The Stochastic Modelling of Social and Territorial Behaviour

by Paul Gavin Blackwell

A Thesis Submitted
to the University of Nottingham
for the Degree of
Doctor of Philosophy

September 1990

ACKNOWLEDGEMENTS

I would like to thank Frank Ball, my supervisor, for all his advice and support, and Philip Bacon for the valuable help and motivation he has provided. I am grateful to Hans Kruuk and David Macdonald for their willingness to discuss the biological background to the models in this thesis. I would also like to thank Caitlin Buck, Jon Forster, Sue Hills, Cliff Litton, Nick Polson, Gareth Roberts, Adrian Smith, David Stephens and Jon Wakefield for their help and encouragement at various stages of this work, and Jacqui Pate for putting up with so much.

This work was supported by an SERC research studentship.

WORK REPORTED ELSEWHERE

Some of the work in Chapters 2, 3 and 6 of this thesis has been accepted for publication as follows:

Bacon, P. J., Ball, F. G., and Blackwell, P. G. (1990). Analysis of a Model of Group Territoriality based on the Resource Dispersion Hypothesis. *J. Theor. Biology*, to appear.

Bacon, P. J., Ball, F. G., and Blackwell, P. G. (1990). A Model for Territory and Group Formation in a Heterogeneous Habitat. *J. Theor. Biology*, to appear.

Blackwell, P. G. (1990). Deterministic and Stochastic Models of Social Behaviour Based on the Resource Dispersion Hypothesis. *IMA J. Math. Appl. Med. Biol.*, to appear.

To Jacqui and my family.

Table of Contents

ABSTRACT

This thesis considers mathematical models of the interaction between social and territorial behaviour in animals, mainly by probabilistic methods.

Chapter 1 introduces the Resource Dispersion Hypothesis, which suggests that territorial behaviour plus dispersed food resources can explain the existence of social groups, and describes an existing model of the process, due to Carr and Macdonald.

In Chapter 2 the model of Carr and Macdonald is analysed, and in Chapter 3 an improved model is suggested and its main properties derived, primarily using renewal theory.

Chapters 4 and 5 consider various spatial models for territory formation, and the effect of spatial factors on social behaviour, using analytic and simulation-based methods.

Chapter 6 considers the evolution of social behaviour using both discrete-time deterministic models and branching processes to investigate the viability of different strategies of social behaviour in the presence of dispersed resources.

Chapter 1 Introduction

1.1 Territorial Behaviour

A territory, for the purposes of this thesis, is an area which, along with the resources such as food that it contains, is monopolised by one or more members of a particular species, to the exclusion of other members of that species. Territorial behaviour, the setting up and holding of territories, is found in a wide range of different species, including insects, birds, fish and mammals (see for example Hölldobler and Lumsden (1980), Maynard Smith (1974), Noakes and McNicol (1982) and Macdonald (1983) respectively). Furthermore,

> "Territorial behaviour, in any animal which displays it, may well be the most important factor stabilising population numbers." (Maynard Smith, 1974).

As a consequence, territorial behaviour has been extensively studied, primarily by means of models of territory size: see for example Davies and Houston (1984), Maynard Smith (1974) and Schoener (1983).

1.2 Sociality in Territorial Species

The territories used by most territorial species are held by individual animals or mated pairs (Carr and Macdonald, 1986). In other species, territories may be held by larger groups. Usually there are clear advantages to such group living. Typical examples include co-operative hunting in wolves (Fox, 1970) and co-operative defence against predators in mongooses (Rood, 1975). However, in some species, such advantages are absent. Examples include foxes (Macdonald, 1981) and badgers (Kruuk, 1978a,b). In these species, an alternative explanation is needed for the existence of social groups.

1.3 The Resource Dispersion Hypothesis

It has been suggested for some territorial species which live in groups, but in which there are no obvious advantages from such grouping, that the existence of the groups, and more generally the sizes of the groups and the territories themselves, can be explained in terms of the distribution in space and time of resources, usually food, e.g. Bradbury and Vehrencamp (1976), Kruuk and Parish (1982), Macdonald (1981). In particular, in some of these cases, it has been suggested that the spatial dispersion of resources determines territory size, and the richness of the resources independently determines group size.

Horn (1968) put forward a related model for nesting colonies in birds, but his model is not related to territorial behaviour since it does not incorporate the monopolisation of resources, and it will not be considered further.

Macdonald (1983; see also Kruuk and Macdonald, 1985) has put forward a hypothesis which generalises the above ideas, and which is intended to explain the existence of group living, even in the absence of obvious benefits. This "Resource Dispersion Hypothesis" (R.D.H.) proposes that

> "groups may develop where resources are dispersed such that the smallest economically defensible territory for a pair ... can also sustain additional animals" (Carr and Macdonald, 1986).

Note that once territorial groups are formed in accordance with the R.D.H., it is relatively easy for a species to develop more sophisticated forms of social interaction and co-operation, which might not otherwise arise (Macdonald and Carr, 1989). Thus the R.D.H. may also be an explanation of the evolutionary origins of social behaviour even in species where other advantages of grouping are now present.

The R.D.H. may apply for many different reasons. For example, the smallest viable territory for a hyaena might be one which is just large enough for the hyaena to chase and catch its prey in. The amount of resources in such a minimum territory might be sufficient to support a group of hyaena (Kruuk and Macdonald, 1985). A more general case in which the R.D.H. might apply is when resource availability is spatially and temporally heterogeneous, and Macdonald (1983) goes on to suggest that in this case, the R.D.H. may lead to the independence between territory size and group size mentioned above.

A recent paper by Carr and Macdonald (1986) presents an explicit model for the process of territory and group formation, concentrating on the heterogeneous case. The paper includes examples to show that groups can be formed, given such heterogeneity, but does not determine the predictions of the model in general. In this thesis we will explore their model, develop related models which consider more general types of habitat and extend these models to cover spatial and evolutionary factors. The aim is to understand the applicability and consequences of the R.D.H. in the general case in which resources are spatially and temporally heterogeneous.

1.4 The Carr and Macdonald Model

The model presented by Carr and Macdonald (1986), with minor changes in notation, is as follows. Food availability is assumed to be heterogeneous in both space and time: specifically, food is assumed to occur in discrete patches, each of which may or may not be available at a given time. So, dividing time into discrete "feeding periods", the total amount of

food available in each period is a random variable, R. A territory is set up by a primary group, taken to be the "smallest, reproductively viable social unit"- typically a monogamous pair. These primary animals each have food requirements of R_α per period. For a given territory, the food security of the primaries in that territory, S_α, is defined to be the probability of achieving these requirements, i.e.

$$S_\alpha = \Pr(R \geq 2R_\alpha).$$

The primary pair are assumed to need food security at least equal to some critical level C_α; and they will choose the smallest territory that will provide this, because of the costs of territory defence.

A potential secondary animal similarly has food requirements R_β, and critical level of food security C_β. Hence, if the territory set up by the primary pair satisfies

$$S_\beta \geq C_\beta,$$

where

$$S_\beta = \Pr(R \geq 2R_\alpha + R_\beta), \tag{1.1}$$

then a secondary animal can also obtain its food requirement from the territory, even if it accepts subordinate status and never competes with the primary pair for food. So in these circumstances, a secondary animal can share the territory at no cost to the primaries.

Let k be the amount of resources in a patch when it is available, which we call its richness. Let X be a random variable with a Bernoulli(p) distribution, i.e. X takes the value 1 with probability p, and the value 0 with probability $1-p$. Then the yield from a patch with richness k and probability of availability p is the random variable kX. If a territory consists of n patches, independent and identical, then the yield will be $R = kY$, where Y has a Binomial(n,p) distribution. Thus the primaries will set up a territory consisting of n patches, where n is the smallest number such that

$$R/k \sim \text{Binomial}(n,p) \Rightarrow \Pr(R \geq 2R_\alpha) \geq S_\alpha.$$

For example, consider the case $R_\alpha = R_\beta = k = 1$, for simplicity, and $C_\alpha = 0.95, C_\beta = 0.9$, as suggested by Carr and Macdonald (1986). If $p = 0.8$, then $n = 4$ is the smallest territory which will satisfy the primary pair, giving $S_\alpha = \Pr(R \geq 2) = 0.97$. In this case, $S_\beta = \Pr(R \geq 3) = 0.82$, so no secondary animal can be accommodated. Similarly, if $p = 0.9$, then $n = 3$ is the smallest satisfactory territory, giving $S_\alpha = 0.97, S_\beta = 0.73$, so again there is no secondary. However, if $p = 0.86$, the primaries still require $n = 4$, but now $S_\beta = 0.903$, which means the territory will also support a secondary animal.

1.5 Outline of the Thesis

In Chapter 2, we will analyse the Carr and Macdonald model in some depth, and in Chapter 3 we consider a modified version. In subsequent chapters, we consider extensions of our modified version of the model to cover spatial location of resources (Chapters 4 and 5) and the evolution of group behaviour in the presence of resource heterogeneity (Chapter 6).

Chapter 2 Analysing the Carr and Macdonald Model

The examples given by Carr and Macdonald (1986) show that their model proves that in some circumstances, spatial discreteness and temporal variability of resources can lead to the formation of groups occupying territories, without assuming any social interaction. However, the overall behaviour and predictions of the model are not explored in that paper. In this chapter, we consider their model in more detail, to determine how widely it is applicable, in what cases it predicts the formation of groups, and the relationships it implies between group size, territory size, and resource richness.

2.1 Behaviour of the Basic Model.

In this section we examine the behaviour of the main model described by Carr and Macdonald (1986), in which the total yield R from a territory in a given period has a Binomial distribution, scaled by some constant richness k, as described in Section 1.4. In Section 2.1.1 we consider the qualitative behaviour of the model, and in Section 2.1.2 we give some numerical results, expressed in tabular and graphical form.

2.1.1 Analytic Results

The simplest case is that of "very rich" patches, by which we mean that any one patch, if available, provides more than enough resources for the primary pair and one or more secondaries, i.e.

$$k \geq 2R_\alpha + R_\beta.$$

In this case, since R can only take the values $0, k, 2k, \ldots$ we have

$$\begin{aligned} &\Pr(R \geq 2R_\alpha) \\ &= \Pr(R \geq 2R_\alpha + R_\beta) \\ &= \Pr(R \neq 0) \\ &= 1-(1-p)^n. \end{aligned} \tag{2.1}$$

If we also assume $C_\beta \leq C_\alpha$, i.e. that the food security required by a secondary is no greater than that required by the primary pair, then

$$\Pr(R \geq 2R_\alpha) \geq C_\alpha$$

implies that

$$\Pr(R \geq 2R_\alpha + R_\beta) \geq C_\beta.$$

So in the case of very rich patches, provided $C_\beta \leq C_\alpha$, any territory set up by a primary pair could support a secondary animal, regardless of the value of p.

In fact, more than one secondary may sometimes be supported. The argument of the previous paragraph is unchanged for any number of secondaries s satisfying

$$k \geq 2R_\alpha + sR_\beta.$$

So by just considering the probability of at least one patch being available, we see that the number of secondaries which can be supported is at least

$$s = \left\lfloor \frac{k - 2R_\alpha}{R_\beta} \right\rfloor, \tag{2.2}$$

where $\lfloor x \rfloor$ is the largest integer not greater than x. For certain parameter values, the probability of at least two patches being available is also at least C_β, which will result in additional secondaries being supported. This is discussed in Section 2.1.2. In most cases, however, the predicted number of secondaries is given by equation (2.2). For example, if $k = 5.5$, and $R_\alpha = R_\beta = 1$, then provided $C_\beta \leq C_\alpha$, any territory set up by a primary pair could support three secondary animals.

On the other hand, from equation (2.1), the territory size, represented by the number of patches n, does not depend on the richness of the patches (as long as they remain "very rich"), but is determined by C_α and p, and is given by

$$n = \left\lceil \frac{\ln(1 - C_\alpha)}{\ln(1 - p)} \right\rceil,$$

where $\lceil x \rceil$ is the smallest integer not less than x.

So in the case of very rich patches, if we compare different habitats, represented by different values of k and p, the model predicts that territory size is a function of probability of availability p, whereas group size is a function of richness k, and is typically independent of p, and so territory size and group size may appear independent.

Groups may also occur with lower values of k, i.e. where patches are not "very rich". This case is not so straightforward to analyse, since the various parameters can interact. In general, R_α, C_α, k and p all affect territory size, and all the parameters affect group size.

We need the following notation :

$$r = \left\lceil \frac{2R_\alpha}{k} \right\rceil,$$

the smallest integer not less than $2R_\alpha/k$, i.e. the requirements of the primary pair in patches; and

$$r' = \left\lceil \frac{2R_\alpha + R_\beta}{k} \right\rceil,$$

the requirements of the primary pair plus one secondary animal, in patches. Note that since the patch richness k is explicitly included in this formulation, we can take $R_\alpha = 1$, without loss of generality.

In this general case, n will be chosen by the primaries to be the smallest integer such that

$$\Pr(Y \geq r) \geq C_\alpha,$$

where Y has a Binomial(n,p) distribution, and so n depends on k through r. The cases where a group can be formed, and more generally the size of such a group, are more difficult to determine. We need to look at the effect of the parameters R_β, C_α, k and p on the food security S_β of potential secondary animals.

For reference, the meanings of the various symbols we have defined are collected in Table 2.1.

2.1.1.1 Changing the amount of food required by a secondary, R_β

Since equation (1.1) can be rewritten in the above notation as

$$S_\beta = \Pr(Y \geq r'),$$

it can be seen that R_β only affects S_β through r', and so changes in R_β which do not change r' have no effect. Changes in R_β which do change r' will cause a discontinuous change in S_β. This will occur whenever $(2R_\alpha + R_\beta)/k$ is an integer, that is when

$$R_\beta = jk - 2R_\alpha, \tag{2.3}$$

for some integer j. As R_β increases past each of these values, S_β will decrease: the exact values before and after depend on k and p. Given a value for C_β , and fixing k and p, this means there will be some value R^* such that

$$R_\beta \leq R^* \Rightarrow S_\beta \geq C_\beta,$$

$$R_\beta > R^* \Rightarrow S_\beta < C_\beta;$$

Table 2.1

Meanings of symbols

Symbol	Meaning
C_α	The level of food security required by the primary pair
C_β	The level of food security required by a secondary animal
k	The richness of a patch i.e. the amount of resources it gives when it is available
n	The number of patches in a territory
p	The probability of availability, in a given feeding period, of each patch
R	The total amount of food available in a given feeding period
R_α	The food requirements per period of a primary animal
R_β	The food requirements per period of a secondary animal
r	The food requirements of the primary pair expressed in patches
r'	The food requirements of the primary pair plus one secondary animal expressed in patches
S_α	The food security of the primary pair in a given territory
S_β	The food security of a potential secondary animal in a given territory
X	A random variable indicating the availability ($X = 1$) or non-availability ($X = 0$) of a particular patch in a given feeding period
Y	The number of patches available in a given feeding period

and R^* will be of the form given in equation (2.3).

2.1.1.2 Changing the required primary food security, C_α

C_α only affects S_β through the value of n, although the values at which changes in n occur are not as easily characterised as the values of R_β at which r' changes. When n increases, S_β will increase, so there will be some value of C_α, C^* say, below which secondary animals are absent, and above which they are present. Of course, it is possible that for given values of $R_\alpha, R_\beta, C_\beta, k$ and p, the value C^* might be unrealistically low, in which case secondary animals will never be present with those parameter values.

2.1.1.3 Changing the patch richness, k

Similarly, k only has an effect on S_β through r and r'. Increasing k only affects S_β when r changes, at

$$k = 2R_\alpha/j,$$

for some integer j, when S_β may decrease (if n increases), or when r' changes, at

$$k = (2R_\alpha + R_\beta)/j,$$

for some integer j, when S_β will increase. Note that these two sets of critical values are interspersed, so it is not generally possible to find a value k^* such that

$$k \leq k^* \Rightarrow S_\beta \geq C_\beta;$$

$$\text{and } k > k^* \Rightarrow S_\beta \leq C_\beta;$$

or any similar simple relationship. Some examples are given in Section 2.1.2.

2.1.1.4 Changing the probability of patch availability, p

The effect of varying p depends on whether this causes n to change or not. An increase in p which reduces n, because it means the primaries are satisfied with fewer patches, will obviously decrease S_β; whereas an increase in p which does not alter n will increase S_β. Thus, as for the parameter k, it is not possible to summarise the effect of p in any simple form such as "secondary animals are present if and only if $p \geq p^*$".

2.1.1.5 Summary of the effects of changes of parameter values

The complexity of these responses to changes in the parameters is largely due to discontinuities, which are a consequence of the discrete distribution used for the yield from a territory. The parameterisation used both here and in Carr and Macdonald (1986), although natural, also makes the effects of some parameters more complex. For example, changing p affects the expected yield per patch, the coefficient of variation of yield, and the shape of the distribution. If the change in p causes a change in n, the effect is even more complex.

These results show that this more general model does not conform to the simple relationships which describe the case of "very rich" patches. In particular, the hypothesis that territory size is independent of richness and group size is an increasing function of richness, independent of territory size, no longer holds. Territory size is now predicted to decrease with richness. Group size can be increasing with richness at some values of richness, and decreasing with richness at other values, even if all other parameters are held constant, i.e. group size can vary non-monotonically with richness. If habitats of differing richness are compared, almost any relationship between territory size and group size might be observed.

2.1.2 Numerical Results

The relationship between S_β, the secondary food security, and k, the patch richness, is complicated, as mentioned above. Figure 2.1 shows the graph of S_β against k in the case $R_\alpha = R_\beta = 1$, $C_\alpha = 0.95$, and $p = 0.65$. The critical value $S_\beta = C_\beta = 0.9$ is also shown for comparison. Since S_β depends on k only through the integers r and r', the graph has discontinuities at the points where r or r' changes, and is constant between these points.

Similarly, Figure 2.2 shows the graph of S_β against p, the probability of patch availability, in the case where $r = 4, r' = 5$. Discontinuities occur in Fig. 2.2 at those values of p that cause n, the number of patches in the territory, to change. Except at these points, S_β is increasing with p, since with a given number of patches, a higher probability of availability leads to higher food security.

As k or p varies, by comparing S_β with C_β we can determine when secondary animals will be present. This information, for all combinations of p and k, is summarised in Figure 2.3. The shaded regions in the figure indicate those values of p and k for which secondary animals are predicted. An important conclusion to be drawn from Fig. 2.3 is that apart from the case when $r' = r$, the parameters suggested in Carr and Macdonald (1986), $C_\alpha = 0.95$, $C_\beta = 0.9$, $R_\alpha = R_\beta = 1$, actually result in few cases where secondary animals are present.

These results are further summarised in Table 2.2, which shows those values of p for which secondaries are predicted, given different values of k.

Table 2.2

Values of p giving secondary animals

k	r	r'	p
$0.50 \leq k < 0.60$	4	6	None
$0.60 \leq k < 0.67$	4	5	Some $p \geq 0.4$
$0.67 \leq k < 0.75$	3	5	None
$0.75 \leq k < 1.00$	3	4	Some $p \geq 0.65$
$1.00 \leq k < 1.50$	2	3	Some $p \geq 0.75$
$1.50 \leq k < 2.00$	2	2	All
$2.00 \leq k < 3.00$	1	2	Some $p \geq 0.8$
$k \geq 3.00$	1	1	All

"Some p" in the table means that secondaries are predicted for values of p in part of the region indicated, possibly only a small part, and not necessarily an interval, as can be seen

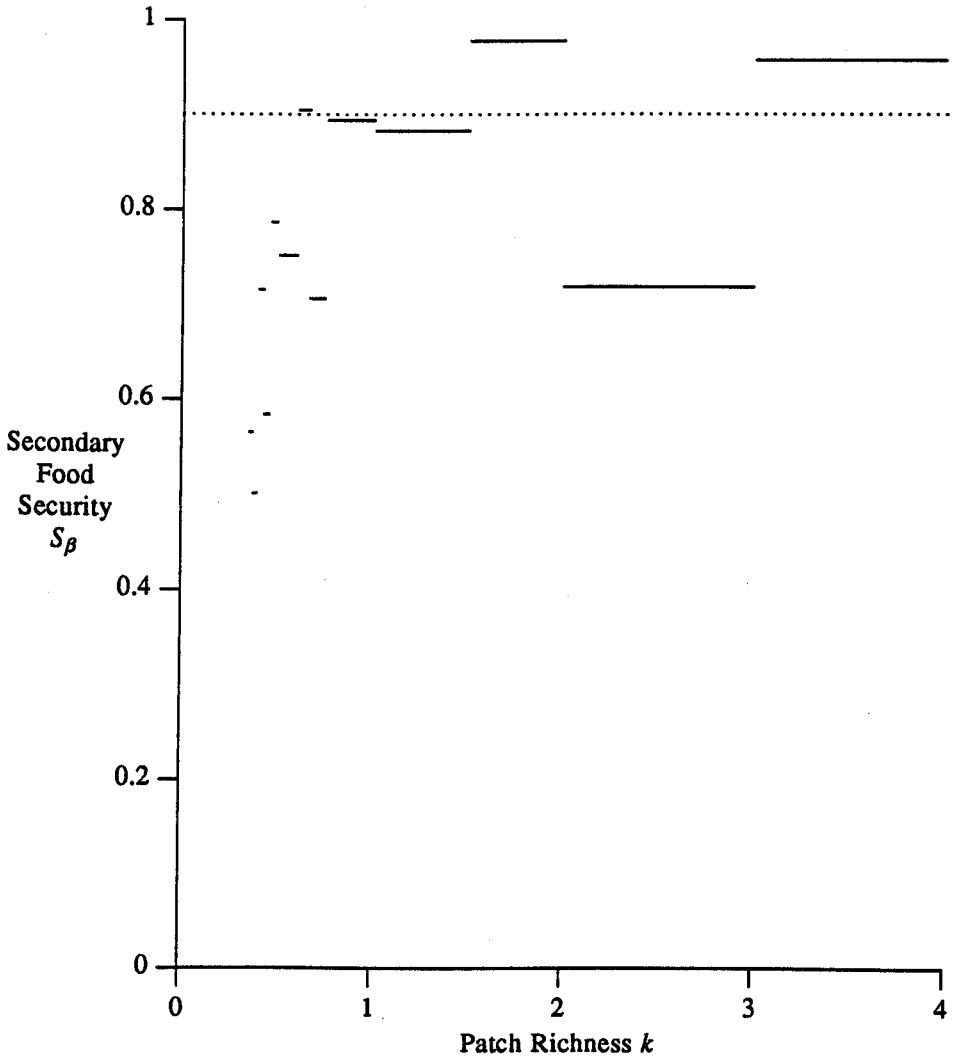

Figure 2.1

The graph of secondary food security S_β against patch richness k when $p = 0.65$, shown for the interval $0.4 < k \leqslant 4.0$, (solid); and the line corresponding to the critical value $S_\beta = C_\beta = 0.90$, (dotted). Secondary animals will be present when $S_\beta \geqslant C_\beta$. Values corresponding to $0.0 \leqslant k \leqslant 0.4$ are omitted for clarity. For all $k \geqslant 3.0$, S_β is constant.

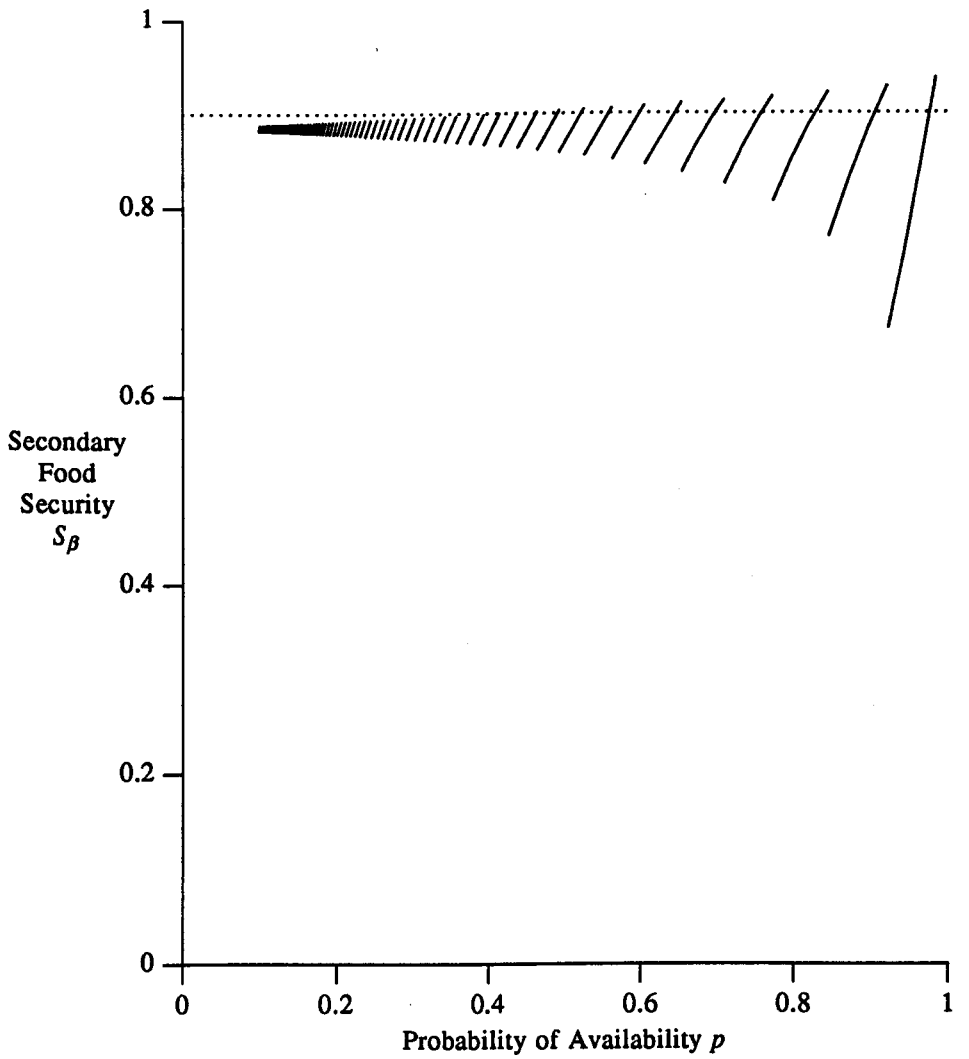

Figure 2.2

The graph of secondary food security S_β against probability of availability p when $r = 4$, $r' = 5$, shown for the interval $0.08 \leqslant p < 0.99$, (solid); and the line corresponding to the critical value $S_\beta = C_\beta = 0.90$, (dotted). Secondary animals will be present when $S_\beta \geqslant C_\beta$. For $p < 0.08$, the trend continues, but values of S_β are omitted for clarity. For $0.99 \leqslant p \leqslant 1.00$, $S_\beta = 0.0$ (not shown).

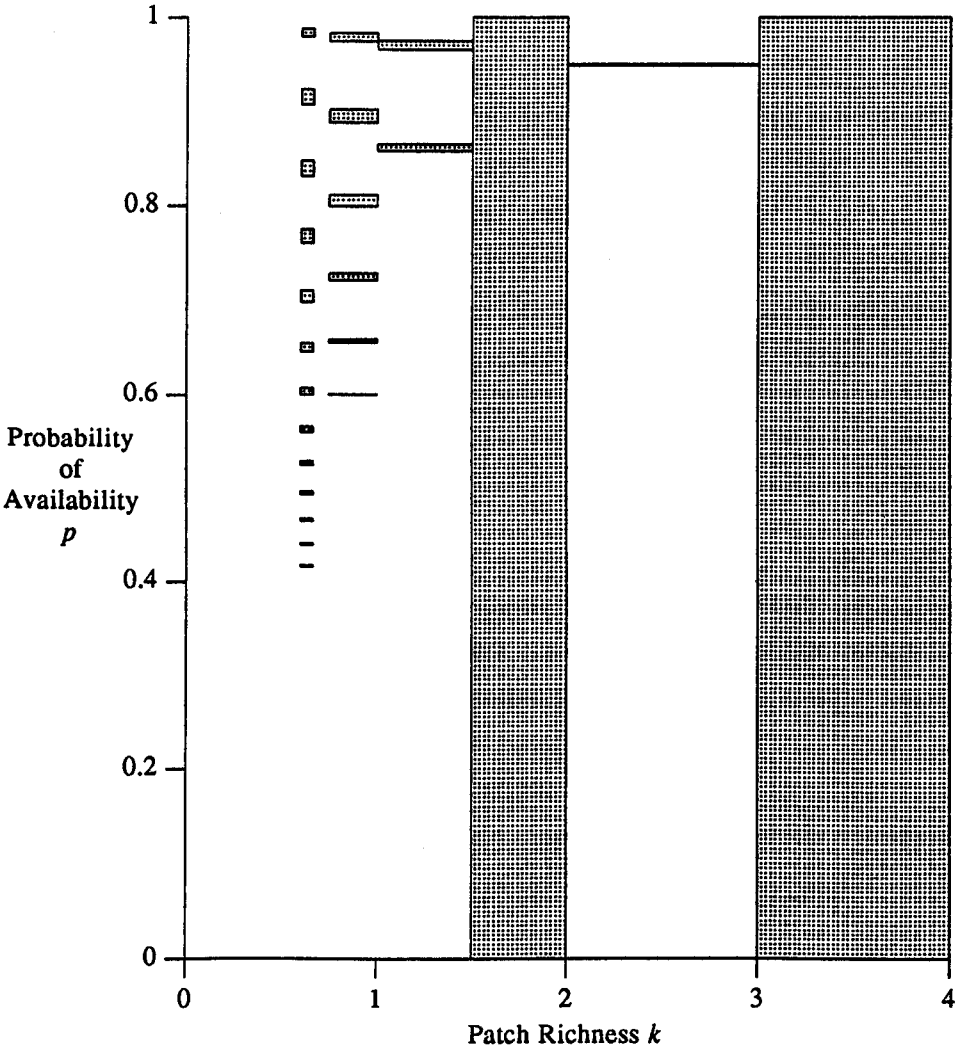

Figure 2.3

The combinations (shaded regions) of patch richness k, and probability of availability, p, for which $S_\beta \geq C_\beta = 0.9$, i.e. for which secondary animals are present. For $k \geq 3.0$, secondaries are always present.

from Fig. 2.3. For $k \leqslant 0.5$, r and r' vary, but always satisfy $r' - r \geqslant 2$, and no secondaries are predicted.

We are now in a position to reconsider briefly the case of "very rich" patches. In such a case, additional secondaries, above the number given in equation (2.2), can be supported if and only if the probability of at least two patches being available is at least C_β. But this condition is equivalent to a single secondary being supported in the more general model, when $r = 1, r' = 2$. From Table 2.2, for the given values of C_α and C_β, the latter event can only occur for certain values of $p \geqslant 0.8$. Hence, we can generally take the number of secondaries in the case of "very rich" patches to be given by equation (2.2).

2.2 Alternatives to the Binomial Distribution of Yield.

The Binomial distribution for the total yield from a territory follows from the assumption that a territory consists of a finite number of individual patches, with yields given by independent and identical Bernoulli random variables. A different distribution for total yield might be obtained either by relaxing some of these assumptions, or as an approximation to the Binomial case.

2.2.1 The Poisson Distribution

The Poisson distribution with rate parameter λ ($\lambda \geqslant 0$) is defined by:

$$\Pr(R = j) = \lambda^j e^{-\lambda}/j!, \quad j = 0, 1, 2, \ldots .$$

Its use to represent the distribution of total yield from a territory is discussed briefly in Carr and Macdonald (1986), and can arise in two possible ways.

Firstly, the Poisson may be used as an approximation to the Binomial with large n and small p. More precisely, for $l = 1, 2, 3, \ldots$ let Y_l have a Binomial distribution with parameters n_l and p_l, and let $\lim_{l \to \infty} n_l = \infty$, $\lim_{l \to \infty} p_l = 0$ and $\lim_{l \to \infty} n_l p_l = \lambda > 0$. Then $Y_l \xrightarrow{D} Y$ as $l \to \infty$, where "$\xrightarrow{D}$" denotes convergence in distribution, and Y is a Poisson random variable with rate λ. Hence the Binomial distribution with parameters n and p, where n is large and p is small can be approximated by the Poisson distribution with rate $\lambda = np$.

If $R = kY$ represents the yield from a territory in the Carr and Macdonald model, the rate λ will be determined by the primaries' choice of territory size, i.e. λ will take the smallest value such that $S_\alpha \geqslant C_\alpha$. Since the Poisson approximation is only appropriate in cases where p is small, it follows from the results in Table 2.2 that, if $C_\alpha = 0.95$, $C_\beta = 0.9$, $R_\alpha = R_\beta = 1$, a secondary will be present if and only if $r = r'$.

Secondly, the Poisson distribution may occur in its own right in the case where equal-sized patches of food appear at random in space and time. The simplest such model assumes

that in each feeding period, patches occur as a two-dimensional homogeneous Poisson point process (see e.g. Cox and Isham, 1980), i.e. completely at random in space, and that all feeding periods are independent. In a territory of fixed area, it then follows that the numbers of patches available in successive feeding periods are independent Poisson random variables, with rate $\lambda = \eta A$, where η is the intensity of the point process, and A is the area of the territory. The calculation of the value of λ chosen by the primaries will be the same as in the case where the Poisson is used as an approximation to the Binomial, so again the presence of secondary animals is not predicted for parameter values close to those suggested by Carr and Macdonald.

Note that Figure 6 in Carr and Macdonald (1986) is incorrect. The two curves should be smooth with both S_α and S_β approaching 0 as $\bar{R} \to 0$ and approaching 1 as $\bar{R} \to \infty$.

2.2.2 The Normal Distribution.

The Normal distribution can be used as an approximation to the Binomial, in the case when n is large and p is in the open interval $(0,1)$. The precise result is DeMoivre's theorem (see Seneta (1982) for a discussion of its history): if for $l = 1,2,3,\ldots$, Y_l has a Binomial distribution with parameters l and p ($p \in (0,1)$), then

$$\frac{Y_l - lp}{\sqrt{lp(1-p)}} \xrightarrow{D} Y, \text{ as } l \to \infty, \tag{2.4}$$

where Y has a standard Normal distribution, i.e. $Y \sim N(0,1)$. Equation (2.4) is simply a special case of the Central Limit Theorem, since a random variable with the Binomial(n,p) distribution can be thought of as the sum of n independent and identically distributed Bernoulli(p) random variables. So, if n is large, and p is in $(0,1)$, equation (2.4) implies that a Binomial(n,p) distribution can be approximated by the Normal distribution with mean np and variance $np(1-p)$. Hence if R represents the yield from a territory, with $R = kY, Y \sim$ Binomial(n,p) we can approximate the distribution of R by a Normal distribution with mean μ and variance σ^2, where

$$\mu = knp, \quad \sigma^2 = k^2 np(1-p); \tag{2.5}$$

A similar argument shows that the Normal distribution can also be used to approximate a Poisson distribution with large rate λ multiplied by some richness k, with the Normal distribution having mean and variance given by

$$\mu = k\lambda, \quad \sigma^2 = k^2\lambda. \tag{2.6}$$

Equations (2.5) and (2.6) restrict the range of Normal distributions which represent valid approximations to Binomial or Poisson distributions. Although Carr and Macdonald (p.1541) use Normal distributions to approximate Binomial or Poisson distributions of yield, they do

not take into account these restrictions. For example, consider Figure 2 of Carr and Macdonald (1986). We have $R_\alpha = R_\beta = 1$, $C_\alpha = 0.95$, $C_\beta = 0.9$, $\mu = 6.9$, $\sigma = 3$, and n, p, k unspecified. If this is a Normal approximation to a Binomial distribution scaled by some richness k, we have

$$\begin{aligned} k &= (k^2)np(1-p)/knp(1-p) \\ &= \sigma^2/\mu(1-p) \\ &\geqslant \sigma^2/\mu \\ &= 1.3. \end{aligned}$$

But $k \geqslant 1.3$ means that the richness of individual patches is at least 1.3, so the Normal curve shown will give inaccurate values for $\Pr(R \geqslant 2)$ and $\Pr(R \geqslant 3)$. Hence the curve shown cannot represent a useful Normal approximation to **any** scaled Binomial distribution.

Another way of looking at this restriction is that, for a good approximation, r and r' must be reasonably large. From the numerical results in Section 2.1.2, this means that the cases where the Normal approximation is valid do not result in secondaries being present, at least when the parameters have values close to those suggested by Carr and Macdonald (1986).

2.2.3 Interdependent Patches.

The distributions for total yield discussed so far are based on independent identically distributed yields from distinct patches. The case where patches are not identical would be extremely complicated within the current model, and is not pursued here, although differences between patches are considered in the models of Chapter 3, where the mechanism for variation over time is different. However, the assumption of independence between patches can be relaxed, and we explore some alternatives in this section.

Interdependence between patch yields seems likely in reality, and has been documented in some specific cases. For example, the patches of earthworms on which badgers feed were found by Kruuk (1978a) to have interdependent yields. Carr and Macdonald give other references, and suggest possible reasons why either positive or negative correlation between patches might occur. However, their discussion of the effects of interdependence on their model is misleading, since they incorrectly assume that interdependence of patches is directly linked to the skewness of the overall distribution. In fact, there is no such connection. The Binomial distribution, which is based on independent patches, may be symmetric or skewed depending on the value of p. The shape of an interdependent distribution, conversely, need not be skewed, as will be seen in the examples below.

To discuss interdependence in more detail, we need to consider specific models. As mentioned at the beginning of this section, we wish at this stage to retain the idea of a simple unstructured environment, with indistinguishable food patches. This implies that the random variables $X_1,\ldots,X_n$ representing yields from individual patches not only have identical distributions, but are exchangeable. That is to say, for any $1 \leqslant i_1 < i_2 < \ldots < i_k$, the distribution of $(X_{i_1},\ldots X_{i_k})$ depends only on the value of k, and not on the choice of $i_1,\ldots,i_k$. In addition, since the Carr and Macdonald model involves choosing the value of n to satisfy a particular criterion, with no upper bound on n, it is necessary to write down a model for the environment from which the patches come which does not involve a specific value for n. The natural mathematical formulation of this condition is that it should be possible to embed $X_1,\ldots,X_n$ in an infinite sequence of exchangeable binary random variables, (Kingman, 1978; Ball and Donnelly 1988, Section 5). De Finetti's theorem (de Finetti, 1937) then states that

$$\Pr(X_1 = x_1,\ldots,X_n = x_n) = \int_0^1 p^y(1-p)^{n-y}\,dF(p) \quad \text{for } x_1,\ldots,x_n = 0,1, \tag{2.7}$$

where $y = \sum_{i=1}^{n} x_i$, and F is the distribution of some random variable P on $[0,1]$. Hence

$$\mathrm{Cov}(X_i,X_j) = \mathrm{Var}(P) \geqslant 0,\ i \neq j,$$

so the covariance (and correlation) between patch yields is non-negative whenever $X_1,\ldots,X_n$ can be embedded in an infinite exchangeable sequence.

This suggests that models involving indistinguishable patches cannot adequately represent negative correlation between patch yields. Such negative correlation can be incorporated into models where the environment has some explicit structure, e.g. where patch locations are modelled (see Besag, 1974, for a discussion of models of spatial interaction; his Section 4.2.1 considers models for collections of binary random variables) or where patches are labelled as being of different types, corresponding to resources likely to be available under different conditions. (Multi-type models are discussed briefly in Section 3.3.2.) For present purposes, however, we concentrate on models involving positive correlation between patches, satisfying equation (2.7). We are primarily interested in the distribution of $Y = \sum_{i=1}^{n} X_i$, and so we can rewrite equation (2.7) as

$$\Pr(Y = y) = \binom{n}{y}\int_0^1 p^y(1-p)^{n-y}\,dF(p).$$

Hence the distribution of Y can be written as a mixture of Binomial(n,p) distributions, keeping n constant, and mixing with respect to p. Thinking of Y as such a mixture corresponds closely to Carr and Macdonald's (1986) idea of "micro-climatic phenomena" (p. 1545), in that, in each period, P is sampled once, representing the current local micro-climate, and the patch yields are conditionally independent given that $P = p$.

We consider only the case where P has a Beta(a,b) distribution, i.e. where P has density

$$f(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} \quad a > 0, b > 0, 0 \leq p \leq 1.$$

The resulting mixture distributions for Y form the family of Pólya-Eggenberger or Binomial-Beta distributions (Johnson and Kotz, 1969, pp. 78-79, 159, 189, 229-232), given by

$$\Pr(Y = y) = \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y-1)} \cdot \frac{\Gamma(y+a)\Gamma(n-y+b)}{\Gamma(n+a+b)}, \quad y = 0, 1, \ldots, n.$$

This is a rich family, including bimodal, (discrete) uniform and unimodal distributions, and including Binomial distributions as a limiting case.

Figure 2.4 shows some examples of the Pólya-Eggenberger distribution, with $n = 10$, and with $a = b$ to illustrate the point made above that interdependent distributions may be symmetric rather than skewed. Note that the distribution in Fig. 2.4a, with $a = b = \infty$, is just the Binomial distribution with $n = 10, p = 0.5$. The effect of these Pólya-Eggenberger distributions on the Carr and Macdonald model, when they are interpreted as distributions for total yield, is shown in Table 2.3. The parameters held constant are $r = 1, r' = 2$, and $C_\alpha = 0.95$. The table shows the effect on the number of patches n required for a territory, and the secondary food security S_β, of different values of $a = b$.

