# Collective Behaviour

Jack Walton

March 2020

*Thesis submitted for the degree of*
*Doctor of Philosophy*

*to the*

*School of Mathematics & Statistics*
*Newcastle University*
*Newcastle upon Tyne*
*United Kingdom*

# *Contents*

# 1

# *Introduction*

## 1.1 Motivation

Many of us have been struck by the inherent beauty of animals moving collectively; starlings gathering at dusk in huge numbers to perform the most mesmerising of ballets, the entire flock moving as if some fluid object; fish forming tight milling structures in defence against predation, changing direction in the blink of an eye and with a flash of silver. At different length scales, and in both the living and non-living domains, startling examples of collective behaviours have been observed (Parrish & Edelstein-Keshet, 1999).

Over the years collective behaviour has become a thriving topic of multidisciplinary research, capturing the imaginations of physicists (Vicsek *et al.*, 1995), biologists (Couzin *et al.*, 2002), mathematicians (Lukeman *et al.*, 2010) and statisticians (Mann, 2011). Our understanding has evolved significantly from early suggestions that collective behaviour results from thought-transference and telepathy between individuals (Selous, 1931). Though we can often explain why animal aggregations are evolutionary advantageous (Giardina, 2008), much less is known about how these structures are formed and maintained.

Much work has been invested in developing theoretical models which seek to explain emergent behaviour by interactions at an individual level. Such models have shown that individual interactions are sufficient to produce group-level structures. Many different simulations, implementing disparate interaction rules, are able to produce behaviour reminiscent of real flocking systems. However, these models have largely only been verified with comparison to empirical observation at a qualitative level, and a thorough quantitative comparison between field data and theory has been lacking.

Figure 1.1: A particularly startling example of a starling murmuration, captured near Gretna in the Scottish Borders. Photograph: Owen Humphreys/PA.

In recent years technological and methodological advances have made it possible to capture the movements of large groups of animal aggregates (Ballerini *et al.*, 2008). With this data, it is only now that we are in a position to make a robust comparison between model predictions and real-world observations.

## 1.2 Overview of thesis

We begin this thesis by giving the reader a review of the literature surrounding collective behaviour. Important results and ideas of the field are introduced and discussed. After relaying the main results from the literature we discuss open problems and the future of research in the field. Finally, we explain how our work will fit into the current research landscape using techniques from Bayesian statistics.

Chapter 3 will introduce the reader to the field of Bayesian statistics. Important results, techniques and algorithms from the field will be outlined as well as problems that the practitioner may encounter.

# 2

# *Literature review*

There is a large body of literature relating to the phenomenon of collective behaviour. Particularly unique to this literature is the variety of backgrounds in which the authors are trained. Biologists, physicists, applied mathematicians and statisticians have all made significant contributions to the field.

In this chapter we shall discuss some of the most important ideas and results from the literature surrounding collective behaviour.

## 2.1 Biological function

Behaving as a group can bring many advantages to the individuals involved. One classically considered benefit of aggregation is in the defence against predation. Shoaling fish groups have the ability to confuse predators, as predators have difficulty selecting an individual target (Landeau & Terborgh, 1986). As well as a confusion effect groups of individuals can be more vigilant than a single individual, allowing for the earlier detection of predators (Pitcher & Parrish, 1993). Despite these advantages, groups may in fact attract predators (Wittenberger & Hunt, 1985).

As well as providing defence against predation grouping can aid in foraging for food; collections of individuals are able to gather more information than lone individuals. In addition to foraging collective motion aids group navigation and migration (Simons, 2004). For birds group navigation often brings an energetic advantage as individuals can work to form aerodynamically efficient shapes (Weimerskirch *et al.*, 2001).

## 2.2 Theoretical models

Models of collective behaviour can largely be divided into two classes: Lagrangian and Eulerian. These descriptions are analogous to the models of fluid dynamics, where Lagrangian models consider the flow in terms of interactions of fluid parcels and Eulerian models consider the changing fluid properties at a given point in space and time. In the context of collective behaviour, Lagrangian models simulate the movements of interactions and Eulerian models typically model the density of a group through space and time.

### 2.2.1 *Lagrangian models*

So called agent-based models (ABMs), also referred to as Lagrangian models, have proven a useful tool in modelling collective behaviours. In these models the behaviour of an agent is modelled at the individual level. An agent's behaviour is determined by social interactions with neighbouring individuals. Examples of typical interactions include the desire to move in the same direction as neighbours (alignment, or orientation), the desire to avoid collisions (repulsion) and a desire to remain close to neighbours (attraction). Along with specifying social behaviours ABMs suggest methods by which agents identify neighbours to interact with. Models also specify how an individual identifies neighbours with which to interact as; within a certain distance (metric distance); positioned inside a field of vision or as one of a fixed number of closest agents (topological distance).

In a pioneering paper, Aoki (1982) developed an ABM to simulate the movements of schooling fish in two-dimensions. Here it was shown that collective behaviour can arise from simple interactions at an individual level and without the need of a leader. The model simulated zonal interactions in which the area around an individual is partitioned into zones of repulsion, alignment and attraction. The partitioning of space in this way is illustrated in Figure 2.1 and has remained a popular idea in following literature. As well as zonal interactions this model accounted for fish having incomplete fields of vision. Further models were also devised to simulate fish schools (Okubo, 1986; Huth & Wissel, 1992).

