MODELS FOR ANIMAL MOVEMENTS

Peter Leith Chesson, B.Sc. (Hons).

Department of Statistics,
The University of Adelaide.

Submitted for the degree of Ph.D.
December, 1976.

CONTENTS

Summary iv
Declaration vi
Acknowledgement vii

1 Introduction 1
  The interest of movements in population ecology 3
  Models for animal movements 6
  Why model? 9
  The models in this thesis 11

2 Discrete space models 17
  2.1 Two simple deterministic models 25
  2.2 The basic stochastic models 29
  2.3 Differences between individuals 43
    The distribution of $\underset{\sim}{N}(t)$ as $t \to \infty$ 52
    Continuous time models 58
    A construction 66
  2.4 Environmental and other effects 68
    Special cases 72
    Conclusion 74
  2.5 Movements and asymptotic distributions 76
    Means and variances 76
    Covariances 77
    Asymptotic distributions 79

3 General space 81
  3.1 The general structural model 82
  3.2 The number of animals in an area 87
    Generating functionals 97
    Does N_t preserve structure? 100
  3.3 Joint distributions in time 105
  3.4 The limits of the distribution of N_t 111
    Limits of joint distributions 128
  3.5 Construction of models from Markov processes 128
  3.6 The population is a sample from a large finite population 134

4 Finding Markov process models 142
  4.1 Self adjoint transition operators 144
    Processes dependent on movements 147
  4.2 Models based on diffusion processes 150
    Markovian kineses 154
    Movement in response to a potential 156
    Passive movement by currents 157
  4.3 Other methods of generation and modification of self adjoint transition operators 157
    Bivariate distributions and Markov processes 158
    Transformation of the state space 167
    Subordination of Markov processes 169

5 Modelling in finer detail 175
5.1 Continuous piecewise linear paths 177
5.2 Differences between individuals and asymptotic properties 185

6 Application and inference 198
6.1 Testing independence 198
Testing independence in a radially symmetric environment 206
A test of independence 213
Derivation of the asymptotic distribution of T_m 214
6.2 Observations on *Artemia salina* (L.) 224
Apparatus and experimental conditions 225
Testing independence 229
Test of discrete space models 232
Continuous space experiments 246
Conclusion 258

Appendix 1 Formulae for Chapter 6 260

Appendix 2 Discrete space experiments 265

Appendix 3 Continuous space experiments 291

Appendix 4 The canonical decomposition of bivariate distributions 312

Bibliography 330

Figures

1 Idealised patchy environment 20
2 Projection of a helical path onto a circle 212
3 $\Phi^{-1}(\hat{F}_{60})$ 222
4 $\Phi^{-1}(\hat{F}_{120})$ 223
5 Chamber A 226
6 Chamber B 227
7 Scatter plot for the test of radial symmetry 230
8 Discrete space 233
9 Continuous space 247

SUMMARY

Animal movements are being recognised increasingly as important in population ecology. Whilst this thesis concentrates on animal movements as a worthwhile study for its own sake, it is hoped that the ideas presented are useful to population ecology.

Consider a population of n animals living in a region $\mathcal{X}$. Let $X_\ell(t)$ be the position of the ℓth animal at time t. The functions $X_1,..,X_n$ can be thought of as stochastic processes and it is the purpose of this thesis to

1. develop useful models for the joint distribution of $X_1,..,X_n$;
2. develop a general theory for the study of a broad class of models;
3. investigate ways of applying models and testing their validity;
4. demonstrate that the theory developed is actually useful in the study of real animal populations.

Consideration is restricted to the situation where $X_1,..,X_n$ are conditionally independent given the environment. A general theory for models satisfying this property is developed with the main emphasis on the random measure N_t, where $N_t(G)$ is the number of animals in the set G at time t. The extent to which the distribution of N_t characterises the distribution of $X_1,..,X_n$ is investigated. Three main classes of random measures arise and it is shown that each of these classes is closed in the weak topology. Ideas from the theory of weak convergence are also found useful in identifying situations where approximate independence, conditional on the environment, can be expected.

Discrete space models, where X is effectively finite, provide simple examples of models within the framework of the general theory. The models given are modifications of existing Markov process models and aim to incorporate the following observations of real animal populations.

a. Animals differ from each other in the distribution of their movement processes.

b. Animals have random fluctuations in time in their rates of movement.

c. A randomly varying environment induces dependence between different animals' movements.

To construct analogous models for continuous spaces, Markov transition operators are required. It is argued that the most useful of these are self adjoint and have a countable spectral decomposition. Such operators are investigated with a view to generating flexible classes of operators to facilitate the fitting of models to real data.

Sometimes Markov processes are too coarse to be useful models. Thus models for the situation where animals have piecewise linear paths are developed. The covariance structure for these models is determined and they are shown to be asymptotically consistent with Markov process models.

Laboratory studies on the Brine Shrimp *Artemia salina* illustrate the application of the theory. Existing statistical techniques are used where ever available but for the test of independence conditional on the environment, it is necessary to develop a new procedure. For this purpose a rank test assuming environmental symmetry is given.

Declaration.

This thesis represents my own work and no part of it has been submitted for any other degree. As far as I know and believe no material is included which has been published or written by any other person except where reference is made in the text.

Peter Chesson, 21 December, 1976.

Acknowledgements.

Among the many people who helped me I want to thank particularly my supervisors Dr. A.J. Butler and Dr. G.M. Tallis, the friendly people in the Department of Statistics and Zoology, the Department of Physics for the use of the Bucone film reader, and Jean Chesson for writing computer programs and generally assisting putting this thesis in its physical form. Finally I am indebted to Katherine Halsey for cheerfully working long hours to complete the typing of the manuscript.

1 INTRODUCTION

This study is about the development of models for the scientific investigation of animal movements.

Movement is a characteristic of animals. At some stage in their lives all animals move about from place to place. There are several aspects of movement. For example one can consider the mechanics of movement (locomotion), the guidance of movement (animal orientation and navigation), the function of movement, the effect of movement and the description of movement. This thesis deals almost entirely with the last aspect. To describe movements we consider the path traced by an animal as a function of time. We can try to describe the joint movements of the animals in a population or simply concentrate on the movements of an individual. Both ways of looking at movements are considered here but the latter is included mainly for its bearing on the former.

By a population I mean a group of animals, generally of the same species but possibly from a special subset such as females or adults. Occasionally we shall come across organisms not regarded as animals, in modern systematics, but whose motility suggests their inclusion in the discussion. Examples are some bacteria, and the amoeboid stage of cellular slime molds.

Three basic kinds of movements can be recognised. There are the ordinary or "every day" movements, forming part of the normal activities of living, and there are two kinds of extraordinary movements which I shall call dispersal and migration.

Dispersal is a movement occurring infrequently in the life of an individual which takes it away from an area where it has lived to another area, this latter area being unpredictable.

Migration also occurs infrequently in an individual's life but takes it away to another area which is reasonably predictable. The seasonal movements of the Monarch Butterfly (Johnson, 1969, p.533) is an example of migration. Seasonal movements of birds and fish commonly called "migrations" also come within this definition. The boundaries of these definitions are not precise and some movements may be difficult to place. Note also that it is possible for daily movements to be dispersive in the sense that they tend to spread the population more evenly in space, particularly when the animals do not have homes, even though daily movements are regarded as distinct from dispersal.

Johnson (1969) discusses the variation in meaning of the terms migration and dispersal as used by different authors as well as giving a comparative discussion of different kinds of movements and his own classification. Southwood (1962) gives a different classification. The classification chosen here is that which I believe is most suitable from a modelling point of view.

The theory and models in this thesis are developed with ordinary movements in mind which is the reason why it is often assumed that stochastic processes describing movement can be obtained by a transformation of a process which is asymptotically stationary, however that is not to say that such models are always applicable to ordinary movements and never applicable to extraordinary movements.

Many parts of the thesis, especially chapter 3, are applicable to all three kinds of movements but the orientation is probably best suited to ordinary movements.

The final general restriction on the subject of this thesis is that it assumes that the population of animals under consideration is invariant in time meaning no new individuals are added to the population and none are lost. Thus, we can either assume that the time interval of consideration is so short that few such changes will take place or, in a probabilistic sense, condition on such changes not occurring. In practical studies one may, when possible, discard from consideration new individuals or individuals which die or disappear, or break up the period of observation into intervals within which the population does not change. The reason for imposing this restriction is the belief that movements are sufficiently complicated themselves without trying to model changes in population simultaneously. The latter problem is probably even more complicated but it is a problem to which I hope the study of movements will contribute.

The interest of movements in population biology

Most animals are unable to satisfy all the needs of their lives at the one point in space so that movements are important to them. People are interested in how animals move and particularly how well movements enable them to satisfy these needs. For example, Dudzinski (1969) and Young (1972) consider the movements of grazing animals in relation to food and water and how these are affected by environmental conditions. Mosimann (1958) and Philip (1957) consider the role of movements in the ability of organisms to find mates and Palaheimo (1971) and Murdie and Hassel (1973) study the ability of a predator to find its prey.

Population ecologists have long been interested in movements of a dispersive kind. First of all the movement capabilities of organisms determine their abilities to extend the limits of their distribution (Krebs 1972) and it is almost tautological to say that movements have an important effect on the density of organisms within the limits of their distribution.

Andrewartha (1970, p.190) believes that animals need to be constantly dispersing as a normal part of their lives and that those individuals able to disperse well will be favoured by natural selection. Andrewartha discusses the role of movements in population dynamics in a changing environment where organisms need to find new places to live. Even though the present environment is still favourable it is advantageous for some individuals to seek new homes. See also Hutchinson (1951). As an example Andrewartha cites the case of a predator pressing heavily on its prey where the relative dispersive abilities of the predator and prey play a particularly important role.

Movements of one species can have significant effects on the density, dispersal, and spatial pattern of other organisms. This is clear in predator prey example cited above and we may also add that dispersive ability is regarded as an important factor in determining the success of an agent of biological control (Krebs 1972, p.373).

The movements of vectors of disease are clearly an important factor in the spread of disease (Krebs 1972, p.26, Andrewartha 1970, p.69) and Krebs (1972 pp. 519-520) explains the hypothesised role of movement behaviour of seed and seedling eating animals in the spatial pattern of forest trees.

Even though movements may be of no direct interest to a person they often enter problems of estimating population size (Arnason 1970, Andrewartha, 1970, p.166-167) and are particularly important if one wishes to estimate spatially dependent mortality and fecundity (Beverton and Holt, 1957, part III, Dempster, 1957).

For the development of models in population biology interest in movements seems to fall into two categories, needless to say these categories are not completely distinct. First field environments commonly exhibit spatial variation so that mortality, fecundity and the fitness of genotypes have come to be regarded as functions of the position in space. People develop models for the ways spatial variation and movements interact to determine the characteristics of populations both as functions of position in space and of the population as a whole. For example, Roughgarden (1974) studies a model incorporating spatially varying carrying capacity for which he shows that the effect of movements is to produce a local averaging of carrying capacity. In theories of speciation movements enter the question of the degree to which spatial variation in selective forces can be reflected by spatial variation in gene frequency (Slatkin, 1973). Skellam's now classical study of the effects of movements (Skellam, 1951) considers the combined effect of spatial variation and movements on population density. There are many other studies of these kinds of which a few are Comins and Blatt (1975), Fleming (1975) and Roff (1974 a and b).

When spatial variation is disregarded movements remain of interest in modelling the degree to which subunits of the population are isolated by distance. The studies of Hadeler et al (1974) and Murray (1975) seek to discover whether movements and population interaction can combine to produce spatial variation in population density. In population genetics there has long been interest in the effect of movements on inbreeding and the so-called effective population size (Wright 1969, chapter 12). A related matter is the correlation between gene frequencies at different points in space (Fleming and Su, 1974).

Richardson (1970) lists other problems in population biology where movements are of interest. See also Sanderson (1966), Brown and Orians (1970), Southwood (1962) and Johnson (1969).

Models for animal movements

In all of the models for movements in the population studies cited above individual animals move independently according to Markov processes all having the same transition probabilities, or if the model is deterministic, as is the case with most, it is the deterministic analogue of some such stochastic model - section 2.1 explains what is meant by a deterministic analogue of a stochastic model. The Markov processes underlying these models are either simple random walks, diffusion processes - mostly with zero drift and constant variance - or if the space is divided into a finite number of areas the processes are homogeneous Markov chains. The majority of movement models are of these kinds. Arnason (1970) reviews the literature to that date.

When the primary interest is movements the models are sometimes more sophisticated, understandably because one does not have to cope with other complications. Stroock (1974), Nossal and Weiss (1974) and Lovely and Dahlquist (1975) independently developed stochastic models for bacterial chemotaxis. In these models the path traced by an individual is ~~piecewise~~ PIECEWISE linear and is not a Markov process (chapter 5). Non Markov models of a simpler kind, where an animal moves in a straight line after a randomly chosen initial direction, are sometimes used (Wilson and Findley, 1973 and Cox and Smith, 1957).

In a few models the movements are density dependent. For discrete space these are Arnason (1970), Roff (1974a,b) and Whittle (1967) and continuous space examples are the deterministic models of Kiester and Slatkin (1974) and Gurney and Nisbet (1975). An interesting stochastic model for the aggregation of amoebae in the formation of cellular slime molds is developed by Novak and Seelig (1976).

Saunders (1975) considered introducing random environmental variation into the stochastic compartment model which is commonly used to describe movements of animals between discrete areas (Arnason, 1970). It is not difficult to show by the methods employed in this thesis that Saunders' model is equivalent to conditional independence of the movements of different animals, given their environment.

There is one final class of population models involving movements that should be mentioned, namely the Markov branching-diffusion process models. The movement processes in these models are generally equivalent to the specification that animals move independently of each other according to Markov processes. A review is given in the article by Davis (1970).

The development of theoretical models in population biology has vastly outpaced the empirical testing of such models. In science the testing of models is a fundamental process. If a model has any meaning at all we should want to determine the validity of its major assumptions and since movement forms an important part of so many models it is essential to see how good the movement components of these models really are. Unfortunately, most movement studies that have been carried out are incapable of giving a reasonable test of the movement process involved or the experimenter has no movement model in mind to compare with the data. One of the major difficulties is the collection of adequate data in the field ; however, for some animals, and experimenters with sufficient funds and/or ingenuity technological innovations may overcome these problems (Siniff and Jessen, 1969, Marten, 1972, 1973 and Mossing, 1975).

The earliest and most common movement models are of the random walk or Brownian motion kinds. The best evaluation of these models comes from studies of *Drosophila* reviewed and analysed by Crumpacker and Williams (1973). See also Wright (1968) for an interpretation of the data of Dobszhansky and Wright (1947) in terms of random walk. An optimistic account of the early tests of diffusion models is given by Beverton and Holt (1957).

More recently Berg and Brown (1972) conducted a detailed study of the movements of a bacterium and although they had no model, their data provided a good basis for model construction resulting in the independent appearance of three different models (chapter 5).

In discrete space settings the most complete tests of models are those given by Bryant (1969) and Wierzbowska (1972). Other workers give evaluations of a less complete nature, for example Siniff and Jessen (1969), Kitching (1971), Jones (1959).

Why model?

In this thesis a model is a mathematical statement meant to approximately represent real world phenomena. The utility of a mathematical formulation is that it enables precise and complex reasoning otherwise not possible and also the language of mathematics enables models to be useful summaries of knowledge or belief.

There are different kinds of models depending on the purpose. First of all there are the highly specific models which are developed in relation to specific populations, are fairly complicated and not very amenable to transferrance to other situations. Maynard Smith (1974) referred to such models as "simulations". Secondly, there are models of an intermediate kind which are less complicated and can be transferred to different situations with a change of parameters but no drastic modification. They are generally intended to be fitted to data. Finally, there are theoretical models which are not generally intended to be fitted to data.

Theoretical models can be of two kinds. There are models which make very highly specific assumptions so that they completely describe a system apart from a finite number of parameters - the classical Lotka-Volterra predator prey model is an example - and theoretical models which do not specify the system completely but seek to explore the relationships between specified postulates. Models of the latter kind are developed by May (1973). In this thesis we seek models of the intermediate kind and of the second kind of theoretical model.

There are many uses for models. They may be used as a means of analysing data and also suggesting the kinds of data needed for a specific purpose. Some models may make predictions of direct practical value or suggest further lines of research. Models lead to precise formulation of vague theoretical or even practical notions and they enable the testing of the compatibility and consequences of assumptions and the adequacy of explanations ; in short they are part of theory building. I am going to advocate a more mundane use of models as well.

A large part of scientific research involves a somewhat random accumulation of knowledge. This is necessary for a foundation on which to build research that will ultimately be recognised as useful. However, a large amount of research becomes lost because the purpose and results are too vague to be conveniently summarised and communicated beyond the original report. According to Sanderson (1966) animal movement studies are no exception.

Models are helpful in overcoming this problem because they enable the identification of specific questions relating to the assumptions and predictions of the model. Because specific and well defined questions are asked the identification, summary and communication of the results is facilitated.

The models in this thesis

As has been pointed out most models of animal movements make assumptions equivalent to assuming that different animals move independently. Gurney and Nisbet (1975) regarded this as an untenable hypothesis. However, it must be remembered that Gurney and Nisbet are working with an ecological time scale, that is significant changes in population size take place over the period of interest to them ; and the space scale is such that only movements which are dispersive are of concern, especially dispersal if it leads to the establishment of a new home. When competition for space exists movements of this kind are likely to be dependent. This does not mean that ordinary movements will be dependent, but, of course, they can be (see Kiester ~~Keister~~ and Slatkin, 1974). Some animals interact very little or infrequently with other members of the same species. This is likely to be so in sparse populations of solitary organisms as a consequence of infrequency of meeting. However, one does not need to postulate infrequency of meeting to expect that there is not likely to be significant interaction, but simply that the individuals or the population as a whole gain nothing from such behaviour. People are often led to believe that animal movements will be dependent because they are very obviously so in some common vertebrates such as mammals and birds, but for invertebrates dependence is much less obvious.

In most studies of animal movements no comment at all is made about dependence or independence unless dependence involving interesting behaviour is apparent. See Jewell and Loizos (1966) for some interesting examples of dependence.

Only occasionally does one find statements of independent behaviour, such as the following statement of Dobzhansky and Wright (1943).

> Active dispersal involves movements, usually uncorrelated in direction in different individuals of a given insect species between places where these individuals are born and where they find food water and oviposition sites.

It is not entirely clear what kind of movements Dobzhansky and Wright are referring to but from the purpose of their investigation and the kind of data they discuss I am led to believe that at least for *Drosophila* these are what I have called ordinary movements.

Wilkinson (1952) decided to regard homing behaviour in birds as independent because birds come home singly even when released in a group.

For a long time there has been considerable interest in "randomness" of spatial distributions of organisms (Southwood 1972). The kinds of dependence expected of animals should be apparent from spatial distributions in cases where the animals regard the space as completely homogeneous (chapter 6). However, the spatial homogeneity condition is rarely satisfied and so it is not possible to determine dependence from tests of randomness except in the case where the distribution is too regular (overdispersion) which may suggest that the animals stay away from each other.

This thesis mostly assumes that there is little interaction between individuals in the population but it stops short of assuming complete independence because if animals' movements depend on an environment which is randomly varying one expects movements to be dependent (section 2.4). However, one can speak of independence conditional on the environment. For this case I present a variety of models containing features which are believed to be useful for the study of animal movements but are not found in any existing models. The thesis tries to present a unifying theory for such models and at the same time some progress is made in the problems of practical application.

For the case where animals interact significantly the task of developing suitable models seems quite difficult because of the enormous variety of ways in which individuals might interact. However, in some cases it may be possible to reduce dependence to a kind of conditional independence. Section 3.6 makes some progress in this direction for large populations where individuals are not distinguished. Brown (1966) describes the interactions in a community of field mice (*Apodemus sylvaticus*). The dominant male exerts a very great influence on the rest of the individuals in the population. Brown makes little mention of interaction between other individuals and so as a first approximation we might model the movements of the rest of the population as conditionally independent given the movements of the dominant male.

This thesis is written with the hope that a reader evaluating the results for personal use will be prepared to adapt and modify, and where necessary, to think locally in time and space and to follow the principle that it is better to have a model with known deficiencies than no model at all.

Chapter 2 presents a method for modelling the movements of animals between a finite number of areas taking into account differences between individuals, in the parameters of their movement processes, and the effects of random environmental variation. As is shown in chapter 3, the ideas presented do not really depend on the finiteness of the space ; but the finite space setting enables an elementary presentation as well as producing some useful models which are specific examples from the broader class of models discussed in chapter 3. The models of chapter 2 are likely to be the easiest to apply. They are less demanding in the kind of observations required to estimate their parameters and since almost any desired degree of parameterisation is available they are particularly useful for handling spatially complex situations.

Chapter 3 develops a theory of animal movement models satisfying the requirement of independence conditional on the environment. It can also be regarded as an investigation of the structure obtained under the postulate of independence conditional on the environment. In addition to extending the ideas of chapter 2 to general space an investigation of the interrelationships between different simplifications of the general structure is made. The purpose is to help clarify thought so that we know when different models are really the same, or are quite distinct ; and so we can tell what kinds of observations might be used to distinguish between models or to estimate their parameters. Results of this nature are also considered in chapter 2 but the approaches in these two chapters are meant to be in part complementary in both the results proved and the methods of proof.

Chapter 3 suggests a general method of constructing models for animal movements on the basis of Markov processes and so chapter 4 takes up the challenge of finding suitable Markov processes, treating this basically as the problem of finding suitable Markov transition operators in a Hilbert Space setting. A discrete spectral representation is the most convenient form in which to obtain an operator for application to the models in the thesis and also for many other purposes. The discussion is restricted to the self-adjoint case which arises for several different models of movement behaviour.

In continuous time and space, models based on Markov processes are necessarily a little crude, for observations of a qualitative nature reveal discrepancies. Sometimes animals' movements are amenable to modelling more closely, for example when the paths traced by the animals consist of connected straight line segments. Chapter 5 discusses models of this type, which are referred to as fine detail models, for the case of uniform space. Fine detail models are also found particularly useful for discussing models of differences between individual animals' movements.

Chapter 6 shows how the theory of the previous chapters is helpful in learning about the movements of real animals.

In this chapter a test for independence of different animals' movements is developed and applied to experiments using the Brine shrimp, *Artemia salina*. Further experiments examine the validity of the discrete space models of chapter 2 for *Artemia salina*, and also the utility of the ideas in chapter 4 for studying movements in continuous space.

2. DISCRETE SPACE MODELS

It has long been recognised that the spatial environment of many animal populations contains discontinuities or is "patchy". The ecological significance of this property has been discussed by Andrewartha and Birch (1954), for example, and more recently by Levin (1974). Levin believes that even if the environment is initially uniform the activities of the animals will tend to make it non uniform, that is, patchiness should be widespread.

Patchiness is sometimes thought of as the existence of areas favourable to a particular species, in the midst of unfavourable areas, so that the population divides into subpopulations existing in favourable areas (e.g. Kitching, 1971). Sometimes patchiness is simply a subdivision of the spatial environment into areas or patches having different qualities in terms of their effect on the animal population (e.g. Doyle, 1975).

The main ecological interest in a patchy environment results from the movements of organisms, at some stage in their lives, from one patch to another, for, if there were never such movements each patch could be treated largely as an independent system. Dependence would only appear through exchanges of non living material.

Models of movements between patches, or discrete areas, have been incorporated in many theoretical population biology investigations. Examples are Gadgil (1971), Roff (1974a, 1974b) and Chewning (1975), who study growth and stability of single species populations, Levin (1974) who investigates competition, and Tallis (1966) and Gillespie (1975) in genetical studies. These authors all consider a finite number of discrete areas. Bailey (1968) and Adke (1969) have considered birth death and migration on countable discrete spaces, and Kimura and Weiss (1964) have undertaken genetical studies for such spaces. The main interest in all of these investigations is the effect of movements on some other property of the populations. The movement process itself does not receive much attention and the movement models are generally very simple In particular movement is often treated deterministically in the manner of the deterministic models presented in this chapter, or if stochastic it is essentially model 2c of this chapter.

For the discrete space models I discuss, it is assumed that the region in which the animals' movements take place is partitioned into k disjoint "areas". The areas are referred to as area 1, area 2, etc. No assumptions are made about the nature of the areas for the development of the theory since movement within an area is ignored. However in a given practical situation it will be important to consider the nature of the areas in conjunction with the behaviour of the animals in deciding if it is reasonable to neglect movements within areas. In this regard I believe the discrete space approach will be more appropriate when the boundaries of the areas correspond to genuine discontinuities in the environment which are perceptible to the animals in such a way that the crossing of a boundary is an active process. Because of the underlying Markov nature of the models presented here, the situations in which the models should be best are when either an animal comes near the boundary between areas many more times than it actually crosses, or the crossing of the boundary is a process which is largely independent of movements within an area.

In some cases it may be appropriate for the areas to be disconnected sets. Consider a region made up of many patches of a few types for which interest centres on movements between different types of patch. Furthermore suppose that, to a reasonable approximation, for each given type of patch the accessibility of all types of patch does not vary from patch to patch. It would seem quite reasonable in this situation to regard an area as the union of all patches of a particular type. The areas are thus likely to be disconnected; see fig. 1. The advantage in defining the areas in this way is hopefully a more manageable model retaining the detail of interest.

In a situation where the Markov postulates underlying the theory I develop are not appropriate, but a second or higher order Markov postulate is, some of the theory can be salvaged by altering the definition of the areas. As an example consider a discrete time second order homogeneous Markov process X_n, then the vector process (X_{n-1},X_n) is a Markov process, so that to apply Markov process theory one alters questions of the kind "$X_n \in$ area i?" to questions of the kind "$(X_n,X_{n-1}) \in$ area i × area j?". Hence in this situation $\{\text{area } i,\ i = 1,\ldots,k\}$ is replaced by $\{\text{area } i \times \text{area } j,\ i = 1,\ldots,k,\ j = 1,\ldots,k\}$.

Finally one may wish to use a discrete space model in a genuinely continuous situation, as an approximation, particularly if the space is very complicated. Discrete approximation is the basis of many numerical methods and

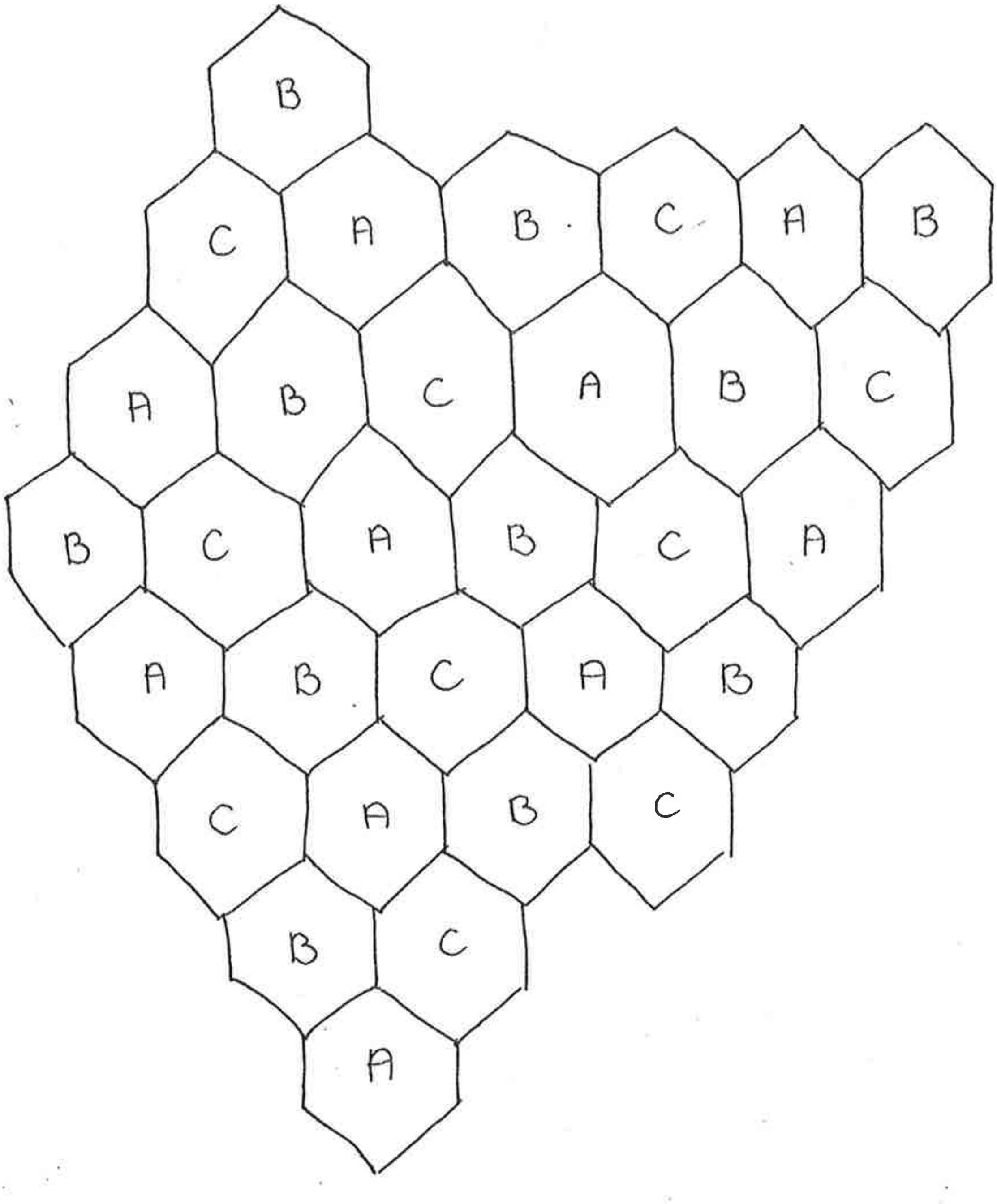


Fig. 1. Idealised patchy environment.

I believe it can be similarly useful in movement studies. Discrete space approximation is a frequent practice in modelling biological systems. As an example consider the Leslie matrix model of population growth which is developed for discrete age classes but is also very useful in approximating continuous age situations. In the context of fish movement Beverton and Holt (1957) discuss the application of a discrete space model to a genuinely continuous situation. They attempt to justify the use of model 1d given here by appeal to the asymptotic behaviour of a diffusion process.

The models in this thesis assume that there are n animals in the region as a whole, and n is fixed. In order to speak of the ℓth animal the animals are regarded as numbered 1 to n but it is certainly not assumed that individuals are distinguishable to an observer. A number of variables are associated with these animals. $X_\ell(t)$ is the number of the area that the ℓth animal occupies at time t. $\underset{\sim}{X}(t) = (X_1(t),..,X_n(t))'$. $N_j(t)$ is the number of animals in area j at time t. $\underset{\sim}{N}(t) = (N_1(t),..,N_k(t))'$. In the stochastic models these variables are random variables and it will be assumed in any discussion that all random variables mentioned are defined on a common probability space (Ω,A,P). Sometimes time is discrete and sometimes continuous. As a notational convenience we shall write $\underset{\sim}{N}$ for the process $\{\underset{\sim}{N}(t), t \geq 0\}$ or $\{\underset{\sim}{N}(t), t=0,1,..\}$ etc.

The approach to stochastic models in this chapter is basically naïve in postulating properties of processes without worrying too much about existence of such processes. Generally existence is not difficult to establish or is a standard result. One construction for a less obvious case is performed.

The emphasis is on the conditional distribution of $\underset{\sim}{N}(t)$ given $\underset{\sim}{N}(0)$, that is the approach is basically predictive although prediction need not be one's purpose. It is mainly a convenient vehicle by which to discuss the different models without becoming involved in too many complications. The process $\underset{\sim}{N}$ is singled out because it is probably the most interesting biologically and often the easiest to observe, however it may not be the best for experimentally determining movement behaviour. This is taken up further in chapter 3.

In cases where $\underset{\sim}{N}$ is a homogeneous Markov process the conditional distribution of $\underset{\sim}{N}(t)$ given $\underset{\sim}{N}(0)$, together with a distribution for $\underset{\sim}{N}(0)$ specifies the entire distribution of $\underset{\sim}{N}$. For other cases the asymptotic on n distribution of $\underset{\sim}{N}$ is given, but if results are required for finite n it is generally not difficult, although possibly tedious, to obtain them.

We need some notation and definitions.

A Markov matrix P is a $k \times k$ non negative matrix for which $P\underset{\sim}{1} = \underset{\sim}{1}$; $\underset{\sim}{1}$ is the vector $(1,..,1)'$. A rates matrix Q is a $k \times k$ real matrix with non negative off diagonal elements satisfying $Q\underset{\sim}{1} = \underset{\sim}{0}$, (a_{ij}) is the matrix with ij^{th} element a_{ij}. diag $(\underset{\sim}{p})$ is the diagonal matrix satisfying $\text{diag}(\underset{\sim}{p})\underset{\sim}{1} = \underset{\sim}{p}$. $I = \text{diag}(\underset{\sim}{1})$. $\underset{\sim}{e}_j$ is the j^{th} column of I. P' is the transpose of P and if Q is a matrix

$$e^Q = \exp\{Q\} = \sum_{n=0}^{\infty} \frac{Q^n}{n!}$$

The indicator function of a set A is denoted by I_A and is defined by

$$I_A(x) = \begin{cases} 0, & x \in A \\ 1, & x \notin A. \end{cases}$$

A multinomial random variable $\underset{\sim}{Y}$, with parameters $(m,\underset{\sim}{p})$, is a multivariate discrete random variable having probability generating function

$$\sum_{\underset{\sim}{y}} \prod_{1}^{k} s_i^{y_i} \, P(\underset{\sim}{Y}=\underset{\sim}{y}) = (\underset{\sim}{p}'\underset{\sim}{s})^m ,$$

m is a positive integer and $\underset{\sim}{p}$ is a non negative vector summing to 1. If $m = 1$ the distribution is called multiple Bernoulli with parameters $\underset{\sim}{p}$.

If $\underset{\sim}{Y}$ is a convolution of multinomials then $\underset{\sim}{Y}$ has probability generating function

$$\prod_{i=1}^{r} (\underset{\sim}{p}_i'\underset{\sim}{s})^{m_i}$$

for some $r \times k$ matrix P with rows $\underset{\sim}{p}_i$. We use the notation

$$M(\underset{\sim}{m},P,\underset{\sim}{s}) \overset{\text{def}}{=} \prod_{i=1}^{r} (\underset{\sim}{p}_i'\underset{\sim}{s})^{m_i}$$

Let $\underset{\sim}{U}$, $\underset{\sim}{V}$ be vector valued random variables then the variance matrix $V\underset{\sim}{U}$ of $\underset{\sim}{U}$ is defined as

$$E(\underset{\sim}{U}-E\underset{\sim}{U})(\underset{\sim}{U}-E\underset{\sim}{U})' = E(\underset{\sim}{U}-E\underset{\sim}{U})\underset{\sim}{U}' = \big(E\ (U_i-EU_i)(U_j-EU_j)\big)$$

The covariance matrix of $\underset{\sim}{U}$ and $\underset{\sim}{V}$ is

$$C(\underset{\sim}{U},\underset{\sim}{V}) = E(\underset{\sim}{U}-E\underset{\sim}{U})(\underset{\sim}{V}-E\underset{\sim}{V})'$$

A stochastic process $X = \{X(t),t \geq 0\}$ or $\{X(t),t=0,1,\ldots\}$ with a finite state space J will be called a homogeneous Markov chain if for every $j \in J$, and $t,\ s \geq 0$, $P\big(X(t+s)=j \,|\, X(u),u \leq t\big) = p\big(s,X(t),j\big)$ for some fixed function p of three variables.

Occasionally we shall need the notion of a regular conditional distribution.

Let $(\Theta_1,\mathcal{B}_1)$ and $(\Theta_1,\mathcal{B}_2)$ be measurable spaces.

A function $q : \Theta_1 \times \mathcal{B}_2 \to [0,1]$ is called a regular conditional distribution if $q(\theta,\cdot)$ is a probability measure on (Θ,B) for every $\theta \in \Theta_1$ and $q(\cdot,B)$ is a measurable function on $(\Theta_2,\mathcal{B}_2)$ for every $B \in \mathcal{B}_1$. Other common names for q are "transition measure" and "random probability measure".

If π is a probability measure on $(\Theta_1,\mathcal{B}_1)$ then by a well known result a probability measure ρ on

$$(\Theta_1{\times}\Theta_2,\mathcal{B}_1{\times}\mathcal{B}_2)$$

is defined by the equation

$$\rho(B_1{\times}B_2) = \int_{B_1} q(\theta,B_2)d\pi(\theta)$$

for $B_1 \in \mathcal{B}_1$, $B_2 \in \mathcal{B}_2$.

Let $(\Omega,\mathcal{A},\mathcal{P})$ be a probability space and Ψ_1 and Ψ_2 be $(\Theta_1,\mathcal{B}_1)$ and $(\Theta_1,\mathcal{B}_2)$ valued random variables, i.e. $\Psi_i^{-1}(\mathcal{B}_i) \subset \mathcal{A}$, if q is a regular conditional distribution such that

$$q(\Psi_1,B_2) = \mathcal{P}(\Psi_2{\in}B_2|\Psi_1)$$

for all $B_2 \in \mathcal{B}_2$ then q is called a regular conditional distribution for Ψ_2 given Ψ_1.

2.1 Two simple deterministic models.

I give here the only deterministic models presented in the thesis; but they serve to illustrate some of the major principles of deterministic modelling of animal movements.

The advantage of a deterministic model is that it is generally simpler to develop than a stochastic model. On the other hand there is so much variability apparent in the majority of movement processes that one may wonder whether deterministic models can be of much use at all. Suppose we are observing a large population, perhaps an invertebrate population, and we are interested in the numbers of animals in different areas. Provided the dependence between individual animals is not too great it can be expected that some law of large numbers will ensure that the proportions of the animals in different areas are approximately deterministic. The same applies to numbers of animals if we consider only relative errors.

In the above example a stochastic element plays an essential role in allowing different animals to be in different areas. The limiting values of the proportions are really probabilities that randomly chosen animals are in those areas. It is easy to see that a deterministic model for these proportions is the means of the proportions in some stochastic model. Conversely every stochastic model yields a deterministic model by taking the means of the stochastic proportions. Thus to obtain deterministic models from the stochastic models in this thesis simply evaluate the appropriate means. This thesis emphasises variability and the stochastic element is generally essential to the formulation of the model but that does not mean that in the presence of large numbers of animals it can not be treated as deterministic.

In continuous space deterministic models, a function giving the density of animals at each point in the space takes the place of proportions. Dividing this density by the total number of animals we obtain a probability density. Lapidus and Schiller's 1975 study of bacterial chemotaxis is a good example of a continuous space deterministic model. The density function in their model arises from the probability density for a diffusion process describing an individual animal's movements. This kind of deterministic movement model is frequently incorporated in population models involving other processes, for example Skellam's work (1951).

In some cases the paths traced by animals' movements are more important than proportions. This is often the case for long distance migration of animals where sometimes the variation from the mean path is insignificant compared with the distance travelled and so can be ignored. Deterministic models of long distance migration of fish have been developed by Seckel (1972) who reports that field observations showed variation from the mean path could be neglected.

Model 1d

Suppose that at each time point, $t = 0,1,2,\ldots$ a proportion p_{ij} of the animals in area i move to area j. Basically this says that the movements are not density dependent, are homogeneous in time and future movements depend only on current positions. If there are m_i animals to begin with in area i, $i = 1,..,n$, then at time 1 there are $\sum_i m_i p_{ij}$ animals in area j. $N_j(t)$ is the number of animals in area j at time t and so for $\underset{\sim}{N}(t)$ we obtain

$$\underset{\sim}{N}(t) = (P^t)'\underset{\sim}{m} \tag{1}$$

where $P = (p_{ij})$. For any times $s < t$

$$\underset{\sim}{N}(t) = (P^{t-s})'\underset{\sim}{N}(s) \tag{2}$$

Since animals are not allowed to be lost or gained we stipulate $p_{ij} \geqslant 0$ and $\sum_j p_{ij} = 1$, that is, P is a Markov matrix.

Model 1c

A continuous time analogue to 1d says that the instantaneous rate of movement from area i to area j, per animal in area i, is a constant q_{ij}. That is

$$\frac{d}{dt}N_j(t) = \sum_i N_i(t)q_{ij}.$$

Since animals are conserved $q_{ij} \geqslant 0$ for $i \neq j$ and $\sum_j q_{ij} = 0$ The solution of these equations, subject to $\underset{\sim}{N}(0) = \underset{\sim}{m}$, is

$$\underset{\sim}{N}(t) = e^{Q't}\underset{\sim}{m} \tag{3}$$

and

$$\underset{\sim}{N}(t) = e^{Q'(t-s)}\underset{\sim}{N}(s) \tag{4}$$

for all non negative t and s.

Now e^{Qt} is non negative since $q_{ij} \geqslant 0$ for $i \neq j$ (Bellman, 1970, p.176). Also

$$\frac{d}{dt}(e^{Qt}\underset{\sim}{1}) = e^{Qt}Q\underset{\sim}{1} = \underset{\sim}{0}$$

since $\sum_j q_{ij} = 0$. Thus $e^{Qt}\underset{\sim}{1}$ is a constant, but as $e^{Q0}\underset{\sim}{1} = I\underset{\sim}{1} = \underset{\sim}{1}$, e^{Qt} is a Markov matrix. Hence if t is restricted to integer values a special case of 1d is obtained, or model 1c can be considered to interpolate model 1d when $P = e^{Q}$. Model 1d is more restrictive since it requires the animals' behaviour to be continuously homogeneous and continuously Markovian, the latter meaning future values depend only on current values. Both requirements will be violated frequently by real animals. In discrete time it may be possible to avoid inhomogeneities by a natural choice of time unit such as days or years. On the other hand the continuous time model may be appropriate when the interval of observation is too short for inhomogeneities to appear. Later, various ways of modifying continuous time models to take account of inhomogeneities will be considered.

Both of the deterministic models appear as means of stochastic models to be presented later. Expressions (1) and (3) are the means of $\underset{\sim}{N}(t)$ for models 2d and 2c and (2) and (3) the conditional means of $\underset{\sim}{N}(t)$ given $\underset{\sim}{N}(s)$.

Usher and Williamson (1970) have incorporated the discrete time model 1d in a birth death and migration model for animals living in a homogeneous bounded region. Their test of this model using some data for the flour beetle *Tribolium confusum* gave a reasonable fit when females only were considered. It is interesting to note that the flour beetles were living in a genuinely continuous space so that Usher and Williamson's model is really a discrete space approximation.