Table 2.3

Territory Sizes and Values of Secondary Food Security with Interdependent Patch Yields

$a = b$	n	S_β
∞	5	0.812
5	6	0.832
1	19	0.900
0.5	125	0.924

The table shows a clear trend: as $a(=b)$ decreases, that is as the correlation between patch yields increases and the variability of total yield increases, both the size of a territory (expressed as a number of patches) and the food security of a potential secondary animal increase.

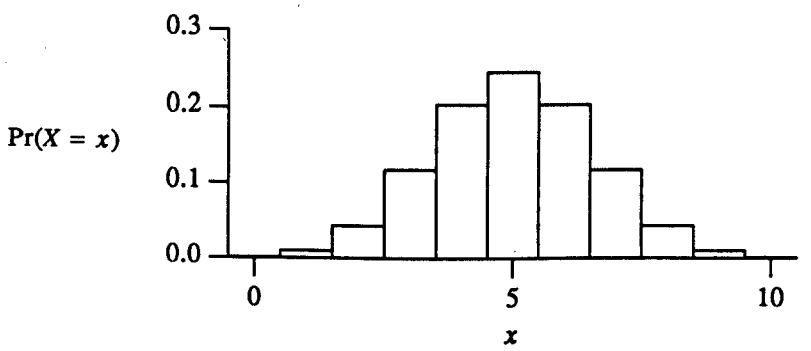

Figure 2.4a Beta-binomial Distribution with $a = b = \infty$

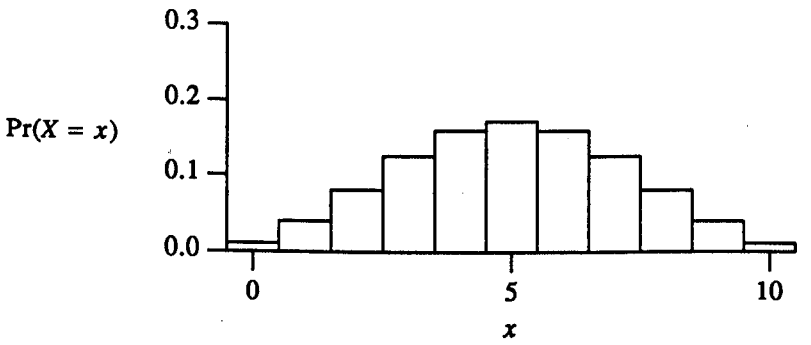

Figure 2.4b Beta-binomial Distribution with $a = b = 5$

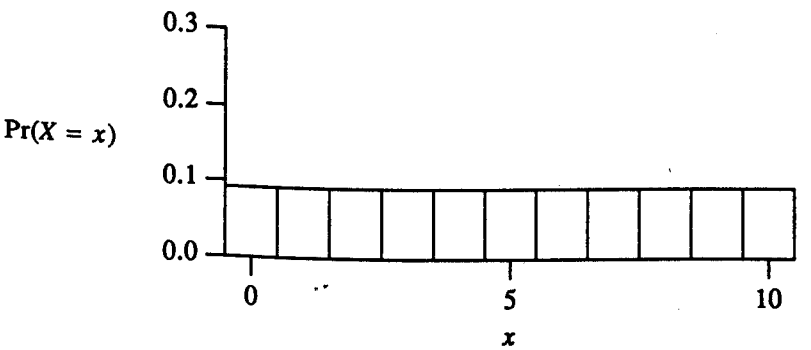

Figure 2.4c Beta-binomial Distribution with $a = b = 1$

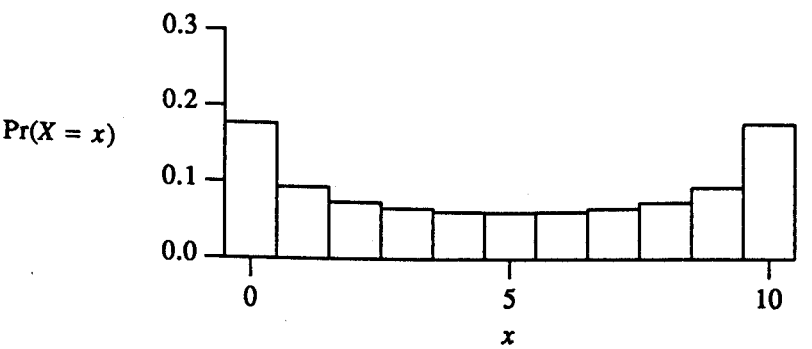

Figure 2.4d Beta-binomial Distribution with $a = b = 0.5$

2.3 Discussion

The model presented by Carr and Macdonald (1986) shows that in some circumstances, the minimum territory needed by a pair of animals may, because of heterogeneity of resources, support one or more extra animals. However, the analysis in the current chapter shows that except in the case of "very rich" patches, the range of circumstances in which such groups will be formed is quite narrow. Another prediction of the model is that independence between territory size and group size would only be expected to apply in the case of "very rich" patches. Finally, the behaviour of the food security of secondaries as the parameters of the model vary is generally complicated and non-monotonic.

However, the predictions from the model appear to be very dependent on the underlying assumptions about the distribution of yield, in particular the assumption of identical patches, each with yield following a scaled Bernoulli distribution. These assumptions also make the model rather intractable, except in the simplest cases.

The Carr and Macdonald model therefore shows explicitly that the R.D.H. can explain the existence of groups, and, particularly in the case of "very rich" patches, indicates the relationships that might be expected to exist between territory size, group size, and richness. For a more general analysis of the problem, however, it seems that a more flexible, tractable model is needed. One possible way to achieve this would be to model the yield from an individual patch as a continuous random variable, rather than as just available or not available. Such a model might be more realistic in many cases, and would give a continuous distribution for the total yield from a territory, which should be more amenable to analysis than a discrete distribution. A continuous distribution of total yield might also make it possible to consider variations in richness between patches, which again might be more appropriate but which, with a Bernoulli distribution as the starting point, would lead to an even more complicated overall distribution. In Chapter 3, we consider a model with a continuous distribution of yield.

Chapter 3. Simple Models based on a Continuous Distribution of Yield

The model described in Chapter 2 assumes that the yield from a territory in a given feeding period is a discrete random variable, arising as the sum of a number of Bernoulli random variables (independent or otherwise), multiplied by some scale factor representing the richness of all patches in the habitat. While such a model may be realistic for some habitats, there will clearly be cases where the distribution of yield will have a more general form. In addition, the results of Section 2.1 show that models based on a discrete distribution are often intractable, and that some of the results which can be obtained may be artefacts of the precise assumptions leading to the use of, for example, a Binomial distribution. We attempt to solve some of these problems in the current chapter by considering a continuous distribution for yield. In Section 3.1 we look at criteria for a satisfactory territory in a habitat giving a continuous distribution of yield, in Section 3.2 we look at a specific simple model based on those criteria, and in Section 3.3 we consider the effect of relaxing two of the simplifying assumptions: that the variance of total yield is proportional to mean total yield, and that only a single resource is important.

3.1 Criteria based on a continuous distribution of yield

3.1.1 A Single Time Period

We start by considering the criterion which determines whether a territory is large enough to support a primary pair of animals over a single feeding period. Let X be a continuous random variable denoting the yield from a territory, let a be a constant representing the required level of resources, and let c be a constant representing the required "food security" of the primary animals. Then, following closely the approach of Carr and Macdonald (1986), as described in Chapter 1, we say that a territory is satisfactory for the primary group if and only if

$$\Pr(X \geq a) \geq c. \tag{3.1.1}$$

Furthermore, let us assume that X has a Normal distribution with mean μ and variance σ^2 (this assumption is further discussed in Section 3.2). Then we have

$$\Pr(X \geq a) = \Pr\left(\frac{X-\mu}{\sigma} \geq \frac{a-\mu}{\sigma}\right).$$

But by definition,

$$(X-\mu)/\sigma \sim N(0,1)$$

and hence

$$\Pr(\frac{X-\mu}{\sigma} \geq \frac{a-\mu}{\sigma}) = 1-\Phi(\frac{a-\mu}{\sigma}) = \Phi(\frac{\mu-a}{\sigma})$$

where

$$\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} e^{-u^2/2}\, du$$

is the distribution function of the standard Normal distribution. Hence a territory is satisfactory for the primaries if and only if

$$\Phi(\frac{\mu-a}{\sigma}) \geq c$$

$$\Leftrightarrow \frac{\mu-a}{\sigma} \geq w, \qquad (3.1.2)$$

where we define

$$w = \Phi^{-1}(c).$$

Note that w is an increasing function of c, and is therefore another measure of required food security. Similarly, define a' to be the total amount of resources required by a minimum-sized group (typically a primary pair plus one secondary animal) and c' to be the food security required by a potential secondary animal. Then a territory with yield X will be able to support a secondary animal if and only if

$$\Pr(X \geq a') \geq c'$$

$$\Leftrightarrow \frac{\mu-a'}{\sigma} \geq w', \qquad (3.1.3)$$

by the same argument as before, where

$$w' = \Phi^{-1}(c').$$

Throughout the model, we ignore cases in which a secondary could only attain its requirements by competing with the primaries, i.e. we assume that the primaries can and will expel any secondary if it is in their interests to do so. We would generally expect $c' \leq c$, since the requirements of a secondary animal would be expected to be no more stringent than those of a primary animal, especially if the alternatives available to the secondary involve a high risk. In addition, if we write

$$\gamma = a'/a$$

then we have $\gamma \geq 1$. In addition, if the primary group is a breeding pair, then since the requirement of the secondary is not likely to be greater than that of an individual primary, we expect $\gamma \leq 1.5$. Other values of γ can be interpreted similarly.

3.1.2 Long-term Territories

The model described in Section 3.1.1 considers only a single feeding period. The parameter c is effectively the minimum required probability that the territory yields sufficient resources during that period, and its value is chosen by the primary pair to optimise the level of risk to be taken in that period.

In general, however, a territory will be intended to support animals for a number of feeding periods. Let $X_t, t = 1, \ldots, T$ be a sequence of continuous random variables representing the yields from a given territory over the sequence of feeding periods of interest. Let $\boldsymbol{X}$ denote the vector of length T formed from $X_1, \ldots, X_T$, and for any two vectors $\boldsymbol{u}, \boldsymbol{v}$ of equal dimension, write $\boldsymbol{u} \leq \boldsymbol{v}$ to mean $u_i \leq v_i$ for all i. Then the territory is satisfactory during each of T periods if and only if

$$\Pr(\boldsymbol{X} \geq a\mathbf{1}) \geq c^*, \tag{3.1.4}$$

where $\mathbf{1}$ is a vector of 1's, and c^* is the critical value of the probability of having sufficient resources in **all** T periods. Assuming $X_1, X_2, \ldots, X_T$ to be independent and identically distributed, we have

$$\begin{aligned}\Pr(\boldsymbol{X} \geq a\mathbf{1}) &= \Pr(X_t \geq a, t = 1, 2, \ldots, T) \\ &= \prod_{t=1}^{T} \Pr(X_t \geq a) \\ &= \Pr(X_1 \geq a)^T.\end{aligned}$$

Hence the criterion in equation (3.1.4) for a satisfactory territory becomes

$$\Pr(X_1 \geq a) \geq (c^*)^{1/T}, \tag{3.1.5}$$

which is clearly of the same form as (3.1.1), with

$$c = (c^*)^{1/T}.$$

Hence if a number of independent periods are considered separately, the criterion given in equation (3.1.2) is still appropriate.

In reality, however, a low yield in a given period is less likely to be unsatisfactory to the territory holders than a number of consecutive periods with low yields. To incorporate this

idea into the model, in a simple way, we define a moving average of the yields in individual periods by

$$Y_t = \frac{1}{l}\sum_{j=1}^{l} X_{t+j}, \tag{3.1.6}$$

where l is the number of periods over which we wish to average. If Y is the vector of Y_t's, of length $T+1-l$, then the criterion for a satisfactory territory becomes

$$\Pr(Y \geq b1) \geq c^*, \tag{3.1.7}$$

for some constant b, and some probability c^*. If $X_1, X_2, \ldots, X_T$ are independent and identically distributed, each with the Normal distribution $N(\mu,\sigma^2)$, then Y will have a multivariate Normal distribution, which can be written as $N(\mu 1, \frac{\sigma^2}{l}\Lambda)$, where Λ is the $(T+1-l)\times(T+1-l)$ correlation matrix given by

$$\Lambda_{ij} = \begin{cases} \frac{l-|i-j|}{l} & \text{if } |i-j| < l; \\ 0 & \text{if } |i-j| \geq l. \end{cases} \tag{3.1.8}$$

Furthermore, if we write

$$Z = \frac{1}{\sigma}(Y-\mu 1), \tag{3.1.9}$$

then Z also has a multivariate Normal distribution,

$$Z \sim N(0, \frac{1}{l}\Lambda), \tag{3.1.10}$$

where 0 is the vector of zeroes, of length $T+1-l$. The parameters of the distribution of Z depend only on l and T, and not on μ and σ^2. Hence we have

$$\begin{aligned} &\Pr(Y \geq b1) \geq c^* \\ \therefore \quad &\Leftrightarrow \Pr(Z \geq \frac{1}{\sigma}(b-\mu)1) \geq c^* \\ &\Leftrightarrow \frac{1}{\sigma}(b-\mu) \leq -v \\ &\Leftrightarrow \frac{\mu-b}{\sigma} \geq v \end{aligned} \tag{3.1.11}$$

where v is chosen such that

$$\Pr(Z \geq -v1) = c^*. \tag{3.1.12}$$

The criterion for a satisfactory territory is therefore given by equation (3.1.11), which is clearly of the same form as (3.1.2), with $b = a, v = w$. So considering a moving average of yields simply affects the interpretation of the parameters v and b in our model. The analysis

of this section can be repeated to express the criterion for the presence of secondaries in a similar form, thus giving corresponding relationships for v' and b'.

Note that even if the individual X's are not Normally distributed, the distribution of the Y's will be close to Normal, for moderate values of l, by the Central Limit Theorem, provided the distribution of the X's is unimodal and not too skewed. So this moving average formulation provides some justification for the initial assumption of Normality.

To use the criterion (3.1.11), we need to know the value of v. Unfortunately, the calculation of v involves the cumulative distribution function of a $(T+1-l)$-dimensional Normal distribution, and is analytically intractable. The following three sections describe ways in which a suitable value of v can be estimated, given the number of feeding periods of interest, T, the number of periods over which resources are averaged, l, and the required long term food security c^*.

Note that, from the meaning of T and c^*, it can be seen that, provided T is large compared with l, v depends on c^* and T only through $(c^*)^{1/T}$, so different values of T and c^* can give rise to the same value of v. The values of T and c^* used below were chosen for ease of interpretation. We assume each feeding period is 1 day, and take $T = 365$, so that c^* represents the probability of an adequate food supply for an entire year, and we take $l = 6$, which means, roughly, that 6 consecutive days of low yield would be fatal. We take $c^* = 0.99$ for primary animals, and $c^* = 0.97$ for secondary animals.

3.1.3 Naive Simulation

One way to estimate v is simply to simulate the multivariate Normal process $Z_0,\ldots,Z_{T-l}$. We have

$$\begin{aligned}
Z_t &= \frac{1}{\sigma}(Y_t-\mu) \\
&= \frac{1}{\sigma}\left(\left(\frac{1}{l}\sum_{j=1}^{l} X_{t+j}\right)-\mu\right) \\
&= \frac{1}{\sigma}\left(\frac{1}{l}\left(\sum_{j=1}^{l} X_{t+j}-l\mu\right)\right) \\
&= \frac{1}{\sigma}\left(\frac{1}{l}\left(\sum_{j=1}^{l} \{X_{t+j}-\mu\}\right)\right) \\
&= \frac{1}{l}\sum_{j=1}^{l}\left\{\frac{X_{t+j}-\mu}{\sigma}\right\} \\
&= \frac{1}{l}\sum_{j=1}^{l} W_{t+j}, \qquad (3.1.13)
\end{aligned}$$

where $W_1,\ldots,W_T$ are independent standard Normal random variables. Hence it is straightforward to simulate W and calculate the corresponding vector Z. For any given value v, we can

then estimate the corresponding value of c^* in equation (3.1.12) by the proportion of simulations in which

$$\min\{Z_t : t = 0,\ldots,T-l\} \geq -v. \tag{3.1.14}$$

If we write Z_t^k for the value of Z_t in the kth of n simulation runs, and

$$\zeta_k = \min\{Z_t^k : t = 0,\ldots,T-l\},$$

then the estimate of c^* is given by

$$\hat{c}^* = \frac{1}{n}\sum_{k=1}^{n} I_{\{\zeta_k \geq -v\}}, \tag{3.1.15}$$

where I_A denotes the indicator random variable of the event A. Clearly, c^* can be estimated with arbitrary accuracy by this method, and so it is also possible to determine the value of v corresponding to a specific value of c^*. This simulation has been implemented in a short computer program, using routine G05DDC from the N.A.G. Pascal library (Numerical Algorithms Group, 1986) to generate pseudo-random Normal variates. Numerical results are shown in Table 3.1.

Table 3.1

Estimates and standard errors for c^* as a function of v, when $T = 365, l = 6$, based on 5000 simulation runs.

v	$\hat{c}^*$	S.E.
0.8	0.013	0.0016
0.9	0.075	0.0037
1.0	0.246	0.0061
1.1	0.469	0.0071
1.2	0.691	0.0065
1.3	0.845	0.0051
1.4	0.928	0.0037
1.5	0.974	0.0022
1.6	0.989	0.0015
1.7	0.996	0.0009

3.1.4 Conditional Simulation.

The technique described in the previous section for estimating c^* for a given v uses only a small amount of information from each simulation run. As can be seen from equation (3.1.15), the estimate of c^* depends only on the value of ζ from each run. It is, however, possible to use more information from each simulation, as follows. From each simulation run, instead of generating a single value ζ, we generate the triple

$$(\tau, \zeta', \eta) \tag{3.1.16}$$

where τ is the value of t such that Z_t is a minimum over that run, and ζ' and η are given by

$$\zeta' = \min\{Z_t : t = 0,\ldots,\tau-1,\tau+l,\ldots,T-l\}, \tag{3.1.17}$$

and

$$\eta = \frac{1}{l}\sum_{j=1}^{l-1} W_{\tau+j}. \tag{3.1.18}$$

Thus ζ' is the minimum of those Z_t's which do not involve $W_{\tau+l}$, which is the last component of $\zeta = Z_\tau$, and $l\eta$ is the sum of all the other components of ζ. Note that the actual value of ζ is not included in the information we use from each simulation; instead, given the event described by the triple (3.1.16), ζ has a conditional distribution which we can write down analytically, in terms of the distribution of $W_{\tau+l}$. The marginal distribution of $W_{\tau+l}$ is just the standard Normal, and $W_{\tau+l}$ is dependent on τ, ζ' and η only through the condition

$$\zeta \leq \zeta',$$

from the definition of ζ. So for a particular value v of interest we have

$$\begin{aligned}
&\Pr(\zeta \geq -v \,|\, \tau, \zeta', \eta) \\
&= \Pr(Z_\tau \geq -v \,|\, \zeta', \eta) \\
&= \Pr(Z_\tau \geq -v \,|\, Z_\tau \leq \zeta', \eta) \\
&= \Pr(W_{\tau+l} + l\eta \geq -lv \,|\, W_{\tau+l} + l\eta \leq l\zeta') \\
&= \Pr(W_{\tau+l} \geq -lv - l\eta \,|\, W_{\tau+l} \leq l\zeta' - l\eta) \\
&= \begin{cases} 0 & -v \geq \zeta' \\ 1 - \Phi(-lv - l\eta)/\Phi(l\zeta' - l\eta) & -v < \zeta'. \end{cases}
\end{aligned} \tag{3.1.19}$$

The true probability c^* could be written as

$$\iiint_{\tau,\zeta',\eta} \Pr(\zeta \geq -v \,|\, \tau, \zeta', \eta) f(\tau, \zeta', \eta)\, d\tau d\zeta' d\eta$$

if the density $f(\tau, \zeta', \eta)$ were known. Instead, we estimate c^* after n simulations by

$$\hat{c}^* = \frac{1}{n} \sum_{k=1}^{n} \Pr(\zeta_k \geq -v \,|\, \tau_k, \zeta'_k, \eta_k). \tag{3.1.20}$$

The above method has been implemented using N.A.G. Pascal library routines (Numerical Algorithms Group, 1986) to generate random variates and to calculate values of $\Phi(x)$. Table 3.2 below shows some numerical results for $\hat{c}^*$ as a function of v, and also the reduction in variance achieved by using the "conditional" method described in this section instead of the "naive" method of Section 3.1.3.

Table 3.2

Estimates and standard errors for c^* as a function of v,
and variance reduction compared with naive simulation (Table 3.1),
when $T = 365, l = 6$, based on conditional simulation with 5000 runs.

v	$\hat{c}^*$	S.E.	Variance Reduction
0.8	0.013	0.0010	2.4
0.9	0.079	0.0027	1.9
1.0	0.224	0.0045	1.8
1.1	0.475	0.0053	1.8
1.2	0.690	0.0048	2.0
1.3	0.842	0.0036	2.2
1.4	0.928	0.0025	2.2
1.5	0.971	0.0015	2.2
1.6	0.989	0.0008	3.2
1.7	0.996	0.0005	3.5

The extra work involved in calculating

$$\Pr(\zeta \geq -v \,|\, \tau, \zeta', \eta)$$

using equation (3.1.19) increases the time taken for simulation by a factor of approximately 1.1, so the net variance reduction obtained varies from $1.8/1.1 \approx 1.6$ to $3.5/1.1 \approx 3.2$. See Ripley (1987) for a more general discussion of conditional simulation and other variance reduction techniques.

The simulation technique described above generalises to apply to any series of the form

$$Z_t = \sum_{j=1}^{l} \theta_j W_{t+j}, \quad t = 0, 1, \ldots, T-l, \tag{3.1.21}$$

where $l \ll T$ and $W_1,\ldots,W_T$ are independent and identically distributed, with common distribution function $F(.)$. We choose an arbitrary $m, 1 \leq m \leq l$. Then let

$$\zeta = \min\{Z_t : t = 0,\ldots,T-l\},$$

let τ be such that

$$Z_\tau = \zeta,$$

and let

$$\zeta' = \min\{Z_t : t = 0,\ldots,\tau+m-l-1,\tau+m,\ldots,T-l\},$$

$$\eta_i = \sum_{j \neq i} \theta_j W_{\tau+m-i+j},\ i = 1,\ldots,l.$$

Then we wish to evaluate

$$\Pr(\zeta \leq u \,|\, \tau,\zeta',\eta) \tag{3.1.22}$$

Since we are only looking at the distribution of ζ, the conditioning in equation (3.1.22) reduces to

$$\zeta = \theta_m W_{\tau+m} + \eta_m \leq \zeta'$$

and

$$\zeta = \theta_m W_{\tau+m} + \eta_m \leq Z_{\tau+m-i} = \theta_i W_{\tau+m} + \eta_i,\ i \neq m.$$

Rearranging each condition, we get

$$W_{\tau+m} \leq \frac{\zeta' - \eta_m}{\theta_m},$$

$$W_{\tau+m} \leq -\left(\frac{\eta_m - \eta_i}{\theta_m - \theta_i}\right) \quad \text{for } i : \theta_i < \theta_m,$$

$$W_{\tau+m} \geq -\left(\frac{\eta_m - \eta_i}{\theta_m - \theta_i}\right) \quad \text{for } i : \theta_i > \theta_m,$$

plus a trivial condition for any $i : \theta_i = \theta_m$. Note that in the special case considered above, $\theta_i \equiv \frac{1}{l}$, so all conditions except the one involving ζ' were trivial.

Hence if we define

$$\alpha = \max_i \left\{ -\left(\frac{\eta_m - \eta_i}{\theta_m - \theta_i}\right) : \theta_i > \theta_m \right\},$$

$$\gamma = \min_i \left\{ -\left(\frac{\eta_m - \eta_i}{\theta_m - \theta_i}\right) : \theta_i < \theta_m \right\},$$

$$\beta = \min\left\{\frac{\zeta' - \eta_m}{\theta_m}, \gamma\right\},$$

then since

$$\zeta \leqslant u \;\Leftrightarrow\; W_{\tau+m} \leqslant \frac{u-\eta_m}{\theta_m},$$

we have

$$\begin{aligned}
&\Pr(\zeta \leqslant u \,|\, \tau, \zeta', \eta) \\
&= \Pr(W_{\tau+m} \leqslant \frac{u-\eta_m}{\theta_m} \,|\, \alpha \leqslant W_{\tau+m} \leqslant \beta) \\
&= \begin{cases} 0 & \frac{u-\eta_m}{\theta_m} \leqslant \alpha \\ \dfrac{F(\frac{u-\eta_m}{\theta_m}) - F(\alpha)}{F(\beta) - F(\alpha)} & \alpha \leqslant \frac{u-\eta_m}{\theta_m} \leqslant \beta \\ 1 & \frac{u-\eta_m}{\theta_m} \geqslant \beta \end{cases}
\end{aligned} \tag{3.1.23}$$

Thus we can construct a simulation-based estimator, as in equation (3.1.20).

In addition to its use for variance reduction, it seems likely that the simulation technique described in this section would be of use in estimating the extreme lower tail of the distribution of ζ, when straightforward simulation would fail because of the extreme rarity of events in the tail.

3.1.5 Approximation

When T is large compared with l, we can use an analytic approximation for the distribution of ζ, instead of simulation. Using the results described in Leadbetter, Lindgren and Rootzén (1983), Chapter 4, we can approximate

$$\Pr(\mathbf{Z} \geqslant -v\mathbf{1})$$

by

$$\Pr(\hat{\mathbf{Z}} \geqslant -v\mathbf{1}),$$

where $\hat{Z}_0, \ldots, \hat{Z}_{T-l}$ have the same marginal distributions as $\hat{Z}_0, \ldots, \hat{Z}_{T-l}$ but are independent. To use this approximation in a particular case, we need to know its accuracy, and this is investigated by Rootzén (1983). Rootzén gives bounds on the error involved in approximating a stationary Normal sequence with an independent sequence: specifically, he shows that

$$R_n^-(u) \;\leqslant\; \Pr(M_n \leqslant u) - \Phi(u)^n \;\leqslant\; R_n^+(u). \tag{3.1.24}$$

where M_n is the maximum of a stationary Normal sequence of length n, with zero mean, unit variance, and covariance r_t at lag t, and R^-, R^+ are functions defined by Rootzén which depend on n, u and the covariances of the sequence. (Equation (3.1.24) is a corrected version of Rootzén's expression immediately preceeding his equation (4.1): note that the "ϕ" in Rootzén's expression should be a "Φ", as elsewhere in his paper.) The general forms of R^+ and R^- are

$$R_n^+(u) = n \sum_{t=1}^{k} r_t^+ f(r_t, u) + R,$$

$$R_n^-(u) = -n \sum_{t=1}^{k} r_t^- f(r_t, u) - R,$$

where

$$R = \frac{n}{2\pi(1-\delta_k^2)^{\frac{1}{2}}} \exp\left\{-\frac{u^2}{1+\delta_k}\right\} \sum_{t=k+1}^{n} |r_t|,$$

$$f(r, u) = \frac{1}{2\pi u^2 r}\left\{c'(r)e^{-u^2/(1+r)} - e^{-u^2}\right\},$$

$$r_t^+ = \max(0, r_t), \quad r_t^- = \max(0, -r_t),$$

$$\delta_k = \sup_{t>k} |r_t|,$$

$$c'(r) = \frac{(1+r)^{3/2}}{(1-r)^{1/2}},$$

and $k \geq 1$ is "a suitably chosen integer" (Rootzén, 1983).

We can make use of equation (3.1.24) as follows. Let

$$\xi_t = -Z_t\sqrt{l},$$

so that $\xi_0, \ldots, \xi_{T-l}$ have zero mean, unit variance, and correlation matrix Λ (equation (3.1.8)), and hence satisfy the conditions of Rootzén's result. Then we have

$$\begin{aligned} c^* &= \Pr(\mathbf{Z} \geq -v\mathbf{1}) \\ &= \Pr(\xi \leq v\sqrt{l}\mathbf{1}) \\ &= \Pr(M_{T+1-l} \leq v\sqrt{l}), \end{aligned}$$

where

$$M_{T+1-l} = \max\{\xi_t : t = 0, \ldots, T-l\}.$$

So from equation (3.1.24) we have

$$R_{T+1-l}^-(v\sqrt{l}) \leq c^* - \Phi(v\sqrt{l})^{T+1-l} \leq R_{T+1-l}^+(v\sqrt{l}),$$

where the covariance function used to calculate R^- and R^+ is that implied by equation (3.1.8) for A, i.e.

$$r_t = \begin{cases} \dfrac{l-t}{l} & \text{if } t < l; \\ 0 & \text{if } t \geq l. \end{cases}$$

In this case, by taking $k = l-1$, we can greatly simplify the expressions for R^- and R^+. We have

$$\sum_{t=k+1}^{n} |r_t| = 0,$$

so $R = 0$, and we have $r_t^- \equiv 0$, so $R^- = 0$. Also $r_t^+ \equiv r_t$, and so we have

$$\begin{aligned} R_n^+(u) &= n \sum_{t=1}^{k} r_t f(r_t, u) \\ &= n \sum_{t=1}^{l-1} \frac{1}{2\pi u^2} \left\{ c'(r_t) e^{-u^2/(1+r_t)} - e^{-u^2} \right\} \\ &= \frac{n}{2\pi u^2} \left[\sum_{t=1}^{l-1} \left\{ c'(r_t) e^{-u^2/(1+r_t)} \right\} - (l-1) e^{-u^2} \right], \end{aligned} \tag{3.1.25}$$

where we now have

$$c'(r_t) = \sqrt{\left(2 - \frac{t}{l}\right)\frac{l}{t}}.$$

A short computer program has been written to evaluate the bounds $\Phi(v\sqrt{l})^{T+1-l} + R^-_{T+1-l}(v\sqrt{l})$ and $\Phi(v\sqrt{l})^{T+1-l} + R^+_{T+1-l}(v\sqrt{l})$ for c^* in this special case. Some numerical values of the bounds are shown in Table 3.3. Note that for some values of v, the bounds contain very little information, but in the cases currently of interest, i.e. $c^* \approx 0.99$ and $c^* \approx 0.97$, they are informative.

3.1.6 Comparison of Simulation and Approximation Methods

The variance reduction calculations in Section 3.1.4 clearly show that conditional simulation is preferable to naive simulation for the current problem. The choice between conditional simulation and the numerical approximation in Section 3.1.5 depend on the particular parameter values of interest and the accuracy required. The numerical approximation should be used in cases where the bounds are sufficiently narrow, since it is quicker to calculate. Otherwise, simulation should be carried out, using the conditional method from Section 3.1.4.

In practice, we wish to estimate v using $T = 365$, $l = 6$, $c^* = 0.99$ and v' using $T = 365$, $l = 6$, $c^* = 0.97$. By interpolation, intervals for v corresponding to particular values of c^* can be obtained from the numerical approximation. It can be shown that the true values for v and v' lie in the intervals (1.58, 1.65) and (1.42, 1.54) respectively. These intervals are rather

Table 3.3.

Upper and lower bounds for c^* as a function of v.

v	Lower Bound	Upper Bound
0.8	0.000	1.000
0.9	0.006	1.000
1.0	0.073	1.000
1.1	0.276	1.000
1.2	0.548	1.000
1.3	0.767	0.977
1.4	0.895	0.967
1.5	0.957	0.981
1.6	0.984	0.991
1.7	0.994	0.996

wide, but can nevertheless ensure that we use plausible parameter values. Thus for these particular cases, the numerical approximation is preferable to simulation. The values used in the numerical examples in subsequent sections are $v = 1.6$ and $v' = 1.5$.

3.2 A Simple Model with a Continuous Distribution of Yield

3.2.1 Description of the Model

Having derived a criterion (3.1.11) for a satisfactory territory in terms of μ and σ^2, which we assume to completely characterise a territory, we must now specify the possible combinations of these parameters which are available to the primary pair. In a given real situation, one would expect some relationship, either deterministic or stochastic, between the mean and the variance for different choices of territory, thus restricting the primaries' choice. In addition, we need some "order of preference" between the different possibilities, so that the primaries' choice of a minimum territory is well-defined.

As a simple case of the spatial heterogeneity mentioned in the introduction, we assume that resources occur in discrete patches, which can only be incorporated into the territory in a fixed sequence. The only decision available to the primary pair is when to truncate the sequence. We assume that choosing a minimum territory corresponds to having as few patches as possible in the territory, i.e. truncating the sequence as soon as possible, subject to the condition that the territory must contain sufficient resources. This definition of a minimum

territory follows from identifying the number of patches in a territory with some measure of its size, and hence the cost of defending it.

Label the patches 1,2,3,... as they become available to the primary pair. Let R_j be the mean yield over time of patch j, i.e. its richness, and let V_j be the variance over time of its yield. Then, if we assume the yields from distinct patches to be independent, the distribution of the total yield from a territory containing n patches will have mean and variance given by

$$\mu = \sum_{j=1}^{n} R_j \text{ and } \sigma^2 = \sum_{j=1}^{n} V_j, \tag{3.2.1}$$

respectively. The territory set up by the primaries will then consist of N patches, where N denotes the smallest value of n such that the parameter values given in equation (3.2.1) satisfy the condition (3.1.11).

In any given habitat, we would expect some relationship between the mean, R_j, and the variance, V_j, of the patch yield. Initially we assume that

$$V_j = f^2 R_j \quad (j = 1,2,\ldots), \tag{3.2.2}$$

where f is the coefficient of variation of the yield from a patch of unit richness. This assumption implies that the total mean, μ, and the total variance, σ^2, are functionally related, with

$$\sigma^2 = f^2 \mu, \tag{3.2.3}$$

and hence the variability of yield from a territory is determined solely by its mean total yield, μ, and does not depend on, for example, the way the resources are divided into patches. The effect of relaxing this assumption is considered in Section 3.3.1.

If we substitute equation (3.2.3) into the condition (3.1.11) for a satisfactory territory we obtain the new criterion

$$\frac{\mu - b}{f\sqrt{\mu}} \geq v. \tag{3.2.4}$$

Note that b and f are clearly non-negative from their respective definitions, and v can be assumed to be non-negative, as discussed in Section 3.1.2. Hence condition (3.2.4) is satisfied if and only if

$$\mu - b \geq fv\sqrt{\mu}$$

or equivalently if and only if

$$\Leftrightarrow (\mu - b)^2 \geq f^2 v^2 \mu \tag{3.2.5}$$

and also

$$\mu \geq b. \tag{3.2.6}$$

We can solve (3.2.5) to obtain

$$\mu \geq b+\frac{f^2v^2}{2}+\sqrt{(\frac{f^2v^2}{2})^2+bf^2v^2} \tag{3.2.7}$$

or

$$\mu \leq b+\frac{f^2v^2}{2}-\sqrt{(\frac{f^2v^2}{2})^2+bf^2v^2}. \tag{3.2.8}$$

Hence we require (3.2.6) and either (3.2.7) or (3.2.8). But equation (3.2.8) implies

$$\mu-b-\frac{f^2v^2}{2} \leq -\sqrt{(\frac{f^2v^2}{2})^2+bf^2v^2} < -\frac{f^2v^2}{2}$$

$$\Rightarrow \mu < b,$$

which contradicts (3.2.6); and equation (3.2.7) implies

$$\mu-b-\frac{f^2v^2}{2} \geq \sqrt{(\frac{f^2v^2}{2})^2+bf^2v^2} > \frac{f^2v^2}{2}$$

$$\Rightarrow \mu-b > f^2v^2 > 0,$$

and hence (3.2.7) implies (3.2.6). So provided $b, f, v \geq 0$, conditions (3.2.4) and (3.2.7) are equivalent. Note that we can rewrite (3.2.7), our new criterion for a satisfactory territory, as

$$\mu \geq b+\frac{f^2v^2}{2}+fv\sqrt{\frac{f^2v^2}{4}+b}. \tag{3.2.9}$$

We denote the critical value of μ on the right hand side of (3.2.9) by m. Similarly, we can show that the territory can then support a secondary animal if and only if

$$\mu \geq m' = b'+\frac{f^2v'^2}{2}+fv'\sqrt{\frac{f^2v^2}{4}+b'} \tag{3.2.10}$$

Given the conditions above, whether or not secondary animals will be present depends on the values of $R_1, R_2, \ldots$, the richnesses of the individual patches. The simplest case is when the patches are all identical. Then the R_j's are all equal, and without loss of generality we can take $R_j = 1$ for all j (simply let the unit of yield be the yield of a single patch). So a territory will contain exactly $\lceil m \rceil$ patches (the smallest integer not less than m), and a secondary animal will be present if and only if

$$\lceil m' \rceil \leq \lceil m \rceil.$$

This condition is easy to evaluate in any specific case, but its dependence on the parameters b, γ, f, v and v' is complicated and discontinuous. Slight uncertainty about those parameter values to which the model is most sensitive can result in complete ignorance about whether

secondaries should be present in all territories or absent from all territories. In fact, the model with $R_j \equiv 1$ has many of the disadvantages of the model due to Carr and Macdonald (1986), discussed in Chapter 2.

A more tractable alternative is to allow differences in mean richness between patches. This is possibly more realistic: for example, it is true of the areas where badgers forage for earthworms, as described in Kruuk (1978a). Formally, we assume that R_j, the mean richness, over time, of patch j is a continuous random variable and that $R_1, R_2, \ldots$ are independent and identically distributed. Again, we assume that the primaries decide when to truncate the random sequence of patches to be included in their territory, but they may not choose between individual patches. The number of patches in the territory will then be the random variable

$$N = \min\{n: \sum_{j=1}^{n} R_j \geq m\}.$$

Since the distribution of R_j is continuous, with probability 1 we have

$$N = 1+N(m),$$

where

$$N(t) = \max\{n: \sum_{j=1}^{n} R_j \leq t\}. \tag{3.2.11}$$

The process $\{N(t); t \geq 0\}$ is a renewal process, so we can use results from renewal theory. Full definitions, and the results we will need, are given in Section 3.2.2. Given the value of N, a secondary animal will be present if and only if

$$\sum_{j=1}^{N} R_j \geq m'.$$

In terms of renewal theory, this condition can be written as

$$E(m) = \sum_{j=1}^{1+N(m)} R_j - m \geq m' - m, \tag{3.2.12}$$

where $E(m)$ is the excess life at m of the renewal process.

Since the mean patch yields $R_1, R_2, \ldots$ are random, condition (3.2.12) will be satisfied with some probability p. So, unlike the case of identical patches, the model in which $R_1, R_2, \ldots$ are random variables predicts that a territory formed under given conditions will have a certain probability of (permanently) supporting a secondary animal, and hence in a given habitat, a certain proportion of territories will support secondary animals. We adopt the latter model henceforth.

Note that since the distributions of $R_1, R_2, \ldots$ are continuous, and since

$$E(m) = \sum_{j=1}^{1+N(m)} R_j - m,$$

it follows that

$$\Pr(E(m) = m' - m) = 0.$$

Combining this result with the condition in equation (3.2.12), we can write the probability of one or more secondary animals being present as

$$p = \Pr(E(m) > m' - m), \tag{3.2.13}$$

a form which will be used in Section 3.2.2.

Since the amount of resources required by the primaries is measured in arbitrary units, we can, without loss of generality, assume $R_1, R_2, \ldots$ to have unit mean. This means that we regard b as measuring the requirements of the primary pair in terms of the mean patch richness. Hence differences in the spatial discreteness of food resources, i.e. in the mean patch richness relative to the requirements of the primary pair, are incorporated into the model by considering different values of b.

In subsequent sections we look at the behaviour of the model described here. Firstly, in Section 3.2.2, we review the ideas and results from renewal theory that we will need, and make some extensions to known results. Secondly, we look at the probability, p, of one or more secondaries being present, i.e. of a spatial group being formed. It is difficult to obtain exact, general results for this probability, but we can make progress in two ways. For some choices of the distribution of the patch richness R_j, exact values for p can be obtained. Some numerical results in these cases are discussed in Section 3.2.3. More generally, we obtain some analytic results based on the limiting distribution of the excess lifetime, $E(m)$, for large m, in Section 3.2.4. As well as the probability p of group formation, we are interested in the size of the group, and the number of food patches in a territory. In Section 3.2.5, we look at group size, territory size, and the relationship between them. Finally Section 3.2.6 is aimed at understanding the accuracy of the limiting approximation used in Section 3.2.4.

3.2.2 Renewal Theory

In this section, we collect together the definitions and results from renewal theory which we will need. These results are well known: see for example Cox (1967), and Grimmett and Stirzaker (1982). Proofs are generally omitted.

