \tau\}$ is conditionally independent of $\{X(t), t \leq \tau\}$ given X_τ. The treatment here draws on results in Oksendal (1998) and Stoock and Varadhan (1979), and is based on the approach in Karlin and Taylor (1981) which however deals only with diffusion processes with sample paths in $\mathbb{R}$.

Definition. The *infinitesimal mean*, or *drift*, of a diffusion process $X(t)$ is

$$\mu(x, t) = \lim_{h \downarrow 0} (1/h) \mathbb{E}[X(t + h) - X(t) \mid X(t) = x],$$

and the *infinitesimal variance* is

$$\Gamma(x, t) = \lim_{h \downarrow 0} (1/h) \mathbb{E}[(X(t + h) - X(t))(X(t + h) - X(t))^T \mid X(t) = x],$$

when these limits exist. Here the superscript "T" refers to the operation of matrix transposition.

A remarkable property is that the law of a diffusion process is specified by $\mu(x, t)$ and $\Gamma(x, t)$ provided the limits exist, excepting behavior at any boundary and assuming a higher moment condition that for some $r > 2$ and all x and t

$$\lim_{h \to 0} (1/h) \mathbb{E}[|X(t + h) - X(t)|^r \mid X(t) = x] = 0.$$

Examples are Brownian motion in $\mathbb{R}^1$, with $\mu(x, t) = 0$ and $\Gamma(x, t) = \sigma^2$, and the Ornstein–Uhlenbeck process, with $\mu(x, t) = -\alpha x$ and $\Gamma(x, t) = \sigma^2$.

We will be concerned with diffusion processes as the solutions to certain stochastic differential equations (SDEs) of the form

$$dX(t) = b(X(t), t)dt + \gamma(X(t), t)dW(t), \qquad (1.1)$$

where $b(X(t), t)$ is a vector in $\mathbb{R}^m$, $\gamma(X(t), t)$ is an $m \times m$ matrix and $W(t)$ is a standard Brownian motion in $\mathbb{R}^m$, i.e., a diffusion process with infinitesimal drift of zero and infinitesimal variance I. There is more than one way to give a formal meaning to the infinitesimal equation (1.1), leading to different possible solutions. The Itô solution is written

$$X(t) = X(0) + \int_0^t b(X(s), s)ds + \int_0^t \gamma(X(s), s)dW(s)$$

where integration is carried out using the so-called Itô integral. The Itô integral is defined, for a stochastic process $\{Y(t), t > 0\}$, when the limit exists, as

$$\int_0^t Y(s)dW(s) = \lim_N \sum_{n=0}^{N-1} Y(t_n)(W(t_{n+1}) - W(t_n)).$$

The limit is taken over all sequences of partitions whose largest intervals tend to zero, in other words, over collections $\{t_i, i = 1, 2, \ldots, N\}$ with $0 = t_0 < t_1 < \cdots < t_N = t$ and $\sup_n |t_{n+1} - t_n| \to 0$. An alternative interpretation, the Stratonovich solution, is written

$$X(t) = X(0) + \int_0^t b(X(s), s)ds + S-\int_0^t \gamma(X(s), s)dW(s)$$

where integration is carried out using the Stratonovich integral defined as

$$S-\int_0^t Y(s)dW(s) = \lim \sum_{n=0}^{N-1} [(1/2)(Y(t_n) + Y(t_{n+1}))(W(t_{n+1}) - W(t_n))].$$

For the Itô integral, a key part of the definition is that the argument t_n of $Y(\cdot)$ is not in the interval (t_n, t_{n+1}). The Stratonovich integral acquires a more symmetric appearance by replacing $Y(t_n)$ with $\frac{1}{2}(Y(t_n) + Y(t_{n+1}))$. However, since $Y(t_{n+1})$ depends itself on $\{W(t), t_n \leq t \leq t_{n+1}\}$, the Stratonovich integral can be interpreted as looking infinitesimally into the future. The Stratonovich integral has been reported as being more appropriate for modeling phenomena in the physical world than its Itô counterpart (Brillinger, 1997; Karlin and Taylor, 1981).

The Itô solution is a diffusion with infinitesimal coefficients $\mu(x,t) = b(x,t)$ and $\Gamma(x,t) = \gamma(x,t)\gamma^T(x,t)$. The Stratonovich solution is also a diffusion, but with infinitesimal coefficients $\mu_i(x,t) = b_i(x,t) + (1/2)\sum_{j=1}^m \sum_{k=1}^m \frac{\partial \gamma_{ij}}{\partial x_k}\gamma_{kj}$ and $\Gamma(x,t) = \gamma(x,t)\gamma^T(x,t)$. Section 2.2 gives an example of a situation where the difference between the Itô and Stratonovich solutions becomes a relevant scientific issue.

Two results we mention here for use later are Itô's lemma, which explains how the infinitesimal coefficients change under a transformation of variables, and the Girsanov theorem, which enables calculation of the likelihood ratio of diffusion processes when this ratio exists.

Theorem 1.1 (Itô's lemma) *Let $X(t)$ be the Itô solution to*

$$dX(t) = b(X(t), t)dt + \gamma(X(t), t)dW(t).$$

For $g(x)$ a twice continuously differentiable function $\mathbb{R}^m \to \mathbb{R}$ and $Y(t) = g(X(t))$, under mild regularity conditions $Y(t)$ satisfies the infinitesimal equation

$$dY(t) = \sum_i \frac{\partial g}{\partial x_i}(X(t))dX_i(t) + \frac{1}{2}\sum_{ij} \frac{\partial^2 g}{\partial x_i \partial x_j}(X(t))dX_i(t)dX_j(t),$$

where $dW_i(t)dW_j(t) = \mathbf{1}_{\{i=j\}}dt$ and $dW_i(t)dt = dtdW_i(t) = 0$. Sufficient regularity conditions are that for all i, j

$$P\left[\int_0^t |b_i(X(s), s)|ds < \infty \text{ for all } t \geq 0\right] = 1$$

$$P\left[\int_0^t (\gamma(X(s), s)\gamma^T(X(s), s))_{ij}ds < \infty \text{ for all } t \geq 0\right] = 1.$$

The following version of the Girsanov theorem is based on Oksendal (1998, Theorem 8.6.5), restated in a form convenient for the statistical inference application to be presented in Chapter 2.

Theorem 1.2 (The Girsanov Theorem) *Let $X(t) \in \mathbb{R}^m$ be the Itô solution to*

$$dX(t) = b(X(t), t)dt + \gamma(X(t), t)dW(t),$$

with $\gamma(X(t), t)$ an invertible $m \times m$ matrix. Under regularity conditions, the law P_T of $\{X(t), 0 \le t \le T\}$ has a density with respect to the law Q_T of the Itô solution for $0 \le t \le T$ of

$$dX(t) = \gamma(X(t), t)dW(t).$$

This density g_T is then a real-valued functional on functions $[0, T] \to \mathbb{R}^m$ which are continuous with finite quadratic variation, and is given by the expression

$$
\begin{aligned}
g_T(x) \;=\; & \exp\left(\int_0^T b^T(x(s), s)(\gamma(x(s), s)\gamma^T(x(s), s))^{-1}dx(s) \right. \\
& \left. -\frac{1}{2}\int_0^T \|\gamma^{-1}(x(s), s)b(x(s), s)\|^2 ds \right).
\end{aligned}
$$

A sufficient regularity condition is that

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \|\gamma^{-1}(X(s), s)b(X(s), s)\|^2 ds \right) \right] < \infty.$$

$\square$

Often the SDE (1.1) will not have an algebraic solution, but the SDE can then often be approximately solved numerically. The simplest numerical technique, the stochastic Euler method, sets

$$X((n+1)\delta) = X(n\delta) + \delta b(X(n\delta), n\delta) + \sqrt{\delta}\gamma(X(n\delta, n\delta))\epsilon_n \qquad (1.2)$$

for some small $\delta > 0$ and with $\epsilon_n \sim N[0, I]$. The notation used here, and subsequently, is that $N[\mu, \Gamma]$ corresponds to a normal distribution whose mean is μ and whose co-variance matrix is Γ. More elaborate techniques are available, based on stochastic Taylor series expansions and implicit methods, similar to those for ordinary differential equations (Kloeden et al., 1992). For the requirements of this thesis, the Euler method appeared to be adequate.

1.3   State Space Models

The general setup for a state space model consists of a so-called *state process* $\{X_n, n = 1, \ldots, N\}$ which is a Markov chain taking values in a *state space* $\mathcal{X}$, and

an *observation process* $\{Y_n = f_n(X_n, \epsilon_n)\}$, where $\{\epsilon_n\}$ is a sequence of independent random variables taking values in $\mathcal{E}$ and $f_n : \mathcal{X} \times \mathcal{E} \to \mathcal{Y}$. Here $\mathcal{Y}$ is called the *observation space*. In particular applications, the sets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{E}$ are often either discrete (finite or countable) or subsets of some metric space.

State space models occur naturally in many scientific and engineering problems. In the case that $\mathcal{X}$ is a discrete or countable set, a state space model is called a hidden Markov model (HMM). HMMs are widely used in applications such as speech recognition (Rabiner, 1989) and locating genes associated with diseases (Lander and Green, 1987). In the case that (X_n, Y_n) is a Gaussian process and $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{E}$ are Euclidean spaces, a state space model arises in the so-called LQG model central to control theory (Whittle, 1996). It appears that real world phenomena are often well modeled by some Markov process with sufficiently rich state space, perhaps constructed according to physical or chemical or economic principles, about which we can make only incomplete or noisy observations. This corresponds to the situation, suggested by the terminology introduced, where we record the observation process and are interested in understanding the behavior of the unobserved state process conditional on the observation process.

The general treatment of state space models described here is based on Kitagawa (1996). No algebraic structure is required on $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{E}$ but we do assume that all required densities exist with respect to some appropriate measures on these spaces. The basic problems of state space models are prediction, filtering and smoothing. These may be described as:

Prediction: Find the conditional density $p(x_n \mid y_1, \ldots, y_{n-1})$, of X_n given observations up to time $n - 1$.

Filtering: Find $p(x_n \mid y_1, \ldots, y_n)$.

Smoothing: Find $p(x_n \mid y_1, \ldots, y_N)$ for $n < N$.

A fourth related problem of particular interest to statistical inference is

Likelihood: Find the density of the complete observation sequence, $p(y_1, \ldots, y_N)$.

These problems can be solved in $O(N)$ computational time, given the initial den-

sity $p(x_0)$, using the following recursion equations for prediction and filtering

$$p(x_n \mid y_1, \ldots, y_{n-1}) = \int_{\mathcal{X}} p(x_{n-1} \mid y_1, \ldots, y_{n-1})p(x_n \mid x_{n-1})dx_{n-1} \qquad (1.3)$$

$$p(x_n \mid y_1, \ldots, y_n) = \frac{p(x_n \mid y_1, \ldots, y_{n-1})p(y_n \mid x_n)}{p(y_n \mid y_1, \ldots, y_{n-1})} \qquad (1.4)$$

$$p(y_n \mid y_1, \ldots, y_{n-1}) = \int_{\mathcal{X}} p(x_n \mid y_1, \ldots, y_{n-1})p(y_n \mid x_n)dx_n. \qquad (1.5)$$

The smoothing problem can then be solved by noting that

$$
\begin{aligned}
p(x_n \mid y_1, \ldots, y_N) &\propto p(x_n, y_n, y_{n+1}, \ldots, y_N \mid y_1, \ldots, y_{n-1}) \\
&= p(y_n, \ldots, y_N \mid x_n)p(x_n \mid y_1, \ldots, y_{n-1})
\end{aligned}
$$

where $p(y_n, \ldots, y_N \mid x_n)$ can be calculated by recursively applying the prediction and filtering operations backwards, starting at time N, namely

$$p(y_n, \ldots, y_N \mid x_n) = p(y_n \mid x_n)p(y_{n+1}, \ldots, y_N \mid x_n)$$

$$p(y_{n+1}, \ldots, y_N \mid x_n) = \int p(y_{n+1}, \ldots, y_N \mid x_{n+1})p(x_{n+1} \mid x_n)dx_{n+1}.$$

The likelihood problem is solved using the quantities calculated in (1.5):

$$p(y_1, \ldots, y_N) = \prod_{n=1}^{N} p(y_n \mid y_1, \ldots, y_{n-1}).$$

For the hidden Markov model with finite state space, the integrals in the recursion equations become sums which may be evaluated exactly. In the linear-Gaussian case there is a closed form solution called the Kalman filter (Box and Jenkins, 1970). In most other cases numerical approximations must be used to solve the recursion equations. A popular method is to make a linear, Gaussian approximation to the actual model. This approximation, termed the Extended Kalman Filter, often works surprisingly well in practice. There are also several numerical approaches that aim to solve approximately the exact recursion equations, rather than solving exactly some approximating equations. Numerical integration can be used (Kitagawa, 1987) to calculate approximately the integrals in (1.3) and (1.5). Monte Carlo Markov chain methods (Carlin et al., 1992) and importance sampling (Durbin and Koopman, 2000)

provide other alternatives. The approach used for the data analysis in Chapter 2 is the sequential Monte Carlo method (Kitagawa, 1996), also known as the Particle Filter.

To describe sequential Monte Carlo, suppose that at time n we have J random particles $X_{n,1}^{(f)}, \ldots, X_{n,J}^{(f)}$ whose marginal distributions approximate the solution to the filtering problem at time n corresponding to the density $p(x_n \mid y_1, \ldots, y_n)$. The prediction problem at time $n + 1$ is then approximated by the marginal distribution of $X_{n+1,j}^{(p)}$ obtained by moving particle $X_{n,j}^{(f)}$ according to the transition probabilities of the state space Markov chain, i.e., draw $X_{n+1,j}^{(p)}$ from $p(x_{n+1} \mid x_n = X_{n,j}^{(f)})$. The filtering problem at time $n + 1$ is approximated by the marginal distribution of $X_{n+1,j}^{(f)}$ obtained by resampling the particles $\{X_{n+1,j}^{(p)}, 1 \leq j \leq J\}$ with weights proportional to the conditional likelihood given Y_{n+1}, i.e., set $X_{n+1,j}^{(f)} = X_{n+1,k}^{(p)}$ with probability

$$p(Y_{n+1} \mid X_{n+1,k}^{(p)}) \Big/ \sum_{l=1}^{J} p(Y_{n+1} \mid X_{n+1,l}^{(p)}) \,.$$

This recursive definition of $X_{n,i}^{(f)}$ and $X_{n,i}^{(p)}$ corresponds to an approximate solution to equations (1.3) and (1.4). Del Moral (1996) shows that this approximation error tends to zero as the number J of particles increases. Del Moral and Guionnet (1999) find a central limit theorem for this convergence. The likelihood can then be estimated via

$$\hat{p}(y_n \mid y_1, \ldots, y_{n-1}) = \frac{1}{J} \sum_{j=1}^{J} p(y_n \mid X_{n,j}^{(f)}).$$

For many applications, including some of the situations encountered in Chapter 2, a linear state space model is adequate. In this case, a state space model is equivalent to the much used ARMA model described in Box and Jenkins (1970), with an implementation in S-Plus described in Venables and Ripley (1995).

1.4 Models for Cell Shape

In this section, a sequence of mathematical descriptions of cell shape is introduced with decreasing generality and increasing simplicity. The ability of these models

to describe features of interest for scientific research is discussed. In Section 2.4 particular parametric forms are suggested and a data analysis presented.

Some previous work has studied cell shape more indirectly than is done in this thesis. Dunn and Brown (1990) use moments to quantify cell shapes, but comment on the difficulty of interpreting third and higher order moments. Noble (1990) calculates a skeleton spanning the shape of a cell in order to identify lamellipodia. Soll et al. (1997) describe a three-dimensional dynamic image analyzing system which computes "more than 100 parameters of motility and dynamic morphology". Writing down a formal stochastic model can help by identifying appropriate parameters, showing relations between the parameters and making efficient use of available data for statistical inference. Some pitfalls in using an assorted collection of parameters to quantify and describe shapes, without an appropriate model, are discussed in Freedman et al. (1998, Section 12.3).

The field of statistical shape theory, surveyed in Small (1996), defines the shape of an object, data set or image as the total of all information that is invariant under translations, rotations and isotropic rescalings. In our situation, symmetries of the experiment may or may not lead to translation and rotation invariance. For example, motion in a uniform electric field should have the same symmetries as the field itself, namely translation but not rotation invariance. The idea of invariance to isotropic rescaling is appealing. Each cell could then have its own scale and behave in the same way up to its scale factor. Whether this assumption is justified in practice will be decided by the data. At any rate, we do not intend to be restricted by the above definition of Small (1996).

We suppose the interior of a cell at time t is given by a subset S_t of $\mathbb{R}^n$, giving rise to a shape process $S = \{S_t, 0 \leq t \leq T\}$. Usually n is either 2 or 3. If one supposes that S takes a value which is a closed (or equivalently, open) subset of $\mathbb{R}^n \times [0, T]$ then S can be formally constructed as a stochastic process using the methods of Matheron (1975). In this general form, S_t could be a disconnected subset of $\mathbb{R}^n$, which is a biologically plausible possibility as the cell may divide or may leave some part of itself behind, stuck on the substrate, when it moves. However we might ignore these complications and suppose that S_t is a continuous deformation of S_0, which is in turn

a continuous deformation of the unit ball in $\mathbb{R}^n$. S_t is then said to be homotopic to the unit ball. This means that S can be represented as the image of a shape function

$$s : B_n \times [0, T] \rightarrow \mathbb{R}^n \tag{1.6}$$

where $B_n = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| \leq 1\}$, the unit ball, and where $s(\cdot, \cdot)$ is continuous in space and time. A reference for the topological terms and results used in this section is Hocking and Young (1961). S_t is defined to be the image of $s(\cdot, t)$, which makes S_t closed, compact and connected. The parameterization is far from unique, as many different shape functions give the same image. A natural way of constructing a shape function is for $s(\mathbf{x}, t)$ to give the position at time t of an imaginary particle attached to the cell at time $t = 0$ and position $s(\mathbf{x}, 0)$. In experimental situations, the imaginary particle can be replaced by a small bead which allows observation of the local movement of the cell cytoplasm or membrane (Anderson et al., 1996). Depending on the coating applied to a bead, it can either stay attached to the surface of the cell membrane or undergo endocytosis. The shape function defined in this natural way contains more information than the shape process alone. One feature of interest that it can describe is the ruffling of a lamella. The leading edge of the cell may fold on itself, causing a so-called ruffle (Hinz and Brosteanu, 1997). This results in the shape function being not one to one as its image must fold back on itself.

Another approach is to model the boundary of the cell in terms of a boundary function

$$b : \sum_{n-1} \times [0, T] \rightarrow \mathbb{R}^n$$

where $\sum_{n-1} = \{\mathbf{x} \in \mathbb{R}^n : |x| = 1\}$, the unit n-sphere embedded in $\mathbb{R}^n$, and $b(\cdot, \cdot)$ is continuous in space and time. Supposing that $b(\cdot, t)$ is one to one, it defines a topological transformation, or homeomorphism, of the sphere. It is more appropriate to think of topological rather than homeotopic mappings as models of cell boundaries, as the former preserve the sense of inside and outside, by the Jordan–Brouwer separation theorem. The boundary function retains the interpretation of the shape function in terms of the motion of particles attached to the cell surface if there is no ruffling or recycling of the cell membrane. A further assumption, which produces an attractively

simplified model, is to suppose that the boundary function has a center-radius form. For the case $n = 2$, appropriate for many microscopy images, this means that the boundary function has the form

$$b(\mathbf{x}, t) = \mathbf{c}(t) + r(\theta, t)\mathbf{x}$$

where $\mathbf{x} = (\cos\theta, \sin\theta)^T$. This restrictive form is adequate for cell types that are near convex in shape, such as keratinocytes, fish keratocytes and Xenopus mesoderm. It is not so suitable for some other commonly studied cell types such as fibroblasts and PMN (human polymorphonuclear white blood cells). The center, $\mathbf{c}(t)$, may be an arbitrary measure of cell location. A convenient specification of the center is the centroid, defined as

$$\bar{c}(t) = \int_{S_t} \mathbf{x} dS \left\{ \int_{S_t} dS \right\}^{-1}.$$

When observing a two-dimensional image, the centroid may be far from the true center of mass of the cell as cell thickness varies enormously between the thick cell body and the thin lamella. If the cell body can be clearly distinguished from the image, its centroid may be closer to the center of mass.

A well-studied statistical problem requiring stochastic shape models is Bayesian object recognition. The object of interest is modeled by a prior distribution on some shape space. The shapes considered may be deformations of a template incorporating knowledge of the form of the objects being looked for (Grenander and Keenan, 1993) or may be a space of random polygons used to detect boundaries of arbitrary objects (Pievatolo and Green, 1998). These prior distributions are candidate models for cell shapes. For example, Grenander and Manbeck (1993) model potatoes using deformations of an ellipse coming from a multivariate von Mises distribution. Grenander and Miller (1994) model mitochondria using a polygon whose edge vectors are multivariate normal. Many other objects such as hands, brains and stomaches have been modeled in similar ways.

Chapter 2

Statistical Analysis of Cell Motion

Studies involving observation of cell motion may be categorized, as in Dickinson and Tranquillo (1995), by the length and time scales under primary consideration. On the scale of *locomotion* the basic actions of cell motion are apparent: this is the scale on which a cell may extend individual protrusions (broad lamellae, thinner lamellipodia or hairlike filopodia) that can be used to pull itself along. On a longer time and length scale, termed *translocation*, one observes the displacement of the cell due to one or several motion cycles. As the resolution of detail about the motion decreases, gross tendencies such as directional preferences can become more apparent. On the scale of *migration*, the cumulative effect of many motion cycles is observed. Although on the migration scale one loses the ability to observe directly the mechanisms of cell motion, it is the behavior of cells on this scale which is of primary interest in applications, such as development, cancer metastasis, and tissue engineering. Furthermore, some assays such as the Boyden diffusion chamber (Byrne et al., 1998) entail observations of populations of cells for which only the behavior on the migration scale can be directly observed. Dickinson and Tranquillo (1995) develop mathematical methods to relate models on different scales, using the method of adiabatic elimination of fast variables (Gardiner, 1983).

The Keller–Segel model of chemotaxis (Keller and Segel, 1971; Byrne et al., 1998) gives a widely accepted approach for modeling cells on the migration scale, using Fokker–Planck equations. These models can be written as stochastic differential equa-

tions, and are discussed in Section 2.1. Section 2.1 also gives a thorough discussion
of models for translocation, using methods based on stochastic differential equations
to develop qualitative and quantitative understanding of these models. Section 2.2
discusses some situations where the models introduced in Section 2.1 clarify and con-
tribute to questions of scientific interest. Section 2.3 addresses hypothesis testing, and
Section 2.4 parameter estimation for the models of Section 2.1. Section 2.5 presents
an investigation of cell motion on the scale of locomotion.

2.1   Models of Cell Translocation and Migration

In this section we look at some models which quantify concepts arising in the
studies of cell motion discussed in Section 1.2. Each model describes the vector-
valued position or velocity of a single cell moving in the plane. The reader is asked
to suspend skepticism about the validity of these models until Section 2.2, where
it is indicated how such models may be used to help rather than hinder scientific
understanding even with the assurance that none provides a true representation of
reality. Also, we will look at ways to quantify the concepts of biological interest in
ways that do not rely on the validity of a particular model.

When the distribution of the velocity process has rotational symmetry about the
origin, and does not depend on position, the model is called *isotropic*. Such models,
suitable when the cell experiences no directional stimuli, are discussed first. When
the velocity process has directional asymmetry or depends on position then the cell is
said to perform *taxis*. The cell must then be picking up some locational or directional
cue from its environment.