Reynolds (1987) formulated a theoretical model, motivated by the production of computer animations, which described the movement of flocking birds in three-dimensional space. To produce more aesthetically pleasing animations, the software, "Boids", included sophistications such as banking during turns. This focus on developing simulations which produce elegant behaviour made rigorous scientific analysis difficult. Interestingly, Tim Burton's 1992 Batman Returns used a modified version of the Boids software to simulate

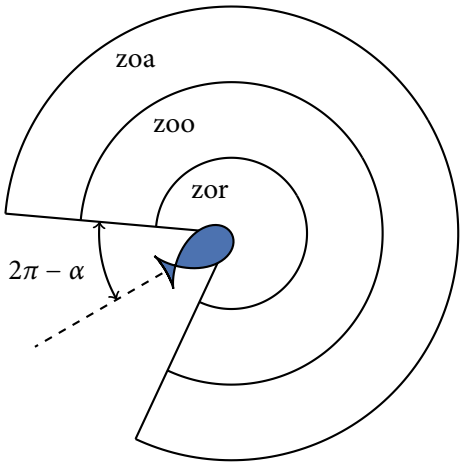



Figure 2.1: An illustration of the area around an agent partitioned into multiple zones. zor: zone of repulsion, zoo: zone of orientation (or alignment), zoa: zone of attraction. The missing segment behind the agent represents a blind zone into which it cannot see.

animations of bat swarms and penguin flocks.

Motivated by research within statistical physics, Vicsek *et al.* (1995) introduced a simple two-dimensional model in which self-propelled particles move with a fixed absolute velocity and align with neighbours within an interaction zone. This model is commonly referred to as "Vicsek Model" (VM), which we shall later use to formulate our "Generalised Vicsek Model" (GVM). Despite its simplicity this model produces complex-behaviour resembling that of a real biological system. Vicsek *et al.* (1995) investigated the phase transition between ordered and disordered motion as the density of particles and noise in the system varied.

Later, models were developed to explore the movements of mammals and other vertebrate groups. Couzin *et al.* (2002) showed major group-level behavioural changes as minor changes in individual interaction rules were made. With small changes in the model parameters, groups transitioned from disordered, swarm like behaviour, to toroidal milling structures, to forming dynamic and highly parallel groups. Further research was made by Couzin *et al.* (2005) which investigated how leaders influence the motion of travelling groups. This work showed that only a small proportion of leaders are necessary to guide groups with a high degree of accuracy. Further results investigated how groups of individuals make collective decisions in the face of conflicting desires.

As a method for exploring collective behaviour, Lagrangian models are very appealing in their intuitiveness and in the ease of implementing explicit behavioural rules. Though for many years the simulation and exploration of these models was limited by computing power; modern computation allows for the simulations of large groups over many time

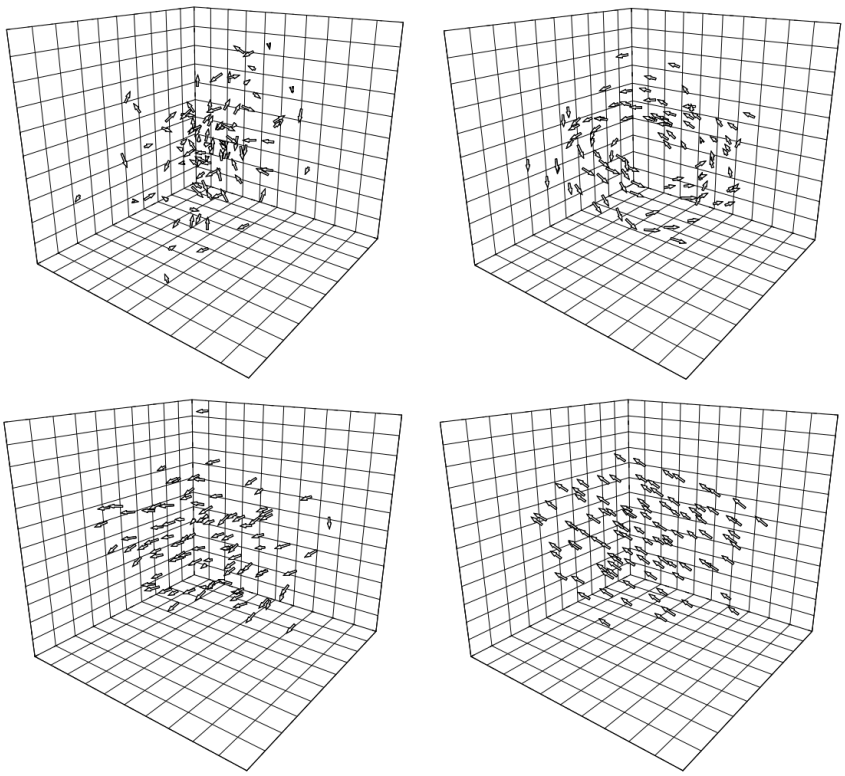


Figure 2.2: Taken from Couzin *et al.* (2002), the different steady-state solutions (swarm, torus, dynamically parallel and highly parallel) obtained by making small changes to model parameters of a three-dimensional flocking model.

steps. With these advances in computing, and a growing interest in the field, a significant proportion of the literature focuses on the analysis and exploration of agent-based models.

### 2.2.2 *Eulerian models*

Sometimes known as continuum models, Eulerian models are complementary to the Lagrangian method. This approach describes a group by the density of organisms at a point in space. Eulerian models are typically constructed of a set of partial differential equations which describe how density develops over time.

One such Eulerian approach by Guernon & Levin (1993) modelled the movements of large groups of wildebeests. The predictions of the model were compared with aerial observations of migrating wildebeest in the Serengeti. The large-scale front patterns seen in the aerial photography were reproduced by the model. However, the ability of a model to reproduce an observable pattern isn't sufficient evidence that the model is correct.

Eulerian models have also been used to analyse vortex solutions (Topaz & Bertozzi,

2004) and stationary clump solutions (Topaz *et al.*, 2006).

However, the Eulerian approach is limited. Most analyses are restricted to 1-dimension and the approach has not proven appropriate for modelling groups of low densities. With this in mind, and with the advantages of the Lagrangian approach, in this thesis we will concentrate entirely on modelling in the Lagrangian framework.