A renewal process is a random process of the form

$$N(t) = \max\{n : \sum_{i=1}^{n} X_i \leq t\}, \tag{3.2.14}$$

where $X_1, X_2, \ldots$ are independent and identically distributed non-negative random variables. We will sometimes refer to $X_1, X_2, \ldots$ as inter-arrival times or as lifetimes. Within the current section, $f_l(.)$ and $F_l(.)$ will represent the probability density function and cumulative distribution function respectively of $\sum_{i=1}^{l} X_i$, with $f(.) \equiv f_1(.), F(.) \equiv F_1(.)$. We define the renewal function to be

$$H(t) = \mathrm{E}(N(t)), \tag{3.2.15}$$

and the renewal density to be

$$h(t) = \frac{\mathrm{d}H(t)}{\mathrm{d}t}. \tag{3.2.16}$$

By considering events of the form $\{\sum_{i=1}^{l} X_i \leq t\}$, we can write

$$H(t) = \sum_{l=1}^{\infty} F_l(t), \tag{3.2.17}$$

and hence

$$h(t) = \sum_{l=1}^{\infty} f_l(t), \tag{3.2.18}$$

Alternatively, by conditioning on the value of X_1, we obtain what is known as the renewal equation,

$$H(t) = F(t) + \int_0^t H(t-x)f(x)\,\mathrm{d}x. \tag{3.2.19}$$

Taking Laplace transforms gives

$$H^*(s) = \frac{F^*(s)}{1-F^*(s)} = \frac{1}{s}\cdot\frac{f^*(s)}{1-f^*(s)} \tag{3.2.20}$$

and hence

$$h^*(s) = \frac{f^*(s)}{1-f^*(s)} \tag{3.2.21}$$

where "*" denotes Laplace transforms. Equations (3.2.20) and (3.2.21) may sometimes be of use in calculating $H(.)$ and $h(.)$.

If we further define

$$G(t) = \mathrm{E}[N(t)^2], \tag{3.2.22}$$

then conditioning on X_1 gives

$$G(t) = F(t) + 2\int_0^t H(t-x)f(x)\mathrm{d}x \int_0^t G(t-x)f(x)\mathrm{d}x, \tag{3.2.23}$$

and taking Laplace transforms gives

$$G^*(s) = \frac{f^*(s)}{1-f^*(s)}\left(\frac{1}{s}+2H^*(s)\right). \tag{3.2.24}$$

The excess life of a renewal process, or the time to the next event, is

$$E(t) = \sum_{i=1}^{N(t)+1} X_i - t. \tag{3.2.25}$$

Again, by conditioning on X_1, we obtain an integral equation:

$$\Pr(E(t) \leq y) = F(t+y) + \int_0^t h(x)\{1-F(t+y-x)\}\,dx \quad (y > 0), \tag{3.2.26}$$

Although the distribution of $E(.)$ generally depends on t, in most cases the distribution approaches a limit as $t \to \infty$. To state the result precisely, we need the following definition. A random variable X is **arithmetic** if, for some $\lambda > 0$, we always have

$$X \in \{n\lambda : n = 0, \pm 1, \ldots\}. \tag{3.2.27}$$

Theorem 3.1

If $E(.)$ is the excess lifetime of a renewal process with interarrival times $R_1, R_2, \ldots$, and R_1 is non-arithmetic with cumulative distribution function $F(.)$ and finite expectation, then

$$\Pr(E(t) \leq y) \to \{\mathrm{E}[R_1]\}^{-1}\int_0^y \{1-F(x)\}\,dx \quad \text{as } t \to \infty. \tag{3.2.28}$$

A particularly simple renewal process is one in which the lifetimes have an Exponential(λ) distribution, i.e.

$$F(x) = 1-e^{-\lambda x}. \tag{3.2.29}$$

The renewal process is then simply a Poisson process, and

$$\Pr(N(t) = n) = \frac{\lambda t^n e^{-\lambda t}}{n!}, \quad n = 0, 1, \ldots . \tag{3.2.30}$$

By the "lack of memory" property of the exponential distribution, the excess lifetime $E(t)$ of such a process is itself Exponential(λ), for any $t \geq 0$.

3.2.3 Numerical Results for p

In this section, we look at two types of distribution for the patch richness R_j for which exact results can be obtained: Erlang distributions and mixtures of two exponential distributions.

The Erlang distribution with parameters λ and k is the distribution of the sum of k independent and identically distributed random variables, each having the Exponential(λ) distribution. It is a special case of the Gamma distribution, and in turn contains the exponential distribution as a special case, when $k = 1$; (see e.g. Johnson and Kotz, (1970)). Its probability density is

$$f(x) = \lambda^k x^{k-1} e^{-\lambda x}/(k-1)!, \quad x > 0, \lambda > 0, k = 1,2,3,.... \tag{3.2.31}$$

It is always positive-valued and unimodal: its exact shape depends on the parameter k, which also determines its coefficient of variation, $1/\sqrt{k}$. It has mean k/λ, and so, since we require R_j to have unit mean, we always take $\lambda = k$.

A mixture of two exponential distributions (or just a mixture distribution, as we shall refer to it within the current section) has probability density

$$f(x) = \rho_1 \alpha_1 e^{-\rho_1 x} + \rho_2 \alpha_2 e^{-\rho_2 x}, \tag{3.2.32}$$

where $\alpha_1, \alpha_2 \geq 0$ are the proportions of the two components of the mixture, and $\rho_1, \rho_2 \geq 0$ are the rate parameters of the two components. We require $\alpha_1 + \alpha_2 = 1$, and in addition, to ensure that R_j has unit mean, $\dfrac{\alpha_1}{\rho_1} + \dfrac{\alpha_2}{\rho_2} = 1$.

From equation (3.2.13), the probability of at least one secondary being present is given by

$$\begin{aligned} p &= \Pr(E(m) > m' - m) \\ &= 1 - \Pr(E(m) \leq m' - m). \end{aligned} \tag{3.2.33}$$

A standard result from renewal theory, as given in equation (3.2.26), states that the cumulative distribution function of the excess $E(m)$ is given by

$$\Pr(E(m) \leq y) = F(m+y) + \int_0^m h(x)\{1 - F(m+y-x)\}\, dx \quad (y > 0). \tag{3.2.34}$$

In the Erlang case, if $k = 1$, so that R_j has the exponential distribution, we can calculate $h(.)$ directly. From equation (3.2.18), $h(u) = \sum_{l=1}^{\infty} f_l(u)$, and then the integral in equation (3.2.34) is straightforward. We find that

$$p = \begin{cases} e^{-(m'-m)} & \text{if } m' > m \\ 1 & \text{if } m' \leq m. \end{cases} \tag{3.2.35}$$

If $k = 2$, similarly,

$$p = \begin{cases} e^{-2(m'-m)}(1 + (m'-m)(1 + e^{-4m})) & \text{if } m' > m \\ 1 & \text{if } m' \leq m. \end{cases} \tag{3.2.36}$$

For general values of k, the integral (3.2.34) is hard to evaluate, but we can make progress by

exploiting the interpretation of a random variable with an Erlang(λ, k) distribution as the sum of k independent Exponential(λ) random variables.

Let $X_1, X_2, \ldots$ be independent and identically distributed Exponential(k) random variables, so that

$$R_j = \sum_{i=1+k(j-1)}^{kj} X_i \tag{3.2.37}$$

and

$$\sum_{j=1}^{n} R_j = \sum_{i=1}^{nk} X_i. \tag{3.2.38}$$

Let

$$C(t) = \max\{n : \sum_{i=1}^{n} X_i \leq t\} \tag{3.2.39}$$

be the renewal process with lifetimes given by $X_1, X_2, \ldots$. Then $E(m)$, the excess at m of the renewal process $\{N(t); t \geq 0\}$ defined in equation (3.2.11), is given by

$$\begin{aligned} E(m) &= \sum_{j=1}^{N(m)+1} R_j - m \\ &= \sum_{i=1}^{k(N(m)+1)} X_i - m \\ &= \sum_{i=1}^{L} X_i - m \end{aligned}$$

where $L = k\left\lceil \frac{C(m)+1}{k} \right\rceil$

$$\begin{aligned} &= \sum_{i=1}^{C(m)+1} X_i + \sum_{i=C(m)+2}^{L} X_i - m \\ &= E_C(m) + \sum_{i=C(m)+2}^{L} X_i - m, \end{aligned} \tag{3.2.40}$$

where $E_C(.)$ represents the excess of $\{C(t); t \geq 0\}$. The first term of equation (3.2.40) is the excess of an Exponential(k) renewal process, which as stated in Section 3.2.2 also has the Exponential(k) distribution. The second term is independent of the first, and is the sum of

$$k\left\lceil \frac{C(m)+1}{k} \right\rceil - C(m) - 1$$

independent Exponential(k) random variables. Hence the excess $E(m)$ has a distribution given by a mixture of Erlang distributions, with common rate parameter k. The mixture is over the second parameter of the Erlang distribution, corresponding to the number of exponential terms, which is given by the random variable

$$S = k\left\lceil \frac{C(m)+1}{k} \right\rceil - C(m), \tag{3.2.41}$$

obtained by combining both terms of (3.2.40).

We can uniquely write

$$C(m) = A+Bk,\ 0 \leq A < k,\ B \geq 0, \tag{3.2.42}$$

to give

$$\begin{aligned} S &= k\left\lceil \frac{A+Bk+1}{k} \right\rceil - A - Bk \\ &= k\left\lceil \frac{A+1}{k} \right\rceil - A \\ &= k-A. \end{aligned}$$

Hence $s \in \{1,\ldots,k\}$, and

$$\Pr(S = k-a) = \Pr((C(m) \bmod k) = a),\quad a = 0,\ldots,k-1, \tag{3.2.43}$$

where $i \bmod j$ denotes the remainder when i is divided by j. Now $C(m)$ is just the value at a certain point m of a renewal process with Exponential(k) inter-arrival times, so

$$C(m) \sim \text{Poisson}(km).$$

Hence $C(m) \bmod k$ has the wrapped Poisson distribution on $\{0,\ldots,k-1\}$ (see e.g. Mardia, 1972). The probabilities in equation (3.2.43) can readily be written as infinite sums, but the following finite form appears to be new.

Theorem 3.2

Let A have the Poisson(λ) distribution. The corresponding wrapped Poisson distribution on $\{0, 1,\ldots,k-1\}$ has probabilities given by

$$\Pr((A \bmod k) = a) = \frac{1}{k} e^{-\lambda} \sum_{j=0}^{k-1} e^{\omega_k^j \lambda} \omega_k^{-aj}, \tag{3.2.44}$$

$$= \frac{1}{k} e^{-\lambda} \sum_{j=0}^{k-1} \exp(\lambda\cos(\theta j))\cos(\theta aj + \lambda\sin(\theta j)), \tag{3.2.45}$$

for $a = 0,\ldots,k-1$, where $\theta = \dfrac{2\pi}{k}$, and ω_k is the complex kth root of unity $\cos(\theta)+i\sin(\theta)$.

Proof

Consider the continuous-time Markov chain $\{X(t): t \geqslant 0\}$ which cycles through the states $\{0,\ldots,k-1\}$, so that the only permissible transitions are from state j to state $j+1$, $j = 0,\ldots,k-2$, and from state $k-1$ to state 0. Let all permissible transitions have transition rates equal to λ, and let $X(0) = 0$. Since the number of transitions up to time t has the Poisson(λt) distribution, the distribution of $X(1)$ is just the wrapped Poisson distribution on $\{0,\ldots,k-1\}$ with rate λ.

The Markov chain X has generator

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot & 0 \\ 0 & & & \cdot & \cdot & \lambda \\ \lambda & 0 & \cdot & \cdot & 0 & -\lambda \end{pmatrix} \tag{3.2.46}$$

and transition probabilities

$$P_{ij}(t) = \Pr(X(t) = j \,|\, X(0) = i)$$

given by

$$P_t = \exp(tQ).$$

Writing $\pi^{(t)}$ for the probability distribution of $X(t)$, we know that $\pi^{(0)} = (1,0,\ldots,0)$, and we are interested in

$$\pi^{(1)} = \pi^{(0)}\exp(Q). \tag{3.2.47}$$

The matrix Q is circulant (Bellman, 1960; Lancaster, 1969) i.e. it is of the form

$$\begin{pmatrix} \lambda_0 & \lambda_1 & \cdot & \cdot & \cdot & \lambda_{k-1} \\ \lambda_{k-1} & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & \cdot & \lambda_1 \\ \lambda_1 & \cdot & \cdot & \cdot & \lambda_{k-1} & \lambda_0 \end{pmatrix}.$$

Thus Q has distinct eigenvalues

$$\lambda_j = \lambda(\omega_k^j - 1), \quad j = 0,\ldots,k-1,$$

with right eigenvectors

$$v_0 = \begin{pmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}, \quad v_1 = \begin{pmatrix} 1 \\ \omega_k^1 \\ \omega_k^2 \\ \cdot \\ \cdot \\ \omega_k^{k-1} \end{pmatrix}, \ldots, v_{k-1} = \begin{pmatrix} 1 \\ \omega_k^{k-1} \\ \omega_k^{2(k-1)} \\ \cdot \\ \cdot \\ \omega_k^{(k-1)(k-1)} \end{pmatrix}$$

(Bellman, 1960; Lancaster, 1969). Define Ω to be the symmetric $k \times k$ matrix (with indices running from 0 to $k-1$) with columns $v_0, \ldots, v_{k-1}$, so that

$$\Omega_{ij} = \omega_k^{ij}. \tag{3.2.48}$$

It can be shown by direct calculation that the inverse of Ω is given by

$$(\Omega^{-1})_{ij} = k^{-1}\omega_k^{-ij}. \tag{3.2.49}$$

From the results in, for example, Chapter 2 of Lancaster (1969), the rows of Ω^{-1}, which we denote by $w_j, j = 0, \ldots, k-1$, are left eigenvectors of Q,

$$Q = \Omega \Lambda \Omega^{-1},$$

where $\Lambda = \text{diag}\{\lambda_0, \ldots, \lambda_{k-1}\}$, and

$$Q = \sum_{j=0}^{k-1} \lambda_j E_j, \tag{3.2.50}$$

the spectral decomposition of Q, where

$$E_j = v_j w_j. \tag{3.2.51}$$

Furthermore, the matrices $E_0, \ldots, E_{k-1}$ have the properties

$$\sum_{j=0}^{k-1} E_j = I,$$

$$E_i E_j = 0, \quad i \neq j,$$

$$E_j^2 = E_j.$$

Thus

$$\exp(tQ) = = \sum_{j=0}^{k-1} \exp(t\lambda_j) E_j,$$

From equations (3.2.48), (3.2.49) and (3.2.51), we have

$$(E_j)_{il} = k^{-1}\omega_k^{(i-l)j}.$$

So from equation (3.2.47), the distribution of $X(1)$ is given by

$$\pi^{(1)} = \pi^{(0)}\exp(Q)$$

$$= \pi^{(0)}\sum_{j=0}^{k-1} \exp(\lambda_j)E_j,$$

giving

$$\Pr(X(1) = i) = \pi_i^{(1)}$$

$$= \sum_{j=0}^{k-1} e^{\lambda_j}(E_j)_{0i}$$

$$= \sum_{j=0}^{k-1} \exp(\lambda(\omega_k^j - 1))k^{-1}\omega_k^{-ij}$$

$$= \frac{1}{k}e^{-\lambda}\sum_{j=0}^{k-1} e^{\omega_k^j\lambda}\omega_k^{-ij}, \qquad (3.2.52)$$

as required. Using the fact that these probabilities are real, they can readily be rewritten, as shown in the statement of the theorem, in a form which is less concise but which is more useful for calculation, since it only involves real functions (3.2.45).

The following trivial corollary will be useful later on in this section. Define

$$J_k^a(\lambda) = \sum_0^\infty \frac{\lambda^{a+bk}}{(a+bk)!}, \quad a = 0,\ldots,k-1. \qquad (3.2.53)$$

Corollary to Theorem 3.2

$$J_k^a(\lambda) = \sum_{b=0}^\infty \frac{\lambda^{a+bk}}{(a+bk)!} = \frac{1}{k}\sum_{j=0}^{k-1} e^{\omega_k^j\lambda}\omega_k^{-aj} \qquad (3.2.54)$$

Proof

The corollary follows immediately from the obvious infinite expression for the wrapped Poisson distribution in the theorem:

$$\Pr((A|k) = a) = e^{-\lambda}\sum_{b=0}^\infty \frac{\lambda^{a+bk}}{(a+bk)!}.$$

The result in equation (3.2.5) is in fact known: the functions $J_k^a(.)$ are known as kth order hyperbolic functions, and arise in the solution of the differential equation

$$d^k y/dx^k - y = 0.$$

See Kaufman (1955) for a bibliography. Clearly equation (3.2.54) can be used to give an alternative, non-probabilistic proof of Theorem 3.2, but in fact the connection between the

wrapped Poisson distribution and the higher order hyperbolic functions does not appear to have been made before.

We can now evaluate the right hand side of equation (3.2.43), that is

$$\Pr(S = k-a) = e^{-km} J_k^a(km). \tag{3.2.55}$$

Thus we have evaluated the proportions in the mixture of Erlang distributions which gives the distribution of the excess $E(m)$. Each component of the mixture has an Erlang(k,s) distribution, for some $s \in \{1,\ldots,k\}$, and so has a cumulative distribution of the form

$$F(x) = 1 - e^{-kx} \sum_{j=0}^{s-1} \frac{(kx)^j}{j!}. \tag{3.2.56}$$

Hence if R_1 has the Erlang(k,k) distribution, the probability of secondary animals is

$$\Pr(E(m) > m'-m) = \sum_{s=1}^{k} e^{-km} J_k^{k-s}(km) \left\{ 1 - e^{-k(m'-m)} \sum_{j=0}^{s-1} \frac{(k(m'-m))^j}{j!} \right\}$$

$$= 1 - e^{-km'} \sum_{s=1}^{k} \left\{ J_k^{k-s}(km) \sum_{j=0}^{s-1} \frac{(k(m'-m))^j}{j!} \right\}. \tag{3.2.57}$$

In the case where the distribution of R_1 is a mixture of two exponential distributions, the cumulative distribution function of R_1 is

$$F(x) = 1 - \alpha_1 e^{-\rho_1 x} - \alpha_2 e^{-\rho_2 x}.$$

The renewal density $h(.)$ can be obtained from equation (3.2.21), as follows. The probability density function for R_1 is given by equation (3.2.32), and thus

$$f^*(z) = \frac{\alpha_1\rho_1}{\rho_1+z} + \frac{\alpha_2\rho_2}{\rho_2+z}$$

$$\Rightarrow h^*(z) = \frac{\rho_1\rho_2 + z(\alpha_1\rho_1+\alpha_2\rho_2)}{z^2 + z(\alpha_2\rho_1+\alpha_1\rho_2)}$$

$$= \frac{\rho_1\rho_2}{z(\alpha_2\rho_1+\alpha_1\rho_2)} + \frac{\alpha_1\alpha_2(\rho_1-\rho_2)^2}{(\alpha_2\rho_1+\alpha_1\rho_2)(z+\alpha_2\rho_1+\alpha_1\rho_2)}.$$

Hence

$$h(x) = \frac{\rho_1\rho_2}{\alpha_2\rho_1+\alpha_1\rho_2} + \frac{\alpha_1\alpha_2(\rho_1-\rho_2)^2}{\alpha_2\rho_1+\alpha_1\rho_2} e^{-(\alpha_2\rho_1+\alpha_1\rho_2)x}$$

$$= 1 + \frac{\alpha_1\alpha_2}{\rho_1\rho_2}(\rho_1-\rho_2)^2 e^{-\rho_1\rho_2 x}, \tag{3.2.58}$$

since we have

$$\frac{\alpha_1}{\rho_1}+\frac{\alpha_2}{\rho_2}=1 \quad \Rightarrow \quad \alpha_1\rho_2+\alpha_2\rho_1=\rho_1\rho_2.$$

Thus

$$\begin{aligned} p &= \Pr(E(m) > m'-m) \\ &= 1-F(m')+\int_0^m h(x)(1-F(m'-x))\mathrm{d}x \\ &= \alpha_1 e^{-\rho_1 m'}+\alpha_2 e^{-\rho_2 m'} \\ &+\int_0^m \left(\alpha_1 e^{-\rho_1(m'-x)}+\alpha_2 e^{-\rho_2(m'-x)}\right)\left(1+\frac{\alpha_1\alpha_2}{\rho_1\rho_2}(\rho_1-\rho_2)^2 e^{-\rho_1\rho_2 x}\right)\mathrm{d}x, \end{aligned}$$

which after integration and rearranging gives

$$\begin{aligned} p &= e^{-\rho_1 m'}\left(\alpha_1+\frac{\alpha_1}{\rho_1}(e^{\rho_1 m}-1)+\frac{\alpha_1^2\alpha_2}{\rho_1^2\rho_2}\frac{(\rho_1-\rho_2)^2}{(1-\rho_2)}(e^{\rho_1(1-\rho_2)m}-1)\right) \\ &+e^{-\rho_2 m'}\left(\alpha_2+\frac{\alpha_2}{\rho_2}(e^{\rho_2 m}-1)+\frac{\alpha_2^2\alpha_1}{\rho_2^2\rho_1}\frac{(\rho_2-\rho_1)^2}{(1-\rho_1)}(e^{\rho_2(1-\rho_1)m}-1)\right). \end{aligned} \tag{3.2.59}$$

We now wish to look at some actual numerical values for p. We start with a standard set of parameters, as follows: $b = 1.0$, chosen to represent the case where the mean yield of an average patch is equal to the requirements of the primary pair; $\gamma = 1.5$, implying that the amount of resources needed by a secondary is equal to that of a single primary; $v = 1.6$, $v' = 1.5$, values which were derived in Section 3.1; and $f = 0.5$, meaning that the standard deviation of the yield from a patch over time is equal to half its mean yield. If R_1 has an Exponential distribution with parameter 1, these standard parameters lead to $p = 0.57$. The effect on p when the distribution of R_1 is a mixture of two exponential distributions is shown in Table 3.4. Note that for our purposes, the family of mixtures of two exponentials is two-dimensional, defined by the four parameters $\alpha_1, \alpha_2, \rho_1$ and ρ_2 and the two constraints

$$\alpha_1+\alpha_2=1, \quad \frac{\alpha_1}{\rho_1}+\frac{\alpha_2}{\rho_2}=1. \tag{3.2.60}$$

Table 3.4 also shows δ, the coefficient of variation of the distribution of R_1, in each case. For particular cases selected from those in Table 3.4, the effect on p of varying the parameters v, γ, b and f is explored in Table 3.5, and compared with the exponential case. Values for m and m' are also given.

Similarly, Table 3.6 shows the effect on p of different values of the parameter k of the Erlang distribution, for different values of v, γ, b and f. For reference, the definitions of these parameters are collected in Table 3.7.

An immediate conclusion from Tables 3.5 and 3.6 is that the probability of supporting a secondary animal is reasonably high over a wide range of parameter values and distributions

Table 3.4

Values of p and coefficients of variation δ of R_1
for different values of $\alpha_1, \alpha_2, \rho_1$ and ρ_2
satisfying equations (3.2.60)

$\alpha_1 : \alpha_2$	$\rho_1 : \rho_2$					
	2 : 1		5 : 1		10 : 1	
	p	δ	p	δ	p	δ
1 : 20	0.57	1.01	0.58	1.03	0.58	1.04
1 : 10	0.57	1.02	0.59	1.06	0.59	1.08
1 : 5	0.58	1.04	0.60	1.11	0.61	1.15
1 : 2	0.58	1.08	0.61	1.24	0.64	1.32
1 : 1	0.58	1.11	0.62	1.37	0.67	1.53
2 : 1	0.58	1.12	0.61	1.52	0.67	1.80
5 : 1	0.57	1.10	0.57	1.61	0.58	2.15
10 : 1	0.57	1.07	0.55	1.56	0.52	2.25
20 : 1	0.57	1.04	0.56	1.42	0.52	2.15

of R_1, which suggests that spatial groups could be formed in this way over a broad range of different resource patterns.

Table 3.6 shows that, within the Erlang family, the probability p decreases with k. This is explained by the decreasing variability of the Erlang distribution as k decreases. When patch size is less variable, the chance of an excess large enough to accommodate a secondary is reduced.

Within the family of mixture distributions, however, there is no clear pattern to the effect of the parameters $\alpha_1, \alpha_2, \rho_1$ and ρ_2 on p, either directly or in terms of δ, the coefficient of variation of R_1. With a few exceptions, mentioned below, values of p within the mixture family are closer to the exponential case than the values in the Erlang case.

The effect of varying v, the required primary food security (cases 1, 2 and 3 in Tables 3.5 and 3.6), is that in all cases, p is an increasing function of v. Such an effect is intuitively

Table 3.5

Selected combinations of parameter values, and corresponding exact values for the probability p of a secondary animal being present

Case	Parameters (all values are as in case 1 except where noted)	m	m'	Probability p				
				$\rho_1 = 1$ $\rho_2 = 1$	$\frac{\alpha_1}{\alpha_2} = 1$ $\frac{\rho_1}{\rho_2} = 10$	$\frac{\alpha_1}{\alpha_2} = 2$ $\frac{\rho_1}{\rho_2} = 10$	$\frac{\alpha_1}{\alpha_2} = 10$ $\frac{\rho_1}{\rho_2} = 10$	$\frac{\alpha_1}{\alpha_2} = 20$ $\frac{\rho_1}{\rho_2} = 10$
1	$b = 1.0$, $\gamma = 1.5$, $\upsilon = 1.6$, $\upsilon' = 1.5$, $f = 0.5$	2.18	2.74	0.57	0.67	0.67	0.52	0.52
2	$\upsilon = 1.5$	2.08	2.74	0.52	0.63	0.64	0.47	0.47
3	$\upsilon = 1.8$	2.39	2.74	0.70	0.76	0.76	0.66	0.66
4	$\gamma = 1.25$	2.18	2.42	0.79	0.82	0.82	0.75	0.75
5	$\gamma = 2.00$	2.18	3.38	0.30	0.47	0.51	0.32	0.28
6	$b = 0.10$	0.83	0.84	0.99	0.99	0.99	0.99	0.99
7	$b = 0.50$	1.47	1.74	0.76	0.80	0.78	0.70	0.72
8	$b = 2.00$	3.50	4.61	0.33	0.49	0.53	0.39	0.33
9	$b = 10.00$	12.87	18.20	0.00	0.05	0.10	0.19	0.15
10	$f = 0.0$	1.00	1.50	0.61	0.68	0.64	0.51	0.53
11	$f = 1.0$	4.33	4.78	0.64	0.72	0.72	0.63	0.61
δ	-	-	-	1.00	1.53	1.80	2.25	2.15

clear, since increasing υ means that the total yield m required by the primaries will be greater, and the requirements of a potential secondary are unchanged, so it is more likely that a secondary can be accommodated.

The effect of varying γ (cases 1, 4 and 5), which measures the amount of extra resources required by a secondary animal, is again intuitively clear. Increasing γ means that the amount of resources m' required by a secondary is greater, and so the probability of a secondary being

Table 3.6

Selected combinations of parameter values, and corresponding exact values for the probability p of a secondary animal being present

Case	Parameters (all values are as in case 1 except where noted)	m	m'	Probability p			
				$k = 1$	$k = 2$	$k = 5$	$k = 10$
1	$b = 1.0$, $\gamma = 1.5$, $v = 1.6$, $v' = 1.5$, $f = 0.5$	2.18	2.74	0.57	0.51	0.38	0.18
2	$v = 1.5$	2.08	2.74	0.52	0.44	0.30	0.15
3	$v = 1.8$	2.39	2.74	0.70	0.67	0.59	0.34
4	$\gamma = 1.25$	2.18	2.42	0.79	0.77	0.70	0.49
5	$\gamma = 2.00$	2.18	3.38	0.30	0.20	0.08	0.03
6	$b = 0.10$	0.83	0.84	0.99	0.99	0.99	0.97
7	$b = 0.50$	1.47	1.74	0.76	0.74	0.63	0.42
8	$b = 2.00$	3.50	4.61	0.33	0.23	0.12	0.02
9	$b = 10.00$	12.87	18.20	0.00	0.00	0.00	0.00
10	$f = 0.0$	1.00	1.50	0.61	0.56	0.46	0.38
11	$f = 1.0$	4.33	4.78	0.64	0.59	0.57	0.42

accommodated is smaller. Hence p is a decreasing function of γ. Low values of γ might be relevant if the amount of food required by a secondary is small compared with the amount required by the primaries, for example because the primaries are breeding. A high value of γ is relevant if the amount of food required by a secondary is comparable with that required by the primaries, for example in the case where there is only a single primary, or where we consider a secondary pair.

The effect of varying b (cases 1, 6, 7, 8 and 9) is of particular importance, since b can be regarded as measuring the scale of spatial patchiness of the resource distribution, and may well vary between the different habitats of a given species. For this reason, a wide range of numerical values of b is considered. It can be seen that low values of b give rise to high values of p. Such low values of b represent those cases where the requirements of the

Table 3.7

Definitions of main parameters

Parameters	Meaning
b	The resource requirements of the primary pair, in units in which the mean patch richness is 1
c	The food security required by the primary pair
c'	The food security required by an individual secondary
γ	The ratio of (amount of resources needed by primaries plus one secondary) to (amount of resources needed by primaries)
f	The coefficient of variation of the yield from a patch of unit richness i.e. the standard deviation of its yield over time divided by its mean yield
k	The shape parameter of the Erlang distribution describing the distribution of mean richness of individual patches
m	The total mean yield which a territory must have to support the primary pair
m'	The total mean yield which a territory must have to support a secondary animal in addition to the primary pair
p	The probability that a territory formed in given conditions will support a secondary animal
v	$\Phi^{-1}(c)$; a measure of required primary food security
v'	$\Phi^{-1}(c')$; a measure of required secondary food security
α_1, α_2	The weights of the components in a mixture distribution
ρ_1, ρ_2	The rate parameters of the components in a mixture distribution
δ	The coefficient of variation of the distribution of mean richness

primary pair are small compared with the average patch richness, i.e. cases of high spatial patchiness. Hence the numerical results support the intuitive idea that greater spatial patchiness gives a greater chance for secondaries to be accommodated. For small values of b ($b = 0.1$, case 9), the probability p is very small ($p < 0.005$), except where the distribution of R_1 is a mixture of exponentials, in which case p may be much greater. This suggests that in an environment where the variability of patch means is great, spatial groups may occur even though the overall mean patch richness is small.

Finally we consider the effect of f (cases 1, 10 and 11), which measures the variability over time of the yield from a resource patch, and hence from a territory. When $f = 0$, all yields are constant over time, i.e. there is no temporal heterogeneity. However, secondary animals can still sometimes be accommodated (case 10 in the tables), because of the spatial heterogeneity in the model. If f is now increased, the effect initially is to decrease the probability p (case 1), for most distributions of R_1 (Exponential, Erlang, and some mixtures). This result is highly counter-intuitive, since it means that in some cases, increased temporal heterogeneity results in a reduced chance of secondary animals being present. If f is further increased, then the value of p increases again (case 11). Hence, for many distributions of R_1, there is some level of temporal heterogeneity (here close to $f = 0.5$) which gives a minimum value for the probability of a secondary, and either higher or lower values of f give increased values for p. A partial explanation for this is given in Section 3.2.4. For some mixture distributions, p is simply an increasing function of f.

All these changes, apart from the effect of f on p when R_1 has certain mixture distributions, can be qualitatively explained by regarding the probability of at least one secondary being present, p, as a function of the additional mean yield required to support a secondary animal, $m' - m$. Such an approach is explored in Section 3.2.4.

3.2.4 Analytic results for p

As well as numerical results for p, we can obtain some analytic results which do not depend on the exact distribution of R_1. We need to use a result from renewal theory which gives the limiting distribution of the excess lifetime, $E(m)$, as defined in equation (3.2.25). From Theorem 3.1 in Section 3.2.2, we know that, under certain regularity conditions,

$$\Pr(E(m) \leq y) \rightarrow \{\mathrm{E}[R_1]\}^{-1} \int_0^y \{1 - F(x)\}\, \mathrm{d}x \quad \text{as } m \rightarrow \infty.$$

For our purposes, R_1 will always have a continuous distribution and expectation 1, so the theorem will apply. The theorem states that for large m, the excess lifetime, $E(m)$, has a limiting distribution which is independent of m. Thus p the probability of a secondary animal being supported, is a decreasing function of $m' - m$, the additional mean yield required to

support the secondary animal. For many distributions of R_1, not necessarily within the families described in Section 3.2.2, this limiting distribution is a good approximation even for quite small values of m, so we can determine the behaviour of the model in many cases from the behaviour of the difference $m'-m$. Hence, in this section, we examine the behaviour of $m'-m$ analytically. The accuracy of this approximation is discussed in Section 3.2.6.

From equations (3.2.4) and (3.2.9), we have

$$\frac{m-b}{f\sqrt{m}} = v, \tag{3.2.61}$$

or equivalently

$$m = b+\frac{f^2v^2}{2}+fv\sqrt{\frac{f^2v^2}{4}+b}, \tag{3.2.62}$$

and similar equations for m'. We want to investigate the effects of the parameters v, v', γ, b, and f.

The effect of changing v or v' is straightforward. Increasing v increases m, decreasing $m'-m$, and increasing p; increasing v' increases m', increasing $m'-m$, and decreasing p. The magnitude of the effect depends mainly on f. This can be seen by differentiating equation (3.2.61) and rearranging to obtain

$$\frac{dm}{dv} = \frac{f\sqrt{m}}{1-\frac{fv}{2\sqrt{m}}}$$

Note that since $m > f^2v^2$, the denominator is always between $\frac{1}{2}$ and 1, so dm/dv is positive.

Similarly, increasing γ increases m' and so decreases p, since

$$\frac{dm'}{d\gamma} = \frac{b}{1-\frac{fv'}{2\sqrt{m'}}}.$$

is positive.

The effect of changing b is less obvious, but it can be shown that provided

$$\gamma \geq (v/v')^2, \tag{3.2.63}$$

$m'-m$ is an increasing function of b. Condition (3.2.63) is always true if $v = v'$, and is true for most of the parameter sets considered here.

The behaviour as f changes, however, is not monotonic. No global analytic results have been obtained, but limiting cases can be considered. The exact expression for the derivative is

$$\frac{d(m'-m)}{df} = f(v'^2-v^2)+v'(\frac{f^2v'^2}{4}+\gamma b)^{\frac{1}{2}}+\frac{f^2v'^3}{4}(\frac{f^2v'^2}{4}+\gamma b)^{-\frac{1}{2}}$$

$$-v(\frac{f^2v^2}{4}+b)^{\frac{1}{2}}-\frac{f^2v^3}{4}(\frac{f^2v^2}{4}+b)^{-\frac{1}{2}}.$$

For small f,

$$\frac{d(m'-m)}{df} = (v'\sqrt{\gamma}-v)\sqrt{b} + O(f).$$

So near $f = 0$, $m'-m$ is increasing with f if and only if the inequality (3.2.63) holds. Since condition (3.2.63) will normally hold for realistic parameter values, this means that if variability of yield is low, increasing it will reduce the probability of forming spatial groups, which seems counter-intuitive.

For large f, on the other hand, we have

$$\frac{d(m'-m)}{df} = 2f(v'^2-v^2) + O(f^{-1})$$

and

$$m'-m = 2(\gamma-1)b+f^2(v'^2-v^2) + O(f^{-2}).$$

From the definition of v and v', we expect $v' \leqslant v$. If $v' = v$, then for large f

$$m'-m \approx 2(\gamma-1)b > 0,$$

so p approaches a limit, satisfying $0 < p < 1$, as f tends to infinity. In this case, p may be decreasing with f for all f. In the more usual case of $v' < v$, $m'-m$ is decreasing for sufficiently large f, and

$$m'-m \to -\infty \text{ as } f \to \infty.$$

Thus $p = 1$ for sufficiently large f.

The results for large f are sufficient to show that the behaviour of p as f changes is not necessarily monotonic. In fact, as has been seen in Section 3.2.2, numerical results show that the value of f at which p is a minimum can be quite low, say $\frac{1}{2}$, which is clearly a meaningful value for this parameter.

To try to understand these results, we can make a crude approximation to $d(m'-m)/df$. We differentiate equation (3.2.61), to get

$$\frac{dm}{df} = \frac{v\sqrt{m}}{1-\frac{fv}{2\sqrt{m}}},$$

and then assume that the second term in the denominator is small compared with 1, i.e. that $\frac{dm}{df} \approx v\sqrt{m}$. Combining this with a similar result for m' gives

$$\frac{d(m'-m)}{df} \approx v'\sqrt{m'} - v\sqrt{m}.$$

For given values of v and v', the sign of this expression depends on m'/m. At low values of f, m' is clearly larger than m; in fact $m'/m \approx \gamma$. As f increases, both m and m' increase, but the ratio m'/m decreases. Eventually, for large f, we reach a point where $v'\sqrt{m'} - v\sqrt{m} < 0$. In fact, it can be shown that the true expression for $d(m'-m)/df$ changes sign earlier, i.e. for lower values of f.

3.2.5 Territory Size and Group Size

As well as p, the probability of a spatial group being formed, we are interested in the size of the group. To calculate the mean number of secondaries, we need to take into account the probability of more than one secondary in a territory. This can be done by considering increasing values of γ. For simplicity, if we assume that all individuals have the same requirements, and the primary group is just a pair, then we have to consider γ taking the sequence of values 1.5, 2.0, 2.5,... and calculate the corresponding sequence of values of p. By a standard result of probability theory (see e.g. DeGroot (1986), p. 192), since the number of secondaries is a non-negative integer, its expectation is

$$\sum_{s=1}^{\infty} \Pr(s \text{ or more secondaries})$$

$$= \sum_{s=1}^{\infty} p(1+\frac{s}{2}),$$

where $p(x)$ is the value of p when $\gamma = x$. So the mean number is just the sum of this series of values of p, which can be easily approximated since the terms approach zero quite quickly.

We are also interested in territory size, represented in this model by the number of patches. This is $N = 1+N(m)$, as given above, and its expectation can be calculated from renewal theory. For example, in the Erlang case, if $k = 1$, the mean number of patches per territory is

$$EN = 1+H(m) = 1+m,$$

and if $k = 2$,

$$EN = 1+H(m) = m+\tfrac{3}{4}+\tfrac{1}{4}e^{-4m}.$$

Such results can be combined with the expected group sizes to give the population density, and hence the overall efficiency of resource use, predicted by the model. Examples for $k = 2$, and $f = 0.0$, 0.5, and 1.0 are given in Table 3.8 below.

Table 3.8

Territory Sizes and Group Sizes

	f=0.0	f=0.5	f=1.0
Expected territory size in patches	1.75	2.93	5.08
Expected number of secondaries per territory	1.05	0.82	0.88
Total expected number of individuals per territory	3.05	2.82	2.88
Expected number of individuals per patch	1.74	1.78	0.57

Note that the theoretical maximum value for the total number of (primary and secondary) individuals per patch is 2, since in theory the average resource requirement would be 1 patch for every 2 animals (since b=1), if there were no variability between patches or over time. From the figures for mean group size, it can be seen that, in these cases, the model predicts that secondary animals may make up roughly a third of the adult population (ignoring any individuals who are completely non-territorial).

For general k, the techniques given in Section 3.2.2 for the calculation of $\Pi(.)$ are rather difficult. As in Section 3.2.3, we use an alternative method which exploits the special properties of the Erlang distribution. Let $\{C(t): t \geq 0\}$ be as in equation (3.2.39). Then

$$C(m) \sim \text{Poisson}(\lambda),$$

where we write $\lambda = km$ for brevity. Now

$$N = \left\lceil \frac{C(m)+1}{k} \right\rceil = \left\lceil \frac{A+bk+1}{k} \right\rceil,$$

say, so

$$\sum_{a+bk \geq 0} \sum \frac{\lambda^{a+bk} e^{-\lambda}}{(a+bk)!} \left\lceil \frac{a+bk+1}{k} \right\rceil$$

$$= e^{-\lambda} \sum_{a+bk \geq 0} \sum \frac{\lambda^{a+bk}}{(a+bk)!} (b+1)$$

$$= e^{-\lambda} \sum\sum_{a+bk \geq 0} \frac{\lambda^{a+bk}}{(a+bk)!}\left(\frac{a+bk}{k} + \frac{k-a}{k}\right)$$

$$= \frac{e^{-\lambda}}{k}\left\{ \sum\sum_{a+bk \geq 0} \frac{\lambda^{a+bk}}{(a+bk)!}(a+bk) + \sum\sum_{a+bk \geq 0} \frac{\lambda^{a+bk}}{(a+bk)!}(k-a)\right\}$$

$$= \frac{e^{-\lambda}}{k}\left\{ \sum\sum_{a+bk > 0} \frac{\lambda^{a+bk}}{(a+bk-1)!}(a+bk) + \sum_{a=0}^{\infty}(k-a)\sum_{b=0}^{\infty} \frac{\lambda^{a+bk}}{(a+bk)!}\right\}$$

$$= \frac{1}{k}\left\{\lambda + e^{-\lambda}\sum_{a=0}^{\infty}(k-a)J_k^a(\lambda)\right\}. \qquad (3.2.64)$$

In the mixture case, we have already obtained the renewal density $h(.)$, in equation (3.2.58), so we can write down

$$\begin{aligned} EN &= 1 + H(m) \\ &= 1 + \int_0^m h(x)\,dx \\ &= 1 + \left[x - \frac{\alpha_1\alpha_2}{\rho_1^2\rho_2^2}(\rho_1-\rho_2)^2 e^{-\rho_1\rho_2 x}\right]_0^m \\ &= 1 + m + \frac{\alpha_1\alpha_2}{\rho_1^2\rho_2^2}(\rho_1-\rho_2)^2(1-e^{-\rho_1\rho_2 x}). \end{aligned} \qquad (3.2.65)$$

Thus we have exact expressions for mean territory sizes for the distributions introduced in Section 3.2.3.

So far, we have looked at differences in probabilities and mean sizes of spatial groups between different sets of parameter values, representing different habitats or different species. We are also interested in variations within a habitat, which arise from the stochastic nature of our model. Macdonald (1983) discusses field data which enables both within and between habitat relationships to be explored.