Two characteristics used to describe isotropic cell motion are speed and persis-
tence. The story is that in the short term cells are observed to move with slowly
varying direction and speed. After a while they appear to forget their initial ori-
entation. This time scale is termed the *persistence* of the cell. The main model
for isotropic translocation in the literature is the so-called correlated random walk
(Alt, 1990; Dunn, 1983) where the velocity $\mathbf{v}_t = (v_x(t), v_y(t))$ follows an Ornstein–
Uhlenbeck process. This process is defined by the stochastic infinitesimal equation

(see Section 1.2)

(M1)
$$d\mathbf{v}_t = -a\mathbf{v}_t dt + b d\mathbf{W}_t$$

for positive constants a and b and a two-dimensional Brownian motion $\mathbf{W}_t$. The coefficient $-a\mathbf{v}_t$ is called the infinitesimal drift and b^2 the infinitesimal variance. Heuristically, a gives the rate at which the velocity regresses to zero and b gives the magnitude of the random innovations which tend to push the velocity away from zero. The root mean square speed can be calculated as $\sqrt{b^2/a}$, and a measure of persistence is $1/a$. One feature sometimes observed for cells is that their direction of motion changes most rapidly when their speed is small (Shenderov and Sheetz, 1997). This behavior is a property of (M1), as is most clearly seen by transforming to polar coordinates (r_t, θ_t) for the velocity by applying Itô's lemma (Section 1.2). This leads to the infinitesimal equations

$$
\begin{aligned}
dr_t &= \left(-ar_t + \frac{b^2}{2r_t} \right) dt + b dW_t^{(r)} \\
d\theta_t &= \left(\frac{b}{r_t} \right) dW_t^{(\theta)}
\end{aligned}
$$

for two independent one-dimensional Brownian motions $W_t^{(r)}$ and $W_t^{(\theta)}$. In Shenderov and Sheetz (1997) it is also observed that for some cell types the velocity has an oscillatory behavior, with high correlation between velocities at certain time lags. They propose a model

(M2)
$$d\mathbf{v}_t = \left(-a\mathbf{v}_t - c \int_{-\infty}^{t} \mathbf{v}_s e^{-k(t-s)} ds \right) dt + b d\mathbf{W}_t,$$

where the evolution of the velocity process depends on the past values of the process. Shenderov and Sheetz show that (M2) has oscillatory behavior for $4c > (a-k)^2$ in the sense that the solutions to the related ordinary differential equation

$$\frac{dx}{dt} + ax(t) + c \int_0^t x(s) e^{-k(t-s)} ds = 0, \quad t \geq 0$$

are oscillatory. (In fact they miscalculate by a factor of $\sqrt{2\pi}$, but we have not reproduced their error.) A natural probabilistic approach to generalizing (M1) to include

oscillatory behavior is to notice that (M1) is in the form of a continuous time autoregressive model of first order, CAR(1), studied in Jones (1981). The CAR(2) model is the solution to the coupled equations

(M3)
$$
\begin{aligned}
d\mathbf{v}_t &= \dot{\mathbf{v}}_t dt \\
d\dot{\mathbf{v}}_t &= (-\alpha d\mathbf{v}_t - \beta \dot{\mathbf{v}}_t) dt + b d\mathbf{W}_t.
\end{aligned}
$$

One can rewrite (M2) to look more like (M3) by setting $\mathbf{u}_t = \int_{-\infty}^t \mathbf{v}_s e^{-k(t-s)} ds$ which turns (M2) into the coupled system

$$
\begin{aligned}
d\mathbf{v}_t &= (-a\mathbf{v}_t - c\mathbf{u}_t) dt + b d\mathbf{W}_t \\
d\mathbf{u}_t &= (-k\mathbf{u}_t + \mathbf{v}) dt.
\end{aligned}
$$

In the sense of Shenderov and Sheetz (1997) (M3) has oscillatory behavior for $4\alpha > \beta^2$, the condition required for the matrix $\begin{pmatrix} 0 & 1 \\ -\alpha & -\beta \end{pmatrix}$ to have complex eigenvalues. A possible advantage of (M3) over (M2) is that it introduces only one rather than two extra parameters into the model to describe the one extra concept of oscillatory frequency.

In the non-isotropic case many ways have been proposed by which a cell might respond to a stimulus on the scale of cell translocation, and these are termed *modes of taxis* (Dickinson and Tranquillo, 1995). *Topotaxis* occurs when a cell turns preferentially toward a stimulus. *Orthotaxis* is said to occur if the magnitude of the velocity of the cell increases when the direction is toward a stimulus. *Klinotaxis* occurs when the rate of turning decreases while traveling toward a stimulus. These three modes depend on the direction of a stimulus, but there are further two modes that depend only on the magnitude of a stimulus. *Orthokinesis* occurs when the magnitude of the velocity decreases with the magnitude of a stimulus. *Klinokinesis* occurs when the rate of turning increases with the magnitude of a stimulus.

The reader may wonder whether these are the only possible modes, and whether the observed motion of a cell toward a stimulus can be uniquely characterized as some combination of these modes. In fact these questions are of basic scientific interest, since modes of taxis are experimentally testable consequences of models at the

mechanistic level for the biochemistry and biophysics of cell motion. Unfortunately there has been some confusion in the literature about how to decide empirically upon the modes of taxis, based on observations from a system. Doucet and Dunn (1990) discuss this problem and give the example of classifying the mode of taxis of a snake whose head can detect the level of a chemical attractant. By moving its head from side to side this snake detects the gradient of the chemical and moves up it. The whole snake appears to be capable of topotaxis, while mechanistically it can only measure the magnitude of the stimulus and so should be capable only of a kinesis.

To formalize modes of taxis mathematically one can avoid the snake paradox by defining modes of taxis as properties of models rather than biophysical mechanisms. The time and length scale on which we are modeling a process can determine the characterization of the behavior. Recall the three scales identified, of locomotion, translocation and migration. On the scale of locomotion, in which the biophysical properties of a cell result in the extension of pseudopodia and traction along a substrate, modes of taxis lose meaningfulness as a way to characterize the process. On the scale of translocation the position and velocity of a cell, but not its internal processes, are modeled. One can then attempt to define modes of taxis. If there is a scale on which the internal processes regulating the velocity of a cell has negligible memory the velocity process may be modeled by the infinitesimal equations

$$
\text{(M4)} \quad
\begin{aligned}
dr_t &= \mu_r(r_t, \theta_t, s_t, \phi_t)dt + \sigma_r(r_t, \theta_t, s_t, \phi_t)dW_t^{(r)} + \tau_r(r_t, \theta_t, s_t, \phi_t)dW_t^{(\theta)} \\
d\theta_t &= \mu_\theta(r_t, \theta_t, s_t, \phi_t)dt + \sigma_\theta(r_t, \theta_t, s_t, \phi_t)dW_t^{(\theta)} + \tau_\theta(r_t, \theta_t, s_t, \phi_t)dW_t^{(r)}.
\end{aligned}
$$

Here (r_t, θ_t) are the polar coordinates for the velocity $\mathbf{v}_t$. The coordinate pair $(s_t, \phi_t) = (s_t(\mathbf{x}_t), \phi_t(\mathbf{x}_t))$ gives the magnitude and direction of the stimulus at the location $\mathbf{x}_t$ of the cell. For multiple stimuli, s_t and ϕ_t take vector values. $W_t^{(r)}$ and $W_t^{(\theta)}$ are two independent Brownian motions. Assuming the process $(\mathbf{v}_t, \mathbf{x}_t)$ is continuous, Markov and time homogeneous it is a small restriction to suppose it has a representation as a solution to (M4) (Karlin and Taylor, 1981). The exact way in which a solution is found for the infinitesimal equations in (M4) becomes relevant as the two major competitors—the Itô and Stratonovich solutions—differ when σ_r, τ_r, σ_θ or τ_θ are non-constant. For models (M1), (M2) and (M3) the two solutions

coincide. Some consequences of the choice of solution will be discussed later.

The coefficients in model (M4) can be given biological interpretations, under some further assumptions. A reason for writing down the model so generally in the first place was to make these assumptions explicit. Assuming the cell has no reason to rotate in a particular direction without a directional cue, (see Alt (1990) for a counterexample), $\mu_\theta(r_t, \theta_t, s_t, \phi_t)$ fits the description of a topotaxis term. If $\mu_r(r_t, \theta_t, s_t, \phi_t)$ can be written as

$$\mu_r(r_t, \theta_t, s_t, \phi_t) = \mu_r^{(1)}(r_t) + \mu_r^{(2)}(r_t, s_t) + \mu_r^{(3)}(r_t, \theta_t, s_t, \phi_t)$$

then $\mu_r^{(2)}(r_t, s_t)$ has the form of an orthokinesis term and $\mu_r^{(3)}(r_t, \theta_t, s_t, \phi_t)$ has the form of an orthotaxis term. Similarly, if $\sigma_\theta(r_t, \theta_t, s_t, \phi_t)$ can be written as

$$\sigma_\theta(r_t, \theta_t, s_t, \phi_t) = \sigma_\theta^{(1)}(r_t) + \sigma_\theta^{(2)}(r_t, s_t) + \sigma_\theta^{(3)}(r_t, \theta_t, s_t, \phi_t)$$

then $\sigma_\theta^{(2)}(r_t, s_t)$ can stake a claim as a klinokinesis term and $\sigma_\theta^{(3)}(r_t, \theta_t, s_t, \phi_t)$ as a klinotaxis term. The remaining terms σ_r, τ_r and τ_θ have no clear roles to play in the existing modes of taxis, indicating that these modes form an incomplete picture of the possible directional behavior in (M4). For example a change in the random variation in speed, caused by a varying level of a ligand that interacts with the speed regulation mechanisms of a cell, might cause directional behavior through a term $\sigma_r(r_t, s_t)$.

On the scale of migration when the location $\mathbf{x}_t$ is supposed to have negligible memory one can write down an analogue to (M4), namely

(M5) $$d\mathbf{x}_t = \mu(\mathbf{x}_t)dt + \gamma(\mathbf{x}_t)d\mathbf{W}_t.$$

Here $\gamma(\mathbf{x}_t)$ is a 2×2 matrix. Since the stimulus is assumed to depend only on position, there is no need to include it explicitly in (M5). In this model the two concepts of rate of turning depending on position and of speed depending on position are linked together in the matrix $\gamma(\mathbf{x}_t)$. Indeed since the sample paths are not differentiable one has to take a broad minded view about "speed" and "rate of turning" to recognize $\gamma(\mathbf{x}_t)$ as a combined kinesis term for both klinokinesis and orthokinesis. Similarly $\mu(\mathbf{x}_t)$ can be thought of as a taxis term, combining topotaxis, orthotaxis and klinotaxis. An alternative interpretation of the parameters in (M5) would come

from applying a rescaling argument to (M4). The technique of adiabatic elimination of fast variables (Gardiner, 1983) can provide such a rescaling for certain particular cases of (M4).

Choosing between the interpretations of persistence, periodicity, speed and modes of taxis given by different models requires more precise definitions of these concepts than are currently available in the biological literature. The goal here has been to present some options, rather than to come down heavily in favor of any one model.

2.2 Consequences and Applications

Three situations are covered that further develop aspects of the previous section. A simple version of model (M5) is used to address the question of whether a pure kinesis is a viable mechanism for directional cell motion, and incidentally to compare Itô and Stratonovich integrals. Some model free definitions of speed, persistence and periodicity for cells are suggested. A version of model (M5) is used to quantify galvanotaxis, the motion of a cell in an electric field, and to justify a model free test.

2.2.1 Do kineses work?

There has been some controversy about whether a cell can move up a gradient of a ligand (a small signaling molecule) just by adjusting its speed or rate of turning according to the concentration of the ligand (Doucet and Dunn, 1990). In other words, do orthokinesis and klinokinesis work as a way of moving up concentration gradients, or must the cell in fact have some memory or ability to detect gradients. A careful theoretical study of klinokinesis where the velocity of a bacterium is treated as a Markov process is undertaken in Stroock (1974). The resulting model fails to give convincing evidence that kineses can work. The best result obtained there is that if $\{x_t\}$ is the $\mathbb{R}$-valued process considered in Stroock (1974), giving the position of the bacterium up an increasing gradient of a ligand, and $f : \mathbb{R} \to \mathbb{R}$ is a convex function then $\mathbb{E}[f(x_t)]$ is monotone increasing with time. This would also be true if $\{x_t\}$ were a random walk, or a martingale.

Kinesis may be modeled in a simple but instructive way by considering a stochastic process $\{x_t\}$, taking values in $(0, \infty)$, defined by the infinitesimal equation

$$dx_t = \sigma x_t dW_t. \qquad (2.1)$$

This is a particular case of (M5). The Itô solution is $x_t^I = x_0 e^{W_t - t/2}$ and the Stratonovich solution is $x_t^S = x_0 e^{W_t}$. The reader is referred to Karlin and Taylor (1981) and Oksendal (1998) for the background on stochastic differential equations beyond the brief introduction in Section 1.2. Calculating expectations gives

$$
\begin{aligned}
\mathbb{E}[x_t^I \mid x_0] &= x_0 \\
\mathbb{E}[x_t^S \mid x_0] &= x_0 e^{t/2}.
\end{aligned}
$$

So for the Stratonovich solution kinesis works, while for the Itô solution it does not! There is little scientific reason for preferring one solution to the other, and this result suggests that there is equally little reason to decide whether a kinesis results in motion up the gradient or whether there must be an additional taxis for this to occur. It may be valuable to determine whether speed and rate of turning vary with stimulus level but an attempt to assign motion up a gradient to this phenomenon has no scientific basis within the framework of (M5).

2.2.2 Model-based and model-free methods

On many occasions fitting and assessing a model is of direct interest. Examples of this include generating a model to use for simulation of part of a complex system, and when a theory to be tested makes explicit claims concerning a particular model. Fitting models (M1)–(M5) is discussed further in Section 2.4 and Chapter 3. Now we comment on the other situation where the model itself is secondary to the scientific question at hand.

Scientists, particularly in the field of biology, are often concerned with comparing particular experimental groups. Quantities measured to compare these groups do not necessarily have to make direct substantive sense outside the experiment being carried out. For example, if cell location is measured every 10 minutes for one hour

giving measurements $\mathbf{x}_i$, $0 \leq i \leq 6$ then the mean speed of a cell could be quantified as $S = \frac{1}{6}\sum_{i=1}^{6}|\mathbf{x}_i - \mathbf{x}_{i-1}|$. This is a common and not unreasonable measure for comparing mean speed between experimental groups all observed at 10-minute intervals. With a sufficiently experienced eye one can compare such a result with results from similar experiments where it may have been convenient to record at 5- or 15-minute intervals. A more subtle and serious problem with the statistic S is that it does not always act as a reliable proxy for the physical concept of mean cell speed. To demonstrate this suppose that (M1) holds for a pair of control and treatment groups having parameters (a_0, b_0) and (a_1, b_1) respectively. If $\frac{b_0^2}{a_0} = \frac{b_1^2}{a_1}$ but $a_2 \neq a_1$ then the estimates S_0 and S_1 arising from evaluating S for control and treatment groups have $\mathbb{E}S_0 \neq \mathbb{E}S_1$. For model (M1) we can calculate explicitly

$$\mathbb{E}[S] = \mathbb{E}\left|\mathbf{v}_0\frac{(1-e^{-a})}{a} + b\int_0^1 \frac{1}{a}(1 - e^{-a(1-t)})d\mathbf{W}_t\right|$$

which, after routine algebra, leads to

$$\mathbb{E}[S] = \sqrt{\frac{\pi b^2}{2a}\left(\frac{e^{-a} - 1 + a}{a^2}\right)}. \tag{2.2}$$

When a is large, so the persistence of the cell is low, we see that for a given value of b^2/a the value of $\mathbb{E}[S]$ becomes small.

From this example we see that if one believes model (M1) and yet uses statistic S then one might be led to conclude that the cell speed varies between treatment and control when in fact only the persistence varies. To avoid embarrassments of this kind it is necessary to bear in mind that S is only an observable proxy for a more objective quantity such as the root mean square velocity

$$S^* = \left(\frac{1}{60}\int_0^{60}|\mathbf{v}_t|^2 dt\right)^{1/2}.$$

To check that S is doing its job of substituting for S^* one could either do further experiments to find out what happens when S is calculated using smaller time intervals, or fit an appropriate model to estimate what would happen. For example, in the context of model (M1),

$$\mathbb{E}[(S^*)^2] = \frac{b^2}{a}. \tag{2.3}$$

If a is constant across an experiment then both S and S^* scale linearly with b. If not, the two quantities S and S^* are less comparable, as can be seen by comparing (2.2) and (2.3).

2.2.3   A model for galvanotaxis

Recall the experiment described in Example 1 of Section 1.1, where cells are moving on a microscopic slide in a uniform electric field. The influence of the electric field on the motion of a cell is termed *galvanotaxis*. It is of interest to quantify galvanotaxis to help describe how it varies with experimental treatments. Empirically one notices that the speed of the cells is not much affected by the electric field (Nishimura et al., 1996). Theory and observation suggest that changes in cell direction are governed by local behavior around the edges of the leading Lamella (Dunn et al., 1997). A model consistent with these considerations as well as the symmetry and translation invariance of the experiment is (M6), below. This model is an extension of (M1) and a special case of (M4). The electric field is taken to have magnitude $\mathcal{E}$ in the direction of the positive x-axis, which coincides with the direction $\theta = 0$. The velocity $\mathbf{v}_t = (v_x(t), v_y(t))^T$ has polar representation (r_t, θ_t)

$$
\text{(M6)} \qquad\qquad d\mathbf{v} = \begin{pmatrix} -\alpha & \beta \sin \theta_t \\ -\beta \sin \theta_t & -\alpha \end{pmatrix} \mathbf{v}_t dt + \gamma d\mathbf{W}_t
$$

An application of Itô's lemma shows the polar representation of the infinitesimal equation defining model (M6) to be

$$
\begin{aligned}
dr_t &= \left(-\alpha r_t + \frac{2\gamma^2}{r_t} \right) dt + \gamma dW_t^r \\
d\theta_t &= -\beta \sin \theta_t dt + \left(\frac{\gamma}{r_t} \right) dW_t^\theta.
\end{aligned}
$$

The magnitude of the velocity is governed by the same equation as for model (M1). The directional behavior of model (M6) is seen to be a rotation of the direction of motion at rate $\beta \sin \theta_t$ toward $\theta = 0$. Thus (M6) fits the description of a topotaxis.

One could quantify galvanotaxis by fitting (M6), as discussed in Section 2.4, and using an estimate $\hat{\beta}$ of β. Another possibility, attractive for its simplicity, is to

calculate an approximation to the so-called score statistic. The likelihood function, $L(\alpha, \beta)$, when $\mathbf{v}_t$ is observed for t in the interval $[0, T]$ and γ is known, is taken to be the density of the process (M6) having parameters (α, β, γ) with respect to the process (M6) with parameters $(0, 0, \gamma)$, evaluated at $\{\mathbf{v}_t, t \in [0, T]\}$. This density, which in formal probabilistic language is termed a Radon-Nikodym derivative, is given by the Girsanov Theorem (Section 1.2) as

$$
\begin{aligned}
L(\alpha, \beta) \;=\; & \exp \left\{ \frac{-\beta}{\gamma^2} \int_0^T r_t^2 \sin \theta_t d\theta_t \right. \\
& \left. - \frac{\beta^2}{2\gamma^2} \int_0^T r_t^2 \sin^2 \theta_t dt - \frac{\alpha}{\gamma^2} \int_0^T r_t dr_t - \frac{\alpha^2}{2\gamma^2} \int_0^T r_t^2 dt + 2\alpha T \right\}.
\end{aligned}
$$

The partial derivative of the logarithm of the likelihood with respect to β evaluated at $\beta = 0$ is termed the (Fisher) score statistic for testing the null hypothesis that $\beta = 0$. The score statistic then, up to an unimportant constant factor, is

$$
Z = \int_0^T r_t^2 \sin \theta_t d\theta_t. \tag{2.4}
$$

In Cartesian coordinates this becomes

$$
Z = \int_0^T \frac{1}{|\mathbf{v}_t|} (v_x(t) v_y(t) dv_y(t) - v_y^2(t) dv_x(t)).
$$

From symmetry considerations, Z has expectation zero when $\beta = 0$ (for any value of α) and so if i.i.d. replicates are available the t statistic can be used to test the hypothesis that $\beta = 0$. This suggests a statistic for a discretely observed process formed by replacing the integral in (2.4) by a finite sum (Kloeden et al., 1996). When the locations $\{\mathbf{x}_t = (x_t, y_t), \; t = 0, 1, \ldots, T\}$ are observed, an approximation to Z is given by setting $x_{t+1} - x_t = \hat{r}_t \cos \hat{\theta}_t$, $y_{t+1} - y_t = \hat{r}_t \sin \hat{\theta}_t$ and then constructing the statistic

$$
Z_1 = \sum_{t=1}^{T-1} \hat{r}_t^2 \sin \hat{\theta}_t (\hat{\theta}_{t+1} - \hat{\theta}_t).
$$

Biologists currently use the statistic (Nishimura et al., 1996; Fang et al., 1999)

$$
Z_2 = (x_T - x_0)/|\mathbf{x}_T - \mathbf{x}_0|.
$$

From symmetry considerations, Z_1 and Z_2 both have zero expectation whenever the velocity process has rotationally symmetric distribution, so they can readily be used

to test for homogeneity. When the true behavior of the cells is similar to model (M6), the statistic Z_1 gives rise to an approximate score test. The score test is asymptotically equivalent to a likelihood ratio test (Rao, 1973), and so has similar asymptotic optimality properties. These statistics are compared in practice as part of the data analysis in Section 2.3.

2.3   Inference from Cell Tracking Data

For an experiment studying the behavior of isolated cells moving on a microscope slide, on the scale of translocation, the data consist of M time series each of length N,

$$\{\mathbf{x}_j^{(i)},\ 1 \le i \le M,\ 1 \le j \le N\}. \tag{2.5}$$

Each time series $\mathbf{x}^{(i)} = \{\mathbf{x}_j^{(i)},\ 1 \le j \le N\}$ gives the location of a cell in $\mathbb{R}^2$, measured in an appropriate way, at each of N equally spaced time points. Cells that do not come close enough to a neighbor to interact directly (roughly two cell body diameters) are presumed to be independent, and, by restricting attention to such cells, the M time series may be considered independent replicates.

Extensions to this situation include experiments where $\mathbf{x}_j^{(i)}$ takes values in $\mathbb{R}^3$ (Noble, 1990), measurement of additional features beyond cell location (Section 2.5), dependence between time series, and the case without replications where only one cell is observed (usually for a longer time).

The location data derive from time lapse microscopy images. An example of one time frame is given in Figure 2.1. Converting this image data into the form of equation (2.5) is called the *cell tracking problem.* One approach to cell tracking is the manual method of following each cell from frame to frame by eye and making some visually determined center of the cell as its location. There are existing computer programs to automate cell tracking (Soll and Wessels, 1998), though none are widely available or in common use for data sets of the type considered in this thesis. In practice, an extension of the image processing techniques described in Chapter 4 was used to produce a satisfactory cell tracking program. An outline of the algorithm is
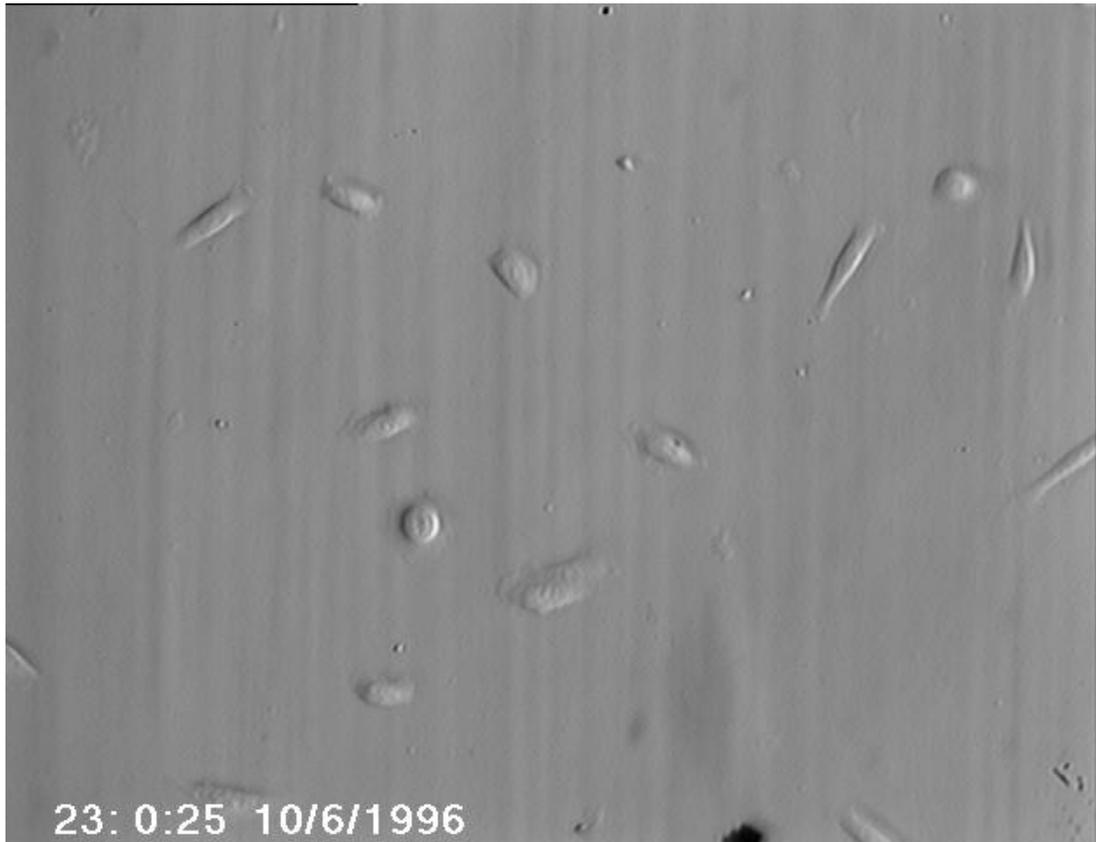
Figure 2.1: One time frame of human keratinocytes moving on a microscope slide, prepared as in Fang *et al. (1999) and viewed using differential interference contrast microscopy.*

as follows.