## 2.3 Empirical studies

Real data of animal aggregations is essential to ensure that theoretical models are falsifiable. The emergence of a desired pattern from simulation is not sufficient evidence that a model is correctly capturing the interactions of individuals. This observation is further compounded by the understanding that models employing different local interactions can produce similar looking behaviour at the group level.

Thorough comparison between real data and model predictions have proven difficult largely because of the scarcity of appropriate data. The collection of suitable data can be a complicated and convoluted process. Taking observations in the field is technically demanding, requiring the precise calibration of sensitive measurement equipment, not to mention the additional difficulty of the typically three-dimensional nature of animal aggregations. Collecting data in a laboratory setting seems an obvious workaround, however this imposes restrictions on the types of behaviour which can be captured. A laboratory may be an appropriate environment to capture the movements of fish in a tank, but it certainly isn't appropriate to capture the movements of flocking of birds. Despite the difficulties associated with collecting data, significant effort has been made to track the movements and dynamics of groups of individuals.

Initial work was limited to tracking small numbers of individuals in groups. In these studies individuals were not linked through frames and hence the collected data had no dynamic component. The first breakthrough came from Cullen *et al.* (1965) who used stereo photography to record the positions of fish in three dimensions.

Fish are an appealing subject to study as experiments are easily conducted in a laboratory setting. Furthermore, the movements of fish can effectively be restricted to two dimensions by conducting the experiments in shallow water. Because of these benefits, further research also concentrated on fish (Van Long *et al.*, 1985; Partridge *et al.*, 1980). Having collected empirical data, these studies investigate properties such as the distance of individuals to their nearest neighbour, or the direction toward their nearest neighbour. Empirical studies were also made of small groups of flocking birds, with similar statistics and properties

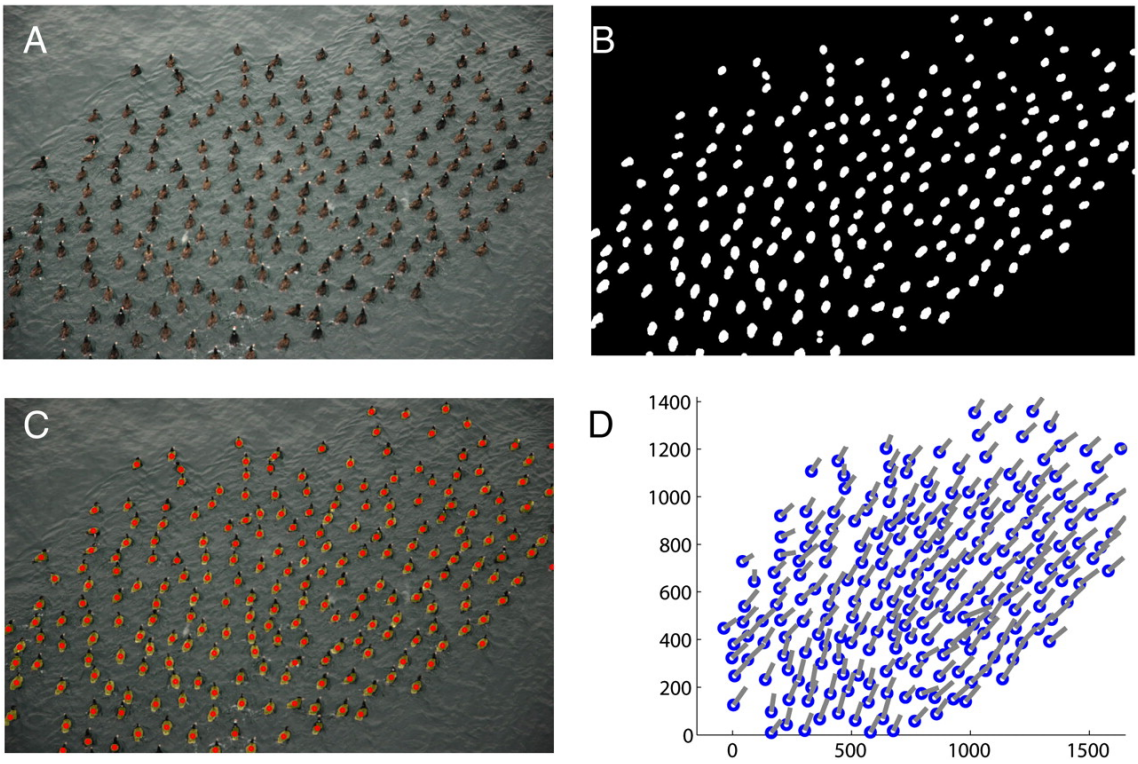



Figure 2.3: Image field data and the process of transformations to extract positions of a flock of surf scoters (Lukeman *et al.*, 2010).

realised (Major & Dill, 1978; Budgey, 1998).

More recently, a breakthrough study by Ballerini *et al.* (2008) reconstructed the three dimensional positions of flocks of starlings consisting of up to 3, 000 individual members. The study made a static analysis of the resulting 3D dataset. Their analysis suggested that interactions are not dependent on metric distance (interactions with agents within a fixed distance), as most models in the literature assume, but on a topological distance (interaction with a fixed number of closest agents, irrespective of distance). This analysis suggested that on average a starling interacts with between six and seven of its closest neighbours. They argue that a topological interaction improved group cohesion when under attack from predation.

A significant contribution to the field was made by Lukeman *et al.* (2010), whom collected and analysed data of large numbers of diving ducks interacting on the surface of a lake. Crucially, this dataset tracked individuals between frames and therefore allowed the reconstruction of a bird's trajectory through space and time. This data showed an increase by factor of 10 the number of individuals which could be reliably tracked though time (Lukeman, 2009). The resulting analysis focused on fitting zonal models which were best able

to reproduce the spatial and angular neighbour-distributions seen in the collected data. It was concluded that a zonal repulsion-alignment-attraction model with an additional frontal interaction was best able to reproduce the desired spatial and angular neighbour-distributions.