The variance of the number of patches in a territory in the Erlang case can be calculated in the same way as the expectation E[N]. Firstly, we write

$$(b+1)^2 = \frac{1}{k^2}\{(a+bk)(a+bk-1) + (2k-2a+1)(a+bk) + (k-a)^2\}.$$

Then

$$e^{\lambda}k^2 E[N^2] = \sum\sum_{a+bk \geq 0} \frac{\lambda^{a+bk}}{(a+bk)!}\{(a+bk)(a+bk-1+(2k-2a+1)(a+bk)+(k-a)^2\}$$

$$= \lambda^2 e^{\lambda} + (2k+1)\lambda e^{\lambda} + \sum_{a=0}^{k-1}(k-a)^2 J_k^a(\lambda) - 2\sum_{a=0}^{k-1} a\sum_{b=0}^{\infty} \frac{\lambda^{a+bk}}{(a+bk)!}(a+bk).$$

But

$$\sum_{a=0}^{k-1} a \sum_{b=0}^{\infty} \frac{\lambda^{a+bk}}{(a+bk)!}(a+bk) = \lambda \sum\sum_{a+bk \geq 0} a \frac{\lambda^{a+bk-1}}{(a+bk-1)!}$$

$$= \lambda \sum_{a=0}^{k-1} \sum_{b=0}^{\infty} a \frac{\lambda^{a+bk-1}}{(a+bk-1)!}$$

$$= \lambda \sum_{a=0}^{k-2} (a+1) J_k^a(\lambda)$$

$$= \lambda \sum_{a=0}^{k-1} (a+1) J_k^a(\lambda) - \lambda k J_k^{k-1}(\lambda).$$

Hence

$$\mathrm{E}[N^2] = \frac{e^{-\lambda}}{k^2}\left\{\lambda^2 e^{\lambda} + (2k+1)\lambda e^{\lambda} + \sum_{a=0}^{k-1} (k-a)^2 J_k^a(\lambda) \right.$$

$$\left. - 2\lambda \sum_{a=0}^{k-1} (a+1) J_k^a(\lambda) + 2\lambda k J_k^{k-1}(\lambda)\right\}. \qquad (3.2.66)$$

If $k = 1$, then R_1 is distributed exponentially, and its variance is just equal to m. If $k = 2$, the variance of the number of patches is

$$\tfrac{1}{2}m + \tfrac{1}{16} - me^{-4m} - \tfrac{1}{16}e^{-8m}.$$

Expressions for higher values of k are increasingly complicated, but in general the variance decreases with k. So, for case 1 of Table 3.5, with $m = 2.18$, the number of patches has mean 3.18, and standard deviation 1.48, if $k = 1$, and mean 2.73, and standard deviation 0.56, if $k = 10$.

In the mixture case, we can use equation (3.2.24) to obtain an expression for $G^*(s)$, where

$$G(t) = \mathrm{E}[N^2(t)].$$

Unfortunately, the expression obtained for $G^*(s)$ cannot readily be inverted to find $G(t)$, and so numerical results are difficult to obtain for the mixture case.

The variance of group size can be calculated in a similar way to the calculation of the mean group size at the beginning of Section 3.2.5. For example, for case 1 of Table 3.5, group size (including primary pair) has mean 3.23, standard deviation 1.60, if $k = 1$, and mean 2.21, standard deviation 0.48, if $k = 10$.

It is also of interest to look at the correlation r between group size and territory size (in patches) within a habitat. In general, the correlation is difficult to calculate, but we can make some progress with special cases. In the where R_1 has the Exponential distribution, r is zero, as the distribution of $E(m)$, and hence the probability of a given number of secondaries, is

independent of the number of patches. This follows from the "lack of memory" property of the exponential distribution (see e.g. Grimmett and Stirzaker, 1982). In other cases, exact calculations taking into account the possibility of more than one secondary animal are difficult. However, in the Erlang case, it is possible to calculate the correlation r' between the presence or absence of secondaries, and the number of patches in a territory. When $k = 2$, r' is found to be

$$(m'-m)e^{-2(m'-m)}(1-8me^{-4m}-e^{-8m})$$

$$\times(8m+1-16me^{-4m}-e^{-8m})^{-\frac{1}{2}}$$

$$\times(1-e^{-2(m'-m)}+(m'-m)(1+e^{-4m})+e^{-2(m'-m)}(m'-m)^2(1+e^{-4m})^2)^{-\frac{1}{2}}. \qquad (3.2.67)$$

Numerical evaluation shows that with $k = 2$, r' is generally small, with a maximum value of approximately 0.08 . So in the case $k = 2$, group size and territory size will appear approximately uncorrelated. For higher values of k, numerical results show that r' increases markedly, with values of up to approximately 0.5 when $k = 10$, for example. So in the case of high values of k, group size and territory size will be positively correlated.

An intuitive explanation for this correlation is that if a territory contains a large number of patches, it is more probable that only a small part of the contribution made by the final patch was needed to complete the territory. Hence it is more probable that the excess, i.e. the rest of the contribution from the final patch, is large enough to support at least one secondary. Note that this argument depends on the shape of the distribution of R_1, but is valid for any Erlang distribution with $k \geq 2$.

The results here show that, provided the distribution of R_1 is close to exponential, territory sizes (in patches) and group sizes within a habitat are highly variable, and approximately independent. This is a possible explanation for the observed independence of these quantities for some species, as summarised in Macdonald (1983). In contrast with territory size, the total resource yield of a territory will always be highly correlated with group size. In fact, group size is a monotonically increasing function of total yield, because of the basic criterion in the model that a given number of secondaries will be present if and only if the total yield exceeds a given level, as discussed in Section 3.1.

3.2.6 Rate of Convergence of $E(m)$.

In Section 3.2.4, we used the fact that $E(m)$, the excess total mean yield in a territory, is approximately independent of m, the total mean yield required by the primary pair, to study the behaviour of p, the probability of the territory supporting secondaries. For cases where we have exact results, this approximation is very accurate. (In fact, for the exponential case it is exact, but independence does not hold exactly for any other distribution - see e.g. Grimmett

and Stirzaker (1982), p289.) But how accurate is it in general? We give a partial answer here, by giving an analytic upper bound on the difference between the exact value of p and the approximation.

Theorem 3.3

Consider a renewal process with lifetimes $R_1, R_2, \ldots$, satisfying the conditions of Theorem 3.1, with $\mathrm{E}[R_1] = 1$. Let $f(.)$ and $F(.)$ denote the probability density function and cumulative distribution function respectively of R_1, and $E(.)$ and $h(.)$ denote the excess lifetime and renewal density respectively of the process, as in Section 3.2.2. Write $G(u)$ for $1-F(u)$, let

$$p_m(x) = \Pr(E(m) > x),$$

$$p_e(x) = \lim_{m \to \infty} \Pr(E(m) > x).$$

Then for any x,

$$|p_m(x) - p_e(x)| \leq \int_0^m |h(u)-1| G(m-u)\, du + \int_m^\infty |f(u)-G(u)|\, du. \qquad (3.2.68)$$

Proof.

From equation (3.2.26), the exact distribution of the excess at m is given by

$$p_m(x) = G(m+x) + \int_0^m h(u) G(m+x-u)\, du$$

and from Theorem 3.1, the equilibrium distribution, independent of m, is given by

$$p_e(x) = \Pr(E > x) = \int_x^\infty G(u)\, du.$$

Hence

$$\begin{aligned} |p_m(x) - p_e(x)| &= \left| G(m+x) + \int_0^m h(u) G(m+x-u)\, du - \int_x^\infty G(u)\, du \right| \\ &= \left| G(m+x) + \int_0^m (h(u)-1) G(m+x-u)\, du - \int_{m+x}^\infty G(u)\, du \right| \\ &\leq \left| \int_0^m (h(u)-1) G(m+x-u)\, du \right| + \left| G(m+x) - \int_{m+x}^\infty G(u)\, du \right|. \end{aligned}$$

But

$$\begin{aligned} &\left| G(m+x) - \int_{m+x}^\infty G(u)\, du \right| \\ &= \left| \int_{m+x}^\infty f(u)\, du - \int_{m+x}^\infty G(u)\, du \right| \end{aligned}$$

$$= \left| \int_{m+x}^{\infty} (f(u)-G(u))\, du \right|$$

$$\leq \int_{m+x}^{\infty} |f(u)-G(u)|\, du$$

$$\leq \int_{m}^{\infty} |f(u)-G(u)|\, du; \qquad (3.2.69)$$

and

$$\left| \int_{0}^{m} (h(u)-1)G(m+x-u)\, du \right|$$

$$\leq \int_{0}^{m} |h(u)-1|G(m+x-u)\, du$$

$$\leq \int_{0}^{m} |h(u)-1|G(m-u)\, du. \qquad (3.2.70)$$

So

$$|p_m(x)-p_e(x)| \leq \int_{m}^{\infty} |f(u)-G(u)|\, du + \int_{0}^{m} |h(u)-1|G(m-u)\, du,$$

as required.

The right hand side of equation (3.2.68) does not depend on x, so provided the renewal density $h(.)$ can be evaluated, a bound can be calculated on the difference between the exact value of p and the approximation, which depends only on m. For values of m for which this bound is small, the approximation discussed at the beginning of Section 3, that p is a decreasing function of $m'-m$, can be used.

For example, when R_1 has the Erlang distribution with $k = 2$, we have from equation (3.2.68) that

$$|p_m(x)-p_e(x)| \leq 2me^{-2m} \text{ if } m \geq \tfrac{1}{2},$$

with a rather more complicated expression if $m < \frac{1}{2}$. In this case the bound has maximum value $e^{-1} \approx 0.368$, when $m = \frac{1}{2}$, and decreases rapidly with m, when $m > \frac{1}{2}$. For example, if $m = 2$, the value of the bound is 0.073. So although we have calculated exact values in this case, the approximation would have performed well.

The bound given in equation (3.2.68) is not tight, i.e. it is not the best possible such bound. The bound is independent of x but depends on m and $F(.)$. Thus, writing $B(m,F)$ for the right hand side of equation (3.2.68), for the bound to be tight we would require

$$\sup_x |p_m(x)-p_e(x)| = B(m,F) \qquad (3.2.71)$$

for all m and $F(.)$. Clearly, under fairly weak regularity conditions, consideration of the inequalities (3.2.69) and (3.2.70) shows that if equation (3.2.71) holds, the supremum must be

attained at $x = 0$. In such cases, we would therefore require, for tightness,

$$B(m,F) = |p_m(0) - p_e(0)| = 0,$$

which is clearly not true in general. Thus the bound is not tight, in general. However the method of proof of Theorem 3.3, which relies on constructing a bound which is a monotonic function of x, and then eliminating x as in the inequalities (3.2.69) and (3.2.70), does not appear to extend to any improved bound. Thus the bound in Theorem 3.3 appears to be the best which can be obtained by this approach, and possibly the best which is simple enough to be useful.

3.3 Generalisations of the Model

3.3.1 Constant Coefficient of Variation for Patches

In Section 3.2.1, equation (3.2.2), we assumed for simplicity that the means and variances of the yield from individual patches are related by

$$V_j = f^2 R_j, \tag{3.3.1}$$

so that the total mean

$$\mu = \sum_{j=1}^{N} R_j$$

and the total variance

$$\sigma^2 = \sum_{j=1}^{N} V_j$$

satisfy the simple relationship

$$\sigma^2 = f^2 \mu.$$

In practice, the relationship (3.3.1) may not always hold. A plausible alternative is

$$V_j = f^2 R_j^2 \tag{3.3.2}$$

so that individual patches all have coefficient of variation f, regardless of richness. This may be more realistic for certain types of resources. Since the variance in yield from a territory now depends on the number and relative sizes of its component patches, as well as their total richness, this model is harder to analyse, but it can be simulated.

Table 3.9 compares such simulated values of p, the probability of secondary animals being present, in the case where equation (3.3.2) holds (column A), with exact values in the case where equation (3.3.1) holds (column B, copied from Table 3.6), in the case $k = 1$. The

parameter values in case 1 are $b = 1.0, \gamma = 1.5, \upsilon = 1.6, \upsilon' = 1.5, f = 0.5$.

Table 3.9

Comparison of values for p, the probability of secondary animals, using two different models for the variability of resource yield, assuming (A) a constant coefficient of variation, or (B) a variance proportional to the mean.

Case	Parameters	A	B
1	As in text	0.50	0.57
2	$\upsilon = 1.5$	0.40	0.52
3	$\upsilon = 1.8$	0.70	0.70
4	$\gamma = 1.25$	0.77	0.79
5	$\gamma = 2.00$	0.16	0.30
6	$b = 0.10$	0.91	0.99
7	$b = 0.50$	0.70	0.76
8	$b = 2.00$	0.23	0.33
9	$b = 10.00$	0.00	0.00
10	$f = 0.0$	0.58	0.61
11	$f = 1.0$	0.67	0.64

The main conclusions are that the behaviour is broadly similar in the two models, but the simulated model generally gives slightly lower probabilities, and is more sensitive to the values of υ and γ. It appears that in most cases, the simpler model is likely to be adequate.

3.3.2 More Than One Resource Type

Another generalisation of the model is to look at a territory which must be large enough to give adequate supplies of more than one type of resource. A typical example might be where the territory itself is fixed throughout the year, but food resources are strongly seasonal, with different food types used at different times of the year.

We assume that the requirements for different resource types can be considered individually. Using suffices to indicate the s types, we define b_i, for example, to be the amount of resources of type i required by the primary pair, for $i = 1, \ldots, s$. The meanings of b_i', f_i, v_i etc. are similarly defined, by analogy with the parameters introduced in Section 3.2.1. Given such a description of the requirement for each food type, we can calculate the required total means m_i, $i = 1, \ldots, s$. The territory is complete when $\mu_i \geq m_i$ for all i, i.e. when the territory has sufficient mean yields of **each** food type.

We also assume that each patch contains resources of a single type, i.e. a patch of type i only affects the value of μ_i, with each patch being of type i with probability q_i, independently of all other patches. Thus $q_1 + q_2 + \ldots + q_s = 1$. Again, the patches are assumed to be incorporated into the territory sequentially. If i_N is the type of the final patch, then the amount of resources of that type will be as in the single resource case, whereas in general other types may have a greater excess than in the simple model, because some patches of type $i, i \neq i_N$, may be added after $\mu_i \geq m_i$.

Exact calculations in this case are difficult, but we can obtain some crude bounds on the probability of secondary animals being present. We define:

A_i to be the event $\mu_i \geq m_i'$ in the single resource model, i.e. $\mu_i \geq \mu_i'$ even if we ignore patches added after $\mu_i \geq m_i$;

A_S to be the event that A_i occurs for all $i = 1, \ldots, s$ (i.e. $A_S = \bigcup_{i=1}^{s} A_i$);

A_N to be the event $\mu_{i_N} \geq m_{i_N}'$, i.e. the territory has sufficient resources of the type of the final patch to support at least one secondary; and

A^* to be the event that for all i, the final value of μ_i in the multiple resources model satisfies $\mu_i \geq m_i'$, i.e. secondary animals are predicted in the multiple resources model.

Then we have

$$A_S \Rightarrow A^* \Rightarrow A_N,$$

So, with the obvious notation for probabilities, we have

$$\prod_{i=1}^{s} p_i \leq p^* \leq \max_{i=1,\ldots,s} \{p_i\},$$

since the probability of A_N is a weighted sum of the p_i's, with the weights given by the (unknown) probabilities that $i_N = i$, $i = 1, 2, \ldots s$.

These bounds may be quite wide, especially if s, the number of food types, is large, or if p_i, the probability of a group being formed in the single resource model including only patches of type i, varies greatly with i. Nevertheless, they may prove useful. For example, if $s = 2$, $p_1 = p_2 = 0.5$, then we know that $0.25 \leq p^* \leq 0.5$, which is sufficient to make some

prediction about behaviour. But if $p_1 = 0.1$, $p_2 = 0.9$, we can only say that $0.09 \leq p^* \leq 0.9$, which gives very little information.

These bounds also ignore the relative frequencies of patches of different types, determined by the q_i's. For example, if patches of type j are rare, i.e. q_j is small, then it is likely that in the completed territory, there will be extra patches of type i, for all $i \neq j$. Hence we expect $\mu_i > m_i'$, for all $i \neq j$, and so the probability of secondaries in the multiple resources model is close to the probability obtained by considering type j alone, i.e. $p^* \approx p_j$. However, no analytic bounds for p^* which take into account the probabilities of different food types, $q_1, q_2, \ldots, q_s$, have been found.

Chapter 4 Spatial Models for Individual Territories

4.1 Introduction

4.1.1 Motivation

The spatial discreteness of the distribution of resources was important in the models in Chapters 2 and 3, but the spatial location of resources has so far been ignored. We will now consider models which do take into account the spatial distribution of resources. In the present chapter we consider the properties of individual territories, and in Chapter 5 we look at interactions between territories.

Setting up a territory is clearly a spatial process. The location of resources affects the cost, in time or energy, of defending a territory or feeding from it, which may in turn affect the amount of resources needed for survival (Covich, 1976; Davies and Houston, 1984; Maynard Smith, 1974). In particular, we wish to consider the effect of the spatial distribution of resources on the models in Chapters 2 and 3. Spatial factors may affect both the setting up of a territory by primary animals and the possibility of secondary animals sharing the territory.

4.1.2 Existing Models

There are a large number of existing models for territory shape and size. Almost all assume a habitat with a uniform distribution of resources. Any one of a wide range of optimisation rules will then lead to a circular shape for a single isolated territory, with the radius of the territory depending on the resource density in the habitat and the particular rule chosen. Possible rules leading to circular territories in a uniform habitat include minimising perimeter for a territory of a given area, minimising the total travelling time from a central point (e.g Getty, 1981) to a territory of a given area, or maximising the weighted difference between area and perimeter (e.g. Hölldobler and Lumsden, 1980). Covich (1976) reviews many such models. Clearly, we would expect circular territories whenever there is a uniform habitat and any kind of isotropic, distance minimising criterion for evaluating a territory.

In reality, however, it is clear that not all animal territories are even approximately circular. Even ignoring the distortions due to neighbouring territories (the subject of Chapter 5), an individual territory will often be far from symmetric. This is illustrated for foxes and badgers, species for which the R.D.H. is thought to be particularly important, by Hersteinsson and

Macdonald (1982) and Kruuk (1978a,b), respectively. Hölldobler and Lumsden (1980) discuss the strikingly non-circular territories of harvester ant colonies, which consist of a nest plus 1-dimensional trails to between 1 and 4 point sources of food. Ewer(1968) generalises this idea, as follows:

> "... if eyesight is relatively unimportant or if cover reduces visibility, the home range may consist essentially of a number of places which are of importance - feeding places, drinking places, resting places, sunning or wallowing spots and so on - linked by a series of pathways. It is thus possible for neighbouring ranges to interpenetrate without significant overlap." (p.64).

Finally, Covich (1976) mentions the idea of a territory which is the union of a small number of discs, representing foraging areas, each centred on a burrow entrance or other refuge from predation.

Despite this evidence that models with non-uniform habitats and non-circular territories are necessary, there are few such existing models. Hölldobler and Lumsden (1980) give a simple model for the harvester ant territories described above, which is interesting, but rather specialised, and not fully explored. Don and Rennolls (1983) also consider habitats with particular points which are important ('nuclei', in their terminology), but their aim is to estimate or model the distribution of the location of an individual animal, given the nuclei it uses, rather than modelling the selection of the set of nuclei. Similarly, Getty (1981) considers a continuous non-uniform habitat, from the point of view of space-use patterns. Noakes and McNicol (1982) represent the territories of juveniles of a particular species of fish (brook charr, *Salvelinus fontinalis*) as cardioid curves, but their model is purely empirical, and seems to have little mechanistic or intuitive appeal.

Thus there are no satisfactory models for the formation of territories in non-uniform habitats. In this chapter we will consider some possible models, concentrating on the case of a single isolated territory. These models will represent generalisations, to a non-uniform habitat, of models leading to circular territories in a uniform habitat.

4.1.3 Mathematical Background

The definitions and properties given in the current section are well-known: see for example Stoyan et al. (1987).

A point process on $\mathbb{R}^d$ is a random variable taking values in a measurable space $(\mathbf{N}, \mathfrak{N})$, where $\mathbf{N}$ is the family of subsets of $\mathbb{R}^d$ which are locally finite, and $\mathfrak{N}$ is a sigma-algebra on $\mathbf{N}$. A set of points ψ in $\mathbb{R}^d$ is defined to be locally finite if any bounded subset of $\mathbb{R}^d$ contains only a finite number of the elements of ψ.

Given a locally finite set ψ, almost all points $x \in \mathbb{R}^d$ have a unique nearest point $n(x)$ in ψ. For a point $y \in \psi$, define the tile of y

$$T(y) = \{x \in \mathbb{R}^d : n(x) = y\}. \tag{4.1.1}$$

If every T is bounded, then each one is a polygon, and

$$\mathfrak{B}(\psi) = \{T(y) : y \in \psi\} \tag{4.1.2}$$

is a tessellation of $\mathbb{R}^d$, known as the Voronoi tessellation, or sometimes, if $d = 2$, the Dirichlet or Thiessen tessellation. If ψ is a realisation of a stationary point process of finite intensity, then with probability 1, every T is bounded, and the Voronoi tessellation exists.

Given the tessellation $\mathfrak{B}(\psi)$ based on a set ψ, we can define the Delaunay triangulation $\mathfrak{D}(\psi)$ which is its dual. Two points $x, y \in \psi$ are said to be connected, or to be neighbours, in the Delaunay triangulation if their Voronoi tiles $T(x), T(y)$ have a boundary segment in common.

The Delaunay triangulation can also be defined directly. Sibson (1978) shows that, except for the arbitrary choices associated with degeneracies, the Delaunay triangulation is the unique triangulation satisfying Lawson's (1972, 1977) criterion, which can be stated as follows (Sibson, 1978).

> "If two triangles in the triangulation share a common edge, they define a quadrilateral with that common edge as a diagonal. If that quadrilateral is strictly convex (that is, each vertex is an extremal point of it) then replacement of the chosen diagonal by the alternative one must not increase the minimum of the six angles in the two triangles making up the quadrilateral, and this must hold for all such strictly convex quadrilaterals."

Less formally, the criterion seeks to divide the convex quadrilaterals occuring in the triangulation in such a way as to make the resulting triangles as close as possible to equilateral. The degeneracies mentioned above are not a problem in the current context: their handling is discussed in detail by Sibson (1978).

4.1.4 Defining a Territory

We shall represent the locations of resources by a point process Ψ in the plane $\mathbb{R}^2$, or a region of the plane. Such a model is the obvious formalisation of the patterns discussed by Don and Rennolls (1983), Ewer (1968), and Hölldobler and Lumsden (1980). It is also a natural spatial version of the patch-based models of Carr and Macdonald (1986), Kruuk and Parish (1982), and Chapter 3 of this thesis. It is natural to assign to each point of the process some quantity of resources, thus forming a marked point process (Stoyan et al., 1987). The

marks will be taken to be independent of the point process, and identically distributed. (In some cases the distribution will be degenerate, so that all marks are equal.) The point process itself we will take to be a homogeneous Poisson process, representing complete spatial randomness. While real resource distributions will generally not be Poisson, the actual pattern depends on the type of resource in question. We will use the Poisson process as a simple, well-understood starting point for spatial models. We will normally be concerned with only a single realisation ψ of Ψ, representing the fixed locations of resource patches. The quantity of resources available at a point may, however, vary over time, i.e. we may interpret the mark at a point as the mean of some process over time.

Given that resources are described by a point process Ψ, a territory will include some subset of the points of the process. We think of the territory as a region of the plane containing a particular subset τ of ψ, the realisation of Ψ. For the models in this chapter, we only consider a single territory, so the boundary between territories, and hence the total region making up a territory, is ill-defined. Instead, as a summary of the geometry of a territory containing a given subset of ψ, we use the convex hull of that subset. We can think of the convex hull as corresponding to the "core area" of a territory (Ewer, 1968). In the absence of information on neighbouring territories, it gives an indication of the shape of the territory of interest. The use of the convex hull implies that territories, or at least their core areas, will tend to be convex. This seems to be a reasonable assumption for species likely to be of interest, such as foxes (Hersteinsson and Macdonald, 1982) and badgers (Kruuk, 1978b), and in keeping with most existing models, though it does not apply to the Hölldobler and Lumsden (1980) model of harvester ant territories.

Given that we represent a territory by a set of points and their convex hull, we must determine how these points will be chosen. The process can be modelled in two steps.

4.1.5 The Quality Of A Territory.

There has been considerable discussion in the biological literature of models of the net value of a territory, usually aimed at determining territory size, and not taking into account shape. Many of these models are discussed by Maynard Smith (1974), Covich (1976), Schoener (1983) and Davies and Houston (1984). Other specific models are given by Hölldobler and Lumsden (1980) and Jones and Krummel (1985). The important factors in such models are usually the area of the territory, representing total resources in a uniform habitat, and some measure of the cost of occupying the territory, such as the defence cost or travelling time for feeding. In our models, resources will be available at specific points rather than uniformly, so we will consider the total of the marks of resource points in the territory instead of the area of the territory. As a measure of the cost of holding a territory, we will use the perimeter of the core area. It is an intuitively reasonable criterion, and is easily defined

mathematically. Some existing models use perimeter as a measure of cost (e.g. Hölldobler and Lumsden, 1980), and in many others cost can be thought of as an increasing function of perimeter. The perimeter is related to the effort required for defence and, less directly, to travel times within the territory. Most existing models could be generalised to a non-uniform habitat by describing a territory in terms of its total resource content (sum of marks) and an increasing, quadratic function of perimeter.

We will generally use a definition of the value of a territory which is a simple, natural extension of the criterion used in Section 3.2.1. A territory τ with sum of marks $\mu(\tau)$ and perimeter $p(\tau)$ has quality

$$q(\tau) = \begin{cases} -p(\tau), & \mu(\tau) \geq m \\ -\infty, & \text{otherwise.} \end{cases} \tag{4.1.3}$$

That is, the best territory is the one with smallest perimeter out of those with $\mu(\tau) \geq m$, and territories with $\mu(\tau) < m$ are regarded as untenable.

4.1.6 Searching for a Territory

Assuming that we can measure territory quality as above, we want to model the process of choosing a territory. Recall that for the purposes of this chapter, we consider only a single territory. Nevertheless, it is unreasonable to assume that a primary pair would choose the best possible territory from a whole habitat. Such a choice would depend largely on the size of habitat, and be undefined in the limiting case of an infinite habitat. Furthermore, in practice animals will localise their search for a territory (P. J. Bacon, pers. comm.). We can model the local nature of the search for an acceptable territory in two ways : by constraining the territory to contain particular resource points; or by choosing a locally optimal territory, in some sense.

4.1.6.1 Constrained Choice of Points

If we are constraining the selection of points for the territory, the simplest, most natural way to localise the search is to pick an initial resource point at random, and allow only territories including that point to be chosen. Since we only consider stationary processes, this constraint is equivalent to conditioning on the existence of a point at the origin, and considering only territories which include the origin. In a model based on the Poisson process, this constraint can easily be incorporated, since such conditioning does not affect the distribution of points, except of course at the origin itself (see e.g. Diggle, 1983).

4.1.6.2 Local optimality

If the search is for a locally optimal territory, we need to define some spatial structure on the set of possible territories. Firstly, we define two resource points $x, y \in \psi$ to be neighbours if they are connected in the Delaunay triangulation $\mathfrak{D}(\psi)$ based on ψ. Given a set of points $\tau \subset \psi$, we say that $x \in \psi$ is adjacent to τ if there exists some $y \in \tau$ such that x and y are neighbours. Finally, two sets $\tau, \upsilon \subset \psi$ are said to be neighbours if one of the following holds:

(i) $\upsilon = \tau \setminus \{x\}$;

(ii) $\upsilon = \tau \cup \{y\}$;

(iii) $\upsilon = \tau \cup \{y\} \setminus \{x\}$;

where $x \in \tau$, and y is adjacent to τ.

Condition (iii) seems slightly counter-intuitive, but is deliberately chosen to to allow the possibility of two distinct sets with the same number of elements being neighbours. Note that the definition allows a wide range of configurations to be reached, stepwise, from an initial singleton set. In particular, even if τ is connected within $\mathfrak{D}(\psi)$, a neighbouring set need not be. The territory defined by a set of points τ is then said to be locally optimal if it is at least as good, according to some given criterion, as the territory defined by any neighbouring set υ.

4.1.7 Overview of the chapter

Having considered existing models and reviewed the mathematics which we will need, we next look at some improved models. In Sections 4.2, 4.3 and 4.4, we explore three specific models in which an initial random point must be retained in a territory. In Section 4.5, we consider a model based on selecting a locally optimal territory. These four models are compared in Section 4.6, while numerical and computational aspects are discussed in Section 4.7.

4.2 A Model with Simultaneous Choice of Points

The first way of localising the search for a territory, described in Section 4.1.6.1, involves selecting the first point x_0 of the territory at random. Clearly, to obtain the best territory, subject to the constraint that x_0 must be retained, the primary pair need to choose a territory τ such that

$$q(\tau) = \max\{q(\upsilon) : x_0 \in \upsilon \subset \psi\}.$$

Thus the remaining points in $\tau \setminus \{x_0\}$ should be chosen simultaneously, to maximise $q(\tau)$.

While such a choice will, by definition, give the best territory subject to the given constraints, it may not be a good model of the actual selection made by animals of a particular species. Maximising $q(\tau)$ involves comparing territories which are of very different

configurations, and in fact may have only the point x_0 in common. Rules which are more *ad hoc* may correspond more closely to reality. Nevertheless, the optimal selection is of interest, as an upper bound, a standard of comparison and perhaps an approximation for other rules.

Unfortunately, the properties of the optimum territory seem to be mathematically intractable. Even with the simplest measure of the quality of a territory, as defined in equation (4.1.3), and with all points $x \in \psi$ having the same mark, it does not seem possible to write down the distribution of $\min\{q(\upsilon)\}$. The only exception is if n, the number of points required to make $\mu(\tau) \geq m$, is 1 or 2, in which case the process is trivial, and is subsumed in the simpler model of Section 4.4.

In other cases, it is necessary to simulate the process. The program and techniques used for the simulation are discussed in Section 4.7 below: in the current section, we concentrate on describing the simulations undertaken, and the results obtained. It should be noted, however, that simulation of the current model is computationally expensive, which limits the trials that can be carried out. In particular, only the simplest measure of territory quality, and only territories requiring small numbers of resource points, are considered. Note that we have used a point process with unit intensity i.e. $\lambda = 1$ throughout: different values of λ would simply have the effect of scaling the values of the perimeter by a factor of $\lambda^{-\frac{1}{2}}$.

Two models for the underlying pattern of resources are used. In the first, the marks representing the mean yields of the resource points are all assumed to be equal, and the common value taken to be 1, without loss of generality. Since we are using the rule defined in equation (4.1.3), a territory in such a habitat will contain a fixed number n of resource points, with

$$n = \lceil m \rceil,$$

and we can take $m \in \mathbb{N}$ without loss of generality. Cases with constant marks are of theoretical interest, since there is no 'confounding' between the point process and the process of marks in determining the perimeters of territories. Results can also be compared with the few analytic results available, for this and other models.

The second model for the marks assumes that they all have the exponential distribution with unit mean, independently of each other and of the point process itself. Clearly other distributions could have been used, but the exponential was chosen for ease of comparison with the non-spatial models of Chapter 3, and to give a contrast with the constant-marks model. With random marks, m can take any positive real value. The values of m used in the simulations consist of integers, for comparison with the constant-marks case, along with values obtained from Section 3.2.3 corresponding to particular cases already explored in non-spatial models. Table 4.1 shows the details of the runs selected, and the results obtained. Each case is defined by the choice of a value of m, and of a distribution of marks (constant or

exponential). For each case, the table gives estimates and standard errors for the first two moments EP and EP^2 of the perimeter of a territory, the effective sample size used, and the c.p.u. time needed for each run (see Section 4.7 for details), and for some cases with random marks, estimates and standard errors for the probability of group formation (i.e. $\Pr(\mu \geq m')$) for one or more values of m'. Because the exponential distribution has been used for random marks, the probabilities of group formation can be directly compared with the corresponding values in the column labelled '$k = 1$' in Table 3.6. In addition to the simulation runs, two other cases, both with constant marks, are included in the table for the purpose of comparison: the case $m = 1$, for which the perimeter is identically zero, and the case $m = 2$, for which precise results are obtained in Section 4.4 below.

Table 4.1

Numerical Results for Simultaneous Choice of Points
"Const" indicates marks which are constant;
"Exp" indicates marks which are exponentially distributed

Case	Marks	m	EP	S.E.	EP^2	S.E.	Sample	C.p.u.	m'	p	S.E.
A	Const	1	0.000	-	0.000	-	-	-	-	-	-
B	Const	2	1.000	-	1.274	-	-	-	-	-	-
C	Const	3	1.859	0.040	3.932	0.158	300	0.91	-	-	-
D	Const	4	2.645	0.049	7.641	0.270	270	2.8	-	-	-
E	Const	5	3.448	0.080	12.460	0.555	175	31	-	-	-
H	Exp	1	0.805	0.046	1.411	0.116	360	0.67	1.5	0.612	0.026
J	Exp	2	1.449	0.074	3.110	0.241	190	1.6	-	-	-
Q	Exp	2.18	1.593	0.097	3.644	0.338	120	2.0	2.42	0.632	0.044
"	"	"	"	"	"	"	"	"	2.74	0.420	0.046
"	"	"	"	"	"	"	"	"	3.38	0.260	0.040

The main conclusions from Table 4.1 are in accordance with intuition. Both EP and EP^2 increase with m, and the net effect is that the coefficient of variation of P decreases with

m. The coefficient of variation is higher in the case of exponentially distributed marks than with constant marks. From comparison with Table 3.6, as mentioned above, it can be seen that the current model gives values of p that are approximately equal to (case 10 in Table 3.6 versus case H in Table 4.1; case 1 in Table 3.6 versus case Q, $m' = 2.74$, in Table 4.1) or significantly lower than (cases 4 and 5 in Table 3.6 versus case Q, $m' = 2.42$ and 3.38 respectively, Table 4.1; significant at the 5% level) those in the non-spatial model.

Table 4.1 also shows that the c.p.u. time needed to simulate each territory increases rapidly with m. For this reason, only small values of m have been used.

4.3 Sequential Choice of Points

A natural alternative to the simultaneous choice of the points of $\tau \setminus \{x_0\}$ is to choose those points sequentially. However, the naive procedure of adding, at each step, the point x giving the highest value of $q(.)$, will not in general give useful results. For instance, consider the simple rule described in equation (4.1.3). Write $r_0, r_1, \ldots$ for the marks of $x_0, x_1, \ldots$. Then unless there exists some x_1 such that $r_0 + r_1 \geq m$, all choices of x_1 give $q\{x_0, x_1\} = q\{x_0\} = -\infty$. We could modify $q(.)$ to try to overcome this problem, though the modifications necessary would depend on the nature of the particular function $q(.)$ with which we started. Instead, for the remainder of Section 4.3, we concentrate on a particular sequential procedure, aimed at giving high, though unavoidably sub-optimal, values of the simple function $q(.)$ defined in equation (4.1.3). The procedure does not explicitly use $q(.)$, but is tailored to the particular structure of $q(.)$.

The procedure is simply described as follows. The initial point x_0 is determined randomly, as before (step 0). At the kth step, $k \geq 1$, the point x_k is added, such that

$$p\{x_0, x_1, \ldots, x_k\} = \min\{p\{x_0, \ldots, x_{k-1}, x\} : x \in \psi \setminus \{x_0, \ldots, x_{k-1}\}\}. \qquad (4.3.1)$$

The procedure stops when $\sum_{j=0}^{k} r_j \geq m$. Although the above procedure will in general give a lower value for $q(\tau)$ than the simultaneous choice of $\tau \setminus \{x_0\}$, the hope is that by minimising perimeter at each stage, we will obtain something reasonably close to the optimum.

We now wish to derive some of the properties of the territory obtained by the above sequential procedure. The final number of points N will be a random variable which is independent of the locations of the points. In fact the process determining N, depending only on the marks of the resource points, will be identical to the corresponding process for the non-spatial model in Chapter 3. Thus, the total amount of resources μ in the territory, and the probability p of forming groups, will be the same as in Chapter 3. However, the distribution of the perimeter $p\{x_0, \ldots, x_{N-1}\}$ is difficult to write down.

If we have constant marks, however, then N will also be a constant, and we can make some analytic progress. If n, the number required to make $\mu(\tau) \geq m$, is 1 or 2, then as with the model of the previous section, the process is trivial, and is subsumed in the simpler model of Section 4.4. If $n=3$, however, the selection of the final point, x_2, is non-trivial. The first point, selected at random, is x_0. The next point x_1 is clearly the nearest neighbour of x_0. We write R for the random distance from x_0 to x_1. The distribution of R is given by

$$\Pr(R \leq r) = F_R(r) = 1-e^{-\lambda\pi r^2} \quad (r > 0), \tag{4.3.2}$$

where λ is the intensity of the resource point process Ψ (see e.g. Diggle, 1983, Section 3.3), and thus we have

$$f_R(r) = 2\lambda\pi r e^{-\lambda\pi r^2} \quad (r > 0). \tag{4.3.3}$$

To investigate the location of the next point, x_2, we will redefine the problem as follows. We condition on $R = r$, and transform the plane isometrically so that $x_0 = (0,0)$ and $x_1 = (r,0)$. We write (x,y) for x_2. Clearly since x_1 is the element of ψ nearest to x_0, no other point of ψ can lie in the disc centred at (0,0), with radius r. Subject to that condition, the remaining points of ψ are independent of x_0 and x_1.

We want to find that point x_2 which gives the smallest value of the perimeter of the convex hull of x_0, x_1, x_2. But since we have only three points, the perimeter is

$$d(x_0,x_1)+d(x_0,x_2)+d(x_1,x_2),$$

where $d(.,.)$ simply denotes the distance between two points. We know $d(x_0,x_1) = r$, so we wish to minimise

$$\begin{aligned} & d(x_0,x_2)+d(x_1,x_2) \\ & = \sqrt{y^2+x^2}+\sqrt{y^2+(r-x)^2} \end{aligned} \tag{4.3.4}$$

which we denote by the random variable C, say. Note that the points (x,y) with a given value of C lie on an ellipse, with foci (0,0) and $(r,0)$. Thus

$$\Pr(C \leq c \,|\, R = r) = 1-\Pr(\Psi(E_c \setminus D) = 0),$$

where E_c is the region contained by the ellipse

$$\sqrt{y^2+x^2}+\sqrt{y^2+(r-x)^2} = c \tag{4.3.5}$$

and D is the disc of radius r centred at the origin. Thus

$$\Pr(C \leq c \,|\, R = r) = 1-e^{-\lambda A_c}, \tag{4.3.6}$$

where A_c is the area of $E_c \setminus D$.

The calculation of A_c is straightforward, and depends only on $\frac{c}{r}$. After considerable calculation, we can write

$$A_c = r^2 a\left(\frac{c}{r}\right),$$

and hence

$$\Pr(C \leqslant c \,|\, R = r) = F_{C|R}(c\,|\,r) = 1 - e^{-\lambda r^2 a\left(\frac{c}{r}\right)}, \tag{4.3.7}$$

where

$$a(s) = \begin{cases} 0 & s \leqslant 1 \\ -\frac{\pi}{2} + v\sqrt{1-v^2} + \sin^{-1} v \\ +\frac{1}{4}s\sqrt{s^2-1}\left(\frac{\pi}{2} - w\sqrt{1-w^2} - \sin^{-1} w\right) & 1 < s < 3 \\ \pi\left(\frac{1}{4}s\sqrt{s^2-1} - 1\right) & s \geqslant 3 \end{cases} \tag{4.3.8}$$

and

$$v = \tfrac{1}{2}(1 + 2s - s^2),$$

$$w = 2 - s.$$

Thus

$$f_{C|R}(c\,|\,r) = \lambda r a'\left(\frac{c}{r}\right) e^{-\lambda r^2 a\left(\frac{c}{r}\right)}, \tag{4.3.9}$$

where

$$a'(s) = \begin{cases} 0 & s \leqslant 1 \\ 2(1-s)\sqrt{1-v^2} + \frac{1}{2}s\sqrt{s^2-1}\sqrt{1-w^2} \\ +\frac{1}{4}\left(\frac{\pi}{2} - w\sqrt{1-w^2} - \sin^{-1} w\right)(2s^2-1)/\sqrt{s^2-1} & 1 < s < 3 \\ \frac{\pi}{4}(2s^2-1)/\sqrt{s^2-1} & s \geqslant 3 \end{cases} \tag{4.3.10}$$

Thus when $n = 3$, we know the joint distribution of R and C, and thus the distribution of the total perimeter

$$P = R + C. \tag{4.3.11}$$

In particular, for comparison with other models, we calculate

$$EP = \int_{u>0} \int_{\frac{u}{2}>r>0} u f_R(r) f_{C|R}(u-r|r)\, dr\, du$$

$$= \int_{u>0} \int_{\frac{u}{2}>r>0} 2\lambda^2 \pi r^2 u a'\left(\frac{u-r}{r}\right) e^{-\lambda r^2(\pi + a\left(\frac{u-r}{r}\right))}\, dr\, du, \qquad (4.3.12)$$

and similarly,

$$EP^2 = \int_{u>0} \int_{\frac{u}{2}>r>0} 2\lambda^2 \pi r^2 u^2 a'\left(\frac{u-r}{r}\right) e^{-\lambda r^2(\pi + a\left(\frac{u-r}{r}\right))}\, dr\, du. \qquad (4.3.13)$$

The integrals above can be evaluated numerically: the values obtained when $\lambda = 1$ are included in Table 4.2 below.