1. Low frequency components of the image are removed by subtracting off a heavily smoothed version of the image. This removes only microscopy artifacts, since the cells are small compared to the size of the image.

2. Thresholding, closing, and filling in connected components are applied, as described in Chapter 4, to produce at each time j a group of candidate cell shapes. Any candidate whose size or length to width ratio is implausible for a cell is discounted as being an artifact.

3. Each candidate cell at time j is supposed to correspond to the candidate cell

at time $(j-1)$ closest to its position, as long as there is a candidate plausibly close. Otherwise, the candidate is discounted from the analysis.

4. If two cells at time $j-1$ correspond to the same candidate at time j, the two cells are assumed to have come into contact. They are discounted from the later analysis, which is intended to be carried out only on single, isolated cells.

5. Each of the M candidate cells at time 1 that has a unique correspondence at each time $j \leq N$ gives rise to a time series $\{\mathbf{x}_j^{(i)}, \ 1 \leq i \leq M, \ 1 \leq j \leq N\}$.

6. An interactive video of the proposed solution to the cell tracking problem is checked visually, allowing for the correction of mistaken cell identities.

This algorithm, which was implemented in a MATLAB program available from the author, was sufficient to deal with the two main difficulties of the tracking problem for the data encountered:

(i) The presence of many features in the image not corresponding to cells. These could, for example, be artifacts of the microscopy or fragments of organic matter. Some such features can be noticed in Figure 2.1.

(ii) The tendency of the cells to combine together when they encounter one another and continue moving slowly as a group.

Two examples of resulting collections of time series are shown in Figure 2.2, for treatment and control experiments, with and without an electric field, carried out as described in Example 1 of Section 1.1.

The three models we shall consider here for the unobserved velocity process $\mathbf{v}_t$ are

(N1) $$d\mathbf{v}_t = -\alpha \mathbf{v}_t dt + \sigma d\mathbf{W}_t$$

(N2) $$d\mathbf{v}_t = \begin{pmatrix} -\alpha & \beta \sin \theta_t \\ -\beta \sin \theta_t & -\alpha \end{pmatrix} \mathbf{v}_t + \sigma d\mathbf{W}_t$$

(N3) $$d\mathbf{v}_t = -\alpha (\mathbf{v}_t - \beta (1,0)^T) + \sigma d\mathbf{W}_t.$$

(i)

(ii)
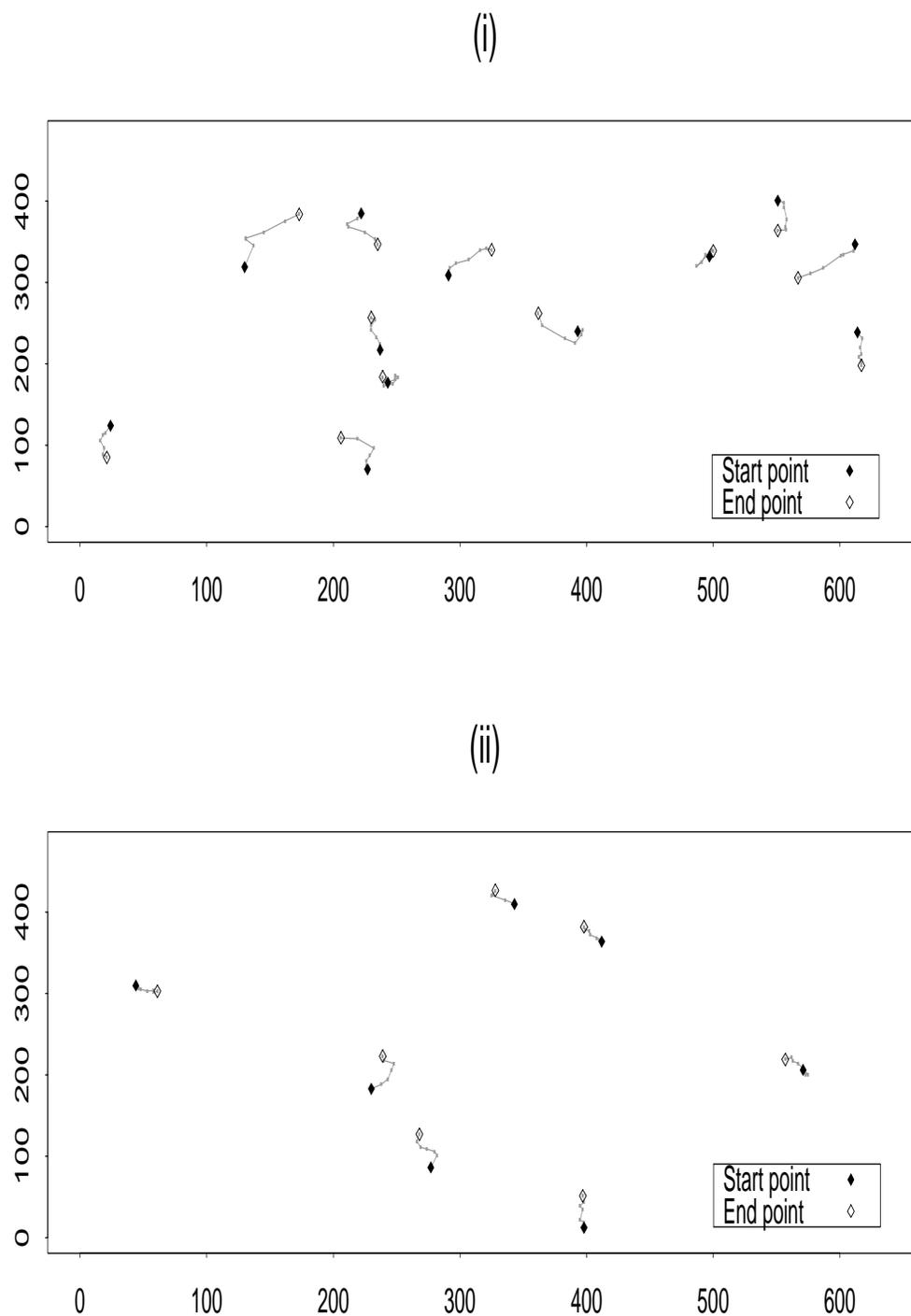
Figure 2.2: Cell paths resulting from applying the tracking algorithm to two micro-scope slides. (i) A control experiment, prepared as in Fang *et al.* (1999), with no electric field. (ii) A treatment experiment, with an electric field off 100 mV/mm. The cathode is at the top of the page.

(N1) is an Ornstein–Uhlenbeck process. (N2) is the model for galvanotaxis proposed in Section 2.2 and (N3) is a recentered Ornstein–Uhlenbeck process. (N1) has rotational symmetry above the origin, whereas (N2) and (N3) show directional behavior for nonzero values of the parameter β. For (N2), β is the rate of rotation toward the positive x-axis. For (N3), β is the expected velocity in the direction of the positive x-axis. For each of these models the time series $\{\mathbf{x}_j,\ 1 \leq j \leq N\}$ for a single cell is modeled by

$$\mathbf{x}_{j+1} - \mathbf{x}_j = \int_{j-1}^{j} \mathbf{v}_t dt + \epsilon_{j+1} - \epsilon_j$$

where $\{\epsilon_j,\ 1 \leq j \leq N\}$ is a sequence of measurement errors which are here taken to be i.i.d. random variables with distribution $N(0, \tau^2 I)$. The particular form for the measurement error does not play a role until Section 2.4.

One interest in deciding whether (N2) or (N3) provides a good description of cell behavior is that, if one wishes to test whether there is directional behavior, different models will suggest different test statistics. In Section 2.2, model (N2) was shown to suggest an approximate score statistic which, written in Cartesian coordinates, takes the form

$$T_2 = \sum_{i=2}^{N-1} \frac{1}{|\hat{\mathbf{v}}(i)|} (\hat{v}_y^2(i)(\hat{v}_x(i+1) - \hat{v}_x(i)) - \hat{v}_x(i)\hat{v}_y(i)(\hat{v}_y(i+1) - \hat{v}_y(i))).$$

Here $\mathbf{x}_i = (x_i, y_i)$, $\hat{v}_x(i) = x_i - x_{i-1}$ and $\hat{v}_y(i) = y_1 - y_{i-1}$. A similar calculation for (N3) leads to the statistic

$$T_3 = x_N - x_1,$$

th total change in the x-coordinate. One important consideration that may affect the preferred method of analysis is whether the model is supposed to describe individual or population behavior. If the model class is parameterized by $\theta \in \Theta$, one could suppose that the population is homogeneous, and each individual has the same parameter θ. If the population is inhomogeneous, with each cell having its own parameter θ' arising from some distribution on Θ, a parameter such as $\theta = \mathbb{E}(\theta')$ can be used to describe the population. For the data considered here, one might suppose that the directional behavior of all cells is similar, but cells vary in their speed. This might

suggest rescaling T_3 to give a statistic

$$T_4 = \frac{x_N - x_1}{|\mathbf{x}_N - \mathbf{x}_1|}.$$

T_4 is the same statistic that was introduced under the different name of Z_2 in Section 2.3, and is the cosine statistic favored in practice by biologists. Based on the discussion of cell speed in Sections 2.1 and 2.2, a more natural way to scale T_3 to allow for inhomogeneous cell speeds might be

$$T_5 = (x_N - x_1) \Big/ \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} |\mathbf{x}_{i+1} - \mathbf{x}_i|^2}.$$

The t-statistics and corresponding p-values for these four test statistics applied to a data set collected by Dr. K. Fang are presented in Table 2.1. The treatment group here consisted of a total of 24 cells prepared on three microscope slides, as previously described in Example 1 of Section 1.1, and exposed to a direct current electric field of 100 mV/mm. The 40 control cells were similarly prepared, on four slides, but were not exposed to the electric field.

	treatment group		control group		t-test
	sample mean	SE	sample mean	SE	p-value
T_2	53.5	18.2	29.8	16.3	0.354
T_3	21.7	5.2	3.6	4.6	0.015
T_4	0.490	0.126	0.095	0.120	0.035
T_5	2.21	0.50	0.22	0.49	0.010

Table 2.1: The statistics T_2, T_3, T_4, T_5 were calculated for each of the 24 cells in the treatment group and 40 in the control group. A two sample t-test was carried out to test the hypothesis that the treatment had no effect, against a general alternative hypothesis. T_2 and T_3 have units of pixels (1 pixel $\approx 1\mu m$). T_4 and T_5 are dimensionless quantities.

All the test statistics are comfortably positive for the treatment group, indicating a preference for the cells to move toward the cathode (negative pole) of the electric field. The statistic T_2 has its treatment group average the fewest SE's from zero, and also detects possible indications of a similar effect for the control group, leading to a large p value for the difference between treatment and control.

On this data set, the approximate score statistic for model (N3), T_3, and is scaled version, T_5, both showed strong statistical evidence for a difference between treatment and control groups. The cosine statistic, T_4, used previously by biologists, gives slightly weaker evidence. The further exploration of the data set in Section 2.4 will help to explain these differences.

2.4   Parameter Estimation for Models of Cell Translocation

Various methods based on moments have been proposed to estimate parameters for models of cell motion such a model (N1) of Section 2.3 (Dickinson and Tranquillo, 1993; DiMilla et al., 1992). In fact, parameter estimation for (N1) is a well-studied statistical problem, as it is a linear, Gaussian state space model. Maximum likelihood estimates are efficient and may be computed using the Kalman filter, as described in Harvey (1989). Model (N3) is similarly linear and Gaussian. For nonlinear models, such as (N2), the likelihood may be calculated using sequential Monte Carlo, as described in Section 1.3. The maximum smoothed likelihood estimator (MSLE) developed in Chapter 3 provides an effective way to estimate parameters and their uncertainties from a Monte Carlo likelihood function which had considerable simulation error even for long computation times. A trick that is available for some Monte Carlo estimation methods, such as the state space methods of Durbin and Koopman (2000), is to use the same seed for the random number generator at each parameter value. When the Monte Carlo likelihood is a continuous function of the parameters for any fixed sequence of random numbers, this trick allows standard numerical maximization techniques to be applied. Sequential Monte Carlo, however, uses random numbers for sequential resampling which leads to a highly discontinuous Monte Carlo likelihood function even for a fixed sequence of random numbers. For sequential Monte Carlo, the simulation error must therefore be dealt with directly.

For the following analysis, the observation error parameter was set to $\tau = 3$, measuring in pixel units (1 pixel $\approx 1\mu m$). This value was based on inspecting the

results of the segmentation algorithm, and in particular noticing that the SD of the location of certain cells which moved little and were presumed to be dead was around this value. The physical interpretation of τ is as the standard deviation of the measurement error. However if τ is estimated from the full data it can compensate for model mis-specification. Large cell displacements, which are rare for the modeled velocity, may be assigned as large observation errors rather than scientifically relevant events.

Estimates of the parameters α, σ and β provide a means to quantify persistence, speed and directionality even when the model is open to question. When the accuracy of the model is in some doubt, but the parameter estimation is still meaningful, the error estimates arising from the Fisher information can be misleading. A simple example is using the sample mean to describe the center of a distribution based on independent draws from a distribution which is modeled as $N(\mu, 1)$ but is in fact $N(\mu, \sigma^2)$ for $\sigma^2 \neq 1$. A solution to this difficulty is to use the form of the error resulting from thinking of the estimate as the root of an estimating equation (Basawa et al., 1997). If n i.i.d. observations $x_1, \ldots, x_n$ are made from a distribution on $\mathbb{R}^d$ with density $f(x \mid \theta)$ for $\theta \in \Theta \subset \mathbb{R}^k$, giving rise to an MLE of $\hat{\theta}$, two estimates of the Fisher information are

$$
\begin{aligned}
I_1 &= \sum_{i=1}^{n} \left(\frac{\partial}{\partial \theta} \log f(x_i \mid \theta)|_{\theta=\hat{\theta}} \frac{\partial}{\partial \theta} \log f(x_i \mid \theta)|_{\theta=\hat{\theta}}^T \right) \\
I_2 &= -\sum_{i=1}^{n} \frac{\partial^2}{\partial \theta \partial \theta^T} \log f(x_i \mid \theta)|_{\theta=\hat{\theta}}.
\end{aligned}
$$

The so-called sandwich estimator of the covariance matrix of $\hat{\theta}$ is

$$
\hat{\Gamma} = I_2^{-1} I_1 I_2^{-1}. \tag{2.6}
$$

This estimator is robust to model mis-specification, provided the observations are independent and the required Taylor series expansion and application of the central limit theorem of White (1982) are justified. The natural extension of this result from MLE to MSLE is to replace derivatives of $\log f(x \mid \theta)$ by the derivatives of a smooth approximation to the likelihood. A theoretical justification for this is left as an open

problem. For a maximum quadratic likelihood approximation estimator (MQLE), calculated as below, this error estimate takes an appealing form. Suppose n i.i.d. random variables are observed, with log likelihoods $\{\lambda_i(\theta),\ 1 \leq i \leq n\}$ giving rise to a likelihood function

$$\lambda(\theta) = \sum_{i=1}^{n} \lambda_i(\theta).$$

Let $\bar{\theta}$ be a preliminary estimate of θ and let $G \subset \Theta \subset \mathbb{R}^k$ be a grid of points around $\bar{\theta}$, as defined and used in Section 3.1 below. Use least squares to fit a quadratic function to $\{\lambda(\theta^*),\ \theta^* \in G\}$, giving rise to a symmetric matrix Q, a vector b and constant c such that

$$\lambda(\theta) \approx (12)\theta^T Q \theta + \theta^T b + c.$$

If the model possesses LAN, one can hope that Q should be negative definite, in which case the MQLE, $\hat{\theta}$, is defined by

$$\hat{\theta} = -Q^{-1}b. \tag{2.7}$$

The linearity of least squares fitting gives an identity

$$b = \sum_{i=1}^{n} b_i,$$

where b_i comes from making a linear fit, using least squares, to $\lambda_i(\theta) - \frac{1}{2n}\theta^T Q \theta$ evaluated on G, written as

$$\lambda_i(\theta) - \frac{1}{2n}\theta^T Q \theta \approx \theta^T b_i + c_i.$$

This suggests an estimated covariance matrix $\hat{\Gamma}$ for $\hat{\theta}$, in terms of the empirical covariance, R, of $\{\sqrt{n}b_i,\ 1 \leq i \leq n\}$, given by the equation

$$\begin{aligned} R &= \sum_{i=1}^{n} b_i b_i^T - \frac{1}{n}bb^T \\ \hat{\Gamma} &= Q^{-1}RQ^{-1}. \end{aligned} \tag{2.8}$$

This construction of $\hat{\Gamma}$ is an extension of the sandwich estimator, given in equation (2.6), to MQLE. This error estimate can be seen to be robust to model misspecification in that it is based on an expression for the covariance that is true for

the sum of any i.i.d. random variables (though $\{b_i,\ 1 \le i \le n\}$ are of course only approximately i.i.d.).

Parameter estimates, with corresponding errors, are presented in Table 2.2 for the same experimental data discussed in Section 2.3, comparing a treatment group of human keratinocyte cells exposed to a DC electric field of 100 mV/mm. with a control group. Model (N1) was only fitted to the control group, as it does not allow for any rotational asymmetry. Both models (N2) and (N3) have parameter estimates differing between treatment and control most noticeably in the asymmetry parameter, β. This supports the belief that the parameters α and σ change little in the presence of an electric field of 100 mV/min. The error estimates labeled SE_1 in Table 2.2 is based on the Hessian estimator, Q^{-1}, for the variance of the parameter vector $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\sigma})^T$, coming out of the general theory of maximum likelihood estimation (White, 1982). SE_2 comes from the sandwich estimator $Q^{-1}RQ$ of equation (2.8). SE_2 is larger than SE_1 in all cases other than for the estimate of α in model (N1). For the estimates of β the discrepancies between SE_1 and SE_2 are not large. The largest discrepancies occur for estimates of σ, and this may be partially explained by recalling that the observation noise parameter was fixed for convenience at $\tau = 3$. Assuming that we know the exact size of the observation noise might be expected to reduce the error in estimating the innovation noise, σ. SE_2, which allows for the possibility of model mis-specification such as $\tau \ne 3$, and which is constructed in a way similar to the usual standard error on a mean via the sample variance, is preferred and will be the error estimate used subsequently.

The estimates of β divided by their SE's (i.e., in SE units) for models (N2) and (N3) are similar to the means of the statistics T_2 and T_3 in SE units. The same comments therefore apply as in Section 2.3, that the model (N3) discriminates more strongly between the asymmetry of the treatment and control groups than does (N2). This may lead one to suspect that the data are better represented by (N3) than by (N2). There are many formal and informal ways of assessing model fit. Formally, one can test model specification by comparing the sandwich estimator of the variance of the parameter estimate with the Hessian estimator (White, 1982). A class of models may be compared with a larger family of models including the original class (Box

		treatment			control		
		estimate	SE_1	SE_2	estimate	SE_1	SE_2
N1	α				0.578	0.064	0.060
	σ				7.15	0.49	0.76
N2	α	0.559	0.097	0.120	0.609	0.066	0.069
	β	0.455	0.125	0.157	0.212	0.100	0.116
	σ	7.94	0.66	1.29	7.34	0.50	0.78
N3	α	0.523	0.118	0.177	0.592	0.092	0.130
	β	4.93	0.93	1.02	0.85	0.71	0.78
	σ	7.22	0.72	1.39	7.24	0.56	0.93

Table 2.2: Parameter estimates for models (N1), (N2) and (N3) based on a treatment group of 24 cells and control group of 40 cells. The MQLE method of equation (2.7) was used. SE_1 corresponds to the Hessian estimator Q^{-1} and SE_2 is the sandwich estimator $Q^{-1}RQ$ of equation (2.8).

and Jenkins, 1970). Estimates, $\{\hat{\epsilon}_j^{(i)}\}$, of the measurement errors, $\{\epsilon_j^{(i)}\}$, are termed *residuals* and may be plotted against possible covariates of interest such as time and the microscope slide label. The clearest evidence found by the author for preferring model (N3) to (N2) for the data presented in this thesis is a plot of magnitude of cell displacement against direction, displayed in Figure 2.3. This plot, shown for the treatment group of cells and for simulations under models (N2) and (N3) with their fitted parameter values, demonstrates qualitative agreement between the data and model (N3). In both cases there are relatively few occasions when cells travel more than 5 pixels in the direction of the anode. Model (N2), however, acquires its anisotropic behavior just by having fewer occurrences of cells traveling in an anodal direction. Those cells that travel toward the anode for (N2) do so with almost as large displacements as those traveling toward the cathode. Observations have been made in the biological literature that directional factors such as DC electric fields often do not affect cell speed and persistence (Nishimura, 1996). Here we find that this is true in the sense that estimates of σ and α are similar for treatment and control, but not in the alternative sense (which is a property of model (N2)) that the marginal distribution of cell speed is independent of cell direction.

This section has demonstrated a methodology suitable for fitting a wide class of

nonlinear models to cell translocation data. For the two models considered in detail here, the linear Gaussian model, (N3), appears more satisfactory than the nonlinear model (N2). However, there are many different types of cells and situations in which one might be interested in studying their motion. It is reassuring to be able to write down, fit and assess plausible nonlinear models, even if only to add evidence for the adequacy of linear ones.

2.5   A Statistical Analysis of Cell Shape

In this section, two models for cell shape are introduced and compared. In each, the shape is defined by a boundary function in center-radius form. This form was developed in Section 1.4, together with comments on its strengths and limitations. First we describe a model based on the principle of local autocatalysis and long range inhibition for the processes governing the extension of the cell boundary. Secondly we introduce a model arising from a cytomechanical model based on equations of fluid dynamics. The forms of these models which appear here are intended to be phenomenological, meaning that they are simplified caricatures of a complex system. The system is considered as a "black box", which one tries to understand by combining physical insights with observational data. A data analysis is carried out to investigate the applicability of these models in a practical setting.

A principal that has been widely used to model biological pattern formation, such as leopard spots and butterfly wings (Murray, 1989), and various other developmental processes such as exotic sea shell patterns (Oster, 1990) is that an agent acts to enhance existing features of the process while suppressing the appearance of new features in some surrounding region. This principal may be applied to cell motion, by modeling the creation, interaction and dispersion of the protrusions that cells employ to locomote. In terms of a radial description $r_t(\phi)$ of cell shape around a center $\mathbf{c}_t$, a model for local autocatalysis and lateral inhibition for a cell moving in an isotropic environment can be written in terms of $a_t(\phi) = \log(r_t(\phi)/\tilde{r})$, where $\tilde{r}$ is

a characteristic radius of the cell, as

$$\text{(S1)} \qquad da_t(\phi) = \left(\int_{-\pi}^{\pi} -h(\psi)a_t(\phi + \psi)d\psi \right) dt + dx_t(\phi).$$

Here $-\pi \in \phi$, $\psi < \pi$ and addition of angles is modulo 2π. The process $x_t(\phi)$ is taken to be a Gaussian process, whose increments are stationary in time and whose distribution is rotationally invariant in ϕ. It would be of interest to extend (S1) to anisotropic situations, for example replacing $h(\psi)$ by $h(\psi, \phi)$ and/or removing the rotational invariance requirement on $x_t(\phi)$, as this would enable a quantification of the anisotropic behavior. A possible parametric form for h is the difference of two Gaussian curves, or "Mexican Hat",

$$h(\phi) = a_1 f(\phi/b_1) - a_2 f(\phi/b_2) \tag{2.9}$$

with $f(\phi) = (1/\sqrt{2\pi})e^{-\phi^2/2}$ and $a_1 > a_2 > 0$, $b_2 > b_1 > 0$.