2.2 The basic stochastic models

Model 2d

Consider just a single animal, for example the ℓth with movement process $X_\ell = \{X_\ell(t), t=0,1,2,...\}$. Unless the animal moves very quickly, relative to a unit of time, the area the animal occupies next will depend on where it is now. Thus the simplest conceivable model for the movement of this animal in a discrete time and space situation is a homogeneous Markov chain. This means that for each t, $P(X(t)=i_t | X(t-1)=i_{t-1},..,X(0)=i_0) = p_{i_{t-1}i_t}$ where $P = (p_{ij})$ is a matrix of constants, or in words the conditional probability distribution for the position of animal ℓ at any time t, given its positions at $t-1, t-2,..,0$, depends only on its position at $t-1$.

Suppose that all the animals exhibit the same kind of movement behaviour, that is X_ℓ, $\ell = 1,..,n$ are homogeneous Markov chains with a common transition matrix P. However we want m_i of the animals initially in area i, $i = 1,..,k$ $(\underset{\sim}{N}(0)=\underset{\sim}{m})$ so assume $P(X_\ell(0)=i) = 1$ occurs for exactly m_i of the X_ℓ. Finally let $X_1,..,X_n$ be independent stochastic processes. In summary we have:

2d

1. The X_ℓ are homogeneous Markov chains in discrete time with common transition matrix P.
2. $P(X_\ell(0)=i) = 1$ for exactly m_i of the X_ℓ, $i = 1,..,k$.
3. $X_1,..,X_n$ are independent.

Theorem 1.

The stochastic process $\underset{\sim}{N}$ is a homogeneous Markov chain. If $(p_{ij}(t)) = P^t$ then the generating function for the transition probabilities is

$$E\left[\prod_{i=1}^{k} s_i^{N_i(t)} \mid \underset{\sim}{N}(0)=\underset{\sim}{m}\right] = \prod_{i=1}^{k}\left(\sum_{j=1}^{k} p_{ij}(t)s_j\right)^{m_i}$$

$$= M(\underset{\sim}{m},P^t,\underset{\sim}{s}) \qquad (5)$$

Proof

The X_ℓ, $\ell = 1,..,n$ are independent homogeneous Markov chains therefore the multivariate process $\underset{\sim}{X} = (X_1,..,X_n)$ is also a homogeneous Markov chain. $\underset{\sim}{N}(t)$ is a function of $\underset{\sim}{X}(t)$. Specifically, defining $\underset{\sim}{Z}_\ell(t) = (Z_{\ell 1}(t),..,Z_{\ell k}(t))'$ by $Z_{\ell j}(t) = I_{\{j\}}(X_\ell(t))$, then

$$\underset{\sim}{N}(t) = \sum_{\ell=1}^{n} \underset{\sim}{Z}_\ell(t) .$$

Now the $\underset{\sim}{Z}_\ell(t)$ are independent multiple Bernoulli random variables and if $P(X_\ell(0)=j) = 1$, $\underset{\sim}{Z}_\ell(t)$ has parameters $(p_{j1}(t),..,p_{jk}(t))$. Hence the distribution of $\underset{\sim}{N}(t)$ is a convolution of k multinomial distributions with parameters $(m_j,p_{j1}(t),..,p_{jk}(t))$, $j = 1,..,k$. Note also that the distribution of $\underset{\sim}{N}(t)$ conditional on $\underset{\sim}{X}(0)$ depends only on $\underset{\sim}{N}(0)$ $(=\underset{\sim}{m})$.

Because $\underset{\sim}{N}(t)$ is a function of $\underset{\sim}{X}(t)$
$P(\underset{\sim}{N}(t)=\underset{\sim}{n}|\underset{\sim}{N}(t-1),..,\underset{\sim}{N}(0)) = E[P(\underset{\sim}{N}(t)=\underset{\sim}{n}|\underset{\sim}{X}(t-1),..,\underset{\sim}{X}(0))|\underset{\sim}{N}(t-1),..,\underset{\sim}{N}(0)]$.
From the Markov property for $\underset{\sim}{X}$ the R.H.S. is $E[P(\underset{\sim}{N}(t)=\underset{\sim}{n}|\underset{\sim}{X}(t-1))|\underset{\sim}{N}(t-1),..,\underset{\sim}{N}(0)]$. The time homogeneity of $\underset{\sim}{X}$ gives

$$P(\underset{\sim}{N}(t)=\underset{\sim}{n}|\underset{\sim}{X}(t-1)=\underset{\sim}{x}) = P(\underset{\sim}{N}(1)=\underset{\sim}{n}|\underset{\sim}{X}(0)=\underset{\sim}{x})$$
$$= P(\underset{\sim}{N}(1)=\underset{\sim}{n}|\underset{\sim}{N}(0)=\underset{\sim}{m})$$

where $\underset{\sim}{x}$ is chosen so that $m_j = \sum_{\ell=1}^{n} I_{\{j\}}(x_\ell)$.

This is true for arbitrary $\underset{\sim}{m}$ and so it is clear that

$$P(\underset{\sim}{N}(t)=\underset{\sim}{n}|\underset{\sim}{N}(t-1),..,\underset{\sim}{N}(0)) = P(\underset{\sim}{N}(t)=\underset{\sim}{n}|\underset{\sim}{N}(t-1))$$

where the R.H.S. depends on t only through $\underset{\sim}{N}(t-1)$. Thus $\underset{\sim}{N}$ is a homogeneous Markov chain and the transition probabilities $P(\underset{\sim}{N}(t)=\underset{\sim}{n}|\underset{\sim}{N}(0)=\underset{\sim}{m})$ coincide with the distribution of $\underset{\sim}{N}(t)$ under the specification 2d2.

This theorem shows that the distribution of $\underset{\sim}{N}(t)$, conditional on $\underset{\sim}{N}(0) = \underset{\sim}{m}$, can be represented as a convolution of k multinomials. Each of the multinomials in the convolution can be identified with the animals initially occupying a particular area. In the special cases where $m_i = n$, $m_j = 0$ for $j \neq i$, or P has rank 1 ($P = \underset{\sim}{1}\underset{\sim}{p}'$ for some vector $\underset{\sim}{p}$) then this convolution of multinomials reduces to a multinomial. The second case, $P = \underset{\sim}{1}\underset{\sim}{p}'$, means that $X_\ell(t)$, $t = 0,1,\ldots$ is a sequence of independent random variables. Such a model may be appropriate if the unit of time is very long compared with the amount of movement that takes place (Wierzbowska, 1972).

The representation as a convolution of multinomials yields

$$E\,\underset{\sim}{N}(t) = (P')^t\underset{\sim}{m} \tag{6}$$

and the variance matrix of $\underset{\sim}{N}(t)$ is

$$V\,\underset{\sim}{N}(t) = \sum_{i=1}^{k} m_i\left(\mathrm{diag}(\underset{\sim}{p}_i(t))-\underset{\sim}{p}_i(t)\underset{\sim}{p}_i{}'(t)\right) \tag{7}$$

where $\underset{\sim}{p}_i(t)$ is the ith row of P^t.

Equation (6) is really a special case of

$$E[\underset{\sim}{N}(t)|\underset{\sim}{N}(s)] = (P')^{t-s}\underset{\sim}{N}(s) \tag{8}$$

for $s < t$. The mean of this process thus yields deterministic model 1d.

<u>Model 2c</u>.

The continuous time analogue to the last model, 2d, specifies each X_ℓ as a homogeneous Markov chain in continuous time. In this case

$$P(X_\ell(t_{c+1}) = i_{c+1} | X_\ell(t_c)=i_c,..,X_\ell(t_0)=i_0) = p_{i_c i_{c+1}}(t_{c+1}-t_c)$$

for every pair of finite sets $i_0,..,i_{c+1}$ and $0 \leq t_0 < t_1 < .. < t_{c+1}$. Assume $\lim_{t\to 0} P(t)\,(=(p_{ij}(t))) = I$ so that X_ℓ is continuous in probability and there is a matrix Q with non negative off diagonal elements satisfying $Q\underset{\sim}{1} = \underset{\sim}{0}$ for which $P(t) = e^{Qt}$. The elements of Q are the transition rates and Q is a rates matrix.

The complete model is:

2c

1. The X_ℓ are homogeneous Markov chains in continuous time with common transition rates Q.
2. $P(X_\ell(0)=i) = 1$ for exactly m_i of the X_ℓ, $i = 1,..,k$.
3. $X_1,..,X_n$ are independent.

Theorem 2.

The stochastic process $\underset{\sim}{N}$ is a homogeneous Markov chain with transition probability generating function

$$E\left[\prod_{i=1}^{k} s_i^{N_i(t)} \,|\, \underset{\sim}{N}(0)=\underset{\sim}{m}\right] = M(\underset{\sim}{m},P(t),\underset{\sim}{s}) \tag{9}$$

The proof of theorem 2 is almost identical to that of theorem 1 and so is omitted.

As in the discrete time model 2d the transition probabilities for $\underset{\sim}{N}$ are convolutions of multinomials and the mean, variance and conditional mean take the same form as in 2d. In particular the mean for this model yields the deterministic model 1c.

If this model is sampled in discrete time it reduces to model 2d. However not every discrete time process can be derived from a continuous time process in this way. This is because an arbitrary Markov matrix, P, is not necessarily representable in the form e^Q with Q a rates matrix for this implies that P is aperiodic (see theorem 6). We shall say P is imbeddable if it does have the above form. For $k = 2$ it is easy to see what imbeddability entails. A rates matrix has the form

$$Q = \begin{bmatrix} -\mu & \mu \\ \lambda & -\lambda \end{bmatrix}$$

Setting

$$P = (p_{ij}) = e^Q = \frac{1}{\lambda+\mu}\begin{bmatrix} \lambda+\mu e^{-\mu-\lambda} & \mu-\mu e^{-\mu-\lambda} \\ \lambda-\lambda e^{-\mu-\lambda} & \mu+\lambda e^{-\mu-\lambda} \end{bmatrix}$$

it is seen that

$$p_{11} > p_{21},\ p_{22} > p_{12}. \tag{10}$$

It is not difficult to show that this is also a sufficient condition. Condition (10) can be restated as $\det P > 0$ which Kingman (1963) has shown is necessary for imbeddability when $k > 2$. Goodman (1970) has shown that $\prod_{i=1}^{k} p_{ii} \geqslant \det P$ is also necessary for imbeddability and Johansen (1973a, b) gives additional results.

We have seen that there are quite significant constraints on an imbeddable transition matrix. This is particularly important to keep in mind when estimating Q from discrete time data. Discrete time methods naturally lead to an estimate $\hat{P}$ of P but it may not be possible to find a rates matrix $\hat{Q}$ for which $\hat{P} = e^{\hat{Q}}$. Since the imbeddability problem has not been solved in a way that is easy to apply, it may be difficult to estimate Q from discrete time data. On the other hand if all one wants is a model which gives a reasonable description of the data there seems to be no good reason not to be content with the discrete time model. It may even be more accurate.

It was suggested in section 2.1 that the discrete time models might be applied in cases where strict time homogeneity required for continuous time models is not satisfied. Periodic changes are features of the environment of many animals; the most obvious having a period of a day or a year. Such changes are often reflected in movement behaviour. It is tempting to model this situation assuming each animal moves according to a non homogeneous continuous time Markov chain having transition rates which are periodic functions of time.

Intuitively, if a non homogeneous continuous time Markov chain has periodic transition rates then sampling at appropriate discrete points in time yields a homogeneous discrete time Markov chain. A proof of this is as follows. Let $X_\ell(t)$ be a Markov chain with transition matrix

$P(s,t) = \left(p_{ij}(s,t)\right)$, $p_{ij}(s,t) = P\left(X_\ell(t)=j \mid X_\ell(s)=i\right)$.

Let P satisfy the forward equation

$$\frac{\partial}{\partial t} P(s,t) = P(s,t)Q(t) \tag{11}$$

with $Q(t) = Q(t+1)$, $t \geqslant 0$, and Q is continuous. Assume P is standard, that is $\lim_{t \downarrow s} P(s,t) = P(s,s) = I$.

For each $s \geqslant 0$ there is a unique solution to (11) subject to $P(s,s) = I$ and so using $Q(t+1) = Q(t)$ it is clear that $P(s+1,t+1) = P(s,t)$. Iterating this $P(s+r,t+r) = P(s,t)$ for positive integers r. Now by the Chapman-Kolmogorov equations $P(0,r) = P(0,1)P(1,2)..P(r-1,r) = \left(P(0,1)\right)^r$ and $P(r,s) = \left(P(0,1)\right)^{s-r}$. Thus sampling at intervals of period 1 gives a homogeneous Markov chain.

By ignoring what happens between sampling points we have simplified the situation and the price paid is loss of information about what is happening between these points. Possibly one is only interested in these discrete points, for example questions may relate to the distribution of the animals at nightfall. In other cases the primary interest may be the variation in behaviour during a period. This can be handled in a discrete time setting by dividing the basic period into a finite number of segments. One then has a non-homogeneous discrete time process

$$P_1 = P\left(0,\frac{1}{r}\right),\ P_2 = P\left(\frac{1}{r},\frac{2}{r}\right),\ldots,P_r = P\left(\frac{r-1}{r},1\right), \text{ and}$$

$P_{r+i} = P_i$. This model is still fairly easy to work with from the point of view of computation and estimation of parameters.

A non-homogeneous model of particular interest is $Q(t) = Qa(t)$ because $a(t)$ can be given direct biological interpretation. Q is a constant rates matrix and $a(t)$ is a non-negative function. We need not assume $Q(t)$ is periodic. $P(s,t)$ takes the particularly simple form

$$P(s,t) = e^{Q\int_s^t a(u)\,du} . \tag{12}$$

From the equation $\frac{\partial}{\partial t} P(s,t) = Qa(t)$ it is apparent that $a(t)$ is a measure of the rate at which the process is going which presumably depends on the "activity" of the animals. Thus we might call $a(t)$ the level of activity at time t. In doing so we are really defining activity for animals that satisfy this model. This definition agrees with the definition of activity, as the mean rate of crossing particular points, used in the experimental studies by Marten

(1973) and Mossing (1975). To see this agreement consider a homogeneous Markov chain $X = \{X(t), t \geq 0\}$ with rates matrix Q and a stationary initial distribution so that X is also a stationary process. Define $N(s,t]$ as the number of times X changes state in $(s,t]$. It is not difficult to see that $\mathcal{E}N(s,t] < \infty$ and, from the stationarity of X, the Lebesgue-Stieltjes measure induced by the interval function $\mathcal{E}N(s,t]$ is invariant under translation and so is proportional to Lebesgue measure. Thus

$$\mathcal{E}N(s,t] = (t-s)\mathcal{E}N(0,1].$$

To prove the result we want we must let X_ℓ have a stationary initial distribution, the same as X, rather than a degenerate distribution as assumed in 2c. A stationary initial distribution will be any vector $\underset{\sim}{\gamma}$ of probabilities satisfying $\underset{\sim}{\gamma}'Q = \underset{\sim}{0}$. X_ℓ has the same distribution as $\{X(\int_0^t a(u)du), t \geq 0\}$ so that

$$\mathcal{E}N_\ell(s,t] = \mathcal{E}N(\int_0^s a(u)du, \int_0^t a(u)du] = \int_s^t a(u)du\,\mathcal{E}N(0,1].$$

Hence $a(t)$ is proportional to the expected rate of changing area at time t.

The model $Q(t) = Qa(t)$ amounts to a time transformation of a homogeneous process since if X_ℓ is a non homogeneous Markov chain with transition rates $Q(t)$ then X_ℓ has the same distribution as $\{X(\int_0^t a(u)du), t \geq 0\}$ where X is a homogeneous Markov chain with transition rates Q, provided the initial distributions are equal. Time transformations of this kind have been very useful in the study of the dynamics of insect populations. The function $a(t)$ is often a function of temperature and $A(t) = \int_0^t a(u)du$ is referred to as "physiological time". Some very well fitting models have been produced. Examples are Wilson *et al* (1972), Gutierrez *et al* (1974) and Hughes and Gilbert (1968).

In the movement context it is evident that many animals show variation in activity with time. Daily rhythms of activity have been widely studied. I cite as an example Hindley's (1975) study of the activity of Banana Prawns. Sometimes it has been possible to relate activity to changing environmental variables; for example Dobzhansky and Wright (1943) found a relationship between activity and temperature for *Drosophila pseudoobscura*. They used two different definitions of activity namely the mean number of flies per trap and the mean distance travelled by the flies. Although different from the definition used here they are closely related to it.

The distribution of $\underset{\sim}{N}$ derived under model 2c has appeared many times before. However the usual derivation ignores individual animals (or particles as they are usually called) and proceeds directly via the differential equations

$$P\left(\underset{\sim}{N}(t+h)=\underset{\sim}{N}(t)+\underset{\sim}{e}_j-\underset{\sim}{e}_i \,\middle|\, \underset{\sim}{N}(t)\right) = N_i(t)q_{ij}h + o(h), \quad i \neq j \qquad (13)$$

and

$$P\left(\underset{\sim}{N}(t+h)=\underset{\sim}{N}(t) \,\middle|\, \underset{\sim}{N}(t)\right) = 1 + \Sigma\, N_i(t)q_{ii}h + o(h)$$

where Q is a rates matrix. $\underset{\sim}{N}$ is assumed to be a Markov process.

The first discussion of (13) of which I am aware is Bartlett (1949). These equations are generally referred to as the "stochastic compartment model". In the animal movement context Arnason (1970) studied (13) for the two area case and generalised this to the non homogeneous situation. He also gives some results for a generalisation in different direction where (13) is replaced by

$$P\left(\underset{\sim}{N}(t+h)=\underset{\sim}{N}(t)+\underset{\sim}{e}_j-\underset{\sim}{e}_i \,\middle|\, \underset{\sim}{N}(t)\right) = f\left(N_i(t)\right)q_{ij}h + o(h),\ i\neq j \quad (14)$$

and

$$P\left(\underset{\sim}{N}(t+h)=\underset{\sim}{N}(t) \,\middle|\, \underset{\sim}{N}(t)\right) = 1 - \Sigma\, f\left(N_i(t)\right)q_{ii}h + o(h)$$

The only additional tractable case proved to be $f(x)=x^2$.

Another generalisation of Arnason's involved adding terms to allow for birth, death and immigration. In a similar manner Radcliff and Staff (1970) considered the addition of death and immigration terms in (13); their discussion deals with an arbitrary finite number of areas.

Bryant (1969) postulated equation (13) for the plant to plant movement of aphids and was reasonably successful in fitting this model to some data of Shiyomi (1967).

The equations (13) do not contain an explicit independence hypothesis. Since independence plays an important role in the construction of models in this thesis it is of interest to see what kinds of dependence are compatible with the derived distribution for $\underset{\sim}{N}$. A related question is to what extent does the distribution of $\underset{\sim}{N}$ determine the distribution of $\underset{\sim}{X}$? The approach here is to construct models starting from individual animals and so the distribution of $\underset{\sim}{X}$ is determined by the models. Whilst this approach is likely to fail in many instances it enables generalisations of the basic models which are not available using an approach such as starting with equations (13).

To begin to answer these questions note that some forms of high dependence between the X_ℓ are compatible with the derived distribution of $\underset{\sim}{N}$. To obtain an example assume model 2c for $Y = (Y_1,..,Y_n)$, that is, Y_ℓ replaces X_ℓ

in the statement of 2c. Now define $\big(X_1(t),..,X_n(t)\big)$ as the order statistics of $\big(Y_1(t),..,Y_n(t)\big)$ so that $X_1(t) \leqslant X_2(t) \leqslant .. \leqslant X_n(t)$ and

$$N_j(t) = \sum_{\ell=1}^{n} I_{\{j\}}\big(X_\ell(t)\big) = \sum_{\ell=1}^{n} I_{\{j\}}\big(Y_\ell(t)\big).$$

Thus the distribution of $\underset{\sim}{N}$ is unchanged but $X_1,...,X_n$ are certainly not independent. This example is even biologically conceivable - the weaker individuals are pushed into the poorer areas as in Newman's 1956 study of the movements of Cutthroat Trout. However in real life it seems unlikely that, although $X_1(t) \leqslant X_2(t) \leqslant .. \leqslant X_n(t)$, the distribution of $\underset{\sim}{X}(t)$ will be an order statistics distribution for independent processes.

When ordering of the variables is excluded a different result is obtained.

Lemma 1

If the random vectors $\underset{\sim}{W}_i$, $i = 1,..,n$ have a jointly symmetric distribution and take values in the set $\{\underset{\sim}{e}_i \mid i=1,..,k\}$ then $\underset{\sim}{W} = \sum_{i=1}^{n} \underset{\sim}{W}_i$ has a multinomial distribution if and only if the $\underset{\sim}{W}_i$ are independent and are marginally distributed multiple Bernoulli.

Proof

Let $p(\underset{\sim}{w}_1,..,\underset{\sim}{w}_n) = P(\underset{\sim}{W}_i=\underset{\sim}{w}_i, i=1,..,n)$.

By "the $\underset{\sim}{W}_i$, $i = 1,..,n$ have a symmetric distribution" it is meant that $p(\underset{\sim}{w}_{\pi_1},..,\underset{\sim}{w}_{\pi_n}) = p(\underset{\sim}{w}_1,..,\underset{\sim}{w}_n)$ for every permutation $(\pi_1,..,\pi_n)$ of $(1,..,n)$, for all $\{\underset{\sim}{w}_i, i=1,..,n\}$. This is equivalent to saying that $\underset{\sim}{W}_1,..,\underset{\sim}{W}_n$ are "exchangeable".

The symmetry of p means that $p(\underset{\sim}{w}_1,..,\underset{\sim}{w}_n)$ depends only on $\underset{\sim}{n} = (n_1,..,n_k)$ where n_i = the number of $\underset{\sim}{w}_j$ equal to $\underset{\sim}{e}_i$. But $\underset{\sim}{n} = \Sigma\, \underset{\sim}{w}_i$. Thus $p(\underset{\sim}{w}_1,..,\underset{\sim}{w}_n) = p'(\Sigma \underset{\sim}{w}_i)$ for some function p'. Now $(\Sigma n_i)!/(n_1!..n_k!)$ is the number of distinct $(\underset{\sim}{w}_1,..,\underset{\sim}{w}_n)$ with common sum $\underset{\sim}{n}$ and so $P(\underset{\sim}{W}=\underset{\sim}{n}) = p'(\underset{\sim}{n})\ (\Sigma n_i)!/(n_1!..n_k!)$. However if $\underset{\sim}{W}$ is multinomial it is clear that $p'(n)$ takes the form $p_1^{n_1}..p_k^{n_k}$ for some vector of probabilities $(p_1,..,p_k)$ summing to 1. Since $p(\underset{\sim}{w}_1,..,\underset{\sim}{}_n) = p'(n_1,..,n_k)$ this means $\underset{\sim}{W}_1,..,\underset{\sim}{W}_n$ are independent and identically distributed as multiple Bernoulli.

To apply this to the movement situation assume $m_i = n$ for some i and $\left(X_1(t),..,X_n(t)\right)$ have a symmetric distribution. The lemma then says that $X_1(t),..,X_n(t)$ are independent if $\underset{\sim}{N}$ has the distribution resulting from model 2c, or equivalently equations (13), further, the joint distribution of $X_1(t),..,X_n(t)$ is determined by the distribution of $\underset{\sim}{N}$. This result is a long way from independence of the stochastic processes $X_1,..,X_n$. A stronger result is available when symmetry is imposed at the level of the transition probabilities.

Theorem 3

Let $\underset{\sim}{X}$ be a homogeneous Markov chain which is continuous in probability and satisfies

1. $P\left(X_\ell(t+h)\neq X_\ell(t) \text{ for more than one } \ell\right) = o(h)$
2. The transition probabilities have the property

$$\lim_{h\to 0}\frac{1}{h}P\left(X_\ell(t+h)=j,\ X_\gamma(t+h)=X_\gamma(t) \text{ when } \gamma\neq\ell \,\middle|\, X_\ell(t)=i,\ \underset{\sim}{X}(t)=\underset{\sim}{x}\right)$$
$$= \lim_{h\to 0}\frac{1}{h}P\left(X_\ell(t+h)=j,\ X_\gamma(t+h)=X_\gamma(t) \text{ when } \gamma\neq\ell \,\middle|\, X_\ell(t)=i,\ \underset{\sim}{N}(t)=\underset{\sim}{n}\right) \qquad (14)$$

for $i \neq j$, whenever $X_\ell(t)=i,\ \underset{\sim}{X}(t)=\underset{\sim}{x},\ \underset{\sim}{N}(t)=\underset{\sim}{n}$ are simultaneously possible, with the limit not depending on ℓ.

3. $\underset{\sim}{N}$ is a homogeneous Markov chain with transition probabilities given by (9).

Then conditional on $\underset{\sim}{X}(0)$, the stochastic processes $X_1,..,X_n$ are independent.

Since the theorem assumes $\underset{\sim}{X}$ is a homogeneous Markov chain on a finite state space, and is continuous in probability, there is a regular version of the process in which the sample paths are step functions. For this regular version no jumps can occur for which the transition rate is 0. Non regular cases are unreasonable for animal movements and so condition 1. ensures $\underset{\sim}{X}$ only changes value at times when $\underset{\sim}{N}$ changes value. On the other hand regular versions of $\underset{\sim}{N}$ only have jumps of the kind $\underset{\sim}{N}(t) \rightarrow \underset{\sim}{N}(t) + \underset{\sim}{e}_j - \underset{\sim}{e}_i$ and so 1. is a necessary and sufficient condition for $\underset{\sim}{X}$ to only change value when $\underset{\sim}{N}$ does.

Condition 2. is a symmetry condition. Basically each animal's movements are symmetrically dependent on the others and so the transition rate on the L.H.S. of (14) depends only on the unordered set of values taken by the $X_\gamma(t)$, $\gamma \neq \ell$. Thus it depends only on $X_\gamma(t)$, $\gamma \neq \ell$, through $\underset{\sim}{N}(t)$. Biologically, an animal does not distinguish one individual from another.

In summary theorem 3 says that if a homogeneous Markov model of the movements of n animals, for which $\underset{\sim}{N}$ signals changes in $\underset{\sim}{X}$, has the transition rates for each animal's movements symmetrically dependent on the other animals, then $\underset{\sim}{N}$ is a homogeneous Markov process with transition probabilities given by (9) only if $X_1,..,X_n$ are independent conditional on $\underset{\sim}{X}(0)$.

<u>Proof of theorem 3.</u>

Fix $i, j, i \neq j$

Let $A_{th} = \{X_\ell(t+h) \neq X_\ell(t)$ for more than one $\ell\}$

$B_{th} = \{X_\ell(t+h)=j,\ X_\ell(t)=i$ for exactly one $\ell\} \cap A^c_{th}$

$$B_{th\ell} = \{X_\ell(t+h)=j, X_\ell(t)=i\} \cap B_{th}$$

$$C_{th} = \{\underset{\sim}{N}(t+h)=\underset{\sim}{N}(t)+\underset{\sim}{e}_j-\underset{\sim}{e}_i\}$$

Now $C_{th} \supset B_{th}$ and $C_{th} - B_{th} \subset A_{th}$ so that

$$P(C_{th}\Delta B_{th}) = o(h) \text{ by } 1.$$

Thus
$$\begin{aligned} P(C_{th}|\underset{\sim}{X}(t)=\underset{\sim}{x}) &= P(B_{th}|\underset{\sim}{X}(t)=\underset{\sim}{x}) + o(h) \\ &= P\Big(\bigcup_{x_\ell=i} B_{th\ell}|\underset{\sim}{X}(t)=\underset{\sim}{x}\Big) + o(h) \\ &= \sum_{x_\ell=i} P(B_{th\ell}|\underset{\sim}{X}(t)=\underset{\sim}{x}) + o(h) \\ &= n_i\, P(B_{th\ell}|X_\ell(t)=i,\underset{\sim}{N}(t)=\underset{\sim}{n}) + o(h). \end{aligned}$$

The last equality is condition 2.

We have

$$\begin{aligned} \lim_{h\to 0} \frac{1}{h}P(C_{th}|\underset{\sim}{X}(t)=\underset{\sim}{x}) &= n_i \lim_{h\to 0} \frac{1}{h} P(B_{th\ell}|X_\ell(t)=i,\underset{\sim}{N}(t)=\underset{\sim}{n}) \\ &= n_i\, q_{ij}(\underset{\sim}{n}) \quad \text{say.} \end{aligned}$$

Since there are only a finite number of possibilities for $\underset{\sim}{x}$ this means $\lim_{h\to 0} \frac{1}{h} P(C_{th}|\underset{\sim}{N}(t)=\underset{\sim}{n}) = n_i\, q_{ij}(\underset{\sim}{n})$.

Using 3. and equations (13) which follow from 3. it is clear that, $q_{ij}(\underset{\sim}{n}) = q_{ij}$. Thus the transition rate

$$\lim_{h\to 0} \frac{1}{h} P(B_{th\ell}|\underset{\sim}{X}(t)=\underset{\sim}{x}) = q_{ij} \quad \text{for all } i,\ j,\ i \neq j.$$

All other transition rates, apart from the no change rates, are 0. Thus under the conditions of the theorem $\underset{\sim}{X}$ has the same transition rates as are determined by model 2c. and so $X_1,..,X_n$ are independent conditional on $\underset{\sim}{X}(0)$.

The proof of theorem 3 does not make use of homogeneity of $\underset{\sim}{X}$ and it holds with only minor modifications for non-homogeneous $\underset{\sim}{X}$ as well. Note also that under the conditions of the theorem the distribution of $\underset{\sim}{N}$ determines the distribution of $\underset{\sim}{X}$ given $\underset{\sim}{X}(0)$.

2.3 Differences between individuals.

The basic stochastic models were constructed on the basis of the way individual animals might behave. One could obtain an individual animal, observe its movements, and estimate the matrix P or the matrix Q. Suppose this process is repeated for another animal. Is it likely that no differences will be detected in the transition probabilities for the two different animals?

It is quite evident when observing many animals that there are fast individuals and slow individuals and sometimes differences can be more drastic such as differences in preferences for different areas. However few movement studies have focused on differences between individuals. Perhaps this is so because differences between individuals are regarded as self-evident and workers have not thought of trying to measure them. In any case the design of movement experiments is frequently incapable of separating the stochastic element of each individual's movement process from variation between individuals.

There seem to be three main ways of detecting variability between individuals. First one might attempt to set up independent trials with the same animals each time so that replicates of each individual's movement process are obtained. This then gives an estimate of the magnitude of the stochastic element of an individual's movement process which can be separated from variation resulting from differences between individuals. A second method is essentially a regression method, specifically show that an individual's movements are related to some known characteristic of the animal such as age or size. Finally one can propose a model for an individual's movement process with a view to seeing if the observed movements are consistent with all individuals having identical values for the parameters of their movement processes.

I am not aware of the first method ever having been used. Perhaps it is not used because it is difficult to make an animal perform independently and identically many times. The second method has been used by Lefkovitch (1967) and Saila (1961). Lefkovitch was able to show that movements of his beetle larvae were related to age but Saila was not able to detect any size dependence in the movements of his adult flounder. The third method appears in the literature on dispersal in the vinegar fly (*Drosophila spp*), for example Dobzhansky and Wright (1943), and also in Jones' (1959) study of Haddock movements. In both cases the observed movement distributions have longer tails than predicted from a theoretical diffusion model. However their experiments were not capable of validating the diffusion model for an individual animal largely because an individual's movement process was only observed at a few points in time. Also the doubt remained that a secondary process produced occasional larger movements for every animal independently, and this appeared like variability between individuals because only short records of movements were available. Wright (1968) seems to accept this latter hypothesis.

When modelling animal movements the magnitude of any differences present of course determines whether they should be taken account of in the model. As noted earlier Bryant (1969) obtained reasonable success in fitting the model 2c to some aphid data which means that differences present did not upset the model. Nevertheless it seems desirable to find ways of introducing differences between individuals' movement processes in movement models. One might think of these differences as phenotypically dependent or in some cases dependent on the past history of the animal.

Consider the discrete time model 2d. A simple way of introducing variability between individuals would be to simply assign each animal a unique, rather than a common, transition matrix, so that animal ℓ has transition matrix P_ℓ.

If one were only ever interested in a particular n individuals, which are distinguishable from each other, then it would be appropriate to regard $P_1,..,P_n$ as non random constants. This is rarely the case since animals do not last very long. They will either die or change. Also, in most experimental studies, the n animals at hand will be only a subset of a larger population of interest. In the latter case one might use a finite sampling approach so that the matrices $P_1,..,P_n$ are regarded as a simple random sample from a population of size p, say, provided the n animals are a simple random sample of the total population. If n is very small compared with p then $P_1,..,P_n$ can be considered approximately independent and identically distributed (i.i.d.). I shall argue that in any case the model that $P_1,..,P_n$ are i.i.d. is an appropriate one.

Populations of animals are transient objects so that information gained from a finite sampling approach may be only applicable for a short time because it is not transferable from one population to another.

The problem is that the focus is on the individuals rather than the circumstances leading to the generation of the present population. A common model in science is to assume that a given set of circumstances leads to an infinite population from which $P_1,..,P_p$ arise as a random sample. The random variables $P_1,..,P_n$ are a random sample from $\{P_1,..,P_p\}$ and are i.i.d. If this model is accepted then a given infinite population can be estimated arbitrarily closely if p can be made arbitrarily large. I assume here that the matrix P_ℓ, for animal ℓ, can be determined for each ℓ. The information gained from a population generated under one set of circumstances is transferable to any population generated under the same set of circumstances.

Henceforth I assume $P_1,..,P_n$ to be i.i.d.

<u>Model 3d.</u>

To cater for differences between animals suppose that for the ℓth animal there is a Markov matrix P_ℓ such that given P_ℓ, X_ℓ is a homogeneous Markov chain in discrete time. Assume that the artificial numbering of the animals is at random with respect to the values of the P_ℓ, and that the P_ℓ, $\ell = 1,..,n$ are then i.i.d. For the sake of simplicity, for the present assume $\underset{\sim}{X}(0)$ and $\{P_\ell, \ell=1,..,n\}$ are independent. Later a more realistic assumption will be considered. We have the model:

3d.

1. Conditional on $P_1,..,P_n$ the stochastic processes $X_1,..,X_n$ are independent homogeneous Markov chains in discrete time with X_ℓ having transition matrix P_ℓ.

2. m_i of the X_ℓ satisfy $P\big(X_\ell(0)=i\big) = 1$.

3. $P_1,..,P_n$ are independent and identically distributed and are jointly independent of $\underset{\sim}{X}(0)$.

It will be recognised that the specification "$\{P_\ell, \ell=1,..,n\}$ is independent of $\underset{\sim}{X}(0)$" is redundant in the presence of 2. since events in the σ-algebra generated by $\underset{\sim}{X}(0)$ have probability 0 or 1. However it serves to emphasise the idea behind the model namely that the initial positions of the animals are independent of their transition matrices from which 3d is obtained conditioning on $\underset{\sim}{X}(0)$.

For this model the generating function for the distribution of $\underset{\sim}{N}(t)$ conditional on $P_1,..,P_n$ is

$$\prod_{i=1}^{k} \; \prod_{X_\ell(0)=i} \; (\underset{\sim}{p}_{i\ell}{}'s) \tag{15}$$

where $\underset{\sim}{p}_{i\ell}{}'$ is the ith row of P_ℓ^t. The generating function for $\underset{\sim}{N}(t)$ is the expected value of (15) and using the independence and identity of distribution of the P_ℓ this is

$$\prod_{i=1}^{k} \; \prod_{X_\ell(0)=i} \; E\underset{\sim}{p}_{i\ell}{}'\underset{\sim}{s} = \prod_{i=1}^{k} \; (E\underset{\sim}{p}_{i\ell}{}'\underset{\sim}{s})^{m_i}$$

or

$$M(\underset{\sim}{m}, EP_1^t, \underset{\sim}{s}) \tag{16}$$

Thus the distribution of $\underset{\sim}{N}(t)$ is a convolution of multinomials with a form similar to that in model 2d, the difference being that P^t is replaced by EP_1^t. Another way of deriving this result is to show that $X_1,..,X_n$ are independent under model 3d which makes it immediately clear that $\underset{\sim}{N}(t)$ must be a convolution of multinomials.

Note that the distribution of $\underset{\sim}{N}(t)$, which can be thought of as the $\underset{\sim}{X}(0)$ conditional distribution, depends on $\underset{\sim}{X}(0)$ only through $\underset{\sim}{N}(0)$ and so it is the $\underset{\sim}{N}(0)$ conditional distribution of $\underset{\sim}{N}(t)$ as well. This will be the case in all of the models in this chapter.

It is important to realise that under model 3d the processes $X_1,\ldots,X_n$ and $\underset{\sim}{N}$ are not necessarily Markov. The following theorem shows part of the picture for the X_ℓ. The notation a.s. (almost surely) is used to denote events which have probability 1.

Theorem 4.

Suppose that, conditional on P_ℓ, X_ℓ is almost surely aperiodic and has all states communicating. Then X_ℓ is a Markov process if and only if P_ℓ is almost surely constant.

The proof of this theorem depends on a well known result which is proved, for example, in Freedman (1971, p74). I state it as Theorem 5.

Theorem 5.

Let $X = \{X(t), t=0,1,\ldots\}$ be a homogeneous Markov chain on the state space $\{1,2,\ldots,k\}$ with transition matrix P and invariant probabilities $\underset{\sim}{\pi}' = \underset{\sim}{\pi}'P = (\pi_1,\ldots,\pi_k)$. Let all states communicate. If f and g are real valued functions for which $\Sigma\pi_j g(j) \neq 0$ then

$$\sum_{s=0}^{t} f\big(X(s)\big) \Big/ \sum_{s=0}^{t} g\big(X(s)\big) \xrightarrow{a.s.} \Sigma\, \pi_j f(j)/\Sigma\, \pi_j g(j) \text{ as } t \to \infty$$

Proof of Theorem 4.

Let F_t be the σ-algebra generated by $\{X_\ell(s), s \leqslant t\}$ and F the σ-algebra generated by $\bigcup_{t\geqslant 0} F_t$. Define $\underset{\sim}{Z}(t) = \big(X_\ell(t-1), X_\ell(t)\big)$. Note that, conditional on P_ℓ, $\underset{\sim}{Z}$ is a homogeneous Markov chain with transition probabilities $P\big(\underset{\sim}{Z}(t)=(j',j) \mid \underset{\sim}{Z}(t-1)=(i',i), P_\ell\big) = p_{ij}\delta_{ij'}$, where $P_\ell = (p_{ij})$. If $\{\pi_i, i=1,\ldots,k\}$ are invariant probabilities for X_ℓ given P_ℓ then $\{\pi_i p_{ij}, i,j=1,..,k\}$ are invariant probabilities for $\underset{\sim}{Z}$ given P_ℓ.

Now define $n_{ij}(t) = \sum_{s=1}^{t} I_{\{(i,j)\}}(\underset{\sim}{Z}(s))$, $n_i(t) = \sum_j n_{ij}(t)$, $\hat{p}_{ij} = \limsup_{t\to\infty} n_{ij}(t)/n_i(t)$. Conditional on P_ℓ, $\underset{\sim}{Z}$ satisfies theorem 5 with $f = I_{\{(i,j)\}}$, $g = \sum_j I_{\{(i,j)\}}$ and so $\hat{p}_{ij} \overset{a.s.}{=} p_{ij}$ in the P_ℓ conditional distribution. Unconditionally $\hat{p}_{ij} \overset{a.s.}{=} p_{ij}$ remains true. Note that $\hat{p}_{ij}$ is $\mathcal{F}$ measurable.

Now

$$\mathcal{P}(X_\ell(t+1)=j|\mathcal{F}_t) = \mathcal{E}[\mathcal{P}(X_\ell(t+1)=j|\mathcal{F}_t,P_\ell)|\mathcal{F}_t]$$
$$= \mathcal{E}[p_{ij}|\mathcal{F}_t]$$

on the set $\{X_\ell(t)=i\} = A(t,i)$, say. On the other hand setting

$$p_i(t) = \mathcal{P}(X_\ell(t)=i|P_\ell)$$

we have

$$\mathcal{P}(X_\ell(t+1)=j|X_\ell(t)=i) = (\mathcal{E}p_{ij}p_i(t))/\mathcal{E}p_i(t)$$

and for X_ℓ to be ~~Makov~~ MARKOV we must have

$$I_{A(t,i)}\mathcal{E}[p_{ij}|\mathcal{F}_t] \overset{a.s.}{=} I_{A(t,i)}(\mathcal{E}p_{ij}p_i(t))/\mathcal{E}p_i(t) \quad (17)$$

Since $\mathcal{F}_t$ is an increasing sequence and $\hat{p}_{ij}$ is $\mathcal{F}$ measurable then from a standard result (see Breiman 1969, p.93)

$$\mathcal{E}[\hat{p}_{ij}|\mathcal{F}_t] \xrightarrow{a.s.} \hat{p}_{ij}$$

which means

$$\mathcal{E}[p_{ij}|\mathcal{F}_t] \xrightarrow{a.s.} p_{ij}. \quad (18)$$

Using the aperiodicity of X_ℓ in the P_ℓ conditional distribution $p_i(t) \to \pi_i > 0$ a.s. Now $\pi_1,..,\pi_n$ is the P_ℓ conditional limiting distribution for $X_\ell(t)$. It follows from the dominated convergence theorem that $\mathcal{E}[p_{ij}p_i(t)] \to \mathcal{E}[p_{ij}\pi_i]$, $\mathcal{E}p_i(t) \to \mathcal{E}\pi_i$. Define

$C_{ij} = E[p_{ij}\pi_i]/E\pi_i$. Then if (17) holds

$$I_{A(t,i)}\left(p_{ij}-E[p_{ij}p_i(t)]/Ep_i(t)\right)\overset{a.s.}{\to} 0,$$

$$I_{A(t,i)}\left(p_{ij}-C_{ij}\right)\overset{a.s.}{\to} 0.$$

With probability one $\limsup_{t\to\infty} I_{A(t,i)} = 1$ since this is true in almost all P_ℓ conditional distributions of X_ℓ. Thus $p_{ij} \overset{a.s.}{=} C_{ij}$ and so X_ℓ is a Markov process only if P_ℓ is constant a.s.