Analysis of empirical data has so far focused on properties of individuals such as nearest neighbour distances or angular neighbour densities. Research has then focused on fitting models which are best able to replicate these properties. With technological advances we expect that more and more empirical data will become available in the future.

## 2.4 Numerical studies

Mann (2011) acknowledged that an important aspect of model fitting is knowing the associated uncertainty of inferred parameters. He continued to stress the importance of quantifying uncertainty in parameter inference on collective behaviour models, as the associated empirical datasets often have high levels of noise.

With the importance of capturing uncertainty in mind, Mann demonstrated a fully Bayesian approach to parameter inference on data simulated from a collective behaviour model. After ten timesteps the simulated data transitioned from disordered motion to a steady state rotating mill. The agents in this model moved under some weighted sum of alignment and attraction.

Mann (2011) compared the ability to infer the weighting parameter, interaction radius and other properties of the agents in two situations: before and after the achievement of steady state. It was discovered that the interaction radius could not be reliably inferred when the agents had formed the rotating mill structure, although it could be inferred in the disordered motion before steady state.

# 3

# *Bayesian statistics*

In this thesis we utilise techniques from Bayesian inference to fit theoretical models of collective behaviour to real data. Bayesian inference allows the practitioner to capture uncertainty about fitted model parameters. In addition to this the Bayesian framework permits flexible model structures and potential inclusion of expert information via the prior distribution. With this we seek to fit empirical data to a generalisation of a popular model from the literature.

In this chapter we shall introduce and give overviews of some important concepts of Bayesian inference, outline schemes which can be used to infer model parameters, and discuss some of the problems which may arise and how we might address them.

## 3.1 Bayesian inference

Using Bayesian inference we wish to quantify beliefs and uncertainties about parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$, using data $\boldsymbol{x}$ which we observe. Given this observed data, the likelihood function for the parameters is defined

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = f(\boldsymbol{x} \,|\, \boldsymbol{\theta}).$$

The likelihood quantifies the probability distribution of the data in terms of the parameters. We may then specify our prior knowledge about the parameters $\boldsymbol{\theta}$ through the prior distribution $\pi(\boldsymbol{\theta})$. Bayes Theorem can then be used to incorporate both the likelihood function

and our prior beliefs, to form the posterior distribution

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{\pi(\boldsymbol{\theta})L(\boldsymbol{\theta}|\boldsymbol{x})}{\int_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta})L(\boldsymbol{\theta}|\boldsymbol{x})\,\mathrm{d}\boldsymbol{\theta}}. \tag{3.1}$$

Because the integral in the denominator is not a function of $\boldsymbol{\theta}$, we may consider it a constant of proportionality and express our posterior beliefs as proportional to the product of the likelihood and prior, that is

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto \pi(\boldsymbol{\theta}) \times L(\boldsymbol{\theta}|\boldsymbol{x})$$
$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

## 3.2 Markov chain Monte Carlo (MCMC)

For the most part, the normalising constant (given in the denominator of Equation (3.1)) will have multiple dimensions, not produce a density function of standard form, and be difficult to evaluate in all but the most trivial cases. Markov chain Monte Carlo algorithms provide methods to sample from the targeted density $\pi(\boldsymbol{\theta}|\boldsymbol{x})$, whilst avoiding evaluating the troublesome normalising constant.

### 3.2.1 *Gibbs sampling*

One may use the full conditional distributions of parameters to sample from a multivariate density. Doing so is to implement the Gibbs algorithm. So, instead of sampling from the full posterior, we sample from the conditional posteriors of the parameters one at a time. The Gibbs algorithm is useful when the conditional densities can be expressed in standard form and are easy to sample form.

Say we wish to target the density $\pi(\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$, where the full conditional densities are $\pi(\theta_i|\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p)$, for $i = 1, \ldots, p$, then we may use the Gibbs sampler, as described in Algorithm 1.

### 3.2.2 *Metropolis-Hastings*

The Metropolis-Hastings algorithm is another MCMC scheme. The algorithm was introduced by Metropolis *et al.* (1953), and this work was later generalised by Hastings (1970). The algorithm works by constructing a Markov chain which has stationary distribution equivalent to the distribution of interest.

**Algorithm 1** Gibbs

0. Initialise chain with $\boldsymbol{\theta}^0$. Set $j = 1$.
1. Generate $\boldsymbol{\theta}^j$ from $\boldsymbol{\theta}^{j-1}$ by simulating from:

$$\begin{aligned}
{\theta_1}^j &\sim \pi({\theta_1}^j | {\theta_2}^{j-1}, \ldots, {\theta_p}^{j-1}, \boldsymbol{x}) \\
{\theta_2}^j &\sim \pi({\theta_2}^j | {\theta_1}^j, {\theta_3}^{j-1}, \ldots, {\theta_p}^{j-1}, \boldsymbol{x}) \\
&\vdots \\
{\theta_p}^j &\sim \pi({\theta_n}^j | {\theta_1}^j, \ldots, {\theta_{p-1}}^{j-1}, \boldsymbol{x})
\end{aligned}$$

2. Increment $j$ to $j + 1$. Repeat from step 1.

**Algorithm 2** Metropolis-Hastings

0. Initialise chain with $\boldsymbol{\theta}^0$. Set $j = 1$.
1. Propose $\boldsymbol{\theta}^*$ by sampling from $q(\cdot|\boldsymbol{\theta}^{j-1})$, where $q$ is some proposal distribution
2. Construct the acceptance probability $\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{j-1})$ as

$$\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{j-1})} \frac{L(\boldsymbol{\theta}^*|\boldsymbol{x})}{L(\boldsymbol{\theta}^{j-1}|\boldsymbol{x})} \frac{q(\boldsymbol{\theta}^{j-1}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{j-1})}\right\}.$$

3. With probability $\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{j-1})$ set $\boldsymbol{\theta}^j = \boldsymbol{\theta}^*$, otherwise set $\boldsymbol{\theta}^j = \boldsymbol{\theta}^{j-1}$.
4. Increment $j$ to $j + 1$. Repeat from step 1.