A convenient way to study rotationally invariant processes on the circle is through the Fourier transform. This is discussed in more detail in Section 4.2. We write $A_t(k)$, $X_t(k)$ and $H(k)$, $k = 0, 1, 2, \ldots$, for the Fourier transform of $a_t(\phi)$, $x_t(\phi)$ and $h(\phi)$. Further supposing that $X_t(k)$ can be written as $\sigma(k)W_t(k)$ for a series of independent complex valued Brownian motions $\{W_t(k), \ k = 0, 1, 2, \ldots\}$, (S1) becomes

$$dA_t(k) = -H(k)A_t(k)dt + \sigma(k)dW_t(k). \tag{2.10}$$

This can be recognized as the infinitesimal equation for a complex valued Ornstein–Uhlenbeck process.

The frequency domain form of (2.9) is

$$H(k) = a_1 b_1 \phi(b_1 k) - a_2 b_2 \phi(b_2 k).$$

For stability of $A_t(k)$ we require $H(k) > 0$ which is satisfied if $a_1 b_1 > a_2 b_2$. When the continuous time model (S1) is observed at discrete time points $t_0, t_1, \ldots, t_{N-1}$, equally spaced with separation $t_n - t_{n-1} = \Delta$, the evolution equation for the discretely observed shape function $A_n^{\Delta}(k) = A_{t_n}(k)$ is given by

$$A_n^{\Delta}(k) = e^{-\Delta H(k)} A_{n-1}^{\Delta}(k) + \sqrt{(-\sigma^2(k)/2H(k))(1 - e^{-2\Delta H(k)})}\epsilon_n(k), \tag{2.11}$$

where $\{\epsilon_n(k)\}$ are i.i.d. standard normal random variables. The two parameters defined by $H^\Delta(k) = e^{-\Delta H(k)}$ and $\sigma^\Delta(k) = \sqrt{(-\sigma^2(k)/2H(k))(1 - e^{-2\Delta H(k)})}$ may be estimated for each k using standard statistical computing packages, such as S-Plus, since (2.11) is an AR(1) process.

Another way to arrive at a phenomenological model for cell shape is to take a simplified form of a cytomechanical model that claims to represent the biophysical processes involved in cell motion. A stochastic model for receptor-mediated cytomechanics is developed in Tranquillo and Alt (1996). They propose a model for $a(t, \phi) = \log(r(t, \phi)/\tilde{r})$ given by the partial differential equation

(S2) $$\frac{\partial a}{\partial t} + \frac{1}{c_1}\frac{\partial^2}{\partial \phi^2}\left(\frac{c_2}{2}a + c_3\ln\left(1 - \frac{a}{c_4}\right) + \frac{c_5}{2}\frac{\partial^2 a}{\partial \phi^2}\right) = x(t, \phi).$$

The left-hand side models the physical properties of the cell, using a simplified form of the two-phase fluid model for the cell cytoplasm of Dembo (1989). The constants $c_1, c_2, \ldots, c_5$ have interpretations within this model. The right-hand side of (S2) is supposed to represent receptors on the cell membrane which, when activated, drive the cell motion process. When the cell is in a homogeneous environment, $x(t, \phi)$ may be modeled by Gaussian white noise. The model (S2) then becomes a stochastic partial differential equation, as studied in, for example, DaPrato and Zabczyk (1992). We avoid complications by considering only a linearized form of (S2), which in the frequency domain representation can be written

$$dA_t(k) = -H(k)A_t(k) + \sigma dW_t(k), \qquad (2.12)$$

where

$$H(k) = C_1 k^2 + C_2 k^4$$

for

$$C_1 = \frac{c_2}{2c_1} - \frac{c_3}{c_4}, \ C_2 = \frac{c_5}{2c_1}.$$

Here $\{W_t(k)\}$ is a collection of independent Brownian motions, indexed by $k = 0, 1, 2, \ldots$. Note that this model specifies that σ is a fixed constant, independent of k. For stability we require $H(k) > 0$ for all k, and so $C_2 > 0$, $C_1 > -C_2$.

Data on the cell shape process are available as a byproduct of the cell tracking algorithm described in Section 2.3, and some examples are presented in Figure 2.4. A way to relate these data to (S1) and (S2) is to estimate the coefficients $H^\Delta(k)$ and $\sigma^\Delta(k)$ for each k. The Fourier representation of cell shape is used, and compared to alternatives, in Brosteanu et al. (1997). The models considered here can be thought of as simple candidates to describe the evolution in time of the Fourier representation of cell shape. Figures 2.5 and 2.6 show estimates of $\hat{H}^\Delta(k)$ and $\hat{\sigma}^\Delta(k)$ arising from the data in Figure 2.4. An investigation of the residuals $\{\hat{\epsilon}_n(k)\}$ suggests only minor deviation from normality (checked by normal quantile plots such as Figure 2.7) and little autocorrelation (checked by autocorrelation plots, not shown).

A noticeable feature of the plot of $\hat{H}^\Delta(k)$ is that it descends down to around zero for higher frequencies. This corresponds to $H(k)$ becoming large, which qualitatively favors (S2) over (S1). For (S1), on a scale much smaller than that of $h(\theta)$ the shape process should be only lightly damped. On the other hand, model (S2), in the form of (2.12) with only three parameters, cannot explain features of the data such as the high value of $\hat{H}^\Delta(k)$ at $k = 4$. Although the models (S1) and (S2) give different, and complementary, ways to interpret the shape process data presented in Figure 2.4, neither explains the whole story. The cell shape process and is decomposition into frequency components remain descriptive statistics, waiting for a parametric form to accompany them.
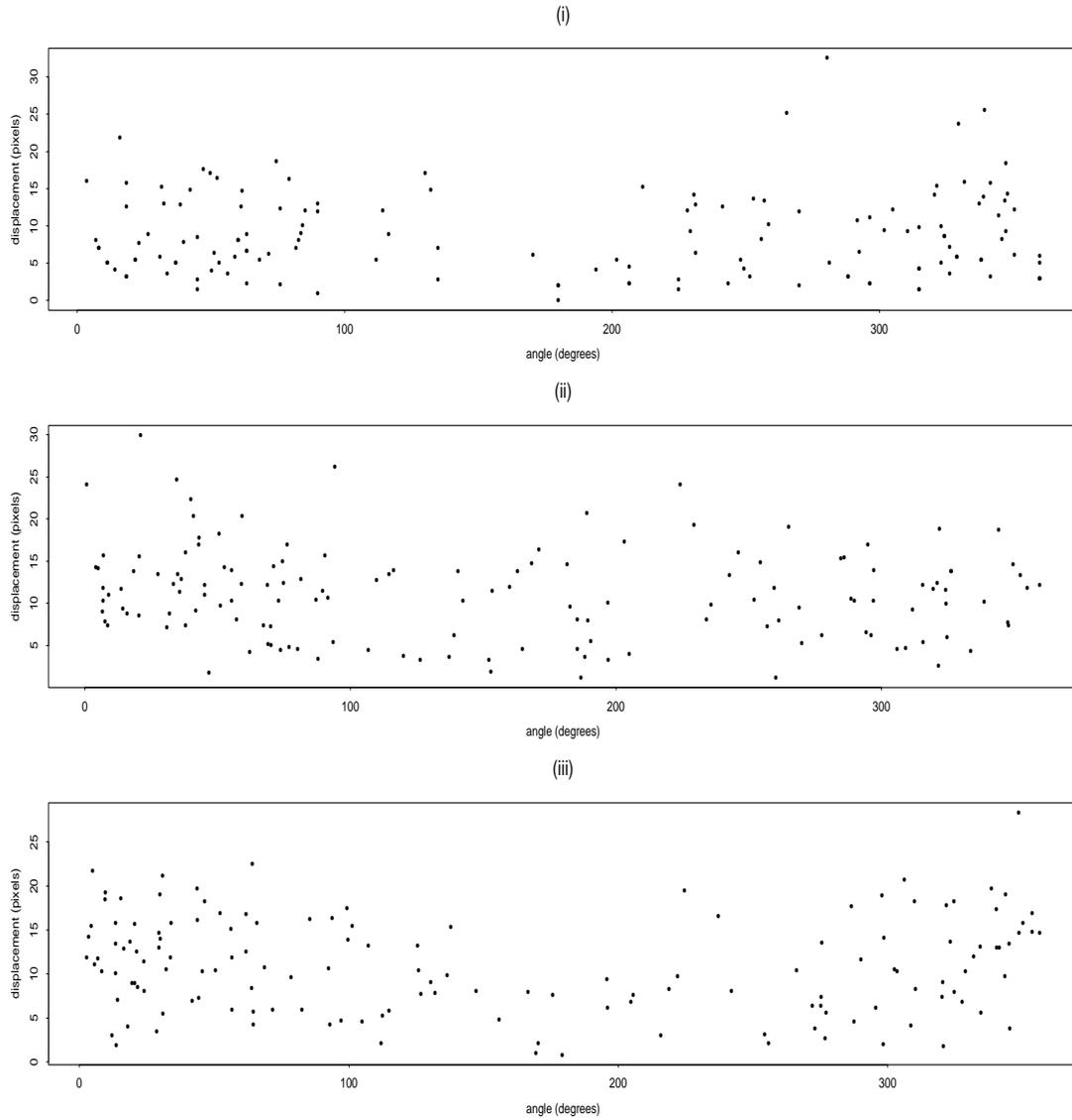
Figure 2.3: Setting $\mathbf{x}_{j+1}^{(i)} - \mathbf{x}_j^{(i)} = (r_j^{(i)}\cos\theta_j^{(i)}, r_j^{(i)}\sin\theta_j^{(i)})^T$, the displacement, $r_j^{(i)}$, is plotted against the angle, $\theta_j^{(i)}$, for each cell i and each time point j. (i) The treatment group. (ii) Simulated data for model (N2), using the fitted parameter values. (iii) Simulated data for model (N3), using the fitted parameter values.

Figure 2.4: The shapes of ten cells from a control experiment, with no electric field. The cell shapes were recorded at seven equally spaced time points (10 minutes apart), using the algorithm for cell tracking described in Section 2.3.
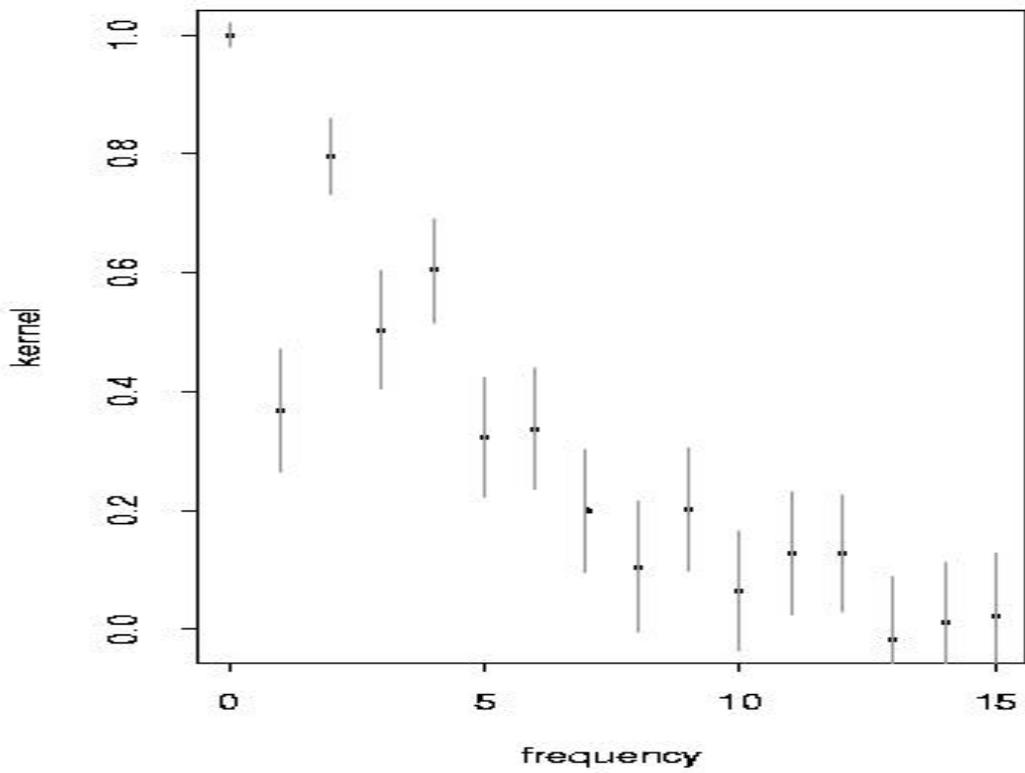
Figure 2.5: Estimates of the convolution kernel, $H^{\Delta}(k)$, at frequency k, with 95% pointwise confidence intervals given by the error bars.
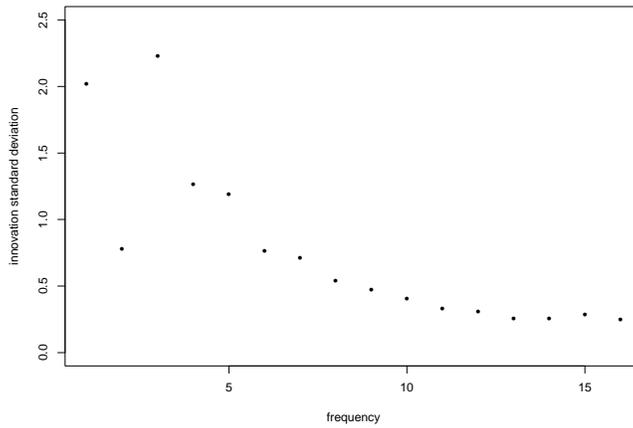
Figure 2.6: Estimates of the innovation standard deviation, $\sigma^\Delta(k)$, at frequency k.
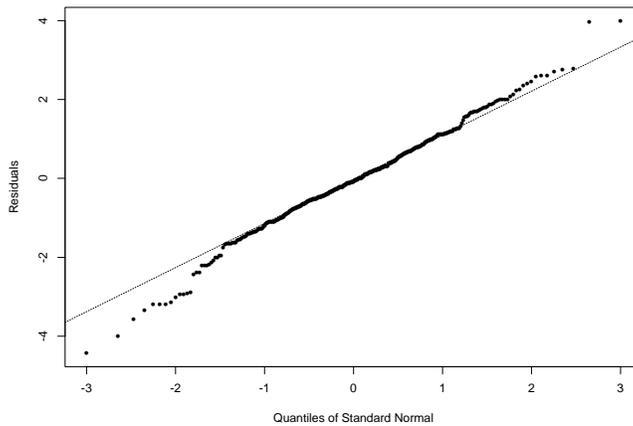
Figure 2.7: A normal quantile plot, for frequency $k = 3$, of the residual process $A_n^\Delta(k) - \exp(-\Delta H(k))A_{n-1}^\Delta(k)$. Both tails for the residuals are seen to be slightly, but noticeably, longer than those of the normal distribution. This indicates that allowance for non-normality might be desirable but would be expected to make little difference in the estimated quantities.

Chapter 3

Asymptotic Theory for Maximum Smoothed Likelihood Estimation and an Application to State Space Models

This chapter develops some asymptotic theory which assists the statistical analysis of Chapter 2 while making a widely applicable contribution to the general theory of maximum likelihood estimation and of state space models. The original motivation of this work was the observation that the best results currently available for the asymptotic properties of the maximum likelihood estimator for general non-linear state space models place heavy restrictions on the form of the model (Bickel et al., 1998; Jensen and Petersen, 2000). An attractive framework for finding a more widely applicable result is based on the local asymptotic normality (LAN) property of Le Cam (1986). Some progress on this project is made in Section 3.2, for a situation where the state process is a discretely observed diffusion process. A more general result based on the preservation properties of LAN under information loss (Le Cam and Yang, 1988) is still an open problem.

The relevance of the LAN framework to applied statistical practice is an important issue to discuss before embarking on a voyage toward this asymptotic limit. This is

a question that has not previously been extensively explored, beyond the comments in Le Cam (1990), so Section 3.1 takes considerable care to motivate the LAN framework introduced. This is done by showing that the LAN property confers desirable asymptotic results on a class of estimators resulting from maximizing a smoothed version of the likelihood in a neighborhood of an effective preliminary estimator.

Maximizing a smoothed version of the likelihood is discussed in Small et al. (2000), in the context of eliminating multiple root problems for estimating equations. Daniels (1960) proposed applying a kernel smoother to the likelihood function to aid numerical evaluation of the maximum. Barnett (1966) found in a simulation study that the method of Daniels could exceed the efficiency of the maximum likelihood estimator (MLE) for a Cauchy location model by up to 10%. Kreimer and Rubinstein (1988) considered smoothing as a general technique for numerical maximization of functions.

Heuristically, there are two main problems that can arise with the maximum likelihood estimator but which an LAN based approach avoids. These are demonstrated in the two examples below.

Example 1 (Normal mixture). A simple situation demonstrating an unbounded likelihood function is the mixture of two normal distributions of ???. Any mixture model can be written as state space model, and this is carried out here since state space models are considered in Section 3.2. The state vector is $X_i = (X_i^{(1)}, X_i^{(2)})$, $1 \leq i \leq n$, where $\{X_i^{(1)}\}$ are i.i.d. $N(\mu, 1)$ and $\{X_i^{(2)}\}$ are i.i.d. $N(\mu, \sigma^2)$ independent of $\{X_i^{(1)}\}$. The observed variables are

$$Y_i = \begin{cases} X_i^{(1)} \text{ with probability } 1/2 \\ X_i^{(2)} \text{ with probability } 1/2 \end{cases}$$

for $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$. Then the likelihood

$$L_n(\mu, \sigma) = \prod_{i=1}^n \left(\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-(Y_i-\mu)^2/2} + \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(Y_i-\mu)^2/2\sigma^2} \right)$$

has an infinite supremum for $\mu = Y_1$ and $\sigma \to 0$.

When fitting such a mixture model, this poor behavior of the MLE can be avoided by maximizing the likelihood in a region where σ is bounded away from zero. This

solution is practical, but theoretically inelegant, and augurs poorly for the existence of good general results for theoretical properties of the MLE in state space models. On the other hand, LAN can be shown to hold for Example 1 using Le Cam's condition of differentiability in quadratic mean introduced in Section 3.2.

When the likelihood function has many local maxima, the global maximum may correspond to a narrow spike distant from the main concentration of the likelihood. This can lead to poor performance of the MLE, and also may make the likelihood function hard to maximize numerically. Furthermore, error estimates based on the second derivative of the likelihood at the maximum are then not applicable. One way to get around these problems is to make a quadratic approximation to the log likelihood in a neighborhood of the true parameters value in a way that avoids small-scale features of the likelihood function. One way of doing this is the so called one-step estimator of Le Cam and Yang (1990). In this thesis, the one-step estimator is more conveniently called the maximum quadratic likelihood approximation estimator (MQLE), and is demonstrated in Example 2 for a shift parameter estimation problem.

Example 2 (Many local maxima). The shift family with densities on $\mathbb{R}$ given by $f(x \mid \theta) \propto \exp(-|x - \theta|^{\alpha})$ with respect to Lebesgue measure can be shown to possess LAN for $\alpha > \frac{1}{2}$ using the criterion of differentiability in quadratic mean discussed in Section 3.2. For $\frac{1}{2} < \alpha \leq 1$ it does not satisfy the Cramér conditions (Cramér, 1946) for the MLE to attain asymptotically the Cramér–Rao bound. An example of the likelihood function based on a simulated sample of size 100 with $\theta = 0$ and $\alpha = 0.6$ is shown in Figure 3.1. The MLE and median are shown, together with the MQLE calculated by the method employed by Le Cam's one-step estimator: a quadratic was fit using the likelihood function values evaluated at the median and points ± 0.3 from the median. This example serves to illustrate the difference between the one-step estimator and an iteration of Newton–Raphson maximization: the quadratic approximation is carried out on a scale that captures the overall shape of the likelihood, without being distracted by small-scale behavior. An iteration of Newton–Raphson, on the other hand, uses a quadratic approximation of a smooth function at an initial estimate to attempt to approach a local maxima of the function. This gives one
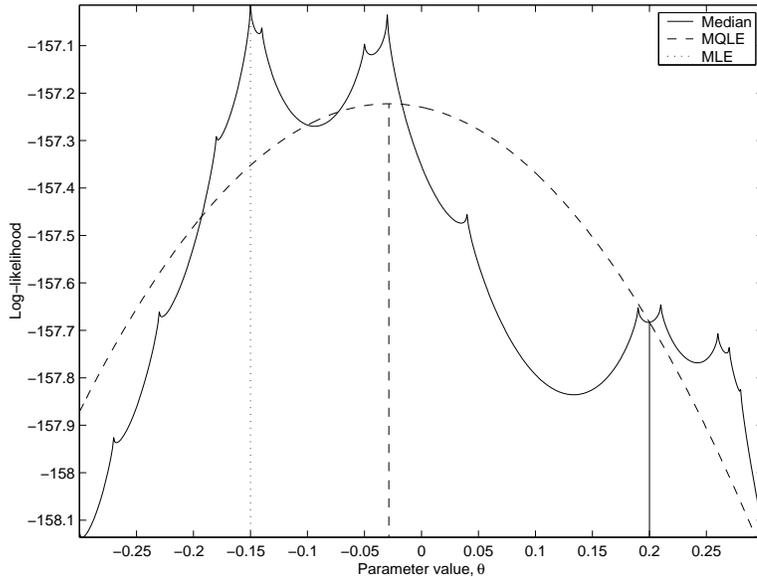
Figure 3.1: The log-likelihood function for a sample of size 100 drawn from the density $f(x|\theta) \propto \exp(-|x - \theta|^{\alpha})$ with $\theta = 0$. The median, MLE and MQLE are shown, together with the approximating quadratic.

answer to the question "Why not repeat the one-step estimator twice?" A more convincing explanation of parameter estimation in the LAN framework is given using the concept of maximum smoothed likelihood in Section 3.1 below.

3.1 Maximum Smoothed Likelihood Estimation

Defining an experiment as a family of probability measures $\{P_\theta, \theta \in \Theta\}$, an extensive theory is detailed in Le Cam (1986) for the convergence of experiments to a limit experiment. An important case is convergence to a Gaussian shift experiment, where statistical inference for θ becomes asymptotically equivalent to estimating the mean of a certain Gaussian random variable. A precise treatment of these concepts, which is not required for reading this thesis, can be found in Le Cam (1986) or Le Cam and Yang (1990). A widely used example of convergence to a Gaussian shift experiment is the condition of Local Asymptotic Normality (LAN), defined below. LAN has found use in many theoretical situations (Bickel et al., 1993; Hallin et al., 1999; Bickel and

Ritov, 1996; Höpfner et al. 1990; Jeganathan, 1995). Later in this section conditions will be found for LAN to hold for the models introduced in Section 2.1. First I am going to argue for the importance of LAN to applied statistics.

The key to understanding the role of LAN in applied statistics is its close relationship to maximum likelihood estimation (MLE). A difficulty with comparing these two concepts is that LAN is an asymptotic property of the likelihood function whereas MLE is a parameter estimation procedure. To level the playing field we must introduce two more acronyms. Write MLAN for the property that the MLE exists, is consistent, and is asymptotically normal, with variance equal to the Cramér–Rao lower bound. An estimation procedure based on the LAN property, for reasons described later, will be called maximum smoothed likelihood estimation (MSLE). Clearly, LAN is most appropriately compared to MLAN and MSLE to MLE.

LAN is a weaker property than MLAN, in the sense that the commonly used sufficient conditions are weaker. In the case of i.i.d. random variables, the Cramér conditions for MLAN imply a condition of differentiability in quadratic mean which in turn implies LAN (Le Cam and Yang, 1990, p. 102). Heuristically, both LAN and MLAN ensure that the log likelihood ratio in a neighborhood of the true parameter value θ_0 is asymptotically approximately quadratic in θ. MLAN further requires that the likelihood function be sufficiently smooth and satisfy a global constraint that the likelihood function should not grow too large for θ distant from θ_0. The surprise about LAN is that it confers similar asymptotic optimality properties on appropriately constructed estimators, confidence regions and tests to those provided by the stronger condition MLAN. A discussion with some more details may be found in Le Cam and Yang (1990, Section 5.8), which builds on a result of Le Cam (1986, Theorem 1 of Section 7.4) that if a sequence of experiments has a Gaussian limit experiment then one cannot achieve asymptotically risk functions that are not achievable on the Gaussian limit.

Although we have seen that the local likelihood approximation provided by LAN has a solid foundation in theory, applications have been restricted by the perceived lack of a practically justified estimator based on the LAN property. In fact, we are going to introduce such a maximum smoothed likelihood estimator (MSLE) and show

that in many cases it corresponds closely to the method used in practice by responsible statisticians claiming to be calculating the MLE.