This result shows also that $\underset{\sim}{N}$ is not a Markov process for the somewhat degenerate case $n = 1$. For the case $n > 1$ it will be seen that $\underset{\sim}{N}$ is not a Markov process for a special model for the distribution of P_ℓ. However for the purpose here it is probably better simply to realise that $\underset{\sim}{N}$ is not necessarily a Markov process rather than attempt to isolate the unusual situations when it is Markov.

We are left with the difficult problem of finding an appropriate distribution for the P_ℓ.

An obvious cause of variation between different animals' movement processes is variation in activity between individuals. Consider the situation where animal ℓ moves in continuous time according to a non homogeneous Markov chain with $Q(t) = Q_\ell a_\ell(t)$ and a_ℓ periodic. Observation in discrete time yields a non homogeneous Markov chain with transition matrix

$$P_\ell = e^{Q_\ell A_\ell}, \quad A_\ell = \int_0^1 a_\ell(u)\,du.$$

Thus if variation in activity is the cause of variation in the P_ℓ a suitable model is A_ℓ, $\ell = 1,..,n$ are i.i.d. random variables and the Q_ℓ are equal to the same non random rates matrix Q. We have

$$P_\ell = e^{QA_\ell} \tag{19}$$

Without reference to any continuous time model one can derive this representation of P_ℓ from the following heuristics:

The transition ~~matrices~~ MATRIX for an animal is a continuous function of some quantity called its activity, $P_\ell = P(A_\ell)$. The activity of animal ℓ, A_ℓ, is a continuous variable. An animal which is n times as active as another moves through the different areas at n times the rate of the other animal. Thus $P(nA) = P(A)^n$ and $P(A) = P(\frac{1}{n}A)^n$. If r and s are positive integers

$$P\left(\frac{r+s}{n}\right) = P\left(\frac{1}{n}\right)^{r+s} = P\left(\frac{1}{n}\right)^{r} P\left(\frac{1}{n}\right)^{s} = P\left(\frac{r}{n}\right)P\left(\frac{s}{n}\right)$$

so that for all rational p and q $P(p+q) = P(p)P(q)$. The postulated continuity of P yields $P(u+v) = P(u)P(v)$ for all u and v. This is Polya's functional equation and from Bellman (1970, p.177) the solution is $P(A) = e^{QA}$ for some MATRIX Q. Thus this intuitive interpretation of differences in activity leads to the model (19).

When (19) holds some very pleasant formulae are available for $\mathcal{E}P_1^t$.

Suppose Q is semi-simple that is $Q = G\,\Lambda\,G^{-1}$ where Λ is a diagonal matrix $\left(\mathrm{diag}(\lambda_1,..,\lambda_k)\right)$, note that $\lambda_1 = 0$. We have

$$e^{Qt} = G\, e^{\Lambda t}\, G^{-1}$$

and so

$$\mathcal{E}P_1^t = G\, \mathcal{E}[e^{\Lambda A_1 t}]G^{-1} \tag{20}$$

$\mathcal{E}e^{\Lambda A_1 t}$ is simply $\mathrm{diag}\left(1, \mathcal{E}e^{\lambda_2 t A_1}, .., \mathcal{E}e^{\lambda_k t A_1}\right)$ $= \mathrm{diag}\left(1, \varphi(\lambda_2 t), .., \varphi(\lambda_k t)\right)$. The function φ is the moment generating function of A_1 which needs to be evaluated at complex arguments with negative real part, because, as will be seen later, the non zero λ_i have negative real part. Since A_1 is non negative φ is always defined for such values of its argument. More complicated formulae are available for non semi-simple matrices via the Jordan canonical form.

If $\varphi(z)$ is defined for all complex z then we obtain the formula

$$E\,P^t = \sum_{n=0}^{\infty} \frac{Q^n t^n E A_1^n}{n!} . \tag{21}$$

To see this we follow Bellman (1970, p.166) and define the matrix norm $\|\cdot\|$ by $\|Q\| = \sum_{ij} |q_{ij}|$. This norm has the readily verifiable properties $\|QR\| \leqslant \|Q\|\,\|R\|$, $\|Q+R\| \leqslant \|Q\| + \|R\|$, $|q_{ij}| \leqslant \|Q\|$ and $\|\alpha Q\| = |\alpha|\,\|Q\|$. It follows that each element of the matrix $\sum_0^m \frac{Q^n t^n A_1^n}{n!}$ is dominated by

$$\sum_{n=0}^{\infty} \frac{\|Q\|^n t^n A_1^n}{n!} = e^{\|Q\| t A_1} .$$

The mean of this is finite and so from the dominated convergence theorem

$$E\left[\sum_{n=0}^{\infty} \frac{Q^n t^n A_1^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{Q^n t^n E A_1^n}{n!}$$

which is the stated result.

The distribution of $\underset{\sim}{N}(t)$ as $t \to \infty$.

Until now I have not made any statement about the behaviour of the distribution of $\underset{\sim}{N}(t)$ as $t \to \infty$. In all the stochastic models so far considered the generating function for $\underset{\sim}{N}(t)$ takes the form

$$M\left(\underset{\sim}{m}, \overset{R(t)}{\cancel{R(r)}}, \underset{\sim}{s}\right) .$$

The matrix $R(t) = (r_{ij}(t))$ is a $k \times k$ Markov matrix. Clearly if $\lim_{t\to\infty} R(t)$ exists, it is a Markov matrix and

$$M(\underset{\sim}{m},R(t),\underset{\sim}{s}) \to M(\underset{\sim}{m},\lim_{t\to\infty}R(t),\underset{\sim}{s})$$

so that the distribution of $\underset{\sim}{N}(t)$ converges and is a convolution of multinomials. By "convergence" I mean all the probabilities converge and the resulting distribution is proper. It is worth noting that in all cases $r_{ij}(t) = P(X_\ell(t)=j)$, for those ℓ for which $P(X_\ell(0)=i)=1$, so that convergence of $R(t)$ amounts to convergence in distribution of all of the $X_\ell(t)$ as $t \to \infty$.

For the model 2d, $R(t) = P^t$. The convergence of this matrix is very well known and there is no need to discuss it here. The model 2c has

$$R(t) = P(t) = e^{Qt}$$

for a rates matrix Q. The limiting behaviour of $P(t)$ is well known but it is difficult to find discussions of this case which show how the limit is obtained from the Jordan canonical form for Q, for an arbitrary $k \times k$ rates matrix Q. Most discussions are either more general or less general. The more general discussions relate directly to the limiting distribution of a continuous time homogeneous Markov chain where the representation $P(t) = e^{Qt}$ is not necessarily available (X_ℓ is not continuous in probability) or the chain has a possibly countable state space. They do not indicate how to find the limit but just its existence and some of its properties. Examples are Chung (1967, II.10), Doob (1953, Chapter 6). Many other treatments assume that Q, although simply a $k \times k$ rates matrix, takes some special form. Ledermann (1950) gives the only full account of which I am aware.

I shall derive the limit of $P(t)$ by a method similar to Lederman's, but which is simpler in the proof that the geometric and algebraic multiplicities of 0, as an eigenvalue of Q, are the same. This result has particular application to model 3d with P_ℓ taking the form

$$P_\ell = e^{QA_\ell} \tag{19}$$

In this case $R(t) = EP_1(t) = Ee^{QA_1t}$. Each element of e^{QA_1t} belongs to $[0,1]$. If $e^{Qt} \to J(Q)$, for some matrix $J(Q)$, and $P(A_\ell>0) = 1$, then the dominated convergence theorem gives

$$\lim_{t\to\infty} E\, e^{QA_1t} = E \lim_{t\to\infty} e^{QA_1t} = EJ(Q) = J(Q).$$

<u>Theorem 6</u>.

Let Q be a rates matrix and $P(t) = e^{Qt}$. If q is the rank of Q then Q has the Jordan canonical form

$$Q = G \left[\begin{array}{c:c} O_q & O \\ \hdashline O & B \end{array}\right] G^{-1}$$

and

$$\lim_{t\to\infty} P(t) = G \left[\begin{array}{c:c} I_q & O \\ \hdashline O & O \end{array}\right] G^{-1} = J(Q).$$

(O_q and I_q denote the $q \times q$ zero and identity matrices respectively.)

<u>Proof</u>.

Let $\underset{\sim}{x} \neq \underset{\sim}{0}$ and $(Q-\lambda I)\underset{\sim}{x} = \underset{\sim}{0}$ then

$$(q_{ii}-\lambda)x_i = -\sum_{j\neq i} q_{ij}x_j$$

and

$$|q_{ii}-\lambda||x_i| \leqslant \sum_{j\neq i} |q_{ij}||x_j|.$$

If $|x_i| = \max_j |x_j|$ then

$$|q_{ii}-\lambda| \leqslant \sum_{j\neq i} |q_{ij}| = \sum_{j\neq i} q_{ij} = -q_{ii}.$$

Since q_{ii} is non-positive this inequality can not hold for any purely imaginary λ or any λ with positive real part.

Thus an eigenvalue of Q is 0 or has negative real part. Note that 0 is always an eigenvalue since $Q\underset{\sim}{1} = \underset{\sim}{0}$.

Write Q in Jordan canonical form

$$Q = GLG^{-1} \quad \text{where,}$$

defining

$$A \oplus B = \left[\begin{array}{c:c} A & 0 \\ \hdashline 0 & B \end{array}\right] ,$$

$L = L_1 \oplus L_2 \oplus .. \oplus L_r$. The L_i take the form $\lambda I_j + N_j$ ~~with~~ WHERE λ is an eigenvalue of Q and N_j the $j \times j$ matrix having $(\ell,\ell+1)$st element 1, $\ell = 1,..,j-1$, and 0's elsewhere:

$$N_j = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix}$$

Note that $N_1 = 0$ and since $N_j^j = 0_j$, $e^{N_j t}$ is simply a polynomial in t.

Now

$$e^{L_1 t} \oplus .. \oplus e^{L_r t} = e^{Lt} = GP(t)G^{-1}.$$

The elements of $GP(t)G^{-1}$ are bounded in t since $P(t)$ is a Markov matrix and so the elements of the $e^{L_i t}$ are bounded. However if $L_i = N_j$ then this can only be so if $j = 1$ as otherwise $e^{N_j t}$ is unbounded in t. Thus the algebraic and geometric multiplicities of 0 as an eigenvalue of Q are equal and if $q = \text{rank } Q$ then

$$L = 0_q \oplus L_s \oplus L_{s+1} \oplus .. \oplus L_r$$

with each of the L_i in this representation having the form $\lambda I_j + N_j$ where λ has negative real part. For such L_i

$$e^{L_i t} = e^{\lambda t} e^{N_j t}$$

from which it is clear that $e^{L_i t} \to 0_j$ as $t \to \infty$.

Hence

$$e^{Lt} = I_q \oplus e^{L_s t} \oplus .. \oplus e^{L_r t} \to I_q \oplus 0_{k-q}$$

as $t \to \infty$ or $P(t) \to G(I_q \oplus O_{k-q})G^{-1}$.

Each of the 0 eigenvalues of Q corresponds to a closed class of a Markov chain with rates Q, and the (i,j)th element of $J(Q)$ is non zero if and only if j is a member of a closed class and j is accessible from i. State j is accessible from i if either $i = j$ or there is a sequence $i_1, i_2, .., i_r$ such that $q_{ii_1} q_{i_1 i_2} .. q_{i_r j} \neq 0$. This occurs because $p_{ij}(t) > 0$ for any and all $t > 0$ if and only if $q_{ij} \neq 0$ or $i = j$. Thus the structure of $J(Q)$ is apparent on inspection of Q.

At first sight it appears unlikely that more than one communicating class will be present in the animal movement context. However it is quite possible for animals to become trapped in some places and for some places to be inaccessible from others and so the full generality of this treatment is not superfluous.

The result of Theorem 6 gives the generating function for the limiting distribution $\underset{\sim}{N}(t)$ as

$$M\left(\underset{\sim}{m}, J(Q), \underset{\sim}{s}\right) \tag{22}$$

which is true for both model 2c and model 3d when equation (18) holds and $P(A_\ell > 0) = 1$. Note that in the most common case, where there is only one closed class for Q, the limiting distribution is simply multinomial.

Model 3d assumes that an animal's transition matrix, P_ℓ, does not depend on its initial position. This assumption can be justified when (19) holds and the animal can not be dead, $\left(P(A_\ell > 0) = 1\right)$, because then $\lim\limits_{t\to\infty} e^{QA_\ell t}$ does not depend on A_ℓ and so $X_\ell(t)$ and A_ℓ are independent in the limiting distribution, $t \to \infty$. Hence the independence assumption is reasonable, provided the system has been in existence sufficiently long to reach equilibrium before observation begins.

Now that the limiting behaviour has been settled it is easy to identify a large class of matrices Q for which $\underset{\sim}{N}$ can not be a Markov process unless A_ℓ is almost surely constant. I shall prove this for a modication of 3d which makes the proof much easier.

Let $Q = G\Lambda G^{-1}$ where Λ is diagonal and at least one element of Λ is a negative real number λ. Now $\lim_{t\to\infty} e^{\underset{\sim}{Q}t} = 1\underset{\sim}{\pi}'$. Let $\underset{\sim}{\pi}$ have positive components which means the elements of $e^{\underset{\sim}{Q}t}$ are all positive for $t > 0$. Let model 3d and equation (19) hold except that 2. of 3d is replaced by

2′. $X_1(0),\ldots,X_n(0)$ are i.i.d. with $P\big(X_\ell(0)=j\big) = \pi_j$.

With this initial distribution, conditional on $(A_1,\ldots,A_\ell)$, $X_1,\ldots,X_n$ are independent stationary Markov chains. It follows that $\underset{\sim}{X}$ is a stationary process and the conditional distribution of $\underset{\sim}{N}(t+h)$ given $\underset{\sim}{X}(h)$ depends on h only through $\underset{\sim}{X}(h)$. However this conditional distribution depends on $\underset{\sim}{X}(h)$ only through $\underset{\sim}{N}(h)$ which shows that if $\underset{\sim}{N}$ is a Markov process it is homogeneous. Now $E[\underset{\sim}{N}(t)\,|\,\underset{\sim}{N}(0)] = \big(EP_1(t)\big)\underset{\sim}{N}(0)$ and the homogeneous Markov property gives

$$E[N(t)\,|\,\underset{\sim}{N}(0)] = E[..E[E[\underset{\sim}{N}(t)\,|\,\underset{\sim}{N}(t-1)]\,|\,\underset{\sim}{N}(t-2)]\;..\;|\,\underset{\sim}{N}(0)]$$
$$= \big(EP_1(1)\big)^t\underset{\sim}{N}(0).$$

Since $\underset{\sim}{N}(0)$ can take any value summing to n, with positive probability, this means $EP_1(t) = \big(EP_1(1)\big)^t$ and it follows that $Ee^{\Lambda A_1 t} = (Ee^{\Lambda A_1})^t$, in particular $Ee^{\lambda A_1 t} = (Ee^{\lambda A_1})^t$. Thus the moments of $e^{\lambda A_1}$ are the same as the moments of the constant $Ee^{\lambda A_1}$, which implies $e^{\lambda A_1} \overset{a.s.}{=} Ee^{\lambda A_1}$ and A_1 is almost surely constant.

It seems as though one should be able to show that $\underset{\sim}{N}$ can not be Markov for an unmodified 3d by simply conditioning on $\underset{\sim}{N}(0)$ in the above result. However this method fails because the conditional distribution of $\underset{\sim}{N}(t)$, given $\underset{\sim}{N}(t-1)$ and $\underset{\sim}{N}(0)$, depends on $\underset{\sim}{N}(0)$. Use of the asymptotic stationarity of $\underset{\sim}{N}(.+h)$, as $h \to \infty$, looks more promising.

Continuous Time Models

A continuous time analogue to model 3d is

3cl.

1. There are random rates matrices $Q_1,..,Q_n$ such that, conditional on $Q_1,..,Q_n$, the processes $X_1,..,X_n$ are independent homogeneous Markov chains in continuous time with X_ℓ having rates matrix Q_ℓ.
2. $Q_1,..,Q_n$ are independent and identically distributed and are jointly independent of $\underset{\sim}{X}(0)$.
3. m_i of the X_ℓ satisfy $P\left(X_\ell(0)=i\right) = 1$.

The matrix Q_ℓ can be thought of as a function of the phenotype of animal ℓ.

The generating function for $\underset{\sim}{N}(t)$ is

$$M(\underset{\sim}{m}, Ee^{Q_1 t}, \underset{\sim}{s}) \tag{23}$$

As before I suggest that a useful model may be $Q_\ell = QA_\ell$ so that $P_\ell(t) = e^{QA_\ell t}$, where $A_1,..,A_n$ are i.i.d. The limit $\lim_{t\to\infty} P_\ell(t)$ exists, is non random, and there is some justification for assuming $\underset{\sim}{X}(0)$ and $Q_1,..,Q_n$ to be independent. The distribution of $\underset{\sim}{N}(t)$ has the same limit as its discrete time counterpart.

In continuous time it is possible to examine the meaning of the model $Q_\ell = QA_\ell$ a little more deeply.

A complete set of invariants for the mapping $t \mapsto e^{Qt}$, under scale changes, $t \mapsto \alpha t$, is

$$-q_{ij}/q_{ii}\ ,\quad i,j = 1,..,k\quad (\text{for}\ \ q_{ii} \neq 0)$$

and $$q_{ii}/q_{..}\ ,\quad i = 1,..,k,\quad \text{where}\ \ q_{..} = \Sigma q_{ii}.$$

For a continuous time Markov chain, with rates Q, $-q_{ij}/q_{ii}$ is the probability that the next change of state is to j, given currently in i. If currently in i the waiting time, T, for the next change of state is exponential with parameter $-q_{ii}$ that is

$$P(T>u) = e^{q_{ii}u} = e^{q_{..}(q_{ii}/q_{..})u}.$$

Thus a difference in activity does not affect the probability distribution for the next state but just the waiting time distribution for the next change of state. The waiting time distributions for one animal are a constant scale change of the waiting time distributions for another.

The model $Q_\ell = QA_\ell$ is particularly pleasing because it is capable of interpretation, results in only a small increase in complexity over the earlier models, and leads to tractable mathematics. However there are likely to be many practical situations where it is inadequate and so I shall consider one final model for this section. This model, 3c2, anticipates the developments of the next section.

Consider introducing variability in the case where individuals' movement processes are non homogeneous Markov chains for which

$$P(s,t) = \exp\{Q\int_s^t a(u)du\} = \exp\{Q(A(t)-A(s)\}\ ,\ A(t) = \int_0^t a(u)du.$$

The pair (Q,A) is replaced by the set $\{(Q_\ell,A_\ell),\ \ell = 1,..,n\}$, such that conditional on the (Q_ℓ,A_ℓ), $X_1,..,X_n$ are independent Markov chains with $P_\ell(s,t) = \exp\{Q_\ell(A_\ell(t)-A_\ell(s))\}$. We shall assume A_ℓ is the integral of a_ℓ, the ℓth animal's level of activity function, and $A_\ell(t) \overset{a.s.}{\to} \infty$ as $t \to \infty$.

In order that Q_ℓ and A_ℓ should be identifiable from P_ℓ, some restriction must be placed on them. The discussion of the invariants of e^{Qt} under scale change suggests that a natural restriction is trace Q = a constant, 1 say. With this restriction it is conceivable that A_ℓ and Q_ℓ are independent because Q_ℓ only contains information about the animal's relative preference for each area and relative amount of time spent in each area, whereas A_ℓ depends on something quite different, namely the rate of movement or level of activity at each time point.

In model 3cl, with $P_\ell(s,t) = e^{QA_\ell(t-s)}$, $\lim_{t\to\infty} e^{QA_\ell t}$ is a constant and it has been argued previously that when the system has been in existence a sufficiently long time $X_\ell(t)$ and P_ℓ are independent, for fixed t. This is the jusification for assuming that the conditional distribution of P_ℓ, given $X_\ell(0)$, does not depend on $X_\ell(0)$. Alternatively $X_\ell(0)$ and P_ℓ can be independent by experimental design. In the present context $\lim_{t\to\infty} e^{Q_\ell A_\ell t}$ may not be constant so that independence of P_ℓ and $X_\ell(0)$ can not be justified on the grounds of the age of the system. However it still seems sensible to base the conditional distribution of P_ℓ given $X_\ell(0)$ on the joint limiting distribution of P_ℓ and $X_\ell(t)$, $t \to \infty$, for a randomly chosen animal. I argue as follows :

If the animals are numbered at random then previous considerations lead to the postulate that (Q_ℓ, A_ℓ), $\ell = 1,..,n$ are i.i.d. Use the notation π for the common marginal distribution of the pairs (Q_ℓ, A_ℓ). π is a probability measure on the space (Ω_1, A_1). Ω_1 is the set of all (Q,A)

where $Q = (q_{ij})$ is a $k \times k$ rates matrix and A is the integral of a non negative Lebesgue measurable function. $\mathcal{A}_1$ is the σ-algebra generated by all the sets $\{(Q,A)\in\Omega_1 | q_{ij}\in B\}$, and $\{(Q,A)\in\Omega_1 | A(t)\in B\}$ where B varies over the Borel sets in R^1, $t \geqslant 0$ and $i,j = 1,..,k$.

We shall assume that, $A_\ell(t) \xrightarrow{a.s.} \infty$ as $t \to \infty$, and $\mathcal{J}(Q_\ell)$ has rank 1 a.s. so that $\mathcal{J}(Q_\ell) \overset{a.s.}{=} \underset{\sim}{1}\underset{\sim}{\gamma}_\ell'$ for a vector $\underset{\sim}{\gamma}_\ell = (\gamma_{1\ell},..,\gamma_{k\ell})'$. This is equivalent to there being only one closed class for the Markov chain with rates Q_ℓ. Note that $\mathcal{J}(Q_\ell)$ is a random variable since it equals $\lim_{t\to\infty} e^{Q_\ell t}$. Finally only the case $\mathcal{E}\gamma_{j\ell} > 0$ will be considered since otherwise the animal leaves area j never to return a.s.

Whatever the joint distribution of $X_\ell(0)$, Q_ℓ, A_ℓ $\lim_{t\to\infty} \mathcal{P}(X_\ell(t)=j | Q_\ell, A_\ell) = \gamma_{j\ell}$. Thus, applying the dominated convergence theorem,

$$\mathcal{P}(X_\ell(t)=j, (Q_\ell,A_\ell)\in D) \to \mathcal{E}\gamma_{j\ell} I_D(Q_\ell,A_\ell)$$

for $D \in \mathcal{A}_1$. This means

$$\mathcal{P}((Q_\ell,A_\ell)\in D | X_\ell(t)=j) \to \mathcal{E}\gamma_{j\ell} I_D(Q_\ell,A_\ell)/\mathcal{E}\gamma_{j\ell}$$

which equals $\int_D \gamma_j d\pi / \int \gamma_j d\pi = \pi_j(D)$ say. γ_j is a function on Ω_1 defined by $\underset{\sim}{\gamma}' = \underset{\sim}{1}'\mathcal{J}(Q)/k$. The measure π_j is a plausible conditional distribution for (Q_ℓ,A_ℓ) when $X_\ell(0) = j$. We can now state model 3c2.

3c2.

1. There are random rates matrices $Q_1,..,Q_\ell$ and random functions $A_1,..,A_\ell$ such that conditional on (Q_1,A_ℓ), $\ell = 1,..,n$ the stochastic processes $X_1,..,X_n$ are independent continuous time Markov chains with transition matrices

$$P_\ell(s,t) = \exp\{Q_\ell(A_\ell(t)-A_\ell(s))\}.$$

2. m_i of the X_ℓ satisfy $\mathcal{P}(X_\ell(0)=i) = 1$.

3. Conditional on $\underset{\sim}{X}(0)$, the pairs (Q_ℓ,A_ℓ), $\ell = 1,..,n$, are independent with (Q_ℓ,A_ℓ) having $\underset{\sim}{X}(0)$ conditional distribution $\pi_{X_\ell(0)}$.

Note that under model 3c2 (Q_ℓ,A_ℓ), $\ell = 1,..,n$, are not identically distributed. This apparent conflict with the previous assumption comes about because model 3c2 is essentially a model which conditions on $\underset{\sim}{X}(0)$. One could replace postulates 2. and 3. by

2! $(X_\ell(0),Q_\ell,A_\ell)$, $\ell = 1,..,n$, are i.i.d. with distribution $\mathcal{P}(X_\ell(0)=j,(Q_\ell,A_\ell)\in D) = \int_D \gamma_j d\pi$.

Model 3c2 then comes about by conditioning on $\underset{\sim}{X}(0)$.

Form the matrix $R(t)$ by selecting for its jth row the jth row of the matrix

$$\int e^{QA(t)} d\pi_j .$$

Under model 3c2 the distribution of $\underset{\sim}{N}(t)$ is a convolution of multinomials with generating function

$$\mathcal{M}(\underset{\sim}{m},R(t),\underset{\sim}{s}) \tag{24}$$

Since $A(t) \to \infty$ a.s. (π_j) it follows that $\int e^{QA(t)} d\pi_j \to \int \underset{\sim}{1}\underset{\sim}{\gamma}' d\pi_j$, by the dominated convergence theorem. Define $R(\infty)$ as the matrix with jth row $\int \underset{\sim}{\gamma}' d\pi_j$ then $R(\infty) = \lim_{t\to\infty} R(t)$ and the limiting distribution of $\underset{\sim}{N}(t)$ is a convolution of multinomials with generating function

$$\mathcal{M}(\underset{\sim}{m},R(\infty),\underset{\sim}{s}) \tag{25}$$

This convolution of multinomials only reduces to a multinomial when $\underset{\sim}{\gamma}$ is constant a.s. (π). To see this note the obvious result that (25) only represents a multinomial when $R(\infty)$ has equal rows. The latter is proved, in any case, in chapter 3. For $R(\infty)$ to have equal rows

$$\int\gamma_j d\pi_j = \int\gamma_j d\pi_i \quad \text{for all} \quad i,j.$$

Multiplying by $\int\gamma_i d\pi$ and summing this yields

$$\int\gamma_j^2 d\pi / \int\gamma_j d\pi = \int\gamma_j d\pi$$

since $\int\gamma_j d\pi_i = \int\gamma_j\gamma_i d\pi / \int\gamma_i d\pi$ and $\Sigma\gamma_i = 1$. Thus $\int\gamma_j^2 d\pi = (\int\gamma_j d\pi)^2$ implying that γ_j is constant a.s. (π).

In all previous limiting distributions for $\underset{\sim}{N}(t)$ a convolution of multinomials, not reducing to a multinomial, was a consequence of the existence of several closed classes in the Markov chains used to construct the movement processes. However (24) is purely a consequence of variation between individuals.

This model of variation between individuals is helpful in thinking about the home range concept.

The individuals of many populations of animals are found to possess what is generally referred to as a home range. This means that the individual animal does not stray far from a particular place. The precise definition of the home range varies. Burt (1943) says the home range is "that area traversed by an individual in its normal activities of food gathering, mating, and caring for its young. Occasional sallies outside the area, perhaps exploratory in nature, should not be considered as part of the home range". Reviews of this concept and methods of measuring parameters associated with the home range are given by Sanderson (1966), Brown and Orians (1970) and Wierzbowska (1975).

I believe that a theoretically more satisfactory concept is a "home distribution". In this discrete space setting the home distribution, H_ℓ, of the ℓth animal can be defined as the limiting distribution of $X_\ell(t)$, $t \to \infty$, conditional on Q_ℓ. In the case where $A_\ell(t) = a_\ell t$, for some positive random variable a_ℓ, Markov chain theory shows that an equivalent definition of the home distribution is

$$H_\ell(D) = \lim_{t\to\infty} \frac{1}{t}\int_0^t I_D\bigl(X_\ell(u)\bigr)du \ , \tag{25}$$

for D a subset of $\{1,2,..,k\}$, because the right hand side converges a.s. to $\sum_{j\in D} \gamma_{j\ell}$. Thus the home distribution is the proportion of time animal ℓ spends in each area. If (25) is interpreted as a limit in mean square (m.s.) this definition is available whenever

$$\lim_{t\to\infty} A_\ell(tv) - A_\ell(tu) \overset{a.s.}{=} \infty \tag{26}$$

for all $v > u \geqslant 0$. Basically (26) insists that the animal does not slow down too quickly.

Theorem 7

If (26) holds then

$$\frac{1}{t}\int_0^t I_D\bigl(X_\ell(u)\bigr)du \overset{m.s.}{\to} \sum_{j\in D} \gamma_{j\ell}$$

Proof

If p is a polynomial and α,β are real numbers, with $\alpha > 0$, then $\left|e^{-2\alpha t+i\beta t}p(t)\right| = e^{-2\alpha t}\left|p(t)\right| \to 0$ as $t \to \infty$. Thus defining $K = \sup_t e^{-\alpha t}\left|p(t)\right| < \infty$ we have $\left|e^{-2\alpha t+i\beta t}\right| \leqslant K\,e^{-\alpha t}$. Using this fact and the proof of theorem 6 it is clear that

$\|\underset{\sim}{1}\underset{\sim}{\gamma}_\ell{}' - e^{Q_\ell t}\| \leqslant K\,e^{-\alpha t}$, for K, $\alpha > 0$ depending on Q_ℓ. Thus for $v > u > 0$

$$\left|P\bigl(X_\ell(u)=j, X_\ell(v)=j\,|\,Q_\ell,A_\ell\bigr) - \gamma_{j\ell}P\bigl(X_\ell(u)=j\,|\,Q_\ell,A_\ell\bigr)\right| \leqslant K\,e^{-\alpha\left(A(v)-A(u)\right)},$$

and so defining $Z(t) = I_{\{j\}}\bigl(X_\ell(t)\bigr) = \max\bigl(0, 1-|X_\ell(t)-j|\bigr)$, it follows that $E[Z(tu)Z(tv)\,|\,Q_\ell,A_\ell] \to \gamma_{j\ell}^2$ as $t \to \infty$. However $E[Z(tu \text{ or } v)\,|\,Q_\ell,A_\ell] \to \gamma_{j\ell}$ and so $C\bigl(Z(tu),Z(tv)\,|\,Q_\ell,A_\ell\bigr) \to 0$.

Let $J(t) = \frac{1}{t}\int_0^t Z(u)du$, then $J(t) = \int_0^1 Z(tu)du$. Conditional on (Q_ℓ, A_ℓ), X_ℓ is continuous in probability and this also holds for Z because $Z(t)$ is a continuous function of $X_\ell(t)$. Additionally $|Z(t)| \leqslant 1$ and so $C(Z(t), Z(s)|Q_\ell, A_\ell)$ is continuous in (s,t). It follows that $\int_0^1 Z(tu)du$ exists as an integral in mean square in the sense of Cramér and Leadbetter (1967). Furthermore

$$E[\int_0^1 Z(tu)du|Q_\ell,A_\ell] = \int_0^1 E[Z(tu)|Q_\ell,A_\ell]du$$
$$= \int_0^1 P(X_\ell(tu)=j|Q_\ell,A_\ell)du \quad \text{and}$$
$$V(\int_0^1 Z(tu)du|Q_\ell,A_\ell) = \int_0^1\int_0^1 C(Z(tu),Z(tv)|Q_\ell,A_\ell)dudv.$$

All integrands above are bounded and so the dominated convergence theorem yields

$$E[J(t)|Q_\ell,A_\ell] \to \gamma_{j\ell}, \quad V(J(t)|Q_\ell,A_\ell) \to 0$$

Hence $J(t) \overset{m.s.}{\to} \gamma_{j\ell}$ where, because of boundedness of $V(J(t)|Q_\ell,A_\ell)$, the m.s. convergence holds both conditional on Q_ℓ, A_ℓ and not conditional on Q_ℓ, A_ℓ.

To see that (26) can not be weakened substantially consider the simple example where $A_\ell(t) = \log(t+1)$ and

$$Q = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}.$$

$$P_\ell(s,t) = \tfrac{1}{2}\begin{bmatrix} 1+b & 1-b \\ 1-b & 1+b \end{bmatrix}, \quad b = \frac{s+1}{t+1}$$

and so $P_\ell(ut,vt)$ converges to

$$\tfrac{1}{2}\begin{bmatrix} 1+\frac{u}{v} & 1-\frac{u}{v} \\ 1-\frac{u}{v} & 1+\frac{u}{v} \end{bmatrix}.$$

Thus whether $j = 1$ or 2

$$P(X_\ell(tu)=j, X_\ell(tv)=j) = P(X_\ell(tv)=j \mid X_\ell(tu)=j)\ P(X_\ell(tu)=j)$$
$$\to \tfrac{1}{4}\left(1+\frac{u}{v}\right) \quad \text{as} \quad t \to \infty,$$

so that $C(Z(tu),Z(tv)) \to \frac{1}{4}\frac{u}{v} = \frac{1}{4}\min(u,v)/\max(u,v)$.

As in the theorem $V\left(\int_0^1 Z(tu)du\right) = \int_0^1\int_0^1 C\left(Z(tu),Z(tv)\right)dudv$ $\to \frac{1}{4}\int_0^1\int_0^1 \min(u,v)/\max(u,v)\,dudv > 0$. Hence $J(t) = \frac{1}{t}\int_0^t Z(tu)du = \int_0^1 Z(tu)du$ does not converge in mean square to $\gamma_{j\ell}$. Since $J(t)$ is bounded, $J(t) \not\to \gamma_{j\ell} = \frac{1}{2}$ in any common sense.

A construction.

Model 3c2 was developed on the basis of the probability space (Ω_1, A_1, π). However it was not made clear that model 3c2 would always make sense when only the restrictions $\pi\{(Q,A)\,|\,\text{rank } J(Q)=1\} = 1$ and $\int \gamma_j d\pi > 0$, for each j, are imposed. The construction below shows that no further restrictions are necessary. This construction also indicates how to construct processes satisfying the other models in this chapter, in particular, the models of the next section. The main interest in such constructions is the demonstration that there is a non trivial class of processes satisfying the model.

Let Ω_2 be the set of all functions $x : [0,\infty) \to \{1,..,k\}$. Define A_2 to be the σ-algebra generated by the sets $\{x \in \Omega_2 \,|\, x(t) \in B\}$ for $t \geqslant 0$ and $B \subset \{1,..,k\}$. Define

$$P(s,t) = e^{Q\left(A(t)-A(s)\right)}.$$

$P(s,t)$ satisfies the Chapman-Kolmogorov equations $P(r,s)P(s,t) = P(r,t)$ so that by a well known theorem (see Breiman 1968, p.321) there is a measure μ on (Ω_2, A_2) such that the process $Y = \{Y(t), t \geqslant 0\}$, defined on (Ω_2, A_2) by

$$Y(t,x) = x(t), \quad x \in \Omega_2,$$

is a Markov process with transition matrix $P(s,t)$ and initial distribution $\mu\left(Y(0)=j\right) = \gamma_j$.

The measure μ is a function on the space Ω_1. In order to obtain the desired construction it is necessary to show that μ has the properties of a regular conditional distribution for Y given (Q,A). To complete this all that has to be shown is that $\mu(E)$ is an $\mathcal{A}_1$ measurable function for every E in $\mathcal{A}_2$.

Now $P(s,t)$ and $\underset{\sim}{\gamma}$ are certainly $\mathcal{A}_1$ measurable and if $E \in \mathcal{A}_2$ depends on only a finite number of coordinates, that is $E = \{x | (x(t_1),..,x(t_p)) \in B\}$ for a finite set $t_1,..,t_p$, and $B \subset \{1,..,k\}^p$, then $\mu(E)$ can be expressed in terms of finite sums of finite products of elements $P(s,t)$, $t > s \geqslant 0$ and $\underset{\sim}{\gamma}$. For such $E, \mu(E)$ is $\mathcal{A}_1$ measurable. To prove $\mathcal{A}_1$ measurability of $\mu(E)$ for all $E \in \mathcal{A}_2$ we use the well known monotone class theorem which is stated in chapter 3. Let $\mathcal{C}$ be the set of all $E \in \mathcal{A}_2$ such that $\mu(E)$ is $\mathcal{A}_1$ measurable. We know that the set of members of $\mathcal{A}_2$ depending only on a finite number of coordinates is a subset of $\mathcal{C}$, generates $\mathcal{A}_2$, and is closed under intersection. Clearly if $\{E_n\}$ is a increasing sequence in $\mathcal{C}$, $\bigcup_{n=1}^{\infty} E_n \in \mathcal{C}$; and $E_1, E_2 \in \mathcal{C}$, $E_1 \subset E_2$ implies $E_2 - E_1 \in \mathcal{C}$. Finally $\Omega_2 \in \mathcal{C}$ which proves $\mathcal{C} = \mathcal{A}_2$ as required.

Since μ has the properties of a regular conditional distribution, a probability measure τ on $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$ is defined by the equation

$$\tau(E_1 \times E_2) = \int_{E_1} \mu(E_2) d\pi, \; E_1 \in \mathcal{A}_1, \; E_2 \in \mathcal{A}_2$$

The CARATHÉODORY ~~Catheodory~~ extension theorem ensures that this equation is sufficient to define τ on all of $\mathcal{A}_1 \times \mathcal{A}_2$.

Now define $X(t)$ to be the random variable on $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2, \tau)$ for which $X(t,Q,A,x) = x(t)$. For simplicity of notation, Q and A denote the mappings $(Q,A,x) \mapsto Q$ and $(Q,A,x) \mapsto A$. By construction, conditional on (Q,A), $X = \{X(t), t \geq 0\}$ is a Markov chain with transition matrix $P(s,t)$ and initial distribution $\tau(X(t)=j|Q,A) = \gamma_j$.

Forming the product space $((\Omega_1 \times \Omega_2)^n, (A_1 \times A_2)^n, \tau^n)$ we obtain a space on which processes X_ℓ are defined and 1. of 3c2 and 2.$'$ are satisfied.

Finally let $\tilde{\tau}(E) = \frac{\tau^n(E \cap F)}{\tau^n(F)}$, $E \in (A_1 \times A_2)^n$ where $F = \{X_\ell(0)=i, \sum_{j=1}^{i-1} m_j < \ell \leq \sum_{j=1}^{i} m_j, i=1,\ldots,k\}$. The space $((\Omega_1 \times \Omega_2)^n, (A_1 \times A_2)^n, \tilde{\tau})$ satisfies model 3c2.

2.4 Environmental and other effects.

Animals movements may be affected by the weather, or more generally the environment, as was discussed in section 2.2 in relation to the model

$$P(s,t) = \exp\{Q \int_s^t a(u)\,du\} \tag{12}$$

where a was regarded as a fixed function. It seems satisfactory to regard a as fixed in the following situations.

1. The function a can be predicted accurately

or 2. The function a can be measured and one is only interested in estimating past events.

or 3. The animals' movements are measured together with environmental variables to determine the dependence of a on the environmental variables. This is essentially a regression situation.

It will only be in special circumstances, such as laboratory experiments that a, regarded as dependent on environmental variables, will be completely predictable; and it may be difficult to measure a or to measure environmental variables on which a might depend. These considerations suggest that it is desirable to model a as a random function even though the problem of assigning a distribution to a may be very difficult.

In line with the thinking of section 2.3 a will be allowed to vary from one individual to another so that it is represented by a function a_ℓ for animal ℓ. However the environment is regarded as common to all animals so that it will affect their movements in a dependent manner. Thus a means of specifying dependence between the a_ℓ is required.

So far the kinds of variation that have been considered are variation between individuals (section 2.3) and variation common to all individuals (environmental effects). A final source of variation is within individuals. The animals may have bursts of activity interspersed with periods of rest or lower activity. In some of the periods of lower activity an animal simply may be preoccupied with something apart from moving. This kind of variation is most likely not deterministic and may be specific to the individual. The movement processes themselves seem to provide within individual variation since if Q_ℓ and a_ℓ are given (held fixed) and X_ℓ is a Markov chain with transition matrix $P_\ell(s,t) = \exp\{Q_\ell \int_s^t a_\ell(u)\,du\}$ then X_ℓ has variation which can only be described as within individual. However the Markov property ensures that conditional on $X_\ell(t)$ the waiting time for the next change of state, after t, is independent of waiting times for changes of state which occur before t. If there are genuine bursts of activity and periods of low activity, then this independence of waiting times seems extremely unlikely. Thus the Markov process

X_ℓ with Q_ℓ, a_ℓ held fixed does not provide sufficiently general within individual variation. I shall seek ways of introducing within individual variation through the function a_ℓ. Such within individual variation will be referred to as "random activity".

A number of studies have focused on variation of the three kinds considered here. Of these I cite three examples where the results are more definite. Hindley's (1975) study of Banana Prawns has already been mentioned in relation to environmentally dependent activity. He also found variation of the nature of random activity and between individual variation. Lipton and Sutherland (1970) were mostly interested in variation in activity between individuals; but as well as demonstrating such variation, the activity graphs in their paper show considerable within individual variation. Finally, a complex pattern of environmentally dependent and random activity was found by Marten (1973) in a field study of a small rodent *Peromyscus*.

Model 4c.

To incorporate all of the different kinds of variation in a single model assume that for each individual animal there is a triple (Q_ℓ, a_ℓ, S_ℓ). Q_ℓ is a random rates matrix for which rank $J(Q_\ell) = 1$ as in section 2.3. $a_\ell = \{a_\ell(t,u,v),\ t \geqslant 0,\ u,v \in R'\}$ is a non-negative random function of three variables which interprets the effect of the environment and random activity for animal ℓ, and regarded as phenotypically determined $S_\ell = \{S_\ell(t), t \geqslant 0\}$ is a random activity process which is independent of (Q_ℓ, a_ℓ).

$S_1,..,S_n$ are i.i.d. The triples (Q_ℓ,a_ℓ,S_ℓ), $\ell = 1,..,n$ are independent and if animals ℓ and ℓ' are initially in the same area then the distributions of (Q_ℓ,a_ℓ,S_ℓ) and $(Q_{\ell'},a_{\ell'},S_{\ell'})$ are identical. Finally there is an environment process $\sigma = \{\sigma(t),t\geq 0\}$ which is independent of (Q_ℓ,a_ℓ,S_ℓ), $\ell = 1,..,n$.