If we have constant marks with $n > 3$, or random marks, the moments of the perimeter in the sequential model can only be determined by simulation. The organisation of these simulations is similar to that used in Section 4.2. We take $\lambda = 1$ throughout, we consider the cases of constant marks and exponential marks, and we consider a range of values of m. Note however that the current model is less costly to simulate than the model of Section 4.2, so a wider range of cases can be considered. As in Table 4.1, we give estimates and standard errors for the first two moments of the perimeter, plus the effective sample size of the simulation and the c.p.u. time required. As well as results from simulation, the table includes three other cases, with constant marks: the trivial case $m = 1$; the case $m = 2$, for which results are obtained in Section 4.4 below; and the case $m = 3$, for which EP and EP^2 have been obtained more directly, from equations 4.3.12 and 4.3.13. Table 4.2 also includes values of $p = \Pr(\mu \geq m')$, for selected cases. As noted above, these probabilities are the same in the current model as in the non-spatial model in Chapter 3. The values of p are thus the same as in Table 3.6, and are included in Table 4.2 for ease of comparison with the other models in the current chapter.

Not included in the table are simulation results for case C, carried out to check the agreement between simulation and exact results. These simulations give $EP = 1.920$ (S.E. 0.043) and $EP^2 = 4.245$ (S.E. 0.182), so they are consistent with the exact results. The effective sample size for these simulations is 300, at a c.p.u. time of 1.1s per territory.

The definition of the sequential model means that the values of EP and EP^2 it gives are equal to the corresponding values for the simultaneous model (Section 4.2) in cases A and B in the table, and not less than the values for the simultaneous model in other cases. Intuitively, we would expect the values for the sequential model to be strictly greater (except in cases A and B), and the simulations provide evidence for this (statistically significant differences for EP in cases C, H, J and Q, at the $2\frac{1}{2}\%$ level). These differences appear to be quite

Table 4.2

Numerical Results for the Sequential Choice of Points

"Const" indicates marks which are constant;

"Exp" indicates marks which are exponentially distributed

Case	Marks	m	EP	S.E.	EP^2	S.E.	Sample	C.p.u.	m'	p
A	Const	1	0.000	-	0.000	-	-	-	-	-
B	Const	2	1.000	-	1.274	-	-	-	-	-
C	Const	3	1.940	-	4.283	-	-	-	-	-
D	Const	4	2.774	0.055	8.510	0.328	275	1.4	-	-
E	Const	5	3.526	0.065	13.484	0.476	250	1.8	-	-
H	Exp	1	0.976	0.049	1.951	0.153	340	0.39	1.5	0.607
J	Exp	2	2.001	0.093	6.514	0.497	280	0.72	-	-
K	Exp	3	2.483	0.100	8.231	0.629	210	1.5	-	-
L	Exp	4	3.517	0.109	15.564	0.901	270	3.0	-	-
Q	Exp	2.18	2.012	0.090	6.226	0.449	270	0.76	2.42	0.791
"	"	"	"	"	"	"	"	"	2.74	0.571
"	"	"	"	"	"	"	"	"	3.38	0.302

small in cases with constant marks, but larger in cases with exponential marks (2%~5% and 21%~38% respectively, based on the ratios of values of EP in Tables 4.1 and 4.2).

The c.p.u. time per simulated territory increases with m, but the rate of increase is relatively low compared with the rate in the simultaneous model.

4.4 A Nearest-Neighbour Model.

4.4.1 The Model

The sequential model described above can be further simplified, and hopefully made more tractable, by replacing the minimisation of perimeter at each stage with a simpler rule. One possibility, which we consider in some detail, is to choose the nearest neighbours of x_0 to

make up τ, giving the following procedure.

Firstly, x_0 is chosen at random, as before. We will refer to x_0 as the centre of the territory. Secondly, the point in $\psi \setminus x_0$ which is nearest to x_0 is chosen as the next resource point in the territory. Thirdly, continue adding resource points, in order of their proximity to the centre, until there are sufficient resources in the territory. Note that under this rule, the number of points N needed to supply a given amount of resources is again independent of the locations of the points, provided the marks are independent of the point locations. Thus, without loss of generality, we can condition on some value, $N=n$, say. Note that all resource points in the territory lie in the closed disc D_{n-1} centred on the first point chosen, with radius equal to the $(n-1)$th nearest neighbour distance from that centre, and all other resource points lie outside the open disc D_n with the same centre, and radius given by the nth nearest neighbour distance from the centre.

4.4.2 The Unit Disc

Since the resources form a Poisson point process, the n points in the territory can be thought of as a uniform random sample of $n-1$ points on the disc D_n plus a single point at the centre of the disc, independently of the location and radius of D_n. So it is sufficient to consider the characteristics, and in particular the perimeter of the convex hull, of such a set of points on the unit disc.

Let $\tilde{P}_n$ be the perimeter of the convex hull H_n of the set of n points consisting of the origin plus a uniform random sample of $n-1$ points on the open disc with radius 1 centred at the origin. The distribution of $\tilde{P}_n$ seems intractable, but it is possible to make progress in evaluating its moments.

Efron (1965) considers the related case in which all n points form a random uniform sample, and obtains an integral expression for the expected perimeter. We adopt a similar approach. For any line in the plane, let $(p,\theta), 0 \leq \theta < \pi$, be the signed length and direction of the normal from the line to the origin. Within the infinite strip defined by $-\infty < p < \infty, 0 \leq \theta < \pi$, consider the region of pairs (p,θ) corresponding to lines which intersect H_n, and let J_n be the area of that region. Then it is well known (e.g. Kendall and Moran, 1963, p. 58) that $\tilde{P}_n = J_n$.

A given line (p,θ) will intersect H_n unless all the $n-1$ random points are on the same side of (p,θ) as the origin. Thus the probability that (p,θ) intersects H_n is

$$1-\Lambda^{n-1}(p,\theta) \tag{4.4.1}$$

where $\Lambda(p,\theta)$ is the probability that a given random point is on the same side of (p,θ) as the origin. Since the distribution of the points is isotropic, we can write $\Lambda(p)$ for $\Lambda(p,\theta)$, with

$\Lambda(p) = \Lambda(-p)$, and since all points lie on the unit disc we have $\Lambda(p) = 1$ for all $p \geq 1$. Then we can write

$$\begin{aligned} E\tilde{P}_n &= EJ_n \\ &= \int_{-\infty}^{\infty}\int_0^{\pi} \{1-\Lambda^{n-1}(p,\theta)\}d\theta dp \\ &= 2\pi\int_0^1 \{1-\Lambda^{n-1}(p)\}dp. \end{aligned} \tag{4.4.2}$$

The probability $\Lambda(p)$ is just the cumulative distribution function of the marginal distribution along any diameter of the uniform distribution on the unit disc, and is given by

$$\begin{aligned} \Lambda(p) &= \int_{-1}^{p} \frac{2}{\pi}\sqrt{1-q^2}dq, \quad 0 \leq p \leq 1 \\ &= \frac{1}{2} + \frac{1}{\pi}(p\sqrt{1-p^2} + \sin^{-1}p), \quad 0 \leq p \leq 1, \end{aligned} \tag{4.4.3}$$

(c.f. Efron (1965), equation 7.8). Thus we have

$$\begin{aligned} E\tilde{P}_n &= 2\pi\int_0^1 1-\{\frac{1}{2} + \frac{1}{\pi}(p\sqrt{1-p^2} + \sin^{-1}p)\}^{n-1}dp \\ &= 2\pi - 2\pi^{2-n}\int_0^1 (\frac{\pi}{2} + p\sqrt{1-p^2} + \sin^{-1}p))^{n-1}dp. \end{aligned} \tag{4.4.4}$$

Substituting $x = \frac{\pi}{2} + \sin^{-1}p$ gives

$$E\tilde{P}_n = 2\pi[1-\pi^{1-n}\int_{\frac{\pi}{2}}^{\pi} (x-\sin x\cos x)^{n-1}\sin x dx] \tag{4.4.5}$$

To evaluate the integral in (4.4.5), we proceed as follows. Define

$$I(n,l) = \int_{\frac{\pi}{2}}^{\pi} (x-\sin x\cos x)^n \sin^{2l+1}x dx \tag{4.4.6}$$

so that we require an expression for $I(n,0)$. Using results from Gradshteyn and Ryzhik (1980), we can write for $n \geq 2$, after considerable algebra,

$$I(n,l) = A(n,l) + 4n(n-1)\sum_{m=0}^{l} \frac{a(l,m)}{2m+3} I(n-2,m+2) \tag{4.4.7}$$

where

$$A(n,l) = \pi^n b(l) + 2n(\frac{\pi}{2})^{n-l}\sum_{m=0}^{l} \frac{a(l,m)}{2m+3},$$

$$a(l,m) = \frac{-2^{l-m}}{2l+1} \cdot \frac{l(l-1)\ldots(m+1)}{(2l-1)(2l-3)\ldots(2m+1)},$$

$$b(l) = \frac{2^l l!}{(2l+1)(2l-1)\ldots 1}.$$

In addition, we have

$$I(0,l) = A(0,l) = b(l), \tag{4.4.8}$$

$$I(1,l) = A(1,l), \tag{4.4.9}$$

by direct calculation. Using equations (4.4.5) to (4.4.9), we can get exact values for $E\tilde{P}_n$. Further details are given, for completeness, in Section 4.4.5.

Similarly, we can calculate

$$\begin{aligned} E\tilde{P}_n^2 &= EJ_n^2 \\ &= \iint_{p,q \geq 0} \iint_{\theta,\phi \in [0,2\pi)} \Pr(\text{Both } (p,\theta) \text{ and } (q,\phi) \text{ intersect } H_n)\, d\theta\, d\phi\, dp\, dq \\ &= 4\pi \iint_{p,q \in [0,1]} \int_{\psi \in [0,2\pi)} f(p,q,\psi) d\psi\, dp\, dq \end{aligned} \tag{4.4.10}$$

by symmetry (and the fact that H_n lies within the unit disc), where $f(p,q,\psi)$ is the probability that any two lines (p,θ) and (q,ϕ), satisfying $\min\{|\theta-\phi|, 2\pi-|\theta-\phi|\} = \psi$, both intersect H_n. By considering the locations of the $n-1$ random points relative to the two lines, we eventually obtain

$$f(p,q,\psi) = 1 - A^{n-1}(p) - A^{n-1}(q) - K^{n-1}(p,q,\psi), \tag{4.4.11}$$

where $A(.)$ is defined as in equation 4.4.3,

$$K(p,q,\psi) = \begin{cases} A(\min\{p,q\}) & 0 \leq \psi \leq |\omega-\chi| \\ A(\cos\rho) + \dfrac{2\pi^{-1}\sin^2\rho}{\cot(\rho-\chi)+\cot(\rho-\omega)} & |\omega-\chi| \leq \psi \leq \omega+\chi \\ A(p)+A(q)-1 & \omega+\chi \leq \psi \leq \pi \end{cases} \tag{4.4.12}$$

and

$$\chi = \cos^{-1}p,$$

$$\omega = \cos^{-1}q,$$

$$\rho = \frac{\psi+\chi+\omega}{2}.$$

The integral (4.4.10) can then be evaluated numerically.

4.4.3 Using the Results from the Unit Disc.

Having obtained values for $E\tilde{P}_n$ and $E\tilde{P}_n^2$, we need to relate them to the perimeter of the core are of the model described at the beginning of Section 4.4.

Let $\tilde{R}_0,\ldots,\tilde{R}_{n-1}$ denote the ranked distances from the origin to each of the n points in the unit disc described in Section 4.4.2. Then we have

$$\tilde{R}_0 = 0,$$

by definition, and we can write

$$\tilde{R}_j^2 = U_{(j)}, \quad j = 1,\ldots,n-1, \tag{4.4.13}$$

where $U_{(1)} \leq U_{(2)} \leq \ldots \leq U_{(n-1)}$ are the order statistics of a sample of $n-1$ independent and identically distributed uniform random variables on $[0,1]$. We let

$$\tilde{A}_j = \pi\tilde{R}_j^2, \quad j = 0,\ldots,n-1. \tag{4.4.14}$$

Define $\mathfrak{P}_n(.,.)$ to be the deterministic function which takes as arguments a vector of areas, $a = a_0,\ldots, a_{n-1}$, and a vector of angles $\theta = \theta_0,\ldots, \theta_{n-1}$, and gives the perimeter of the convex hull of the points having polar co-ordinates $(\sqrt{\frac{a_0}{\pi}},\theta_0),\ldots,(\sqrt{\frac{a_{n-1}}{\pi}},\theta_{n-1})$. We will always take $\theta_0,\ldots,\theta_{n-1}$ to be independent and identically distributed random variables on $[0,2\pi)$, and so we define $\mathfrak{P}_n(.)$, with the second argument omitted, to be the corresponding random function. Thus we can write

$$\begin{aligned}\tilde{P}_n &= \mathfrak{P}_n(\tilde{A}),\\ &= \mathfrak{P}_n(\pi(0, U_{(1)},\ldots, U_{(n-1)})),\end{aligned} \tag{4.4.15}$$

using relations (4.4.13) and (4.4.14).

Now consider the model in Section 4.4.1, and let R_j, $j = 0,\ldots, n-1,n$ be the distance from the initial point, or centre, of the territory to the jth nearest neighbour of that point (so that $R_0 \equiv 0$). Let

$$A_j = \pi R_j^2. \tag{4.4.16}$$

From the basic properties of the Poisson process (see e.g. Diggle, 1983),

$$A_n \sim \Gamma(\lambda^{-1}, n), \tag{4.4.17}$$

where λ is the intensity of the Poisson process, and Γ denotes the gamma distribution. Furthermore, from the independence properties of the Poisson process, we can write

$$A_j = U_{(j)}A_n, \quad j = 1,\ldots, n-1, \tag{4.4.18}$$

where $U_{(1)},\ldots, U_{(n-1)}$ are as in equation (4.4.13) above, and are independent of A_n. Clearly, $A_0 \equiv 0$. Thus the perimeter of the core area of the territory in the current model is

$$\begin{aligned} P_n &= \mathfrak{P}_n(A) \\ &= \mathfrak{P}_n(A_n(0, U_{(1)},\ldots, U_{(n-1)})). \end{aligned} \tag{4.4.19}$$

Now for any scalar α,

$$\mathfrak{P}_n(\alpha a) = \alpha^{\frac{1}{2}}\mathfrak{P}_n(a), \tag{4.4.20}$$

so

$$\begin{aligned} P_n &= A_n^{\frac{1}{2}}\mathfrak{P}_n((0, U_{(1)},\ldots, U_{(n-1)})) \\ &= \left(\frac{A_n}{\pi}\right)^{\frac{1}{2}}\mathfrak{P}_n(\pi(0, U_{(1)},\ldots, U_{(n-1)})) \\ &= \left(\frac{A_n}{\pi}\right)^{\frac{1}{2}}\tilde{P}_n \\ &= R_n\tilde{P}_n. \end{aligned} \tag{4.4.21}$$

We can use (4.4.21) to relate the first two moments of P_n to the corresponding moments of $\tilde{P}_n$. From (4.4.17),

$$\begin{aligned} & A_n \sim \Gamma(\lambda^{-1}, n), \\ \Rightarrow\quad & 2\lambda A_n \sim \chi^2_{2n} \\ \Rightarrow\quad & (2\pi\lambda)^{\frac{1}{2}}R_n \sim \chi_{2n}, \end{aligned} \tag{4.4.22}$$

where χ_ν represents the chi-distribution with ν degrees of freedom, i.e. the distribution of the positive square root of a random variable with the chi-square distribution with ν degrees of freedom (see e.g. Johnson and Kotz (1970), Chapter 17, Section 8.3). Hence from e.g. Johnson and Welch (1939) (or from Johnson and Kotz (1970), Chapter 17, equation (10); but note that equation (63) of the same chapter gives an incorrect version of the same result), we have

$$\mathrm{E}[(2\pi\lambda)^{\frac{1}{2}}R_n] = \frac{\Gamma(\frac{1}{2}(2n+1))}{\Gamma(n)}\cdot 2^{\frac{1}{2}}.$$

Hence

$$\begin{aligned} \mathrm{E}R_n &= \frac{\Gamma(\frac{1}{2}(2n+1))}{(\pi\lambda)^{\frac{1}{2}}\Gamma(n)} \\ &= \frac{\frac{1}{2}\times\frac{3}{2}\times\ldots\times\frac{2n-1}{2}}{\lambda^{\frac{1}{2}}(n-1)!}. \end{aligned} \tag{4.4.23}$$

So from (4.4.21), by independence,

$$\begin{aligned} EP_n &= ER_n E\tilde{P}_n \\ &= \frac{\frac{1}{2}\times\frac{3}{2}\times\ldots\times\frac{2n-1}{2}}{\lambda^{\frac{1}{2}}(n-1)!}\cdot 2\pi[1-\pi^{1-n}I(n-1,0)], \end{aligned} \tag{4.4.24}$$

from Section 4.4.2. Also

$$EP_n^2 = ER_n^2\, E\tilde{P}_n^2, \tag{4.4.25}$$

where $E\tilde{P}_n^2$ can be obtained numerically from equation 4.4.10.

4.4.4 Numerical Results

We collect here numerical results for the nearest-neighbour model, considering essentially the same cases as in Sections 4.2 and 4.3. In cases with constant marks, results can be taken directly from Section 4.4.3. In cases with random marks, we can think of the results in Section 4.4.3 as being conditional on a particular value n of the random variable N, representing the number of resource points in the territory. The distribution of N is known, from Chapter 3: in the case of exponential marks,

$$N-1 \sim \text{Poisson}(m).$$

Thus we can obtain the overall distribution of the perimeter from the conditional distributions. As with the model in Section 4.3, the values of

$$p = \Pr(\mu \geq m')$$

are the same for the current model as for the non-spatial model in Chapter 3. They are repeated here for convenience.

Table 4.3 gives values for EP and EP^2, for constant and exponential marks and for various values of m, and in some cases values for p. All values are exact (i.e. a closed form is available) or have been determined by numerical integration with much greater precision than the simulation results in other sections of this chapter.

Intuitively, the nearest-neighbour model should, from its definition, give expected perimeters strictly greater than those given by the sequential model (Section 4.3), except in cases A and B when the two models are equivalent. Table 4.3 shows that the intuitive relationship does hold in case C, for which exact results are available for both models, and appears to hold in most other cases, when the exact results in Table 4.3 are compared with the simulations in Table 4.2 (statistically significant differences in cases D, E, K, L, Q and R, at the $2\frac{1}{2}\%$ level). The differences, based on ratios of the values for EP in the tables, vary considerably (from 3% to 22%); the only case known exactly, case C, gives an increase of 8% in EP going from the sequential to the nearest-neighbour model.

Table 4.3

Numerical Results for the Nearest-Neighbour Model

Case	Marks	m	EP	EP^2	m'	p
A	Constant	1	0.000	0.000	-	-
B	Constant	2	1.000	1.274	-	-
C	Constant	3	2.099	5.185	-	-
D	Constant	4	3.144	11.105	-	-
E	Constant	5	4.104	18.399	-	-
F	Constant	10	7.871	64.302	-	-
G	Constant	20	12.951	170.35	-	-
H	Exponential	1	1.028	2.489	1.5	0.607
J	Exponential	2	2.056	7.015	-	-
K	Exponential	3	3.038	13.067	-	-
L	Exponential	4	3.958	20.246	-	-
M	Exponential	5	4.814	28.259	-	-
N	Exponential	10	8.323	75.209	-	-
Q	Exponential	2.18	2.237	8.005	2.42	0.791
"	"	"	"	"	2.74	0.571
"	"	"	"	"	3.38	0.302
R	Exponential	4.33	4.247	22.810	4.78	0.638

4.4.5 Calculating the Expected Value of $\tilde{P}_n$.

For very small n, we can calculate E$\tilde{P}_n$ easily from equations (4.4.5) to (4.4.9). However, the complexity of the expression for E$\tilde{P}_n$ increases rapidly with n. Thus for moderate values of n, we use computer algebra to obtain exact expressions for E$\tilde{P}_n$. A computer algebra package manipulates algebraic expressions, unlike a conventional computing language, in which only numerical values are used. Using the package REDUCE 3.3, exact expressions for E$\tilde{P}_n, n \leq 20$, have been obtained. A listing of the simple program used is given in the

appendix. Thus we obtain for example,

$$E\tilde{P}_1 = \tfrac{4}{3},$$

$$E\tilde{P}_2 = \frac{(4(15\pi+32))}{(45\pi)},$$

$$E\tilde{P}_3 = \frac{(1575\pi^2+13440\pi-13088)}{(1575\pi^2)},$$

$$E\tilde{P}_4 = \frac{(2(55125\pi^3+1411200\pi^2-1374240\pi-2768896))}{(165375\pi^3)},$$

$$E\tilde{P}_5 = \frac{(63669375\pi^4+4346496000\pi^3-3174494400\pi^2-25584599040\pi+25516279808)}{(152806500\pi^4)}.$$

The REDUCE package can also produce its output in a format suitable for immediate inclusion in a FORTRAN program, which can then be used to obtain very precise numerical values.

4.5 A Model using Local Optimality

In this section, we consider a model in which the territory chosen is locally optimal, as defined in Section 4.1.6. As in the previous three sections, we consider only the simple measure $q(.)$ of territory quality defined in equation (4.1.3), although other criteria could be used.

The procedure for selecting a territory is as follows. Firstly, select a resource point at random, x_0. Let the 'current' territory be $\tau_0 = \{x_0\}$. Note that x_0 need not be an element of the final territory, unlike in previous models. Secondly, if the current territory is τ_k, consider all sets υ which are neighbours of τ_k, in the sense defined in Section 4.1.6. If there are no neighbours υ which are better than τ_k, then the process stops, with $\tau = \tau_k$ as the chosen territory. If there is a neighbouring set υ which is better than τ_k, then choose τ_{k+1}, the new 'current territory', to be the best of the neighbouring territories, and repeat the second step.

Within the above procedure, we have to define carefully what we mean by one territory being better than another. Clearly if τ_* or any neighbour υ has $\mu \geq m$, then we can choose the best set to be the one with the highest value of $q(.)$. But typically $\tau_0 = \{x_0\}$ will have $q(\tau_0) = -\infty$, and possibly $q(\upsilon) = -\infty$ for all neighbours υ of τ_0. So clearly some other way of choosing between territories is needed in the initial stages of the search. The problem is similar to that encountered in Section 4.3. However, since we are comparing territories with differing numbers of points, the minimisation of perimeter, used in Section 4.3, is not useful here. Instead, when choosing between territories with $\mu < m$ (and hence $q = -\infty$), we prefer the territory with the larger value of μ. (Note that such a criterion would not have worked in Section 4.3: in that model, with random marks, it would lead to very large perimeters, since the search was over the whole habitat.) Of two territories which have equal values of μ, as will happen frequently when all points have unit mark, or which both have $\mu \geq m$, we prefer the one with the smaller perimeter.

The above procedure will eventually select a territory τ satisfying

$$q(\tau) > -\infty \quad \text{and} \quad q(\tau) \geq q(\upsilon)$$

for any υ which is a neighbour of τ, except in the pathological case in which the sum of marks in the whole habitat is less than m. Unfortunately, the properties of the resulting territory are difficult to determine.

Table 4.4

Numerical Results for Locally Optimal Territories

"Const" indicates marks which are constant;

"Exp" indicates marks which are exponentially distributed

Case	Marks	m	EP	S.E.	EP^2	S.E.	Sample	C.p.u.	m'	p	S.E.
A	Const	1	0.000	-	0.000	-	-	-	-	-	-
B	Const	2	0.720	0.023	0.687	0.041	330	0.83	-	-	-
C	Const	3	1.552	0.037	2.818	0.133	300	2.9	-	-	-
D	Const	4	2.395	0.052	6.450	0.275	270	8.1	-	-	-
E	Const	5	3.198	0.058	10.85	0.392	185	14.3	-	-	-
H	Exp	1	0.031	0.011	0.044	0.018	380	0.32	1.5	0.583	0.031
J	Exp	2	0.367	0.039	0.586	0.087	330	0.82	-	-	-
K	Exp	3	1.083	0.064	2.390	0.201	300	1.9	-	-	-
L	Exp	4	1.892	0.091	5.817	0.475	270	3.7	-	-	-
M	Exp	5	2.641	0.109	9.937	0.687	250	4.2	-	-	-
Q	Exp	2.18	0.514	0.043	0.874	0.112	330	0.97	2.42	0.809	0.022
"	"	"	"	"	"	"	"	"	2.74	0.584	0.027
"	"	"	"	"	"	"	"	"	3.38	0.265	0.025
R	Exp	4.33	2.120	0.093	6.754	0.456	260	4.7	4.78	0.503	0.031

In the absence of any analytic results, Table 4.4 shows simulation results for the local optimisation model, covering a similar range of cases to the tables in previous sections of this

chapter.

An initial look at the values of EP in Table 4.4 shows that they are all lower than the corresponding values in Table 4.1 (the simultaneous choice model) and hence lower than the values for either of the other two models. The differences are statistically significant for all cases occuring in both Tables 4.1 and 4.4 (except the trivial case A), and are quite large, particularly in cases with exponential marks and small m (cases H, J, Q).

The probabilities p of group formation in the local optimisation model are surprisingly similar to those in the non-spatial model (and hence in the sequential and nearest-neighbour models) in cases H and Q. In case R, the probability in the local optimisation model (0.503, S.E. 0.031) is significantly lower than in the non-spatial model (0.638), but the difference is still fairly small considering the difference in expected perimeter.

4.6 Comparison of the models.

In this section, we discuss and summarise the comparisons between numerical results for different models made in Sections 4.2 to 4.5.

All the numerical results are based on the function $q(.)$ in equation (4.1.3). An important question is what quality of territories do the models produce, i.e for satisfactory territories, how small a perimeter can be obtained?

Firstly, consider cases with constant marks. These can be thought of as limiting cases as the variation between the marks, which represent mean (over time) yields from patches, decreases to zero. When $m=1$ all the models give a perimeter which is identically zero. When $m=2$ the first three models always agree, giving a territory consisting of the initial random part and its nearest neighbour, while the local optimisation model gives a smaller expected perimeter. For $m = 3,4,5$, the remaining cases for which full results are available, the local optimisation model always gives the smallest perimeter. The order of the the values of EP for the remaining three models is in accordance with intuition, so that the overall ordering for EP in the different models is

local optimisation < simultaneous < sequential < nearest-neighbour.

Secondly, in the cases involving exponentially distributed marks, the same ordering holds. The differences in these cases are much more pronounced, since with random marks, the simultaneous and local optimisation models may well have fewer resource points than the sequential and nearest neighbour models. The local optimisation model in particular does much better than any of the other models, presumably since all the other models are constrained to include an initial random point which may contribute little to the resource content of the territory.

So the local optimisation model clearly performs better than the others, in the sense of giving territories with acceptable levels of resources and smaller perimeters, especially with random marks. These results suggest that, if all the above procedures for selecting a territory were feasible for animals to actually use, there would be a definite advantage in using the local optimisation procedure.

The other important characteristic, apart from perimeter, of the territories formed by these models is the excess of the actual sum of marks (or mean yield of resources) μ over the required value m. The distribution of that excess determines the probability of social groups being formed in these territories. As noted above, for the sequential and nearest neighbour models, the distribution of the excess is identical to the distribution in the non-spatial models of Chapter 3. Comparing values for $p = \Pr(\mu \geq m')$ in Tables 4.1 and 3.6 shows that the excess is generally smaller, and p is generally larger, in the simultaneous choice model than in the non-spatial model, at least in case Q (m = 2.18). On the other hand comparing Table 4.4 with Table 3.6 shows that with the local optimisation model, the probability of forming groups is very close to the corresponding probability in the non-spatial model, at least in cases H (m=1) and Q (m = 2.18). In Case R, however, the local optimisation model shows a decrease in the probability of group formation, compared with the non-spatial model, which is statistically significant (at the 5% level).

Thus the main conclusions for these single-territory models are: the local optimisation model (Section 4.5) gives the smallest perimeters, and gives probabilities of group formation similar to those of the well-understood non-spatial model of Chapter 3; the simultaneous choice model (Section 4.2) gives perimeters which are much higher than in the local optimisation model, and probabilities of group formation which are lower than in the non-spatial model or any of the other spatial models; and the simplified models of Sections 4.3 and 4.4, while somewhat more tractable, are limited in use because the territories they produce have much higher perimeters than in either the local optimisation model or the simultaneous choice model.

4.7 Computation.

The purpose of this section is to indicate briefly some of the important points about the numerical computation required to obtain the results of the current chapter (algebraic computing has already been discussed in Section 4.4.5).

Where possible, standard numerical packages have been used. In particular, extensive use has been made of the N.A.G. Library (Numerical Algorithms Group, 1986) for efficient numerical integration and sorting. The results of Section 4.4 (the nearest-neighbour model) could thus be obtained fairly easily. For the remaining models, extensive programming was necessary. The programs used were actually written to enable the simulation of multiple,

interacting territories (Chapter 5) and are rather less efficient at simulating single isolated territories than a tailor-made program would be. The c.p.u. times (in seconds, per territory simulated) given in Tables 4.1, 4.2, and 4.4 should only be interpreted as relative costs for simulating the different models.

For each of these three models (in Sections 4.2, 4.3, and 4.5), the simulation follows the same basic steps.

(i) Simulate an area of habitat, i.e. a realisation over a finite area of a homogeneous Poisson process, with marks which may or may not be random. The area is taken to be circular, to minimise the simulated area which is unavailable due to edge-effects, as described in (ii) below.

(ii) Select a random point of the process from within the simulated area, to serve as the initial resource point in a territory. To avoid edge-effects, this initial point should not be too close to the edge of the simulated area. The minimum allowable distance from the initial point to the edge was taken to be $4\sqrt{\frac{m}{\pi}}$ i.e. twice the diameter of a disc having an expected resource content equal to the requirement for a territory. This value was chosen so that the territory was very unlikely to reach the edge of the simulated region: informal trials suggested that it was a large enough distance to make edge effects negligible.

(iii) Given the initial point from (ii), choose a territory according to the "rules" for the required model, and record its final resource content and the perimeter of its convex hull.

In practice, the simulated habitat can be used as the basis for many territories, so each time step (i) is carried out, steps (ii) and (iii) are executed a number of times. The re-use of a simulated area affects the inference from the simulation, and improves the overall efficiency of simulation. In the simulation of k territories let P_j be the perimeter of the jth territory and I_j be the indicator random variable of the event

$$\{\mu \geq m' \text{ for territory } j\}$$

(for whatever m' is of interest), for $j = 1,\ldots,k$. If the simulated territories were independent, we would estimate the mean perimeter by

$$\bar{P} = k^{-1}\sum_{j=1}^{k} P_j$$

with estimated standard error

$$k^{-\frac{1}{2}}.\frac{\sum_{j=1}^{k} P_j^2 - k\bar{P}^2}{k-1},$$

and the probability p of group formation by

$$\bar{I} = k^{-1} \sum_{j=1}^{k} I_j$$

with estimated standard error

$$k^{-\frac{1}{2}} \bar{I}(1-\bar{I}).$$

However, since we are re-using the simulated habitat, our observations are not independent. Instead, they constitute a random sub-sample (with replacement, because of the structure of the program) from a finite random sample determined by the simulated habitat. Since the models for territory formation are deterministic (once the habitat and the starting point are chosen), the size of the latter sample (determined by the habitat) is just the number of resource points in the simulated habitat which satisfy the requirement in (ii) above, say K. The inter-dependence between observations increases the variances of $\bar{P}$ and $\bar{I}$ by a factor of $1+(k-1)K^{-1}$, giving estimated standard errors of

$$\left\{ \frac{1+(k-1)K^{-1}}{k} \right\}^{\frac{1}{2}} \cdot \frac{\sum_{j=1}^{k} P_j^2 - k\bar{P}^2}{k-1}$$

and

$$\left\{ \frac{1+(k-1)K^{-1}}{k} \right\}^{\frac{1}{2}} \bar{I}(1-\bar{I})$$

respectively. We can think of $\frac{k}{1+(k-1)K^{-1}}$ as the effective sample size, and values are given in the tables of simulation results above.

The choices of k and K affect the efficiency of the simulation. Initial trials showed that in a wide range of cases, simulation is most efficient, as measured by effective sample size divided by c.p.u. time, when $k \approx K$, and so this relationship has been used throughout the current chapter. The best value of k and K to use for a given case and model, is determined by experimentation. In some cases, more than one set of simulations at the most efficient value of k is necessary to give a sufficiently large effective sample size.

The above considerations apply to all the simulations in this chapter. There are also specific techniques which can improve the efficiency of the simulation of a particular model. With the simultaneous choice model (Section 4.2), the territory chosen will in theory be the best possible from the whole, infinite habitat. In practice, as described above, we simulate a finite part of the habitat, and assume that the optimum territory (based on an initial point not too close to the boundary) lies within the simulated area. Nevertheless, the number of potential territories which need to be evaluated may be very large. In the simplest case, with constant marks and m an integer, the number of potential territories increases approximately as the $(m-1)$th power of the number of points in the simulated habitat. However, most of these

potential territories can be eliminated without the computationally expensive step of calculating the perimeters of their convex hulls.

The key is to note that, since (in this model) the territory is constrained to contain the initial point x_0, the perimeter of any territory containing another point x cannot be less than $2d(x_0,x)$. So provided a viable territory τ_0 containing x_0 exists, we can disregard any point x such that

$$d(x_0,x) > \tfrac{1}{2}p(\tau_0).$$

The procedure actually used can be outlined as follows.

(a) Select an initial point x_0.

(b) Select a "reasonable" territory τ_0, containing x_0, quickly (by using the nearest neighbours of x_0).

(c) Define p^* to be the lowest perimeter so far, initially equal to $p(\tau_0)$.

(d) Define σ to be the set of resource points x with

$$d(x_0,x) \leq \tfrac{1}{2}p^*.$$

(e) Consider each subset τ of σ which contains x_0, which has $\mu(\tau) \geq m$, and which is minimal in the sense that there is no strict subset υ of τ with $x_0 \in \upsilon, \mu(\upsilon) \geq m$. If $p(\tau) < p^*$, then set p^* equal to $p(\tau)$, and eliminate from further consideration any point x with

$$d(x_0,x) > \tfrac{1}{2}p^*,$$

and thus any subset of σ containing x.

(f) When all subsets of σ have been considered (or eliminated), the chosen territory is the one giving the final value of p^*.

The method is further speeded up by labelling the elements of $\sigma \setminus \{x_0\}$ as $x_1,x_2,\ldots$, so that $d(x_0,x_1) \leq d(x_0,x_2) \leq \ldots$. Then (at step (e)) we consider the subsets of σ in lexical order so that: all subsets containing x_1 are examined before the subsets without x_1; within each of those categories, subsets that contain x_2 are examined before those that do not; and so forth. Because the better territories are likely to be those consisting of points near x_0, this ordering means that points in σ are more likely to be eliminated early on, so less subsets will have to be evaluated.

For the "sequential" model in Section 4.3, a similar approach is applied, at each stage in the formation of a territory, to the choice of the next point to include in the territory. The scope for saving processing time is not so great for the "sequential" model as for the "simultaneous" model, since the number of possibilities to be tried does not increase so quickly with

the number of points in the habitat.

In the local optimisation model (Section 4.5), it is necessary to calculate the Delaunay triangulation of the resource points, in order to determine which pairs of sets of points are neighbours. Since the triangulation involves all the resource points, it is potentially expensive to compute. To minimise the time spent, the highly efficient algorithm of Green and Sibson (1978) is used.

Another potential problem is that the value, $q(\tau)$ say, of a particular subset may be needed at more than one stage in the search for a local optimum. If $\tau_1, \tau_2, \ldots$ is the sequence of "current" territories (in the terminology and notation of Section 4.5), then τ may be a neighbour of τ_i and τ_j $(i \neq j)$, and a naive approach would result in $q(\tau)$ being calculated twice. The method used to avoid this recalculation is to store known values of $q(.)$ in a tree-like data structure which can be easily searched to find out whether $q(\tau)$ has been calculated, and retrieve its value if so.

Chapter 5 Models for Multiple Territories

5.1 Introduction

The models in Chapter 4 considered the random configuration of one isolated territory. While such models give some tractability, and are directly comparable with many models mentioned above (Sections 1.2, 3.2.1 and 4.1.2), it is clear that in reality, territories will interact. In the current chapter, we will discuss some existing models for interacting territories, and then consider how the spatial models of Chapter 4 can be extended to include more than one territory.

5.1.1 Existing Models

The existing models for multiple territories, like those for single territories, nearly all assume a uniform habitat (Jones and Krummel, 1985; Maynard Smith, 1974; see also the review by Covich, 1976). These models are typically stated to predict a pattern of equal-sized regular hexagons as territories. A more accurate statement, at least for Jones and Krummel (1985) and Maynard Smith (1974), is that a pattern of regular hexagons is clearly a stable equilibrium of the model described, but not enough detail on dynamics and/or initial conditions is given to determine whether, or with what probability, that particular equilibrium would be obtained. In fact, Hasegawa and Tanemura (1976) show by simulation that, for a simple, intuitive choice of the dynamics and initial conditions, a pattern of regular hexagons does not occur. Instead, a stable pattern is reached in which territories are polygonal, and the number of sides has some non-constant distribution, (the modal value, six, having a probability of approximately a half). Their model is further discussed in Hasegawa and Tanemura (1980) and Tanemura and Hasegawa (1980); the conclusion is essentially unchanged. The same authors, Hasegawa and Tanemura (1980) and Tanemura and Hasegawa (1980), also give an alternative model for territory formation, based on the Voronoi tessellation (Section 4.1.3) of the centres of randomly packed circles. Again, a uniform habitat is assumed, and the model is more concerned about the separation between centres than with any resource-based criterion.

Thus all existing models seem to assume uniform habitats, and, with the exception of the work of Hasegawa and Tanemura, are imprecisely defined.

5.1.2 Modelling Interacting Territories

The models we will use are directly derived from those in Chapter 4, modified to allow multiple territories to exist and interact. Most of the necessary changes apply equally to the four models of Sections 4.2 to 4.5.

As before, we have a realisation ψ of a (marked) point process Ψ, representing resources. Unlike the models in Chapter 4, we consider only a finite region A of the plane, for simplicity. Now instead of determining a single territory, we form territories in sequence, continuing until no more viable territories are possible. The initial point of each new territory is selected at random from those points not already included in a territory. These models thus have some features in common with "hard-core" sequential inhibition processes (see e.g. Diggle, 1983, or Stoyan et. al., 1987). However the models described here are more complex, and less tractable, since the final result is a collection of sets of points, rather than just a collection of points.

The first territory τ_1 generated by such a model will have the same properties as the single territory formed in the corresponding model from Chapter 4. Denote the common distribution by $T(\psi,A)$, and write

$$\tau_1 \sim T(\psi,A).$$

Note that τ_1 takes values in the power set $\{0,1\}^{\Psi}$, and since the process of selecting a territory is deterministic once the initial point is specified, its distribution assigns probability $\frac{1}{\psi(A)}$ (where $\psi(A)$ is the number of points of ψ in the whole region) to each of $\psi(A)$ subsets of ψ, corresponding to the equally likely choices of initial point, and probability 0 to any other subset.

Clearly we wish to prohibit subsequent territories from containing points which have already been used. Such points are thus excluded from selection as starting points for the search for any further territory, and from membership of intermediate or final territories in the search.

We could define the kth territory τ_k in a region A to have the same distribution as τ_1 except that the points of $\tau_1,\ldots,\tau_{k-1}$ are ignored, giving

$$\tau_k \sim T(\psi \setminus (\tau_1 \cup \ldots \cup \tau_{k-1}),A).$$

However, simple examples show that configurations with positive probability in $T(\psi \setminus (\tau_1 \cup \ldots \cup \tau_{k-1}),A)$ may, when thought of as territories, be biologically implausible given the existence of territories represented by $\tau_1,\ldots,\tau_{k-1}$.

For example, in the situation shown in Figure 5.1, given the prior existence of territories τ_1 and τ_2, a territory containing both the points x and y is unlikely.