A practical procedure for likelihood-based parameter estimation from a complicated likelihood function might include the following steps

(P1) Take several starting values, θ_k, $1 \leq k \leq K$. Hopefully knowledge of the particular application will suggest some reasonable values of θ_k. These might also come from the method of moments, a convenient but usually inefficient estimation procedure, discussed in Basawa et al. (1997).

(P2) For each θ_k, run a numerical optimization procedure starting at θ_k to attempt to find the maximum of the likelihood function. Hopefully this algorithm will terminate under a reasonable convergence criterion to give an estimate $\hat{\theta}_k$.

(P3) If all the $\hat{\theta}_k$ are close, use their common value $\hat{\theta}$ for an estimate of θ_0. An estimate of the error on $\hat{\theta}$ can come from numerical calculation of the second derivative of the likelihood function at $\hat{\theta}$, using asymptotic properties of the likelihood function. If and when there are enough data, it may be preferable to calculate an error estimate in a more data-driven way, such as bootstrap and jack-knife methods.

(P4) If the values of $\hat{\theta}_k$ for $1 \leq k \leq K$ vary considerably, try to use knowledge of the subject matter, the form of the likelihood function and the numerical algorithm used to understand why. Possibly one or more of the starting values may be rejected as unreasonable.

(P5) In the event of either (P3) or (P4), it will do no harm to plot the region of interest of the likelihood function, or to try to find some graphical representation such as marginal plots if the parameter space is of too high dimension to allow a standard plot.

If MLAN is proved for some asymptotic limit of the model in question, then the statistician who follows the above procedure can sleep at night safe in the knowledge that there is probably not much better that could have been done. He can claim to

have approximated the MLE, which has asymptotic optimality properties. It only remains to check the modeling assumptions and perhaps do some simulations to investigate the finite sample properties of his estimation algorithm and compare it to competitors.

The MSLE is defined to be the value of θ maximizing a smooth approximation to the log likelihood coming from evaluation of the likelihood function on a finite grid, G, of points which with high probability lie in a neighborhood of θ_0 and which are a subset of a discretization Θ^* of Θ. Further notation and details follow later. The MSLE is an extension of the method of centering variables described in Le Cam and Yang (1990, Section 5.3). The method of centering variables, also called Le Cam's one-step estimator, and here the MQLE, attempts to reconstruct from the likelihood function the approximating quadratic whose existence is asymptotically assured by LAN. Fitting a second degree polynomial is a special case of smoothing the log-likelihood, and so MQLE is a special case of MSLE. An auxiliary estimator $\bar{\theta}$ is required to allow the identification of a grid, G, and a sufficient number of points from G are used to fit a second degree polynomial with a symmetric quadratic term (alternatively, the quadratic term may be assumed to take on its known asymptotic value at $\bar{\theta}$). The quadratic approximation estimator is the value $\tilde{\theta}$ at which this quadratic is maximized. The theory of Le Cam (1986, Chapter 11) shows that under LAN the sequence of experiments corresponding to observing $\tilde{\theta}$ has the same Gaussian limit experiment as the original sequence of experiments, which provides asymptotic optimality properties for $\tilde{\theta}$ and associated tests and confidence intervals.

The MSLE is strikingly similar to the method carried out in (P1)–(P5). In particular, a plot of the likelihood function in a region around the estimated value gives an approximation to the likelihood function based on evaluations on a grid of points, as used for the MSLE. The requirement that G take values on a discretization Θ^* prevents evaluation of the likelihood function at particular points where it might be badly behaved. Such a discretization will be unimportant if the likelihood is smooth, but if not it provides a useful trick to avoid evaluating the likelihood at particular points where the likelihood may have peculiarities (for example, the median or an MLE of a location parameter for a density with a singularity).

The similarity between MSLE and good practice is strong enough that MSLE theory can help support practice. Although people who have tried to use numerical methods to maximize a likelihood know how important the initial values can be, having the requirement of a good auxiliary estimator cast into the theory suggests that more time should be spent justifying initial values than, say, arguing for the smoothness of the likelihood function which turns out to be unnecessary, if a discretization trick is used.

Another major reason to prefer the LAN/MSLE theory to MLAN/MLE is that in many situations MLAN does not hold. A humorous account of things that can go wrong with MLE is given in Le Cam (1990). The well-known sufficient conditions for MLAN are restrictive, and not easy to check for complicated likelihood functions. In fact it is quite possible that the statistician employing (P1)–(P5) slept soundly not because he had proven that MLAN holds in his model, but because MLE is a principle which is frequently justified, in a somewhat circular argument, by its widespread use. However, since the method (P1)–(P5) more closely resembles MSLE, the statistician's application of MLE could in fact be marked up as another successful use of MSLE.

Finally, suppose that our statistician has a theoretical bent, and manages to show that MLAN does indeed hold for his model. He could have saved himself time and effort by checking the weaker condition of LAN instead. For an example of this, compare the two papers Bickel and Ritov (1996) and Bickel, Ritov and Ryden (1998), in which conditions are found for hidden Markov models to have LAN and MLAN respectively. Another example of LAN being considerably more convenient to work with than MLAN, that will play a role in this thesis, is the preservation of LAN under information loss in quite general situations (Le Cam and Yang, 1988). No such results appear to exist for MLAN.

The time has come to formalize some of the concepts introduced above. A family $\{P_\theta, \theta \in \Theta\}$ of probability measures on $(\Omega, \mathcal{F})$ indexed by a set Θ forms an experiment in the sense of Le Cam (1986, Chapter 1). Following Le Cam and Yang (1990, Chapter 5) we shall take Θ to be an open subset of a fixed Euclidean space $\mathbb{R}^d$, and limit ourselves to indicating possible generalizations. The quantity $\frac{dP_\phi}{dP_\theta}$, $\phi \in \Theta$, denotes the Radon–Nikodym derivative of the absolutely continuous part of P_ϕ with

respect to P_θ. Now let $\{P_{\theta,n}, \theta \in \Theta\}$ be a family of measures on $(\Omega_n, \mathcal{F}_n)$ for each integer $n > 0$, and define the log likelihood ratio

$$\Lambda_n(\phi, \theta) = \log \frac{dP_{\phi,n}}{dP_{\theta,n}}.$$

To simplify notation we use the o_p convention. For a sequence $\{\xi_n\}$ of random variables and a real sequence $\{\alpha_n\}$, write $\xi_n \sim o_p(\alpha_n; \theta)$ if $\xi_n/\alpha_n \to 0$ in $P_{\theta,n}$ measure, i.e., for any $\epsilon > 0$, $\lim_{n\to\infty} P_{\theta,n}\{|\xi_n/\alpha_n| > \epsilon\} = 0$. Conventionally, one also abuses the notation by writing $\xi_n = o_p(\alpha_n; \theta)$.

Definition. The family $\{P_{\theta,n}, \theta \in \Theta\}$ has *local asymptotic normality* (LAN) at θ_0 if there exists a positive definite matrix K and a sequence of random variables $\{\Delta_n\}$ termed *centering variables*, such that for any bounded sequence $\{t_n \in \mathbb{R}^d\}$

(i)
$$\Lambda_n\left(\theta_0 + \frac{t_n}{\sqrt{n}}, \theta_0\right) = t_n^T K^{1/2} \Delta_n - \frac{1}{2} t_n^T K t_n + o_p(1; \theta_0)$$

(ii)
$$\Delta_n \xrightarrow{d} N(0, I) \text{ under } P_{\theta_0,n}.$$

Slight variations on the definition of LAN are used, with different tradeoffs between simplicity and generality, in Basawa and Prakasa Rao (1980), Ibragimov and Has'minski (1981), Le Cam (1986, Section 11.7) and Le Cam and Yang (1990). One can consider rates of convergence other than $\sqrt{n}$, or allow for random matrices K_n (the local asymptotic quadratic condition of Le Cam and Yang (1990)).

We proceed to construct an MSLE and show that it has the same asymptotic properties as the method of centering variables, while having more finite sample justification. Let $\Theta_n^* = \frac{1}{\sqrt{n}}\mathbb{Z}^d$ be a discretization of Θ for each integer n. For some $\sqrt{n}$-consistent estimator $\bar{\theta}$, let $\bar{\theta}^* = \arg\min_{\theta^* \in \Theta_n^*} |\bar{\theta} - \theta^*|$ and define a grid G_n to be the set
$$G_n = \bar{\theta}_n^* + \frac{1}{\sqrt{n}}\{-M, -M+1, \ldots, M-1, M\}^d$$
for some integer $M > 0$. The fixed integer M can be chosen so that the grid $G_n \subset \Theta_n^*$ surrounds the true parameter value θ_0 with high probability. For any $\epsilon > 0$, we assume that M and n_0 can be chosen so that $P_{\theta_0,n}\left\{|\bar{\theta}_n^* - \theta_0| < \frac{M\sqrt{d}}{\sqrt{n}}\right\} < \epsilon$ for all

$n > n_0$. This follows from, or can be taken as a definition of, the $\sqrt{n}$-consistency of $\bar{\theta}_n$.

A smoother, S, takes a function on a grid $G \subset \mathbb{R}^d$ to functions on $\mathbb{R}^d$. Let $G = \{-M, -M+1, \ldots, M-1, M\}^d$ and define an arbitrary function $g : G \to \mathbb{R}$ with $t^* \mapsto g(t^*)$. Suppose S has the following properties, for each $t \in \mathbb{R}^d$ and any fixed g:

(S1) $s(t) = S(g)(t)$ is continuously differentiable.

(S2) If q is a second degree polynomial on $\mathbb{R}^d$ with symmetric quadratic part, there is a constant C such that for $g(t^*) = q(t^*) + \epsilon(t^*)$,

$$\left| \frac{dS}{dt} - \frac{dq}{dt} \right| < \max_{t^* \in G} \epsilon(t^*) \cdot C(1 + |t|).$$

Property (S2) formalizes a requirement that if $g(t^*)$ can be interpolated approximately by a second degree polynomial $q(t)$ then the smoother produces a function $S(g)$ close to q. A trick that may modify a smoother S' (such as the loess program in the package S-plus described in Venables and Ripley (1995)) to satisfy (S2) is first to choose $1/2(d+1)(d+2)$ distinct points of G and interpolate them by a second degree polynomial q^* with symmetric quadratic part. Then let

$$S''(g) = q^*(r) + S'(g(t^*) - q^*(t^*)),$$

where $q^*(t^*)$ is a mildly abusive abbreviation for the restriction of q^* to G. The modification S'' of S' will satisfy (S2) as long as S' takes small functions on G to small, smooth functions on Θ which one would expect of any reasonable smoother.

The particular function we are interested in finding a smooth approximation to is the rescaled log likelihood

$$\lambda_n(t) = \Lambda_n \left(\bar{\theta}_n^* + \frac{t}{\sqrt{n}}, \bar{\theta}_n^* \right).$$

Evaluating $\lambda_n(t^*)$ on $t^* \in G$ is equivalent to evaluating $\Lambda_n(\theta^*, \bar{\theta}_n^*)$ on $\theta^* \in G_n$. Now set

$$\hat{t} = \operatorname*{argmax}_{t^* \in G} S(\lambda_n(t^*))$$

and define the MSLE by

$$\hat{\theta}_n = \bar{\theta}_n^* + \frac{\hat{t}}{\sqrt{n}}.$$

Lemma 3.1 *Suppose $\{P_{\theta,n}\}$ has LAN at θ_0 with matrix K and centering variables $\{\Delta_n\}$. Let $Q_{\theta,n}$ be the law under $P_{\theta,n}$ of an $\mathbb{R}^d$-valued random variable T_n. If $|T_n - \Delta_n| = o_p(1; \theta_0)$ then $\{Q_{\theta,n}\}$ also has LAN at θ_0 with matrix K.*

Proof. The lemma is a special case of Theorem 4 of Le Cam and Yang (1988). □

This result says heuristically that if $|T_n - \Delta_n|$ becomes small then T_n is an asymptotically sufficient statistic for θ. The proof is based on formalizing this idea of asymptotic sufficiency. A sequence $\{T_n\}$ of statistics is called *distinguished* if there is no asymptotic information loss from observing the statistic T_n rather than the outcome of the experiment $P_{\theta,n}$. A formal definition of distinguished statistics and some powerful properties are given in Le Cam (1986, Section 7.3). Perhaps a surprising feature of Lemma 3.1 is that T_n approximating Δ_n under $P_{\theta_0,n}$ is enough to show that $\frac{dQ_{\theta,n}}{dQ_{\theta_0,n}}$ approximates $\frac{dP_{\theta,n}}{dP_{\theta_0,n}}$ for θ in a neighborhood of θ_0.

Theorem 3.1 *Let $Q_{\theta,n}$ be the law of the MSLE, $\hat{\theta}_n$, under $P_{\theta,n}$. If $\{P_{\theta,n}\}$ has LAN at θ_0 with matrix K, then $\{Q_{\theta,n}\}$ also has LAN at θ_0 with the same matrix K.*

Proof. Writing $T_n = \sqrt{n}K^{1/2}(\hat{\theta}_n - \theta_0)$ and letting Δ_n be a centering variable for $P_{\theta,n}$, in order to apply Lemma 3.1 it is enough to show that $|T_n - \Delta_n| = o_p(1; \theta_0)$.

Let $q_n(t) = t^T K^{1/2}\Delta_n - 1/2 t^T K_t$ be the quadratic identified in the definition of LAN for $\{P_{\theta,n}\}$. As G is finite, from LAN we have

$$\max_{t^\star \in G}(\lambda_n(t^*) - q_n(t^*)) = o_p(1; \theta_0).$$

The maximum of $q_n(t)$ occurs at $\overset{\circ}{t}_n = K^{-1/2}\Delta_n$. Set $S_n(t) = S(\lambda_n(t^*))(t)$, with a maximum at $\hat{\theta}_n$

$$\left|\frac{d}{dt}S_n(t - \overset{\circ}{t}_n)\right| > \left|\frac{d}{dt}q_n(t - \overset{\circ}{t}_n)\right| - \left|\frac{d}{dt}S_n(t - \overset{\circ}{t}_n) - \frac{d}{dt}q_n(t - \overset{\circ}{t}_n)\right|.$$

The first term on the right grows linearly in t, and the second term grows as $(1 + |t|)o_p(1; \theta_0)$ by (S2). Thus for $\epsilon > 0$,

$$\lim_{n \to \infty} P_{\theta_0, n} \left(\inf_{|t - \overset{\circ}{t}_n| > \epsilon} \left| \frac{d}{dt} S_n(t - \overset{\circ}{t}_n) \right| > 0 \right) = 1.$$

From (S1), any maximum of $S_n(t)$ must have $\frac{d}{dt} S_n(t) = 0$, giving $|\hat{t}_n - \overset{\circ}{t}_n| = o_p(1; \theta_0)$ and hence $|T_n - \Delta_n| = o_p(1; \theta_0)$. $\qquad \square$

An alternative approach to demonstrating asymptotic properties of an MSLE, $\hat{\theta}_n$, is to notice that it is asymptotically close to Le Cam's quadratic approximation estimator, which we denote as $\tilde{\theta}_n$. The quadratic approximation estimator is a particular example of an MSLE, so the argument used to prove Theorem 3.1 shows that $|\hat{\theta}_n - \tilde{\theta}_n| = o_p(1/\sqrt{n}; \theta_0)$. The estimator $\hat{\theta}_n$ thus inherits from $\tilde{\theta}_n$ the properties demonstrated in Le Cam and Yang (1990) of being consistent and asymptotically normal with large n asymptotic variance $(nK)^{-1}$. One can either use the theoretical value of K, if it is known, or find an estimate $\hat{K}_n$. Such an estimate is found and used in the extension of the quadratic approximation estimator to the case of random matrices K_n described in Le Cam and Yang (1990, Chapter 5). Under the additional assumptions (S3) and (S4) below on the smoother, S, the second derivative matrix at $\hat{\theta}_n$ can be used to estimate K.

(S3) $S(t) = S(g)(t)$ is twice continuously differentiable.

(S4) If q is a linear-quadratic function on $\mathbb{R}^d$, with symmetric quadratic part, there is a constant C such that for $g(t^*) = q(t^*) + \epsilon(t^*)$ and $\| \cdot \|$ the matrix supremum norm

$$\left\| \frac{d^2 s}{dt^2} - \frac{d^2 q}{dt^2} \right\| \leq \max_G \epsilon(t^*) C.$$

Taking $q(t)$ as the linear-quadratic function appearing in the LAN condition, as in the proof of Theorem 3.1, (S4) allows $\frac{d^2 s}{dt^2}$ evaluated at any point and therefore at $\hat{\theta}_n$ to be used as an estimator $\hat{K}_n$ with $|K - \hat{K}_n| = o_p(1; \theta_0)$, using the supremum norm for matrices given by $|A| = \sup_{i,j} |A_{ij}|$. The condition (S4) looks like a derivative of

(S2). When (S2) and (S3) hold, it is therefore not much more to ask that (S4) hold also.

A generalization of LAN, replacing K by random matrices K_n and requiring that Δ_n conditional on K_n is asymptotically normal, is called local asymptotic mixed normality (LAMN). Many results for LAN can be extended to LAMN, which occurs naturally in the study of certain non-stationary processes such as branching processes (Jeganathan, 1995). Under an additional contiguity requirement the quadratic approximation estimator keeps many of its asymptotic properties under LAMN (Le Cam and Yang, 1990, Chapter 5), and one might expect the MSLE to do likewise.

Another possible extension would be to relax the requirement for LAN that Θ is an open subset of $\mathbb{R}^d$. Allowing for θ_0 to take values on the boundary of Θ is useful for hypothesis testing but rather less so for the parameter estimation problem developed here. Many of the complexities arising from an inability to approach θ_0 from all directions are dealt with in Le Cam (1986). The definition of LAN extends readily to more general spaces than $\mathbb{R}^d$ equipped with an inner product and a quadratic form K. The quadratic approximation estimator however makes use of the ability to span Θ using a finite subset. LAN has been used effectively for non-parametric problems with Θ an infinite dimensional space (Bickel et al., 1993), though our focus on parametric problems solved by methods similar to (P1)–(P5) does not lead us in that direction.

Last in this list of issues, for which the reader is directed elsewhere for details, is the use of Theorem 3.1 to make asymptotic optimality claims concerning $\hat{\theta}_n$. The matrix K is thought of as the information about θ in the experiment, and coincides with the Fisher information under regularity conditions (Le Cam, 1986, Chapter 17.3). One might say that $\hat{\theta}_n$ contains asymptotically all the information about θ in $P_{\theta,n}$. To formalize this we state a result analogous to an asymptotic form of the Cramér–Rao lower bound, called Hájek's convolution theorem and given in this form in Le Cam and Yang (1990, Section 5.6).

Theorem 3.2 *Suppose $\{P_{\theta,n}\}$ is LAN at θ. For the experiments $\mathcal{E}_{t,n} = \{P_{\theta+t/\sqrt{n},n}, t \in \mathbb{R}^d\}$ let T_n be an estimate of At for a given non-random matrix A. Assume that $T_n - At$ tends in distribution under $P_{\theta+t/\sqrt{n},n}$ to some random vector H whose distri-*

bution does not depend on t. Then H shares its distribution with $AK^{-1/2}Z + U$ where Z is $N(0,1)$ independent of a random vector U.

The property that $U = 0$, enjoyed by $\tilde{\theta}_n$ and therefore also $\hat{\theta}_n$, might reasonably be called asymptotic efficiency, a term usually reserved for asymptotically achieving the Cramér–Rao bound. Clearly Theorem 3.2 implies a variance bound which coincides with the Cramér–Rao bound when MLAN holds. The usual statement of the Cramér–Rao bond requires an estimator to be unbiased, though there are extensions for biased estimators (Casella and Berger, 1990, Section 7.3). The condition for the convolution theorem analogous to asymptotic unbiasedness is that the estimator T_n is assumed to have an asymptotic shift invariance property, as the distribution of H does not depend on t.

3.2   Checking LAN for State Space Models

A useful way to check that LAN holds is through a condition, introduced in Le Cam (1986), called differentiability in quadratic mean (DQM). After discussing DQM a version of DQM for conditional experiments will be introduced and used to show that LAN holds for the models in Section 2.1. DQM is a property of a single experiment $\{P_\theta, \theta \in \Theta\}$ and in this setting $P_{\theta,n}$ is the product experiment of n independent versions of P_θ. Here, as before, θ takes values in an open subset Θ of $\mathbb{R}^d$. We are now making an independence requirement, that $P_{\theta,n}$ be a product measure, which was not assumed in Section 3.1.

Definition. A family of probability measures $\{P_\theta, \theta \in \Theta\}$ is said to be differentiable in quadratic mean (DQM) at θ_0 with derivative V if $\mathbb{E}[|V|^2] < \infty$ and

(i)
$$\lim_{\theta \to \theta_0} \frac{1}{|\theta - \theta_0|^2} \mathbb{E}\left[\sqrt{\frac{dP_\theta}{dP_{\theta_0}}} - 1 - (\theta - \theta_0)^T V\right]^2 = 0$$

(ii) For $\beta(\theta, \theta_0)$ the mass of the P_{θ_0} singular part of P_θ,

$$\lim_{\theta \to \theta_0} \frac{1}{|\theta - \theta_0|^2} \beta(\theta, \theta_0) = 0.$$

DQM (ii) is a contiguity condition, a form of asymptotic absolute continuity of measures, which is required to make the following result hold. Theorem 3.3 shows that DQM implies LAN, and gives a partial converse result.

Theorem 3.3 (Le Cam, 1986) *If $\{P_\theta\}$ is DQM at θ_0 with derivative V then the product experiment $\{P_{\theta,n}\}$ has LAN at θ_0 with $K = \mathbb{E}[VV^T]$. A converse holds, that if $\{P_{\theta,n}\}$ has LAN at θ_0 and Δ_n can be written as*

$$\Delta_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k$$

for $\{X_k\}$ i.i.d. copies of a random vector X, then $\{P_\theta\}$ is DQM at θ_0.

In multivariate situations, checking that DQM holds can be awkward (Le Cam and Yang, 1990, Section 6.3). In this section, we introduce a conditional form of DQM and show how it forms a convenient tool for checking DQM in multivariate situations with a natural conditional structure. Let $\{P_{X,Y}(\theta)\}$ be a family of joint distributions for random variables X and Y. Assuming regular conditional probabilities exist (Durrett, 1991, p. 198) they can be used to define a conditional experiment $\{P_{Y|X}(\theta)\}$ where Y takes on its distribution conditional on the outcome of the random variable X.

Definition. $\{P_{Y|X}(\theta)\}$ is conditionally DQM given X at θ_0 with derivative W if $\mathbb{E}[W^2] < \infty$ and

(i) $$\lim_{\theta \to \theta_0} \frac{1}{|\theta - \theta_0|^2} \mathbb{E}\left[\sqrt{\frac{dP_{Y|X}(\theta)}{dP_{Y|X}(\theta_0)}} - 1 - (\theta - \theta_0)^T W \right]^2 = 0$$

(ii) For $\gamma(\theta, \theta_0, X)$ the mass of the $P_{Y|X}(\theta_0)$ singular part of $P_{Y|X}(\theta)$,

$$\lim_{\theta \to \theta_0} \frac{1}{|\theta - \theta_0|^2} \mathbb{E}[\gamma(\theta, \theta_0, X)] = 0.$$

The following proposition shows how DQM follows in a bivariate situation where DQM can be shown for one variable and conditionally for the other. A proof is left to Appendix A.

Proposition 3.1 *If $\{P_X(\theta)\}$ is DQM at θ_0 with derivative V and $\{P_{Y|X}(\theta)\}$ is conditionally DQM at θ_0 with derivative W then $\{P_{X,Y}(\theta)\}$ is DQM at θ_0 with derivative $V + W$.*

When P_θ is the law of a time series $\{Y_j, 1 \le j \le N\}$, as for the models proposed in Section 2.1 for cell tracking data, Proposition 3.1 allows us to check the DQM condition by showing that DQM holds for Y_1 and $Y_j \mid Y_1, \ldots, Y_{j-1}$ for $2 \le j \le N$.

A property of DQM that will become useful is the preservation of DQM under information loss. Proposition 3.2 is a statement to this effect taken from Le Cam and Yang (1988, Section 7).

Proposition 3.2 (Le Cam and Yang, 1988) *Let $\{P_\theta, \theta \in \Theta\}$ be a family of measures on a sigma algebra $\mathcal{A}$ and let Q_θ be the restriction of P_θ to a sigma algebra $\mathcal{B} \subset \mathcal{A}$. If $\{P_\theta\}$ satisfies DQM at θ_0 then so does $\{Q_\theta\}$.*

A final general result concerning DQM gives sufficient conditions for DQM in terms of the Fisher information, when the required derivatives exist.