The model is essentially that, conditional on $(Q_\ell,a_\ell,S_\ell,\sigma)$, X_ℓ is a Markov chain with transition matrix

$$P_\ell(s,t) = \exp\{Q_\ell \textstyle\int_s^t a_\ell(u,S_\ell(u),\sigma(u))du\}.$$

Some conditions are required to ensure that the integral will exist and be a random variable. Recall that the underlying probability space is $(\Omega,\mathcal{A},P)$ and so assume that a_ℓ is $\mathcal{B}^3 \times \mathcal{A}$ measurable, where $\mathcal{B}^p$ is the Borel sets in R^p. This means that for each $B \in \mathcal{B}^1$,

$$\{(t,u,v,\omega)\,|\,a_\ell(t,u,v,\omega)\in B\} \in \mathcal{B}^3 \times \mathcal{A}.$$

Furthermore let both S_ℓ and σ be $\mathcal{B}^1 \times \mathcal{A}$ measurable then the composition $a_\ell(u,S_\ell(u),\sigma(u),\omega)$ as a function of (u,ω) is $\mathcal{B}^1 \times \mathcal{A}$ measurable. The integral

$$A_\ell(t) - A_\ell(s) = \int_s^t a_\ell(u,S_\ell(u),\sigma(u))du$$

is defined as an ordinary Lebesgue integral and is $\mathcal{A}$ measurable, that is, it is a random variable. It will also be assumed that the integral is finite a.s. and $\to \infty$ as $t \to \infty$. Random functions which satisfy the measurability conditions above are discussed by Blumenthal and Getoor (1968, p.34).

With these preliminaries model 4c can be stated.

4c.

1. There are random triples (Q_ℓ,a_ℓ,S_ℓ), $\ell = 1,..,n$ and a random process σ such that conditional on σ, (Q_ℓ,a_ℓ,S_ℓ), $\ell = 1,..,n$, the stochastic processes

$X_1,..,X_n$ are independent Markov chains with transition matrices

$$P_\ell(s,t) = \exp\{Q_\ell \int_s^t a_\ell(u,S_\ell(u),\sigma(u))du\}.$$

Q_ℓ is a rates matrix for which $J(Q_\ell) = 1$ a.s.

2. m_i of the X_ℓ satisfy $P(X_\ell(0)=i) = 1$.

3. Conditional on $\underset{\sim}{X}(0)$ the process and triples σ, $(Q_1,a_1,S_1),..,(Q_n,a_n,S_n)$ are mutually independent. Furthermore S_ℓ and (Q_ℓ,a_ℓ) are independent and the S_ℓ, $\ell = 1,..,n$ are i.i.d. The conditional distribution of (Q_ℓ,a_ℓ) depends only on $X_\ell(0)$.

The model for the distribution of (Q_ℓ,a_ℓ) is essentially the same as that for (Q_ℓ,A_ℓ) in 3c2. If S_ℓ and σ are non random model 4c reduces to model 3c2.

Define $R(t)$ to be the random matrix which has jth row equal to the jth row of the matrix

$$E[\exp\{Q_\ell \int_s^t a_\ell(u,S_\ell(u),\sigma(u))du\}|\sigma],$$

for ℓ such that $P(X_\ell(0)=j) = 1$. Conditional on σ, $\underset{\sim}{N}(t)$ has generating function

$$M(\underset{\sim}{m},R(t),\underset{\sim}{s}) \qquad (28)$$

so that $\underset{\sim}{N}(t)$ is conditionally a convolution of multinomials.

Unconditionally the generating function of $\underset{\sim}{N}(t)$ is

$$EM(\underset{\sim}{m},R(t),\underset{\sim}{s}) \qquad (29)$$

The distribution of $\underset{\sim}{N}(t)$ is a mixture of convolutions of multinomials. However, since we assume $A_\ell(t) \to \infty$ as $t \to \infty$, $\lim_{t\to\infty} R(t) = R(\infty)$ does not depend on σ and so the limiting distribution is simply a convolution of multinomials.

Special cases

Several specialisations of model 4c are of interest.

A. Q_ℓ is equal to a constant matrix Q and correspondingly the a_ℓ are identically distributed. The generating function of $\underset{\sim}{N}(t)$ is

$$EM\left(\underset{\sim}{m}, E[e^{QA_\ell(t)}|\sigma], \underset{\sim}{s}\right) \tag{30}$$

and the limiting distribution is a mixture of multinomials.

B. Q_ℓ is as in A. and also σ is a non random function. The generating function is

$$M\left(\underset{\sim}{m}, R(t), \underset{\sim}{s}\right) \tag{31}$$

which is of the same form as the generating function for $\underset{\sim}{N}(t)$ under model 3c2. If one observes realisations of $\underset{\sim}{N}(t)$, $\underset{\sim}{N}_1(t), \ldots, \underset{\sim}{N}_m(t)$, say, using a different set of animals for each realisation of the process, then the models 3c2 and 4cB are indistinguishable; but if the same set of animals is used for each realisation then the between realisation variability is greater under model 4cB than 3c2. This difference results from the random activity processes S_ℓ.

C. Q_ℓ as in A, $a_\ell\left(u, S_\ell(u), \sigma(u)\right) = S_\ell(u)$ and specify $\int_{t-1}^{t} S_\ell(u)du$, $t = 0, 1, \ldots,$ i.i.d., for each ℓ.

$P\left(X_\ell(t)=j|X_\ell(t-1), \ldots, X_\ell(0), S_\ell(u), t\geqslant 0\right)$ = (i,j)th element of $\exp\{Q\int_{t-1}^{t} S_\ell(u)du\}$ when $X_\ell(t-1) = i$. Thus

$P\left(X_\ell(t)=j|X_\ell(t-1), \ldots, X_\ell(0)\right)$ = (i,j)th element of $E\exp\{Q\int_{t-1}^{t} S_\ell(u)du\}$ when $X_\ell(t-1) = i$, and since $E\exp\{Q\int_{t-1}^{t} S_\ell(u)du\}$ does not depend on i or ℓ, $X_1, \ldots, X_n$ are i.i.d. homogeneous Markov chains when sampled in discrete time. This is model 2d.

D. A plausible model for a_ℓ is

$a_\ell\left(u, S_\ell(u), \sigma(u)\right) = \alpha(u)A_\ell S_\ell(u)\sigma(u)$ where α is a fixed function and A_ℓ is a random variable. This is a multiplicative model. Model $S_\ell(u)$ as a stationary process $\int_{-\infty}^{t} g(t-u)\,dY_\ell(u)$ where g is a non negative function satisfying $\int_0^\infty g(u)\,du < \infty$ and Y_ℓ is a process with non negative stationary independent increments for which $E\big(Y_\ell(1)-Y_\ell(0)\big) < \infty$. This is a special case of a model discussed by Bartlett and Kendall (1951). Let f be a non negative function and λ a complex number with negative real part, then

$$\begin{aligned} & E\exp\{\lambda \int_0^t f(u) \int_{-\infty}^{u} g(u-v)\,dY_\ell(v)\,du\} \\ &= E\exp\{\lambda \int_{-\infty}^{t}\int_v^t f(u)g(u-v)\,du\,dY_\ell(v)\} \\ &= \exp\{\int_{-\infty}^{t}\psi\big(\lambda \int_v^t f(u)g(u-v)\,du\big)dv\} \end{aligned}$$

where ψ is the logarithm of the moment generating function of $Y_\ell(1) - Y_\ell(0)$. Thus putting $f = \alpha(u)A_\ell\sigma(u)$ and letting $Q_\ell = G_\ell\Lambda_\ell G_\ell^{-1}$, Λ_ℓ diagonal, we obtain a formula for $E[\exp\{Q_\ell A_\ell(t)\}|A_\ell,Q_\ell,\sigma]$ in terms of ψ.

An alternative to a multiplication model is an additive model $a_\ell\big(u,S_\ell(u),\sigma(u)\big) = \alpha(u) + A_\ell + S_\ell(u) + \sigma(u)$ which I think is probably best regarded as a first order Taylor approximation to a more complicated form. This represents a substantial simplification since $\exp\lambda\big(\alpha(u)+A_\ell+S_\ell(u)+\sigma(u)\big)$ is a product of independent terms.

Models of these kinds make the evaluation of the moment formulae given in the next section, a feasible proposition.

Conclusion

We have begun with the simplest models that reasonably might be proposed for animal movements. These have been examined for the short-comings that they have and generalisations have been proposed within certain bounds. The main bounds are the discrete space structure and independence between individual animals' movements.

The major concern has been with variation of different kinds. Three main kinds are recognised.

1. Environmental variation, common to all animals.
2. Between individual variation.
3. Within individual variation.

Random substitution of the time parameter proved to be a useful method for incorporating variation in movement models, but at the same time it is recognised that there are many other ways that these three kinds of variation might appear.

The most obvious way of generalising the models presented here is in the method of incorporating variation, particularly in relation to environmental variation, for the environment may well affect different areas differently. To do this one can return to the model of section 2.2 where the X_ℓ are non homogeneous Markov chains with transition rates a function of time. The transition rates can be made a function of the environment and in general stochastic. Saunders (1975) presented a model of a two compartment system where the transition rates are functions of countable Markov chains. He modelled $N_1(t)$ and the transition rates jointly as functions of a Markov chain with a special structure and showed that the distribution of $N_1(t)$ is a mixture of binomials when $N_1(0) = n$ and $N_2(0) = 0$.

The models of this chapter serve a number of purposes. They illustrate how one can approach the modelling of animal movements. They provide a means of summarising the ideas one might have about a given real situation. In these two respects the discrete space restriction is more apparent than real. Finally they provide models, both specific and general which can be applied to real situations. I mean one can collect data, estimate parameters and attempt to draw conclusions.

The next and last section of this chapter provides quantitative results in terms of moments and asymptotic distributions. The derivations of these results serve as a reminder of the interrelationships between the models.

2.5 Moments and asymptotic distributions.

Means and variances.

The generating functions for the distribution of $\underset{\sim}{N}(t)$, which can be thought of as conditional distributions for $\underset{\sim}{N}(t)$ given $\underset{\sim}{N}(0) = \underset{\sim}{m}$, fall into two classes:

1. $$g(\underset{\sim}{s}) = M(\underset{\sim}{m}, P, \underset{\sim}{s}),$$

where P is a $k \times k$ Markov matrix;

2. $$g(\underset{\sim}{s}) = EM(\underset{\sim}{m}, \tilde{P}, \underset{\sim}{s}),$$

where $\tilde{P}$ is a $k \times k$ random Markov matrix.

The first case represents a convolution of multinomials with parameters $(m_i, p_i') = (m_i, p_{i1}, \ldots, p_{ik})$, $i = 1, \ldots, k$. Hence $EN_j(t) = \Sigma\, m_j p_{ij}$ and so

$$E\underset{\sim}{N}(t) = P'\underset{\sim}{m} \tag{32}$$

The variances and covariances are

$$VN_j(t) = \Sigma\, m_i p_{ij}(1-p_{ij})$$

and $$C\big(N_j(t), N_\ell(t)\big) = -\,\Sigma\, m_i p_{ij} p_{i\ell}$$

Thus

$$V\underset{\sim}{N}(t) = \Sigma\, m_i\big(\mathrm{diag}(\underset{\sim}{p}_i) - \underset{\sim}{p}_i \underset{\sim}{p}_i'\big) \tag{33}$$

For the second class of generating functions $\underset{\sim}{N}(t)$ is a mixture of convolutions of multinomials. The mean is thus

$$E\underset{\sim}{N}(t) = E\tilde{P}'\underset{\sim}{m} \tag{34}$$

To calculate the variance matrix we use the well known formula

$$V\underset{\sim}{N}(t) = EV\left(\underset{\sim}{N}(t)\,|\,\tilde{P}\right) + VE\left(\underset{\sim}{N}(t)\,|\,\tilde{P}\right).$$

Substituting

$$V\underset{\sim}{N}(t) = \sum_i m_i E\left(\operatorname{diag}(\tilde{\underset{\sim}{p}}_i) - \tilde{\underset{\sim}{p}}\tilde{\underset{\sim}{p}}_i'\right) + E\tilde{P}'\underset{\sim}{m}\underset{\sim}{m}'\tilde{P}' - E\tilde{P}'\underset{\sim}{m}\underset{\sim}{m}'E\tilde{P}. \qquad (35)$$

Covariances.

In the calculation of the mean and variance of $\underset{\sim}{N}(t)$ it was possible to consider Markov and non Markov models together. However for the calculation of the covariance of $N(s)$ and $\underset{\sim}{N}(t)$, $t \neq s$, separate approaches are needed. These separate approaches serve to re-emphasise the essential differences between the models.

In the models of section 2.2 $\underset{\sim}{N}$ is a Markov process. Let $s < t$ and define P^* as the matrix with (i,j)th element $P\left(X_\ell(t)=j\,|\,X_\ell(s)=i\right)$. Now

$$E[\underset{\sim}{N}(t)\,|\,\underset{\sim}{N}(s)] = P^{*\prime}\underset{\sim}{N}(s)$$

$$\begin{aligned} C\left(\underset{\sim}{N}(s),\underset{\sim}{N}(t)\right) &= E\left(\underset{\sim}{N}(s)-E\underset{\sim}{N}(s)\right)\underset{\sim}{N}(t)' \\ &= EE[\left(N(s)-E\underset{\sim}{N}(s)\right)\underset{\sim}{N}(t)'\,|\,\underset{\sim}{N}(s)] \\ &= E\left(\underset{\sim}{N}(s)-E\underset{\sim}{N}(s)\right)\underset{\sim}{N}(s)'P^* \\ &= \left(V\underset{\sim}{N}(s)\right)P^* \end{aligned}$$

For the discrete time model 2d $P^* = P^{t-s}$ and

$$C\left(\underset{\sim}{N}(s),\underset{\sim}{N}(t)\right) = \left(V\underset{\sim}{N}(s)\right)P^{t-s}. \qquad (36)$$

The continuous time models of section 2, for which $P(s,t) = e^{Q\left(A(t)-A(s)\right)}$, have

$$C\left(\underset{\sim}{N}(s),\underset{\sim}{N}(t)\right) = \left(V\underset{\sim}{N}(s)\right)e^{Q\left(A(t)-A(s)\right)} \qquad (37)$$

The models of section 3 are non Markov. To derive the covariance we use the fact that the $Z_\ell(t) = \left(I_{\{1\}}(X_\ell(t)),\ldots,I_{\{k\}}(X_\ell(t))\right)$, $\ell = 1,\ldots,n$ are independent and so

$$C(\underset{\sim}{N}(s),\underset{\sim}{N}(t)) = \sum_{\ell=1}^{n} C(\underset{\sim}{Z}_\ell(s),\underset{\sim}{Z}_\ell(t)) \tag{38}$$

Consider first the discrete time model 3d with $P_\ell = e^{QA_\ell}$. Conditional on P_ℓ, $\underset{\sim}{Z}_\ell$ or equivalently X_ℓ, is a homogeneous Markov chain with transition matrix P_ℓ. If $P(\underset{\sim}{Z}_\ell(0)=\underset{\sim}{e}_i) = 1$ formula (36) applies with $n = m_i = 1$ and so

$$C(\underset{\sim}{Z}_\ell(s),\underset{\sim}{Z}_\ell(t)|P_\ell) = \{\text{diag}(\underset{\sim}{p}_i(s)) - \underset{\sim}{p}_i(s)\underset{\sim}{p}_i'(s)\}\, P_\ell^{t-s},$$

$\underset{\sim}{p}_i'(s)$ being the ith row of P_ℓ^s. Also $E[\underset{\sim}{Z}_\ell(t)|P_\ell] = (P_\ell')^t\underset{\sim}{e}_i$

Now

$$\begin{aligned} C(\underset{\sim}{Z}_\ell(s),\underset{\sim}{Z}_\ell(t)) &= EC(\underset{\sim}{Z}_\ell(s),\underset{\sim}{Z}_\ell(t)|P_\ell) + C(E[\underset{\sim}{Z}_\ell(t)|P_\ell],\ E[\underset{\sim}{Z}_\ell(t)|P_\ell]) \\ &= E(\text{diag}(\underset{\sim}{p}_i(s)) - \underset{\sim}{p}_i(s)\underset{\sim}{p}_i'(s))P_\ell^{t-s} + EP_\ell^{s'}\underset{\sim}{e}_i\underset{\sim}{e}_i'P_\ell^t - (EP_\ell^{s'}\underset{\sim}{e}_i)(\underset{\sim}{e}_i'EP_\ell^t) \end{aligned}$$

But $\underset{\sim}{p}_i(s)\underset{\sim}{p}_i'(s)P_\ell^{t-s} = P_\ell^{s'}\underset{\sim}{e}_i\underset{\sim}{e}_i'P_\ell^sP_\ell^{t-s} = P_\ell^{s'}\underset{\sim}{e}_i\underset{\sim}{e}_i'P_\ell^t$ and so

$$C(\underset{\sim}{Z}_\ell(s),\underset{\sim}{Z}_\ell(t)) = E\ \text{diag}(\underset{\sim}{p}_i(s))P_\ell^{t-s} - EP_\ell^{s'}\underset{\sim}{e}_i\underset{\sim}{e}_i'EP_\ell^t$$

Noting $\sum_{i=1}^{k} m_i\underset{\sim}{e}_i\underset{\sim}{e}_i' = \text{diag}(\underset{\sim}{m})$ and applying (38) we obtain

$$C(\underset{\sim}{N}(s),\underset{\sim}{N}(t)) = \sum_{i=1}^{k} m_i\, E\,\text{diag}(\underset{\sim}{p}_i(s))P_1^{t-s} - EP_1^{s'}\text{diag}(\underset{\sim}{m})EP_1^t. \tag{39}$$

The continuous time analogue to the last model is model 3c1 with $Q_\ell = QA_\ell(t)$. The derivation of the covariance is essentially the same and defining $\underset{\sim}{p}_i'(s)$ as the ith row of $e^{QA_1(s)}$ we have

$$C(\underset{\sim}{N}(s),\underset{\sim}{N}(t)) = \sum_{i=1}^{k} m_i E\text{diag}(\underset{\sim}{p}_i(s))e^{Q(A_1(t)-A_1(s))} - Ee^{Q'A_1(s)}\text{diag}(\underset{\sim}{m})Ee^{QA_1(t)} \tag{40}$$

For model 3c2, where Q is also random, define E_i to mean expectation with respect to the distribution of (X_1,Q_1,A_1) when animal 1 is initially in area i, that is, $P(X_1(t)=i)=1$.

$$C(\underset{\sim}{N}(s),\underset{\sim}{N}(t)) = \sum_{i=1}^{k} m_iE_i\text{diag}(\underset{\sim}{p}_i(s))e^{Q_1(A_1(t)-A_1(s))} - R'(s)\text{diag}(\underset{\sim}{m})R(t), \tag{41}$$

$\underset{\sim}{p}_i'(s)$ now being the ith row of $e^{Q_1A_1(s)}$ and $E_i\underset{\sim}{p}_i'(s)$ the ith row of $R(s)$.

The general model 4c involves another level of complication. Let E_i denote expectation with respect to the distribution of (X_1, Q_1, A_1, σ). Now (41) shows that the conditional covariance

$$C\left(\underset{\sim}{N}(s), \underset{\sim}{N}(t) \mid \sigma\right) = \sum_{i=1}^{k} m_i E_i\left[\operatorname{diag}\left(\underset{\sim}{p}_i(s)\right) e^{Q_1\left(A_1(t) - A_1(s)\right)} \mid \sigma\right] - R'(s) \operatorname{diag}(\underset{\sim}{m}) R(t),$$

where the ith row of $R(s)$ is $E_i[\underset{\sim}{p}_i'(s) \mid \sigma]$. Also $E[\underset{\sim}{M}(t) \mid \sigma] = R'(t)\underset{\sim}{m}$. Thus

$$C\left(\underset{\sim}{N}(s), \underset{\sim}{N}(t)\right) = \sum_{i=1}^{k} m_i E_i \operatorname{diag}\left(\underset{\sim}{p}_i(s)\right) e^{Q_1\left(A_1(t) - A_1(s)\right)} - ER'(s)\{\operatorname{diag}(\underset{\sim}{m}) - \underset{\sim}{m}\underset{\sim}{m}'\}R(t)$$
$$- ER'(s)\underset{\sim}{m}\underset{\sim}{m}'ER(t). \qquad (42)$$

Asymptotic distributions

For all models except 4c $\underset{\sim}{N}(t) = \sum_{\ell=1}^{n} \underset{\sim}{Z}_\ell(t)$, where the $\underset{\sim}{Z}_\ell = \{\underset{\sim}{Z}_\ell(t), t \geq 0\}$ are independent. If $X_\ell(0) \overset{a.s.}{=} X_{\ell'}(0)$ then $\underset{\sim}{Z}_\ell$ and $\underset{\sim}{Z}_{\ell'}$ are identically distributed. Thus if $n \to \infty$ and $m_i/n \to \alpha_i$, $i = 1,..,k$, the joint distribution of $\frac{1}{\sqrt{n}}\left(\underset{\sim}{N}(t_1) - E\underset{\sim}{N}(t_1), .., \underset{\sim}{N}(t_r) - E\underset{\sim}{N}(t_r)\right)$ is asymptotically multivariate normal with means $\underset{\sim}{0}$ and variances and covariances obtained by substituting α_i for m_i in the variance and covariance formulae for the $\underset{\sim}{N}(t_i)$.

In model 4c a normal asymptotic distribution does not necessarily pertain. Consider $n \to \infty$, $m_i/n \to \alpha_i$, $i = 1,..,k$. Conditional on σ, whose distribution does not depend on $m_1, .., m_k$, $\frac{1}{n}\underset{\sim}{N}(t) \overset{a.s.}{\to} R'(t)\underset{\sim}{\alpha}$. Thus the joint distribution of $\frac{1}{n}\left(\underset{\sim}{N}(t_1), .., \underset{\sim}{N}(t_r)\right)$ converges to the joint distribution of $\left(R'(t_1)\underset{\sim}{\alpha}, .., R'(t_r)\underset{\sim}{\alpha}\right)$. However if we do not condition on $\underset{\sim}{X}(0)$ 2. and 3. of 4c are replaced by

2'. $\left(X_1(0), Q_1, a_1, S_1\right), .., \left(X_n(0), Q_n, a_n, S_n\right)$ are i.i.d. Furthermore S_ℓ and $\left(X_\ell(0), Q_\ell, a_\ell, S_\ell\right)$ are independent and the distribution of $(X_\ell(0),Q_\ell,a_\ell)$ takes the form

$$\mathcal{P}(X_\ell(0)=j,(Q_\ell,a_\ell)\in D) = \int_D \gamma_j d\pi$$

where π is a measure on the space of all (a,Q) such that a is a non negative $\mathcal{B}^3$ measurable function and Q is a rates matrix with $\mathcal{J}(Q) = \underset{\sim}{1}(\gamma_1,..,\gamma_k)$.

For this specification $\mathcal{P}(X_\ell(t)=j) = \mathcal{E}\sum_i \int \gamma_i\tilde{p}_{ij}(t)d\pi$ where $\tilde{p}_{ij}(t)$ = the (i,j) element of $\exp\{Q\int_0^t a(u,S_\ell(u),\sigma(u))du\}$. However $\sum_i \gamma_i\tilde{p}_{ij}(t) = \gamma_j$ and so $\mathcal{P}(X_\ell(t)=j) = \int\gamma_j d\pi = q_j$, say. In a similar manner it can be seen that $X_1(t),..,X_n(t)$ are i.i.d. Thus $\mathcal{E}\underset{\sim}{N}(t) = n\underset{\sim}{q}$ and $\mathcal{V}\underset{\sim}{N}(t) = n(\mathrm{diag}(\underset{\sim}{q})-\underset{\sim}{q}\underset{\sim}{q}')$. $\underset{\sim}{N}(t)$ is a multinomial and $\frac{1}{\sqrt{n}}(\underset{\sim}{N}(t)-nq)$ is asymptotically multivariate normal. However the stochastic processes $X_1,...,X_n$ are not necessarily independent. By analogy with earlier results

$$\mathcal{C}(\underset{\sim}{Z}_\ell(s),\underset{\sim}{Z}_\ell(t)\,|\,\sigma) = \mathcal{E}[\mathrm{diag}(\underset{\sim}{\gamma}_1)e^{Q_1(A_1(t)-A_1(s))}|\sigma]-\underset{\sim}{q}\underset{\sim}{q}'$$

where $\underset{\sim}{1}\gamma_1' = \mathcal{J}(Q_1)$. Thus

$$\mathcal{C}(\underset{\sim}{N}(s),\underset{\sim}{N}(t)\,|\,\sigma) = n\{\mathcal{E}[\mathrm{diag}(\gamma_1)e^{Q_1(A_1(t)-A_1(s))}|\sigma] - qq'\} \tag{43}$$

$\mathcal{C}(\underset{\sim}{N}(s),\underset{\sim}{N}(t))$ is simply the expected value of (43). Conditional on σ, $\frac{1}{\sqrt{n}}(\underset{\sim}{N}(t_1)-n\underset{\sim}{q},..,\underset{\sim}{N}(t_r)-n\underset{\sim}{q})$ is asymptotically multivariate normal with means $\underset{\sim}{0}$ and covariance given by (43) with $n = 1$. Thus the asymptotic distribution of $\frac{1}{\sqrt{n}}(\underset{\sim}{N}(t_1)-n\underset{\sim}{q}_1,..,\underset{\sim}{N}(t_r)-n\underset{\sim}{q})$ is a mixture of multivariate normal distributions.

3. GENERAL SPACE

The purpose of this chapter is to present the ideas of chapter 2 in a general setting and to develop a framework for thinking about animal movement processes.

Independence between individual animals played an important role in the development of the discrete space theory; however it was not always complete independence but independence conditional on the environment. It was always assumed that the animals do not directly affect each other's movements but that dependence is induced by an external factor, namely a random environment. This situation may be thought of as a low degree of interaction between individuals, which I shall call simply a low degree of interaction.

In many situations the environment will be subject to fluctuations affecting movements and unless these contain no stochastic element they will introduce dependence between individual animal's movements. If there is a low degree of interaction, or independence conditional on the environment, then there are basically two problems:

1. What are the conditional distributions of individual animal's movements, given the environment?
2. How is the environment varying?

The first question is the more basic since it indicates the degree of importance of the second. In most situations at best incomplete answers will be obtained. A good answer to the first question achieves a great deal because it really means that the animals' movements are understood. The consequences of different environmental regimes can be explored and estimation of past events from measurement of the environment is possible. On the other hand a complete answer to the second is useless in the absence of any knowledge of the first.

The fact that these questions may be difficult to answer does not mean there is no point in developing a general theory, on the contrary, it emphasises the need for one. One needs to be aware of the possibilities in various situations and in particular to isolate conclusions which follow from weak assumptions. A good general theory should then enable one to decide what kinds of information one needs to collect for a specific purpose.

The general theory given here aims at specifying the properties expected of a probability space suitable for describing the movements of animals in a population where the individuals are independent conditional on the environment. Various simplifications of the general structure are examined with a view to determining the extent to which they are distinguishable. The analogous functions to the $\underset{\sim}{N}(t)$ of chapter 2 are defined and form the major part of the discussion. It is shown how models of the general form may be constructed from Markov processes. The chapter concludes with a discussion of a situation in which the general theory of this chapter applies as an approximation.

3.1 The general structural model.

Consider a probability space (Ω, A, P) on which there are defined n stochastic processes $X_\ell = \{X_\ell(t), t \in T\}$, $\ell = 1,\dots,n$, corresponding to n animals. The set of times, T, is either $\{0,1,\dots\}$ or $[0,\infty)$. The random variable $X_\ell(t)$ is the position of the ℓth animal at time t and it takes its values in the measurable space $(\mathcal{X},\mathcal{G})$. This means $X_\ell(t,\omega) \in \mathcal{X}$, for every $\omega \in \Omega$, and for every $G \in \mathcal{G}$, $\{\omega | X_\ell(t,\omega) \in G\} \in \mathcal{A}$. One says $X_\ell(t)$ is $\mathcal{A}/\mathcal{G}$ measurable. The space $\mathcal{X}$ is the place where the animals live and $\mathcal{G}$ is an appropriate σ-algebra of subsets of $\mathcal{X}$.

The set $\mathcal{X}^T$ is the set of all functions from T into $\mathcal{X}$. Thus for each ω, the animals' paths X_ℓ are elements of $\mathcal{X}^T$. In order to ask questions about X_ℓ more general than "$X_\ell(t) \in G$?", for fixed t, the σ-algebra $\mathcal{G}^T$ is needed. This is defined as the smallest σ-algebra of subsets of $\mathcal{X}^T$ containing all sets of the form $\{x \in \mathcal{X}^T | x(t) \in G\}$, $t \in T$, $G \in \mathcal{G}$. The sets $\{X_\ell \in G\}$ belong to $\mathcal{A}$ for $G \in \mathcal{G}^T$.

All animals are different. Let us call those properties of animals, morphological, physiological and behavioural which characterise the differences between individuals the "phenotype". Only those properties of the phenotype which affect movements are needed here, so that for modelling purposes, the phenotype could be represented as a much simpler object. On the other hand occasions arise when, for the purpose of discussing movements, other properties of an animal, such as where it establishes its home, are conveniently included as part of the phenotype. Suppose there is a space Z consisting of all possible phenotypes for the animals under consideration. Let $\mathcal{Z}$ be a σ-algebra of subsets of Z.

Defined on $(\Omega,\mathcal{A},P)$ are $(Z,\mathcal{Z})$ valued random variables $z_1,\ldots z_n$, which are the phenotypes of the n

animals. The matrices Q_ℓ, $\ell = 1,\ldots,n$, of discrete space model 3c2, were thought of as functions of the phenotype, or possibly an animal's past history. These were i.i.d. in the setting up of 3c2, but as model 3c2 conditioned on $\underset{\sim}{X}(0)$, the equality of distribution was destroyed although the independence remained. To allow for such conditional models in the present discussion assume that the animals are recognised as belonging to r subpopulations with m_i individuals in the ith subpopulation. For animals belonging to the same subpopulation, the z_ℓ are identically distributed. The entire set $z_1,\ldots,z_n$ will be assumed independent.

Other situations might arise in which different subpopulations are recognised for the purpose of assigning phenotypic distributions, for example some property of the home distribution may be known or the individuals may be of several origins.

A random variable σ defined on $(\Omega,\mathcal{A},\mathcal{P})$ and taking values in the space $(E,\mathcal{S})$ represents the environment. Generally σ can be thought of as a random function $\sigma = \{\sigma(t),\ t \in T\}$, where $\sigma(t)$ represents the "state of the environment at time t", but for the purposes of this section such representation is not needed. It is assumed that σ and $(z_1,\ldots,z_n)$ are independent.

The animals have a low degree of interaction so that any particular animal's movements are not affected by the movements of the rest of the population. From a behavioural point of view this statement applies conditional on the phenotypes of the animals involved, that is, it applies given any particular set of real animals of the kind under consideration. The idea is expressed mathematically as:

S1. The processes $X_1,\ldots,X_n$ are independent conditional on σ and $z_1,\ldots,z_n$.

Since σ and $z_1,\ldots,z_n$ are not necessarily any of the commonly encountered kinds of random variables, to avoid any possible ambiguity, I shall state S1 in more basic mathematical terms. Let $\mathcal{S}^*$ be the σ-subalgebra of $\mathcal{A}$ generated by σ , that is, $\mathcal{S}^*$ is the smallest σ-algebra containing all sets of the form $\{\omega \in \Omega | \sigma(\omega) \in S\}$ for $S \in \mathcal{S}$. If $\mathcal{C}$ and $\mathcal{D}$ are two σ-algebras define $\mathcal{C} \vee \mathcal{D}$ as the smallest σ-algebra containing both $\mathcal{C}$ and $\mathcal{D}$. Let Z^*_ℓ be the σ-algebra generated by z_ℓ and define $\mathcal{Z}^* = Z^*_1 \vee Z^*_2 \vee \ldots \vee Z^*_n$. Then S1. means

$$P\left(\bigcap_{\ell=1}^{n} \{X_\ell \in G_\ell\} | \mathcal{S}^* \vee \mathcal{Z}^*\right) = \prod_{\ell=1}^{n} P(X_\ell \in G_\ell | \mathcal{S}^* \vee \mathcal{Z}^*)$$

whenever $G_1,\ldots,G_n$ belong to $\mathcal{G}^T$. The equality means almost sure equality, here, as in later relations between conditional expectations.

In keeping with the idea of little interaction between individuals, the movements of any animal should not be affected by the phenotypes of the rest of the animals. This leads to

S2. $P(X_\ell \in G | \sigma, z_1,\ldots,z_n) = P(X_\ell \in G | \sigma, z_\ell)$ whenever $G \in \mathcal{G}^T$; $\ell = 1,\ldots,n$.

Finally the conditional distribution of X_ℓ given σ and z_ℓ should be expressible as a function of σ and z_ℓ , independent of ℓ .

S3. There is a function $\pi : G^T \times E \times Z \to [0,1]$ such that

$P(X_\ell \in G|\sigma, z_\ell) = \pi(G;\sigma,z_\ell)$, $G \in G^T$; $\ell = 1,\ldots,n$.

It is assumed that π is a regular conditional distribution so that for σ, z_ℓ fixed, π is a probability measure. This is not just a mathematical convenience but very much part of the model since it is natural to admit a probability distribution for $X_1,\ldots,X_n$ in situations where the environment and phenotypes are determined. The model described here then comes about by allowing σ and $z_1,\ldots,z_n$ to vary stochastically so that the distribution of $(X_1,\ldots,X_n)$ is a mixture of distributions.

The specifications S1. - 3. can be summarised conveniently as

S1′. $P(\bigcap_{\ell=1}^{n} \{X_\ell \in G_\ell\}|\sigma,z_1,\ldots,z_n) = \prod_{\ell=1}^{n} \pi(G_\ell;\sigma,z_\ell)$, for $G_1,\ldots,G_n \in G^T$. As a result of S1′. it is clear that

$$P(\bigcap_{\ell=1}^{n} \{X_\ell \in G_\ell\}|\sigma) = E[\prod_{\ell=1}^{n} \pi(G_\ell;\sigma,z_\ell)|\sigma].$$

However $z_1,\ldots,z_n$ are independent and jointly independent of σ and so the R.H.S. is

$$\prod_{\ell=1}^{n} E[\pi(G_\ell;\sigma,z_\ell)|\sigma] = \prod_{\ell=1}^{n} P(X_\ell \in G_\ell|\sigma)$$

Thus $X_1,\ldots,X_n$ are conditionally independent given σ . The independence, conditional on the environment, that was postulated for individual animals of fixed phenotype, is preserved upon averaging over the possible phenotypes. Let ρ_i be the distribution of z_ℓ, $\rho_i(W) = P(z_\ell \in W)$ for $W \in Z$, when animal ℓ belongs to subpopulation i and define

$$\pi^*(G;\sigma_0,i) = \int_Z \pi(G;\sigma_0,z)\, d\rho_i(z) \quad , \text{ for } \sigma_0 \in E,$$

then $\pi^*(.;\sigma,i)$ is a regular conditional distribution for X_ℓ given σ. This can be summarised as:

Theorem 1. The stochastic processes $X_1,\ldots,X_n$ are independent conditional on σ and the conditional distribution of X_ℓ is $\pi^*(.;\sigma,i)$ when animal ℓ belongs to subpopulation i.

The theorem shows that the distribution of $(X_1,\ldots,X_n)$ agrees with the situation where there are no differences between the distributions of individual's movement processes conditional on their phenotypes, for animals belonging to the same subpopulation. However the distributions π and π^* may be quite different in nature, for example, Theorem 4 of Chapter 2 gives a situation where π is always Markov but π^* is not Markov. The proof of this theorem shows that there is a consistent estimator of $\pi(.;\sigma,z_\ell)$ based on a single realisation of $\{X_\ell(t),t \in T\}$. However one would rarely expect to find a reasonable consistent estimator of π^* based on $\{X_\ell(t),\ t \in T\}$.

3.2 The number of animals in an area.

In many practical situations the easiest quantities to observe are the numbers of animals in different places and if one is interested in an effect that the animals' presence has then these are the natural quantities. This section looks at the properties of the numbers of animals in different places in the general setting of the previous section.

Define

$$N_t(G) = \sum_{\ell=1}^{n} I_G(X_\ell(t)) \text{ , for } G \in \mathcal{G} . \qquad (1)$$

$N_t(G)$ is the number of animals in the set G at time t and is a random variable on $(\Omega, \mathcal{A}, \mathcal{P})$. On the other hand if ω is fixed $N_t(.,\omega)$ is a measure on $(\mathcal{X}, \mathcal{G})$. Hence N_t is a random measure.

In order to talk about the distribution of a random measure we need the measurable space $(U, \mathcal{U})$ where U is the set of all finite measures on $(\mathcal{X}, \mathcal{G})$ and $\mathcal{U}$ is the σ-algebra of subsets of U generated by all sets of the form

$$\{M \in U | M(G_1) \in B_1, \ldots, M(G_k) \in B_k\} \qquad (2)$$

where $G_1, \ldots, G_k \in \mathcal{G}$ and $B_1, \ldots, B_k \in \mathcal{B}^1$, with $\mathcal{B}^1$ being the Borel sets of the real line R^1.

The well known monotone class theorem is also needed. It is stated as theorem 2.

Theorem 2. Let $\mathcal{F}$ be a set of subsets of some set Θ which is closed under finite intersections. Suppose $\mathcal{H}$ is the smallest set containing Θ which satisfies

1. $\Theta \in \mathcal{H}$.
2. If $A, B \in \mathcal{H}$, $A \subset B$ then $B - A \in \mathcal{H}$.
3. If $\{A_n\}$ is an increasing sequence of elements of $\mathcal{H}$ then $\cup A_n \in \mathcal{H}$.

Then $\mathcal{H}$ is the σ-algebra generated by $\mathcal{F}$.

A corollary to theorem 2 is that if $\mathcal{F}$ is closed under finite intersections and $\mathcal{X}$ is the σ-algebra generated by $\mathcal{F}$ then two probability measures on $\mathcal{H}$, which agree on $\mathcal{F}$, also agree on $\mathcal{H}$. This corollary shows that any probability measure on $(U,\mathcal{U})$ is determined by its values on the sets (2) since these sets are closed under finite intersections.

Because $N_t(G)$ is $\mathcal{A}$ measurable the set $\{\omega | N_t(.,\omega) \in V\}$ belongs to $\mathcal{A}$ when V is any set of form (2) and it follows that N_t is $\mathcal{A}/\mathcal{U}$ measurable. Thus the equation

$$\mu(V) = P(N_t \in V) , \text{ for } V \in \mathcal{U},$$

defines a measure μ on $(U,\mathcal{U})$ called the distribution of N_t. Note that the distribution of N_t is determined by its values on the sets (2), or equivalently, it is determined by the distribution of $(N_t(G_1),\ldots,N_t(G_k))$ for every finite set $G_1,\ldots,G_k \in \mathcal{G}$. However if $G_1,\ldots,G_k$ are any k' sets belonging to $\mathcal{G}$ then there are disjoint sets $G'_1,\ldots,G'_{k'}, \in \mathcal{G}$, $k' < \infty$, such that each

$$G_i = \bigcup_{G'_j \subset G_i} G'_j .$$

which means

$$N_t(G_i) = \sum_{G'_j \subset G_i} N(G'_j)$$

so that the distribution of $(N_t(G_1),\ldots,N_t(G_k))$ is determined by the distribution of $(N_t(G'_1),\ldots,N_t(G'_{k'}))$. Thus the distribution of N_t is determined by the distribution of $(N_t(G_1),\ldots,N_t(G_k))$ for $G_1,\ldots,G_k \in \mathcal{G}$, $G_i \cap G_j = \emptyset$, if $i \neq j$, $\cup G_j = X$. Such $G_1,\ldots,G_k$ will be referred to simply as a partition of X.

The simplest situation for the distribution of N_t arises when there is only one subpopulation and σ is almost surely constant, that is σ is non random, but if expressed as a function of time it may vary deterministically in time. Let $G_1,\ldots,G_k$ be a partition of X, then in this simplest situation $X_1(t),\ldots,X_n(t)$ are i.i.d. and so $(N_t(G_1),\ldots,N_t(G_k))$ has a multinomial distribution with parameters $(n,\pi_t(G_1),\ldots,\pi_t(G_k))$, where π_t is the common distribution of $X_\ell(t)$. Thus N_t is in a sense a generalisation of a multinomial. It seems appropriate to call this generalisation a continuous multinomial.

Definition. Let M be a random measure then M is called a continuous multinomial (CM) if for every partition $G_1,\ldots,G_k$ of X, $(M(G_1),\ldots,M(G_k))$ is a multinomial. M is said to have a CM distribution.

The following theorem shows that every CM distribution can be expressed as the distribution of a random measure M^* of the form

$$M^*(G) = \sum_{j=1}^{n} I_G(Y_j), \quad G \in \mathcal{G}, \tag{2}$$

where $Y_1,\ldots,Y_n$ are i.i.d. $(X,\mathcal{G})$ valued random variables. The random measure M^* is of the same form as N_t and so a CM distribution is no more general than is needed here for studying animal movements.

Theorem 3. Let M be a continuous multinomial. There is a positive integer n and a probability measure π on $(X,\mathcal{G})$ such that M has the same distribution as M^*, given by (2), where $Y_1,\dots,Y_n$ are i.i.d. with common distribution π .

Proof. Define $n = EM(X)$. Let $G_1,\dots,G_n$ be a partition of X , then $(M(G_1),\dots,M(G_k))$ has a multinomial distribution and so $\sum_{j=1}^{k} M(G_j)$ is almost surely constant. However $\sum_{j=1}^{k} M(G_j) = M(X)$ so that $n \overset{a.s.}{=} \sum_{j=1}^{k} M(G_j)$. Thus the multinomial $(M(G_1),\dots,M(G_k))$ has parameters $(n, EM(G_1)/n,\dots,EM(G_k)/n)$. For $G \in G$ define $\pi(G) = EM(G)/n$. Now $0 \leqslant \pi(G) \leqslant 1$, $\pi(X) = 1$ and if $\{G_n\}$ is a disjoint sequence in G then the random measure property of M and the monotone convergence theorem show that

$$n\sum_{j=1}^{\infty}\pi(G_j) = \sum_{j=1}^{\infty} EM(G_j) = E\sum_{j=1}^{\infty} M(G_j) = EM(\bigcup_{j=1}^{\infty} G_j) = n\pi(\bigcup_{j=1}^{\infty} G_j)$$

and so π is a probability measure on (X,G). A set of n i.i.d. random variables $Y_1,\dots,Y_n$, with common distribution π , can be defined on the space (X^n,G^n,π^n); and if M^* is defined by (2) it follows that $(M(G_1),\dots,M(G_k))$ and $(M^*(G_1),\dots,M^*(G_k))$ have the same distribution for every partition $G_1,\dots,G_k$ of X. Hence M and M^* have the same distribution.