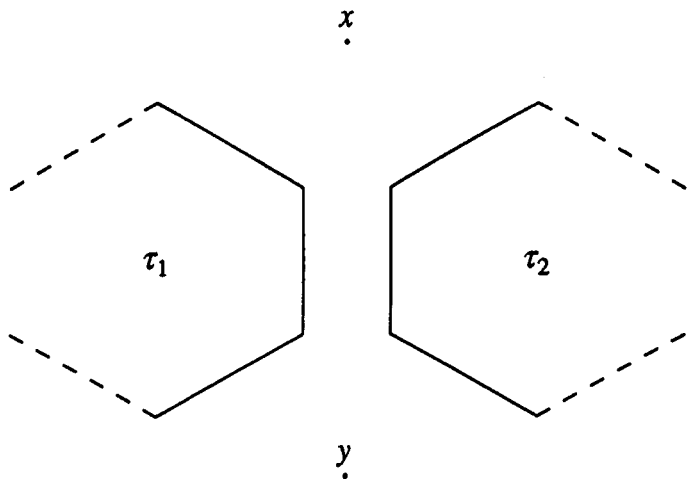

Figure 5.1 The Interaction of Territories

The process is modified in two ways to take account of these biological constraints. Firstly, we avoid territories which are potentially divided into two or more parts by the presence of other territories. We wish a territory to be connected in some sense: the definition we use is based on the Delaunay triangulation $\mathfrak{D}(\psi)$, as defined in Section 4.1.3. We require all territories to be connected within $\mathfrak{D}(\psi)$; more precisely, given $x,y \in \tau \subset (\psi \cap A)$, then for some m there must exist $x_1,\ldots x_m \in \tau$ such that the pairs $\{x,x_1\},\{x_m,y\}$ and $\{x_i,x_{i+1}\}$ $(i = 1,\ldots,m-1)$ are all pairs of neighbours in $\mathfrak{D}(\psi)$. Any τ not satisfying the above criterion is prohibited: we set its quality $q(\tau) = -\infty$. Note that, unlike the models in Chapter 4, models for multiple territories may result in all territories τ based on an initial point x_0 having quality $q(\tau) = -\infty$. For example, consider the case where the point x_0 has mark $r_0 < m$, and where all the neighbours of x_0 in the Delaunay triangulation already belong to other territories. Then no territory τ containing x_0 has $q(\tau) > -\infty$, and the initial set $\{x_0\}$ has no neighbours in the sense of Section 4.1.6.2.

Secondly, we wish the cost of defending a territory τ, which affects $q(\tau)$, to reflect the presence of other territories. For a single territory τ, the core area $C(\tau)$ coincides with $H(\tau)$, the convex hull of τ, as discussed in Section 4.1.4, and we are interested in $p(\tau)$, the perimeter of $H(\tau)$. Within a model for multiple territories, $H(\tau)$ has the same definition, but the definition of the core area depends on existing territories, say $\tau_1,\tau_2,\ldots,\tau_k$. We will define the core area, denoted by $C(\tau|\{\tau_1,\tau_2,\ldots,\tau_k\})$, inductively.

When $k = 0$, there are no previous territories, and the appropriate definition is

$$C(\tau|\varnothing) = H(\tau).$$

For $k \geq 1$, we proceed as follows.

(i) Determine $\mathfrak{D}(\tau)$, the Delaunay triangulation of the points of τ.

(ii) Partition $H(\tau)$ into a set $\mathfrak{P}$ of line segments (the edges of $\mathfrak{D}(\tau)$) and open triangles (the components of the set formed by deleting the edges of $\mathfrak{D}(\tau)$ from $H(\tau)$).

(iii) Assume that core areas

$$C(\tau_i|\{\tau_1,\ldots,\tau_{i-1}\}),\ i = 1,\ldots,k$$

have already been defined for existing territories. Then we can define $\mathfrak{C}$, the set of elements of $\mathfrak{P}$ which are disjoint from

$$C = \bigcup_{i=1}^{k} C(\tau_i|\{\tau_1,\ldots,\tau_{i-1}\}),$$

the combined core area of the existing territories.

(iv) Define $C(\tau|\{\tau_1,\ldots,\tau_k\})$ to be the union of the elements of $\mathfrak{C}$.

The perimeter relevant to the cost of defending τ, and used in determining the quality of τ, is $p(\tau|\{\tau_1,\ldots,\tau_k\})$, the perimeter of $C(\tau|\{\tau_1,\ldots,\tau_k\})$.

There is one case in which the core area defined by the above process is rather unrealistic. It is possible for the perimeter of some existing territory, τ_j say, to separate $C(\tau|\{\tau_1,\ldots,\tau_k\})$ into more than one component. The perimeter calculated for τ is then just the sum of the perimeters of the components, whereas it would be more reasonable to disallow τ completely. Normally, the above problem is prevented by the requirement that τ itself be connected. Under that constraint, the problem can only occur when τ_j completely encircles not only another territory, τ_i say, which existed before τ_j, but also at least one unused resource point which can then be included in τ. Even then, the configuration required for the problem to occur is extremely unlikely. In view of the very high extra computational cost that would be required to avoid the situation described, the low probability of it occurring, and the minor effect that it is likely to have if it did occur, no allowance has been made for the possibility.

Since we are considering a sequence of territories in a finite habitat, there will eventually come a stage at which no further territories can be added. Clearly this is the case if all resource points have been included in territories, so that no initial point can be found for any further territory. In addition, we also consider a habitat containing territories $\tau_1, \tau_2, \ldots, \tau_k$ to be full, i.e. incapable of supporting any more territories, if the best territory τ which could be formed with any remaining point $x \in \psi \setminus (\tau_1 \cup \ldots \cup \tau_k)$ as initial point has quality $q(\tau) = -\infty$.

Thus the habitat is full when each component of the graph induced on $\psi \setminus (\tau_1 \cup \ldots \cup \tau_k)$ by $\mathfrak{D}(\psi)$, has total yield (sum of marks) less than m, the yield needed for a viable territory.

5.1.3 Overview of the chapter

The remainder of the current chapter parallels Chapter 4. In Sections 5.2 to 5.5, we consider the four models of territory formation from Chapter 4, each extended to allow for multiple territories. In Section 5.6 we summarise and compare the numerical results from those models, and in Section 5.7 we discuss computational aspects.

5.2 Simultaneous Choice of Points

The model based on the simultaneous choice of the points in a territory, introduced in Section 4.2, generalises readily to the case of multiple territories, in the way described in Sections 5.1.2. Given an initial point x_0, a territory τ is chosen such that

$$q(\tau) = \max\{q(\upsilon) : x_0 \in \upsilon \subset \psi\},$$

provided that $q(\tau) > -\infty$. When no more such territories exist, the habitat is full.

The properties of the territories formed are even more difficult to obtain than in the corresponding single territory model. Except in the most trivial case ($m = 1$, constant marks), simulation appears to be necessary, but it is computationally expensive. In only a few cases can territories readily be simulated in sufficient quantities to give meaningful results, and these are shown in Table 5.1 below. For more computational details, see Section 5.7.

Table 5.1

Numerical Results for Territories based on Simultaneous Choice

"Const" indicates marks which are constant;

"Exp" indicates marks which are exponentially distributed

Case	Marks	m	EP	S.E.	EP^2	S.E.	Sample	C.p.u.	m'	p	S.E.
A	Const	1	0.000	-	0.000	-	-	-	-	-	-
B	Const	2	1.295	0.019	2.324	0.075	1720	0.26	-	-	-
H	Exp	1	1.271	0.077	3.630	0.396	344	7.1	1.5	0.578	0.027

As with the single territory models in Chapter 4, case A is trivial, giving perimeters which are identically zero. Case B, as in Chapter 4, has exactly the same properties for the

three 'constrained choice' models (Section 4.1.6.1): simultaneous choice, sequential choice, and nearest neighbour. However, with multiple territories, it seems that case B of these models is no longer mathematically tractable. Instead, Table 5.1 includes simulations of case B, which show an increase in mean perimeter, as compared with the single territory case (e.g. Table 4.1), which is statistically significant. Case H of the simultaneous choice model also shows a higher value for EP in the multiple territory model (Table 5.1) than in the single territory model (Table 4.1). Such an increase is clearly to be expected, since many of the territories, in any multiple territory model, will be forced by pre-existing nearby territories to take shapes that are not optimal. In contrast, there is no evidence from case H in Table 5.1 that the probability of group formation is any different in the multiple territory model from in the single territory model in Section 4.2. Unfortunately, the range of cases of the current model that can be simulated is severely restricted by the computation time required.

In addition to results analogous to those for single territory models, we can look at the global properties of a multiple territory model. In particular, we can look at the density of territories in a habitat which is full, which we call ρ. Note that since we use a resource point process of unit intensity, the density ρ can be thought of as representing both territories per unit area and territories per resource point. The density will depend on the expected number of resource points in a territory (which if marks are random need not be the same as in a non-spatial model) and on the amount of resources unused when no more connected territories can be accommodated. Table 5.2 shows the estimated density $\hat{\rho}$ for each of the cases covered in Table 5.1, with the associated standard error, and the sample size (number of areas of habitat simulated, not number of territories) on which it is based. It also compares the density in each case with $\{EN\}^{-1}$, the number of territories per resource point in the non-spatial model from Chapter 3, where N is the number of points in a territory (see Section 3.2.5).

Table 5.2

Estimated Densities of Territories based on Simultaneous Choice

Case	Marks	m	$\{EN\}^{-1}$	$\hat{\rho}EN$	$\hat{\rho}$	S.E.	Sample size
A	Const	1	1.000	1.00	1.000	-	-
B	Const	2	0.500	0.95	0.477	0.011	50
H	Exp	1	0.500	0.95	0.476	0.024	20

The figures for $\hat{\rho}EN$ represent in some sense the efficiency with which territories are 'packed' in the current model, relative to the non-spatial model. The values summarise both the proportion of resource points left unused and the average number of resource points per territory, which, if marks are random, may not be the same as in the non-spatial model. Note that in principle, ρEN could take a value greater than 1, if the number of points per territory was smaller in the multiple territory model than in the non-spatial model.

Regardless of the model being used, case A will clearly have $\rho = EN = 1$; in a full habitat, with exactly 1 point per territory, there is a one-to-one correspondence between resource points and territories. Cases B and H in Table 5.2 have $\hat{\rho}EN$ close to 1, which means that the density of territories is close to that expected from the non-spatial model.

5.3 Sequential Choice of Points

The sequential model can also be adapted as a model for multiple territories. However, as in Section 4.3, the choice of successive points cannot be based in a naive way on $q(.)$, the function defining the quality of a territory. Instead, given $x_0,\ldots,x_{k-1}$, we add the point x_k which, taking into account the discussion in Section 5.1, minimises $p\{x_0,x_1,\ldots,x_k\}$ subject to the constraints that

$$x_k \in \psi \setminus \{x_0,\ldots,x_{k-1}\} \setminus \bigcup \tau_j,$$

and $\{x_0,x_1,\ldots,x_k\}$ is connected within the graph induced on $\psi \setminus \bigcup \tau_j$ by $\mathfrak{D}(\psi)$, where the union is taken over any territories already established. As in the sequential model for a single territory, the process starts with a single randomly chosen point x_0, and continues until $\sum_{j=0}^{k} r_j \geq m$, where r_j is the mark associated with x_j.

As with the simultaneous choice model, it does not seem possible to make analytic progress in determining the properties of territories formed in the sequential choice model. However, simulation is not so computationally expensive. Table 5.3 gives numerical results for the sequential choice model, for a range of cases. Case A in Table 5.3 is trivial, and case B, included here for completeness, has already been discussed in Section 5.2. In all other cases, the multiple territory model gives significantly higher mean perimeters (at the $2\frac{1}{2}\%$ level) than the corresponding single territory model (Table 4.2), as expected. In contrast, all the probabilities of group formation in Table 5.3 are very close to those in Table 4.2 (or equivalently, in the non-spatial model), suggesting that interactions between territories will have little effect on social structure, within the current model. Comparison of the sequential model with the simultaneous model for multiple territories is difficult because the information on the latter (as given in Table 5.1) is very limited. Only case H can be compared, and it does not show a significant difference between values of EP or between values of p.

Table 5.3

Numerical Results for Territories based on Sequential Choice

"Const" indicates marks which are constant;

"Exp" indicates marks which are exponentially distributed

Case	Marks	m	EP	S.E.	EP^2	S.E.	Sample	C.p.u.	m'	p	S.E.
A	Const	1	0.000	-	0.000	-	-	-	-	-	-
B	Const	2	1.295	0.019	2.324	0.075	1720	0.26	-	-	-
C	Const	3	2.488	0.036	7.486	0.234	993	0.73	-	-	-
D	Const	4	3.505	0.045	14.242	0.417	958	1.26	-	-	-
E	Const	5	4.309	0.049	21.030	0.541	1007	2.00	-	-	-
H	Exp	1	1.196	0.065	3.676	0.378	532	0.41	1.5	0.602	0.021
J	Exp	2	2.257	0.073	8.778	0.519	684	0.89	-	-	-
Q	Exp	2.18	2.403	0.073	9.685	0.578	735	0.77	2.42	0.792	0.015
"	"	"	"	"	"	"	"	"	2.74	0.577	0.018
"	"	"	"	"	"	"	"	"	3.38	0.302	0.017

Table 5.4

Estimated Densities of Territories based on Sequential Choice

Case	Marks	m	$\{EN\}^{-1}$	$\hat{\rho}EN$	$\hat{\rho}$	S.E.	Sample size
A	Const	1	1.000	1.00	1.0000	-	-
B	Const	2	0.500	0.95	0.4772	0.0109	50
C	Const	3	0.333	0.92	0.3068	0.0061	30
D	Const	4	0.250	0.89	0.2218	0.0046	30
E	Const	5	0.200	0.93	0.1865	0.0038	30
H	Exp	1	0.500	0.98	0.4912	0.0188	30
J	Exp	2	0.333	0.95	0.3163	0.0097	30
Q	Exp	2.18	0.314	0.99	0.3119	0.0076	30

Table 5.4 summarises information on estimated territory density in the sequential multiple territory model. As in the previous model, all the values of $\hat{\rho}EN$ are close to 1, so that territory density is always close to the value predicted by the simple non-spatial model.

5.4 The Nearest-Neighbour Model

The nearest-neighbour model, described in Section 4.4, is affected by the ideas in Section 5.1 in a very similar way to the sequential model (Section 5.3). Given points $x_0,\ldots,x_{k-1}$, the kth step in choosing a territory is to add the point x_k which minimises $d(x_0,x_k)$, i.e. the point closest to the nominal centre of the territory, subject to the constraints that

$$x_k \in \psi \setminus \{x_0,\ldots,x_{k-1}\} \setminus \bigcup \tau_j,$$

and $\{x_0,x_1,\ldots,x_k\}$ is connected within the graph induced on $\psi \setminus \bigcup \tau_j$ by $\mathfrak{D}(\psi)$, where the union is over any territories already established. As with the sequential model, x_0 is randomly chosen, and the process continues until $\sum_{j=0}^{k} r_j \geq m$, where r_j is the mark associated with x_j. Note that because of the requirement of connectedness, the points are not necessarily the nearest neighbours of x_0, even within the appropriate component of the graph. Nevertheless, the model just described does seem to be the most appropriate generalisation to multiple territories of the nearest-neighbour model in Section 4.4.

Table 5.5 gives simulation results for the model described above. As expected, the figures for mean perimeters in the multiple territory model, given in Table 5.5, are all significantly higher (at the 5% level) than in the corresponding single territory model (Table 4.3). The figures for EP are also significantly greater in Table 5.5 than in Table 5.3 (the sequential model), in most cases. However, none of the probabilities p of group formation are significantly different in Table 5.5 from in Tables 4.3 or 5.3. This suggests that the nearest neighbour, multiple territory model leads to similar social behaviour to the sequential choice, multiple territory model and the nearest neighbour, single territory model.

Table 5.6 gives estimated densities for territories according to the current model. As with the previous multiple territory models, all values of $\hat{\rho}EN$ are reasonably close to 1, so that the density is close to that suggested by the non-spatial model.

Note that, as mentioned in Section 5.2, it is possible for ρEn to be greater than 1, implying a higher density in the spatial model than predicted by the non-spatial model. However, as can be seen from the S.E. for $\hat{\rho}$, there is no evidence that $\rho EN > 1$ in case K of Table 5.6: the simulation results are consistent with $\rho EN \leq 1$.

Table 5.5

Numerical Results for Nearest Neighbour Territories

"Const" indicates marks which are constant;

"Exp" indicates marks which are exponentially distributed

Case	Marks	m	EP	S.E.	EP^2	S.E.	Sample	C.p.u.	m'	p	S.E.
A	Const	1	0.000	-	0.000	-	-	-	-	-	-
B	Const	2	1.295	0.019	2.324	0.075	1720	0.26	-	-	-
C	Const	3	2.714	0.041	8.959	0.272	960	0.52	-	-	-
D	Const	4	3.769	0.049	16.576	0.448	979	0.70	-	-	-
E	Const	5	4.803	0.054	25.847	0.595	965	1.00	-	-	-
F	Const	10	8.957	0.075	85.680	1.696	969	6.26	-	-	-
H	Exp	1	1.299	0.062	4.337	0.372	688	0.29	1.5	0.624	0.018
J	Exp	2	2.547	0.083	11.106	0.715	664	0.44	-	-	-
K	Exp	3	3.306	0.081	16.274	0.734	818	0.66	-	-	-
L	Exp	4	4.595	0.102	29.430	1.526	793	1.02	-	-	-
M	Exp	5	5.552	0.108	40.653	1.702	841	1.54	-	-	-
Q	Exp	2.18	2.601	0.085	11.765	0.745	698	0.47	2.42	0.784	0.016
"	"	"	"	"	"	"	"	"	2.74	0.567	0.019
"	"	"	"	"	"	"	"	"	3.38	0.322	0.018
R	Exp	4.33	4.953	0.100	32.476	1.308	793	1.10	4.78	0.629	0.017

Table 5.6

Estimated Densities of Nearest Neighbour Territories

Case	Marks	m	$\{EN\}^{-1}$	$\hat{\rho}EN$	$\hat{\rho}$	S.E.	Sample size
A	Const	1	1.000	1.00	1.0000	-	-
B	Const	2	0.500	0.95	0.4772	0.0109	50
C	Const	3	0.333	0.89	0.2966	0.0070	30
D	Const	4	0.250	0.91	0.2266	0.0040	30
E	Const	5	0.200	0.89	0.1787	0.0026	30
F	Const	10	0.100	0.90	0.0898	0.0010	30
H	Exp	1	0.500	0.95	0.4764	0.0135	40
J	Exp	2	0.333	0.92	0.3071	0.0073	30
K	Exp	3	0.250	1.01	0.2527	0.0064	30
L	Exp	4	0.200	0.92	0.1836	0.0033	30
M	Exp	5	0.167	0.93	0.1557	0.0031	30
Q	Exp	2.18	0.314	0.94	0.2962	0.0089	30
R	Exp	4.33	0.188	0.91	0.1698	0.0035	30

5.5 The Local Optimisation Model

The local optimisation model in Section 4.5 adapts readily to multiple territories. The key point to note is that the requirement that territories should be connected and should not contain points in pre-existing territories applies to all sets considered in the search for a local optimum: it is not merely a constraint on the territory finally selected.

Table 5.7 gives simulation results for the resulting model. In the discussion in this section, all significance levels are 5%. In both cases B and H, the mean perimeter in the local optimisation model is significantly lower than in any of the other three models. The same appears to hold in case C, although no value for the simultaneous choice model is available in that case.

Table 5.7

Numerical Results for Locally Optimal Territories
"Const" indicates marks which are constant;
"Exp" indicates marks which are exponentially distributed

Case	Marks	m	EP	S.E.	EP^2	S.E.	Sample	C.p.u.	m'	p	S.E.
A	Const	1	0.000	-	0.000	-	-	-	-	-	-
B	Const	2	1.170	0.021	1.960	0.074	1329	0.51	-	-	-
C	Const	3	2.306	0.038	6.715	0.253	953	1.40	-	-	-
H	Exp	1	0.946	0.112	4.070	0.746	256	2.07	1.5	0.426	0.031

The values of EP are, however, significantly greater than in the corresponding single territory model, the difference being particularly marked in case H.

The only probability of group formation available for the current model is for case H, and it is significantly lower than for any other model; the reason for this difference is not clear.

Table 5.8

Estimated Densities of Locally Optimal Territories

Case	Marks	m	$\{EN\}^{-1}$	$\hat{\rho}EN$	$\hat{\rho}$	S.E.	Sample size
A	Const	1	1.000	1.00	1.0000	-	-
B	Const	2	0.500	0.92	0.4609	0.0086	40
C	Const	3	0.333	0.88	0.2945	0.0065	30
H	Exp	1	0.500	1.05	0.5249	0.0309	40

Territory densities for the current model are shown in Table 5.8. All values of $\hat{\rho}EN$ are close to 1, as has been found for the other multiple territory models. Although the estimate in case H seems rather high, it is consistent with $\rho EN = 1$, $\rho EN < 1$ or $\rho EN > 1$.

5.6 Comparison of the Models

The aim of this section is to summarise the main comparisons which have been made in the current chapter, between models and with the models in Chapter 4.

Firstly, consider the results obtained for the expected perimeter of a territory, EP. The comparisons between multiple territory models are rather weak, largely because of the high cost of simulating most such models. There are no significant differences between the simultaneous choice model and the sequential choice model. Both models appear to give larger values than the local optimisation model, and smaller values than the nearest neighbour model, in the sense that all estimated differences are in the given direction, and a number of them are statistically significant (at the $2\frac{1}{2}\%$ level).

Comparing the multiple territory models of the current chapter with their respective single territory versions in Chapter 4 gives a rather stronger, though unsurprising, conclusion. In all cases for which information is available, multiple territory models give larger mean perimeters than the corresponding single territory models, with all differences being significant (at the $2\frac{1}{2}\%$ level).

These relationships are illustrated in a simple way by Tables 5.9 and 5.10, which give values for EP in cases C and H respectively. These are the two cases for which the most information is available. All values are taken from Tables 4.1, 4.2, 4.3, 4.4, 5.1, 5.3, 5.5 and 5.7, which contain more precise estimates and standard errors.

Table 5.9

Estimates of EP in case C (constant marks, $m = 3$); all models

Model	Single Territory	Multiple Territories
Simultaneous Choice	1.86	-
Sequential Choice	1.94	2.49
Nearest Neighbour	2.10	2.71
Local Optimisation	1.56	2.31

Estimates in Table 5.9 (case C) have standard errors of approximately 0.04, except the figures for the single territory, sequential choice / nearest neighbour models, which are exact.

Table 5.10

Estimates of EP in case H (exponential marks, m = 1); all models

Model	Single Territory	Multiple Territories
Simultaneous Choice	0.81	1.27
Sequential Choice	0.98	1.20
Nearest Neighbour	1.03	1.30
Local Optimisation	0.03	0.95

Estimates in Table 5.10 (case H) have standard errors of between 0.05 and 0.08, except the figure for the single territory, nearest neighbour model, which is exact, and the two figures for the local optimisation model, which have standard errors of 0.01 and 0.11.

The other main feature of importance in these models is the probability p of group formation. The relationships between values of p in different models are not easily summarised; we will present and discuss some examples. Tables 5.11 and 5.12 give the percentages of territories which will contain social groups, according to each model, in cases H and Q, with appropriate values of m'. Again, the figures are based on Tables 4.1, 4.2, 4.3, 4.4, 5.1, 5.3, 5.5 and 5.7.

Estimates in Table 5.11 (case H) have standard errors of 2–3%, except the figures for the single territory, sequential choice / nearest neighbour models, which are exact.

Estimates in Table 5.12 (case Q) have standard errors of 2–5%, except the figures for the single territory, sequential choice / nearest neighbour models, which are exact.

In each of the two tables of percentages above, all entries are similar (not significantly different) except one (significantly different at the 1% level). However, the two values which are different do not come from the same model in each case. In Table 5.11, the value from the multiple territory, local optimisation model is lower than all other values, and in Table 5.12, the value from the single territory, simultaneous choice model is lower than the rest. Note that interpretation of Table 5.12 is complicated by the fact that some figures are unavailable due to limitations on computing time.

Table 5.11

Estimated percentages of territories containing social groups in case H (exponential marks, $m = 1, m' = 1.5$); all models

Model	Single Territory	Multiple Territories
Simultaneous Choice	61	58
Sequential Choice	61	60
Nearest Neighbour	61	62
Local Optimisation	58	43

Table 5.12

Estimated percentages of territories containing social groups in case Q (exponential marks, $m = 2.18, m' = 2.74$); all models

Model	Single Territory	Multiple Territories
Simultaneous Choice	42	-
Sequential Choice	57	58
Nearest Neighbour	57	57
Local Optimisation	58	-

The general conclusion is therefore that the level of social behaviour, as indicated by the probability p, is not greatly affected by the exact spatial model chosen, with a few exceptions which seem difficult to predict.

5.7 Computation

As mentioned in Section 4.7, the main programs used in Chapter 4 were written to enable multiple territories to be simulated. Thus the same programs are used, in a rather different way, to produce the numerical results in the present chapter.

The basic steps are as follows, regardless of the particular model being simulated.

(i) Generate a random habitat over a finite circular region. Initially all points are regarded as 'available'.

(ii) Select a random resource point from those 'available' within the habitat to serve as the initial point of a territory.

(iii) Choose a territory based on that initial point, according to the particular model required, if possible. Flag any points included in a territory, or unsuccessfully tried as initial points, as 'unavailable'.

(iv) Repeat steps (ii) and (iii) until no more points are available.

(v) To estimate expected perimeter etc, use only those territories with initial points at least a distance of $4\sqrt{\frac{m}{\pi}}$ from the edge of the simulated region. Trials suggest that this approach overcomes edge effects, and is more efficient than actually making points close to the edge of the region unavailable as initial points.

Note that, for all models except the nearest-neighbour model, step (iii) requires the calculation of the perimeter of many potential territories, taking into account any pre-existing territories. These calculations would be computationally expensive if carried out naively, since (denoting the potential territory by τ) each one involves checking each line segment in $\mathfrak{D}(\tau)$, and each triangle formed by $\mathfrak{D}(\tau)$, for intersection with the existing territories. Since new territories cannot develop wholly inside old ones, because of the nature of the models used, the existing territories can be represented by the resource points they contain and the line segments forming their perimeters (not necessarily their convex hulls). All triangles formed by $\mathfrak{D}(\tau)$ must be checked for the presence of resource points belonging to existing territories. A line segment in $\mathfrak{D}(\tau)$ need only be checked for intersection with the perimeter of existing territories if (a) it is on the perimeter of $H(\tau)$, (b) it is adjacent to a triangle containing points of an existing territory, or (c) it has an endpoint in common with a line segment which does intersect existing territories. These checks together are sufficient to define the core area of τ in a reasonably efficient way, provided the number of resource points in existing territories is not too large.

The approach is illustrated in Figure 5.2, in which $\tau = \{x_0, x_1, \ldots, x_6\}$. All six of the triangles formed by $\mathfrak{D}(\tau)$ are checked, and $x_0x_2x_3$ is found to contain y_0, a point of the existing territory $\upsilon = \{y_0, y_1, y_2, \ldots\}$. The line segments $x_1x_2, x_2x_3, \ldots x_5x_6, x_6x_1$ are checked in accordance with criterion (a) above, and the segment x_0x_2 is checked in accordance with (b). Since x_1x_2 and x_0x_2 both intersect υ, the segment x_0x_1 is also checked, in accordance with (c). Thus the core area $C(\tau|\upsilon)$ can be defined: it consists of the whole of $H(\tau)$ except the triangles $x_0x_1x_2$ and $x_0x_2x_3$, and the line segments x_0x_2 and x_1x_2. Its perimeter is given by

$$p(\tau|v) = d(x_0,x_3)+2d(x_2,x_3)+d(x_3,x_4)+d(x_4,x_5)+d(x_5,x_6)+d(x_6,x_1)+d(x_3,x_4);$$

note that the segment x_2x_3 is counted twice, to give a closed curve.

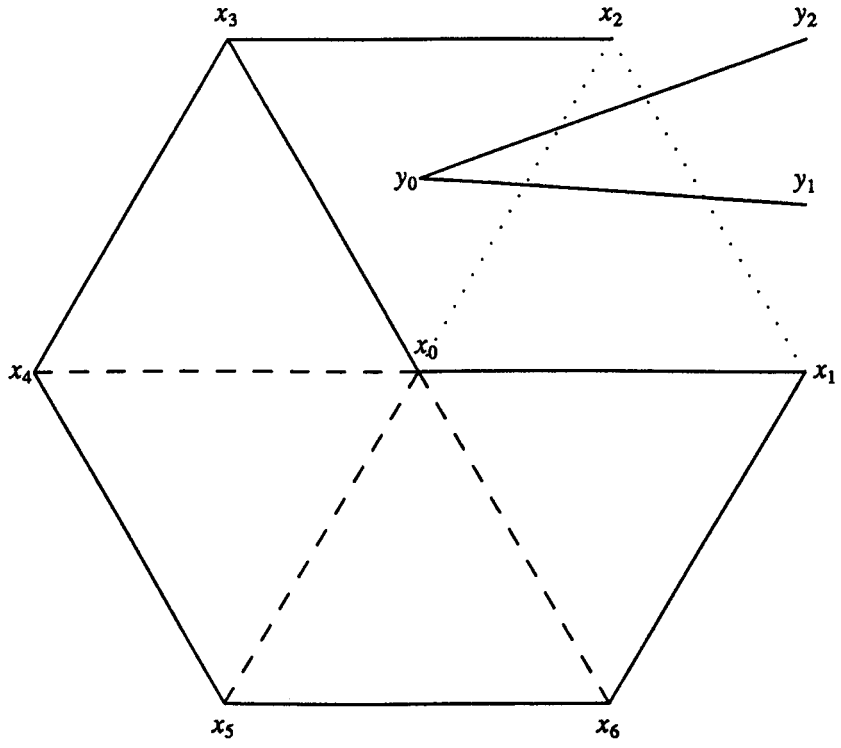

Figure 5.2 Calculating the Perimeter of a Territory

If the number of resource points and perimeter line segments belonging to existing territories is large, then involving all such points and lines in the above checks is computationally expensive. Instead, we can do some preliminary calculations to reduce the number of checks. If τ is the territory currently being considered, define $\mathfrak{R}$ to be the smallest rectangle, with sides parallel to the co-ordinate axes being used, which contains τ. Given a point x of a pre-existing territory, it is straightforward to determine whether or not $x \in \mathfrak{R}$; and if not, then clearly x cannot affect the core area or the perimeter of τ. Similarly, given a perimeter line segment of a pre-existing territory, with end-points x_1,x_2 say, it is often straightforward to see that x_1x_2 does not intersect $\mathfrak{R}$, and therefore cannot affect the core area or the perimeter of τ. These preliminary calculations decrease computing time considerably, especially when a large area of habitat is being simulated.

Chapter 6 Evolutionary Models

6.1 Introduction

6.1.1 Motivation

After the more general consideration of spatial heterogeneity and territoriality in Chapters 4 and 5, we now return to the fundamental idea of the Resource Dispersion Hypothesis, or R.D.H.. The models we have discussed so far have the following form. Firstly, we describe how an individual animal or a mated pair of animals (referred to as primary animals) set up a territory in a given habitat. Secondly, we ask if one or more further animals (referred to as secondary animals) can share the territory, in such a way that there is no extra cost to the primaries, and if so, what probability of survival do the secondaries have? The true underlying question, however, is whether animals will actually behave in this way, i.e. whether groups will ever be formed in the manner suggested by the R.D.H..

To answer this fundamental question, we need to model the decision making process of the animals, or the mechanism which determines their behaviour. We must also determine the consequences, for an individual, of alternative patterns of behaviour, bearing in mind that those consequences will in general depend on the behaviour of other individuals, as well as on environmental and random factors.

6.1.2 Modelling Individual Behaviour

In describing the behaviour of an individual animal, we will use the concept of a strategy, as used in game theory (see von Neumann and Morgenstern, 1944). A strategy is a description of the actions which an individual would take (and their probabilities, if random actions are allowed) in each possible situation in which the individual may find itself.

Different individuals might adopt different strategies for a variety of reasons: because of environmental, genetic, or other, perhaps random, factors, or any combination of these. For the purposes of this thesis, we will assume that an individual's strategy is inherited, through some genetic mechanism. The details of some possible mechanisms are discussed in Section 6.1.4. Such an assumption is widespread in the literature concerning the theoretical basis of behaviour. A number of references will be given in the course of this chapter.

6.1.3 Modelling the Evolutionary Process

Given the assumption of inherited strategies, it follows that the consequences of different strategies, and the answer to the question which motivates this chapter, depend on the long-term reproductive success of different members of the population. Determining these levels of reproductive success is a major part of modelling any particular situation, and is considered in Sections 6.3, 6.4 and 6.5. Once these levels of success are known, we can attempt to determine the final numbers of individuals adopting different strategies in a population, and in particular to determine which strategy or strategies will be successful, and therefore present in the population, in the long term.

In our models, and in many other models concerned with behavioural questions, the principal factors affecting an individual's success are its own strategy and the composition, by strategy, of the population of which it is a member. The study of such systems is known as evolutionary game theory.

One well-established approach to such problems is to use the concept of an evolutionarily stable strategy, or E.S.S., pioneered by Maynard Smith and Price (1973) (see also Maynard Smith, 1982, and Hines, 1987, for reviews). We will use the notation of Taylor and Jonker (1978) to describe the basic ideas of E.S.S. theory in the usual case where any strategy can be expressed as a stochastic mixture of a finite number n of pure (i.e. non-random) strategies. The state of the population is described by a probability vector $\boldsymbol{p}$, with p_i being the overall probability of strategy i being played by a randomly selected individual. The population state space is then

$$K = \{\boldsymbol{p} : \sum_{i=1}^{n} p_i = 1, p_i \geq 0, i = 1,\ldots,n\}.$$

We measure the reproductive success of an individual playing pure strategy i in a population in state $\boldsymbol{p}$ by its fitness $F(i|\boldsymbol{p})$. We assume that the fitness of an individual playing a mixed strategy $\boldsymbol{q}$, i.e. playing strategy i with probability q_i, is given by

$$F(\boldsymbol{q}|\boldsymbol{p}) = \sum_{i=1}^{n} q_i F(i|\boldsymbol{p}).$$

The key concept is that of an E.S.S. .

Definition. A state $\boldsymbol{p}$ is called an **E.S.S.** if for every state $\boldsymbol{q} \in K \setminus \{\boldsymbol{p}\}$, if we let $\bar{\boldsymbol{p}} = (1-\varepsilon)\boldsymbol{p}+\varepsilon\boldsymbol{q}$, then $F(\boldsymbol{q}|\bar{\boldsymbol{p}}) < F(\boldsymbol{p}|\bar{\boldsymbol{p}})$ for all sufficiently small $\varepsilon > 0$.

The definition of an E.S.S. contains no reference to the dynamics of the population. Taylor and Jonker (1978) explored ways of defining a dynamic on the state space K, and compared the stable equilibria of the resulting dynamical system with the corresponding E.S.S.s. In particular, they considered two natural dynamics: the continuous time dynamic given by

$$\dot{s}_i = s_i[F(i|s) - F(s|s)]; \tag{6.1.1}$$

and the discrete time dynamic given by

$$\Delta s_i = s_i \frac{[F(i|s) - F(s|s)]}{F(s|s) + 1}. \tag{6.1.2}$$

They found that under certain regularity conditions, E.S.S.s are stable points of the continuous time dynamic system defined by equation (6.1.1), but that the result does not extend to the discrete time system defined by equation (6.1.2). Perhaps more importantly, the converse does not hold for either dynamic, so there generally may be stable equilibria which are not E.S.S.s. These ideas were also explored by Rowe, Harvey and Hubbard (1985), and Zeeman (1979, 1981). Several of these papers discuss the consequences of the difference between stable points and E.S.S.s. In practice, however, the E.S.S. approach has become widely established.

For the purposes of this chapter, we do not use the E.S.S concept. The principal reason is that the particular phenomenon of current interest is the social grouping of the population. Social grouping can affect the reproduction and mortality of individuals in the population, and so the state of the population must record such social grouping, as well as the numbers of individuals inheriting different strategies. Furthermore, to determine the dynamics of such a system, it is not sufficient to know the fitness of each strategy given a particular state: we need to know the rates of formation and break-up of social groups. Thus an E.S.S. approach would not give a complete description of the system of interest. In fact, we avoid the use of any concept of fitness: instead, to define a model, we will directly write down a dynamic on the extended state space in which both social grouping and individual strategy are recorded. We will then look for stable equilibria of that dynamic, avoiding the uncertainty of interpretation of an E.S.S..

6.1.4 Genetics

To describe the genetic mechanisms which we wish to consider, we need some basic biological definitions (see for example King, 1972). A locus is a particular location on a chromosome, i.e. the location of a particular gene. An allele is a particular form of a given gene. The genotype of an individual refers to the alleles present in that individual, whereas the phenotype of an individual refers to the observable properties of the individual, such as behaviour, appearance, etc.. We will normally assume that a given aspect of phenotype is controlled by a single locus or a small number of loci, so that, for example, which strategy an individual uses is determined by which alleles are present at the relevant locus or loci. The justification for such an assumption is discussed by Grafen (1984): in practice, it is a very common assumption in theoretical studies, made either implicitly or explicitly (e.g. Feldman and Eshel, 1982, Cressman and Dash, 1985, Thomas, 1985). New alleles, and hence new

phenotypes, will occur as a result of rare mutations, independently of the current state of the population.

There are two main ways in which alleles can determine phenotype, depending on the number of alleles at each locus, which is essentially constant for any given species.

The simplest case is a haploid system. Each individual has a single allele at each locus, so that if a particular behavioural feature is controlled by a single locus, an individual will have a single allele, indicating one particular phenotype, at the relevant locus. In this system, individuals are not classed into sexes: each individual reproduces separately, and in the absence of mutation, passes on copies of all its alleles to each of its offspring. This is the easiest genetic system to analyse, and for that reason is widely used in theoretical models. In particular, it is an implicit assumption in the original work, and much of the subsequent work, on E.S.S.s (Maynard Smith and Price, 1973, Maynard Smith, 1982, Hines, 1987). Grafen (1984) gives an extensive discussion of the status of this assumption, with particular reference to behavioural models.

The alternative is a diploid system. This is known to be the true underlying mechanism in most species of behavioural interest, but it is more complicated than the haploid case. In a diploid species, each individual has two alleles at each locus, and these both affect phenotypic characteristics of the individual, such as strategy choice. The interaction of the alleles at a locus may be completely general, although the alleles are not thought of as ordered, so that n possible alleles lead to $n(n+1)/2$ possible phenotypes (e.g. Rowe, 1988). Often there will be fewer distinct phenotypes. For example, if there are two possible alleles A and a at a locus, then the genotypes $Aa (=aA)$ and AA might lead to the same phenotype, distinct from that given by aa, in which case allele A is said to be dominant, and allele a recessive. Diploid species reproduce sexually: at a given locus, each offspring will have one allele from each of its parents, each one chosen at random from the two alleles that the parent has at that locus. The selections from the two parents are independent, but the selections at two different loci from the same parent are not necessarily independent. Thus genetic systems in which the important aspects of the phenotype are controlled by more than one locus in a diploid species can be very complicated.

6.1.5 Biological Concepts: the Dynamics of Social Groups

Having discussed the general approach that we will adopt to evolutionary questions, we now explain the additional biological ideas needed to define the required dynamical system.

Dispersal

In the absence of social behaviour, it is assumed that juvenile animals will disperse from the territory in which they were born into the surrounding habitat, and will attempt to establish

territories of their own, though their probability of success may not be very high.

Group Formation

We will assume that the social groups of interest are formed by juveniles remaining in the territory of their birth, rather than dispersing. Potentially, the decision that this takes place will be made jointly by parent(s) and young. Note that all groups formed by this mechanism will consist of relatives. This reflects observations on a number of species (Macdonald, 1983), and is a common assumption in the theoretical literature, for example in the models of Macdonald and Carr (1989), Lindström (1986) and Emlen (1982).

Breeding

If a territory is occupied by a primary mated pair of animals and one or more secondaries, the usual assumption is that only the primary pair take part in breeding. Under the resource dispersion hypothesis, we are particularly interested in the case where the reproductive success of the primaries is unaffected by the secondaries, but we can also consider the case where it is decreased (interference) or increased (helping) (Macdonald, 1983).

Mortality

We assume that once a group is formed, it remains as a group until the death of one or more of its members. Mortality may be independent for all individuals (Macdonald and Carr, 1989), or it may reflect the different (and interdependent) mortality rates for primaries and secondaries suggested by the Resource Dispersion Hypothesis (see Lindström (1986) for an extreme version of this case, which will be discussed later).

Alternatives to this assumption are possible: for example, Emlen (1982) considers secondary animals which automatically leave the group after one year, with an increased probability of successful dispersion due to their experience.

Inheritance

The final stage in population dynamics concerns the surviving individuals from a group in which one of the members has died. If a secondary has died, it can generally be replaced. If one of a pair of primaries has died, and the survivor can remate, then the survivor and its new mate will become the new primary pair. If a surviving primary does not remate, if both primaries die or if we are considering a population in which only one sex is territorial (or a haploid population in which there is only one sex), then the primary position in the territory becomes vacant. In the absence of secondary animals, the territory would then be available for settlement by a dispersing juvenile from another territory. If secondary animals are present, however, one of them is assumed to inherit the territory, and is thus 'promoted' to primary status. This concept of territory inheritance is used by Macdonald and Carr (1989), and is the key concept in the model proposed by Lindström (1986) as an alternative to the R.D.H..

6.2 Previous Work

Macdonald and Carr (1989) present a model which has the same objective as the present work, but which uses a rather different approach. The authors assume a diploid population in which only females can become secondaries, and only the female primary can influence the behaviour of potential secondaries. They then calculate the fitness of both the primary and secondary females, under the two possible strategies of remaining or dispersing available to the secondary, and under different assumptions about the number of female offspring present. However, for a decision by a secondary animal at a given time, Macdonald and Carr consider only the fitness due to offspring born (to either the primary female or an inheriting or successfully dispersing juvenile) in the next breeding season: they do not consider the fitness due to animals which may survive and breed again, even though the relevant probabilities may be affected by strategy choices. Similarly, Macdonald and Carr consider only the formation, and not the survival or dissolution, of groups. Thus they do not have any implicit or explicit dynamic for the social organisation of the population. The models presented below are intended to rectify that problem, and to avoid some of the technicalities involved in calculating fitness in a complex social structure (see the discussion in Grafen (1984) and Dawkins (1982)). Some specific conclusions from Macdonald and Carr (1989) are compared with results from our dynamic model in Section 6.5.