Proposition 3.3 *Suppose that for $\theta \in \Theta \subset \mathbb{R}^d$, P_θ has a density $p(x, \theta)$ with respect to a dominating measure ν. Suppose further that*

(i) *$p(x, \theta)$ is continuously differentiable in θ for ν almost all x with derivative $\mathbf{p}_\theta(\theta)$.*

(ii) *$\mathbf{s}(\theta) = \frac{\mathbf{p}_\theta(\theta)}{p(\theta)} 1_{\{p(\theta) > 0\}}$ satisfies $|\mathbf{s}(\theta)| \in L^2(P_\theta)$.*

(iii) *The Fisher information defined by*

$$I(\theta) = \int \mathbf{s}(\theta)\mathbf{s}^T(\theta) dP_\theta$$

is a continuous function of θ.

Then $\{P_\theta\}$ satisfies DQM with derivative $\mathbf{s}(\theta)$.

Proof. This result is Proposition 2.1.1 of Bickel et al. (1993) combined with their statement on page 15 on the equivalence between DQM and their definition of a regular parametric model.

Application of Proposition 3.2 to a state space model setting shows that often one need check DQM only for the state process to guarantee DQM for the observation process. This requires that the observation process be a non-random function of the state process and that the function should not depend on the parameter θ. An example occurs when v_t is velocity, $x_t = x_0 + \int_0^t v_t dt$ is position for $t \geq 0$ and (v_t, x_t) forms a Markov process. A discrete time state space model is defined for $n = 1, \ldots, N$ by

$$
\begin{aligned}
X_n &= (v_n, x_n, \epsilon_n) \\
Y_n &= x_n + \epsilon_n
\end{aligned}
$$
(L1)

for independent random vectors $\{\epsilon_n\}$ which might represent measurement error.

When (v_t, x_t) is a continuous process defined as the solution to an infinitesimal equation such as those in models (M1)–(M4), one is led to look for sufficient conditions for (v_1, x_1) to satisfy DQM conditional on (v_0, x_0). We demonstrate in Proposition 3.4 a set of conditions (D1)–(D5) below. These conditions are designed to include the Ornstein–Uhlenbeck process (model (M1)) when the diffusion coefficients are smoothly parameterized by $\theta \in \Theta$, and other similar though not necessarily linear diffusions. The requirement made that the log transition densities have quadratic tails is very restrictive though it does cover many cases of practical interest. The method of proof for Proposition 3.4 could be applied to other situations, but we were unable to construct a simple and more general statement. Thus conditions (D1)–(D5) are best thought of as one example for which the method of proof in Proposition 3.4 can be used.

There is a reasonably large literature on statistical inference for discretely observed diffusion processes, though we are not aware of previous results similar to Proposition 3.4 below. Dacunha–Castelle and Florens–Zmirou (1986) give asymptotic results when the discretization intervals tend to zero. Bibby and Sorensen (1995) use martingale estimating functions and obtain asymptotic results with the discretization interval fixed. Jensen and Pedersen (2000) look at diffusions that can be written as transformations of a linear diffusion, which when observed discretely become the well studied AR process.

We consider a family of diffusion processes in $\mathbb{R}^d$ parameterized by $\theta \in \Theta$, an open subset of $\mathbb{R}^k$, satisfying the infinitesimal equations

$$dX_t = \mu(X_t, \theta)dt + \gamma(X_t, \theta)dW_t.$$

Here W_t is a standard Brownian section in $\mathbb{R}^d$, $\gamma = \gamma(X_t, \theta)$ is a $d \times d$ matrix, and we set $\Gamma(X_t, \theta) = \gamma\gamma^T$. Let $P_{X_1|X_0}(\theta)$ be the conditional law of X_1 given X_0, and let $p(\varphi, x, t)$ be the density of X_t at x when X_0 is distributed with a density φ. In this notation the dependence of p on θ is suppressed, x takes values in $\mathbb{R}^d$, t lies in the interval $[0, 1]$ and φ is a fixed probability density.

(D1) μ and Γ have uniformly continuous partial derivatives $\mu_\theta, \mu_x, \mu_{\theta x}, \Gamma_{\theta\theta}, \Gamma_{xx}, \Gamma_{\theta xx}$ with respect to x and θ. Subscripts are used here to denote vectors and higher dimensional arrays of partial derivatives, so $\mu_x = \left(\frac{\partial\mu}{\partial x_1}, \dots, \frac{\partial\mu}{\partial x_d} \right)^T$.

(D2) $p(\varphi, x, t)$ has continuous first and second partial derivatives p_x and p_{xx}, and there exist symmetric positive definite matrices $A = A(\theta)$, $B = B(\theta)$ and positive scalars $a = a(\theta)$, $b = b(\theta)$ such that for $0 \le t \le 1$ and $x \in \mathbb{R}^d$

$$ae^{-x^T Ax} \;<\; p(\varphi, x, t) < be^{-x^T Bx}$$

$$be^{-x^T Ax} \;>\; \left| \sum_i \frac{\partial}{\partial x_i}(\mu_\theta p) - \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j}(\Gamma_\theta p) \right|$$

(D3) For any function η in $L^1(\mathbb{R}^d)$ with $|\eta(x)| \le be^{-x^T Bx}$, there is a scalar $c = c(\theta)$ such that for $0 \le t \le 1$

$$|p(\eta, x, t)| \le ce^{-x^T Bx}$$

where $p(\eta, x, t)$ is defined for η in $L^1(\mathbb{R}^d)$ by linear extension from η in the set of probability densities $\{\eta : \eta(x) \ge 0, \ \int \eta dx = 1\}$.

(D4) $a(\theta), b(\theta), c(\theta), A(\theta), B(\theta)$ are continuous functions of θ.

(D5) $2B(\theta) - A(\theta)$ is positive definite for all $\theta \in \Theta$.

Example. Let X_t be an Ornstein–Uhlenbeck process in $\mathbb{R}^1$ with infinitesimal equation

$$dX_t = \theta_1 X_t dt + (\theta_2)^{1/2} dW_t$$

and $X_0 \sim N[\mu_0, \sigma_0^2]$. Let $A = (\inf_{0 \le t \le 1} \text{var}(X_t))^{-1}$, $B^* = \left(\sup_{0 \le t \le 1} \text{var}(X_t)\right)^{-1}$ and $\Theta = \{(\theta_1, \theta_2) : 2B^* - A > 0\}$. (D1) clearly holds, and since the marginal distributions are all Normal one can readily see that (D2)–(D5) hold for $\theta \in \Theta$ with A as above and $B = B^* - \epsilon$ for any $\epsilon > 0$.

Proposition 3.4 *Suppose* $P_{X_1|X_0}(\theta)$ *and the density* φ *of* X_0 *satisfy conditions* (D1)–(D5) *at* θ_0. *Then* $P_{X_1|X_0}$ *satisfies* DQM *at* θ_0.

Proof. The density $p(\varphi, x, t)$ of X_t satisfies the forward equation for $t > 0$ (Karlin and Taylor, 1981)

$$\frac{\partial p}{\partial t} - Lp = 0 \tag{3.1}$$

with $L = \sum_i \frac{\partial}{\partial x_i} \mu - \frac{1}{2} \sum_{ij} \frac{\partial}{\partial x_i \partial x_j} \Gamma$ and initial condition $p(\varphi, x, 0) = \varphi(x)$. Following Stroock (personal communication) we note that, if it exists, the partial derivative with respect to θ, denoted with a subscript, satisfies an equation derived from (3.1)

$$\frac{\partial p_\theta}{\partial t} - Lp_\theta = L_\theta p \tag{3.2}$$

with $L_\theta = \sum_i \frac{\partial}{\partial x_i} \mu_\theta - \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \Gamma_\theta$ and initial condition $p_\theta(\varphi, x, 0) = 0$. The forward equation fell out of fashion for probabilistic applications because of difficulties with existence and particularly uniqueness of solutions. In this setting, (D1)–(D2) are strong enough to apply general results for parabolic equations, such as Lieberman (1996, Theorem 5.15), to give the existence of a unique solution to (3.1) and (3.2). Denoting the right-hand side of (3.2) as $f(x, t)$ the solution to (3.2) can be written in terms of the solution to (3.1) using the linearity of L,

$$p_\theta(\varphi, x, t) = \int_0^t p(f(\cdot, s), x, t - s) ds.$$

Condition (D2) gives $f(x, t) < be^{-x^T Bx}$ and then (D3) gives

$$|p_\theta(\varphi, x, t)| < ce^{-x^T Bx}. \tag{3.3}$$

We now proceed to verify the three requirements needed to apply Proposition 3.3, writing $p(x, \theta)$ for $p(\varphi, x, 1)$.

(i) The continuity of $p_\theta(x, \theta)$ as a function of θ can be seen to hold under (D1) by considering a diffusion (X_t, θ) on $\mathbb{R}^d \times \Theta$. The θ component is a trivial diffusion—it remains constant—but one can still write down analogous equations to (3.1) and (3.2). Uniform continuity of $L_\theta p$ as a function of θ and then of p_θ as a function of θ then comes from applying the Hölder continuity bounds of Lieberman (1996, Theorem 5.15) to the initial conditions of (3.1) and the right-hand side of (3.2) respectively. This trick was suggested by Stoock (personal communication).

(ii) The inequality (3.3) together with conditions (D2) and (D5) show that the score function $s(\theta) = p_\theta(x, \theta)/p(x, \theta)$ is in $L^2(p(x, \theta))$, as

$$|s(\theta)| < (c^2/a) \exp\{-x^T(2B - A)x\}.$$

(iii) The Fisher information is

$$I(\theta) = \int \frac{(p_\theta(x, \theta))^2}{p(x, \theta)} dx.$$

From part (i), the integrand converges pointwise in θ for almost all x. Using (D4), the bound from part (ii) provides a uniform bound in a neighborhood of θ (recall that Θ is open) and so by dominated convergence the integral $I(\theta)$ is continuous in θ. $\square$

3.3   Concluding Remarks

This chapter has discussed some implications for statistical practice of asymptotic theory built around the property of local asymptotic normality (LAN). This property was used to justify a parameter estimation procedure involving maximization of a smooth approximation to the likelihood function (MSLE), and also to give error estimates for this procedure. The asymptotic results place only weak restrictions on the particular smoothing method used. In practice one has to choose a particular

method, and often a bandwidth parameter to go along with it. A number of methods, including simulation and bootstrap techniques, are available to address these issues (Silverman, 1985).

Section 3.2 developed some results to show how the state space models proposed in Section 2.4 fit into the framework of Section 3.1. State space models have been found useful in many applications, some of which were discussed in Section 1.4. In addition, mixture models can be written in state space form (Example 1, this chapter), as can many methods for nonparametric regression (Wood and Kohn, 1998). The method of MSLE is not however limited to state space models, and might assist in the analysis of many parametric models where the likelihood function is either not smooth or can only be estimated using Monte Carlo simulation. Another situation where these conditions arise is spatial statistics, which involves modeling using random fields (Cressie, 1993).

Estimators not based on the likelihood function are sometimes prefered in practice when evaluating and maximizing the likelihood function is found to be difficult. Examples include the use of moment estimators for cell motion parameters (Dickinson and Tranquillo, 1993b; DiMilla *et al.*, 1992), the use of various forms of pseudo-likelihood for random field models (Besag, 1974), and the use of estimating equations for regression analysis of longitudinal data (Liang and Zeger, 1986). Likelihood based methods, such as MLE and MSLE, have a major advantage that they can be shown to be asymptotically efficient in many situations. It is at least reassuring to be able to compute an estimator that can be shown to be efficient, even if it does not become the method of choice (for example, if the extra computational labor is found not to yield much improvement in the estimator). The MSLE introduced in this chapter, by being both more generally applicable and possibly more readily computable than the MLE, may facilitate likelihood based analysis.

Chapter 4

A Statistical Analysis of Data Arising from Staining Fixed Cells

Fixing and staining cells or tissues is a standard procedure which involves adding a preservative chemical followed by an observable marker that can be used to determine where a particular molecule concentrates in a cell or collection of cells (Lodish et al., 1995, Chapter 5). In the study of cell motion such experiments can complement time lapse video microscopy of living cells (Fang et al., 1999). This chapter correspondingly complements our Chapters 2 and 3 by considering some statistical issues arising from the analysis of digitized images of fixed, stained cells.

In the experiment motivating this chapter and reported in Fang et al. (1999), scientists were concerned whether a protein stained for using a monoclonal antibody is implicated in the mechanism that a cell uses to direct its motion toward a stimulus. Human skin cells were placed on a microscope slide and an electric field applied, as described in Section 1.2. After a certain time interval the cells were fixed with paraformaldehyde, stained and captured as a digital image. An example of a typical observation is shown in Figure 4.1. The stain was expected to attach preferentially to the cell membrane, and indeed a boundary region of high stain intensity was seen. A traditional way to statistically compare the images from different treatment groups is for an observer to estimate some quantity relevant to the hypothesis in question. In our example a cell might be given a score of "1" if the boundary staining
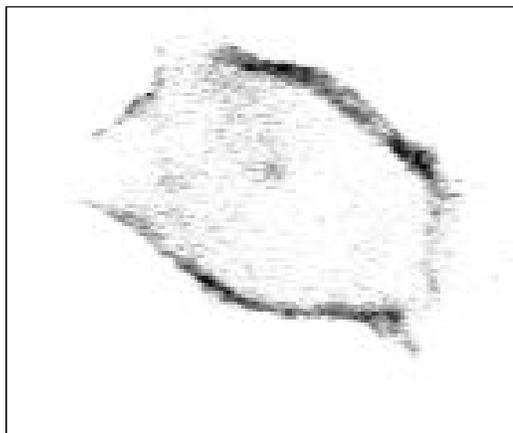
Figure 4.1: A cell from the treatment group. The electric field is oriented with the cathode at the top of the page.

appears concentrated toward the cathode, and "0" if it appears concentrated toward the anode. This technique, termed manual scoring, can be time consuming, simplistic and difficult to make objective. Unless care is taken to ensure the observer is unaware of the treatment group, a procedure termed blinding, unintended observer bias can have important consequences (Freedman et al., 1998). A goal of this chapter is to present a more formal and impartial analysis which produces more informative summaries of experimental results and efficient tests of relevant hypotheses.

Section 4.1 introduces an algorithm that has been successfully implemented in a computer program to estimate the boundary of a cell and the intensity of the stain along this boundary. Section 4.2 proposes some simple statistics to quantify the location of this stain in order to test hypotheses such as whether the stain is located preferentially at the end of the cell facing the cathode. In Section 4.3 we develop some models for the staining process which enable us to discuss optimality properties of tests and estimators. Section 4.4 demonstrates these methods on a data set collected for Fang et al. (1999).

4.1   Computation of the boundary and its staining

This section introduces an algorithm whose input is an image of a stained cell and whose output is a one-dimensional set B called the boundary of the cell and a real-valued function $I(b)$, defined for $b \in B$, called the boundary intensity of the stain. This algorithm can be viewed as a preprocessing step which derives from the image a quantity $I(b)$ of direct scientific interest. The algorithm quantifies the concept of boundary stain intensity and is successful in as much as the output generally has the qualitative features that a cell biologist seeks. The algorithm is not trying to estimate a preexisting quantity such as a parameter in some suitable model for the cell staining process, so the output can be taken to define the boundary intensity of the stain.

The main tool used in the algorithm is mathematical morphology, which is introduced for image processing in Serra (1982). Mathematical morphology is a group of techniques based on the theory of random sets (Matheron,

1975; Stoyan et al., 1995). Notation and further details are introduced later, in a formal description of the algorithm. First we give a short and informal description, illustrated with the example in Figure 4.2.

Boundary Stain Intensity Algorithm – Short Description

1. Threshold the original image to give a purely black or white image whose black area is approximately the shape of the cell. See Figure 4.2(i).

2. Clean up the thresholded image using the mathematical morphology closing operation, to be described below. See Figure 4.2(ii).

3. Fill in the center of the cell and remove small outlying black areas unattached to the cell. See Figure 4.2(iii).

4. Find the grey-scale level measuring the intensity of the stain in a boundary region around the edge of the cell. See Figure 4.2(iv).
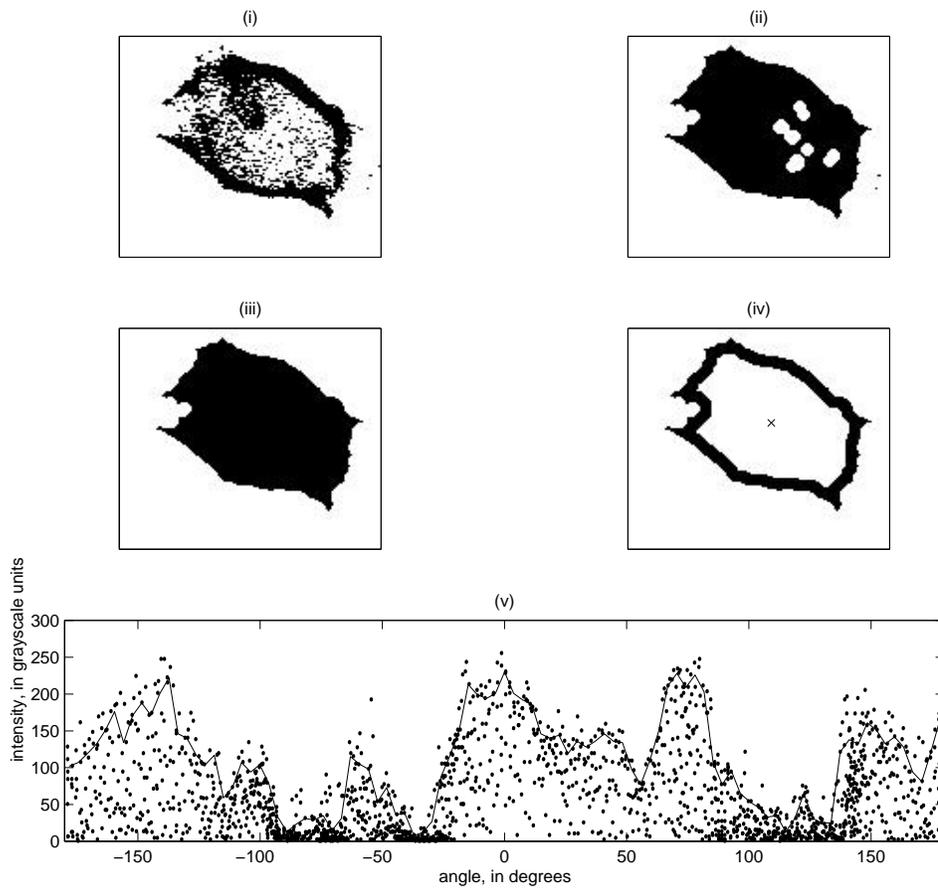
Figure 4.2: (i) Threshold the original image to get a crude black and white image whose black area is approximately the shape of the cell. (ii) Use the closing operation from mathematical morphology to clean up the thresholded image. (iii) Fill in holes in the cell and remove small outlying black areas unattached to the cell. (iv) Find a boundary several pixels wide around the edge of the cell. (v) The gray scale values of the boundary pixels are smoothed using a local quantile smoother to give boundary intensity as a function of angle around the center of the cell.

5. Use these intensities to determine a measure of boundary stain intensity as a function of angle around the center of the cell. See Figure 4.2(v).

Formal Description

The grey-scale image of a cell, such as that in Figure 4.1, is a matrix of pixel values $M = \{M(i,j)\}$, $i = 1, \ldots, N_x$, $j = 1, \ldots, N_y$. The gray-scale level, $M(i,j)$, takes values in $\{0, 1, \ldots, 255\}$, and we suppose that darker, stained areas correspond to higher pixel values.

1. For some predetermined threshold t, and 1_X the indicator function that takes value 1 if X is true and 0 otherwise, set M_1 to be the matrix given by

$$M_1(i,j) = 1_{M(i,j) \geq t}.$$

2. To define the required morphological operation we begin by setting N to be a subset of $\mathbb{Z}^2$, called a neighborhood. In this work we take N to be

$$N = \{(0,0), (0,-1), (0,1), (-1,0), (1,0)\},$$

called the 4-neighborhood of zero. The Minkowski sum of N with a set $S \subset \mathbb{Z}^2$ is

$$S \oplus N = \{x + y \mid x \in S, \; y \in N\}.$$

The Minkowski subtraction of N from S is

$$S \ominus N = (S^c \oplus N)^c$$

where A^c is the complement of A. The operation of *dilation* by N is given by $d_N(S) = S \oplus N$ and *erosion* by $e_N(S) = S \ominus N$. These are combined to produce the two useful operations of *opening* and *closing* given respectively by

$$
\begin{aligned}
o_N(S) &= d_N(e_N(S)), \\
c_N(S) &= e_N(d_N(S)).
\end{aligned}
$$

The closing operation on S can be thought of as filling in holes in S, which suits our purpose for the algorithm. In fact we use a neighborhood N^p defined for a predetermined parameter p as

$$N^p = \overbrace{N \oplus \cdots \oplus N}^{p \text{ times}}.$$

The interpretation of p is that it gives the length scale, in pixel units, on which the boundary can be resolved from the image. We can now write the second step of the algorithm as

$$M_2 = c_{N^p}(M_1).$$

No confusion should arise from the notational convention used here in which a binary matrix is identified with the set of all pairs of indices labeling positions taking value 1 in this matrix.

3. Let S_1 be the largest connected component of M_2 taking the value 1 and S_0 the largest connected component of $(M_2 \cap S_1)$ taking value 0. Computer packages for image analysis, such as the MATLAB Image Analysis Toolkit, provide functions for calculating these connected components. The *silhouette* of the cell is defined by

$$M_3 = S_1 \cup (M_2^c \cap S_2^c).$$

This operation takes the largest connected component of M_2 and fills in, i.e., gives value 1, to any background 0's surrounded by it. The resulting algorithm becomes robust in that it ignores outlying areas of stain, such as adhesion regions left behind during cell motion, and fills in large interior regions that picked up negligible stain. The silhouette M_3 estimates the extent of the cell membrane, or in other words the shape of the cell projected onto a plane. Implicit assumptions for this method to be appropriate are that the silhouette should be connected, and that the background region surrounding the cell in the image is larger than the unstained interior region of the cell.

4. The boundary B is the outline of the silhouette, which can be written formally as $B \subset \mathbb{R}^2$ with $B = \{\mathbf{x} = (x, y) : d(\mathbf{x}, M_3) = 1/2, \ d(\mathbf{x}, M_3^c) = 1/2\}$ where

$d(\mathbf{x}, S) = \inf_{(x',y')\in S} \max\{|x - x'|, |y - y'|\}$. For given w, the boundary region of the cell is calculated by

$$M_4 = M_3 \cap (e_{N^w}(M_3))^c.$$

The parameter w gives the width of the boundary region, which is chosen so that the ring of stain often observed inside the boundary of the cell is included in M_4.

5. We obtain from M_4 a function $I(\phi)$ representing the stain intensity on the boundary at angle ϕ around a point of origin.

In the computations presented, this point of origin is taken to be the "center of mass", $\mathbf{c} = (c_x, c_y)$, of the silhouette M_3, and is given by

$$c_x = \sum_{(i,j)\in M_3} i \bigg/ \left(\sum_{(i,j)\in M_3} 1\right)$$

$$c_y = \sum_{(i,j)\in M_3} j \bigg/ \left(\sum_{(i,j)\in M_3} 1\right).$$

In practice a more robust method for calculating a center, such as the median, could be used. Evaluating the stain intensity as a function of angle around $\mathbf{c}$ is convenient for making comparisons between different cells, and furthermore the angle from the cathode is a quantity with scientific importance for experiments of the type under consideration. These considerations led to the choice of a radial parameterization of the boundary stain intensity, even though it does not give a true parameterization of the boundary set, B, when a ray starting from $\mathbf{c}$ intersects B at more than one point.

Suppose that the points in M_4 are listed as $\{(i_n, j_n), \ 1 \le n \le |M_4|\}$, let the stain intensity of the n^{th}p point be denoted by $V_n = M(i_n, j_n)$, and let Φ_n be the angular coordinate of (i_n, j_n) from $\mathbf{c}$. Determining the stain intensity, $I(\phi)$, can be viewed as smoothing $V = \{V_n\}$ on $\Phi = \{\Phi_n\}$. Figure 4.2(v) shows the points in our example. There are many low values, corresponding to positions in M_4 with little stain. Some but not all of the ring of stain may have fallen in the region M_4. A robust method

with a high breakdown point, such as the moving median (Tukey, 1977) is called for. Here a moving quantile smoother turned out to be effective with

$$I(\phi) = q^{\text{th}} \text{ quantile of } \{V_n : |\phi - \Phi_n| \leq \psi\}$$

for some fixed parameters q and ψ. One could use a weight function such as the bisquare (Tukey, 1977) to form a locally weighted quantile smoother, paying more attention in the calculation of $I(\phi)$ to the points (V_n, Φ_n) with Φ_n close to ϕ.