In the case where there are r subpopulations but σ remains almost surely constant then N_t can be written

$$N_t(G) = \sum_{j=1}^{r} N_{jt}(G), \quad G \in G ,$$

where the N_{jt} are independent continuous multinomials.

Definition A random measure M is a continuous convolution of multinomials (CCM) if for every partition $G_1,\dots,G_k$ of X , $(M(G_1),\dots,M(G_k))$ is a convolution of multinomials.

It is quite clear that N_t is a CCM, but it is not obvious that every CCM distribution can be represented as the distribution of a sum of independent continuous multinomials. The next theorem proves that every CCM distribution does have such a representation and furthermore it is in a certain sense unique.

Theorem 4. If M is a CCM then there is a unique set of positive integers $m_1, \ldots, m_r$, and corresponding to these, r distinct measures, $\pi_1, \ldots, \pi_r$ on (X, G), such that M has the same distribution as $\sum_{j=1}^{r} M_j$ where $M_1, \ldots, M_r$ are independent continuous multinomials with parameters (m_i, π_i), $i = 1, \ldots, r$.

Proof. First we need a lemma.

Lemma 1. A convolution of multinomials has a unique, apart from order, representation as the convolution of a minimal number of multinomials. Furthermore this minimal representation is the unique representation for which parameters $(m_1, \underset{\sim}{p}_1), \ldots, (m_r, \underset{\sim}{p}_r)$ of the constituent multinomials, have $m_i > 0$ and $\underset{\sim}{p}_i \neq \underset{\sim}{p}_j$ for $i \neq j$.

Proof of lemma 1. The generating function for a convolution of multinomials has the form

$$\prod_{i=1}^{c} (\underset{\sim}{s}' \underset{\sim}{q}_i)^{n_i} \quad , \quad n_j > 0.$$

This is a polynomial in the several variables $s_1, \ldots, s_k$, over the field of real numbers and so has a unique, apart from order, representation as the product of suitably normalised irreducible polynomials. The linear polynomials $\underset{\sim}{s}' \underset{\sim}{q}_i$ are irreducible and the normalising condition $\underset{\sim}{1}' \underset{\sim}{q}_j = 1$ ensures the uniqueness of the factors.

Thus collecting identical factors a unique minimal representation is obtained:

$$\prod_{i=1}^{r} (\underset{\sim}{s}' \underset{\sim}{p}_i)^{m_i}, \quad \text{where} \quad m_i > 0 \quad \text{and} \quad \underset{\sim}{p}_i \neq \underset{\sim}{p}_j ; \text{ for } i \neq j.$$

In the sequel the parameters $(m_1, \underset{\sim}{p}_1), \ldots, (m_r, \underset{\sim}{p}_r)$ are referred to as the minimal parameters.

Now define $n = \mathcal{E}M(\mathcal{X})$. Let $G_1, \ldots, G_k$ be a partition of $\mathcal{X}$, then $(M(G_1), \ldots, M(G_k))$ is a convolution of multinomials. If the minimal parameters are $(m_1, \underset{\sim}{p}_1), \ldots, (m_r, \underset{\sim}{p}_r)$ then $r \leqslant n$ and so there is a maximum possible r. If this maximum is achieved for a partition $G_1, \ldots, G_k$ then the partition is called maximal.

Suppose partitions $G_1, \ldots, G_k$ and $H_1, \ldots, H_\ell$ are such that

$$G_i = \bigcup_{H_j \subset G_i} H_j \quad , \; i=1, \ldots, k$$

then $H_1, \ldots, H_\ell$ is said to refine $G_1, \ldots, G_k$.
We need a second lemma.

<u>Lemma 2</u>. Suppose the partition $H_1, \ldots, H_\ell$ refines the maximal partition $G_1, \ldots, G_k$, then $H_1, \ldots, H_\ell$ is maximal and furthermore, if $(m_1, \underset{\sim}{p}_1), \ldots, (m_r, \underset{\sim}{p}_r)$ and $(n_1, \underset{\sim}{q}_1), \ldots, (n_t, \underset{\sim}{q}_t)$ are the minimal parameters for $G_1, \ldots, G_k$ and $H_1, \ldots, H_\ell$ respectively, then there is a unique reordering of the $(n_i, \underset{\sim}{q}_i)$ such that

$$m_i = n_i \quad \text{and} \quad p_{ij} = \sum_{H_c \subset G_j} q_{ic} \quad , \; i = 1, \ldots, r, j = 1, \ldots, k, \tag{3}$$

where $\underset{\sim}{p}_i = (p_{i1}, \ldots, p_{ik})$, $\underset{\sim}{q}_i = (q_{i1}, \ldots, q_{i\ell})$. When this is so the parameters $(n_1, \underset{\sim}{q}_1), \ldots, (n_t, \underset{\sim}{q}_t)$ are said to be ordered relative to $(m_1, \underset{\sim}{p}_1), \ldots, (m_1, \underset{\sim}{p}_r)$. The relationship of one set of parameters being ordered relative to another set is transitive for minimal parameters of maximal partitions.

Proof of Lemma 2. Consider the case where $H_1, \ldots, H_\ell$ has one more set than $G_1, \ldots, G_k$. For definiteness assume $G_i = H_i$, $i \leqslant k-1$, $H_k \cup H_{k+1} = G_k$. Since $M(H_k) + M(H_{k+1}) = M(G_k)$ we have

$$\prod_{i=1}^{r} (\underset{\sim}{s}' \underset{\sim}{p}_i)^{m_i} = \prod_{i=1}^{t} \Big(\sum_{j=1}^{k-1} s_j q_{ij} + s_k (q_{ik} + q_{i\,k+1}) \Big)^{n_i} .$$

Lemma 1 applies giving $r = t$ and a unique reordering of the $(n_i, \underset{\sim}{q}_i)$ such that $m_i = n_i$ and

$$\underset{\sim}{p}_i = (q_{i1}, \ldots, q_{i\,r-1}, q_{ir} + q_{i\,r+1}) .$$

The general case is proved by induction as follows

Let $K_1, \ldots, K_p$ refine $H_1, \ldots, H_\ell$ and contain just one more set than $H_1, \ldots, H_\ell$, where the latter partition refines $G_1, \ldots, G_k$ is maximal, and there is a unique ~~recording~~ REORDERING of the parameters for $H_1, \ldots, H_\ell$ so that (3) holds. Let $K_1, \ldots, K_p$ have minimal parameters $(\ell_1, \underset{\sim}{t}_1), \ldots, (\ell_u, \underset{\sim}{t}_u)$. Then the maximality of $H_1, \ldots, H_\ell$ implies the maximality of $K_1, \ldots, K_p$. Furthermore the orderings of the $(\ell_i, \underset{\sim}{t}_i)$ relative to the $(n_i, \underset{\sim}{q}_i)$, and the latter relative to the $(m_i, \underset{\sim}{p}_i)$, serve to construct an ordering of the $(\ell_i, \underset{\sim}{t}_i)$ relative to the $(m_i, \underset{\sim}{p}_i)$. However if there is a different ordering of the $(\ell_i, \underset{\sim}{t}_i)$ relative to the $(m_i, \underset{\sim}{p}_i)$, then this ordering, and the ordering of the $(\ell_i, \underset{\sim}{t}_i)$ relative to the $(n_i, \underset{\sim}{q}_i)$, serve to construct a different ordering of the $(n_i, \underset{\sim}{q}_i)$ relative to the $(m_i, \underset{\sim}{p}_i)$, which is a contradiction. This proves the lemma except for transitivity which is proved in the same way as the construction of a unique ordering between the $(\ell_i, \underset{\sim}{t}_i)$ and the $(m_i, \underset{\sim}{p}_i)$ above.

Let $G_1, \ldots, G_k$ be a fixed maximal partition with minimal parameters $(m_1, \underset{\sim}{p}_1), \ldots, (m_r, \underset{\sim}{p}_r)$. Define the probability measure π_i on $A(G_1, \ldots, G_k)$, the σ-algebra generated by $G_1, \ldots, G_k$, by the relation $\pi_i(G_j) = p_{ij}$. Suppose $H_1, \ldots, H_\ell$ refines $G_1, \ldots, G_k$ and has minimal parameters $(m_1, \underset{\sim}{q}_1), \ldots, (m_r, \underset{\sim}{q}_r)$ where the $(m_i, \underset{\sim}{q}_i)$ are ordered relative to the $(m_i, \underset{\sim}{p}_i)$. The π_i can be extended to $A(H_1, \ldots, H_k)$ by setting $\pi_i(H_j) = q_{ij}$. Relation (3) guarantees that this extension of the probability measures π_i is consistent.

Now suppose $K_1, \ldots, K_p$, which refines $H_1, \ldots, H_\ell$, has minimal parameters $(m_1, \underset{\sim}{t}_1), \ldots, (m_r, \underset{\sim}{t}_r)$. From lemma 2, if both the $(m_i, \underset{\sim}{q}_i)$ and the $(m_i, \underset{\sim}{t}_i)$ are ordered relative to the $(m_i, \underset{\sim}{p}_i)$, then the $(m_i, \underset{\sim}{t}_i)$ are ordered relative to the $(m_i, \underset{\sim}{q}_i)$. Now applying relations (3) to the $(m_i, \underset{\sim}{q}_i)$ and $(m_i, \underset{\sim}{t}_i)$, it follows that the extensions of the π_i to $A(K_1, \ldots, K_p)$ and $A(H_1, \ldots, H_\ell)$ are consistent. Next if $K_1, \ldots, K_p$ and $H_1, \ldots, H_\ell$ are arbitrary partitions refining $G_1, \ldots, G_k$ then consideration of a partition $L_1, \ldots, L_t$ refining both $K_1, \ldots, K_p$ and $H_1, \ldots, H_\ell$ shows that the extensions of π_i to $A(K_1, \ldots, K_p)$ and $A(H_1, \ldots, H_\ell)$ are consistent because both are consistent with the extension to $A(L_1, \ldots, L_t)$. Thus the π_i are defined consistently as probability measures on every finite σ-subalgebra of G, which shows they are finitely additive probability measures on G.

To prove they are countably additive all we need show is that they are continuous at $\emptyset$.

$EM(G)$, $G \in \mathcal{G}$, is a measure, as in the proof of theorem 3 , and $\pi_i(G) \leqslant EM(G) \leqslant n$. If $\{G_m\}$ is a decreasing sequence of sets in $\mathcal{G}$ with empty intersection then $\lim_{m \to \infty} EM(G_m) = 0$ implying $\lim_{m \to \infty} \pi_i(G_m) = 0$, which is the required continuity at $\emptyset$.

On the space $(\mathcal{X}^n, \mathcal{G}^n, \pi_1^{m_1} \times \ldots \times \pi_r^{m_r})$ there are independent random variables $Y_1, \ldots, Y_n$ such that m_i of the Y_j have distribution π_i .

Define

$$M_i(G) = \sum I_G(Y_j) \ , \quad G \in \mathcal{G}$$

where the summation is over those j for which Y_j has distribution π_i . The M_i are independent continuous multinomials and $\sum M_i$ has the same distribution as M.

Finally the unordered set $\{(m_1, \pi_1), \ldots, (m_r, \pi_r)\}$ $\pi_i \neq \pi_j$ for $i \neq j$ and $m_i > 0$, for which this representation is possible, is unique. Suppose there were another set $\{(m_1', \pi_1'), \ldots, (m_t', \pi_t')\}$ with these properties. Let $G_1, \ldots, G_k$ be a maximal partition then $(M(G_1), \ldots, M(G_k))$ has, as alternative sets of parameters, the $(m_i, \pi_i(G_1), \ldots, \pi_i(G_k))$ and the $(m_i', \pi_i'(G_1), \ldots, \pi_i'(G_k))$. Lemma 1 shows that $r = t$ and there is a unique reordering of the (m_i', π_i') so that $m_i = m_i'$ and $\pi_i'(G_j) = \pi_i(G_j)$. Lemma 2 then shows that this reordering implies $\pi_i = \pi_i'$ on any partition refining $G_1, \ldots, G_n$. Hence $\pi_i = \pi_i'$, and the two sets $\{(m_1, \pi_1), \ldots, (m_r, \pi_r)\}$ and $\{(m_1', \pi_1'), \ldots, (m_r', \pi_r')\}$ are equal.

Theorems 3 and 4 eliminate the possibility of any ambiguity concerning continuous multinomials and continuous convolutions of multinomials by showing that what they appear to be, when examined on partitions, is what they are. In addition theorem 4 shows that, when σ is constant a.s., then the distribution of N_t determines the distributions of the $X_\ell(t)$, and also, if these distributions are distinct, for animals belonging to different subpopulations, the numbers of animals in the subpopulation are determined. In particular the distribution of N_t determines whether the $X_\ell(t)$ are i.i.d.

The final possibility that can arise for the distribution of N_t is a mixture of CCMs. Consider the general case when σ is not necessarily constant a.s. Conditional on σ N_t is a CCM with $E[N_t(\mathcal{X})\,|\,\sigma] = n$. If μ_t^σ is the conditional distribution for N_t, given σ, then the distribution of N_t is ν_t defined by

$$\nu_t(V) = E\mu_t^\sigma(V), \quad V \in \mathcal{U}.$$

In this case N_t, and its measure ν_t, are referred to as mixtures of continuous convolutions of multinomials or in short MCCMs.

Generating functionals.

The random measure N_t is a counting process in the sense of Moyal's "general theory of stochastic population processes" (1962). In Moyal's theory a counting process M is defined in terms of $(\mathcal{X},\mathcal{G})$ valued random variables $Y_1,\ldots,Y_n$ as

$$M(G) = \sum_{\ell=1}^{n} I_G(Y_\ell).$$

Moyal considered the more general case where n is possibly random, but that is not needed here.

The probability generating functional of M is defined as

$$g[s] = \mathcal{E}s(Y_1)..s(Y_n) = \mathcal{E}\exp\{\int_X \log s(x)dM(x)\},$$

where s is a bounded measurable complex function on (X,G), log is the principal branch of the logarithm when $s(x) \neq 0$ and $\log 0 = -\infty$. Defining

$$s = \sum_{i=1}^{k} e^{t_i} I_{G_i},$$

for $G_1,..,G_k$ a partition of X, it is clear that g determines the joint moment generating function of $(M(G_1),..,M(G_k))$ and so determines the distribution of $(M(G_1),..,M(G_k))$ for every partition $G_1,..,G_k$. Hence g determines the distribution of M. On the other hand the distribution of M determines g because the complex or $-\infty$ valued random variable $\int \log s(x)dM(x)$ is a U measurable function of M. To see this note that $\int I_G(x)dM(x)$ is U measurable, $G \in G$, and the measurability of $\int \log s(x)dM(x)$ follows, in the usual way, from the linearity and continuity properties of integrals. This also shows that the generating functional of an arbitrary finite random measure can be defined without reference to a set of random variables $Y_1,..,Y_n$, although more restrictions on s are necessary.

Moyal shows that, for counting processes, the generating functional determines the symmetrised distribution of $(Y_1,..,Y_n)$. This symmetrised distribution is equivalent to the probabilities $P((Y_1,..,Y_n) \in G_s)$, for G_s a symmetric set in G^n, that is $G_s = \{(x_1,..,x_n) \mid (x_{j(1)},..,x_{j(n)}) \in G_s\}$ for every permutation $(j(1),..,j(n))$ of $(1,..,n)$. Moyal sets up a probability distribution for the unordered set $\{Y_1,..,Y_n\}$ which is equivalent to the symmetrised distribution of $(Y_1,..,Y_n)$. Furthermore he shows that there is a 1-1 correspondence between M and $\{Y_1,..,Y_n\}$; and the two classes of sets $\{M(G)=j\}$, $G \in \mathcal{G}$, $j = 0,1,...,$ and $\{(Y_1,..,Y_n) \in \mathcal{G}_s\}$, G_s a symmetric member of $\mathcal{G}^n$, generate the same σ-subalgebra in the underlying probability space. Thus studying M is essentially equivalent to studying the unordered set $\{Y_1,..,Y_n\}$. Little explicit use is made of this fact but it is implicit in the proof of lemma 1 of chapter 2 and hence theorem 7 of this chapter.

To derive the generating functional of N_t let π^σ_{ti} denote the σ conditional distribution of $X_\ell(t)$, when animal ℓ belongs to subpopulation i. If σ is constant a.s., omit the superscript and if there is only one subpopulation omit the subscript i. The generating functional for N_t in the single subpopulation and constant σ situation is

$$\left(\int s(x)\,d\pi_t(x)\right)^n; \tag{4}$$

for σ constant and several subpopulations

$$\prod_{i=1}^{r} \left(\int s(x)\,d\pi_{ti}(x)\right)^{m_i}; \tag{5}$$

and in general

$$\int \prod_{i=1}^{r} \left(\int s(x)\,d\pi^\sigma_{ti}(x)\right)^{m_i} d\mu(\sigma), \tag{6}$$

where μ is the distribution of σ.

Moyal shows how the probability generating functional is related to various moments. It is perhaps worth noting the fairly obvious fact that, if f is a real valued measurable function, then the mean of $\int f\,dN_t$ or equivalently the mean of $\sum_{\ell=1}^{n} f(X_\ell(t))$ is

$$E \int f\,dN_t = \sum_{i=1}^{r} m_i \int f(x)\,d\tau_{ti}(x) \ , \tag{7}$$

where $\tau_{ti}(G) = \int \pi^\sigma_{ti}(G)\,d\mu(\sigma)$.

Does N_t preserve structure?

Three main classes for the distribution of N_t have been presented. An important question is when are these distributions distinguishable? If two different models give the same distribution for N_t then clearly they can not be distinguished from observations on N_t. On the other hand if the two models give distinct distributions for N_t then observation of N_t can lead to further knowledge.

Theorem 4 analysed the structure of a CCM and in particular showed that the single and several subpopulation cases are not confused, unless of course π_{ti} does not depend on i.

The next question is can the σ constant a.s. case be distinguished from the other cases? This question is answered with the help of some lemmas.

Lemma 3.

A non degenerate mixture of multinomials with common n parameter, $n > 1$, can not be a convolution of multinomials.

Proof

The factorial moment generating function for a convolution of binomials is of the form

$$\prod_{\ell=1}^{n} (\rho_\ell t+1) . \tag{8}$$

For the special case of a binomial this becomes $(tp+1)^n$ and in the case of a mixture of binomials the generating function is

$$E(tp+1)^n \tag{9}$$

where E is expectation with respect to a probability distribution for p. Equating coefficients in the polynomials (8) and (9) we obtain

$$\mathcal{E}np = \sum_{i=1}^{n} \rho_i \text{ and } \mathcal{E}\binom{n}{2} p^2 = \sum_{i<j} \rho_i \rho_j .$$

Thus

$$\begin{aligned}
\mathcal{V}p &= \frac{2}{n(n-1)} \sum_{i<j} \rho_i \rho_j - \frac{1}{n^2}\left(\sum_{i=1}^{n} \rho_i\right)^2 \\
&= 2 \sum_{i<j} \rho_i \rho_j \left(\frac{1}{n(n-1)} - \frac{1}{n^2}\right) - \frac{1}{n^2} \sum_{i=1} \rho_i^2 \\
&= - \frac{1}{n^2(n-1)} \left[(n-1) \sum_{i=1}^{n} \rho_i^2 - 2 \sum_{i<j} \rho_i \rho_j \right] \\
&= - \frac{1}{n^2(n-1)} \sum_{i<j} (\rho_i - \rho_j)^2
\end{aligned}$$

from which it follows that $\rho_1 = .. = \rho_n$ and $\mathcal{V}p = 0$.

To obtain the lemma we note that the margins of a multinomial, and a convolution of multinomials, are respectively binomial and a convolution of binomials. Furthermore if the parameters of the multinomial are $(n, p_1, .., p_k)$ the ith margin has parameters (n, p_i). Thus if a mixture of multinomials is a convolution of multinomials we obtain $\mathcal{V}p_i = 0$, for each i, so that the mixture must be degenerate.

The factorial moment generating function of a mixture of convolutions of binomials is

$$\mathcal{E} \prod_{i=1}^{r} (tp_i + 1)^{m_i}, \tag{10}$$

where $p_1, .., p_r$ are random variables, not all degenerate. From lemma 3 it is clear that (3) can not represent a convolution of binomials for all values of $\underset{\sim}{m} = (m_1, .., m_r)$ summing to $n \geqslant 2$, for one can put $\underset{\sim}{m} = n\underset{\sim}{e}_i$ to show that p_i is degenerate, $i = 1, .., r$. However a more interesting result is available.

Lemma 4.

Consider probability generating functions

$$\mathcal{E} \prod_{i=1}^{n} (\underset{\sim}{s}' \underset{\sim}{p}_i) \tag{11}$$

and

$$\prod_{i=1}^{r} (\underset{\sim}{s}'\underset{\sim}{\rho}_i)^{m_i} \tag{12}$$

$m_i > 0$, $i = 1,..,r$. Suppose (11) and (12) are equal for $\underset{\sim}{m} \in M_1 \cup M_2$ where $M_1 \cap M_2 = \emptyset$, M_1 and M_2 consist of r linearly independent vectors and each element of M_2 is a rational multiple of an element of M_1, then the random vectors $\underset{\sim}{p}_i$, $i = 1,..,r$ are degenerate.

Proof

Denote p_{ij} by p_i, ρ_{ij} by ρ_i and put $s_j = t + 1$, $s_i = 0$, $i \neq j$ to obtain the factorial moment generating functions

$$E\prod_{i=1}^{r} (tp_i+1)^{m_i} \quad \text{and} \quad \prod_{i=1}^{r} (t\rho_i+1)^{m_i}.$$

Equating coefficients as in lemma 3 we obtain

$$E\ \Sigma\ m_i p_i = \Sigma\ m_i \rho_i$$

and

$$E\Big(\sum_i \tfrac{1}{2}m_i(m_i-1)p_i^2 + \sum_{i<j} m_i m_j p_i p_j\Big) = \sum_i \tfrac{1}{2}m_i(m_i-1)\rho_i^2 + \sum_{i<j} m_i m_j \rho_i \rho_j .$$

It follows that

$$\begin{aligned} V(\Sigma m_i p_i) &= \Sigma\ m_i(Ep_i^2 - \rho_i^2) + V(\Sigma m_i \rho_i) \\ &= \Sigma\ m_i(Ep_i^2 - \rho_i^2). \end{aligned}$$

Now if (11) and (12) remain equal when $\underset{\sim}{m}$ is replaced by $q\underset{\sim}{m}$ for some rational $q \neq 1$

$$V(\Sigma q m_i p_i) = \Sigma\ q m_i(Ep_i^2 - \rho_i^2) = q\ V(\Sigma m_i p_i);$$

but $V(\Sigma q m_i p_i) = q^2\ V(\Sigma m_i p_i)$. Thus $V(\Sigma m_i p_i) = 0$. Since this holds for r linearly independent vectors $\underset{\sim}{m}$, $p_1,..,p_r$ must be degenerate.

Theorem 5

Suppose there is just one subpopulation, then N_t can only be a CCM if $X_\ell(t)$ is independent of σ, $\ell = 1,..,n$.

Proof

If N_t is a CCM then $N_t(G)$ is a convolution of binomials. However $N_t(G)$ is the mixture of binomials with generating function

$$E\left(s\pi_t^{\sigma}(G)+1-\pi_t^{\sigma}(G)\right)^n .$$

Applying lemma 3 we see that $\pi_t^{\sigma}(G)$ is constant a.s.. This holds for all $G \in \mathcal{G}$ and so σ and $X_{\ell}(t)$ are independent.

Theorem 5 says two things. First, if it is known that there is only one subpopulation, then it is in principle possible to determine environmental dependence of the $X_{\ell}(t)$ from the distribution of N_t. Secondly the case of a single subpopulation, with $X_{\ell}(t)$ depending on σ, can not be confused with the case of several subpopulations with the $X_{\ell}(t)$ independent of σ or σ constant a.s.

In the case of several subpopulations the distribution of N_t is still capable of revealing environmental dependence of the $X_{\ell}(t)$ provided the numbers in the subpopulations can be varied without altering the π_{ti}^{σ} or the distribution of σ. If $\underset{\sim}{m}$ can be chosen equal to $n\underset{\sim}{e}_i$, $i = 1,..,r$, $n \geqslant 2$, then Theorem 5 applies to show that σ and $X_{\ell}(t)$ are independent for each ℓ. By varying $\underset{\sim}{m}$ in this way one is effectively isolating subpopulations. If subpopulations can not be isolated lemma 4 can be used to show the following.

Theorem 6

Suppose π_{ti}^{σ} and the distribution of σ do not depend on the subpopulation numbers $m_1,..,m_r$. Let M_1, M_2 be as in lemma 4 and $\tau_1,..,\tau_r$ fixed probability measures on $(\mathcal{X},\mathcal{G})$. If N_t has the generating functional

$$\prod_{i=1}^{r} \left(\int s(x)\, d\tau_i(x)\right)^{m_i} ,$$

for every $\underset{\sim}{m} \in M_1 \cup M_2$, then $X_1(t),..,X_n(t)$ and σ are independent.

The questions answered by theorems 5 and 6 relate to comparisons within the structure laid down in section 3.1. The question of the appropriateness of the structure will often arise and in this regard the most obvious hypothesis to test is independence conditional on the environment. This will be easiest to test if the environment is deterministic for then dependence induced by the environment is not confused with the kinds of dependence one is trying to detect. Consider only the single subpopulation case.

Theorem 7

Let $X_1(t),..,X_n(t)$ be jointly symmetrically distributed and suppose that the measure N_t is a continuous multinomial, then $X_1(t),..,X_n(t)$ are independent.

Proof

Let $G_1,..,G_k$ be a partition of $\mathcal{X}$. Lemma 1 of chapter 2 applies showing that $\left(I_{G_1}(X_\ell(t)),..,I_{G_k}(X_\ell(t))\right)$, $\ell = 1,..,n$ are i.i.d. multiple Bernoulli random variables. Hence, for any sets $G_1,..,G_n \in \mathcal{G}$, such that for all i,j either $G_i = G_j$ or $G_i \cap G_j = \emptyset$, then

$$P\left(\bigcap_{\ell=1}^{n} \{X_\ell(t) \in G_\ell\}\right) = \prod_{\ell=1}^{n} P\left(X_\ell(t) \in G_\ell\right).$$

It follows readily from this that $X_1(t),..,X_n(t)$ are independent.

Note that if the animals are labelled at random then $X_1(t),..,X_n(t)$ are necessarily symmetrically distributed. However labelling at random prevents the detection of asymmetric dependence which disappears on symmetrisation. Whether such dependence is likely or interesting in real situations is debatable.

The several subpopulation case provides difficulties in the detection of dependence because of the possible effects of subpopulation numbers.

3.3. Joint distributions in time

The previous section considered the distribution of N_t, for fixed $t \in T$. This might be considered relating more to the outcome of the movement processes than the movement processes themselves. To obtain a real understanding of the movement processes one would expect to have to look at joint distributions for different time points.

In some cases N_t is a homogeneous Markov process, for example the simplest discrete space models, and the distributions for N_t derived previously can sometimes be thought of as conditional distributions for N_t given N_0. Together with an initial distribution, this information provides the distribution of $\{N_t, t \in T\}$. However in general the situation is more complicated.

If τ is a subset of T define $\mathcal{G}^\tau$ to be the smallest σ-subalgebra of $\mathcal{G}^T$ containing all sets of the form

$$\{x \in X^T | x(t) \in G\}, \quad t \in \tau,\ G \in \mathcal{G}.$$

Sometimes the animals are tagged individually in which case it may be possible to observe random variables

$$N_\tau(G) = \sum_{\ell=1}^{n} I_G(X_\ell) \quad , \quad G \in \mathcal{G}^\tau.$$

$N_\tau(G)$ is the number of animals whose paths belong to G. For fixed τ, N_τ is a random measure. Furthermore, all the theory for N_t carries over to N_τ with obvious modifications. Of particular note is an analogue of theorem 7.

Theorem 8

Let $X_1,..,X_n$ be jointly symmetrically distributed and suppose the random measure N_τ is a continuous multinomial for all finite subsets τ of T, then the stochastic processes $X_1,..,X_n$ are independent.

Often it is not possible to distinguish individuals and N_τ can not be observed when τ is not a singleton. Thus the joint distribution of N_t, $t\in\tau$, meaning the distribution of $\{N_t,t\in\tau\}$, must be considered. As noted earlier the sets $\{N_t\in V\}$, $V\in\mathcal{U}$, belong to $\mathcal{A}$. Define U^τ as the set of functions which map τ into U and $\mathcal{U}^\tau$ as the smallest σ-algebra generated by the sets of the form

$$\{\{M_t,t\in\tau\}\in U^\tau|M_t\in V\},\ t\in\tau,\ V\in\mathcal{U}.$$

It follows that $\{N_t,t\in\tau\}$ is $\mathcal{A}/\mathcal{U}^\tau$ measurable and so it is appropriate to regard the distribution of $\{N_t,t\in\tau\}$ as the measure

$$\mathcal{P}(\{N_t,t\in\tau\}\in V),\quad V\in\mathcal{U}^\tau$$

Note that $\mathcal{U}^\tau$ is generated by the class of sets of the form

$$\{\{M_t,t\in\tau\}\in U^\tau|M_{t_i}(G_{ij})\in B_{ij},j=1,..,k_i,i=1,..,p\}\qquad(13)$$

where $G_{ij}\in\mathcal{G}$, $B_{ij}\in\mathcal{B}^1$, $t_1,..,t_p\in\tau$, and k_i, p are arbitrary non negative integers. The sets (13) are closed under finite intersections which means that the distribution of $\{N_t,t\in\tau\}$ is determined on the sets (13). Thus the distribution of $\{N_t,t\in\tau\}$ is determined by the distribution of $\{N_{t_i}(G_{ij}),j=1,..,k_i,i=1,..,p\}$ for all $G_{ij}\in\mathcal{G}$, $t_1,..,t_p\in\tau$, k_i,p. However, as with N_t, we need only consider cases where $G_{i1},..,G_{ik_i}$ is a partition of $\mathcal{X}$, for each i.

When τ is finite, $\tau = \{t_1,..,t_p\}$, $t_i < t_j$ for $i < j$, define the probability generating functional of $\{N_t, t \in \tau\}$ as

$$g_\tau[s] = \mathcal{E}\exp\{\sum_{i=1}^{p} \int \log s_i(x)\, dN_{t_i}(x)\}$$

$$= \mathcal{E} \prod_{\ell=1}^{n} \prod_{i=1}^{p} s_i\big(X_\ell(t_i)\big) ,$$

where s represents a vector $(s_1,..,s_p)$ of bounded complex $\mathcal{G}$ measurable functions.

Let $G_{i1},..,G_{ik_i}$, $i = 1,..,p$ be partitions of $\mathcal{X}$ and consider

$$s_i = \sum_{j=1}^{k_i} s_{ij} I_{G_{ij}} ,$$

where s_{ij} is a complex variable. We obtain

$$g_\tau[s] = \mathcal{E} \prod_{i=1}^{p} \prod_{j=1}^{k_i} s_{ij}^{\,N_{t_i}(G_{ij})}$$

so that g_τ determines the joint probability generating function of $\{N_{t_i}(G_{ij}), j=1,..,k, i=1,..,p\}$ and hence the joint distribution of these random variables. Thus if τ is any subset of T then the distribution of $\{N_t, t \in \tau\}$ is determined by g_ν for all finite subsets ν of τ.

Define $\pi^{\sigma}_{\tau i}$ as the conditional distribution of $\{X_\ell(t), t \in \tau\}$ when animal ℓ belongs to subpopulation i. The symbols σ and i are omitted when they are redundant.

In the single subpopulation and constant σ case we have

$$g_\tau[s] = \mathcal{E} \prod_{\ell=1}^{n} \prod_{i=1}^{p} s_i\big(X_\ell(t_i)\big)$$

$$= \prod_{\ell=1}^{n} \mathcal{E} \prod_{i=1}^{p} s_i\big(X_\ell(t_i)\big)$$

$$= \Big(\int \prod_{i=1}^{p} s_i(x_i)\, d\pi_\tau(\underset{\sim}{x})\Big)^n \qquad (14)$$

where $\underset{\sim}{x}$ represents a vector $(x_1,..,x_p)$. Clearly g_τ determines π_τ and we have the somewhat surprising result that the distribution of $\{N_t, t\in\tau\}$ determines the distribution of $\{X_\ell(t), t\in\tau\}$, for finite τ. This clearly generalises to arbitrary τ and so includes T. Thus we have:

<u>Theorem 9</u>

If there is only one subpopulation and σ is constant a.s. then the distribution of $\{N_t, t\in\tau\}$ determines the distribution of $\{X_\ell(t), t\in\tau\}$. In particular the distribution of $\{N_t, t\in T\}$ determines the distribution $\{X_\ell(t), t\in T\}$.

When there are several subpopulations and σ is constant a.s.

$$g_\tau[s] = \prod_{i=1}^{r} \left(\int \prod_{j=1}^{p} s_j(x_j)\, d\pi_{\tau i}(x)\right)^{m_i} \tag{15}$$

In general the distribution of $\{N_t, t\in\tau\}$ does not determine the distributions of the $\{X_\ell(t), t\in\tau\}$. Consider the case of just two subpopulations with one individual in each. Let $X_1(1)$, $X_1(2)$, $X_2(1)$, $X_2(2)$ be mutually independent so that the distributions of $(X_1(1), X_1(2))$ and $(X_2(1), X_2(2))$ are product measures, $\pi_1 \times \pi_2$ and $\pi_3 \times \pi_4$, for example. From the distribution of $\{N_t, t\in\tau\}$ it is impossible to distinguish the true set of distributions for the $\{X_\ell(t), t\in\tau\}$, namely $\{\pi_1\times\pi_2, \pi_3\times\pi_4\}$, from the set $\{\pi_1\times\pi_4, \pi_3\times\pi_2\}$.

The distribution of $\{N_t, t\in\tau\}$ does provide some information about the distribution of $\{X_\ell(t), t\in\tau\}$, in particular it is possible to distinguish between the single and several subpopulation cases.

<u>Theorem 10</u>

If σ is constant a.s. then the generating functional of $\{N_t, t\in\tau\}$ can be expressed in the form

$$g_\tau[s] = \left(\int \prod_{i=1}^{p} s_i(x_i)\, d\pi(\underset{\sim}{x})\right)^n \tag{16}$$

for some probability measure π if and only if $\pi_{\tau i}$ does not depend on i.

Proof

The sets $G_1 \times .. \times G_p$ belonging to G^p, are closed under intersection, and generate G^p. Thus if the $\pi_{\tau i}$ are not all equal there is a set $G_{11} \times .. \times G_{p1} \in G^p$ such that the numbers $\pi_{\tau i}(G_{11} \times .. \times G_{p1})$, $i = 1,..,r$, are not all equal. There are sets $G_{1j} \times .. \times G_{pj} \in G^p$, $j = 2,..,q$, where q is some positive integer, such that $G_{1j} \times .. \times G_{pj}$, $j = 1,..,q$ forms a partition of X^p.

Fix j for the present and define $s_i = I_{G_{ij}}$ then equating (15) and (16) we find

$$\pi(G_{1j} \times .. \times G_{pj}) = \left\{\prod_{i=1}^{r}\left[\pi_{\tau i}(G_{1j} \times .. \times G_{pj})\right]^{m_i}\right\}^{\frac{1}{n}}$$

which expresses π as a geometric mean of the other measures. However the geometric mean is strictly less than the arithmetic mean unless the numbers being averaged are all equal. Thus

$$1 = \sum_{j=1}^{q} \frac{1}{n}\left\{\sum_{i=1}^{r} m_i\, \pi_{\tau i}(G_{1j} \times .. \times G_{pj})\right\} > \sum_{j=1}^{q}\left\{\prod_{i=1}^{r}[\pi_{\tau i}(G_{1j} \times .. \times G_{pj})]^{m_i}\right\}^{\frac{1}{n}}$$

$$= \sum_{j=1}^{q} \pi(G_{1j} \times .. \times G_{pj}) = 1.$$

The contradiction $(1>1)$ means that the $\pi_{\tau i}$ do not depend on i when (15) and (16) are equal.

When environmental dependence is allowed $\{N_t, t\in\tau\}$ has generating functional

$$g_\tau[s] = \int \prod_{i=1}^{r}\left(\int \prod_{j=1}^{p} s_j(x_j)\, d\pi_{\tau i}^{\sigma}(\underset{\sim}{x})\right)^{m_i} d\mu(\sigma) \tag{17}$$

The distribution of $\{N_t, t\in\tau\}$ is capable of revealing environmental dependence of the $X_\ell(t), t\in\tau\}$, in the case of a single subpopulation.

Theorem 11

If there is only one subpopulation and $n \geqslant 2$ then g_τ can only take the form (16) if $\{X_\ell(t), t \in \tau\}$ is independent of σ.

Proof

In the case of a single subpopulation

$$g_\tau[s] = \int\left(\int_{i} \prod_{=1}^{p} s_i\,(x_i)\,d\pi_\tau^\sigma\right)^n d\mu(\sigma) \tag{18}$$

Let $G_1 \times .. \times G_p \in \mathcal{G}^p$ and $s_i = I_{G_i}$ then if (16) and (18) are equal

$$\int\left\{\pi_\tau^\sigma(G_1 \times .. \times G_p)\right\}^n d\mu(\sigma) = \left(\pi(G_1 \times .. \times G_p)\right)^n .$$

However applying Jensen's inequality we obtain

$$\int \pi_\tau^\sigma(G_1 \times .. \times G_p)\,d\mu(\sigma) \leqslant \left\{\int [\pi_\tau^\sigma(G_1 \times .. \times G_p)]^n d\mu(\sigma)\right\}^{\frac{1}{n}} = \pi(G_1 \times .. \times G_p) \tag{19}$$

with strict inequality if $\pi_\tau^\sigma(G_1 \times .. \times G_p)$ is not constant a.s. with respect to μ.

Now let $G_{1j} \times .. \times G_{pj}$, $j = 1,..,q$ form a partition of $\mathcal{X}^p$. Substituting in (19) and summing we obtain

$$1 \leqslant \sum_{j=1}^{q} \left\{\int [\pi_\tau^\sigma(G_{1j} \times .. \times G_{pj})]^n d\mu(\sigma)\right\}^{\frac{1}{n}} = 1$$

and since the inequality is strict, if any $\pi_\tau^\sigma(G_{1j} \times .. \times G_{pj})$ is not constant a.s., we conclude that each $\pi_\tau^\sigma(G_{1j} \times .. \times G_{pj})$ is constant a.s. For any set $G_{11} \times .. \times G_{p1} \in \mathcal{G}^p$ there are sets $G_{1j} \times .. \times G_{pj} \in \mathcal{G}^p$, $j = 2,..,q$, so that $G_{1j} \times .. \times G_{pj}$, $j = 1,..,q$ forms a partition of $\mathcal{X}^p$. Thus $\pi_\tau^\sigma(G_{11} \times .. \times G_{p1})$ is constant a.s. for all $G_{11} \times .. \times G_{p1} \in \mathcal{G}^p$ which means $\{X_\ell(t), t \in \tau\}$ is independent of σ.

When there are several subpopulations and $\underset{\sim}{m}$ can be varied to take on the values $n\underset{\sim}{e}_i$, $i = 1,..,r$, theorem 11 applies to determine environmental dependence of the

$\{X_\ell(t), t\in\tau\}$ provided the $\pi^\sigma_{\tau i}$ and the distribution of σ do not depend on $\underset{\sim}{m}$.

If the environment is kept constant it is not generally possible to tell whether $\{X_\ell(t), t\in\tau\}$, $\ell = 1,..,n$ are independent from the distribution of $\{N_t, t\in\tau\}$ even when the distribution of $(X_1,..,X_n)$ is symmetric. Theorem 7 does not generalise to cover the $\{X_\ell(t), t\in\tau\}$ and $\{N_t, t\in\tau\}$ as the following counter example shows.

Consider the discrete space model 2d or 2c. Let $Y_1,..,Y_n$ replace $X_1,..,X_n$ in that model and furthermore suppose $P(Y_\ell(0)=1) = 1$ for each ℓ. Now let $Z_\ell(t)$ be the ℓth order statistic of $(Y_1(t),..,Y_n(t))$ and $(j(1),..,j(n))$ a random permutation of $(1,..,n)$ which is independent of the Z_ℓ. Define $X_\ell = Z_{j(\ell)}$. $\{N_t, t\in\tau\}$ has a generating function of the form (16) and the $\{X_\ell(t), t\in\tau\}$ are symmetrically distributed; but the $\{X_\ell(t), t\in\tau\}$ are generally highly dependent. Exceptions occur when $\tau = \{t\}$ or $\tau = \{0,t\}$ which are both cases of independence.

In some cases restrictions more stringent than symmetry enable the distribution of $\{N_t, t\in\tau\}$ to indicate independence of the X_ℓ. Theorem 3 of chapter 2 is an example.

3.4 The limits of the distribution of N_t.

It was pointed out in chapter 2 that if the σ conditional distributions of the $X_\ell(t)$ converge, as $t \to \infty$, then the distribution of $\underset{\sim}{N}(t)$, in the notation of that chapter, converges to a mixture of convolutions of multinomials which variously simplifies according to the dependence of these conditional distributions on σ and the subpopulations. However it is certainly possible to obtain convergence to a convolution of multinomials without convergence of the σ conditional distributions of the $X_\ell(t)$. To see how this occurs consider the simple discrete space discrete time model 2d. Let the transition matrix P have all states communicating and all states periodic with period 2. If half the animals start in one phase and half in the other then a limiting convolution of multinomials results with generating function

$$M\left(\left(\frac{n}{2},\frac{n}{2}\right),\ \begin{bmatrix}\underset{\sim}{p}'_1\\ \underset{\sim}{p}'_2\end{bmatrix},\underset{\sim}{s}\right)$$

where $\underset{\sim}{1}\underset{\sim}{p}'_1 = \lim_{n\to\infty} P^{2n}$ and $\underset{\sim}{1}\underset{\sim}{p}'_2 = \lim_{n\to\infty} P^{2n+1}$.

However $\underset{\sim}{p}_1 \neq \underset{\sim}{p}_2$ and so the distributions of the $X_\ell(t)$ do not converge.

This example, although somewhat pathological and not really of much interest in the movement situation, indicates a need for a more general investigation. The first task is to show that no distributions outside the class of mixtures of convolutions of continuous multinomials can appear.