The algorithm begins by initialising the chain with parameters $\boldsymbol{\theta}^0$. Next, the algorithm proposes new values $\boldsymbol{\theta}^*$ from a proposal distribution, $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{j-1})$, which is chosen to have the same support as the target distribution. After this, the proposed values $\boldsymbol{\theta}^*$ are either accepted or rejected, depending on the evaluation of the acceptance probability $\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{j-1})$. Because the acceptance probability depends on a ratio of $\pi(\cdot|\boldsymbol{x})$, the normalising constants cancel and therefore the target distribution only has to be known to a constant of proportionality. Metropolis-Hastings is described more formally in Algorithm 2.

### *Choosing a Proposal Distribution*

The practitioner must choose a suitable proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$. Ideally the choice of proposal distribution will give rapid convergence to $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ and efficiently explore the support of $\pi(\boldsymbol{\theta}|\boldsymbol{x})$.

A special case of Metropolis-Hastings arises when the proposal distribution is symmetric, that is

$$q(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = q(\boldsymbol{\theta}|\boldsymbol{\theta}^*).$$

In this case we observe cancellation in the acceptance ratio, as it simplifies to become

$$\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{j-1})}\frac{L(\boldsymbol{\theta}^*|\boldsymbol{x})}{L(\boldsymbol{\theta}^{j-1}|\boldsymbol{x})}\right\}.$$

Another special case of Metropolis-Hastings is the random walk sampler. In this case proposals are realised as

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}^{j-1} + \boldsymbol{\omega}^{j-1},$$

where the $\boldsymbol{\omega}$ are drawn from

$$\boldsymbol{\omega}^{j-1} \sim \mathcal{N}_p(\mathbf{0}, \Sigma),$$

and $\mathcal{N}_p$ denotes a $p$-dimensional multivariate normal distribution. The parameter $\Sigma$ is called the tuning parameter and controls how the chain moves around the parameter space. Mixing describes how efficiently the chain moves around the sample space and how long it takes for the chain to converge to the target distribution.

Crucially then, the parameter $\Sigma$ controls the mixing of the chain. So, naturally, we wish to select some optimum $\Sigma$ to try and improve mixing. If the target distribution is Gaussian, it has been shown that 0.234 is an optimum acceptance probability (Roberts & Rosenthal, 2001). In an attempt to tune $\Sigma$ to obtain the optimum acceptance probability, a common technique is to use

$$\Sigma = \frac{2.38^2}{2}\widehat{\text{Var}}(\boldsymbol{\theta}|\boldsymbol{x}).$$

### *Convergence Diagnostics*

Though there are theoretical methods to assess the convergence of chains, it is an attractive idea to analyse the output of our schemes in an attempt to assess whether the chains have converged. One of the simplest informal methods to assess convergence is to inspect the trace plots of the scheme and check for any irregularities. It is also good to use autocorrela-

tion plots to assess autocorrelation between samples at different lags.

One way to lower autocorrelation between samples is to thin the output. When thinning, every $k$-th sample from a chain is kept and the remaining samples are discarded. Another common technique is to allow for a burn-in period. The purpose of a burn-in period is to discard any samples from before the chain has converged.

### *Blocking Parameters*

In the schemes considered so far the proposal, acceptance and rejection of the entire parameter space $\boldsymbol{\theta}$ happened simultaneously. This approach becomes inefficient for high-dimensional problems. Consider that as the dimension of the problem increases, the chances of proposing a value $\theta_i^*$ in the tails of the posterior distribution increases. Increasing the chance of proposing a component $\theta_i$ out in the tails of the distribution in turn decreases the acceptance rate of the chain and leads to slower convergence.

To overcome this problem the parameter space $\boldsymbol{\theta}$ can be split into blocks of parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d$ which are proposed and accepted or rejected separately. There are no theoretical results which determine how best to block the parameter space, though typically blocks are chosen to contain related parameters.

Blocking doesn't reduce the risk of a parameter being proposed in the tails of the distribution, however when such a proposal does occur only a subset of all the chains are affected. In this way blocking can alleviate the lower acceptance rates that are associated with high dimensionality.

There is, however, an additional computational cost that comes with blocking parameters. Consider that if the parameter space $\boldsymbol{\theta}$ is partitioned into $d$ blocks, then for each iteration of the scheme the acceptance ratio, and therefore the likelihood, proposal density and prior distribution must be evaluated $d$ times. In application the practitioner will likely seek some compromise between acceptance rate and computation time.

# 4

# *Directional statistics*

Circular data arises naturally in the study of collective behaviour. In particular, a researcher in the field will undoubtedly encounter data describing the direction of motion of individuals. Given any dataset, the first instinct of the practitioner is to summarise and visualise the data. However, such a researcher should proceed with caution — circular data cannot be treated as if it were its linear counterpart.

In this short chapter we shall consider why standard techniques, methods and summaries are inappropriate to use with circular data. After this realisation, we proceed to introduce some useful techniques which can be used to handle and visualise directional data.