Recall that the algorithm presented requires several parameters to be specified, namely t, p, w, q and φ. At the moment these parameters are chosen by the scientist, since the algorithm is ad hoc and justified by producing a scientifically useful quantity as an output. The subjectivity introduced by the choice of parameters seems minimal compared to that for manual scoring methods, particularly when one finds (as in Fang et al., 1999) that the same values of these parameters are appropriate for every image collected under similar experimental conditions.

4.2   Some statistics to measure stain location

Having constructed the quantity $I(\phi)$, measuring the boundary stain intensity as a function of angle around the cell center, we are interested in using it in two distinct ways. The first use is to quantify and test hypotheses about the location of the stain. We demonstrate this by considering the pair of hypotheses of interest in Fang et al. (1999), namely the null hypothesis

H_0: the distribution of the stain intensity is rotationally invariant.

and the alternative hypothesis

H_1: the stain intensity is concentrated in the upper half plane, $-\pi/2 < \phi < \pi/2$.

The second use is to model the distribution of the stain intensity and its relation to other experimental factors in order to aid understanding of the system under investigation. The link between these two is that optimality properties for parameter

estimates and statistical tests come from and depend on a model postulated for the process.

For testing H_0 against H_1 it is convenient to have a single real-valued statistic measuring the asymmetry of the function $I(\phi)$. The particular statistic chosen in Fang et al. (1999), called the *asymmetry index*, is the correlation of $I(\phi)$ with some function $f(\phi)$, defined as

$$A = \frac{\int_{-\pi}^{\pi} I(\phi)f(\phi)d\phi}{\sqrt{\int_{-\pi}^{\pi}(I(\phi) - \bar{I})^2 d\phi \int_{-\pi}^{\pi} f(\phi)^2 d\phi}} \tag{4.1}$$

where $\bar{I} = \frac{1}{2\pi}\int_{-\pi}^{\pi} I(\phi)d\phi$ and f is chosen so that $\bar{f} = \int_{-\pi}^{\pi} f(\phi)d\phi = 0$. The statistical properties of A are simplified if $f(\phi)$ is fixed rather than data determined, and our current choice is $f(\phi) = \cos(\phi)$.

The asymmetry index is invariant to linear transformations of the gray scale $I \mapsto aI + b$. This is an appropriate property as the gray scale is somewhat arbitrary, depending on the amount of stain that entered the cell and the particular lighting conditions. In addition, the asymmetry index has the property that, when H_0 holds, for any given $f(\phi)$,

$$\mathbf{E}[A] = 0.$$

When the cells are assumed to be independent, this allows the use of the familiar t-statistic to test H_0 against H_1.

4.3 A stochastic model

In the previous section no assumptions were made about the nature of the process $I(\phi)$. To develop the discussion further we postulate a model

$$(M1) \qquad\qquad I(\phi) = \mu(\phi) + \eta(\phi),$$

where $\mu(\phi)$ is a non-random function on the circle, $-\pi \leq \phi < \pi$, and $\eta(\phi)$ is a mean zero noise process on the circle with rotationally invariant distribution. One could also suppose that the additive relation in $M1$ holds instead for a transformation of

$I(\phi)$, such as the logarithm. The model is set up so that $I(\phi)$ has a rotationally invariant distribution when $\mu(\phi)$ is constant.

A frequency domain approach to analyzing model $M1$ is developed in Dufour and Roy (1976). We will follow their approach, based on observing $I(\phi)$ at N equally spaced points $\phi_n = \frac{2\pi n}{N}$, $n = 0, 1, \ldots, N - 1$.

The covariance of η is written, using addition modulo 2π on the circle, as

$$R(\psi) = \mathbb{E}[\eta(\phi)\eta(\phi + \psi)].$$

The spectrum $\{S_k\}$ is then defined as the coefficients of the Fourier series for $R(\psi)$

$$R(\psi) = \sum_{k=0}^{\infty} S_k \cos(R\psi).$$

From N equally spaced points we define the covariance $R_n = R\left(\frac{2\pi n}{N}\right)$, which has the corresponding spectrum $\{S_k^N\}$ given by

$$R_n = \sum_{k=0}^{\lfloor N/2 \rfloor} S_k^N \cos(2\pi kn/N),$$

where $\lfloor N/2 \rfloor$ denotes the largest integer less than or equal to $N/2$. S_k^N differs from S_k due to the aliasing phenomenon.

As for the situation on the line (i.e., a time series), $\{S_k^N\}$ can be estimated using the finite Fourier transform

$$d_\eta^N(k) = \sum_{n=0}^{N-1} \eta(\theta_n) e^{-k\theta_n}.$$

The result that makes spectral analysis worthwhile is Theorem 4.1 below in which $N^c(\mu, \sigma^2)$ denotes a complex normal random variable with mean μ and variance σ^2.

Theorem 4.1 (Dufour and Roy, 1976) *Let $\eta(\phi)$ be a mean zero stationary Gaussian process on the circle. Then the random variables $d_\eta^N(k)$, $k = 0, 1, \ldots, \lfloor N/2 \rfloor$ are mutually independent with $d_\eta^N(k) \sim N^c(0, 1/2\ N^2 S_k^N)$ for $0 < k < N/2$ and $d_\eta^N(k) \sim N(0, N^2 S_k^N)$ for $k = 0, N/2$.*

In contrast to the time series case (Brillinger, 1975) the independence of the terms in the Fourier transform is an identity rather than an asymptotic result. Note however that Dufour and Roy make a Gaussian assumption, which can be replaced by a mixing condition for the equivalent asymptotic time series result (Brillinger, 1975, Theorem 4.4.1). The difficulty with removing the Gaussian assumption from Theorem 4.1 is that the natural asymptotic limit occurs from sampling $\eta(\theta)$ increasingly finely on the circle. One could perhaps consider an asymptotic result where $\eta(\phi) = \epsilon(\phi) + \xi^N(\phi)$ with $\epsilon(\phi)$ Gaussian and $\xi^N(\phi)$ mixing on a distance of order $O(1/N)$. Here we do not get sidetracked into such an investigation, which would not substantially assist the exploratory data analysis which we are aiming towards in Section 4.4.

Applying Theorem 4.1 to $(M1)$, assuming $\eta(\theta)$ is a Gaussian process, gives

$$\begin{aligned}
d_I^N(k) &\sim N^c(d_\mu^N(k), 1/2\ N^2 S_k^N) \text{ for } 0 < n < N/2 \\
d_I^N(k) &\sim N(d_\mu^N(k), N^2 S_k^N) \text{ for } n = 0, N/2.
\end{aligned}$$

We can use this result for a pointwise hypothesis test of the null hypothesis that $d_\mu^N(k) = 0$ against the general alternative for a particular value of $k > 0$. For the case $k = 1$ this gives a test somewhat similar to the asymmetry index in Section 4.1. For T i.i.d. observations $I_1, \ldots, I_T$ let $c_i = \langle I_i, \cos \rangle = \sum_{n=0}^{N-1} I_i(\phi_n) \cos(\phi_n)$ and $s_i = \langle I_i, \sin \rangle$. A standard F-test (Venables and Ripley, 1997) can be used to test the null hypothesis that c_i and s_i have mean zero, namely,

$$\frac{(\bar{c}^2 + \bar{s}^2)/2}{\sum_i [(c_i - \bar{c})^2 + (s_i - \bar{s})^2]/(2T - 2)} \sim F_{2, 2T-2} \tag{4.2}$$

where $\bar{c} = \sum_{i=1}^T c_i/T$, $\bar{s} = \sum_{i=1}^T s_i/T$.

The F-test is discussed in Seber (1977) where it is shown to have the theoretical appeal of being a likelihood ratio test and the practical appeal of being robust to the assumption of normality in the particular form it takes in (4.2).

In order to investigate the choice of f in (4.1) we now consider more specific hypotheses than the pair H_0 and H_1 of Section 4.2. When testing two point hypotheses, a likelihood ratio test has the property of being most powerful, according to the Neyman-Pearson lemma (Casella and Berger, 1990). Assuming model (M1) holds,

together with the Gaussian condition of Theorem 4.1, one might write down a null and alternative hypothesis as

H_0: $\mu = 0$.

H_1: $\mu = \tilde{\mu}(\phi)$.

The log-likelihood ratio statistic from observing $\{I(\phi_i), 0 \le i \le N-1\}$ is

$$\Lambda = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} I(\phi_i) Q_{ij}^{-1} \tilde{\mu}(\phi_j),$$

where Q is an $N \times N$ matrix given by $Q_{ij} = R_{|i-j|}$. This can be written (for $S_k^N > 0$) as

$$\Lambda = \langle I, g \rangle \tag{4.3}$$

with

$$g(\phi) = \frac{1}{N} \sum_{k=0}^{N-1} \exp(ik\phi) d_{\tilde{\mu}}^N(k) / S_k^N.$$

We see that, in this particular case, if $\tilde{\mu}(\phi) \propto \cos(\phi)$ then a most powerful test statistic is $\langle I, \cos \rangle$, regardless of the power spectrum of $\eta(\phi)$. In the context of the asymmetry index, A, of (4.1), this suggests that $f(\phi) = \cos(\phi)$ will be appropriate when we expect deviations from the null hypothesis to resemble $\cos(\phi)$. This is found to hold in practice for the data analyzed in Section 4.4. The statistic A differs from (4.3) in that it is scaled by the norm of the intensity function, $\|I\|$. This is motivated by the desire to allow for fair comparison of cells which have varying levels of stain.

An extension of model $M1$ to describe the phenomenon that different cells take up varying quantities of stain is to suppose that $(M1)$ holds for a linear transformation of the grey-scale intensity, leading to a model

$(M2)$ $\qquad\qquad\qquad\qquad I(\phi) = a(\mu(\phi) + \eta(\phi)) + b.$

For $(M2)$ to be well specified we suppose that $\langle \mu, 1 \rangle = \int_0^{2\pi} \mu(\phi) d\phi = 0$ and the spectrum $S = \{S_k\}$ of η satisfies $\|S\|^2 = \sum_{k=0}^{\infty} S_k^2 = 1$. Each observed cell is supposed to have its own value of a and b with some common function $\mu(\phi)$ and an i.i.d.

realization of $\eta(\phi)$. We then test for $c_i/\|I_i - \bar{I}_i\|$ and $s_i/\|I_i - \bar{I}_i\|$ having mean zero, where $\|I_i - \bar{I}_i\|^2 = \langle I_i - \langle I_i, 1 \rangle, \ I_i - \langle I_i, 1 \rangle \rangle$, using the same F statistic given in (4.2). The F statistic only approximately follows the F distribution in this case as $c_i/\|I_i - \bar{I}\|$ and $s_i/\|I_i - \bar{I}_i\|$ are no longer normally distributed.

For the situation (from Fang et al. (1999)) considered in Section 4.2, symmetry of the exponent along the $\theta = 0$ axis should ensure that $s_i/\|I_i - \bar{I}_i\|$ has mean zero. One might then test only for $\mathbb{E}[c_i/\|I_i - \bar{I}_i\|] = 0$. The F statistic for this can be seen to be exactly a t-statistic for the asymmetry index A, with $f(\phi)$ set to $\cos(\phi)$, introduced in Section 4.2.

The tests against a sinusoidal alternative discussed above will be seen in Section 4.4 to be adequate for our purposes here, but we note that smoothing techniques (Hastie and Tibshirani, 1990) can be used to test $\mu(\phi) = $ constant against the alternative that $\mu(\phi)$ is a smooth function of ϕ. A simple application in the context of model $(M1)$ is to apply an arbitrary smoother to $I(\phi) - \bar{I}(\phi)$, resulting in a function $\hat{\mu}(\phi)$ which can be taken as an estimate of $\mu(\phi) - \bar{\mu}(\phi)$.

The quantity $\hat{\mu}(\phi)$ has a norm given by $\|\hat{\mu}\|^2 = \langle \hat{\mu}, \hat{\mu} \rangle$. A two sample t-test may then be used to compare the expected value of $\|\hat{\mu}\|$ for a control group (where $\mu(\phi) = 0$ may be presumed from the symmetry of the experiment) and a treatment group. If the sample mean of the statistic $\|\hat{\mu}\|$ is significantly larger for the treatment group one infers that $\mu(\phi)$ is non-constant for the treatment group. More sophisticated applications of smoothing techniques, taking into account the dependence structure of the random process $\eta(\phi)$, should be more efficient but are beyond the scope of this section.

Tests made against an unspecified smooth alternative are sometimes called "nonparametric". Interestingly, the t-test for $\|\hat{\mu}\|$ proposed above is actually model dependent in an unpleasant way not shared by the F-test for the sinusoidal alternative. The quantity $\mathbb{E}[\|\hat{\mu}\|]$ depends upon the distribution of $\eta(\theta)$, which one supposes is unchanged in the treatment and control groups. If however the treatment group also has $\mu(\phi) = $ constant but the spectrum of $\eta(\phi)$ has more power at low frequencies than the control group this increases $\mathbb{E}[\|\hat{\mu}\|]$. The quantities c_i and s_i introduced for the F-test have their mean unaffected by $\eta(\phi)$ provided only that it have rotationally

invariant distribution.

4.4   Analysis of some experimental data

A data set of 23 cell images is presented in Figure 4.3. It forms a single experimental group from Fang et al. (1999), meaning that each cell received the same experimental treatment. This treatment was exposure for 5 minutes to a DC electric field of 100 mV/mm for isolated cells attached to a microscope slide, after which the cells were fixed and stained for epidermal growth factor receptor (EGFR). We see that the cells stand out fairly clearly from the background, and the eye can detect a ring of stain around the boundary of the cells. Visually we can perhaps make out a trend that there is more stain in general on the top of the images, but it is nothing we would care to swear to. For example, cell (6) has its stain concentrated on the bottom left.

The algorithm described in Section 4.1, where it was demonstrated on a cell that can now be recognized as cell (10) of Figure 4.3, produced for each cell a boundary and boundary stain intensity function that corresponds to the visually observed ring of stain. In Fang et al. (1999, Table 2) it was relevant to check whether there is indeed evidence of asymmetry of the boundary stain. For this purpose the asymmetry index, motivated in Sections 4.2 and 4.3, was calculated for each cell. The average was 0.25 with standard error 0.06, giving the t-test P-value of 0.001. A normal quantile plot (described in Venables and Ripley, 1995) showed no evidence of serious violation of the normality assumptions for this test (Freedman et al., 1998), which strongly suggests that the stain intensity is indeed higher on average toward the cathode in the experimental electric field. The control experiment, with the electric field turned off, gave an average of -0.05 with standard error 0.04, showing no evidence of the effect seen for the treatment experiment, allowing that effect to be attributed to the electric field.

A more complete and informative analysis can be carried out by considering the stain intensity functions $I_i(\phi)$, $1 \leq i \leq n = 23$, rather than just calculating the single summary statistic of the asymmetry index. In the context of the model $M2$, with
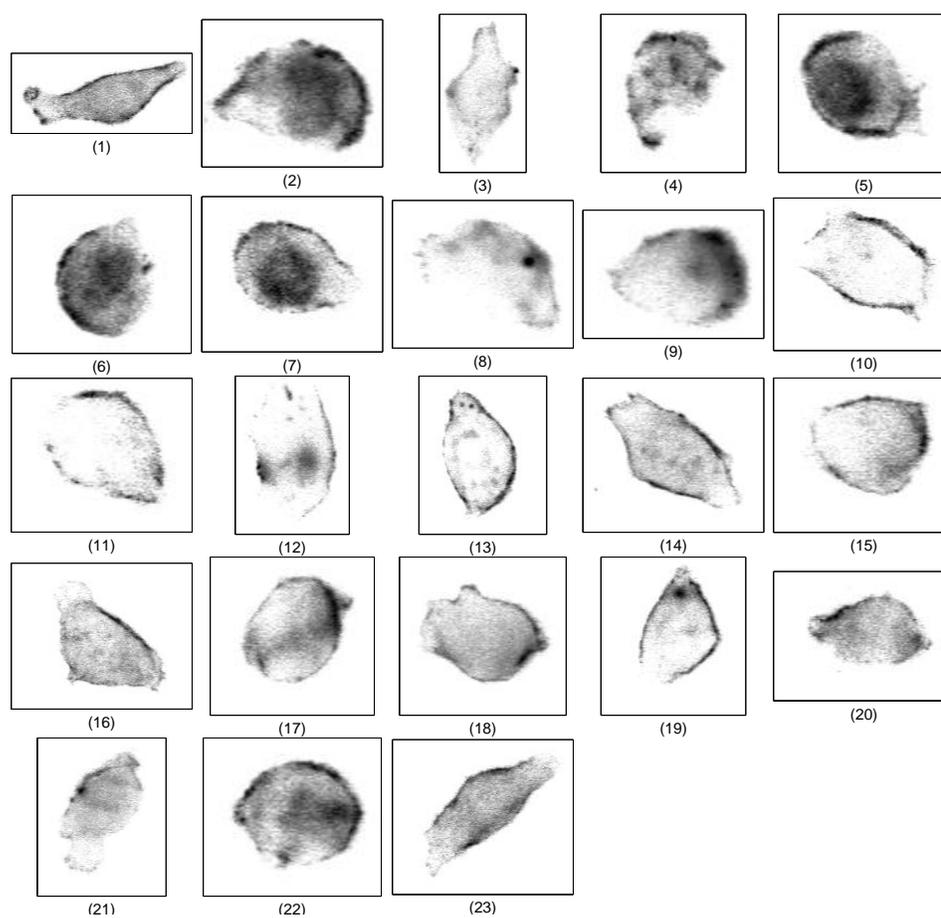
Figure 4.3: A treatment group of 23 cells, exposed to an electric field oriented with the cathode at the top of the page.

unknown scale and shift nuisance parameters, a and b, one can construct standardized stain functions $\tilde{I}_i(\phi) = (I_i(\phi) - \bar{I}_i)/\sqrt{\langle I_i - \bar{I}_i, I_i - \bar{I}_i \rangle}$. Estimates of the mean, covariance and spectrum of the standardized stain functions may be taken as estimates of $\mu(\phi)$ and the covariance and spectrum of $\eta(\phi)$. Estimates are given for $\phi = \phi_0, \phi_1, \ldots, \phi_{N-1}$ by averaging across experiments. In particular,

$$
\begin{aligned}
\hat{\mu}(\phi) &= \frac{1}{n} \sum_{i=1}^{n} \tilde{I}_i(\phi) \\
\hat{\eta}_i(\phi) &= \tilde{I}_i(\phi) - \hat{\mu}(\phi) \\
\hat{R}(\phi) &= \frac{1}{nN} \sum_{i=1}^{n} \sum_{j=1}^{N} \hat{\eta}_i(\phi) \hat{\eta}_i(\phi + \phi_j) \\
\hat{S}(k) &= \frac{\alpha_k}{N} \sum_{m=0}^{\lfloor N/2 \rfloor} \alpha_m \hat{R}(\phi_m) \cos(2\pi mk/N)
\end{aligned}
$$

where

$$
\alpha_k = \begin{cases} 1 & k = 0, N/2 \\ 2 & 0 < k < N/2. \end{cases}
$$

These statistics are presented in Figure 4.4, together with indications of statistical uncertainty. In practice, the estimated covariance is scaled to give an estimated correlation, $\hat{R}(\phi)/\hat{R}(0)$. For the sample mean, a convenient pointwise measure of uncertainty comes from an acceptance region of the hypothesis $\mu(\phi) = 0$, in model $(M2)$, against a pointwise alternative that $\mu(\phi) \neq 0$. This can be constructed by noticing that under the null hypothesis, $\{\tilde{I}_i(\phi), \ i = 1, 2, \ldots, n\}$ are i.i.d., mean zero, unit variance random variables. An approximate pointwise confidence interval for the spectrum $S(k)$ of $\hat{\eta}_i(\phi)$ can be found by an application of Theorem 4.1. This can in turn be taken to indicate a plausible range for the spectrum of $\eta(\phi)$ in $(M2)$, though it suffices here to consider the rescaled process $\tilde{I}_i(\phi)$ in its own right, without concern about the effects of rescaling. That $\hat{R}(\phi)$ is in the form of a convolution, we have an alternative representation of the sample spectrum as

$$
\hat{S}(k) = \frac{1}{n} \sum_{i=1}^{n} |d^N(\hat{\eta}_i)(k)|^2 (1 + \mathbf{1}_{\{n=0, N/2\}}.
$$
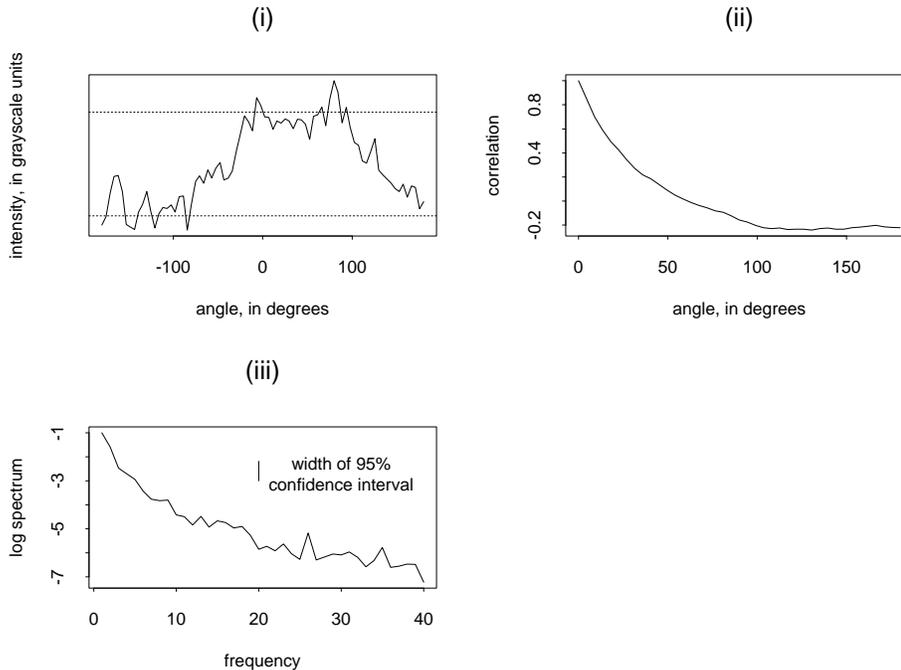
Figure 4.4: (i) Mean standardized stain intensity as a function of ϕ. The dashed lines give a pointwise acceptance region around zero, at a significance level of 5%. (ii) Estimated correlation function. (iii) Estimated log-spectrum.

If $\hat{\eta}_i(\phi)$ were Gaussian then Theorem 4.1 would give

$$
\begin{aligned}
\hat{S}(k) &\sim S(k)\chi^2_{2n}/2n \ \ 0 < R < N/2 \\
\hat{S}(k) &\sim S(k)\chi^2_{n}/n \ \ k = 0, N/2.
\end{aligned}
$$

It is often appropriate to plot $\log(\hat{S}_k)$ and on this scale a $(1-\alpha)$ confidence interval for each $0 < k < N/2$ can be taken as $[\log \hat{S}_k - \delta, \log \hat{S}_k + \delta]$ with $P[|\log(\chi^2_{2n}/2n)| > \delta] = \alpha$. This can be used as an approximate $100(1 - \alpha)\%$ confidence interval, based on the discussion following Theorem 4.1.

In Figure 4.4 we see that the mean function passes above the pointwise 5% significance region around $0°$ and below it around $180°$. The estimated correlation and spectrum tell the same story; that there is a large low frequency contribution to the distribution of $\tilde{I}_i(\phi)$. This implies the sample mean is highly correlated at neighboring values of ϕ, making it hard to interpret the deviation from the pointwise significance

bands for the sample mean. The approximate p-values for the lowest frequencies of the F-test in (4.2) are plotted in Figure 4.5 and show very convincing evidence (p-value = 0.0000) of a non-zero mean at a frequency of one cycle per 360°. Since the higher frequencies are not significant, the data appear insufficient to estimate the mean function to a high accuracy (or else it is simply well modeled by a sinusoid).