In order to discuss convergence in distribution we need to use the theory of weak convergence of probability measures as presented by Billingsley (1968) and Parthasarathy (1967).

Let D be a complete separable metric space and $\mathcal{D}$ the σ-algebra generated by the open sets, that is, $\mathcal{D}$ is the Borel sets. Denote by $C(D)$ the set of all real valued bounded continuous functions on D. A sequence of probability measures $\{\pi_m\}$ on $(D,\mathcal{D})$ is said to converge weakly to a probability measure π if and only if

$$\int f\,d\pi_m \to \int f\,d\pi \tag{20}$$

for all $f \in C(D)$. This convergence is denoted by $\pi_m \Rightarrow \pi$. Equivalent definitions of $\pi_m \Rightarrow \pi$ are

$$\liminf_{m \to \infty} \pi_m(G) \geqslant \pi(G) \quad \text{for all open } G \subset X$$

and

$$\lim_{m \to \infty} \pi_m(G) = \pi(G) \quad \text{for all continuity sets of } \pi.$$

Continuity sets of π are defined as follows. Let $G \in \mathcal{G}$, $\bar{G}$ is the closure of G, G° the interior of G and $\partial G = \bar{G} - G^\circ$ is the boundary of G. G is a continuity set of π if and only if $\pi(\partial G) = 0$.

A set Π of probability measures on $(D, \mathcal{D})$ is said to be tight if for every $\varepsilon > 0$ there is a compact subset K_ε of D such that $\pi(K_\varepsilon) > 1 - \varepsilon$ for every $\pi \in \Pi$. Tightness is important because it is equivalent to relative compactness so that, if Π is tight, then every sequence in Π has a subsequence which converges weakly to a probability measure.

The general structure set down in section 3.1 gives the distribution of $(N_t(G_1), \ldots, N_t(G_k))$ as a mixture of convolutions of multinomials. Only mixtures of convolutions of multinomials can appear as limiting distributions.

Theorem 12

A sequence of mixtures of convolutions of multinomials has a convergent subsequence, in the sense that all the probabilities converge, and the limit is a mixture of convolutions of multinomials.

Proof

Let D be the space of $k \times n$ non negative matrices with row sums 1. D can be regarded as a subspace of the metric space R^{nk} with the Euclidean distance. D is then a separable compact space and $\mathcal{D}$ consists of those Borel sets of R^{nk} which are subsets of D.

The generating function for the mth element of the sequence of mixtures of convolutions of multinomials is

$$\int \sum_{i=1}^{n} (\underset{\sim}{p}_i' \underset{\sim}{s})\, d\mu_m(P) \ ,$$

where μ_m is a probability measure on D. $\{\mu_m\}$ is tight because D is compact and so there is a convergent subsequence $\{\mu_{r(m)}\}$. If the limit is μ then

$$\int \prod_{1}^{n} (\underset{\sim}{p}_i' \underset{\sim}{s})\, d\mu_{r(m)}(P) \rightarrow \int \prod_{1}^{n} (\underset{\sim}{p}_i' \underset{\sim}{s})\, d\mu(P),$$

since for fixed $\underset{\sim}{s}$ $\prod_{i=1}^{n} (\underset{\sim}{p}_i' \underset{\sim}{s}) \in C(D)$. For discrete distributions, convergence of the generating function is equivalent to convergence of all of the probabilities and so the theorem is proved.

Note that in theorem 12 the parameter n, corresponding to the number of animals, is the same for each of the mixtures of convolutions of multinomials in the sequence. This assumption holds throughout the section, in particular all members of any set of MCCMs discussed are assumed to have the same parameter n.

Simple corollaries of Theorem 12 show that the structure of the distribution of $N_t(G_1),..,N_t(G_k)$ is preserved in the limit.

Corollary 1

A sequence of convolutions of multinomials has a convergent subsequence whose limit is a convolution of multinomials.

Proof of Corollary 1

$\mu_{r(m)}$ is concentrated on a single point $x_{r(m)}$. It is proved by Parthasarathy (1967, lemma (a), p.40) that since

$\mu_{r(m)} \Rightarrow \mu$ then $x_{r(m)} \to x$ and μ is concentrated on the point x. The corollary now follows.

Corollary 2

A sequence of multinomials has a convergent subsequence with a multinomial limit.

These results settle the question of the kinds of limits available for the distribution of N_t when G is a finite σ-algebra. When G is infinite it is not clear that the limits given by theorem 12, for different partitions $G_1,..,G_k$, will be consistent, particularly if an infinite number of such partitions are considered simultaneously. Additionally some oddities arise.

Consider the case where there is just one animal and X is R^2. Let $X_1(t) = (t,0)$, so that the animal moves off to the right. $X_1(t)$ does not have a proper limiting distribution and there is no MCCM, with $n = 1$, to which the distribution of N_t converges in any reasonable sense. However if $G_1,..,G_k$ is a partition of X then there is a G_j, and a sequence $\{t_m\}, t_m \to \infty$, such that $X_1(t_m) \in G_j$, for each m. $(N_{t_m}(G_1),..,N_{t_m}(G_k))$ has the same distribution for every m and so certainly has a limiting distribution. This distribution is a multinomial with $n = 1$. Thus there appear to be quite marked differences in the possible limiting behaviour of the distributions of $(N_t(G_1),..,N_t(G_k))$ and N_t.

A separate theory is needed for the discussion of possible limiting behaviour of the distribution of N_t. In particular some way of defining convergence of the distribution of a random measure is required. In the theory I present, a random measure whose distribution is the limiting distribution of N_t has $M(\mathcal{X}) = n$ a.s., where $N_t(\mathcal{X}) = n$. This means that the possibility of the escape of any animals is excluded. In some cases it would be quite useful to have some animals possibly escape but it is not in line with the rest of the theory which assumes that a fixed population is always observable. Jagers(1974) gives a theory of convergence in distribution for random measures where $N_t(\mathcal{X})$ and $M(\mathcal{X})$ need not be the same, in particular N_t converges in distribution when $n = 1$ and $X_1(t) = (t,0)$. See also Harris (1968,1971) for a more general theory.

Let $\mathcal{X}$ be a complete separable metric space and assume $\mathcal{G}$ is the Borel sets. For example, $\mathcal{X}$ could be a closed subset of a Euclidean space, R^p, $p = 1,2,$ or 3, with the usual metric or $\mathcal{X}$ could be an open set in R^p with the metric $d(x,y) = |x-y|+|f(x)-f(y)|$, $f(x) = 1/\inf\{|x-y|\,|y\notin\mathcal{X}\}$. In both cases $\mathcal{G}$ consists of the usual R^p Borel subsets of $\mathcal{X}$ (Kelley, 1955, p.207). The second example is of interest when the animals can not reach the boundary of $\mathcal{X}$.

In this section U and $\mathcal{U}$ are given new meanings, but, as we shall see, the new meanings are hardly any different from their former meanings.

Let U be the space of all probability measures on $(\mathcal{X},\mathcal{G})$. From Billingsley (1968) there is a metric p on U such that weak convergence is convergence according to this metric. Furthermore, with this metric U is complete and separable. If $\mathcal{U}$ is the σ-algebra generated by the open sets of U then the theory of weak convergence of probability measures applies to probability measures on $(U,\mathcal{U})$.

Occasionally it is necessary to refer to a base for the topology on U. It is not difficult to see that a base is given by the sets of the form

$$\{\pi \in U \mid |\int f_i \, d\pi - \int f_i \, d\nu_i| < \varepsilon_i , i=1,..,k\} \tag{21}$$

where p is a non negative integer $\varepsilon_1,..,\varepsilon_k > 0$, $f_1,..,f_k \in C(X)$ and $\nu_1,..,\nu_k \in U$. This base is obtained from a sub-base given by Billingsley (1968, p.236). The topology on U is called the weak topology. Jagers' theory uses the "vague" topology which is defined by specifying compact support for $f_1,..,f_k$ in the base (21) and U is simply the space of measures which are finite on compact sets. Jagers also needs X to be locally compact but that is not required here. The approach of Harris produces a similar topology by replacing the compact sets in the definition of the vague topology by more general sets.

Recently Kallenberg (1975) has shown how weak convergence in distribution, on the space of finite measures with the weak topology, is related to weak convergence in distribution on this same space with the vague topology. One approach to our questions here is then to use Kallenberg's ideas and well known results for the vague topology. However I have chosen to make the presentation self contained because the results are not difficult to prove, I do not need local compactness, and lemma 7 has not appeared elsewhere because it is false for arbitrary finite random measures.

The results we obtain concerning N_t are briefly

1. If N_t is a CM, CCM or MCCM then the only available limiting distributions are respectively CMs, CCMs, or MCCMs. Thus the finite time structure of the model is preserved in limiting distributions. This

does seem intuitively obvious but it is pleasant to know that intuition is correct.

2. Relative compactness of the distributions of N_t is determined by relative compactness of the distributions of the $X_\ell(t)$ or alternatively by relative compactness of the set of measures $\{EN_t\}$.

3. Weak convergence in distribution for N_t is equivalent to the weak convergence of the symmetrised distribution of $(X_1(t),\ldots,X_n(t))$ which immediately explains the discrete space example with which this chapter opened.

Taken together we have a complete characterisation of the possible limiting behaviour of N_t under the model of section 3.1.

For each fixed ω, $\frac{1}{n} N_t(.,\omega)$ belongs to U and since n is fixed the theory of weak convergence of probability measures on $(U,\mathcal{U})$ can be used to study convergence of the distribution of N_t. First we must show that the distribution of $\frac{1}{n} N_t$, as defined in 3.2, is a probability measure on $(U,\mathcal{U})$.

If g is a bounded non negative $\mathcal{G}$ measurable function define g^U to be the mapping $\pi \mapsto \int g d\pi$ for $\pi \in U$. Thus g^U is a function on U.

Lemma 5

If g is a bounded non negative $\mathcal{G}$ measurable function then g^U is $\mathcal{U}$ measurable.

Proof

If $g \in C(X)$ then, using (21), g^U is continuous and hence $\mathcal{U}$ measurable. Let $F \subset X$ be closed. Then from a well known theorem (Billingsley's theorem 1.2) there is a uniformly bounded sequence $\{g_m\}$ in $C(X)$ satisfying $\lim_{m\to\infty} g_m(x) = I_F(x)$ for all $x \in X$. Applying the dominated convergence theorem we have

$\lim_{m\to\infty} g_m^U(\pi) = I_F^U(\pi)$, for all $\pi \in U$, so that I_F^U is $\mathcal{U}$ measurable.

The set $\mathcal{F}$, of closed subsets of X, is closed under finite intersections and $\mathcal{G}$ is the σ-algebra generated by $\mathcal{F}$.

Let $\mathcal{H} = \{G \in \mathcal{G} \mid I_G^U$ is $\mathcal{U}$ measurable$\}$. Clearly $\mathcal{F} \subset \mathcal{H}$; $X \in \mathcal{H}$; $G, H \in \mathcal{H}$ and $G \subset H$ implies $H - G \in \mathcal{H}$; and if G_m is an increasing sequence in $\mathcal{H}$, $\cup G_m \in \mathcal{H}$. Thus theorem 2, the monotone class theorem, applies and $\mathcal{H} = \mathcal{G}$.

Lemma 6

If $\mathcal{V}$ is the σ-algebra generated by the sets of the form

$$\{\pi \in U \mid \pi(G_1) \in B_1, \ldots, \pi(G_k) \in B_k\}, \tag{22}$$

where $G_1, \ldots, G_k \in \mathcal{G}$, $B_1, \ldots, B_k \in \mathcal{B}^1$, then $\mathcal{U} = \mathcal{V}$.

Proof

Lemma 5 shows that $\mathcal{V} \subset \mathcal{U}$.

Any $g \in C(X)$ can be approximated uniformly by a sequence of simple functions $\{g_m\}$

$$g_m = \sum_{j=1}^{m} d_{mj} I_{G_{mj}} \quad , \quad G_{mj} \in \mathcal{G}.$$

The g_m^U are $\mathcal{V}$ measurable, $\{g_m^U\}$ uniformly approximates g^U

and so g^U is $\mathcal{V}$ measurable. It follows that the base (21) is a subset of $\mathcal{V}$.

For every point ν of an open subset V of U there is a set V_ν of the form (21) with $\nu \in V_\nu \subset V$. Since V is open $\bigcup_{\nu\in V} V_\nu = V$. Since U is separable and V_ν is open there is a countable subcollection $\{V_{\nu_m}, m=1,2,\dots\}$ such that $\bigcup_m V_{\nu_m} = V$. Hence every open subset of U belongs to $\mathcal{V}$ and so $\mathcal{U} \subset \mathcal{V}$. $\mathcal{U} = \mathcal{V}$.

Lemma 6 shows the distribution of $\frac{1}{n} N_t$, as defined in section 3.2, is a probability measure on $(U,\mathcal{U})$. Next we need a characterisation of tightness of sets of probability measures on $(U,\mathcal{U})$.

Lemma 7

Let Π be a set of probability measures on $(U,\mathcal{U})$ then Π is tight if and only if the set

$\Gamma = \{\tau \in U \mid \tau(G) = \int \pi(G)\,d\mu(\pi)$, for all $G \in \mathcal{G}$ and some $\mu \in \Pi\}$

is a tight subset of U.

Note that $\pi(G) = I_G^U(\pi)$ is a $\mathcal{U}$ measurable function of π.

Proof

Assume Γ is tight. Choose K_m compact so that $\tau(K_m) > 1 - \frac{1}{m^4}$ for all $\tau \in \Gamma$. Define $G_m = \{\pi \mid \pi(K_m) \leqslant 1 - \frac{1}{m^2}\}$.

For $\mu \in \Pi$

$$\begin{aligned} \tau(K_m) &= \int_{G_m} \pi(K_m)\,d\mu(\pi) + \int_{G_m^c} \pi(K_m)\,d\mu(\pi) \\ &> 1 - \frac{1}{m^4} \end{aligned}$$

and so $\left(1 - \frac{1}{m^2}\right)\mu(G_m) + 1 - \mu(G_m) > 1 - \frac{1}{m^4}$

from which we conclude

$$\mu(G_m) < \frac{1}{m^2} .$$

Hence $\sum_{p=m}^{\infty} \mu(G_p) < \sum_{p=m}^{\infty} \frac{1}{p^2}$. Putting

$$B_m = \bigcap_{p=m}^{\infty} G_p^c = \left(\bigcup_{p=m}^{\infty} G_p\right)^c$$

we have

$$\mu(B_m) > 1 - \sum_{p=m}^{\infty} \frac{1}{p^2} \quad \text{for all} \quad \mu \in \Pi$$

Now B_m is a tight set since, for $\pi \in B_m$, $\pi(K_p) > 1 - \frac{1}{p^2}$ for all $p \geqslant m$, and K_p is compact. Thus the closure of B_m, $\bar{B}_m$, is compact.

$$\mu(\bar{B}_m) > 1 - \sum_{p=m}^{\infty} \frac{1}{p^2} \quad \text{for all} \quad \mu \in \Pi,$$

which means Π is tight.

Assuming Π to be tight choose a compact set $B_\varepsilon \subset U$ such that $\mu(B_\varepsilon) > 1 - \varepsilon$ for all $\mu \in \Pi$. Since B_ε is compact, it is tight and there is a compact set $K_\varepsilon \subset X$ such that $\pi(K_\varepsilon) > 1 - \varepsilon$ for every π in B_ε. Hence if $\tau(G) = \int \pi(G)\, d\mu(\pi)$, so that $\tau \in \Gamma$, then

$\tau(K_\varepsilon) \geqslant \int_{B_\varepsilon} \pi(K_\varepsilon)\, d\mu(\pi) > (1-\varepsilon)^2$, showing that Γ is tight.

The final lemma needed is a characterisation of weak convergence of probability measures on the space $(U, \mathcal{U})$. This relies on theorem 2.2 of Billingsley (1968) which I relate as theorem 13.

Theorem 13

Let $\mathcal{R}$ be a subclass of $\mathcal{U}$ such that

1. $\mathcal{R}$ is closed under finite intersections and
2. each open set in U is a finite or countable union of elements of $\mathcal{R}$. If μ_m, μ are probability measures on $(U, \mathcal{U})$ then $\mu_m(V) \to \mu(V)$ for all $V \in \mathcal{R}$ implies $\mu_m \Rightarrow \mu$.

Lemma 8

Let M_m, $m = 1, 2, \ldots$, and M be random probability measures on $(X, \mathcal{G})$. If the distribution of the (R^1, B^1) valued random variable $g^U(M_m)$ converges weakly to the distribution of $g^U(M)$, for every $g \in C(X)$, then the distribution of M_m converges weakly to the distribution of M.

Note that weak convergence of probability measures on (R^k, B^k), where the metric is the Euclidean distance, is equivalent to the ordinary notion of convergence in distribution given by convergence of characteristic functions.

The proof of lemma 8 is based on the ideas in Billingsley (1968, p.15).

Proof

$g^U(M_m)$ converges in distribution to $g^U(M)$ for every $g \in C(X)$. Thus writing $g = t_1 f_1 + .. + t_k f_k$, for arbitrary real $t_1, .., t_k$ and $f_1, .., f_k \in C(X)$ we see that $(f_1^U(M_m), .., f_k^U(M_m))$ converges in distribution to $(f_1^U(M), .., f_k^U(M))$. Let τ_m, τ denote respectively the distributions of $(f_1^U(M_m), .., f_k^U(M_m))$ and $(f_1^U(M), .., f_k^U(M))$. If B is a Borel set in R^k for which $\tau(\partial B) = 0$, ∂B is boundary of B, then $\tau_m(B) \to \tau(B)$ as $m \to \infty$. Thus, defining $V = \{\pi \in U | (f_1^U(\pi), .., f_1^U(\pi)) \in B\}$ and μ_m, μ to be the distributions of M_m and M, then $\mu_m(V) \to \mu(V)$.

Consider the sets $\mathcal{W}$ of the form

$$\{\pi \in U || \int f_i d\pi - \int f_i d\nu_i | < \varepsilon_i , \ i=1, .., k\} \tag{21}$$

for $f_i \in C(X)$, $\varepsilon_i > 0$, $\nu_i \in U$. Let $\mathcal{R}$ denote the subset of $\mathcal{W}$ consisting of finite insections of sets of the form

$$\{\pi \in U || \int f d\pi - \int f d\nu | < \varepsilon\}$$

$\nu \in U$, $f \in C(X)$, $\varepsilon > 0$, for which

$$\mu\{\pi \in U || \int f d\pi - \int f d\nu | = \varepsilon\} = 0 . \tag{23}$$

Let $B = B_1 \times .. \times B_k$, $B_i = (\int f_i d\nu_i - \varepsilon_i , \int f_i d\nu_i + \varepsilon_i)$, $i = 1, .., k$, and suppose (23) holds for each $(f_i, \nu_i, \varepsilon_i)$, then in the previous notation $\tau(\partial\tilde{B}_i) = 0$, where $\tilde{B}_i = \{(x_1,..,x_k) \in R^k \mid x_i \in B_i\}$. Now $\partial(C \cap D) \subset \partial C \cup \partial D$ so that we have $\tau(\partial B) = 0$. Defining $V = \{\pi \in U \mid (f_1^U(\pi),..,f_k^U(\pi)) \in B\} = \{\pi \in U \mid |\int f_i\, d\pi - \int f_i\, d\nu_i| < \varepsilon_i, i=1,..,k\}$ we see that $\mu_m(V) \to \mu(V)$. Hence $\mu_m(V) \to \mu(V)$ for every $V \in \mathcal{R}$.

If ν and f are fixed then the sets

$$\{\pi \in V \mid |\int f d\pi - \int f d\nu| = \varepsilon\}$$

are disjoint for varying ε. Thus (23) holds for all but countably many ε. It follows that $\mathcal{R}$ is a base for the topology on $\mathcal{U}$ and so every open set is a union of elements of $\mathcal{R}$. However U is separable and so each open set is a countable union of elements of $\mathcal{R}$. Applying theorem 13 we see that $\mu_m \Rightarrow \mu$.

It is worth noting that there is a converse to lemma 8 namely if the distribution of M_m converges weakly to the distribution of M then $g^U(M_m)$ converges in distribution to $g(M_m)$, $g \in C(X)$. This holds because, for each real t, $\exp\{tg^U\} \in C(U)$ and so $E\exp\{tg^U(M_m)\} \to E\exp\{tg^U(M)\}$.

From lemma 8 and its converse it is easy to see that the generating functional of N_t converges to a generating functional, pointwise for each s having $\log s \in C(X)$, if and only if the distribution of N_t converges weakly.

The generating functional of N_t can be written

$$\int \prod_{i=1}^{r} \left(\int s(x)\, d\pi_i(x)\right)^{m_i} d\mu_t \tag{24}$$

In (24) μ_t is a measure on the space $(U_r, \mathcal{U}_r)$ where U_r is the space of probability measures on (X^r, G^r) and $\mathcal{U}_r$ is the Borel sets arising from the topology of weak convergence in U_r. μ_t vanishes outside the set Π_r of product probability measures on $(\mathcal{X}^r, \mathcal{G}^r)$. $\pi_1, .., \pi_r$ correspond to the margins of a measure on $(\mathcal{X}^r, \mathcal{G}^r)$ so that an element of Π_r is expressed $\pi = \pi_1 \times .. \times \pi_r$.

$\mathcal{X}^r$ is a complete separable metric space, because $\mathcal{X}$ is, and $\mathcal{G}^r$ is the Borel sets for this space. It follows that U_r is a complete separable metric space.

We need a little theory concerning weak convergence of probability measures on product spaces (Billingsley 1968, p.21,22). If $\{\pi_n\}$ is a sequence in U_r then $\pi_m \Rightarrow \pi$ if and only if $\pi_m(G_1 \times .. \times G_r) \to \pi(G_1 \times .. \times G_r)$ whenever G_i is a continuity set of the ith margin, $i = 1, .., r$. The ith margin is of course the measure $\pi^{(i)}(G) = \pi\{(x_1, .., x_r) | x_i \in G\}$. If $\pi_m = \pi_{1\,m} \times .. \times \pi_{r\,m}$ then $\pi_m \Rightarrow \pi$ if and only if $\pi_{i\,m} \Rightarrow \pi^{(i)}$, for each i, and $\pi = \pi^{(i)} \times .. \times \pi^{(r)}$. These results show that Π_r is a closed subset of U_r and a subset Γ of U_r is tight only if $\{\pi^{(i)} | \pi \in \Gamma\}$ is tight for each i. Also if $K_1, .., K_r$ are compact subsets of $\mathcal{X}$, $K_1 \times .. \times K_r$ is a compact subset of $\mathcal{X}^r$ and $\pi\left((K_1 \times .. \times K_r)^c\right) \leqslant \sum_{i=1}^{r} \pi^{(i)}(K_i^c)$. It follows that Γ is tight if $\{\pi^{(i)} | \pi \in \Gamma\}$ is tight for each i.

Theorem 14

The distributions of N_t form a tight set if and only if the measures $\tau_i(G) = \int \pi_i(G) d\mu_t$ form a tight set. Furthermore the only possible distributions to which the distribution of N_t converges weakly are MCCMs.

Proof

From lemma 7 the distributions of N_t, $t \in T$ are a tight set if and only if (iff)

$\{\frac{1}{n} \mathcal{E} N_t\}$ is tight ($\mathcal{E}N_t$ means the measure $\mathcal{E}N_t(G), G \in \mathcal{G}$)

iff $\{\frac{1}{n}\sum_1^r m_i \int \pi_i \, d\mu_t\}$ is tight

iff $\{\tau_i = \int \pi_i \, d\mu_t\}$ is tight for each i

iff $\{\tau = \int \pi_1 \times .. \times \pi_r \, d\mu_t\}$ is tight

iff $\{\mu_t\}$ is tight.

The last line implies that there is a subsequence $\{\mu_{t_m}\}$ converging weakly to a probability measure μ on $(U_r, \mathcal{U}_r)$. Π_r^c is an open subset of U_r so that

$0 = \liminf \mu_{t_m}(\Pi_r^c) \geqslant \mu(\Pi_r^c)$. Thus μ vanishes outside Π_r.

For positive $s \in C(X)$ the generating functional is a bounded continuous function of μ_t so that

$$\int \prod_{i=1}^{r} \left(\int s(x)\,d\pi_i(x)\right)^{m_i} d\mu_{t_m} \to \int \prod_{i=1}^{r} \left(\int s(x)\,d\pi_i(x)\right)^{m_i} d\mu .$$

The right hand side represents an MCCM. Furthermore writing $s(x) = e^{tf(x)}$, $f \in C(X)$ we see that $\int f dN_{t_m}$ converges in distribution to $\int f dM$ for every $f \in C(X)$, where M is an MCCM. Thus from lemma 8 the theorem is proved.

It is easy to see (c.f. corollary 2) that if N_t is a CCM(CM) then the only limiting distributions available are CCMs (CMs). Thus the finite time structure is preserved in limiting distributions.

The remaining question is when does the distribution of N_t converge? The answer is not surprising in view of the relationship between N_t and the unordered set $\{X_1(t),..,X_n(t)\}$ studied by Moyal.

Theorem 15

The distribution of N_t converges weakly if and only if the symmetrised joint distribution of $X_\ell(t)$, $\ell = 1,..,n$ converges weakly.

(c.f. Kallenberg, 1975, exercise 4.4)

Proof

Assume that defined on (Ω, A, P) there is a random permutation $(j(1),..,j(n))$ of $(1,..,n)$, independent of $\underset{\sim}{X}(t) = (X_1(t),..,X_n(t))$. Define $Y_\ell(t) = X_{j(\ell)}(t)$. The distribution of $\underset{\sim}{Y}(t)$ is the symmetrisation of the distribution of $\underset{\sim}{X}(t)$; $N_t(G) = \sum_{\ell=1}^{n} I_G(Y_\ell(t))$.

If the distribution of $\underset{\sim}{Y}(t)$ converges weakly then $E \prod_{\ell=1}^{n} s(Y_\ell(t)) \to E \prod_{\ell=1}^{n} s(Y_\ell)$, for some random vector $\underset{\sim}{Y}$ taking values in (X^n, G^n), and $s \in C(X)$. Thus for all positive $s \in C(X)$

$$E\, e^{\int \log s(x)\, dN_t(x)} \to E\, e^{\int \log s(x)\, dM(x)}$$

where $M(G) = \sum_{\ell=1}^{n} I_G(Y_\ell)$. Hence the distribution of N_t converges weakly to that of M.

If the distribution of N_t converges weakly, then by theorem 14, the limiting distribution is a MCCM and so there are random variables $Y_1,..,Y_n$ such that M defined by $M(G) = \Sigma I_G(Y_\ell)$ has this limiting distribution. Without loss of generality $Y_1,..,Y_n$ can be assumed to have a symmetric distribution.

If $\log s \in C(X)$ then $E \exp\{\int \log s(x) dN_t(x)\}$ $\to E \exp\{\log s(x) dM(x)\}$ so that

$$E \prod_{\ell=1}^{n} s(Y_\ell(t)) \to E \prod_{\ell=1}^{n} s(Y_\ell) \,. \tag{25}$$

However any non negative $s \in C(X)$ is uniformly approximated, arbitrarily closely, by functions, s' satisfying $\log s' \in C(X)$, and so (25) holds for all non negative $s \in C(X)$.

Define $s = \sum_{\ell=1}^{n} \lambda_\ell s_\ell$, $s_\ell \in C(X)$, s_ℓ non negative and λ_ℓ a non negative real number, then (25) holds for this s and

$$\prod_{\ell=1}^{n} s(x_\ell) = \sum_{j} \prod_{\ell=1}^{n} \lambda_{j(\ell)} s_{j(\ell)}(x_\ell)$$

where the sum is over all functions $j : \{1,\ldots,n\} \to \{1,\ldots,n\}$. Thus $\mathcal{E} \prod_{\ell=1}^{n} s(Y_\ell(t))$ is a polynomial in $\lambda_1,\ldots,\lambda_n$ which converges to the polynomial $\mathcal{E} \prod_{\ell=1}^{n} s(Y_\ell)$ for non negative $\lambda_1,\ldots,\lambda_n$ and so the coefficient, c_t, of $\lambda_1 \ldots \lambda_n$ in $\mathcal{E} \prod_{\ell=1}^{n} s(Y_\ell(t))$ converges to the coefficient, c, of $\lambda_1 \ldots \lambda_n$ in $\mathcal{E} \prod_{\ell=1}^{n} s(Y_\ell)$. Since the distributions are symmetric

$$c_t = n!\, \mathcal{E} \prod_{\ell=1}^{n} s_\ell(Y_\ell(t)) \quad \text{and} \quad c = n!\, \mathcal{E} \prod_{\ell=1}^{n} s_\ell(Y_\ell).$$

Hence

$$\mathcal{E} \prod_{\ell=1}^{n} s_\ell(Y_\ell(t)) \to \mathcal{E} \prod_{\ell=1}^{n} s_\ell(Y_\ell), \tag{26}$$

for non negative $s_\ell \in C(X)$, which trivially extends to arbitrary $s_\ell \in C(X)$. This is sufficient to prove weak convergence of the distribution of $(Y_1(t),\ldots,Y_\ell(t))$ to the distribution of $(Y_1,\ldots,Y_n)$. I am not sure where this is proved but in any case (26) implies

$$P\big((Y_1(t),\ldots,Y_n(t)) \in G_1 \times \ldots \times G_k\big) \to P\big((Y_1,\ldots,Y_n) \in G_1 \times \ldots \times G_k\big),$$

G_ℓ a continuity set for the distribution of Y_ℓ, as follows.

Suppose

$$\mathcal{E} \prod_{\ell=1}^{j} I_{G_\ell}(Y_\ell(t)) \prod_{\ell=j+1}^{n} s_\ell(Y_\ell(t)) \to \mathcal{E} \prod_{\ell=1}^{j} I_{G_\ell}((Y_\ell)) \prod_{\ell=j+1}^{n} s_\ell(Y_\ell) \tag{27}$$

for G_ℓ a continuity set for the distribution of Y_ℓ, $\ell = 1,\ldots,j$ and $s_\ell \in C(X)$, $\ell = j+1,\ldots,n$. Choose s_ℓ non negative, $\ell = j+1,\ldots,n$ and $s_{j+1} \equiv 1$. Without loss of generality assume that the R.H.S. of (27) is positive. The probability measure

$$\pi_t(G) = \mathcal{E} \prod_{\ell=1}^{j} I_{G_\ell}(Y_\ell(t)) I_G(Y_{j+1}(t)) \prod_{\ell=j+2}^{n} s_\ell(Y_\ell(t)) \Big/ \mathcal{E} \prod_{\ell=1}^{j} I_{G_\ell}(Y_\ell(t)) \prod_{\ell=j+2}^{n} s_\ell(Y_\ell(t)),$$

defined for t sufficiently large, converges weakly to

$$\pi(G) = \mathcal{E} \prod_{\ell=1}^{j} I_{G_\ell}(Y_\ell) I_G(Y_{j+1}) \prod_{\ell=j+2}^{n} s_\ell(Y_\ell) \Big/ \mathcal{E} \prod_{\ell=1}^{j} I_{G_\ell}(Y_\ell) \prod_{\ell=j+2}^{n} s_\ell(Y_\ell)$$

Thus (27) holds for $j+1$ replacing j. Equation (26) proves (27) for $j = 0$ thus (27) holds for $j = n$, which means that the distribution of $(Y_1(t),\ldots,Y_n(t)) \Rightarrow$ to the distribution of $(Y_1,\ldots,Y_n)$.

Limits of joint distributions

The theory for N_t applies equally well to $N_{t+\tau}$, where $t + \tau = \{u+t | u \in \tau\}$, for finite or countable τ because X^p is a complete separable metric space with Borel sets G^p, $p = 1,2,..,\infty$. Furthermore, since N_t, $t \in \tau$ are the margins of N_τ, in the case of finite τ tightness of $\{N_u, u \in t+\tau, t \in T\}$ is equivalent to tightness of $\{EN_{t+\tau}, t \in T\}$ and so some of the theory for $N_{t+\tau}$ can be modified to apply to $\{N_u, u \in t+\tau\}$.

However when τ is uncountable, the situation is more complicated, for restrictions must be placed on the sample paths of the X_ℓ; but weak convergence for all finite subsets of τ can still be considered.

3.5 Construction of models from Markov processes.

In the discrete space models random substitution of the time parameter was found to be a convenient way of introducing variability of various kinds in movement processes. There are obviously many other ways such variability could be expressed. However this method has the advantage that it has an interpretation in terms of the behaviour of the animals, which means it is amenable to experimental testing. Also this method is very simple, and leads to relatively tractable mathematics.

Following a similar approach to that in the discrete space models, examples of the probability space (Ω, A, P) will be constructed from homogeneous Markov processes. The construction depends very little on Markov properties and so also applies to many other processes.

Assume that the space X has a topology for which G is the Borel sets. The set T is $[0,\infty)$.

Definition

A stochastic process $\{X(t), t \geq 0\}$ is homogeneous Markov process if there is a function p on $[0,\infty) \times X \times G$ such that

1. $p(t,x,.)$ is a probability measure on G for fixed $t \geq 0$, $x \in X$.
2. $p(t,.,G)$ is G measurable for fixed $t \geq 0$, $G \in G$.
3. $p(t+s,x,G) = \int p(t,y,G)p(s,x,dy)$, $x \in X, t, s \geq 0$
4. $P(X(t+s) \in G | X(u), u \leq s) = p(t, X(s), G)$, $t, s \geq 0$, $G \in G$.

The function p is called the transition measure.

For the later discrete space models, in addition to randomisation of time, the matrix Q was allowed to vary. The latter variability is of an essentially different nature. Two homogeneous Markov processes will be said to be of the same type if the transition measure of one process can be transformed into the transition measure of the other by a linear time transformation $t \mapsto \alpha t$, where α is a fixed positive number. If two processes are of the same type then they will be said to evolve at different rates if $\alpha \neq 1$. The only differences between homogeneous Markov processes of the same type are the rates at which the processes evolve and the initial distributions. The partitioning of variation into differences of type and differences in rates, developed in chapter 2, will be extended to the more general case.

One reason for restricting consideration to homogeneous Markov processes is the simplicity of the definitions and interpretations of differences in type and differences in the rates of evolution of different processes. Another reason is that homogeneous Markov processes are frequently suggested as models for animal movements.

The result of the construction is easily summarised. We simply write $X_\ell(t) = Y_\ell(A_\ell(t))$ where conditional on z_ℓ, the ℓth animal's phenotype, Y_ℓ is a homogeneous Markov process independent of A_ℓ. A_ℓ has the form $A_\ell(t) = A(t,s_\ell,\sigma,z_\ell)$ where A is a fixed function, s_ℓ represents the random activity of section 2.4, and σ is the environment. $(Y_1,z_1,s_1),\ldots,(Y_n,z_n,s_n)$ are independent and z_ℓ has distribution ρ_i if the ℓth animal belongs to subpopulation i. In this way we obtain a generalisation of section 2.4 and a probability space satisfying the general structural model. Unfortunately the construction which follows is a little technical.

Recall that Z is the space consisting of the phenotypes of the animals and $\mathcal{Z}$ is an appropriate σ-algebra for this space. With the ith subpopulation associate a probability measure ρ_i giving the distribution of the phenotypes of animals in the ith subpopulation. We want the phenotype of the animal to determine the probability distribution of a Markov process and in addition it will be convenient if the Markov processes can be assumed to have right continuous sample paths. An element x of $\mathcal{X}^T$, or equivalently a function $x : T \to \mathcal{X}$, is said to be right continuous if $\lim_{s\downarrow t} x(s) = x(t)$ for every $t \in T$.

Let $\mathcal{X}^*$ denote the subset of $\mathcal{X}^T$ consisting of right continuous functions and $\mathcal{G}^*$ the σ-algebra $\{\mathcal{X}^*\cap G \mid G\in\mathcal{G}^T\}$. If, defined on some probability space $(\Omega_0,\mathcal{A}_0,P_0)$, there is a homogeneous Markov process $Y_0 = \{Y_0(t), t\geq 0\}$ having $Y_0(\cdot,\omega)\in\mathcal{X}^*$ for all $\omega\in\Omega_0$ then it is easy to see that Y_0 is $\mathcal{A}_0/\mathcal{G}^*$ measurable; and there is a homogeneous Markov process Y defined on $(\mathcal{X}^*,\mathcal{G}^*,\pi)$, $\pi(G) = P_0(Y_0\in G)$ for $G\in\mathcal{G}^*$, such that Y has the same distribution as Y_0. Specifically, let Y be defined on X^* by the formula $Y(t,x) = x(t), x \in X^*$. Y is G^*/G^T measurable and, for $G \in G^T$, $\pi(Y \in G) = \pi(G \cap X^*) = P_0(Y_0 \in G \cap X^*) = P_0(Y_0 \in G)$. Thus Y_0 and Y have the same distribution. Y is a homogeneous Markov process because Y_0 is, and of necessity Y has right continuous sample paths. The construction of the process Y shows that there is no loss of generality in restricting consideration to processes defined on (X^*, G^*).

Blumenthal and Getoor (1968 section I 1.) show that the existence of a right continuous version of a Markov process is a very weak requirement and certainly includes all cases of interest in the study of animal movements.

Assume that there is defined on the space (Z, Z) a random probability measure $\tilde{\pi} = \{\tilde{\pi}(G,z), G \in G^*, z \in Z\}$. For each fixed z the measure $\tilde{\pi}(\cdot, z)$ is a probability measure on (X^*, G^*) such that Y is a homogeneous Markov process. The other part of the definition of a random probability measure says that $\tilde{\pi}(G)$ is Z measurable for fixed $G \in G^*$. Thus for each phenotype there is a basic Markov process from which the movement processes of the animals will be constructed. For different values of z, $\tilde{\pi}(\cdot, z)$ may represent Markov processes of different types. However it is desirable that if $\tilde{\pi}(\cdot, z)$ and $\tilde{\pi}(\cdot, z')$ represent Markov processes of the same type then they are equal because differences in rates of evolution are represented in another way in the present construction. In the discrete space models the restriction trace $Q_\ell = 1$ was a natural way of preventing variation in rates of evolution which applies independently of type. Possibly similar methods can be found in other situations.

Also defined on $(Z,\mathcal{Z})$ is a random function A of three variables $t \in T$, $s \in S$ and $\sigma \in E$. $A(t,s,\sigma)$ will be used for random time substitution in the basic Markov process to give differences in rates of evolution and to interpret the effects of random activity, s, and the environment, σ. To provide the random activity and random environment we need probability spaces $(S,\mathcal{Q},\theta)$ and $(E,\mathcal{S},\mu)$. It is assumed that for fixed t, $A(t,\cdot,\cdot)$ is $\mathcal{Q} \times \mathcal{S} \times \mathcal{Z}/\mathcal{B}^1$ measurable, and furthermore for fixed s, σ and z, $A(t,s,\sigma)$ is an increasing function of t.

Since the construction becomes rather complicated if many individuals are considered simultaneously, from this point I shall give the construction for just one individual and indicate at the end how it is modified for many individuals.

For each fixed $z \in Z$ define τ as the probability measure $\tilde{\pi} \times \theta \times \mu$ on $(\mathcal{X}^* \times S \times E, \mathcal{G}^* \times \mathcal{Q} \times \mathcal{S})$. For $G \in \mathcal{G}^*$, $Q \in \mathcal{Q}$, $F \in \mathcal{S}$, $\tau(G \times Q \times F)$ is $\mathcal{Z}$ measurable since the set of all such product sets $G \times Q \times F$ is closed under intersection and generates $\mathcal{G}^* \times \mathcal{Q} \times \mathcal{S}$. The monotone class theorem (theorem 2) applies to show that $\tau(B)$ is measurable for all $B \in \mathcal{G}^* \times \mathcal{Q} \times \mathcal{S}$. τ has the properties of a regular conditional distribution and so the equation

$$P(B \times V) = \int_V \tau(B)\, d\rho_i$$

defines P as a probability measure on the set $\mathcal{A} = \mathcal{G}^* \times \mathcal{Q} \times \mathcal{S} \times \mathcal{Z}$

Let $\Omega = \mathcal{X}^* \times S \times E \times Z$. Define random variables $Y_\ell, s_\ell, \sigma, z_\ell$ on $(\Omega,\mathcal{A})$ by $Y_\ell(t,x,s,\sigma,z) = Y(t,x) = x(t)$, $s_\ell(x,s,\sigma,z) = s$, $\sigma(x,s,\sigma,z) = \sigma$, $z_\ell(x,s,\sigma,z) = z$. Clearly $Y_\ell(t), s_\ell, \sigma, z_\ell$ are respectively $\mathcal{A}/\mathcal{G}$, $\mathcal{A}/\mathcal{Q}$, $\mathcal{A}/\mathcal{S}$ and $\mathcal{A}/\mathcal{Z}$ measurable. Furthermore for each fixed $\omega \in \Omega$, $Y_\ell(\cdot,\omega) \in \mathcal{X}^*$.

To set up the movement process X_ℓ suppose that α is a non negative random variable on $(\Omega,\mathcal{A},\mathcal{P})$ then $Y_\ell(\alpha)$ is an $(\mathcal{X},\mathcal{G})$ valued random variable on $(\Omega,\mathcal{A},\mathcal{P})$. To see this define $\alpha_m = [m\alpha+1]/m$ where $[a]$ is the greatest integer $\leqslant a$. $Y_\ell(\alpha_m)$ is $\mathcal{A}/\mathcal{G}$ measurable since

$$\{Y_\ell(\alpha_m)\in G\}=\bigcup_{j=1}^{\infty}\{Y_\ell(j/m)\in G,\alpha_m=j/m\}. \quad \lim_{m\to\infty} Y_\ell(\alpha_m)=Y_\ell(\alpha)$$

because $Y_\ell(\cdot,\omega)\in\mathcal{X}^*$ for all $\omega\in\Omega$. Hence $Y_\ell(\alpha)$ is $\mathcal{A}/\mathcal{G}$ measurable. An analogous proof shows that Y_ℓ, as a function on $\Omega\times[0,\infty)$, is $\mathcal{A}\times\mathcal{B}^1$ measurable. Define $A_\ell(t)$ as the mapping $(x,s,\sigma,z)\mapsto A(t,s,\sigma,z)$, where $\mathcal{B}^1$ is the Borel sets in R^1. $A_\ell(t)$ is a non negative random variable on $(\Omega,\mathcal{A},\mathcal{P})$. We can now define $X_\ell(t) = Y_\ell(A_\ell(t))$. The process X_ℓ is $\mathcal{A}/\mathcal{G}^T$ measurable.