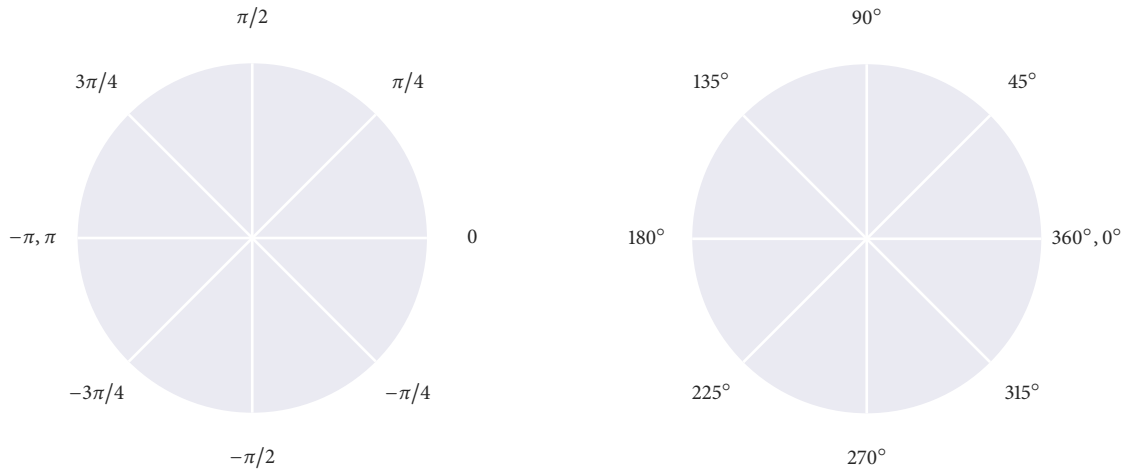
## 4.1 Conventions

Directions can be represented as rotations with respect to some zero–direction, or origin. The practitioner is free to chose the zero–direction as they feel appropriate. In a similar way, the practitioner may choose whether a clockwise or anti–clockwise rotation is taken as the positive direction.

Recall that angles may be represented in units of degrees or radians. To convert between degrees and radians we may multiply by a factor of $\pi/180°$.

In this thesis we shall define the zero angle as the direction from $(0, 0)$ and along the positive $x$–axis. For the most part, we shall measure angles in units of radians, and take anti–clockwise rotations as the positive direction. The schematics of this setup are illustrated in Figure 4.1(a). Occasionally, we shall appeal to degrees and their comparative intuitiveness,

and in these cases we shall use the setup illustrated in Figure 4.1(b).




(a) (b)

Figure 4.1: Comparing (a): radian and (b): degree measuring conventions adopted in this thesis.

## 4.2 Visualisation

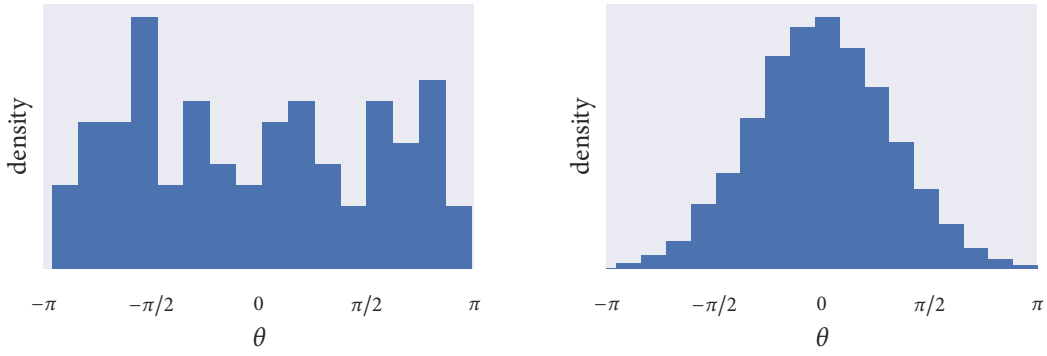
In possession of a dataset, one of the first instincts of the scientist is to visualise their data. The researcher is undoubtedly familiar with a large number of graph types. Yet choosing the most suitable graph to display a given dataset is crucial in making an informative plot.

Traditional histograms are not very good for visualising directional data. Polar histograms (sometimes known as rose plots) make the visualisation of such data easier. Instead of using bars, as the histogram does, the rose plot bins data into sectors of a circle. The area of each sector is proportional to the frequency of data points in the corresponding bin.

To advocate the advantages of the rose plot we shall visualise two randomly generated datasets. The first dataset consists of 100 realisations from a uniform $U(-\pi, \pi)$ distribution, and the second dataset consists of $10,000$ draws from a normal $N(0,1)$ distribution.

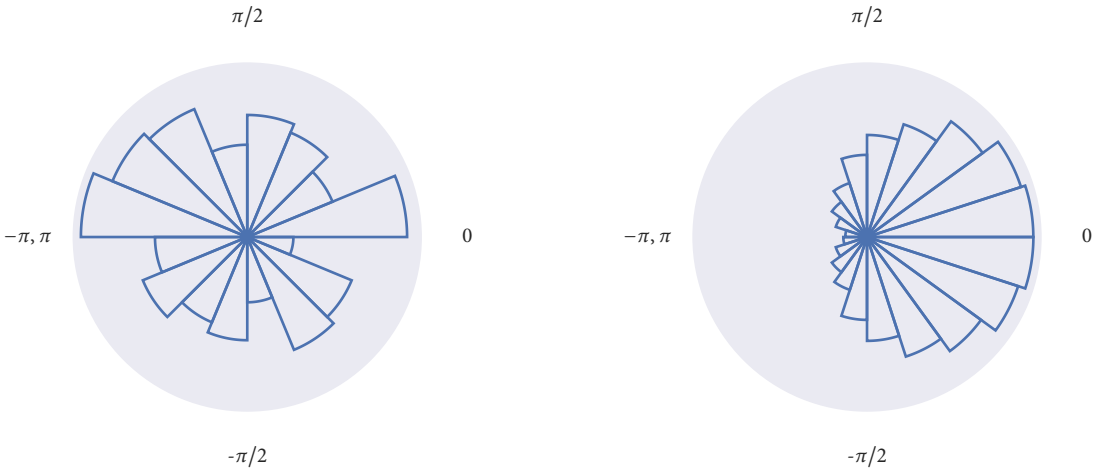
In Figure 4.2 we visualise the two datasets using traditional histogram plots. From this figure we get a good idea of the distribution of the data, however we get no sense of direction. The traditional histogram leaves us to interpret the directions ourselves.