4.5 Conclusions

This chapter has described and demonstrated a methodology for the statistical analysis of anisotropy in stained, fixed cells. Image analysis techniques provide a high quality, fast, objective description of the stain intensity compared to the common alternative of manual scoring methods. A test for anisotropy is proposed which is shown to arise from a particular model but which requires only the weaker condition that each cell is an independent, identically distributed replication of the experiment. This test has the possibility of being used to find convincing evidence of patterns in stain intensity data that are not obvious from a visual inspection.

Instrumental to the practical application of this methodology is computer software that is accessible to the scientists who collect such data. For this purpose, a MAT-LAB program was developed by the author, together with a graphical user interface (GUI). The program runs on a Windows, Mac or UNIX platform. The GUI makes the program easy to learn to use, and encourages playing with the parameter settings which can be changed simply using sliders. Such playing, which could formally be called a sensitivity analysis, is an important part of using an algorithm effectively. For example, it takes a little trial and error to find a threshold that appears appropriate to segment the cells from their background. One should further check that the exact choice of threshold has little effect on the final analysis. Having a convenient threshold slider setting on the GUI means there is little effort involved in trying different threshold values. The code is available from the author on request.

This chapter has focused on a particular experiment, described in Fang et al. (1999), for which the methods were developed. Similar approaches could be successful for variations on this particular problem, such as situations where the stained region

of interest is in the interior of the cell rather than the boundary. Statistical techniques for spatial processes would then be involved. Recent advances in the understanding of cells at a genetic and biochemical level can be expected to suggest many more experiments of this type, where a stain is developed to label a molecule that has been implicated in a cell process of interest.

Appendix A

Some Results on Conditional Differentiability in Quadratic Mean

This appendix develops a proof of Proposition A.1. We use the notation and concepts introduced in Chapter 3 to discuss this proposition and to prove two necessary lemmas extending results of Le Cam to a conditional setting.

Proposition A.1 *If $\{P_x(\theta)\}$ is DQM at θ_0 with derivative V and $\{P_{Y|X}(\theta)\}$ is conditionally DQM at θ_0 with derivative W then the joint law, $\{P_{X,Y}(\theta)\}$ is DQM at θ_0 with derivative $V + W$.*

The proposition may not be a surprising result if one thinks of V as playing the same role as the Fisher score function for an observation with law $P_X(\theta_0)$, and W as the conditional score function for an observation with law $P_{Y,X}(\theta_0)$. One might then expect the score function for $P_{X,Y}(\theta_0)$ to be $V + W$. From another point of view, Proposition A.1 is a little surprising. The condition DQM(i) for $\{P_X(\theta)\}$ is Frechet differentiability, viewed as function $\mathbb{R}^d \to L^2(P_X(\theta_0))$, of $Z_X(t) = \sqrt{\frac{dP_X(\theta_0+t)}{dP_X(\theta_0)}}$ with derivative $dZ_X = W$ (Bickel et al., 1993). For $\{P_{Y|X}(\theta)\}$ it is Frechet differentiability of $Z_{Y|X}(t) = \sqrt{\frac{dP_{Y|X}(\theta_0+t)}{dP_{Y|X}(\theta_0)}}$ with derivative $dZ_{Y|X} = W$. The proposition amounts to a product rule for the Frechet derivative, $d(Z_X Z_{Y|X}) = Z_X dZ_{Y|X} + Z_{Y|X} dZ_X$, since $Z_X(0) = Z_{Y|X}(0) = 1$. However, the product rule does not in general apply to Frechet derivatives. For example, a general product rule for Frechet differentiable functions

$\mathbb{R} \to L^2(P)$ for P Lebesgue measure on $(0,1)$, can be seen to require that if $\epsilon(t)$ is a function $\mathbb{R} \to L^2(P)$ with $\|\epsilon(t)\| \to 0$ as $t \to 0$, and $X \in L^2(\mu)$, then $\|X\epsilon(t)\| \to 0$. A counterexample for which $\|X\epsilon(t)\| < \infty$ is provided by

$$\epsilon(t)(\omega) = \frac{1}{\sqrt{\log(1/t)}\omega^{1/4}}1_{\{\omega>t\}},\ 0 < \omega < 1,$$

$$X(\omega) = \frac{1}{\omega^{1/4}},\ 0 < \omega < 1.$$

Then $\|X\epsilon(t)\| \to 1$ as $t \to 0$, even though $\|\epsilon(t)\| \to 0$. Lemma A.1 below can now be seen as one of the properties of Frechet derivatives peculiar to random variables which are the square root of a Radon–Nikodym derivative and which enable the product rule to hold. Pollard (1997) gives an elegant explanation of how these properties may be used to show that DQM implies LAN, by considering Frechet derivatives of functions onto the unit ball. Our treatment is slightly complicated by not assuming absolute continuity, so $\left\| \sqrt{\frac{dP_X(\theta)}{dP_X(\theta_0)}} \right\| \leq 1$ with equality holding only in the limit as $\theta \to \theta_0$, by the assumption of contiguity.

Lemma A.1 *Suppose $\{P_{Y|X}(\theta)\}$ is conditionally DQM at $\theta_0 \in \Theta$, for Θ an open subset of $\mathbb{R}^d$, with derivative W. Then $\mathbb{E}[W \mid X] = 0$ almost surely.*

Proof. Pick $u \in \mathbb{R}^d$ and set $Z_{Y|X}(\tau) = \sqrt{\frac{dP_{Y|X}(\theta_0+\tau u)}{dP_{Y|X}(\theta_0)}}$ for $\tau \in \mathbb{R}$. From conditional DQM(i), since L^2 convergence implies L^1 convergence, in the limit as $\tau \to 0$

$$\frac{1}{\tau}\mathbb{E}[Z_{Y|X}(\tau) - 1 \mid X] = \mathbb{E}[u^T W \mid X] + o(1). \tag{A.1}$$

However, using conditional DQM(ii) we can write

$$\frac{1}{\tau}\mathbb{E}[Z_{Y|X}(t) - 1 \mid X] = \tau \cdot \left\{ \frac{1}{\tau^2}\mathbb{E}\left[-\frac{1}{2}(Z_{Y|X}(\tau) - 1)^2 \mid X\right] + o(1) \right\}. \tag{A.2}$$

Another application of conditional DQM(i) on the right-hand side of (A.2) gives

$$\frac{1}{\tau}\mathbb{E}[Z_{Y|X}(\tau) - 1 \mid X] = \tau\mathbb{E}\left[-\frac{1}{2}u^T WW^T u \mid X\right] + o(\tau). \tag{A.3}$$

A comparison of (A.1) and (A.3), noticing that u is arbitrary, proves the lemma.

Proof of Proposition A.1. DQM(ii) holds for $\{P_{X,Y}(\theta)\}$ as the $P_{X,Y}(\theta_0)$-singular set for $P_{X,Y}(\theta)$ is the union of the respective singular sets for $\{P_X(\theta)\}$ and $\{P_{Y|X}(\theta)\}$. These have masses bounded by DQM(ii) which is assumed to hold for $\{P_X(\theta)\}$ and conditionally for $\{P_{Y|X}(\theta)\}$.

To demonstrate DQM(i), we follow a standard method (Le Cam, 1986), based on the result that if $\{\xi_n\}$ and ξ are real-valued random variables with $\xi_n - \xi = o_p(1)$ and $\mathbb{E}\xi_n^2 - \mathbb{E}\xi^2 \to 0$ then $\mathbb{E}(\xi_n - \xi)^2 \to 0$. For $t \in \mathbb{R}^d$, let $Z_X(t) = \sqrt{\frac{dP_X(\theta_0+t)}{dP_X(\theta_0)}}$, so $Z_X(t) = 1 + t^T V + \epsilon(t)$ with $\frac{1}{|t|^2}\mathbb{E}\epsilon^2(t) \to 0$. Similarly, $Z_{Y|X}(t) = \sqrt{\frac{dP_{Y|X}(\theta_0+t)}{dP_{Y|X}(\theta_0)}}$, so $Z_{Y|X}(t) = 1 + t^T W + \eta(t)$ with $\frac{1}{|t|^2}\mathbb{E}[\eta^2(t)] \to 0$. The joint Radon–Nikodym derivative is given by

$$\frac{dP_{X,Y}(\theta)}{dP_{X,Y}(\theta_0)} = \frac{dP_X(\theta)}{dP_X(\theta_0)}\frac{dP_{Y|X}(\theta)}{dP_{Y|X}(\theta_0)} = Z_X^2(\theta - \theta_0)Z_{Y|X}^2(\theta - \theta_0). \tag{A.4}$$

We can see from the above expressions for $Z_X(t)$ and $Z_{Y|X}(t)$ that

$$Z_X Z_Y(t) = 1 + t^T(V + W) + o_p(|t|). \tag{A.5}$$

Also,

$$
\begin{aligned}
\mathbb{E}[(Z_X Z_Y(t) - 1)^2] &= \mathbb{E}[Z_X^2 Z_Y^2(t)] - 2\mathbb{E}[Z_X Z_Y(t)] + 1 \\
&= 1 - o(|t|^2) - 2\mathbb{E}[1 + \epsilon(t) + \eta(t)] + o(|t|^2) + 1 \quad \text{(A.6)} \\
&= -2\mathbb{E}[\epsilon(t) + \eta(t)] + o(|t|^2) \\
&= t^T(\mathbb{E}[VV^T] + \mathbb{E}[WW^T])t + o(|t|^2) \quad \text{(A.7)} \\
&= t^T\mathbb{E}[(V + W)(V + W)^T]t + o(|t|^2). \quad \text{(A.8)}
\end{aligned}
$$

To get to line (A.6), DQM(ii) was used together with the Cauchy–Schwarz inequality. Line (A.7) requires the equation (A.3) developed in the proof of Lemma A.1, and line (A.8) requires an application of Lemma A.1. Combining (A.5) and (A.8) gives

$$Z_X Z_Y(t) = 1 + t^T(V + W) + \xi(t)$$

with $\frac{1}{|t|^2}\mathbb{E}[\xi(t)] \to 0$ as $t \to 0$. This gives DQM(i) for $\{P_{X,Y}(\theta)\}$.

Bibliography

Alt, W. (1980). Biased random walk models for chemotaxis and related diffusion approximations. *J. Math. Biol.* **9**, 147–177.

Alt, W., Deutsch, A. and Dunn, G. (1997). *Dynamics of Cell and Tissue Motion.* Birkhäuser, Basel.

Alt, W. and Hoffmann, G. (1990). *Biological Motion.* Springer, Berlin.

Anderson, K. I., Wang, Y. and Small, J. V. (1996). Coordination of protrusion and translocation of the keratocyte involves rolling of the cell body. *J. Cell Biol.* **134**, 1209–1218.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. S., Rogers, W. H. and Tukey, J. W. (1972). *Robust Estimates of Location.* Princeton Univ. Press, Princeton.

Barnett, V. D. (1966). Evaluation of the maximum likelihood estimator where the likelihood equation has multiple roots. *Biometrika* **53**, 151–165.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36**, 192–225.

Bibby, B. M. and Sorensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1***, 17–39.*

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore.

Bickel, P. J. and Ritov, Y. (1996). Inference in hidden Markov models, I: Local asymptotic normality in the stationary case. *Bernoulli* **2**, 199–228.

Bickel, P. J. and Ritov, Y. (1997). Local asymptotic normality of ranks and covariates in transformation models. In *Festschrift for Lucien Le Cam* (D. Pollard, E. Torgersen and G. L. Yang, eds.) 43–54. Springer, New York.

Bickel, P. J., Ritov, Y. and Ryden, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Statist.* **26**, 1614–1635.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control.* Holden–Day, San Francisco.

Bray, D. (1992). *Cell Movements.* Garland Publishing, New York.

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston, New York.

Brillinger, D. R. (1997). A particle migrating randomly on a sphere. *J. Theoret. Probab.* **10**, 429–443.

Brillinger, D. R. and Stewart, B. S. (1998). Elephant-seal movements: Modelling migration. *Canadian J. Statist.* **26**, 431–443.

Brosteanu, O., Plath, P. J. and Vicker, M. G. (1997). Mathematical Analysis of Cell Shape. In *Dynamics of Cell and Tissue Motion* (W. Alt, A. Deutsch and G. Dunn, eds.) 29–32. Birkhäuser, Basel.

Byrne, H. M., Cave, G. and McElwain, D. (1998). The effect of chemotaxis and chemokinesis on leukocyte locomotion: A new interpretation of experimental results. *IMA J. Math. Appl. in Medic. and Biol.* **15**, 235–256.

Carlin, B. P., Polson, N. G. and Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *J. Amer. Statist. Assoc.* **87**, 493–500.

Casella, G. and Berger, R. L. (1990). *Statistical Inference.* Wadsworth, Pacific Grove.

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* Wiley, New York.

Dacunha–Castelle, D. and Florens–Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* **9**, 263–284.

Daniels, H. E. (1960). The asymptotic efficiency of a maximum likelihood estimator. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1**, 151–163. Univ. California Press, Berkeley.

DaPrato, G. and Zabczyk, J. (1992). *Stochastic Equations in Infinite Dimensions.* Cambridge University Press, Cambrige.

Davies, R. B. (1973). Asymptotic inference in stationary Gaussian time-series. *Adv. Appl. Prob.* **5**, 469–497.

Del Moral, P. (1996). Nonlinear filtering using random particles. *Theory Probab. Appl.* **40**, 690–701.

Del Moral, P. and Guionnet, A. (1999). Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.* **9**, 275–297.

Dembo, M. (1989). Field theorems of the cytoplasm. *Comments Theor. Biol.* **1**, 159–177.

Dickinson, R. B. and Tranquillo, R. J. (1993a). A stochastic model for adhesion-mediated cell random motility and haptotaxis. *J. Math. Biol.* **31**, 563–600.

Dickinson, R. B. and Tranquillo, R. J. (1993b). Optimal estimation of cell movement indices from the statistical analysis of cell tracking data. *AIChE J.* **39**, 1995–2010.

Dickinson, R. B. and Tranquillo, R. T. (1995). Transport equations and indices for random and biased cell migration based on single cell properties. *SIAM J. Appl. Math.* **55**, 1419–1454.

DiMilla, P. A., Quinn, J. A., Albelda, S. M. and Lauffenburger, D. A. (1992). Measurement of individual cell migration parameters for human tissue cells. *AIChE J.* **38**, 1092–1104.

Doukhan, P. (1994). *Mixing: Properties and Examples.* Springer, New York.

Dufour, J. and Roy, R. (1976). On spectral estimation for a homogeneous random process on the circle. *Stochastic Process Appl.* **4**, 107–120.

Dunn, G. A. and Brown, A. F. (1987). A unified approach to analyzing cell motility. *J. Cell Sci. Suppl.* **8**, 81–102.

Dunn, G. A. and Brown, A. F. (1990). Quantifying cellular shape using moment invariants. In *Biological Motion* (W. Alt and G. Hoffmann, eds.) 10–34. Springer, Berlin.

Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *J. Roy. Statist. Soc. Ser. B* **62**, 3–56.

Durrett, R. (1991). *Probability: Theory and Examples.* Wadsworth, Belmont.

Fang, K. S., Ionides, E., Oster, G., Nuccitelli, R. and Isseroff, R. R. (1999). Epidermal growth factor receptor relocalization and kinase activity are necessary for directional migration of keratinocytes in DC electric fields. *J. Cell Science* **112**, 1967–1978.

Farrell, B. E., Daniele, R. P. and Lauffenburger, D. A. (1990). Quantitative relationships between single-cell and cell-population model parameters for chemosensory migration responses of alveolar macrophages to C5a. *Cell Motil. Cytoskel.* **16**, 279–293.

Ford, R. M. and Lauffenburger, D. A. (1991). Measurement of bacterial random motility and chemotaxis coefficients: II. Application of single-cell-based mathematical model. *Biotechnology and Bioengineering* **37**, 661–672.

Ford, R. M., Phillips, B. R., Quinn, J. A. and Lauffenberger, D. A. (1991). Measurement of bacterial random motility and chemotaxis coefficients: I. Stopped-flow diffusion chamber assay. *Biotechnology and Bioengineering* **37**, 647–660.

Freedman, D., Pisani, R. and Purves, R. (1998). *Statistics* (3rd edition). W. W. Norton, New York.

Gardiner, C. W. (1983). *Handbook of Stochastic Methods.* Springer, New York.

Grenander, U. and Keenan, D. M. (1993). On the shape of plane images. *SIAM J. Appl. Math.* **53**, 1072–1094.

Grenander, U. and Manbeck, K. M. (1993). A stochastic shape and color model for defect detection in potatoes. *J. Comput. Graph. Statist.* **2**, 131–151.

Grenander, U. and Miller, M. I. (1994). Representation of knowledge in complex systems. *J. R. Statist. Soc. B* **56**, 549–603.

Hallin, M., Taniguchi, M., Serroukh, A. and Choy, K. (1999). Local asymptotic normality for regression models with long-memory disturbance. *Ann. Statist.* **27**, 2054–2080.

Harvey, A. C. (1989). Forecasting, structural time series models and the Kalman filter. *Cambridge University Press.*

Hinz, B. and Brosteanu, O. (1997). Periodicity in shape changes of human epidermal keratinocytes. In *Dynamics of Cell and Tissue Motion* (W. Alt, A. Deutsch and G. Dunn, eds.) 21–28. Birkhäuser, Basel.

Hocking, J. G. and Young, G. S. (1961). *Topology.* Addison–Wesley, Reading, MA.

Höpfner, R., Jacod, J. and Ladelli, L. (1990). Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Th. Rel. Fields* **86**, 105–129.

Huttenlocher, A., Sandborg, R. R. and Horwitz, A. F. (1995). Adhesion in cell migration. *Curr. Opin. Cell Biol.* **7**, 697–706.

Jensen, J. L. and Pedersen, J. (1999). Ornstein–Uhlenbeck type processes with non-normal distribution. *J. Appl. Prob.* **36**, 389–402.

Jensen, J. L. and Petersen, N. V. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Statist.* **27**, 514–535.

Jones, R. H. (1981). Fitting a continuous time autoregression to discrete data. In *Applied Time Series Analysis II* (D. F. Findley, ed.) 651–682. Academic Press, New York.

Karlin, S. and Taylor, H. M. (1981). *A Second Course in Stochastic Processes.* Academic Press, New York.

Kendall, D. G. (1974). Pole-seeking Brownian motion and bird navigation. *J. Roy. Statist. Soc. Ser. B* **36**, 365–417.

Kitagawa, G. (1987). Non-Gaussian state-space modelling of non-stationary time series. *J. Amer. Statist. Assoc.* **82**, 1032–1063.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.* **5**, 1–25.

Kloeden, P. E. and Platen, E. (1992). *Numerical Soluion of Stochastic Differential Equations.* Springer, New York.

Kloeden, P. E., Platen, E., Schurz, H. and Sorensen, M. (1996). On effects of discretization on estimators of drrift parameters for diffusion processes. *J. Appl. Prob.* **33**, 1061–1078.

Kreimer, J. and Rubinstein, R. Y. (1988). Smoothed functionals and constrained stochastic approximation.*SIAM J. Num. Anal.* **25**, 470–487.

Lander, E. S. and Green, P. (1987). Construction of multilocus genetic linkage maps in humans. *Proc. Natl. Acad. Sci. USA* **84**, 2363–2367.

Langer, R. S. and Vacanti, J. P. (1999). Tissue engineering: The challenges ahead. *Scientific American* **280**, No. 4, 86–89.

Lauffenburger, D. A. and Linderman, J. J. (1993). *Receptors: Models for Binding, Tracking and Signalling.* Oxford University Press, New York.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

Le Cam, L. (1990). Maximum likelihood: An introduction. *Int. Statist. Review* **58**, 153–171.

Le Cam, L. and Yang, G. L. (1988). On the preservation of local asymptotic normality under information loss. *Ann. Statist.* **16**, 483–520.

Le Cam, L. and Yang, G. L. (1990). *Asymptotics in Statistics.* Springer, New York.

Levin, S. A. (1986). Random walk models of movement and their implications. In *Mathematical Ecology* (T. G. Hallam and S. Levin, eds.) 149–154. Springer, Berlin.

Liang, K. and Zeger, S. L.. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Lieberman, G. M. (1996). *Second Order Parabolic Differential Equations.* World Scientific, River Edge.

Lodish, H., Baltimore, D., Berk, A., Zipursky, S. L. et al. (1995). *Molecular Cell Biology.* Scientific American Books, New York.

Maheshwai, G. and Lauffenburger, D. A. (1998). Deconstructing (and reconstructing) cell migration. *Microscopy Research and Technique* **43**, 358–368.

Matheron, G. (1975). *Random Sets and Integral Geometry.* Wiley, New York.

Mogilner, A. and Oster, G. (1996). The physics of lamellipodial protrusion. *European Biophysical J.* **25**, 47–53.

Mooney, D. J. and Mikos, A. G. (1999). Growing new organs. *Scientific American* **280**, No. 4, 60–65.

Murray, J. D. (1989). *Mathematical Biology.* Springer, Berlin.

Murray, J., Vaurter–Hugast, H., Voss, E. and Soll, D. R. (1992). Three-dimensional motility cycle in leukocytes. *Cell Motil. Cytoskel.* **22**, 211–223.

Nishimura, K. Y., Isseroff, R. R. and Nuccitelli, R. (1996). Human keratinocytes migrate to the negative pole in DC electric fields comparable to those measured in mammalian wounds. *J. Cell Sci.* **109**, 199–207.

Noble, P. B. (1990). Images of cells changing shape: Pseudopods, skeletons and motile behavior. In *Biological Motion* (W. Alt and G. Hoffmann, eds.) 42–67. Springer, Berlin.

Oksendal, B. (1998). *Stochastic Differential Equations.* Springer, New York.

Oliver, T., Dembo, M. and Jacobsen, K. (1995). Traction force in locomoting cells. *Cell Motil. Cytoskel.* **31**, 225–240.

Oster, G. F. (1990). Lateral inhibition models of developmental processes. *Mathematical Biosciences* **90**, 265–286.

Parenteau, N. (1999). Skin: The first tissue-engineered products. *Scientific American* **280**, No. 4, 83–84.

Pievatolo, A. and Green, P. J. (1998). Boundary detection through dynamic polygons. *J. Roy. Statist. Soc. B* **60**, 609–626.

Pollard, D. (1997). Another look at differentiability in quadratic mean. In *Festschrift for Lucien Le Cam* (D. Pollard, E. Torgersen and G. L. Yang, eds.) 305–314. Springer, New York.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77**, 257–285.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (2nd edition). Wiley, New York.

Shenderov, A. D. and Sheetz, M. P. (1997). Inversely correlated cycles in speed and turning in an amoeba: An oscillatory model of cell locomotion. *Biophysical J.* **72**, 2382–2389.

Silverman, B. (1985). *Density Estimation.* Chapman and Hall, London.

Small, C. G. (1996). *The Statistical Theory of Shape.* Springer, New York.

Small, C. G., Wang, J. and Yang, Z. (2000). Eliminating multiple root problems in estimation. *Statist. Science* **15**, 313–341.

Soll, D. and Wessels, D. (1998). *Motion Analysis of Living Cells.* Wiley–Liss, New York.

Stroock, D. W. (1974). Some stochastic processes which arise from a model of the motion of a bacterium. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **28**, 305–315.

Stoock, D. W. and Varadhan, S. R. S. (1979). *Multidimensional Diffusion Processes.* Springer, New York.

Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochasic Geometry and Its Applications* (2nd edition). Wiley, New York.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series.* Springer, New York.

Tranquillo, R. T. and Alt, W. (1996). Stochastic model of receptor-mediated cytomechanics and dynamic morphology of leukocytes. *J. Math. Biol.* **54**, 361–412.

Tukey, J. W. (1977). *Exploratory Data Analysis.* Addison–Wesley, Reading, MA.

Venables, W. N. and Ripley, B. D. (1997). *Modern Applied Statistics with S-Plus* (2nd edition). Springer, New York.

Wessels, D., Vawter–Hugart, H., Murray, J. and Soll, D. R. (1994). Three-dimensional dynamics of pseudopod formation and the regulation of turning during the motility cycle of dictyostelium. *Cell Motil. Cytoskel.* **27**, 1–12.

White, H. (1982). Maximum likelihood estimation of mis-specified models. *Econometrica* **50**, 1–26.

Whittle, P. (1996). *Optimal Control: Basics and Beyond.* Wiley, Chichester.

Wood, S. and Kohn, R. (1998). A Bayesian approach to robust binary nonparametric regression. *J. Amer. Statist. Assoc.* **93**, 203-213.

Yang, G. L. (1997). Le Cam's procedure and sodium channel experiments. In *Festschrift for Lucien Le Cam* (D. Pollard, E. Torgersen and G. L. Yang, eds.) 411–421. Springer, New York.