To derive conditional probabilities for X_ℓ define $C(z) = \{(x,s,\sigma)\in\mathcal{X}^*\times S\times E \mid (x,s,\sigma,z)\in C\}$. If $C = B\times W$, $B\in\mathcal{G}^*\times\mathcal{Q}\times\mathcal{S}$, $W\in\mathcal{Z}$ then it follows directly from the definition that $\tau(C(z_\ell))$, which equals $\tau(B)I_W(z_\ell)$, is a version of $\mathcal{P}(C|z_\ell)$. The monotone class theorem now applies to show that $\tau(C(z_\ell))$ is a version of $\mathcal{P}(C|z_\ell)$ for all $C\in\mathcal{A}$. Furthermore the conditional probability measure $\tau(C(z_\ell))$, $C\in\mathcal{A}$, is regular. Defining

$$C(s,\sigma,z) = \{x\in\mathcal{X}^* \mid (x,\sigma,s,z)\in C\}$$

it follows that $\tilde{\pi}(C(s_\ell,\sigma,z_\ell),z_\ell)$ is a regular conditional probability measure given (s_ℓ,σ,z_ℓ). With this result it is clear that, conditional on (s_ℓ,σ,z_ℓ), X_ℓ is a non homogeneous Markov process. Let $C = \{X_\ell\in G\}$ for $G\in\mathcal{G}^T$. For $\sigma\in E$, $z\in Z$ define

$$\pi(G;\sigma,z) = \int \tilde{\pi}(C(s,\sigma,z))d\theta(s).$$

Then $\pi(G;\sigma,z_\ell)$ is a regular conditional distribution for X_ℓ given (σ,z_ℓ).

For n animals, m_i in each subpopulation consider a vector $(\tilde{\pi}_1,..,\tilde{\pi}_n)$ of probability measures defined on Z^n such that

$$\tilde{\pi}_j(G,z_1,..,z_n) = \tilde{\pi}(G,z_j), \; G \in G^T, \quad \text{and} \quad (z_1,..,z_n) \in Z^n.$$

For fixed $(z_1,..,z_n) \in Z^n$ define

$$\tau = \tilde{\pi}_1 \times .. \times \tilde{\pi}_n \times \theta^n \times \mu$$

and $\quad P(B \times V) = \int_V \tau(B)\, d\rho_1^{m_1} \times .. \times \rho_r^{m_r}$

where $V \in Z^n$, $B \in (G^*) \times Q^n \times S$. Define, by abuse of notation at times, $\omega = (x_1,..,x_n, s_1,..,s_n, \sigma, z_1,..,z_n)$, $Y_\ell(t,\omega) = x_\ell(t)$, $s_\ell(\omega) = s_\ell$, $z_\ell(\omega) = z_\ell$, $\sigma(\omega) = \sigma$, and $A_\ell(t)$ as the mapping $\omega \mapsto A(t,s_\ell,\sigma,z_\ell)$. Now define

$$\Omega = (X^*)^n \times S^n \times E \times Z^n, \quad A = (G^*)^n \times Q^n \times S \times z^n .$$

Finally $X_\ell(t) = X_\ell\big(A_\ell(t)\big)$.

If the function π is defined as before it is not difficult to see that

$$P\Big(\bigcap_{\ell=1}^{n} \{X_\ell \in G_\ell\} \,|\, \sigma, z_1,..,z_n\Big) = \prod_{\ell=1}^{n} \pi(G_\ell;\sigma,z_\ell)$$

which is S_1' of section 3.1. Thus the construction has been successful.

3.6 The population is a sample from a large finite population.

In this section P and E are given the subscript n to indicate the dependence of these quantities on the size of the population.

Suppose there is a very large population of animals whose movements we wish to study. Because of physical limitations we tag a number of these and this number, although possibly large, is still small compared with the total population. This is a common situation, for example in *Drosophila* dispersal studies (Dobzhansky and Wright, 1943, Crumpacker and Williams, 1973) and in fish migration studies (Jones, 1959).

Frequently the data are analysed as though the animals in the tagged sample have independent movements, without providing any justification for this assumption. Thus an investigation of the kind of dependence expected between the individuals in the sample is warranted. It is also interesting to see if the earlier theory of this chapter is of any use in the present setting. In particular we will seek an example of a situation where the animals in the sample are approximately independent even though they interact with each other and the rest of the population.

The intuition behind this example is that if an individual interacts with many other individuals then some law of large numbers may be in force making the effect of these interactions approximately deterministic. The individuals in the sample then appear to have independent movements.

Suppose that the sample is random, in the usual sense, as is often regarded highly desirable. If there are m animals in the sample they have movement processes $X_1,..,X_m$ which are jointly symmetrically distributed. Without any loss of generality we can think of the movement processes $X_1,..,X_n$, for the entire population, as symmetrically distributed. The random measure $p_n = \frac{1}{n} N_t$ can be thought of as an empirical distribution for the X_ℓ. Also if $G \in (\mathcal{G}^T)^n$

$$P_n\left((X_1,..,X_n) \in G \middle| \frac{1}{n} N_T\right) = \frac{1}{n!} \sum_j I_G(X_{j(1)},..,X_{j(n)}) = H_n(G)$$

the sum being over all permutations j of $(1,..,n)$. The easiest way to see this is to use the result of Moyal, mentioned in 3.2, that the σ-algebra with elements $\{(X_1,..,X_n) \in G_s\}$, G_s being a symmetric element of $(\mathcal{G}^T)^n$ equals the σ-algebra $\mathcal{A}(N_T)$ generated by N_T. It is then obvious that $H_n(G)$ is $\mathcal{A}(N_T)$ measurable and if $A \in \mathcal{A}(N_T)$

$$\mathcal{E}_n H_n(G) I_A = \mathcal{E}_n H_n(G) I_{G_s}(X_1,..,X_n), \quad G_s \quad \text{symmetric},$$

$$= \frac{1}{n!} \sum_j \mathcal{E}_n I_G\big(X_{j(1)},..,X_{j(n)}\big) I_{G_s}\big(X_{j(1)},..,X_{j(n)}\big)$$

$$= \frac{1}{n!} \sum_j P_n\big((X_1,...,X_n) \in G \cap G_s\big)$$

$$= P_n\big((X_1,..,X_n) \in G \cap G_s\big).$$

$$= P_n\big(\{(X_1,..,X_n) \in G\} \cap A\big).$$

The random probability measure H_n is a regular conditional distribution for $X_1,..,X_n$ given N_T or p_n.

We can write

$$P_n\big((X_1,..,X_n) \in G\big) = \mathcal{E}_n H_n(G) \tag{28}$$

Hewitt and Savage (1955) give the result that any symmetric probability measure σ, can be represented (uniquely) as an integral

$$\sigma A = \int_{P_n} \tilde{\pi}(A)\, d\mu(\pi) \tag{29}$$

where P_n is the space of all $\pi = \frac{1}{n}\sum_1^n \pi_i$, π_i is a 0-1 probability measure, $\tilde{\pi} = \frac{1}{n!}\sum_j \pi_{j(1)} \times .. \times \pi_{j(n)}$, with j varying over permutations; μ is a probability measure on P_n. Note that for fixed values of $X_1,..,X_n$, H_n belongs to P_n and (28) and (29) are essentially the same; the main difference is that H_n is a random measure on the basic probability space $(\Omega, \mathcal{A}, P)$ whereas $\tilde{\pi}$ is not.

Define $H_{n,m}(G_1 \times .. \times G_m) = H_n(G_1 \times .. G_m \times \mathcal{X}^T \times .. \times \mathcal{X}^T)$. Hewitt and Savage show

$$\Big|H_{n,m}(G_1 \times .. \times G_m) - \prod_{\ell=1}^{m} p_n(G_\ell)\Big| \leqslant 2\frac{m-1}{n}. \tag{30}$$

Thus for small $\frac{m}{n}$, $X_1,..,X_m$ are approximately independent, conditional on p_n, with common marginal distribution p_n.

Let $\mathcal{X}$ be a complete separable metric space and G as in section 3.4. In the sequel we shall assume that

$T = \{0,1,\ldots\}$ in which case $\mathcal{X}^T$ is a complete separable metric space with Borel sets $\mathcal{G}^T$. If $T \neq \{0,1,\ldots\}$ then the theory applies to countable subsets of T and also applies when $\mathcal{X}^T$ is replaced by a subset which is a complete separable metric space.

Let U_T be the space of probability measures on $(\mathcal{X}^T, \mathcal{G}^T)$ and $\mathcal{U}_T$ the Borel sets generated by the topology of weak convergence on U_T. If μ_n is the distribution of p_n then from (28) and (30) we deduce

$$\left| \mathcal{P}_n(X_\ell \in G_\ell, \ell=1,..,m) - \int \prod_{\ell=1}^{m} \pi(G_\ell)\,d\mu_n(\pi) \right| \leqslant 2\frac{m-1}{n}. \qquad (31)$$

This statement is of particular interest for it indicates that when m/n is small the joint distribution of $X_1,..,X_m$ is close to a mixture of distributions of i.i.d. random variables. The latter is equivalent to the most general kind of dependence induced by the environment. Thus the theorems in sections 3.2 and 3.3 relating to the distinction between environmental dependence and independence would appear to be approximately applicable to the situation assumed in this section.

Now suppose $\mu_n \Rightarrow \mu$ in (31) then the distribution of $X_1,..,X_m$ converges weakly to τ defined by

$$\tau(G) = \int \pi^m(G)\,d\mu(\pi) \quad , \quad G \in (\mathcal{G}^T)^m.$$

Moreover, considering the joint distribution of p_n and $X_1,..,X_m$ on the space $\left((\mathcal{X}^T)^m \times U, (\mathcal{G}^T)^m \times \mathcal{U}_T\right)$, we find that this distribution converges weakly to $\mathcal{P}$ defined by

$$\mathcal{P}(G \times V) = \int_V \pi^m(G)\,d\mu(\pi),\ G \in (\mathcal{G}^T)^m,\ \ V \in \mathcal{U} \qquad (32)$$

This result follows from the equation

$$\begin{aligned} &\mathcal{E}_n g(p_n) \prod_1^m f_j(X_j) - \int g(\pi)\Big(\prod_1^m \int f_j\,d\pi\Big)d\mu(\pi) \\ &= \Big\{\mathcal{E}_n g(p_n)\Big(\prod_1^m f_j(X_j) - \prod_1^m \int f_j\,dp_n\Big)\Big\} \\ &+ \Big\{\int g(\pi)\prod_1^m \int f_j\,d\pi d\mu_n(\pi) - g(\pi)\prod_1^m \int f_j\,d\pi d\mu(\pi)\Big\} \end{aligned}$$

where $f_1, \ldots, f_m$ and g are bounded continuous functions defined on the appropriate spaces and we make use of the fact that Hewitt and Savages' proof of (30) extends without essential change to show

$$\left| \int \prod_1^m f_i \, dH_{n,m} - \prod_1^m \int f_i \, dp_n \right| \leqslant \frac{2m-1}{n}, \quad \text{for} \quad |f_i| \leqslant 1$$

From (32) we see that, in the limiting distribution, $X_1, \ldots, X_m$ are i.i.d. conditional on π. It seems appropriate to regard the randomness of π as a result of random environment and interaction between individuals. Of particular interest in the present study are situations where π is constant conditional on the environment; $X_1, \ldots, X_m$ are then i.i.d. conditional on the environment. To study these cases let us assume that the environment is deterministic, then we simply look for situations where π has a degenerate distribution or equivalently the limiting distribution of p_n is degenerate. First we note that $X_1, \ldots, X_m$ are asymptotically independent with common distribution π_0 if and only if p_n converges in distribution to π_0. This is lemma 9.

Lemma 9.

The joint distribution of $X_1, \ldots, X_m$ converges weakly to π_0^m if and only if $\mu_n \Rightarrow \mu$ where $\mu(\{\pi_0\}) = 1$.

Proof

The "if" part of the statement is proved above. For the "only if" part, the hypothesis implies

$$\mathcal{E}_n f(X_i) f(X_j) \to \left(\int f d\pi_0\right)^2 \quad \text{for} \quad f \in C(X^T), \quad i \neq j. \tag{33}$$

Moreover the convergence is uniform in (i,j) because of the symmetry of distribution. Now $\int f dp_n = \frac{1}{n} \sum_1^n f(X_\ell)$ and

$$\mathcal{E}_n \left\{ \frac{1}{n} \sum_\ell f(X_\ell) - \int f d\pi_0 \right\}^2 = \frac{1}{n^2} \sum_1^n \mathcal{E}_n \left(f(X_\ell) - \int f d\pi_0\right)^2 + \frac{1}{n^2} \sum_{i \neq j} \mathcal{E}_n \left(f(X_i) - \int f d\pi_0\right)\left(f(X_j) - \int f d\pi_0\right)$$

$$= \frac{1}{n} \mathcal{E}_n \left(f(X_1) - \int f d\pi_0\right)^2 + \frac{n(n-1)}{n^2} \mathcal{E}_n \left(f(X_1) - \int f d\pi_0\right)\left(f(X_2) - \int f d\pi_0\right)$$

Thus using (33) we have

$$E_n(\int f dp_n - \int f d\pi_0)^2 \to 0 \quad \text{as} \quad n \to \infty.$$

Since this holds for all $f \in C(X^T)$, lemma 8 yields $\mu_n \Rightarrow \mu$, $\mu(\{\pi_0\}) = 1$.

To obtain a situation where $X_1, .., X_m$ are asymptotically independent postulate that the ℓth animal's next movement is dependent on the current and past positions of itself and the other animals, but in the presence of this information, its next movement does not depend on the next movement of the other animals. Furthermore assume that this dependence on present and past positions can be expressed continuously in terms of $X_\ell(0), .., X_\ell(t)$ and $p_n^{(t)}$ defined by $p_{n(G)}^{(t)} = \frac{1}{n} \sum_{i=1}^{n} I_G(X_i(0), .., X_i(t))$, t being the present. This dependence on $p_n^{(t)}$ is slightly more general than dependence on past and present normalised densities of the population. If U_t is the space of probability measures on (X^t, G^t), with the topology of weak convergence, the model can be stated precisely as

1. $P_n(X_\ell(t+1) \in G_\ell, \ell=1, .., m \mid X_\ell(s), s \leq t, \ell=1, .., n)$

$$= \prod_{\ell=1}^{m} P_n(X_\ell(t+1) \in G_\ell \mid X_\ell(s), s \leq t, p_n^{(t)})$$

2. For every t there is a mapping

$$\gamma^{(t)} : X^{t+1} \times U_{t+1} \times G \to [0,1]$$

such that

$$P_n(X_\ell(t+1) \in G \mid X_\ell(s), s \leq t, p_n^{(t)}) = \gamma^{(t)}(X_\ell(0), .., X_\ell(t), p_n^{(t)}, G)$$

and the mapping

$$(x_0, .., x_t, \pi) \mapsto \gamma^{(t)}(x_0, .., x_t, \pi, \cdot), (x_0, .., x_t) \in X^{t+1}, \quad \pi \in U_{t+1}$$

is a continuous mapping of $X^{t+1} \times U_{t+1}$ into U_1.

For there to be any hope of asymptotic independence of the X_ℓ we must have asymptotic independence of the $X_\ell(0)$ or equivalently convergence in distribution of $p_n^{(0)}$ to a constant. We shall impose this condition and then see if 1. and 2., which concern the joint behaviour of the animals, are sufficient to give asymptotic independence of the X_ℓ.

Theorem 15.

Let 1. and 2. hold. If the distribution of $p_n^{(0)}$ converges weakly to a constant then the joint distribution of $X_1,..,X_m$ converges weakly and in their asymptotic distribution $X_1,..,X_m$ are i.i.d.

Proof.

Induction on t is used to show that $\{X_\ell(s), s \leqslant t\}$, $\ell = 1,..,r$ are asymptotically i.i.d. with common distribution $\pi^{(t)}$, say.

Lemma 9 holds with $\{X_\ell(s), s \leqslant t\}$ replacing X_ℓ, $\ell = 1,..,n$ thus the case $t = 0$ is a direct consequence.

If $\{X_\ell(s), s \leqslant t\}$, $\ell = 1,..,m$ are asymptotically independent with common distribution $\pi^{(t)}$ then lemma 9 implies that $p_n^{(t)}$ converges in distribution to $\pi^{(t)}$ and we also have convergence of the joint distribution of $\{X_\ell(s), s \leqslant t\}$, $\ell = 1,..,m$ and $p_n^{(t)}$.

If $f \in C(X)$ then the mapping $\pi \mapsto \int f d\pi$, $\pi \in U_1$, is continuous. It follows that the mapping

$$(x_0,...,x_t,\pi) \mapsto \int f(x)\gamma^{(t)}(x_0,...,x_t,\pi,dx),$$

$(x_0,...,x_t,\pi) \in X^{t+1} \times U_{t+1}$ is continuous. Now if $f_1,...,f_m \in C(X)$, $g_1,...,g_m \in C(X^{t+1})$ then

$$E_n \prod_{\ell=1}^{m} f_\ell\big(X_\ell(t+1)\big)g_\ell\big(X_\ell(0),..,X_\ell(t)\big)$$

$$= E_n \prod_{\ell=1}^{m} g_\ell\big(X_\ell(0),..,X_\ell(t)\big)\int f_\ell(x)\gamma^{(t)}\big(X_\ell(0),..,X_\ell(t),p_n^{(t)},dx\big)$$

$$\to \prod_{\ell=1}^{m} \int g_\ell(x_0,..,x_t)\big(\int f_\ell(x)\gamma^{(t)}(x_0,..,x_t,\pi^{(t)},dx)\big)d\pi^{(t)}(x_0,..,x_t)$$

Defining the probability measure $\pi^{(t+1)}$ on (x^{t+2}, G^{t+2}) by

$$\pi^{(t+1)}(H\times G) = \int_H \gamma^{(t)}(x_0,..,x_t,\pi^{(t)},G)\,d\pi^{(t)}(x_0,..,x_t)$$

and using results in the proof of theorem 15 we see that the joint distribution of $\{X_\ell(s), s\leqslant t+1\}$, $\ell = 1,..,m$ converges weakly to $(\pi^{(t+1)})^m$.

From the definition of $\pi^{(t+1)}$ we see that regular conditional probabilities for one time given previous times exist. Thus by theorem (c) of Ionescu (1949) there is a probability measure $\pi^{(T)}$ on (X^T, G^T) such that

$$\pi^{(T)}\{x\in X^T \mid (x(0),..,x(t))\in G\} = \pi^{(T)}(G), \quad G\in G^{t+1}, \ t\in T,$$

that is the $\pi^{(t)}$ are finite dimensional distributions for $\pi^{(T)}$. It is now clear from Billingsley (1968, p.22, problem 7) that the joint distribution of $X_1,..,X_m$ converges weakly to $(\pi^{(T)})^m$.

Perhaps the greatest weakness in the conditions 1. and 2. is that the conditional probability does not depend directly on n but only indirectly through $p_n^{(t)}$. A useful generalisation might replace $\gamma^{(t)}$ by $\gamma_n^{(t)}$ where $\gamma_n^{(t)} \to \gamma^{(t)}$ in some sense.

4. FINDING MARKOV PROCESS MODELS

The general discussion of chapter 3 was concerned with the joint structure of the movement processes for a population of animals. The models of that chapter were structural or qualitative as were the conclusions derived from them. To answer more detailed questions about movements the movement processes must be specified more precisely. The purpose of this chapter is to look for ways of generating suitable homogeneous Markov processes from which more detailed models can be built using the construction of section 3.5.

Markov processes are frequently suggested as models for animal movements. If Y is a stochastic process in discrete time giving the movements of an animal we can speak of $Y(1) - Y(0)$, $Y(2) - Y(1), \ldots$ as "steps". Markov process models for animal movements have their origin in the idea of a random walk where each step is independent of previous steps. In non homogeneous space the appropriate generalisation is that the next step depends on previous steps only through the current position which means that the process is Markov. Skellam (1973) explains the origin of Markov process models from random walk ideas, and examples are given by Wilkinson (1952), Saila and Shappy (1963), Siniff and Jessen (1969).

A homogeneous Markov process is about the simplest process imaginable for animal movements. Generally a Markov model ignores random inhomogeneities in time and differences between individuals. In the latter respect it essentially conditions on the phenotype. The theory of section 3.5 suggests ways of constructing more general types of processes.

Since the transition probabilities, together with an initial distribution, specify the distribution of a Markov process it is the transition probabilities or the operator derived from the transition probabilities, with which we shall be concerned. In the search for Markov transition probabilities and methods of generating them we keep in mind that the purpose is to develop models. We are not looking for "truth" but reasonable description. It is desirable to obtain the transition probabilities in a form which enables them to be

1. interpreted in the biological context,
2. modified in the light of further knowledge since this will almost always be necessary,
3. used for estimation of unknown parameters and testing the model,
4. used for calculating probabilities and means of functions associated with the movement process, which are of interest.

This chapter concentrates on criterion 2, because animal movement can be very complicated especially in a field situation where inhomogeneities in the environment are important. This is not to say that we shall seek to take account of every detail but just to maintain sufficient flexibility in modelling to accommodate the more important details.

As part of criterion 2 it is of particular interest to obtain the transition probabilities in a form which facilitates the application of the model constructed in section 3.5. We are not particularly concerned that we sometimes obtain transition operators only applicable to discrete time processes since it is possible to develop a discrete time version of the model of 3.4, although admittedly the discrete version will sometimes be clumsy.

4.1 Self adjoint transition operators.

In the application of the discrete space models it is useful to have P in the form $G \Lambda G^{-1}$ which is a spectral decomposition of P regarded as a linear operator. In a general space setting, using the notation of 3.5, suppose that the transition measure p has an invariant measure π, that is

$$\int p(t,y,G)\,d\pi(y) = \pi(G), \tag{1}$$

which implies

$$\int \left\{\int f(x)p(t,y,dx)\right\} d\pi(y) = \int f(y)\,d\pi(y). \tag{2}$$

It is possible to proceed with an arbitrary invariant measure but since we are assuming that animals do not escape it is appropriate to impose the restriction that π is a finite measure and we can then assume that π is a probability measure. The measure π is an initial distribution for which the Markov process with these transition probabilities is stationary and in many cases π appears as a limit, in some sense, of $p(t,x,\cdot)$ as $t \to \infty$.

Let $\mathcal{L}^2(X,G,\pi) = \mathcal{L}^2(\pi)$ denote the real Hilbert space $\{f|f$ is real valued, G measurable and $\int f^2 d\pi<\infty\}$, with inner product $\langle f,g\rangle = \int fg\,d\pi$. As is customary we regard elements of $\mathcal{L}^2(\pi)$ as equal if they are equal a.s. (π). Define

$$P^t f(x) = \int f(y)p(t,x,dy) \tag{3}$$

From (1)

$$\int P^t f d\pi = \int f d\pi \tag{4}$$

Using Jensens inequality $(P^t f(x))^2 \leqslant P^t f^2(x)$ so that applying (4) with f^2 replacing f we have

$$\int (P^t f)^2 d\pi \leqslant \int f^2 d\pi.$$

Thus P^t is a bounded linear operator on $\mathcal{L}^2(\pi)$ and has norm 1 since $P^t 1 = 1$. The fact that P^t is a linear operator on $\mathcal{L}^2(\pi)$ is widely recognised, see for example Foguel (1969). We shall call an operator, derived from transition probabilities in this way, a transition operator.

With the definition P^0 = the identity I it follows from the Chapman-Kolmogorov equation (3. of section 3.5) that the family of operators P^t, $t \in T$, is a semigroup. The best known spectral decompositions of semigroups on $\mathcal{L}^2(\pi)$ are available when P^t is self adjoint. Then, from Goodrich (1972),

$$P^t = \int_{[-1,1]} \rho^t dE(\rho). \tag{5}$$

where E is a spectral measure in the sense of Halmos (1951) except that our spaces are real Hilbert spaces. When $T = [0,\infty)$, E is supported by $[0,1]$.

For simplicity we shall consider only the case where (5) reduces to

$$P^t = \sum_{m=0}^{\infty} \rho_m^t \langle\cdot,\xi_m\rangle\, \xi_m \tag{6}$$

with $\{\xi_m\}$ an orthonormal sequence in $\mathcal{L}^2(\pi)$ and $\langle\cdot,\xi_m\rangle\, \xi_m$ representing the projection operator $f \mapsto \langle f,\xi_m\rangle\, \xi_m$. We can assume $\rho_0 = 1$.

If the process is in continuous time then $\rho_m > 0$ and we can replace ρ_m by $e^{-\lambda_m}$, $\lambda_m = -\log \rho_m$. There are opportunities for generalisation of the discussion of this chapter beyond the case where (6) holds for one could consider (5) or the decompositions given by Dunford and Schwartz (1971) or generalisations of the decompositions of D.G. Kendall (1959 a and b). However the situation (6) is the only case where I know useful examples.

One might regard equation (6) as revealing how the operator works, but whether this will be biologically interpretable is another matter. It is fairly simple to see the operation of the model of section 3.5. Assuming that there is no variation in the type of process or alternatively conditioning on the type

$$E[\xi_m(X(t))\,|\,X(0)=x] = \xi_m(x)\int e^{-\lambda_m u}d\gamma(u)$$

where γ is the distribution of $A(t)$, and the subscript ℓ has been omitted. If the initial distribution is absolutely continuous with respect to π, and $f \in \mathcal{H}$, then

$$E[f(X(t))\,|\,X(0)=x] = \sum_{m=0}^{\infty}\int e^{-\lambda_m u}d\gamma(u)\ \langle f,\xi_m\rangle\ \xi_m(x).$$

The sum on the right is mean square convergent with respect to π and therefore converges in probability with respect to the initial distribution of X.

If the initial distribution of X is π then

$$E[f(X(t))\,|\,X(s)=x] = \sum\int e^{-\lambda_m u}d\gamma_{t,s}(u)\ \langle f,\xi_m\rangle\ \xi_m(x)$$

where $\gamma_{t,s}$ is the distribution of $A(t) - A(s)$, and if A has stationary increments it is not difficult to see that X is a stationary process. We can use the statistics

$$\frac{1}{M}\sum_{t=0}^{M-1}\xi_j(X(t))\xi_j(X(t+s))$$

and

$$\frac{1}{M}\int_0^M \xi_j\left(X(t)\right)\ \xi_j\left(X(t+s)\right)dt$$

as unbiassed estimators of $\int e^{-\lambda_j u}\ d\gamma_{0,s}(u)$. Chapter 6 discusses other ways in which the representation (6) is useful in estimation and hypothesis testing.

Most of the desirable properties of the representation (6) stem from the functional calculus available for operators in this form (Dunford and Schwartz, 1956). The model below provides a further demonstration of the utility of the representation (6).

Processes dependent on movements

Suppose an animal performs some activity inhomogeneously in space. If we want to know the amount of this activity that has been performed it is probably a good idea to look at movements. Alternatively we might be interested in some spatially inhomogeneous effect that the environment has on the animal. For definiteness assume we are interested in the animal's use of a resource.

If $U(t)$ is the amount of resource used by time t postulate

$$\frac{d}{dt}\ U(t) = f\left(X(t)\right) , \tag{7}$$

f is real valued and G measurable. The model essentially assumes that the resource does not become more difficult to find locally, because of use, and also the animal does not alter its efforts to obtain more of the resource. Such a model might apply in some situations when the animal constantly experiences a relative, but not absolute shortage of the resource. However it might also apply when there is no shortage but the animal is still constantly in need.

The model of section 3.5 is assumed with no variation in type. Also it is assumed that the animal has been present long enough for the initial distribution to be π and for A to have stationary increments.

Let $f \in \mathcal{L}^2(\pi)$, $\langle f, \xi_m \rangle = \beta_m$. Since the initial distribution is stationary

$$E[f(X(t))|A] = \int f d\pi = \beta_0$$

Define $\tilde{f} = f - \beta_0$. When $t \geqslant s$

$$C(\tilde{f}(X(s)), \tilde{f}(X(t))|A) = E\{\tilde{f}(X(s))E[\tilde{f}(X(t))|X(s),A]|A\}$$

$$= \int \tilde{f} \sum_{m=1}^{\infty} \beta_m \xi_m \exp\{-\lambda_m(A(t)-A(s))\}d\pi$$

$$= \sum_{m=1}^{\infty} \beta_m^2 \exp\{-\lambda_m(A(t)-A(s))\}.$$

If φ_{t-s} is the Laplace transform of $A(t)-A(s)$ we obtain

$$C(f(X(s)), f(X(t))) = \sum_{m=1}^{\infty} \varphi_{t-s}(\lambda_m)\beta_m^2$$

using the fact that $E[f(X(t))|A] = Ef(X(t))$ a.s. If A is continuous in probability it follows that $\varphi_h(\lambda_m)$ is a continuous function of h. Now $|\varphi_{r-s}(\lambda_m)| \leqslant 1$ and since $\sum_{m=1}^{\infty} \beta_m^2 < \infty$ the dominated convergence theorem applies to the sum $\sum_{m=1}^{\infty} \varphi_h(\lambda_m)\beta_m^2$ showing that it is continuous in h. It now follows $f(X(t))$ is continuous in mean square and we can define

$$U(t) = \int_0^t f(X(u))du$$

where the integral is interpreted as mean square convergent.

We have

$$E\ U(t) = t\beta_0$$

and

$$C\left(U(s),U(t)\right) = \int_0^t \int_0^s \sum_{m=1}^{\infty} \varphi_{|u-v|}(\lambda_m)\beta_m^2 du dv$$

$$= \sum_{m=1}^{\infty} \beta_m^2 \int_0^t \int_0^s \varphi_{|u-v|}(\lambda_m) du dv$$

The assumption that $U(t)$ is a deterministic function of X is somewhat unrealistic. To include some random variation we can model the conditional moments of $U(t)$ as

$$E[U(t)|X] = \int_0^t f\left(X(u)\right)du$$

and

$$C\left(U(s),U(t)|X\right) = \int_0^s g\left(X(u)\right)du \quad , \quad s \leqslant t,$$ where

the integrals are assumed to exist at least in mean square. The covariance structure represents the situation where the rate of resource use is independent at every moment, that is, conditional on X, U has independent but non stationary increments. Essentially the variation in use of resource is local in time and is also local in space if the animal does not remain in one place for long periods. Let β_m be defined in terms of f as before. We obtain

$$E\ U(t) = t\beta_0$$

and

$$C\left(U(s),U(t)\right) = s\int g d\pi + \sum_{m=1}^{\infty} \int_0^t \int_0^s \varphi_{|u-v|}(\lambda_m)\beta_m^2 du dv.$$

It is not difficult to extend these results to obtain means and covariances for a total $U(t) = \sum_{\ell=1}^{n} U_\ell(t)$ of the resource use of n animals in the same space. Consider just a single subpopulation with dependence between individuals only appearing through dependence between the A_ℓ. We have

$$E\ U(t) = nt\beta_0$$

and $$C\left(U(s),U(t)\right) = \sum_{\ell,\ell'} C\left(U_\ell(s),U_{\ell'}(t)\right).$$

Now, conditional on A_ℓ and $A_{\ell'}$, the processes X_ℓ and $X_{\ell'}$ are independent and so U_ℓ and $U_{\ell'}$ are conditionally independent. Thus $C(U_\ell(s),U_{\ell'}(t)\,|\,A_\ell A_{\ell'}) = 0$ for $\ell \neq \ell'$. Furthermore $E[U_\ell(t)\,|\,A_\ell,A_{\ell'}] = t\beta_0$ a.s. and so from the expression

$$C(U_\ell(s),U_{\ell'}(t))$$
$$= EC(U_\ell(s),U_{\ell'}(t)\,|\,A_\ell,A_{\ell'}) + C(E[U_\ell(s)\,|\,A_\ell,A_{\ell'}],E[U_{\ell'}(t)\,|\,A_\ell,A_{\ell'}])$$

we conclude $C(U_\ell(s),U_{\ell'}(t)) = 0$. Hence

$$C(U(s),U(t)) = nC(U_\ell(s),U_\ell(t)).$$

There are many ways that the situation can be made more complicated preserving the kind of results obtained above.

In the next section we look at ways of obtaining Markov transition operators in the form (6) from diffusion processes.

4.2 Models based on diffusion processes.

By a diffusion process I mean a continuous time homogeneous Markov process for which the state space X is a subset of R^d, $d = 1,2,$ or 3, the sample paths are continuous functions, and the transition probabilities satisfy the backward differential equation

$$\frac{\partial}{\partial t}\hat{p}(t,x,y) = \tfrac{1}{2}\sum_{i,j}\sigma_{ij}(x)\frac{\partial^2}{\partial x_i\,\partial x_j}\hat{p}(t,x,y) + \sum_i \mu_i(x)\frac{\partial}{\partial x_i}\hat{p}(t,x,y) \quad (8)$$

where $\hat{p}(t,x,\cdot)$ is the density of the transition measure $p(t,x,\cdot)$ with respect to Lebesgue measure. The operator

$$S = \tfrac{1}{2}\sum_{i,j}\sigma_{ij}(x)\frac{\partial^2}{\partial x_i\,\partial x_j} + \sum_i \mu_i(x)\frac{\partial}{\partial x_i}$$

is called the differential generator of the process. We shall consider S as an operator on the space $C^2(X)$ of real valued functions which are twice continuously differentiable and satisfy the reflecting barrier condition (KEILSON ~~Kielson~~, 1965) $n(x)'\Sigma(x)\nabla f(x) = 0$, for x belonging to the boundary of X, $n(x)$ being the normal to the boundary at x and ∇ the vector valued operator $\left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_d}\right)$. Thus we restrict consideration to processes with reflection at the boundary described by this condition on the domain of S. Stroock and Varadhan (1971) consider a more general class of boundary conditions.

Diffusion processes have a strong intuitive appeal for modelling animal movements because they have continuous sample paths and can be obtained as limits of random walks. The parameters of the process $\Sigma(x) = (\sigma_{ij}(x))$ and $\mu(x) = (\mu_1(x), \ldots, \mu_d(x))'$ have a very natural interpretation as local means and variances. Defining

$$X_{\varepsilon,x}(t) = X(t)\, I_{(-\varepsilon,\varepsilon)}(|X(t)-x|)$$

we have $\mu(x) = \lim_{t \downarrow 0} \frac{1}{t} E[X_{\varepsilon,x}(t)-x|X(0)=x]$

and

$$\Sigma(x) = \lim_{t \downarrow 0} \frac{1}{t} V(X_{\varepsilon,x}(t)|X(0)=x).$$

However problems arise in applications because the differential equation (8) is very difficult to solve in all but the simplest cases. Numerical methods are available (e.g. Ames (1969)) but those methods which present the solution simply as arrays of numbers corresponding to values of $\hat{p}(t,x,y)$, at a finite number of points, are not really very useful for our purpose because it is difficult to extract structural and analytical information from such solutions. Later we shall consider more informative numerical methods.

For a large class of diffusion processes a decomposition of the form (6) is available. We use the theory of ~~Kielson~~ KEILSON (1964, 1965). ~~Kielson~~ KEILSON imposes regularity conditions on S and X which are stated below in a slightly specialised form

1. X is a closed bounded region of R^d with a simple boundary ∂X which has a continuously turning tangent hyperplane.
2. $\Sigma(x)$ and $\mu(x)$ are analytic functions of $x_1,..,x_d$ at all points $x = (x_1,..,x_d) \in X$.
3. $\Sigma(x)$ is positive definite.
4. The vector $A(x) = \Sigma^{-1}(x)[2\mu(x) - \sum_i \underset{\sim}{e}_i \sum_j \frac{\partial}{\partial x_j} \sigma_{ij}(x)]$ is irrotational, that is $\frac{\partial}{\partial x_i} A_j(x) = \frac{\partial}{\partial x_j} A_i(x)$.

As noted by ~~Kielson~~ KEILSON condition 2. can be weakened substantially. However rather than search for the weakest conditions under which ~~Kielson~~ KEILSON's results hold we shall simply note that Stroock and Varadhan (1971) required very much weaker conditions than 2. in their construction of diffusion processes with boundary conditions.

Condition 4 entails that there exists a real valued function U on X such that $\nabla U = A$

$$\hat{\pi}(x) = e^{U(x)} / \int_X e^{U(x)} dx$$

Letting π be the probability measure having density $\hat{\pi}$ with respect to Lebesgue measure then $C^2(X)$ can be thought of as a subset of $\mathcal{L}^2(\pi)$ and it is also easy to see that it is dense in $\mathcal{L}^2(\pi)$. Thus S is an operator on $\mathcal{L}^2(\pi)$, and, using the conditions above, ~~Kielson~~ KEILSON has shown that S is self adjoint. Thus the semi group P^t of transition operators on $\mathcal{L}^2(\pi)$ generated by S, and arising from the solution of (8), is self adjoint and so satisfies equation (5). In many cases we would also expect the discrete decomposition (6) to pertain. This is the case for all one dimensional processes satisfying the conditions above (~~Kielson~~ KEILSON, 1964) and for some general examples given by ~~Kielson~~ KEILSON (1965).

Note that KEILSON ~~Kielson~~ started with the Kolmogorov forward differential equation and derived an operator L which we can define as

$$Lf(x) = \pi(x)Sf(x)$$

by using the conditions 1. to 4. above. The results in KEILSON ~~Kielson~~'s paper are expressed in terms of L but it seems more natural to work with the differential generator S.

When the representation (6) is available for P^t KEILSON ~~Kielson~~ shows that the sequence $\beta_n = \log \rho_n$ and the sequence ξ_n may be characterised as

$$\beta_n = \sup \langle Sf,f\rangle = \langle S\xi_n, \xi_n\rangle$$

where the supremum is taken over all $f \in C^2(X)$ such that f is orthogonal to $\xi_0, \ldots, \xi_{n-1}$ and $\langle f,f\rangle$ $(=|f|) = 1$. Note that $\xi_0 \equiv 1$, $\rho_0 = 1$, and $\xi_n \in C^2(X)$. KEILSON ~~Kielson~~ suggests the use of this characterisation as a numerical method for estimating the β_n and ξ_n by maximising over finite dimensional subspaces rather than the whole space $C^2(X)$. This is the basis of the now popular Rayleigh-Ritz method of estimating eigenvalues and eigenfunctions (e.g. Weinburger, 1974).

In the Rayleigh-Ritz method one chooses a finite number M, say, of linearly independent functions $f_1, \ldots, f_M$ in an appropriate subspace of $\mathcal{L}^2(\pi)$ and the eigenvalue-eigenfunction problem

$$S\xi = \beta\xi$$

is replaced by the eigenvalue-eigenvector problem

$$S_M \underset{\sim}{v}^{(M)} = \beta^{(M)} A_M \underset{\sim}{v}^{(M)}$$

where $\underset{\sim}{v}^{(M)} \in R^M$, S_M is the $M \times M$ matrix $(\langle Sf_i, f_j\rangle)$ and A_M is the matrix $(\langle f_i, f_j\rangle)$. Weinberger (1974) shows that the estimates $\beta_n^{(M)}$ satisfy $\beta_n^{(M)} \leqslant \beta_n$. He also discusses the choice of the f_i and the accuracy of the estimates $\beta_n^{(M)}$ and $\xi_n^{(M)} = \sum_1^M v_{i\,n}^{(M)} f_i$, $\underset{\sim}{v}_n^{(M)}$ being the eigenvector corresponding to the eigenvalue $\beta_n^{(M)}$.

The numerical procedure for estimating eigenvalues and eigenfunctions of linear operators, known as the finite element method, is the Rayleigh-Ritz method with a special choice of the functions $f_1, f_2, \ldots, f_M$. Strang and Fix (1973) give an introductory account of this method.

The fundamental assumption involved in the discussion above is that A is irrotational. Next we consider a number of examples of models based on postulated animal behaviour where this assumption holds.

Markovian kineses.

Fraenkel and Gunn (1940) set up a classification of animal movement behaviour on the basis of the kind of movement response to environmental stimuli. They define a kinesis as any undirected movement response to a stimulus, that is, the stimulus simply causes a change in the rate of some process associated with movement. Fraenkel and Gunn recognise two different kinds of response.

1. Orthokineses: the stimulus causes a change in speed of movement, but not direction.
2. Klinokinesis: the stimulus causes a change in the rate of turning.

We shall model the stimulus as changing continuously over the space so that if an animal is at a point x in the space it receives a stimulus which is a function of that point. We shall not consider the property of accommodation sometimes associated with klinokinesis because it is essentially non-Markovian in character unless accommodation takes place very quickly.

For a Markov process in continuous time, speed and rate of turning are not necessarily defined as functions of the sample paths. However this does not particularly matter since continuous time Markov processes are only expected to be valid as asymptotic approximations to other processes. In the latter processes speed and rate of turning may have reasonable definitions, see, for example, the models of chapter 5. The transition probabilities of the Markov processes will be used to interpret the idea of kinesis. The methods of chapter 5 can be used to construct distinct models of orthokinesis and klinokinesis.

Suppose $\mathcal{X}$ is a subset of R^d (d=1,2, or 3). If $B \subset \mathcal{X}$, $x \in \mathcal{X}$ and H is an orthogonal matrix define $x + H(B-x)$ to be the set $\{x+H(y-x) \mid y \in B\}$. If an animal's movements are governed by kinesis one expects that

$$P\left(X(t) \in x+H(B-x) \mid X(0)=x\right)$$

should not depend very much on H for t small, that is, when the animal has not had time to move very far, all directions are approximately equivalent. This idea is expressed precisely in the following definition.

Definition

A homogeneous Markov process X is a kinesis if for every x in the interior of $\mathcal{X}$ and every bounded continuous f

$$\lim_{t \downarrow 0} \frac{1}{t} \left\{ \int f\left(x+H(y-x)\right) \, p(t,x,dy) - f(x) \right\} \tag{8}$$

does not depend on the orthogonal matrix H whenever it exists for some orthogonal H.

If X is a diffusion process then (8) exists for $H = I$ whenever $f \in C^2(X)$ and the value of the limit is then $Sf(x)$. For arbitrary H and $f \in C^2(X)$ the value of (8) is

$$\mu'(x)H'\nabla f(x) + H\Sigma(x)H' \cdot \nabla\nabla' f(x)$$

where $\nabla\nabla'$ is the matrix valued operator $\left(\frac{\partial^2}{\partial x_i \, \partial x_j}\right)$ and the notation $A \cdot B$ means $\sum_{i,j} a_{ij} b_{ij}$, for two matrices $A = (a_{ij})$, $B = (b_{ij})$. Since the value of (8) does not depend on H for all $f \in C^2(X)$ we must have $H\mu(x)$ and $H\Sigma(x)H'$ independent of H which can only be true if $\mu(x) = 0$ and $\Sigma(x) = \sigma^2(x)I$. In this case

$$A = -\frac{1}{\sigma^2(x)} \nabla\sigma^2(x)$$

$$= -\nabla \log \sigma^2(x).$$

Thus A is irrotational with $U = -\log \sigma^2(x)$.