A closely related paper is that of Emlen (1982). As mentioned above, however, he assumes that the two strategies available to a juvenile are to disperse immediately, or to disperse after one year spent as a helper. Thus the groups in Emlen's model last only one year, by assumption.

Finally, a paper by Lindström (1986) considers a possible alternative to the R.D.H., which Lindström calls the Territory Inheritance Hypothesis (T.I.H.). The T.I.H. states that groups form because the fitness of a juvenile which remains in its natal territory, in the hope of inheriting that territory, is higher than the fitness of a juvenile which disperses in the hope of establishing a territory of its own. Lindström describes a mathematical model which formalises the T.I.H., studying only a single sex (as in Macdonald and Carr, 1989). He assumes that exactly one juvenile is born in each territory in each year, and considers different population growth rates, and two different sizes of territory. We consider separately the two cases corresponding to these two territory types.

Single-territories.

Lindström defines a single-territory to be a "territory [which] is the optimal size for one adult individual only, of the studied sex". Implicit in this definition seems to be the property that if a juvenile chooses to remain in such a territory, it is not possible for both the juvenile and the occupying adult to survive the subsequent 'winter' i.e. the mortality phase of the life-

cycle. We will regard this property as the characterisation of a single-territory. Lindström obtains two different results for such territories, depending on the overall growth rate in the population. If the population is strictly increasing, i.e. new territories are becoming available, then dispersal is the better strategy. If the population is non-increasing, then remaining and dispersing are equally good strategies. The latter, rather surprising result is not further explored by Lindström, but can be seen to be a (clearly unstable) consequence of two of his assumptions, neither of which has any real justification. The first is that if a primary adult dies, a juvenile which has remained in the territory will always inherit the territory, i.e.

$$\Pr(\text{juvenile dies} \mid \text{adult dies}) = 0.$$

Thus mortality is clearly not independent, but nor does it fit the pattern one would expect if survival was resource-limited, nor even the idea of competition between the two individuals, since the mortality probability of the adult is unaffected by the presence of the juvenile. The second assumption by Lindström that leads to the equivalence of the 'remaining' and 'dispersing' strategies is that each territory produces exactly one juvenile of the studied sex each year. Any change in the number produced would alter the survival probability of dispersing juveniles, and thus give an advantage to one strategy or the other. In Section 6.5, we will use our model to explore the effect of these assumptions, in the particular case of a steady population size.

Double-territories

A double-territory is a territory which is larger than those in the rest of the habitat, and which is "the optimal size for two adult individuals of the same sex" (Lindström, 1986). The key property of such a territory is that it gives some probability of two adults surviving in the territory for one or more complete years. Lindström shows that in such a territory, remaining is better than dispersing, provided adult survival is sufficiently high, and the population is not increasing too quickly. Thus Lindström's model predicts the formation of groups of two adults in double-territories. Note however that this result is only obtained by **assuming** the existence of a double-territory, i.e. a territory set up by a primary individual of the studied sex, but which can support two adults of that sex over the whole year. There are two ways in which this can arise.

Firstly, such a territory may arise naturally, i.e. it may be the smallest territory which is viable for the primary. In this case, the T.I.H. exploits the existence of the double-territory, but does **not** explain it. Note, however, that these are exactly the sort of territories described by the R.D.H., and explored in the models presented by Carr and Macdonald (1986), and in earlier chapters of this thesis. So Lindström's model with a double-territory which is a minimum territory for a primary is equivalent to a model based on the R.D.H., with territory

inheritance as a mechanism by which secondary animals may eventually reproduce. However, models of the latter type, such as that of Macdonald and Carr (1989), have the advantage of explaining, rather than assuming, the existence of a double-territory. Thus, given the first interpretation of a double-territory, territory inheritance seems to be not an alternative to the R.D.H., but a different stage of the same process. The T.I.H., in this version, needs some explanation for double-territories, such as the R.D.H.; the R.D.H. needs some mechanism whereby secondary animals contribute to the survival of their own genes, such as dispersing after some time in a group (Emlen, 1982), 'helping' primary animals, territory inheritance, or some combination of these.

A second, very different way in which Lindström's double-territories may arise is by primaries deliberately occupying larger territories than necessary. In this case,

> "it should ... be noted that the model does not take [into account] the cost of taking up a territory large enough for a group" (Lindström, 1986).

In the absence of effects such as resource dispersion, the extra cost involved could be considerable, and would act against the formation of groups under the T.I.H., making Lindström's conclusions less plausible. It is also important to note that, under this second interpretation of a double-territory, the two strategies being compared are not

> "disperse as a juvenile"

versus

> "remain in natal territory as a juvenile",

but rather

(i) "disperse as a juvenile, then occupy a single-territory as an adult"

versus

(ii) "remain in natal territory as a juvenile, then occupy a double-territory as an adult".

Strategy (ii) has a number of disadvantages. In the very early stages of the establishment of the strategy, juveniles following it may be selected against, if they are 'remaining' in a territory which is not a double-territory. The above argument only holds if the strategy (ii) is thought of as a single step away from strategy (i), the default strategy, but if that is not the case, then there must be some intermediate strategy, such as

> "disperse as a juvenile, then occupy a double-territory as an adult"

or

> "remain in natal territory as a juvenile, then occupy a single-territory as an adult".

But in general, either of these intermediate strategies would also be selected against.

The complexity of the step from strategy (i) to strategy (ii) above indicates another disadvantage of the second version of the T.I.H. as compared with the R.D.H.. The T.I.H. requires that individuals defend larger than optimal territories, and immediately obtain sufficient benefit, through territory inheritance by relatives, to outweigh the extra cost of defence. The R.D.H., on the other hand, suggests that the minimum territories, which the animals must defend, allow the formation of groups, in a single, simple evolutionary step, serving as an "evolutionary catalyst for group living" (Macdonald, 1983). Such groups would then be able to acquire naturally the other possible advantages of group living, perhaps, in a separate evolutionary step, leading to larger than minimum territories (see Kruuk and Macdonald, 1985). The idea that the R.D.H. may form part of the evolutionary history of species to which it does not currently apply is discussed at length by Macdonald (1983).

6.3 A Deterministic Haploid Model.

Having discussed the aims, results and problems of previous models, we wish to construct an alternative model which can answer the questions of interest, but which avoids some of the above problems. We start off by trying to formulate the simplest possible model which incorporates all the essential features of resource-based social behaviour. Hence we initially consider a haploid population, as defined in Section 6.1.4. We assume that there are two types of animals, corresponding to two different strategies: the first type live exclusively as individuals, while the second type sometimes form rudimentary social groups, which have a maximum size of two (one primary animal and one secondary). It is assumed that such a group is formed by a juvenile choosing to remain in its natal territory, and any potential parent-offspring conflict is ignored. All animals are assumed to be territorial, and all territories identical, notwithstanding the conclusions of Section 2.3 above. Finally, we assume an infinitely large population, occupying a constant number of territories, with deterministic dynamics. Although we may describe events at the level of the individual stochastically, we shall simply take expectations when determining the behaviour of the whole system. The corresponding stochastic model will be considered in Section 6.4. We consider the system in discrete time, to correspond with the systems described by Macdonald and Carr (1989) and Lindström (1986). We assume the existence of some initial population in which all animals use the first strategy and live individually, and then consider whether a small number of individuals adopting the second strategy can establish themselves.

The initial individuals have the following life cycle.

(i) At time t, each individual reproduces, with the jth individual having $Y_{j,t}$ surviving young, and with $y_0 = \mathrm{E}[Y_{j,t}]$ for all j, t.

(ii) Each adult then has probability m of dying, thus leaving a vacant territory.

(iii) The territories vacated in stage (ii) are settled by randomly selected juveniles from stage (i) (it is assumed that there are always enough juveniles to occupy all such vacancies). All other juveniles die.

(iv) The surviving adults from stage (ii) and the juveniles which obtained territories in stage (iii) go on to become the reproducing adults at time $t+1$.

Clearly such a system will be in equilibrium. We perturb the system by replacing a proportion of animals with individuals of a second type, which have the following modified life cycle.

(i)′ With probability π_1, a juvenile of this type will attempt to remain in its natal territory, and will succeed in doing so provided that its presence does not increase the group size in that territory above 2.

In a territory which is already occupied by a group of 2 animals at stage (i)′ of the cycle, only the elder of the animals reproduces, but because of possible interference or helping from the younger animal, the expected number of young produced is now y_1.

(ii)′ There are four possible outcomes for a group of 2 animals at the mortality stage in the cycle, as follows:

(a) the juvenile dies but the adult survives, with probability p_α;

(b) the adult dies, but the juvenile survives and inherits the territory, with probability p_β;

(c) both animals survive, and retain the territory as a group, with probability p_γ; or

(d) both animals die and the territory is vacated, with probability p_δ.

Clearly $p_\alpha+p_\beta+p_\gamma+p_\delta = 1$.

(iii)′ Those young which did **not** remain in their natal territories have the same chances of settling vacant territories as young of the initial type.

(iv)′ Surviving groups from (ii)′(c) go on to reproduce in the next period as described in (i)′; surviving individuals from (ii)′(a) and (ii)′(b), and successful juveniles from (iii)′ go on to reproduce in the next period in the same way as individuals of the initial type in (i).

We now wish to determine the dynamics of a population containing individuals of both types. Define $x_{1,t}$ to be the **proportion** of territories occupied by individuals of the initial type at time t, $x_{2,t}$ to be the proportion of territories occupied by individuals of the second type at time t, and $x_{3,t}$ to be the proportion occupied by groups of two animals of the second type.

Note: it is important to distinguish between the classifications of individuals (two types, or strategies: type 1 is the initial type, with no grouping, and type 2 is the 'invading' type, showing some grouping) and of territories (three types, corresponding to the definitions of x_1, x_2, and x_3).

At time t, the numbers of young produced, expressed like x_1 etc., as proportions of the total number of territories, are as shown in Table 6.1.

Table 6.1

Numbers of Young Produced

From territory type:	1	2	3
Young of type 1	$x_1 y_0$	0	0
Young of type 2	0	$x_2 y_0$	$x_3 y_1$

In type 2 territories, there is a chance that a juvenile will remain to form a group. Each juvenile has probability π_1 of attempting to do so, and there are $Y_{j,t}$ juveniles in the jth territory. We assume that exactly one of these will remain, unless none attempt to do so, in which case none will remain. Hence the probability of exactly one juvenile remaining is

$$\pi = 1 - \mathrm{E}[(1-\pi_1)^{Y_{1,1}}], \tag{6.3.1}$$

since $Y_{j,t}$ is identically distributed for all j and t. Note that π depends on the distribution of $Y_{j,t}$, not just on its expectation. However, since π_1 only enters the model through the above equation for π, we shall not use the exact form of their relationship, but instead ignore π_1 and treat π as a parameter of the model. Since the number remaining cannot be greater than the number attempting to do so, taking expectations gives

$$0 \leq \pi \leq \min\{1, y_0\}.$$

Thus, after stage (i), we have individuals and groups as shown in Table 6.2.

Table 6.2

Numbers of Individuals before Mortality

From territory type:	1	2	3
Type 1 adults	x_1	0	0
Type 2 adults	0	$x_2(1-\pi)$	0
Type 2 groups	0	$x_2\pi$	x_3
Type 1 young	$x_1 y_0$	0	0
Type 2 young (dispersing)	0	$x_2(y_0-\pi)$	$x_3 y_1$

Next, we must consider the mortality of adults and groups in stage (ii) of the life cycle. Bearing in mind the outcomes and probabilities given in (ii) and (ii)′ above, we obtain the expressions in Table 6.3.

Table 6.3

Numbers of Individuals after Mortality

Original type of territory	1	2	3
Type 1 adults	$x_1(1-m)$	0	0
Type 2 adults	0	$x_2\{(1-\pi)(1-m)+\pi(p_\alpha+p_\beta)\}$	$x_3(p_\alpha+p_\beta)$
Type 2 groups	0	$x_2\pi p_\gamma$	$x_3 p_\gamma$
Type 1 young	$x_1 y_0$	0	0
Type 2 young	0	$x_2(y_0-\pi)$	$x_3 y_1$
Vacant territories	$x_1 m$	$x_2\{(1-\pi)m+\pi p_\delta\}$	$x_3 p_\delta$

Finally, we need to calculate the proportions of vacancies resettled by individuals of the different types. We assume that there are always enough young to settle all vacant territories: sufficient conditions would be

$$y_0 \geq m, \quad y_0-\pi \geq (1-\pi)m+\pi p_\delta, \quad \text{and} \quad y_1 \geq p_\delta;$$

$$y_0-\pi, y_1 \geq m, p_\delta; \quad \text{or}$$

$$y_0-\pi \geq m \geq p_\delta \leq y_1.$$

Assuming that some such condition holds, the numbers of vacancies settled by type 1 and type 2 individuals will be

$$s_1 = \frac{d_1 v}{d} \quad \text{and} \quad s_2 = \frac{d_2 v}{d}$$

respectively, where d_1, d_2 are the numbers of dispersing young of the two types, d is the total number of dispersing young, and v is the total number of vacancies.

From Table 6.3 we have

$$d_1 = x_1 y_0,$$

$$d_2 = x_2(y_0-\pi)+x_3 y_1,$$

$$v = x_1 m+x_2\{(1-\pi)m+\pi p_\delta\}+x_3 p_\delta$$

and clearly

$$d = d_1 + d_2.$$

Hence we have

$$s_1 = \frac{[x_1 m + x_2\{(1-\pi)m + \pi p_\delta\} + x_3 p_\delta]x_1 y_0}{x_1 y_0 + x_2(y_0 - \pi) + x_3 y_1},$$

$$s_2 = \frac{[x_1 m + x_2\{(1-\pi)m + \pi p_\delta\} + x_3 p_\delta](x_2(y_0 - \pi) + x_3 y_1)}{x_1 y_0 + x_2(y_0 - \pi) + x_3 y_1}.$$

Collecting together terms from stages (ii) and (iii), we have

$$\begin{aligned} x_{1,t+1} &= x_{1,t}(1-m) + s_{1,t} \\ &= x_{1,t}\{y_0 x_{1,t} + y_0(1 - m - \frac{\pi(1-m)}{y_0} + (1-\pi)m + \pi p_\delta)x_{2,t} \\ &\quad + ((1-m)y_1 + p_\delta y_0)x_{3,t}\}(x_{1,t}y_0 + x_{2,t}(y_0 - \pi) + x_{3,t}y_1)^{-1}, \end{aligned} \tag{6.3.2}$$

$$\begin{aligned} x_{2,t+1} &= x_{2,t}\{(1-\pi)(1-m) + \pi(p_\alpha + p_\beta)\} + x_{3,t}(p_\alpha + p_\beta) + s_{2,t} \\ &= \{x_{1,t}x_{2,t}[m(y_0 - \pi) + y_0\{(1-\pi)(1-m) + \pi(p_\alpha + p_\beta)\}] \\ &\quad + x_{1,t}x_{3,t}[y_0(p_\alpha + p_\beta) + y_1 m] \\ &\quad + x_{2,t}^2(y_0 - \pi)(1 - \pi p_\gamma) \\ &\quad + x_{2,t}x_{3,t}[(y_0 - \pi)(1 - p_\gamma) + y_1(1 - \pi p_\gamma)] \\ &\quad + x_{3,t}^2 y_1(1 - p_\gamma)\}\ (x_{1,t}y_0 + x_{2,t}(y_0 - \pi) + x_{3,t}y_1)^{-1}, \end{aligned} \tag{6.3.3}$$

and

$$x_{3,t+1} = x_{2,t}\pi p_\gamma + x_{3,t}p_\gamma. \tag{6.3.4}$$

Given equations (6.3.2), (6.3.3) and (6.3.4), what can we say about the behaviour of the system? Firstly, consider a possible equilibrium point of the system, (x_1, x_2, x_3) say. From equation (6.3.4) we must have

$$\begin{aligned} x_3 &= x_2 \pi p_\gamma + x_3 p_\gamma \\ \Rightarrow x_3 &= \left(\frac{\pi p_\gamma}{1 - p_\gamma}\right)x_2 = kx_2 \end{aligned} \tag{6.3.5}$$

say. The possible 'trivial' equilibria of the system, with only one type of individual present, are

$$x_1 = 1, \quad x_2 = x_3 = 0, \tag{6.3.6}$$

and

$$x_1 = 0, \quad x_2 = \frac{1}{1+k}, \quad x_3 = \frac{k}{1+k}. \tag{6.3.7}$$

To find possible non-trivial equilibria, we substitute equation (6.3.5) into equations (6.3.2) and (6.3.3), and then solve (6.3.2) and (6.3.3) simultaneously, subject to the conditions

$$x_{1,t+1} = x_{1,t}, \quad x_{2,t+1} = x_{2,t}.$$

After substitution we obtain

$$x_1 = \frac{x_1\{y_0x_1 + ax_2\}}{x_1y_0 + x_2(y_0 - \pi + ky_1)},$$

$$x_2 = \frac{x_2\{bx_1 + cx_2\}}{x_1y_0 + x_2(y_0 - \pi + ky_1)},$$

where

$$a = y_0(1 - m - \frac{\pi(1-m)}{y_0} + (1-\pi)m + \pi p_\delta) + k((1-m)y_1 + p_\delta y_0),$$

$$b = m(y_0 - \pi) + y_0((1-\pi)(1-m) + \pi(p_\alpha + p_\beta)) + k(y_0(p_\alpha + p_\beta) + y_1 m),$$

$$c = (y_0 - \pi)(1 - \pi p_\gamma) + k((y_0 - \pi)(1 - p_\gamma) + y_1(1 - \pi p_\gamma)) + k^2 y_1(1 - p_\gamma).$$

Since for a non-trivial equilibrium we require $x_1, x_2 > 0$, we have equilibrium if and only if

$$\frac{y_0x_1 + ax_2}{y_0x_1 + (y_0 - \pi + ky_1)x_2} = \frac{bx_1 + cx_2}{y_0x_1 + (y_0 - \pi + ky_1)x_2} = 1$$

$$\Leftrightarrow y_0x_1 + ax_2 = y_0x_1 + (y_0 - \pi + ky_1)x_2 = bx_1 + cx_2. \tag{6.3.8}$$

A necessary condition for non-trivial equilibrium is therefore

$$a = y_0 - \pi + ky_1$$

$$\Leftrightarrow y_0\left(1 - \frac{p_\delta}{m}\right) + p_\gamma(1+w) = 1, \tag{6.3.9}$$

where we define

$$w = y_1 - y_0$$

to be the change in mean number of young produced due to the presence of a secondary animal. Clearly equation (6.3.9) will not generally be satisfied for a given set of parameters m, p_γ, p_δ, y_0, y_1, and so we conclude that generally there is no non-trivial equilibrium.

When equation (6.3.9) is satisfied, we have

$$b-y_0 = \frac{m\pi}{1-p_\gamma}\left(y_0\left\{1-\frac{p_\delta}{m}\right\}+p_\gamma(1+w)-1\right)$$

$$= 0.$$

In addition, for any parameter values, we have

$$c-(y_0-\pi+ky_1) = (y_0-\pi+ky_1)(\pi p_\gamma-\pi p_\gamma) = 0.$$

Hence from condition (6.3.8), whenever equation (6.3.9) is satisfied we have an equilibrium at any point (x_1,x_2,x_3) satisfying equation (6.3.5),

$$x_3 = kx_2.$$

From equation (6.3.9), however, it is clear that the equilibrium at any of these points could be removed by an arbitrarily small change in the model parameters, i.e. these equilibria are not structurally stable. For a formal definition and discussion of structural stability, see e.g. Hirsch and Smale (1974) or Zeeman (1981), the latter paper being specifically concerned with the dynamics of evolutionary systems.

On the other hand, the equilibria defined by equations (6.3.6) and (6.3.7) will always exist, with the position of the latter depending on the model parameters through the expression for k. Hence for the current system, we can confine our attention to these two equilibria.

We need to determine the stability of the equilibria defined by equations (6.3.6) and (6.3.7). It is not sufficient to merely consider these points as equilibria of the 3-dimensional system defined by equations (6.3.2), (6.3.3) and (6.3.4), since they might for instance be unstable as equilibria of that 3-dimensional system, but stable within the state space currently of interest, the simplex defined by

$$x_1+x_2+x_3 = 1, \qquad x_1,x_2,x_3 \geq 0. \tag{6.3.10}$$

Hence we use a simple reparameterisation of the form

$$z_1 = x_i, \qquad z_2 = x_j,$$

which can be inverted by noting that

$$x_l = 1-z_1-z_2,$$

using equation (6.3.10), where (i,j,l) is some permutation of $\{1,2,3\}$. Such a reparameterisation maps the simplex defined by equation (6.3.10) onto the triangular region in the z_1,z_2-plane with vertices at $(0,0)$, $(0,1)$, and $(1,0)$.

To consider the stability of $x = (1,0,0)$ we let

$$\phi_1 = x_2, \qquad \phi_2 = x_3 \tag{6.3.11}$$

so that the equilibrium point is mapped to $\phi = 0$. Then a sufficient condition for stability of the equilibrium is that $|\lambda| < 1$ for any eigenvalue λ of A, and a sufficient condition for instability is that $|\lambda| \geq 1$ for some eigenvalue λ of A whose eigenspace intersects the state space of the process, where A is defined by

$$a_{ij} = \left.\frac{\partial f_i}{\partial \phi_j}\right|_0,$$

and f_i is such that

$$\phi_{i,t+1} = f_i(\phi_t), \qquad i = 1,2 \tag{6.3.12}$$

The forms of f_1, f_2 can be calculated from equations (6.3.2), (6.3.3) (6.3.4) and (6.3.12), recalling that the reparameterisation (6.3.11) implies

$$x_1 = 1 - \phi_1 - \phi_2.$$

Thus we can calculate A, obtaining

$$A = \begin{bmatrix} m\left(1-\frac{\pi}{y_0}\right)+\pi(p_\alpha+p_\beta)+(1-\pi)(1-m) & p_\alpha+p_\beta+m\frac{y_1}{y_0} \\ \pi p_\gamma & p_\gamma \end{bmatrix} \tag{6.3.13}$$

Except in the pathological case when π or p_γ is zero, we can show that A^2 is strictly positive (i.e. all its elements are strictly positive). Hence we can apply the Perron–Frobenius theorem (see e.g. Karlin and Taylor, 1975).

Theorem 6.1.

Let A be a matrix such that for some integer m, A^m is strictly positive. Then there exists a simple eigenvalue λ_0 of A such that $|\lambda| < \lambda_0$ for any other eigenvalue λ of A, and the eigenvector associated with λ_0 may be taken to be strictly positive.

Proof: See Karlin and Taylor (1975).

Since the eigenvector associated with λ_0 is positive, the eigenspace of λ_0 always intersects the state space of our process in the ϕ_1, ϕ_2-plane. Hence, from the sufficient conditions described above, and the fact that $|\lambda| < \lambda_0$, we have:

$$\lambda_0 < 1 \Rightarrow \text{ stable equilibrium, and } \lambda_0 \geq 1 \Rightarrow \text{ unstable equilibrium.}$$

The characteristic equation of A is

$$|A - \lambda I| = 0 \Leftrightarrow \lambda^2 + b\lambda + c = 0,$$

where

$$b = -\left(1+p_\gamma+\pi\left\{p_\alpha+p_\beta+m-1-\frac{m}{y_0}\right\}\right)$$

$$c = p_\gamma\left(1-\pi-\frac{m\pi}{y_0}\{1+y_1-y_0\}\right).$$

So the Perron–Frobenius eigenvalue is

$$\lambda_0 = \frac{-b\pm\sqrt{b^2-4c}}{2},$$

and the condition for stability is

$$\lambda_0 < 1 \Leftrightarrow \sqrt{b^2-4c} < b+2 \Leftrightarrow b^2-4c < b^2+4b+4,$$

since $b+2 \geqslant 0$,

$$\Leftrightarrow c > -b-1$$

$$\Leftrightarrow y_0\left(1-\frac{p_\delta}{m}\right)+p_\gamma(1+w) < 1 \qquad (6.3.14)$$

after some algebra.

At $x = \left(0, \frac{1}{1+k}, \frac{k}{1+k}\right)$, we use the parameterisation

$$\psi_1 = x_1, \qquad \psi_2 = x_3,$$

so that the equilibrium is mapped to $\left(0, \frac{k}{1+k}\right)$. Then the equilibrium is stable if and only if $|\lambda| < 1$ for any eigenvalue λ of B, where B is defined by

$$b_{ij} = \left.\frac{\partial g_i}{\partial \psi_j}\right|_{\left(0, \frac{k}{1+k}\right)}$$

and each g_i is such that

$$\psi_{i,t+1} = g_i(\psi_t).$$

Note that we need not consider the eigenvectors of B, since the equilibrium point $\left(0, \frac{k}{1+k}\right)$ is not at a vertex of the process state space. We obtain

$$B = \begin{bmatrix} [y_0(1-m+(1-\pi)m+\pi p_\delta)-\pi(1-m)+k((1-m)y_1+p_\delta y_0)](y_0+ky_1-\pi)^{-1} & 0 \\ -\pi p_\gamma & p_\gamma(1-\pi) \end{bmatrix},$$

with the eigenvalues displayed on the diagonal. Clearly in non-pathological cases $p_\gamma(1-\pi) < 1$, so the equilibrium is stable if and only if

$$|[y_0(1-m+(1-\pi)m+\pi p_\delta)-\pi(1-m)+k((1-m)y_1+p_\delta y_0)](y_0+ky_1-\pi)^{-1}| < 1$$

$$\Leftrightarrow y_0\left(1-\frac{p_\delta}{m}\right)+p_\gamma(1+w) > 1 \qquad (6.3.15)$$

after rearranging.

Thus the behaviour of the deterministic model described in this section is summarised by the parameter

$$C = y_0\left(1-\frac{p_\delta}{m}\right)+p_\gamma(1+w).$$

If $C > 1$, the point $(1,0,0)$ is unstable, so users of the second strategy (forming social groups in accordance with the R.D.H.) can invade, and the point $(0,(1+k)^{-1},k(1+k)^{-1})$ is stable, so a population consisting entirely of users of the second strategy cannot be invaded. If $C < 1$, the first strategy (no social groups) can always invade, and cannot be re-invaded. Finally, if $C = 1$, there is an entire line of equilibrium points, but the system is structurally unstable. In no case is there a structurally stable equilibrium in which both strategies are present.

The above results completely describe the dynamics of the system provided we assume that the only attractors of the system are simple equilibria. While this assumption is difficult to prove for such a discrete-time system, it is supported by the results of numerical investigation.

6.4 A Stochastic Model.

The model above assumes deterministic population dynamics, which will clearly not be true of any real, finite population. We will now consider a stochastic model. In all other respects, the model parallels the previous one, and so the description will be more concise.

We need one extra piece of information, which is the distribution of the numbers of offspring for individuals and groups. The simplest choice in some respects would be to assume Poisson distributions, because of the extra independence it would give in the model. However, this would mean that there was always a positive probability of some vacant territories remaining unfilled, and so the population would eventually become extinct. While such a model might in some ways be realistic, its use in answering questions about strategy choice would be difficult. Instead, we constrain the stochastic model, like the deterministic one, so that all vacant territories are filled. The simplest way to do this is to make $Y_{j,t}$ and $Y^*_{j,t}$, the numbers of young produced by solitary individuals and groups respectively, deterministic, so that

$$\Pr(Y_{j,t} = y_0) = 1,$$

$$\Pr(Y^*_{j,t} = y_1) = 1,$$

where y_0 and y_1 are now constrained to be integers, with $y_0 \geq 2$, $y_1 \geq 1$. Note that we now have

$$\pi = 1-(1-\pi_1)^{y_0}. \tag{6.4.1}$$

Having decided on the offspring distributions, we can now write down the dynamics of the stochastic model. Let N be the fixed number of occupied territories. Let $X_{1,t}^N$ be the number of territories occupied by individuals using strategy 1, at time t, $X_{2,t}^N$ be the number occupied by individuals using strategy 2, at time t, and $X_{3,t}^N$ be the number occupied by groups of one primary and one secondary, using strategy 2, at time t.

The total numbers of young produced by the territories in these three possible states are $y_0X_1^N$, $y_0X_2^N$ and $y_1X_3^N$, respectively (c.f. Table 6.1). Note: we will omit the subscript t and superscript N from some variables, for brevity.

Let $R \sim$ Binomial (X_2^N, π) be the number of juveniles remaining to form groups. Then after stage (i) of the life cycle, the population is as shown in Table 6.4.

Table 6.4

Numbers of Individuals before Mortality: Stochastic Model

From territory type:	1	2	3
Type 1 adults	X_1^N	0	0
Type 2 adults	0	X_2^N-R	0
Type 2 groups	0	R	X_3^N
Type 1 young	$y_0X_1^N$	0	0
Type 2 young (dispersing)	0	$y_0X_2^N-R$	$y_1X_3^N$

At stage (ii), mortality, let

$$A \sim \text{Binomial } (X_1^N, m)$$

$$B \sim \text{Binomial } (X_2^N-R, m)$$

$$K \sim \text{Multinomial } (R, p_\alpha+p_\beta, p_\gamma, p_\delta)$$

and

$$L \sim \text{Multinomial } (X_3^N, p_\alpha+p_\beta, p_\gamma, p_\delta).$$

Then after mortality, the state of the population is given by Table 6.5 (c.f. Table 6.3).

Table 6.5

Numbers of Individuals after Mortality: Stochastic Model

Original type of territory	1	2	3
Type 1 adults	$X_1^N - A$	0	0
Type 2 adults	0	$X_2^N - R - B + K_1$	L_1
Type 2 groups	0	K_2	L_2
Type 1 young	$y_0 X_1^N$	0	0
Type 2 young	0	$y_0 X_2^N - R$	$y_1 X_3^N$
Vacant territories	A	$B + K_3$	L_3

We now have a total number of vacancies given by

$$V = A + B + K_3 + L_3 \tag{6.4.2}$$

with D_1, D_2 dispersing young using strategies 1 and 2 respectively, where

$$D_1 = y_0 X_1^N,$$

$$D_2 = y_0 X_2^N + y_1 X_3^N - R.$$

Writing $D = D_1 + D_2$, the numbers of territories resettled by the two types, S_1 and S_2 respectively, are given by

$$S_1 \sim \text{Hypergeometric } (V, D_1, D),$$

$$S_2 = V - S_1.$$

So the numbers at time $t+1$ are

$$X_{1,t+1}^N = X_{1,t}^N + S_{1,t}^N - A_t^N, \tag{6.4.3}$$

$$X_{2,t+1}^N = X_{2,t}^N - R_t^N - B_t^N + K_{1,t}^N + L_{1,t}^N + S_{2,t}^N, \tag{6.4.4}$$

$$X_{3,t+1} = K_{2,t}^N + L_{2,t}^N. \tag{6.4.5}$$

Clearly $\{X_t^N\}$ forms a Markov chain, with state space

$$\{(X_1,X_2,X_3): X_1,X_2,X_3 \in \mathbb{Z}_+, X_1+X_2+X_3 = N\}.$$

It has a single absorbing state at $X_1 = N$, and an irreducible closed set of states with $X_1 = 0$. All other states are transient. Starting from a transient state, $\{X_t^N\}$ must eventually reach a state with $X_1 = 0$ or the single state with $X_1 = N$, corresponding to the extinction of strategy 1 or strategy 2 respectively. We would like to know the probabilities of those 2 possible outcomes, for large N, and we are particularly interested in the case where the initial population consists largely of type 1 individuals, say $X_{1,0}^N = N-1$, $X_{2,0}^N = 1$, $X_{3,0}^N = 0$.

Exact calculation of these probabilities for large finite N, and for given model parameters, is possible but unenlightening. Concise analytic expressions are not available, mainly because of the many interdependencies in the model.

However, in the limit as $N \to \infty$, we can approximate $\{X_t^N\}$ by a particular sort of branching process. An r-type Galton–Watson process with ancestor of type k is defined by

$$X_0 = e_k \tag{6.4.6}$$

(where e_k is the vector with 1 in the kth place, and zeroes elsewhere), and

$$X_{t+1} = \sum_{i=1}^{r} \sum_{j=1}^{X_{i,t}} Z_t^{(i,j)} \tag{6.4.7}$$

where $Z_t^{(i,j)}$ are random vectors, independent of each other and of $X_0,\ldots,X_t$, identically distributed for each i, with

$$\Pr(Z_t^{(i,j)} = z) = p_i(z)$$

for any $z \in \mathbb{Z}_+^r$. See for example Jagers (1975) or Mode (1971).

Theorem 6.2.

For any finite T, if $X_0^N = (N-1,1,0)$ then

$$\{(X_{2,t}^N, X_{3,t}^N): t = 0,1,\ldots,T\} \xrightarrow{a.s.} \{X_t: t = 0,1,\ldots,T\}$$

as $N \to \infty$, where $\{X_t\}$ is a two-type Galton–Watson process, with ancestor of type 1, and $\xrightarrow{a.s.}$ denotes almost sure convergence.

Proof.

Let X_0^N and X_0 be defined on the trivial probability space $(\Omega(0),\mathcal{F}(0),\mathrm{P}(0))$ with sample space $\Omega(0) = \{\omega_0\}$. For the purposes of the proof, we need to define random variables which indicate the fate of each individual during the dispersal and mortality phases of the life cycle. For $i,j,k,t \in \mathbb{Z}_+$, define independent random variables $R_t^{(j)}, B_t^{(j)}, K_t^{(j)}, L_t^{(j)}$ on $(\Omega_t,\mathcal{F}_t,\mathrm{P}_t)$ with

$$R_t^{(j)} \sim \text{Bernoulli } (\pi),$$

$$B_t^{(j)} \sim \text{Bernoulli } (m),$$

$$K_t^{(j)}, L_t^{(j)} \sim \text{Multinomial } (1, p_\alpha + p_\beta, p_\gamma, p_\delta)$$

and $A_t^{(j)}, Q_t^{(i,j,k)}, S_t^{(i,j,k)}$ on $(\Omega_t', \mathcal{F}_t', \mathbf{P}_t')$ with

$$A_t^{(j)} \sim \text{Bernoulli } (m),$$

$$Q_t^{(i,j,k)} \sim \text{Bernoulli } (m/y_0),$$

$$S_t^{(i,j,k)} \sim \text{Hypergeometric } (i,j,k).$$

We will construct $\{X_t^N\}$ and $\{X_t\}$ inductively in terms of the random variables defined on $(\Omega_t, \mathcal{F}_t, \mathbf{P}_t)$, and of random variables, on a probability space $(\hat{\Omega}_t, \hat{\mathcal{F}}_t, \hat{\mathbf{P}}_t)$, having the same distributions as those defined on $(\Omega_t', \mathcal{F}_t', \mathbf{P}_t')$. The branching process $\{X_t\}$ will have the offspring **distribution** defined by the following equations:

$$Z_{1,t}^{(1,j)} = R_t^{(j)} K_{1,t}^{(j)} + (1 - R_t^{(j)})(1 - B_t^{(j)}) + \sum_{k=1}^{y_0 - R_t^{(j)}} Q_t^{(1,j,k)} \tag{6.4.8}$$

$$Z_{2,t}^{(1,j)} = R_t^{(j)} K_{2,t}^{(j)} \tag{6.4.9}$$

$$Z_{1,t}^{(2,j)} = L_{1,t}^{(j)} + \sum_{k=1}^{y_1} Q_t^{(2,j,k)}, \quad \text{and} \tag{6.4.10}$$

$$Z_{2,t}^{(2,j)} = L_{2,t}^{(j)}. \tag{6.4.11}$$

At time $t = 0$, we have $\Omega(0) = \{\omega_0\}$, and

$$X_0^N = (N-1, 1, 0)$$

$$\Rightarrow (X_{2,0}^N, X_{3,0}^N) = (1, 0) = X_0.$$

So the theorem is trivially true for $T = 0$.

Now assume for induction that the theorem holds for T, i.e. that we have $X_t^N, X_t, t = 0, \ldots, T$ defined on $(\Omega(T), \mathcal{F}(T), \mathbf{P}(T))$ say, where $\{X_t^N\}$ is as described above, and $\{X_t\}$ is a branching process with offspring distribution as defined in equations (6.4.8) to (6.4.11), and

$$\{(X_{2,t}^N, X_{3,t}^N): t = 0, \ldots, T\} \xrightarrow{a.s.} \{X_t: t = 0, \ldots, T\}.$$

Define $Z_T^{(i,j)}$, $i = 1, 2$, as in equations (6.4.8) to (6.4.11), and define

$$R_T^N = \sum_{j=1}^{X_{2,T}^N} R_T^{(j)}, \qquad B_T^N = \sum_{j=1}^{X_{2,T}^N} (1 - R_T^{(j)}) B_T^{(j)},$$

$$K_T^N = \sum_{j=1}^{X_{2,T}^N} R_T^{(j)} K_T^{(j)}, \qquad L_T^N = \sum_{j=1}^{X_{2,T}^N} L_T^{(j)};$$

$$S_{2,T}^N = S_T^{(D_{2,T}^N, V_T^N, D_T^N)},$$

where

$$V_T^N = B_T^N + K_{3,T}^N + L_{3,T}^N + \sum_{j=1}^{X_{1,T}^N} A_T^{(j)},$$

$$D_{2,T}^N = y_0 X_{2,T}^N + y_1 X_{3,T}^N - R_T^N,$$

$$D_T^N = y_0 X_{1,T}^N + D_{2,T}^N;$$

and

$$X_{3,T+1}^N = K_{2,T}^N + L_{2,T}^N, \tag{6.4.12}$$

$$U_{T+1} = X_{2,T}^N - R_T^N - B_T^N + K_{1,T}^N + L_{1,T}^N. \tag{6.4.13}$$

Note that from equation (6.4.4), we can write

$$X_{2,T+1}^N \overset{D}{=} U_{T+1} + S_{2,T}^N. \tag{6.4.14}$$

We also have

$$X_{T+1} \overset{D}{=} \sum_{i=1,2} \sum_{i=1}^{X_{i,T}} Z_T^{(i,j)}. \tag{6.4.15}$$

Under the inductive hypothesis, we can take

$$X_T = (X_{2,T}^N, X_{3,T}^N) \tag{6.4.16}$$

for sufficiently large N, with probability 1.

Then

$$X_{1,T+1}^N \overset{D}{=} \sum_i \sum_j Z_{1,T}^{(i,j)} = U_{T+1} + \sum_{j=1}^{X_{1,T}} \sum_{k=1}^{y_0 - R_T^{(j)}} Q_T^{(1,j,k)} + \sum_{j=1}^{X_{2,T}} \sum_{k=1}^{y_1} Q_T^{(2,j,k)}, \tag{6.4.17}$$

$$X_{2,T+1}^N \overset{D}{=} \sum_i \sum_j Z_{2,T}^{(i,j)} = K_{2,T}^N + L_{2,T}^N. \tag{6.4.18}$$

Now consider the terms

$$S_{2,T}^N \quad \text{and} \quad \sum_{j=1}^{X_{1,T}} \sum_{k=1}^{y_0 - R_T^{(j)}} Q_T^{(1,j,k)} + \sum_{j=1}^{X_{2,T}} \sum_{k=1}^{y_1} Q_T^{(2,j,k)}.$$

Under the condition (6.4.16), and conditional on any fixed values of $R_T^{(j)}$, X_T, $B_T^{(j)}$, $K_T^{(j)}$, $L_T^{(j)}$, and hence U_{T+1}, we have

$$\sum_{j=1}^{X_{1,T}} \sum_{k=1}^{y_0-R_T^{(2)}} Q_T^{(1,j,k)} + \sum_{j=1}^{X_{2,T}} \sum_{k=1}^{y_1} Q_T^{(2,j,k)} \sim \text{Binomial}\left(y_0 X_{1,T} - R_T^N + y_1 X_{2,T}, \frac{m}{y_0}\right), \quad (6.4.19)$$

and

$$\begin{aligned} S_{2,T}^N \sim \text{Hypergeometric}\,(& y_0 X_{2,T}^N - R_T^N + y_1 X_{3,T}^N, \\ & B_T^N + K_{3,T}^N + L_{3,T}^N + \sum_{j=1}^{N-X_{2,T}^N-X_{1,T}^N} A_T^{(j)}, \\ & y_0(N - X_{2,T}^N - X_{3,T}^N) + y_0 X_{2,T}^N - R_T^N + y_1 X_{3,T}^N). \end{aligned} \quad (6.4.20)$$

By considering the parameters of the distributions in equations (6.4.19) and (6.4.20), we have

$$S_{2,T}^N \xrightarrow{D} \sum_{j=1}^{X_{1,T}} \sum_{k=1}^{y_0-R_T^{(2)}} Q_T^{(1,j,k)} + \sum_{j=1}^{X_{2,T}} \sum_{k=1}^{y_1} Q_T^{(2,j,k)}$$

on $(\Omega_T', \mathcal{F}_T', \mathbf{P}_T')$ as $N \to \infty$. Hence by the Skorokhod Representation Theorem (see for example Grimmett and Stirzaker, 1982) there exists a probability space $(\hat{\Omega}_T, \hat{\mathcal{F}}_T, \hat{\mathbf{P}}_T)$ and random variables $\hat{Q}_T^{(i,j,k)}, \hat{S}_{2,T}^N$ defined on $(\hat{\Omega}_T, \hat{\mathcal{F}}_T, \hat{\mathbf{P}}_T)$ such that

$$\hat{S}_{2,T}^N \xrightarrow{a.s.} \sum_{j=1}^{X_{1,T}} \sum_{k=1}^{y_0-R_T^{(2)}} \hat{Q}_T^{(1,j,k)} + \sum_{j=1}^{X_{2,T}} \sum_{k=1}^{y_1} \hat{Q}_T^{(2,j,k)}.$$

Finally, define

$$X_{2,T+1}^N = U_{T+1} + \hat{S}_{2,T}^N,$$

$$X_{3,T+1}^N = K_{2,T}^N + L_{2,T}^N$$

(c.f. equations (6.4.12) to (6.4.14)),

$$X_{1,T+1}^N = N - X_{2,T+1}^N - X_{3,T+1}^N,$$

and

$$X_{1,T+1} = U_{T+1} + \sum_{j=1}^{X_{1,T}} \sum_{k=1}^{y_0-R_T^{(2)}} \hat{Q}_T^{(1,j,k)} + \sum_{j=1}^{X_{2,T}} \sum_{k=1}^{y_1} \hat{Q}_T^{(2,j,k)},$$

$$X_{2,T+1} = K_{2,T}^N + L_{2,T}^N$$

(c.f. equations (6.4.17) and (6.4.18)) to get processes which are defined on $(\Omega(T+1), \mathcal{F}(T+1), \mathbf{P}(T+1))$, the product space of $(\Omega(T), \mathcal{F}(T), \mathbf{P}(T)), (\Omega_T, \mathcal{F}_T, \mathbf{P}_T)$ and $(\hat{\Omega}_T, \hat{\mathcal{F}}_T, \hat{\mathbf{P}}_T)$, which satisfy the conditions of the theorem, and for which the result of the theorem is true up to time $T+1$.