In Figure 4.3 we visualise the same data as before. Here we also get a good idea of how the directions are distributed. However, using the rose plot means we get a very intuitive representation of direction. We therefore consider the rose plot a very useful method of displaying circular data which is easy to interpret.




(a) (b)

Figure 4.2: Using histograms to visualise (a): 100 samples from $U(-\pi, \pi)$ and (b): 10,000 draws from $N(0,1)$.




(a) (b)

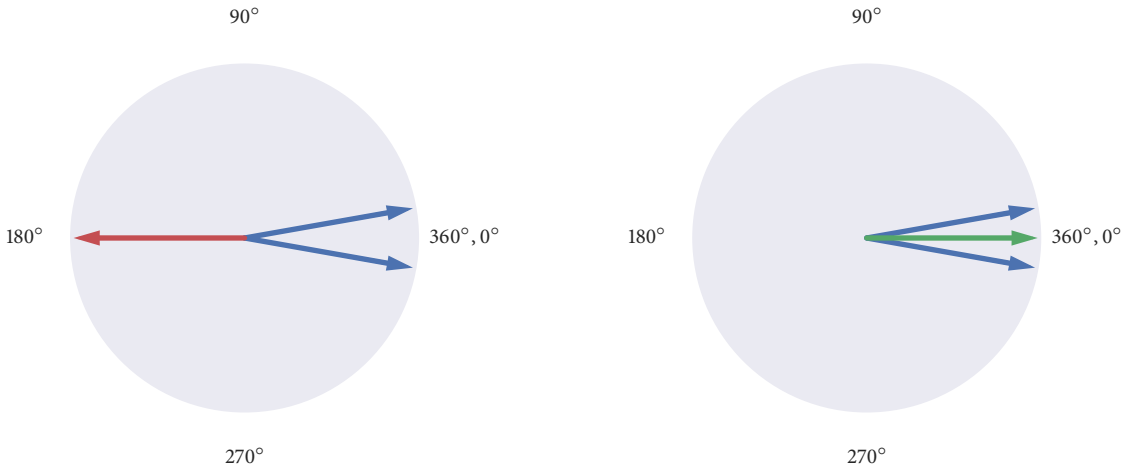
Figure 4.3: Using polar histograms to visualise the dataset generated from (a): $U(-\pi, \pi)$ and (b): $N(0,1)$.

## 4.3 Summary statistics

Summary statistics are a useful tool to give an idea of the general characteristics of a dataset. Probably the first statistic which we learn to compute is the arithmetic mean. The arithmetic mean, however, is not an appropriate measure to use with circular data.

Consider that we wish to take an average of the angles 10°and 350°. Using the arithmetic mean we compute an average of 180°— however, this average points in the opposite direction to which we intuitively expect. In Figure 4.4(a) we visualise this surprising result.

Before introducing the circular mean it is first necessary to introduce the atan2 function.




(a) Arithmetic mean (red).

(b) Circular mean (green).

Figure 4.4: Computing the average of two angles with two different methods. The blue arrows (10°and 350°) represent the directions to be averaged and the green and red arrows show the average computed by the method.

The atan2 function dates back to the Fortran programming language (Organick, 1966). It was introduced to overcome some of the inconveniences inherent in the atan (or $\tan^{-1}$) function. For starters, the inverse tangent function has codomain $(-\pi/2, \pi/2)$, though we are often interested in directions in the range $(-\pi, \pi]$. In addition to this, the arctan function is not quadrant–aware — it cannot distinguish between opposite directions (directions which differ by $\pi$ radians). As an example, consider calculating the direction from the $x$–axis to the ray extending from the origin to $(1, 1)$. Naturally, we'd reach for $\tan^{-1}$ to compute the angle as $\tan^{-1}(1/1) = \pi/4$, as expected. Now, consider that we wish to calculate the direction from the $x$–axis to the ray extending from $(0, 0)$ to $(-1, -1)$. By inspection, or intuitively, we expect an answer of $-3\pi/4$ — however, we compute the answer as $\tan^{-1}(-1/-1) = \pi/4$. The angle calculated using the inverse tangent function points in the opposite direction to what we expect.

The atan2 function, however, does not have these shortcomings. The function is constructed to be quadrant–aware: correcting the computations of $\tan^{-1}$ to return the directions we intuitively expect. It does this by adding a correction term that depends on which quadrant contains the point $(x, y)$. The handling of the four quadrants by atan2 is visualised in Figure 4.5. With these considerations, atan2 can be realised as:

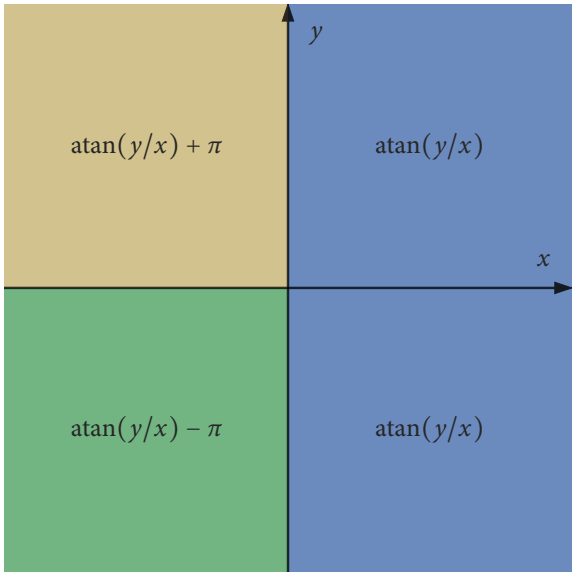



Figure 4.5: An illustration of the quadrant corrections made by atan2.