Hence by induction, the theorem is true for any finite time T.

By applying Theorem 6.2, we can use the known results about multi-type Galton–Watson processes to understand the behaviour of the stochastic model from this section, when

N is large.

Consider an r-type Galton–Watson process with ancestor type k, as defined in equations (6.4.6) and (6.4.7). Let

$$m_{ij} = \mathrm{E}[Z_j^{(i,1)}], \; M = [m_{ij}],$$

and define such a process to be **positively regular** if there is some power M^n of its matrix of expected numbers of offspring which is strictly positive.

Define the probability generating function $f(s)$ for any $s \in [0,1]^r$ by

$$f_i(s) = \sum_{z \in Z_+^r} \left\{ p_i(z) \prod_{j=1}^{r} s_j^{z_j} \right\},$$

and define a process to be **singular** if

$$f_i(s) = \sum_{j=1}^{r} p_{ij} s_j, \quad 1 \leqslant i \leqslant r,$$

for some numbers p_{ij}.

Finally define q_k to be the extinction probability of the process with ancestor type k, and let $q = (q_1, \ldots, q_r)$. Then we have the following result (see for example Jagers, 1975 or Mode, 1971).

Theorem 6.3.

Let $\{X_t\}$ be a positively regular, non-singular r-type Galton–Watson process, with M, f and q defined as above, and let λ be the Perron–Frobenius eigenvalue of the matrix M. Then q is the solution of the equation

$$f(s) = s$$

that lies closest to the origin in the unit cube. If $\lambda \leqslant 1$, then $q_k = 1$ for all k, and if $\lambda > 1$, then $q_k < 1$ for all k.

Proof: See Mode (1971).

We wish to apply Theorem 6.3 to the process $\{X_t\}$ in Theorem 6.2. It is clear from equations (6.4.8) to (6.4.11) that the matrix M of expectations of offspring distributions in $\{X_t\}$ is equal to the matrix A (of partial derivatives at $(0,0)$ of the reparameterised process ϕ_t) which occurred in the analysis of the deterministic model of Section 3. Thus we have $M = A$, where A is given by equation (6.3.13), and from the subsequent remarks we see immediately that $\{X_t\}$ is positively regular. To see that $\{X_t\}$ is also non-singular, it is sufficient to note that $p_i(z) > 0$ when for example $i = 2, z_1 = 2$, and $z_2 = 0$, so f_2 includes a term in s_1^2. So $\{X_t\}$ satisfies the conditions of Theorem 6.3, and furthermore its Perron–Frobenius eigenvalue is the same as that of the matrix A. Hence using equation (6.3.15), and writing

$$C = y_0\left(1 - \frac{p_\delta}{m}\right) + p_\gamma(1+w) \tag{6.4.21}$$

as before, we have the result that if $C \leqslant 1$, then $q_k = 1$ for all k, and if $C > 1$, then $q_k < 1$ for all k. Thus the same criterion which determines the stability of the equilibria of our deterministic model also determines whether the branching process $\{X_t\}$ has a positive probability of survival.

The properties of $\{X_t\}$ carry over to $\{X_t^N\}$, to some extent, through Theorem 6.2. Clearly if $X_T = \mathbf{0}$, then for sufficiently large N, $X_T^N = \mathbf{0}$, and so $q(N)$, the extinction probability for $\{X_t^N\}$ is at least as great as q_1, the relevant extinction probability for $\{X_t\}$. No inequality in the other direction has been proven, but comparison with the deterministic process of Section 3 strongly suggests that

$$q_1 < 1 \Rightarrow \lim_{N \to \infty} q(N) < 1.$$

In fact it is conjectured that

$$\lim_{N \to \infty} q(N) = q_1.$$

Then the parameter C, which determines whether or not $q_1 = 1$, would also determine whether or not $\lim_{N \to \infty} q(N) = 1$, and hence whether social behaviour could become established.

Throughout the current section, we have assumed, for simplicity, that the number of offspring is deterministic, but all the results given, including Theorem 6.2, generalise in an obvious way to the more realistic case of stochastic family sizes, provided

$$\Pr(Y_{j,t} \geqslant 2) = 1,$$

$$\Pr(Y_{j,t}^* \geqslant 1) = 1,$$

so that all vacancies are always filled.

6.5 Diploid Models

6.5.1 Introduction

Although the models of Emlen (1982), Lindström (1986), and Macdonald and Carr (1989) consider social behaviour within a single sex, the fact that the whole population has two sexes, i.e. is diploid, is used explicitly in the calculation of relatedness, fitness etc.. Thus for comparison with the above published models, we should use a diploid version of the models in Sections 6.3 and 6.4.

We make the following assumptions, in addition to those mentioned in Section 6.3.1. The population consists of two sexes, and individuals are identical apart from strategy and sex.

All adults live either as secondary animals or as monogamous pairs; such pairings, once formed, last until the death of one partner (or both), and any remaining partner takes no further part in reproduction. At time t, the jth breeding pair produces $Y_{j,t}^{M}$ male and $Y_{j,t}^{F}$ female young, or $Y_{j,t}^{M^*}$ and $Y_{j,t}^{F^*}$ respectively if a secondary animal is present, where

$$E[Y_{j,t}^{M}] = E[Y_{j,t}^{F}] = y_0 \tag{6.5.1}$$

and

$$E[Y_{j,t}^{M^*}] = E[Y_{j,t}^{F^*}] = y_1 \tag{6.5.2}$$

for all j and t.

Grouping is caused by an allele at one particular locus, which has the effect of causing the juvenile bearing it to remain in its natal territory, when possible, with some probability π_0. We further assume that the allele responsible for grouping is dominant, giving the simplest possible diploid model, as discussed in Section 6.1.4. Thus we assume two alleles, the recessive allele a and the dominant allele A say; three genotypes, aa,Aa (=aA), and AA; and two phenotypes or strategies, the background, non-social type, corresponding to the genotype aa, and the type which may form groups, corresponding to genotypes Aa and AA.

We concentrate on two deterministic models, incorporating different assumptions about secondary animals, using similar techniques to those in Section 6.3. We assume that the population initially has genotype aa, except for a small proportion carrying the mutant allele A.

6.5.2 Secondary Animals of Either Sex.

The first diploid model we consider allows a juvenile of either sex to remain in its natal territory to form a rudimentary social group. Let

$x_{1,t}$ be the proportion of territories held at time t by mated pairs of animals, both of genotype aa,

$x_{2,t}$ be the proportion of territories held at time t by mated pairs, with each pair having exactly one copy of the allele A,

$x_{3,t}$ be the proportion of territories held at time t by groups, each consisting of a primary pair with exactly one copy of A, and a single secondary animal with genotype Aa, and

$x_{4,t}$ be the proportion of territories held at time t by pairs or groups of other genotypes, i.e. including individuals or pairs with more than one copy of the mutant allele A.

Clearly

$$x_{1,t}+x_{2,t}+x_{3,t}+x_{4,t} = 1,$$

and we assume that allele A is initially rare, i.e. at time $t = 0$, we have

$$x_2 = O(\varepsilon), \tag{6.5.3}$$

$$x_3 = O(\varepsilon), \tag{6.5.4}$$

$$x_4 = O(\varepsilon^2), \tag{6.5.5}$$

and hence

$$x_1 = 1 - x_2 - x_3 + O(\varepsilon^2), \tag{6.5.6}$$

for sufficiently small $\varepsilon > 0$. We will show that if equations (6.5.3) to (6.5.6) hold at time t, then they also hold at time $t+1$. The dynamics of the system are similar to those of the model in Section 6.3, and so the explanation here will be more concise. The numbers produced of young of each sex are shown in Table 6.6.

Table 6.6
Numbers of Young of Each Sex: Diploid
Model with Secondaries of Either Sex

From territory type:	1	2	3	4
Young of genotype *aa*	$x_1 y_0$	$\frac{1}{2}x_2 y_0$	$\frac{1}{2}x_3 y_1$	$O(\varepsilon^2)$
Young of genotype *Aa*	0	$\frac{1}{2}x_2 y_0$	$\frac{1}{2}x_3 y_1$	$O(\varepsilon^2)$
Young of genotype *AA*	0	0	0	$O(\varepsilon^2)$

The probability of group formation in a territory of type 2 is now

$$\pi = 1 - \mathrm{E}[(1 - \frac{\pi_0}{2})^{(Y_{j,t}^M + Y_{j,t}^F)}] \tag{6.5.7}$$

for any j and t. After the dispersion of juveniles, mortality takes place. Some care is needed in defining the parameters of mortality. We define m to be the probability that in a given territory of type 1 or 2, in a given time period, at least one of the occupying adults dies. By assumption, this has the same effect as both adults dying. In a territory occupied by a group of two primaries and one secondary, we have the following possibilities:

(a) the secondary dies but the primaries survive and retain the territory, with probability p_α;

(b) at least one of the primaries dies, but the secondary survives and inherits the territory, with probability p_β;

(c) all three animals survive, and retain the territory as a group, with probability p_γ; or

(d) the secondary and at least one of the primaries die, and the territory is vacated, with probability p_δ.

The state of the population is then as shown in Table 6.7.

Table 6.7

Numbers of Individuals after Mortality:

Diploid Model with Secondaries of Either Sex

From territory type:	1	2	3	4
Type 1 pairs	$x_1(1-m)$	0	0	0
Type 2 pairs	0	$x_2(1-\pi)(1-m)+x_2\pi p_\alpha$	$x_3 p_\alpha$	$O(\varepsilon^2)$
Type 3 groups	0	$x_2\pi p_\gamma$	$x_3 p_\gamma$	0
Type 4	0	0	0	$O(\varepsilon^2)$
Dispersers of each sex				
aa	$x_1 y_0$	$\frac{1}{2}x_2 y_0$	$\frac{1}{2}x_3 y_1$	$O(\varepsilon^2)$
Aa	0	$\frac{1}{2}x_2(y_0-\pi)$	$\frac{1}{2}x_3 y_1$	$O(\varepsilon^2)$
AA	0	0	0	$O(\varepsilon^2)$
Vacancies	$x_1 m$	$x_2(1-\pi)m+x_2\pi p_\delta$	$x_3 p_\delta$	$O(\varepsilon^2)$
Inheriting juveniles				
Aa	0	$x_2\pi p_\beta$	$x_3 p_\beta$	$O(\varepsilon^2)$
AA	0	0	0	$O(\varepsilon^2)$

Finally, we resettle the inherited and vacant territories, by assuming that each inheriting secondary selects a mate at random from the dispersing juveniles of the opposite sex, and that mated pairs form at random to occupy vacant territories. These calculations are relatively straightforward because the mutant allele *A* is rare. The final expressions for the numbers of such resettled territories are given in Table 6.8.

Table 6.8

Numbers of Territories: Diploid Model

with Secondaries of Either Sex

	Inherited	Newly Settled
Type 1	0	$x_1 m + x_2\pi(p_\delta - m(1-\frac{1}{y_0})) + x_3(p_\delta - m\frac{y_1}{y_0}) + O(\varepsilon^2)$
Type 2	$x_2\pi p_\beta + x_3 p_\beta + O(\varepsilon^2)$	$x_2 m(1-\frac{\pi}{y_0}) + x_3 m\frac{y_1}{y_0} + O(\varepsilon^2)$
Type 3	0	0
Type 4	$O(\varepsilon^2)$	$O(\varepsilon^2)$

Note: as in Section 6.3.2, we have assumed that there are enough dispersing young; a sufficient condition is that

$$y_0 > m \tag{6.5.8}$$

Thus we have:

$$x_{1,t+1} = x_{1,t} + x_{2,t}(p_\delta - m(1-\frac{1}{y_0})) + x_{3,t}(p_\delta - m\frac{y_1}{y_0}) + O(\varepsilon^2); \tag{6.5.9}$$

$$x_{2,t+1} = x_{2,t}(1+\pi(p_\alpha + p_\beta + m - 1 - \frac{m}{y_0})) + x_{3,t}(p_\alpha + p_\beta + m\frac{y_1}{y_0}) + O(\varepsilon^2); \tag{6.5.10}$$

$$x_{3,t+1} = x_{2,t}\pi p_\gamma + x_{3,t}p_\gamma + O(\varepsilon^2); \tag{6.5.11}$$

$$x_{4,t+1} = O(\varepsilon^2) \tag{6.5.12}.$$

Since we have assumed that equations (6.5.3) to (6.5.5) apply at time t, they clearly also hold at time $t+1$.

Under these assumptions, the key question concerns the stability of the equilibrium

$$x^* = (1,0,0,0) = e_1. \tag{6.5.13}$$

If x^* is unstable, then the population be invaded by the mutant allele A.

We wish to know if x^* is stable under biologically meaningful perturbations of the form

$$(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)$$

where

$$\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4 = 0,$$

$$\varepsilon_1,\varepsilon_2,\varepsilon_3 = O(\varepsilon),$$

$$\varepsilon_4 = O(\varepsilon^2),$$

for sufficiently small ε. Following Section 6.3.3, we reparameterise the system, defining

$$z_1 = x_2,$$

$$z_2 = x_3.$$

The variables z_1, z_2 are sufficient to describe x up to $O(\varepsilon^2)$;

$$x_1 = 1-z_1-z_2+O(\varepsilon^2),$$

$$x_2 = z_1,\ \ x_3 = z_2,\ \ x_4 = O(\varepsilon^2).$$

From equations (6.5.9) to (6.5.12), we then have

$$z_t = O(\varepsilon)$$

$$\Rightarrow z_{t+1} = Az_t+O(\varepsilon^2)$$

where

$$A = \begin{pmatrix} 1+\pi(p_\alpha+p_\beta+m-1-\dfrac{m}{y_0}) & p_\alpha+p_\beta+m\dfrac{y_1}{y_0} \\ \pi p_\gamma & p_\gamma \end{pmatrix}.$$

So the stability of the equilibrium $z^* = (0,0)$, and hence of $x^* = (1,0,0,0)$, depends on the eigenvalues of A. But after rearrangement, it can be seen that A is identical to the matrix of partial derivatives obtained at the corresponding equilibrium of the haploid system in Section 6.3. Thus the equilibrium $x^* = (1,0,0,0)$ is stable if and only if

$$y_0\left(1-\frac{p_\delta}{m}\right)+p_\gamma(1+w) < 1,$$

by equation (6.3.15).

6.5.3 Secondary Animals of Only One Sex

The published models mentioned at the beginning of Section 6.5.1 differ from the model presented in Section 6.5.2 in that they assume all secondary animals to be of the same sex, which without loss of generality we can take to be female. We will show, however, that if the frequency of the mutant allele A is of order $O(\varepsilon)$, the assumption that all secondaries are female has an effect that is of order $O(\varepsilon^2)$, so the conclusions about the model are unchanged.

If all juveniles which remain to become secondaries are female, then instead of the situation shown in Table 6.7, we obtain that shown in Table 6.9 below.

Table 6.9
Numbers of Individuals after Mortality:
Diploid Model with Female Secondaries

From territory type:	1	2	3	4
Territories				
Type 1	$x_1(1-m)$	0	0	0
Type 2	0	$x_2(1-\pi)(1-m)+x_2\pi p_\alpha$	$x_3 p_\alpha$	$O(\varepsilon^2)$
Type 3	0	$x_2\pi p_\gamma$	$x_3 p_\gamma$	0
Type 4	0	0	0	$O(\varepsilon^2)$
Dispersers				
Male *aa*	$x_1 y_0$	$\frac{1}{2}x_2 y_0$	$\frac{1}{2}x_3 y_1$	$O(\varepsilon^2)$
Male *Aa*	0	$\frac{1}{2}x_2 y_0$	$\frac{1}{2}x_3 y_1$	$O(\varepsilon^2)$
Male *AA*	0	0	0	$O(\varepsilon^2)$
Female *aa*	$x_1 y_0$	$\frac{1}{2}x_2 y_0$	$\frac{1}{2}x_3 y_1$	$O(\varepsilon^2)$
Female *Aa*	0	$x_2(\frac{y_0}{2}-\pi)$	$\frac{1}{2}x_3 y_1$	$O(\varepsilon^2)$
Female *AA*	0	0	0	$O(\varepsilon^2)$
Vacancies	$x_1 m$	$x_2(1-\pi)m+x_2\pi p_\delta$	$x_3 p_\delta$	$O(\varepsilon^2)$
Inheriting females				
Aa	0	$x_2\pi p_\beta$	$x_3 p_\beta$	$O(\varepsilon^2)$
AA	0	0	0	$O(\varepsilon^2)$

Assuming that inheriting females select random males as mates, and that the resettlement of vacancies is random, the numbers of territories of different types obtained are the same as in Table 6.8. Thus, up to order $O(\varepsilon^2)$, the dynamics of the system are unaffected by the assumption that all secondaries are of the same sex.

6.5.4 Comparison with Other Models

The assumptions of the model of Section 6.5.3 are very close to those of Macdonald and Carr (1989). In fact, we can find special cases of the two models which have essentially the same assumptions.

Macdonald and Carr assume that mortality of primaries and secondaries is independent, and unaffected by group size, with primaries having survival probability $P_\alpha, (\neq p_\alpha)$ and secondaries having survival probability P_i. Thus we will restrict our attention to the case where p_γ, p_δ and m satisfy

$$p_\gamma = P_\alpha P_i,$$

$$p_\delta = (1-P_\alpha)(1-P_i),$$

$$m = 1-P_\alpha,$$

for some choices of P_α, P_i. Then the criterion for the establishment of group behaviour in our model becomes

$$P_i(y_0+P_\alpha(1+w)) > 1.$$

The Macdonald and Carr model considers general group sizes, so for comparison with our model we must restrict it to the case of a single secondary. This yields the condition

$$P_e y_0 > P_i\{\frac{P_\alpha w}{2}+(1-P_\alpha)y_0\}$$

for 'remaining' to be optimal for a secondary, where P_e is Macdonald and Carr's notation for the survival probability of dispersing juveniles. In order to give a constant number of occupied territories, P_e must be given by

$$P_e = \frac{1-P_\alpha}{y}.$$

Macdonald and Carr also aim to consider conflict between adult and juveniles, by allowing the formation of groups to need the 'consent' of both adult and juvenile. Incorporating such potential conflict into the models developed here would make them rather complicated, so instead we restrict our parameter values so that the conflict is unimportant. With only one secondary allowed, a sufficient condition for the absence of conflict, according to the criteria in Macdonald and Carr, is

$$y_1 \geq y_0.$$

Under the conditions given here, the criterion for the establishment of groups in the Macdonald and Carr model is

$$1-P_\alpha < P_i\{\frac{P_\alpha w}{2}+(1-P_\alpha)y_0\}$$

or

$$P_i\{y_0+\frac{P_\alpha}{1-P_\alpha}.\frac{w}{2}\} > 1,$$

whereas for our model it is

$$P_i\{y_0+P_\alpha(1+w)\} > 1.$$

Clearly these two criteria are similar, but neither one implies the other, as is shown in Table 6.10.

Table 6.10
Comparison between the Female-Secondaries
and Macdonald and Carr Models

P_α	P_i	y_0	w	Groups predicted by:	
				Section 6.5.3	Macdonald and Carr
0.9	0.6	2.0	0.0	Yes	Yes
0.9	0.5	2.0	0.0	Yes	No
0.9	0.4	2.0	0.0	Yes	No
0.9	0.3	2.0	0.0	No	No
0.9	0.3	2.0	1.0	Yes	Yes
0.9	0.2	2.0	1.0	No	Yes
0.9	0.1	2.0	1.0	No	No

The differences between the conclusions of our explicitly dynamic model, and Macdonald and Carr's (1989) fitness-based calculations, show that their model does not apply to groups lasting for more than one year.

The diploid model we have derived can also be compared with Lindström's (1986) model. Our model assumes all territories in the habitat to be identical, so we will only consider Lindström's single-territories (his double-territories have been discussed at length in Section 6.2). As mentioned in Section 6.2, single-territories have the property that they cannot support primary and secondary animals for the whole year, i.e. $p_\gamma = 0$, in our notation.

Our model also assumes a constant number of territories occupied, so we will restrict our attention to that case of Lindström's model ($R = 1$ in his notation). With single-territories and constant population size, Lindström predicted that the two strategies for a juvenile, 'remain' and 'disperse' would have the same fitness. However, his conclusion depends critically on two further assumptions, which in our notation are

$$y_0 = 1 \text{ and } p_\delta = 0.$$

More generally, our model predicts that, in territories with no long-term groups, (i.e. $p_\alpha = 0$), remaining to form groups is advantageous if and only if

$$y_0(1 - \frac{p_\delta}{m}) > 1,$$

as can be seen by considering the form of A when $p_\delta = 0$. Thus it is clearly possible for the strategy of forming groups to be strictly better or strictly worse than that of dispersing, depending on y_0 and $\frac{p_\delta}{m}$, and Lindström's conclusion about single-territories is unjustified.

6.6 Discussion.

We have described four different models for the evolution of social behaviour in territorial animals, based on the Resource Dispersion Hypothesis. The important qualitative features of all four models are summarised in a single, simple expression,

$$C = y_0\left(1 - \frac{p_\delta}{m}\right) + p_\gamma(1+w). \tag{6.6.1}$$

The parameters in equation (6.6.1) are defined in Section 3, for haploid models, and in Section 5, for diploid models, the only differences in meaning being those that are natural in changing between those two types of model. In the three deterministic models, we conclude that social behaviour, as described here, can become established in a population if and only if $C > 1$. In the case of the stochastic model, $C > 1$ is a necessary (and, it is conjectured, sufficient) condition for the probability of establishment to be positive.

Equation (6.6.1) shows the importance of the various parameters of the models. For example, only the ratio, not the values, of p_δ and m is important, and y_0, the expected number of offspring, will be important unless $p_\delta \approx m$. In addition, the parameter π, the probability of a juvenile actually electing to form a group, given that it has the appropriate gene, does not appear in equation (6.6.1). Thus it has no effect on whether social grouping can become established (though it may affect the probability of establishment in a stochastic model).

As described in Section 4, equation (6.6.1) shows that Lindström's (1986) conclusion, that social behaviour gives no advantage when groups cannot over-winter, is not generally valid. Instead a simple (and easily attained) condition for social grouping to be advantageous,

even when groups are necessarily transient, has been derived from equation (6.6.1).

In general, however, our models describe groups which potentially last for a number of years. We can allow for such groups because the models are dynamic, with the state of the population depending both on numbers of individuals and their social organisation. Defining a dynamic on such a state-space enables us to model situations outside the usual range of evolutionary game theory, and avoid some of the complexities of calculating fitness. Further work is needed to extend these models to allow for larger groups and for parent-offspring conflict, but the results here suggest that that the dynamic approach presented here can be useful.

Chapter 7 Conclusions

7.1 Summary of results

We have seen in Chapter 2 that the model proposed by Carr and Macdonald (1986) for the formation of territories and groups, in a habitat of identical food patches with randomly varying yields, makes very strong assumptions, and leads to rather artificial results, which in some cases differ from those stated by the original authors. In Chapter 3, we have shown that by replacing Carr and Macdonald's discrete distribution of yield with a continuous distribution, and allowing random variation between food patches and hence between territories, we can obtain a new model which is more mathematically tractable. The behaviour and predictions of the new model are reasonably robust, in the sense that they are not too sensitive to small changes in parameter values or in the structure of the model. The model leads to a number of important conclusions: we mention some examples here. One is that groups of animals will be formed under a wide range of conditions, lending strength to the key idea of the Resource Dispersion Hypothesis. Another is that independence between group size and territory size, (measured in patches), which Carr and Macdonald suggested follows from the R.D.H., only holds under a restricted range of conditions. Finally, increasing the variability does not always increase the probability of groups being formed, contrary to intuition.

In Chapters 4 and 5, we looked at the effect of the spatial location of patches on the above model. Four different spatial models were used, each one considered both with and without interaction between territories. The models differed widely in the quality, i.e. in the perimeter, of the territories that they produced, but in most cases the predicted probability of group formation is fairly consistent, both between the spatial models and with the above non-spatial model. Thus the most important conclusion of the non-spatial model, that groups will be formed in a wide range of different stochastic environments, is robust to the introduction of spatial aspects into the model, in a number of possible ways.

The above models, strictly speaking, do not show that groups will definitely be formed, but only that opportunities for group formation occur. The aim of Chapter 6 was to determine when these opportunities would be taken, by examining when such behaviour would be evolutionarily successful. We considered four slightly different models (deterministic and stochastic, and with varying assumptions about genetics) based on a simple, abstract model for identical territories. All four models agree in their main conclusions, which can be summarised by a single simple equation that gives information on when group behaviour can actually occur.

7.2 Further Work

There are a number of areas where it would be desirable to extend the work in this thesis. Some of these are considered below, under the headings non-spatial models (related to the material in Chapters 1,2 and 3), spatial models (Chapters 4 and 5) and evolutionary models (Chapter 6).

7.2.1 Non-Spatial Models

The models presented in Chapter 1 (due to Carr and Macdonald, 1986) and Chapter 3 are rather simple models, and it is not clear that there is much to be gained from much further analysis beyond what is given here, apart from some places where the given results might be superseded by more precise analytic results. One exception to this is if there were strong interest in, or detailed data available for, a particular choice of species and habitat. There might then be sufficient reason to further explore a specific case of the model in Chapter 3, e.g. a particular choice of distribution of patch means, or a particular case of the seasonal model mentioned in Section 3.3.2. If sufficiently detailed data were available, it would clearly be desirable to formally test the predictions of the model.

7.2.2 Spatial Models

Since the simulations in Chapters 4 and 5 are in many cases limited by the computing time required, there is clearly some scope for trying to increase the efficiency of simulation, and carrying out further simulation experiments in selected cases. As in Section 7.2.1 above, this is especially true if there is particular interest in a given real case, to which simulations can be tailored. However, a more important step would be to improve the realism of the models by incorporating some notion of readjustment of the territories after the initial settlement. There are two obvious approaches to adjustment. One is to allow each territory to change its perimeter to include or exclude certain points, to achieve some type of local optimality. Clearly this is a natural generalisation of the local optimisation model in Section 4.5, though not necessarily one which is easy to formalise and implement. The other approach to adjustment is to introduce mortality, followed by resettlement, into the model, so that the habitat is never filled permanently, but has vacancies continually occuring. Intuitively, the process would be expected to reach a stochastic equilibrium. Mortality might be completely random, or might vary according to territory quality. Readjustment through mortality seems more natural than the local optimisation discussed above, for most of the spatial models we have considered, and is likely to be easier to implement. More importantly, it suggests a natural link with evolutionary models, which is pursued in Section 7.2.3 below.

7.2.3 Evolutionary Models

Although the evolutionary models in Chapter 6 give a useful insight into the existence of groups sharing territories, they need to be extended to allow for the possibility of larger groups, and to explore the possible conflicts mentioned in Section 6.5.4. For these purposes, it is anticipated that the methods of Chapter 6 will continue to be useful. Some further work is required to prove (or disprove) the conjecture at the end of Section 6.4.

However, perhaps the main limitation of the models considered is that they deal with identical territories. It would be extremely interesting to explore the affect of stochastic territories, as in Chapters 3,4 and 5, on evolutionary models. If the pattern of territories is selected randomly and then fixed, then some analytic progress might be possible, at least in the limit for large numbers of territories. In the more general case, where new territories set up need not correspond exactly to those vacated due to the death of a territory holder, the distribution of territory quality etc. may change over time, and simulation is likely to be necessary. Such simulation could incorporate the spatial models from Chapter 5, and could then be used both to answer evolutionary questions and to investigate the idea of adjustment through mortality mentioned in Section 7.2.2 above.

REFERENCES

Ball, F. G. and Donnelly, P. (1988). A Unified Approach to Variability in Compartmental Models. *Biometrics*, **44**, 685-694.

Bellman, R. (1960). *Introduction to Matrix Analysis*. New York: McGraw-Hill.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *JRSS B*, **36**, 192-236.

Bradbury, J. W. and Vehrencamp, S. L. (1976). Social Organisation and Foraging in Emballonurid Bats II. A Model for the Determination of Group Size. *Behav. Ecol. Sociobiol.*, **1**, 383-404.

Carr, G. M. and Macdonald, D. W. (1986). The Sociality of Solitary Foragers : A Model Based on Resource Dispersion. *Anim. Behav.*, **34**, 1540-1549.

Covich, A. P. (1976). Analyzing shapes of foraging areas: some ecological and economic theories. *Annual Review of Ecology and Systematics*, **7**, 235-257.

Cox, D. R. (1967). *Renewal Theory*. London: Methuen.

Cox, D. R. and Isham, V. (1980). *Point Processes*. London: Chapman and Hall.

Cressman, R. and Dash, A. T. (1985). Evolutionarily Stable Strategies with Two Types of Player I. Two-Species Haploid or Randomly Mating Diploid. *J. Appl. Prob.*, **22**, 1-14.

Davies, N. B. and Houston, A. I. (1984). Territory Economics. In: *Behavioural Ecology: An Evolutionary Approach. 2nd Edition* (ed. J. R. Krebs and N. B. Davies), pp. 148-169. Blackwell.

Dawkins, R. (1982). *The Extended Phenotype*. Oxford: O. U. P.

DeGroot, M. H. (1986). *Probability and Statistics, 2nd edition*. Addison Wesley.

De Finetti, B. (1937). Foresight: Its logical laws, its subjective sources. In: *Studies in Subjective Probability* (ed. H. E. Kyburg and H. E. Smokler). New York: Wiley.

Diggle, P. J. (1983). *Statistical analysis of spatial point patterns*. Academic Press.

Don, B. A. C. and Rennolls, K. (1983). A home range model incorporating biological attraction points. *Journal of Animal Ecology*, **52**, 69-81.

Efron, B. (1965). The convex hull of a random set of points. *Biometrika*, **52**, 331-343.

Emlen, S. T. (1982). The Evolution of Helping. II. The Role of Behavioural Conflict. *The American Naturalist*, **119**, 40-53.

Ewer, R.F. (1968). *Ethology of mammals*. Logos.

Feldman, M. W. and Eshel, I. (1982). On the theory of parent-offspring conflict: a two-locus genetic model. *The American Naturalist*, **119**, 285-292.

Fox, M. W. (1970). *Behaviour of Wolves, Dogs and Related Canids*. London: Jonathan Cape.

Getty, T. (1981). Analysis of central-place space-use patterns : the elastic disc revisited. *Ecology*, **62**, 907-914.

Gradshetyn, I. S. and Ryzhik, I. M. (1980). *Table of integrals, series and products*. New York: Academic Press. Corrected and Enlarged Edition : ed. A. Jeffrey

Grafen, A. (1984). Natural Selection, Kin Selection and Group Selection. In: *Behavioural Ecology: An Evolutionary Approach. 2nd Edition* (ed. J. R. Krebs and N. B. Davies), pp. 62-84. Blackwell.

Green, P. J. and Sibson, R. (1978). Computing Dirichlet tesselations in the plane. *The Computer Journal*, **21**, 168-173.

Grimmett, G. and Stirzaker, D. (1982). *Probability and Random Processes*. Oxford: Oxford University Press.

Hasegawa, M. and Tanemura, M. (1976). On the pattern of space division by territories. *Annals of the Institute of Statistical Mathematics , Part B*, **28**, 509-519.

Hasegawa, M. and Tanemura, M. (1980). Spatial patterns of territories. In: *Recent Developments in Statistical Inference and Data Analysis* (ed. K. Matusita), pp. 73-78. North-Holland.

Hersteinsson, P. and Macdonald, D. W. (1982). Some comparisons between red and arctic foxes as revealed by radio tracking. *Symp. Zool. Soc. Lond.*, **49**, 259-289.

Hines, W. G. S. (1987). Evolutionary Stable Strategies: A Review of Basic Theory. *Theor. Pop. Biol.*, **31**, 195-272.

Hirsch, M. W. and Smale, S. (1974). *Differential equations, dynamical systems and linear algebra*. New York: Academic Press.

Hölldobler, B. and Lumsden, C. J. (1980). Territorial Strategies in Ants. *Science*, **210**, 732-739.

Horn, H. S. (1968). The Adaptive Significance of Colonial Nesting in the Brewer's Blackbird (*Euphagus cyanocephalus*). *Ecology*, **49**, 682-694.

Jagers, P. (1975). *Branching Processes*. Wiley.

Johnson, N. L. and Welch, B. L. (1939). On the calculation of the cumulants of the χ-distribution. *Biometrika*, **31**, 216-218.

Johnson, N. L. and Kotz, S. (1969). *Distributions in statistics: discrete distributions*. New York: John Wiley.

Johnson, N. L. and Kotz, S. (1970). *Distributions in statistics: continuous univariate distributions - 1*. New York: John Wiley.

Jones, D. W. and Krummel, J. R. (1985). The location theory of animal populations: the case of a spatially uniform food distribution. *The American Naturalist*, **126**, 392-404.

Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes (2nd edition)*. New York: Academic Press.

Kaufman, H. (1955). A Bibliographical Note on Higher Order Sine Functions. *Scripta Mathematica*, **28**, 29-36.

Kendall, M. G. and Moran, P. A. P. (1963). *Geometrical Probability*. London: Charles Griffin.

King, R. C. (1972). *A Dictionary of Genetics, 2nd edition*. New York: O.U.P..

Kingman, J. F. C. (1978). Random Partitions in Population Genetics. *Proc. Royal Society, Series A*, **361**, 1-20.

Kruuk, H. (1978a). Foraging and Spatial Organisation of the European Badger, *Meles meles L.. Behav. Ecol. Sociobiol.*, **4**, 75-89.

Kruuk, H. (1978b). Spatial organization and territorial behaviour of the European badger *Meles meles. J. Zool., Lond.*, **184**, 1-19.

Kruuk, H. and Parish, T. (1982). Factors affecting population density, group size and territory size of the European badger, *Meles meles. J. Zool., Lond.*, **196**, 31-39.

Kruuk, H. and Macdonald, D. W. (1985). Group territories of carnivores : Empires and Enclaves. In: *Behavioural Ecology : ecological consequences of adaptive behaviour* (ed. R.M. Sibly & R.H. Smith), pp. 521-536. The 25th Symposium of The British Ecological Society, Reading 1984

Lancaster, P. (1969). *Theory of Matrices*. New York: Academic Press.

Lawson, C. L. (1972). Generation of a triangular grid with application to contour plotting. *California Institute of Technology Jet Propulsion Laboratory*,Technical Memorandum 299.

Lawson, C. L. (1977). Software for C^1 surface interpolation. In: *Mathematical Software III* (ed. J. R. Rice), pp. 161-194. New York: Academic Press.

Leadbetter, M.R., Lindgren, G., and Rootzén, H. (1983). *Extremes and related properties of stationary sequences and processes*. Berlin: Springer-Verlag.

Lindström, E. (1986). Territory inheritance and the evolution of group-living in carnivores. *Anim. Behav.*, **34**, 1825-1835.

Macdonald, D. W. (1981). Resource dispersion and the social organization of the red fox. In: *Worldwide Furbearer Conference Proceedings* (ed. J. A. Chapman & D. Pursley), pp. 918-949.

Macdonald, D. W. (1983). The ecology of carnivore social behaviour. *Nature, Lond.*, **301**, 379-384.

Macdonald, D. W. and Carr, G. M. (1989). Food security and the rewards of tolerance. In: *Comparative Socioecology* (ed. V. Standen and R. A. Foley), pp. 75-99. Oxford: Blackwell.

Mardia, K. V. (1972). *Statistics of Directional Data*. New York: Academic Press.

Maynard Smith, J. and Price, G. R. (1973). The Logic of Animal Conflict. *Nature*, **246**, 15-18.

Maynard Smith, J. (1974). *Models in Ecology*. Cambridge: Cambridge University Press.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Mode, C. J. (1971). *Multitype Branching Processes*. New York: American Elsevier.

Noakes, D. L. G. and McNicol, R. E. (1982). Geometry for the eccentric territory. *Can. J. Zool.*, **60**, 1776-1779.

Numerical Algorithms Group, (1986). *The NAG Pascal Library Handbook*. Oxford: NAG.

Ripley, B. D. (1987). *Stochastic simulation*. Wiley.

Rood, J. P. (1975). Population Dynamics and Food Habits of the Banded Mongoose. *E. Afr. Wildl. J.*, **13**, 89-111.

Rootzén, H. (1983). The rate of convergence of extremes of stationary normal sequences. *Adv. Appl. Prob.*, **15**, 54-80.

Rowe, G. W., Harvey, I. F., and Hubbard, S. F. (1985). The Essential Properties of Evolutionary Stability. *J. Theor. Biol.*, **115**, 269-285.

Rowe, G. W. (1988). To Each Genotype a Separate Strategy - A Dynamic Game Theory Model of a Diploid Genetic System. *J. Theor. Biol.*, **134**, 89-101.

Schoener, T. W. (1983). Simple models of optimal feeding-territory size : a reconciliation. *The American Naturalist*, **121**, 608-629.

Seneta, E. (1982). De Moivre, Abraham. In: *Encyclopedia of Statistical Sciences, Vol. 2* (ed. S. Kotz and N. L. Johnson). New York: Wiley.

Sibson, R. (1978). Locally equiangular triangulations. *The Computer Journal*, **21**, 243-245.

Stoyan, D., Kendall, W.S., and Mecke, J. (1987). *Stochastic geometry and its applications*. Wiley.

Tanemura, M. and Hasegawa, M. (1980). Geometrical Models of Territory 1. Models for Synchronous and Asynchronous Settlement of Territories. *J. Theor. Biol.*, **82**, 477-496.

Taylor, P. D. and Jonker, L. B. (1978). Evolutionarily Stable Strategies and Game Dynamics. *Math. Biosci.*, **40**, 145-156.

Thomas, B. (1985). Genetical ESS-Models. I. Concepts and Basic Model. *Theor. Pop. Biol.*, **28**, 18-32.

Von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behaviour*. Princeton: Princeton University Press. 2nd ed. 1947; 3rd ed. 1953

Zeeman, E. C. (1979). Population Dynamics from Game Theory. In: *Proc. Int. Conf. Global Theory of Dynamical Systems* (ed. Z. Nitecki and C. Robinson). Heidelberg: Springer-Verlag.

Zeeman, E. C. (1981). Dynamics of the Evolution of Animal Conflicts. *J. Theor. Biol.*, **89**, 249-270.

Appendix: The REDUCE program from Section 4.4.5

```
nmax := 20;
array a(nmax,nmax),i(nmax,nmax),b(nmax),p(nmax);
comment a(l,m);
FOR l:= 0:nmax DO (
  FOR m:=0:l DO (
    a(l,m):=(-2^(l-m)/(2l+1))*(for j:=(m+1):l PRODUCT j/(2j-1))
  )
);
comment n = 0;
FOR l:= 0:nmax DO (
  i(0,l) := b(l) := 2^l * (FOR m:=1:l PRODUCT m)
                  / (FOR m:=0:l PRODUCT (2m+1))
);
comment n = 1;
FOR l:=0:nmax DO (
  i(1,l) := pi*b(l) + 2*(FOR m:=0:l SUM (a(l,m)/(2m+3)))
);
comment n > 1;
FOR n:=2:nmax DO (
  FOR l:=0:(nmax-n) DO (
    i(n,l) := pi^n * b(l) + 2n*(pi/2)^(n-1)*(FOR m:=0:l SUM (a(l,m)/(2m+3)))
          + 4n*(n-1)*(FOR m:=0:l SUM (a(l,m)*i(n-2,m+2)/(2m+3)))
  )
);
FOR n:=0:nmax DO (
  write( p(n) := 2pi*(1-i(n,0)/pi^n))
);
on fort;
off period;
FOR n:=0:nmax DO (
  write( p(n) := 2pi*(1-i(n,0)/pi^n))
);
on numval,float;
pi;
off numval;
FOR n:=0:nmax DO (
  write( p(n) := 2pi*(1-i(n,0)/pi^n))
);
end;
```