$$\operatorname{atan2}(y,x) = \begin{cases} \operatorname{atan}(y/x) & \text{if } x > 0, \\ \operatorname{atan}(y/x) + \pi & \text{if } x < 0 \text{ and } y \geq 0, \\ \operatorname{atan}(y/x) - \pi & \text{if } x < 0 \text{ and } y < 0, \\ \pi/2 & \text{if } x = 0 \text{ and } y > 0, \\ -\pi/2 & \text{if } x = 0 \text{ and } y < 0, \\ \text{undefined} & \text{if } x = 0 \text{ and } y = 0. \end{cases} \tag{4.1}$$

When averaging a set of angles we must not use the arithmetic mean. Instead, we must refer to the circular mean. Given a set of angles $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$, we may compute their circular mean as:

$$\langle \boldsymbol{\theta} \rangle = \operatorname{atan2}\left( \frac{1}{n} \sum_{j=1}^{n} \sin\left(\theta_j\right), \frac{1}{n} \sum_{j=1}^{n} \cos\left(\theta_j\right) \right), \tag{4.2}$$

where the atan2 function is defined in Equation (4.1).

The definition of the circular mean given in equation Equation (4.2) works by converting the angles into Cartesian co–ordinates — representing the directions as points on the unit circle. The centre of mass of the Cartesian co–ordinates is then computed, and the resulting position is converted back to a direction, resulting in our mean angle.

# *Bibliography*


Aoki, I. 1982 A simulation study on the schooling mechanism in fish. *Bulletin of the Japanese Society of Scientific Fisheries* **48** (8), 1081–1088.

Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., Viale, M. & Zdravkovic, V. 2008 Interaction ruling animal collective behaviour depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the National Academy of Sciences of the United States of America* **105** (4), 1232–1237.

Budgey, R. 1998 Three dimensional bird flock structure and its implications for birdstrike tolerence in aircraft. *International Bird Strike Proceedings Committee* **24**, 207–220.

Couzin, I. D., Krause, J., Franks, N. R. & Levin, S. A. 2005 Effective leadership and decision making in animal groups on the move. *Nature* **433**, 513–516.

Couzin, I. D., Krausew, J., Jamesz, R., Ruxtony, G. D. & Franksz, N. R. 2002 Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology* **218**, 1–11.

Cullen, J. M., Shaw, E. & Baldwin, H. A. 1965 Methods for measuring the three-dimensional structure of fish schools. *Animal Behaviour* **13** (4), 534–536.

Giardina, I. 2008 Collective behavior in animal groups: theoretical models and empirical studies. *HFSP Journal* **2** (4), 205–219.

Guernon, S. & Levin, S. A. 1993 Self-organization of front patterns in large wildebeest herds. *Journal of Theoretical Biology* **165** (4), 541–552.

Hastings, W. K. 1970 Monte carlo sampling methods using markov chains and their applications. *Biometrika* **57** (1), 97–109.

Huth, A. & Wissel, C. 1992 The simulation of the movement of fish schools. *Journal of theoretical Biology* **156**, 365–385.

Landeau, L. & Terborgh, J. 1986 Oddity and the fhconfusion effectfh in predation. *Animal Behaviour* **34** (5), 1372–1380.

Lukeman, R. 2009 Modelling collective behaviour in animal groups: From mathematical analysis to field work. PhD thesis, University of British Columbia.

Lukeman, R., Li, Y.-X. & Edelstein-Keshet, L. 2010 Inferring individual rules from collective behavior. *Proceedings of the National Academy of Sciences of the United States of America* **107** (28), 12576–12580.

Major, P. F. & Dill, L. M. 1978 The three-dimensional structure of airborne bird flocks. *Behavioral Ecology and Sociobiology* **4** (2), 111–122.

Mann, R. P. 2011 Bayesian inference for identifying interaction rules in moving animal groups. *PLoS ONE* **6** (8).

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. 1953 Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21** (6), 1087–1092.

Okubo, A. 1986 Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Advanced Biophysics* **21**, 1–94.

Organick, E. I. 1966 *A Fortran IV Primer*. Addison-Wesley.

Parrish, J. K. & Edelstein-Keshet, L. 1999 Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science* **284**, 99–101.

Partridge, B. L., Pitcher, T., Cullen, J. M. & Wilson, J. 1980 The three-dimensional structure of fish schools. *Behavioral Ecology and Sociobiology* **6** (4), 277–288.

Pitcher, T. J. & Parrish, J. K. 1993 *Behaviour of Telost Fishes*. Chapman and Hall.

Reynolds, C. W. 1987 Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics* **24** (4), 25–34.

Roberts, G. O. & Rosenthal, J. S. 2001 Optimal scaling for various metropolis-hastings algorithms. *Statistical Science* **16** (4), 351–367.

Selous, E. 1931 *Thought-Transference (Or What?) in Birds*. Constable & Co.

Simons, A. M. 2004 Many wrongs: the advantage of group navigation. *TRENDS in Ecology and Evolution* **19** (9), 453–455.

Topaz, C. M. & Bertozzi, A. L. 2004 Swarming patterns in a two-dimensional kinematic model for biological groups. *SIAM Journal on Applied Mathematics* **65** (1), 152–174.

Topaz, C. M., Bertozzi, A. L. & Lewis, M. A. 2006 A nonlocal continuum model for biological aggregation. *Bulletin of Mathematical Biology* **68**, 1601–1623.

Van Long, L., Aoyama, T. & Inagaki, T. 1985 A stereo photographic method for measuring the spatial position of fish. *Bulletin of the Japanese Society of Scientific Fisheries* **51** (2), 183–190.

Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. & Shochet, O. 1995 Novel type of phase transition in a system of self-driven particles. *Physical Review Letters* **75** (6), 1226–1229.

Weimerskirch, H., Martin, J., Clerquin, Y., Alexandre, P. & Jiraskova, S. 2001 Energy saving in flight formation. *Nature* **413**, 697–698.

Wittenberger, J. F. & Hunt, Jr, G. L. 1985 *Avian Biology*, , vol. 8. Academic Press,.