Movement in response to a potential

Kiester and Slatkin (1974) describe a deterministic model for movement of iguanid lizards in a one dimensional habitat. The lizards move in such a way as to maximise free time after performing essential activities, feeding and interacting with conspecifics. This free time is space dependent. As is our custom we will ignore the interaction or assume that there are sufficient animals present for the interactions to be essentially deterministic in the sense of section 3.6. The basic idea of Kiester and Slatkins' model generalises to the case where direction and rate at which an animal tends to move is determined by the relative attractiveness of different places which is described by a potential function h. There is no need to assume that maximisation of free time is involved but it is a possible mechanism.

A plausible diffusion process model for animal moving in response to a potential function h is

1. $$\mu(x) = c\nabla h(x)$$

If the local variance is not affected by the potential we have

2. $$\Sigma(x) = \sigma^2 I$$ where σ^2 is a positive constant.

Now $A(x) = 2c\sigma^{-2}\nabla h(x)$ which is clearly irrotational.

Rosen (1973) has proposed a diffusion model for chemotactic bacteria which agrees with 1. and 2. In his case $h = \log s$ where s is the concentration of the chemotactic agent.

Passive movement by currents.

Suppose that the animals are carried passively by water or air currents and their own movements, which are non directional, are superimposed on the displacement due to the current and do not depend on position in space. An appropriate diffusion model is $\mu(x) = V(x)$, V being the velocity of the current and $\Sigma(x) = \sigma^2 I$, for some constant σ^2. For this model the vector A is irrotational if and only if the current is irrotational.

4.3 Other methods of generation and modification of self adjoint transition operators.

In this section we look at other ways of generating Markov transition operators in the form (6). The methods are not exhaustive and are intended primarily to indicate the kinds of methods one might employ. In many situations it will be appropriate to employ a combination of these methods.

Bivariate distributions and Markov processes.

Markov processes can be obtained from bivariate distributions and there is a relationship between the theories of bivariate distributions and Markov transition operators.

Let X be a Borel subset of R^d, $d = 1,2,$ or 3 and G those elements of B^d, the Borel sets in R^d, which are subsets of X. With this restriction, (X,G) is a Borel space in the sense of Breiman (1968, p.79) and we can use this fact to infer the existence of regular conditional distributions. The possibilities for (X,G) include all measurable spaces which are likely to be of interest in the study of animal movements. Parthasarathy (1967) provides the necessary theory for a more general presentation.

If τ is a probability measure on (X^2,G^2) we shall say that τ is a bivariate distribution for the random variables $Y_0,Y_1 : X^2 \to X$ defined by

$$(x_0,x_1) = \big(Y_0(x_0,x_1),Y_1(x_0,x_1)\big)$$

for all $(x_0,x_1) \in X^2$. Given τ we can use the theory in Breiman (1968, pp.77-79) to infer the existence of a function $q : X \times G \to [0,1]$ which is a regular conditional probability measure for Y_1 given Y_0 as defined in the introduction to chapter 2. We can use q to define Markov transition probabilities $p(t,x,G)$, $t \in T = \{0,1,\ldots\}$, as follows

$$p(0,x,G) = I_G(x)$$
$$p(t,x,G) = \int p(t-1,y,G)\ q(x,dy).$$

By well known results (e.g. Revuz, 1974, p.16) p is a transition measure of a Markov process with state space $(\mathcal{X},\mathcal{G})$. The transition operator is given by

$$P = \int f(y)q(x,dy)$$

On the other hand let X be a homogeneous Markov process, $T = \{0,1,\ldots\}$, with state space $(\mathcal{X},\mathcal{G})$. The joint distribution of $\left(X(0),X(1)\right)$ on the space $(\mathcal{X}^2,\mathcal{G}^2)$ is a bivariate distribution for (Y_0,Y_1).

Thus we can obtain bivariate distributions from Markov processes and Markov processes from bivariate distributions. Two questions arise, first can the theory of bivariate distributions be used to advantage in the study of Markov processes and secondly does the relationship between bivariate distributions and Markov processes provide a useful method of constructing Markov transition operators? These questions motivate the present discussion. The relationship between bivariate distributions and Markov processes has long been recognised for example Sarmanov (1961) who was also interested in discrete spectral representations.

We need some of the theory of bivariate distributions as presented by Chesson (1976). Let Y_0 and Y_1 have the common marginal distribution π and define $\mathcal{H}$ to be the real Hilbert space $\{(f,g)\,|\,f,g\in\mathcal{L}^2(\pi)\}$ with inner product $\langle(f_1,g_1),(f_2,g_2)\rangle = \frac{1}{2}\int f_1f_2 + g_1g_2\,d\pi$. The following is a special case of a theorem given by Chesson.

Theorem 1.

There is a unique family of subspaces M_ρ of $\mathcal{H}$, $0 \leqslant \rho \leqslant 1$ such that

1. $\bigcap_{\rho>\rho'} M_\rho = M_{\rho'}$, $0 \leqslant \rho' < 1$ and $M_0 = \{0\}$.
2. If $\{(\xi_t^\rho, \eta_t^\rho)\}_{t \in T_\rho}$ is an orthonormal basis for M_ρ then $\{\xi_t^\rho\}_{t \in T_\rho}$ and $\{\eta_t^\rho\}_{t \in T_\rho}$ are orthonormal families of functions on the space $\mathcal{L}^2(\pi)$
3. If $(f,g) \in \mathcal{H}$ and $\{(\xi_t^\rho, \eta_t^\rho)\}_{t \in T_\rho}$ as in 2.

$$Ef(Y_0)g(Y_1) = \int_{(0,1]} \rho \, dQ(\rho),$$

where $Q(\rho) = \sum_{t \in T_\rho} (\int f\xi_t^\rho d\pi)(\int g\xi_t^\rho d\pi)$.

In the present context we are only interested in the case of discrete spectra, that is, $\{M_\rho, 0 \leqslant \rho \leqslant 1\}$ consists of countably many distinct subspaces, and generally only those cases where the orthogonal complement $M_1 \ominus M_\rho$ is finite dimensional for $\rho > 0$. In this case there is a sequence of positive real numbers $\{\rho_m\}$ and orthonormal sequences $\{\xi_m\}$, $\{\eta_m\}$ in $\mathcal{L}^2(\pi)$ such that

$$Ef(Y_0)g(Y_1) = \sum_{m=0}^{\infty} \rho_m \int f\xi_m d\pi \int g\eta_m d\pi \tag{9}$$

from which it follows that

$$E[g(Y_1)|Y_0] = \sum_{m=0}^{\infty} \rho_m \xi_m(Y_0) \int g\eta_m d\pi .$$

Expression (9) will be referred to as the canonical decomposition of (Y_0, Y_1) or their distribution.

Theorem 2.

The operator P obtained from the bivariate DISTRIBUTION for Y_0, Y_1 is self adjoint and has spectral decomposition of the form (6) if and only if the joint distribution of Y_0 and Y_1 is symmetric and satisfies (9).

Proof

Let P be self adjoint and take the form (6) then $\mathcal{E}f(Y_0)g(Y_1) = \sum_{m=0}^{\infty} \rho_m \int f\xi_m d\pi \int g\xi_m d\pi$. Clearly (Y_0,Y_1) have a symmetric distribution and also

$$\mathcal{E}f(Y_0)g(Y_1) = \sum_{m=0}^{\infty} |\rho_m| \int f\xi_m d\pi \int g\{\rho_m/|\rho_m|\xi_m\} d\pi$$

which is of the form (9).

If (Y_0,Y_1) are symmetrically distributed

$$\langle Pf,g\rangle = \langle f,Pg\rangle$$

so that P is self adjoint. If (9) is satisfied also $P\eta_m = \rho_m\xi_m$ and by symmetry $P\xi_m = \rho_m\eta_m$ implying $\eta_m = \rho_m/|\rho_m|\xi_m$. Thus

$$P = \sum_{m=0} \rho_m/|\rho_m| \langle\cdot,\xi_m\rangle \xi_m .$$

The problem we are left with is when does the bivariate distribution have the form (9). Lancaster (1969) gives the condition "φ^2-boundedness" which says that the Radon-Nikodym derivative $d\tau/d\pi^2$ is square summable relative to π^2. This is perhaps the easiest condition to check, but we shall come across examples where (9) is true but τ is not φ^2-bounded.

A criterion implying that a bounded linear operator has a discrete spectrum, which also converges to 0, is compactness. We shall interpret this criterion in terms of bivariate distributions.

Let $\mathcal{H}$ be a Hilbert space, real or complex, then one of the many equivalent definitions of compactness of an operator A is as follows (Helmberg, 1969, p.187).

Definition

A bounded linear operator A is compact if, whenever $\{f_m\}$ is a sequence in $\mathcal{H}$ such that $\{f_m,g\} \to 0$ for all $g \in \mathcal{H}$, then $\langle Af_m,Af_m\rangle \to 0$ as $m \to \infty$.

If A is a positive operator, that is A is self adjoint and $\langle Af,f\rangle \geqslant 0$ for all $f \in \mathcal{H}$, then $A^{\frac{1}{2}}$ exists and is compact if and only if A is compact. Thus in the definition of a compact operator we can replace $\langle Af_m,Af_m\rangle \to 0$ by $\langle Af_m,f_m\rangle \to 0$. This holds for general bounded self adjoint operators $A = \int_{[a,b]} \lambda dE(\lambda)$ by consideration of the positive operators $\int_{[0,b]} \lambda dE(\lambda)$ and $-\int_{[a,0]} \lambda dE(\lambda)$. The interpretation of these results for bivariate distributions is as follows.

Theorem 3.

If (Y_0,Y_1) have a symmetric distribution then representation (9) is available, with $\rho_m \to 0$ as $m \to \infty$, if and only if every sequence $\{f_m\}$ in $\mathcal{L}^2(\pi)$ satisfying

1. $\int f_m g d\pi \to 0$ as $m \to \infty$, for every $g \in \mathcal{L}^2(\pi)$, also satisfies
2. $\mathcal{E} f_m(Y_0) f_m(Y_1) \to 0$ as $m \to \infty$

Proof

The symmetry of the distribution of (Y_0,Y_1) implies that P is self adjoint and when 1. implies 2. P is compact. Thus P has the representation (6) with $\rho_m \to 0$ as $m \to \infty$. On the other hand if (9) holds with $\rho_m \to 0$ then P is compact and 1. implies 2.

We have reduced the problem of finding discrete time Markov transition operators of the form (6) to the problem of finding symmetric bivariate distributions with discrete canonical decomposition. The canonical decomposition is known for many bivariate distributions. Lancaster (1969) gives examples and an extensive bibliography. More recently Griffiths (1970) and Tyan and Thomas (1975) have looked at conditions on the $\{\rho_m\}$ $\{\xi_m\}$ and π for the expansion (9) to define a bivariate distribution.

The following example of a bivariate distribution which yields useful Markov transition operators is due to Lancaster (1969, p.192).

Let Φ denote the normal measure on R^1 with mean 0 and variance 1. H_m denotes the mth Hermite polynomial normalised so that $\int H_m^2 d\Phi = 1$. If Z_0 and Z_1 are one dimensional random variables having a bivariate normal distribution with mean $\underset{\sim}{0}$ and variance

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

it is well known that

$$Ef(Z_0)g(Z_1) = \sum_{m=0}^{\infty} \rho_m \int H_m f d\Phi \int H_m g d\Phi \tag{11}$$

Now let $X = R^d$. Suppose (Y_0, Y_1) are symmetrically distributed as a 2d dimensional normal distribution. Write

$$Y_0 = (Y_{01}, \ldots, Y_{0d}), \quad Y_1 = (Y_{11}, \ldots, Y_{1d}),$$

where the Y_{ij} are 1 dimensional random variables. There is an affine transformation $T : R^d \to R^d$ such that $(TY_0, TY_1) = (Z_{01}, \ldots, Z_{0d}, Z_{11}, \ldots, Z_{1d})$ is a 2d dimensional normal with mean 0 and variance matrix

$$\begin{bmatrix} I_d & B \\ B & I_d \end{bmatrix}, \quad B = \text{diag}\,(\beta_1, \ldots, \beta_d)$$

T is the transformation to the classical normal canonical variables. It follows that the pairs (Z_{0j}, Z_{1j}), $j = 1, \ldots, d$ are independent and (Z_{0j}, Z_{1j}) has variance matrix

$$\begin{bmatrix} 1 & \beta_j \\ \beta_j & 1 \end{bmatrix}.$$

Using the next theorem we shall show that

$$Ef(TY_0)g(TY_1) = \sum_{m=0}^{\infty} \rho_m \int \tilde{\xi}_m f d\Phi_d \int \tilde{\xi}_m g d\Phi_d \tag{12}$$

where $\{\rho_m\}$ is a rearrangement of the d-sequence

$$\{\beta_1^{j_1} \ldots \beta_d^{j_d}\ ;\ j_i = 0,1,\ldots;\ i=1,\ldots,d\}$$

so that $|\rho_m|$ is non increasing. $\{\tilde{\xi}_m(z_1,\ldots,z_d)\}$ is the corresponding rearrangement of $\{H_{j_1}(z_1)\ldots H_{j_d}(z_d)\}$, and Φ_d is the product measure $\Phi\times\ldots\times\Phi$. In theorem 5, $\varphi\times\psi$ denotes the mapping $(x_1,x_2)\mapsto\varphi(x_1)\psi(x_2)$ for $\varphi : \mathcal{X}^{(1)}\to R$, $\psi : \mathcal{X}^{(2)}\to R$.

Theorem 5.

Let $Y_0^{(1)}, Y_1^{(1)}, Y_0^{(2)}, Y_1^{(2)}$ be random variables such that $Y_j^{(i)}$ takes its values in $(\mathcal{X}^{(i)}, \mathcal{G}^{(i)})$ and has distribution $\pi^{(i)}$, $j = 0,1$, $i = 1,2$. Furthermore assume $(Y_0^{(1)}, Y_1^{(1)})$ and $(Y_0^{(2)}, Y_1^{(2)})$ independent, and $(Y_0^{(i)}, Y_1^{(i)})$ have the canonical decomposition

$$\mathcal{E}f(Y_0^{(i)})g(Y_1^{(i)}) = \sum_{m=0}^{\infty} \rho_m^{(i)} \int \xi_m^{(i)} f d\pi^{(i)} \int \eta_m^{(i)} g d\pi^{(i)} .$$

Then $\left((Y_0^{(1)}, Y_0^{(2)}), (Y_1^{(1)}, Y_1^{(2)})\right)$ has canonical decomposition

$$\mathcal{E}f(Y_0^{(1)}, Y_0^{(2)})g(Y_1^{(1)}, Y_1^{(2)})$$
$$= \sum_{m,n} \rho_m^{(1)} \rho_n^{(1)} \left(\int \xi_m^{(1)} \times \xi_n^{(2)} d\pi^{(1)} \times \pi^{(2)}\right)\left(\int \eta_m^{(1)} \times \eta_n^{(2)} d\pi^{(1)} \times \pi^{(2)}\right) \tag{13}$$

Proof

Let $\mathcal{X} = \mathcal{X}^{(1)} \times \mathcal{X}^{(2)}$, $\mathcal{G} = \mathcal{G}^{(1)} \times \mathcal{G}^{(2)}$, $\pi = \pi^{(1)} \times \pi^{(2)}$

We use the definition of $\mathcal{H}$ given before theorem 1. If $\{\varphi_c^{(i)}\}_{c\in C_i}$, $\{\psi_c^{(i)}\}_{c\in C_i}$ are orthonormal bases for $\mathcal{L}^2(\pi^{(i)})$, $i = 1,2$, then $\{\varphi_{c_1}^{(1)} \times \varphi_{c_2}^{(2)}\}_{(c_1,c_2)\in C_1\times C_2}$ and $\{\psi_{c_1}^{(1)} \times \psi_{c_2}^{(2)}\}_{(c_1,c_2)\in C_1\times C_2}$ are orthonormal bases for $\mathcal{L}^2(\pi^{(1)} \times \pi^{(2)})$. Furthermore, if $(f,g) \in \mathcal{H}$

$$\mathcal{E}f(Y_0^{(1)}, Y_0^{(2)})g(Y_1^{(1)}, Y_1^{(2)})$$
$$= \Sigma\left(\int f\varphi_{c_1}^{(1)} \times \varphi_{c_2}^{(2)} d\pi^{(1)} \times \pi^{(2)}\right)\left(\int g\psi_{c_1'}^{(1)} \times \psi_{c_2'}^{(1)} d\pi^{(1)} \times \pi^{(2)}\right) \times$$
$$\mathcal{E}\varphi_{c_1}^{(1)}(Y^{(1)})\varphi_{c_2}^{(2)}(Y^{(2)})\psi_{c_1'}^{(1)}(Y^{(1)})\psi_{c_2'}^{(2)}(Y^{(2)})$$

where the summation is over $(c_1, c_2), (c_1', c_2') \in C_1 \times C_2$. Now $\{\xi_m^{(i)}\}$ $\{\eta_m^{(i)}\}$ can both be completed to form orthonormal bases for $\mathcal{L}^2(\pi^{(i)})$, $i = 1,2$.

$E\xi_m^{(1)}(Y_0^{(1)})\xi_n^{(2)}(Y_0^{(2)})\eta_u^{(1)}(Y_1^{(1)})\eta_v^{(2)}(Y_1^{(2)}) = \rho_n^{(1)}\rho_m^{(2)}\delta_{mu}\delta_{nv}$, and if $\zeta_{ij} \in \mathcal{L}^2(\pi^{(i)})$, $i = 1,2$, $j = 0,1$ and if $\zeta_{i'j'}$ is orthogonal to $\{\xi_m^{(i')}\}$ or $\{\eta_m^{(i')}\}$ for some i', j' then

$$E \prod_{i,j} \zeta_{ij}(Y_j^{(i)}) = 0.$$

This proves the theorem.

The equation (12) is a direct consequence of theorem 5 and it follows that

$$Ef(Y_0)g(Y_1) = \sum_{m=0}^{\infty} \rho_m \int f\xi_m d\pi \int g\xi_m d\pi \tag{14}$$

where $\xi_m = \tilde{\xi}_m \circ T^{-1}$, π_j is the distribution of Y_j, $j = 0,1$. Lancaster justifies (14) by considering products of bivariate normal densities. Our method illustrates the use of theorem 5 which can be used generally to obtain bivariate distributions for R^d valued random variables, with known spectral decomposition, from bivariate distributions on spaces of lower dimensions.

From (14) we obtain the transition operator

$$P = \sum_{m=0}^{\infty} \rho_m < , \xi_m > \xi_m \tag{15}$$

which gives a d dimensional normal Markov process. The 2 dimensional normal distribution has been suggested as a model for the continuous space analogue of the home distribution of chapter 2 (Wierzbowska, 1972). A Markov process with transition operator (15) is an intuitively reasonable model yielding a multinormal home distribution.

Note that if ρ_m is strictly positive we can write

$$P^t = \sum_{m=0}^{\infty} e^{-\lambda_m t} <\cdot, \xi_m> \xi_m$$

where $\lambda_m = -\log \rho_m$. P^t is the transition operator for a Markov process. This process is a d dimensional Ornstein-Uhlenbeck process.

When dealing with data, sometimes we may not have a particular Markov process model in mind but simply the model that the observed movements X have the representation $X = Y(A(t))$ where Y is a homogeneous Markov process and A is a process with stationary non negative increments. When the transition operator for Y has the discrete representation (6) with $\rho_m = e^{-\lambda m}$ we have shown that the spectral decomposition of the bivariate distribution for $X(0)$, $X(t)$ is given by

$$Ef(X(0))g(X(t)) = \sum_{m=0}^{\infty} \varphi_t(\lambda_m) \int f\xi_m d\pi \int g\xi_m d\pi$$

where φ_t is the Laplace transform of $A(t)$. If $\{1\} \cup \{\psi_m\}$ is a complete orthonormal set for $\mathcal{L}^2(\pi)$ we can think about estimating the $\{\xi_m\}$ as finite linear combinations of the $\{\psi_m\}$. From THE observed sequence $X(0), X(1), \ldots, X(M)$ we can form the symmetric estimates

$$s_{ij} = \frac{1}{2M} \sum_{t=0}^{M-1} \psi_i(X(t))\psi_j(X(t+1)) + \psi_j(X(t))\psi_i(X(t+1))$$

of $E\psi_i(X(t))\psi_j(X(t+1))$. From the work of Lancaster (1969, chapter 6) it is known that ξ_m is a function ξ orthogonal to $1, \xi_1, \ldots, \xi_{M-1}$ and satisfying $\int \xi^2 d\pi = 1$ for which $E\xi(X(t))\xi(X(t+1))$ is maximal. Thus we can try to estimate the ξ_m by successive maximisation of the quadratic form

$$\sum_{ij} q_i q_j s_{ij} .$$

This is done by diagonalising S to obtain $S = H \Lambda H'$ with H orthogonal and Λ diagonal. The estimate of ξ_m is then $\Sigma h_{jm} \psi_j$ where $H = (h_{ij})$. Clearly a variety of other estimates based on these ideas are also available. The application in chapter 6 uses a generalisation for the case where π is unknown.

Transformation of the state space

Dynkin (1965, chapter 6) gives a number of methods of transforming Markov processes. One of these methods, namely state space transformation, leads to a simple transformation of the spectral decomposition of the transition operator when it is available. For our purposes it is a useful method for generating new operators from ~~exsiting~~ EXISTING operators. It would be interesting to see if Dynkin's other transformations could be used similarly in some situations.

Let $(\mathcal{X},\mathcal{G})$ and $(\mathcal{Y},\mathcal{H})$ be measurable spaces and let $\gamma : \mathcal{Y} \to \mathcal{X}$ be 1-1 and onto. Furthermore assume that γ and γ^{-1} are $\mathcal{H}/\mathcal{G}$ and $\mathcal{G}/\mathcal{H}$ measurable respectively. If Y is a homogeneous Markov process with state space $(\mathcal{Y},\mathcal{H})$ and X is defined by $X(t) = \gamma(Y(t))$, it is clear that X is homogeneous Markov process with state space $(\mathcal{X},\mathcal{G})$. If q is the transition measure for Y then P^t defined by

$$(P^t f)(x) = \int_{\mathcal{X}} f(\gamma(x)) q(t,\gamma^{-1}(x),dy)$$

is the transition operator for X. Furthermore if the transition operator Q^t, for Y, has the spectral decomposition

$$Q^t = \sum_{m=0}^{\infty} \rho_m < ,\eta_m> \eta_m$$

on the space $\mathcal{L}^2(\tau)$, τ then

$$P^t = \sum_{m=0}^{\infty} \rho_m^t < ,\xi_m> \xi_m ,$$

$\xi_m(x) = \eta_m(\gamma^{-1}(x))$, is the spectral decomposition of P^t on the space $\mathcal{L}^2(\pi)$, $\pi(G) = \tau(\gamma^{-1}(G))$, $G \in \mathcal{G}$.

Most spaces $(\mathcal{X},\mathcal{G})$ that are likely to be of interest in animal movements can be obtained by a transformation γ

of R^d, $[0,1]^d$, $(0,1)^d$, or perhaps the unit ball in R^d or its surface. Thus we can generally find a Markov transition operator in the form (6) on (X,G) by transformation of simple processes on other spaces. Furthermore there are very often many bijective bimeasurable transformations between (X,G) and $(Y,\mathcal{H})$ so that for a given transition operator Q^t a large number of transition operators P^t can be obtained. However it is generally not possible to obtain an arbitrary operator P^t from a given operator Q^t. The case of diffusion processes in one dimension, say restricted to the closed interval $[0,1]$, illustrates this point well. We use the theory in Breiman (1968, chapter 16).

For a given diffusion process $X(t)$ on $[0,1]$ there is a continuous increasing function u, unique up to a linear transformation, such that the diffusion process $u(X(t))$ has $\mu(x) = 0$. The local variance $\tilde{\sigma}^2(u)$ for the new process is expressed $\tilde{\sigma}^2(u(x)) = \sigma^2(x)\left[\frac{d}{dx}u(x)\right]^2$ where $\sigma^2(x)$ is the local variance of X. Thus there is essentially one diffusion process with zero drift that can be constructed by continuous monotonic transformation of a given diffusion process. Having chosen the local mean to be zero the shape of the local variance function is not open to choice.

If one is interested in discrete time movements and the dependence between times is not too great then for many purposes the most important property of the operator P^t is the probability measure π since this is the limiting distribution of $X(t)$, as $t \to \infty$, provided $|\rho_m| < 1$

for $m > 0$, and X will often be modelled as a stationary process with this marginal distribution. To obtain operators with a given measure π one can look for suitable transformations γ such that $\pi(G) = \tau(\gamma^{-1}(G))$ where τ is as above and Y is some known process. We can very often expect to find such transformations. For example let π be absolutely continuous with respect to Lebesgue measure on R^d. Rosenblatt (1952) shows that there is a 1 - 1 bimeasurable transformation such that $\gamma(\underset{\sim}{X}(t))$ is uniformly distributed on the unit hypercube. Using this transformation and its inverse any absolutely continuous π can be transformed into any other such π. However there is no guarantee that the transformation does not have some undesirable property.

Subordination of Markov processes

Let $Y = \{Y(t), t \geqslant 0\}$ be a homogeneous Markov process and $U = \{U_t, t \geqslant 0\}$ a process with non negative stationary independent increments, $U_0 = 0$, Y and U are assumed independent, Y has transition measure q with invariant probability measure π.

Suppose that the σ-algebra G is generated by a topology for which Y is right continuous then, as shown in section 3.5, $X(t) = Y(U_t)$ is a random variable. Furthermore X is a homogeneous Markov process with transition operator P^t defined by

$$P^t f(x) = \int Q^u f(x) dF_t(u), \tag{16}$$

Q^t being the transition operator for Y and F_t the distribution of U_t. This result is proved by Nelson (1958) but I shall present the intuition behind it.

Recall that in section 3.5 it is shown that Y can be considered as a $\mathcal{A} \times \mathcal{B}^1$ measurable function on $\Omega \times [0,\infty)$. ~~$\mathcal{A} \times \mathcal{B}^1$ measurable.~~ Thus if $G_1, \ldots, G_m \in \mathcal{G}$ the function $I_{G_1}(X(u_1,\omega)) \ldots I_{G_m}(X(u_m,\omega))$ is $\mathcal{A} \times \mathcal{B}^m$ measurable in $(\omega, u_1, \ldots, u_m) \in \Omega \times [0,\infty)^m$. It follows that the conditional distribution of $X(t_1), \ldots, X(t_m)$ given U is obtained by substituting $U_{t_i} - U_{t_{i-1}}$ for u_i in the formula for the distribution of $Y(t_1), \ldots, Y(t_m)$. Thus X is a non homogeneous Markov process with

$$P(X(t) \in G \mid X(s), U) = q(U_t - U_s, X(s), G), \quad t \geqslant s.$$

Now for $t \geqslant s$

$$\begin{aligned} P(X(t) \in G \mid X(u), u \leqslant s) &= E[P(X(t) \in G \mid X(u), u \leqslant s, U) \mid X(u), u \leqslant s] \\ &= E[q(U_t - U_s, X(s), G) \mid X(u), u \leqslant s] \end{aligned}$$

Since U has independent increments $\{X(u), u \leqslant s\}$ and $U_t - U_s$ are independent we have

$$P(X(t) \in G \mid X(u), u \leqslant s) = \int q(u, X(s), G)\, dF_{t-s}(u)$$

which means X is a homogeneous Markov process with transition operator (16). $\int \{\int q(u,x,G)\, dF_t(u)\}\, d\pi(x) = \pi(G)$ and so P^t is an operator on $\mathcal{L}^2(\pi)$.

Nelson actually proves the more general result that there is a probability space for which the definition $X(t) = Y(U_t)$ makes X a homogeneous Markov process with transition operator P^t when the right continuity of Y is replaced by $\mathcal{A} \times \mathcal{B}^1$ measurability of $p(t,x,G)$ in (t,x). The process X is called a subordinated process - X is subordinate to Y.

Although the above discussion is presented for $T = [0,\infty)$ and the index set T' for U being $[0,\infty)$ the results continue to apply whenever

$$P(U_t \in T) = 1, \quad t \in T' \tag{17}$$

so that subordination can be used to obtain continuous time processes from discrete time processes and vice versa. When $T = \{0,1,\ldots\}$ we need not be concerned with continuity requirements. In the sequel simply assume (17) holds.

The following theorem is stated by Bochner (1949)

Theorem 6

If $Q^t = \sum_{m=0}^{\infty} \rho_m^t < \, ,\xi_m> \xi_m$ and $P(U_t > 0) = 1$ then

$$P^t = \sum_{m=0}^{\infty} \int \rho_m^u dF_t(u) < \, ,\xi_m> \xi_m$$

Proof

$Q^t \xi_m = \rho_m^t \xi_m$ therefore $P^t \xi_m = \int \rho_m^u dF_t(u) \xi_m$

On the other hand if $f \in \mathcal{L}^2(\pi)$ is orthogonal to $\{\xi_m\}$ then $Q^t f = 0$ implying $P^t f = 0$. Thus

$$P^t f = \sum_{m=0}^{\infty} \int \rho_m^u dF_t(u) <f,\xi_m> \xi_m$$

for $f \in \mathcal{L}^2(\pi)$.

We see that the operator for the subordinated process has the same eigenfunctions as the original operator but that the eigenvalues are changed. In general subordination decreases the rate of decrease of the eigenvalues since if $\rho_n > \rho_m > 0$ and U_t is not degenerate Jensen's inequality implies

$$(E\rho_n^{U_t})^{\log \rho_m / \log \rho_n} < E\rho_m^{U_t}$$

whereas $(\rho_n^t)^{\log \rho_m / \log \rho_n} = \rho_m^t$. Because of this reduction in the rate of decrease the Hilbert-Schmidt property may not be preserved in subordination as we shall see in an example. However compactness is preserved when $P(U_t > 0) = 1$ because $\rho_m \to 0$ as $m \to \infty$ implies $\rho_m^{U_t} \xrightarrow{a.s.} 0$ and by the dominated convergence theorem $E\rho_m^{U_t} \to 0$.

In the case where $\mathcal{P}(U_t = 0) > 0$ assume, for simplicity, that $\mathcal{L}^2(\pi)$ is separable and $\{\tilde{\xi}_m\}$ is the sequence, possibly null, required to complete $\{\xi_m\}$ to a basis for $\mathcal{L}^2(\pi)$. We have

$$P^t = \sum_{m=0}^{\infty} (E\rho_m^{U_t}) < ,\xi_m> \xi + \mathcal{P}(U_t = 0) <,\tilde{\xi}_m> \tilde{\xi}_m .$$

The eigenvalues are bounded below by $\mathcal{P}(U_t = 0)$. Often $\{\xi_m\}$ is a basis for $\mathcal{L}^2(\pi)$, for example when $T = [0,\infty)$ and $\{P^t, t \geq 0\}$ is a strongly continuous semi group, in which case the term involving $\tilde{\xi}_m$ does not appear.

To see the effect that subordination has on sample paths notice that U can be chosen right continuous so that if Y is right continuous so is X. However the sample paths of U have jumps a.s. and so X generally has jumps a.s. Thus subordination creates processes having jumps in their paths from continuous processes, but as pointed out by Feller (1970, X.9), the space of subordinated transition operators is dense in the space of transition operators, when the condition of right continuity is imposed. Thus the fact that sample paths have jumps need not be of great concern. Sometimes these jumps may actually be desirable as models of situations where animals have occasional bursts of high activity or in situations where the scale of observation makes a limiting diffusion approximation inadequate, that is, the local disturbances appear non normal.

I present the idea of subordination simply as a means of modifying a given transition operator. It is possible, using subordination, to generate a whole class of transition operators from a single operator and thus to introduce parameters where there were none before.

The class of processes with non negative stationary independent increments is very broad. If $T' = \{0,1,...\}$ then U_t is simply the partial sums of a sequence of i.i.d. non negative random variables. When $T' = [0,\infty)$, U_t can be characterised as a non negative random variable with an infinitely divisible distribution. The Laplace transform of U_t, $\varphi_t(\lambda) = Ee^{-\lambda U_t}$ takes the form

$$\varphi_t(\lambda) = e^{-t\,\psi(\lambda)}$$

where

$$\psi(\lambda) = \int_{[0,\infty)} \frac{1-e^{-\lambda x}}{x}\, d\mu(x)$$

and μ is any σ-finite measure satisfying

$$\int_{[1,\infty)} \frac{1}{x}\, d\mu(x) < \infty .$$

(Feller, 1970, XIII, 7)

Note that when $T = [0,\infty)$, then $\rho_m > 0$, $\rho_m = e^{-\lambda m}$ say, and so it is the Laplace transform of U_t that we require to obtain the spectral decomposition of P^t.

A process with stationary independent increments is a homogeneous Markov process and subordination of a process with stationary independent increments gives another process with stationary independent increments. Thus a more general class of processes is not obtained by subordination of subordinated processes.

Example

Let Y be the Ornstein-Uhlenbeck process with transition operator

$$Q^t = \sum_{n=0}^{\infty} e^{-n\lambda t} <, H_n > H_n$$

$\pi = \Phi$, the standard normal distribution. This operator was obtained earlier from the bivariate normal.

Let U_t be a gamma process so that

$$Ee^{-\lambda U_t} = (1+\lambda)^{-bt} \quad , \; b > 0 \text{ is a parameter,}$$

then $P^t = \sum_{m=0}^{\infty} (1+m\lambda)^{-bt} <,H_m> H_m$.

P^t depends on one more parameter than Q^t and

$$\sum_{m=0}^{\infty} (1+m\lambda)^{-2bt} = \infty \quad \text{when} \quad t < \frac{1}{2b}$$

so that the Hilbert-Schmidt property of Q^t does not carry over to P^t.

In this particular example it is quite easy to examine the effect of transformation in terms of conditional means and variances because $H_1(x) = x$ and $H_2(x) = 2^{-½}(x^2-1)$

$$E[Y(t)|Y(0)] = e^{-\lambda t} \; Y(0)$$

$$V\big(Y(t)|Y(0)\big) = 1 - e^{-2\lambda t} .$$

However

$$E[X(t)|X(0)] = (Ee^{-\lambda U_t}) \; X(0)$$

$$V\big(X(t)|X(0)\big) = X^2(0) V(e^{-\lambda U_t}) + 1 - Ee^{-2\lambda U_t}$$

so that the conditional variance of the new process is a quadratic function of $X(0)$. Other conditional moments are also readily available from the spectral decomposition of the operators.

5 MODELLING IN FINER DETAIL

If the movement process, X_ℓ, of an animal, is modelled as a Markov process then $P(X_\ell(t+h)\in B|X_\ell(s),s\leq t) = P(X_\ell(t+h)\in B|X_\ell(t))$. For this to be true one has to deny that the direction the animal is travelling, for example, has any influence on its future movements, given its present position. Furthermore, for many Markov processes, the direction of travel can not be defined and so one can not even inquire about its relationship to future movements. The Markov property imposes an irregularity on sample paths that is difficult to imagine being "true" for any real animals.

Models are never perfect descriptions and it is pointless expecting them to be. In continuous space, Markov models for animal movements are generally postulated from some impression of the asymptotic behaviour of the animals' movements. For example, it is often imagined that there is a sufficient degree of randomness associated with each increment, $X_\ell(t+h)-X_\ell(t)$, for small h, that $P(X_\ell(t+h')\in B|X_\ell(s)s\leq t)$ depends little on $\{X_\ell(s),s<t\}$ for h' large. If a Markov model is arrived at in this way the right to use the model to inquire about the behaviour of $X_\ell(t)$ over small time intervals is forfeited. Models which do incorporate information about short periods of time will be referred to as fine detail models.

In several instances it is desirable to obtain fine detail models. First one may wish to relate movements to more basic behaviour of the animals to gain a better understanding and the ability to make sharper predictions.

Secondly fine detail modelling can justify models, such as Markov models, as asymptotic forms of more believable models. Thirdly one may have observations continuous, or near continuous, in time and desire to make full use of these observations. In this context also, because of limitations in the distance over which an animal can be followed, it may be difficult to obtain observations sufficiently spaced in time for other kinds of models to be applicable. Thus one must take observations over short periods of time and use a fine detail model. Finally I shall use fine detail models to explore different ways of incorporating differences between individual animals' movements.

The movements over short periods of time, for many animals, are so complicated as to defy any kind of fine detail modelling. However some animals' paths are made up approximately of pieces of curves of a definite kind such as straight lines or circles. It is with these situations that we shall be concerned.

Croll and Blair (1973) found that nematode larvae produce tracks consisting of pieces of straight and approximately circular arcs. They were interested in whether movements of this kind have adaptive value for dispersal and so they built a fine detail simulation model to test this question.

Observations of the movements of the bacterium *Escherichia coli* by Berg and Brown (1972) showed that its paths could be approximated by straight line segments. These observations led to the development of fine detail models by Stroock (1974), Lovely and Dahlquist (1974) and Nossal and Weiss (1974). Stroock was interested in developing a general model which reproduces bacterial chemotaxis and indicates optimal strategies for movement behaviour of a bacterium. In another paper Baggett and Stroock (1974) show that the distribution of X_ℓ, in one of Stroock's models, is asymptotically equal to the distribution of a homogeneous Markov diffusion process. Lovely and Dahlquist use their models to find means of various quantities used for detecting and measuring bacterial chemotaxis. The models of Stroock and Lovely and Dahlquist belong to the general setting of section 5.1, however the kinds of specific assumptions made and the results sought are different. These authors had specific applications in mind whereas I have the simple purpose of demonstrating another way of modelling animal movements.

I shall take the attitude that fine detail modelling is of most use for understanding movements locally in space and time and that movements over extended periods are best studied by some asymptotic model. Therefore I shall assume that the space is homogeneous meaning that the distribution of X_ℓ is invariant under translation of the underlying space. However it is not generally assumed that the distribution is invariant under rotation of the space - animals can have preferred directions.

5.1 Continuous piecewise linear paths.

Suppose an animal moves in straight lines at constant speed except at a discrete set of points in time when it changes direction and speed. The animal can be thought of as having velocity $\underset{\sim}{U}_0$ in an interval $(0,T_1]$, $\underset{\sim}{U}_1$ in an interval $(T_1,T_2]$ and so on.

To build a mathematical model for this animal's movements let $\underset{\sim}{U}_0,\underset{\sim}{U}_1,\ldots$ be a sequence of random vectors in R^d. The dimension of the space, d is 1,2 or 3. Let M be a point process on $[0,\infty)$, that is, M is a random measure for which $M(B)$ is a non negative integer valued random variable when B is a bounded Borel set in $[0,\infty)$. Let $M(t) = M[0,t)$, then the velocity of the animal at time t is defined as

$$\underset{\sim}{V}(t) = \sum_{m=0}^{\infty} \underset{\sim}{U}_m \, I_{\{m\}}\big(M(t)\big) \qquad (1)$$

Thus the animal's velocity is $\underset{\sim}{U}_m$ while $M(t) = m$ and the times at which the changes of velocity occur are $T_m = \sup\{t | M(t) \leqslant m-1\}$, $m = 1,2,\ldots$. In general M may have multiple points so that it is possible for $T_m = T_{m+1} = \ldots = T_{m+r}$ for some m,r. When this occurs $\underset{\sim}{V}$ does not take the values $\underset{\sim}{U}_m, \underset{\sim}{U}_{m+1}, \ldots, \underset{\sim}{U}_{m+r-1}$ but jumps from $\underset{\sim}{U}_{m-1}$ to $\underset{\sim}{U}_{m+r}$.

Given this velocity process, the position process, $\{\underset{\sim}{X}(t), t \geqslant 0\}$ giving the animal's position as a function of time, is

$$\underset{\sim}{X}(t) = \int_0^t \underset{\sim}{V}(u)\,du \quad ; \qquad (2)$$

and the integral exists as an ordinary Lebesque or Riemann integral. Since $\underset{\sim}{X}(t) = \lim_{n\to\infty} \frac{t}{n} \sum_{i=1}^{n} \underset{\sim}{V}\left(\frac{it}{n}\right)$, it is a random variable. In many cases the integral (2) exists in the sense of mean square (m.s.) convergence which enables moments of $\underset{\sim}{X}$ to be related to moments of $\underset{\sim}{V}$. Mean square convergence of the integral follows from mean square continuity of $\underset{\sim}{V}$.

Theorem 1.

1. Let $E|\underset{\sim}{V}(t)|^2$, $t \in [a,b]$ be bounded, for all non negative $a \leqslant b$.
2. $P(M\{t\} = 0) = 1$ for each t. Then $\underset{\sim}{V}$ is continuous in mean square.

Proof

For $s,t \in [a,b]$ we can assume $E|\underset{\sim}{V}(s) - \underset{\sim}{V}(t)|^2 <$ a positive constant K.

We have

$$\begin{aligned}K \geqslant E|\underset{\sim}{V}(s)-\underset{\sim}{V}(t)|^2 &= E|\underset{\sim}{V}(s)-\underset{\sim}{V}(t)|^2\, I_{\{M(t)\neq M(s)\}}\\ &= E \sum_{m\neq n} |\underset{\sim}{U}_m-\underset{\sim}{U}_n|^2 I_{\{m\}}(M(t))\, I_{\{n\}}(M(s))\\ &= \sum_{m\neq n} E|\underset{\sim}{U}_m-\underset{\sim}{U}_n|^2 I_{\{m\}}(M(t))\, I_{\{n\}}(M(s))\end{aligned}$$

The last step follows from the non negativity of the summands. Since $M\{t\} = 0$ a.s. and M is a random measure we have $|\underset{\sim}{U}_m-\underset{\sim}{U}_n|^2 I_{\{m\}}(M(t))\, I_{\{n\}}(M(s)) \xrightarrow{a.s.} 0$ as $s \to t$. Thus applying the dominated convergence theorem twice we have

$$E|\underset{\sim}{U}_m-\underset{\sim}{U}_n|^2 I_{\{m\}}(M(t))\, I_{\{n\}}(M(s)) \to 0$$

and $$E|\underset{\sim}{V}(s)-\underset{\sim}{V}(t)|^2 \to 0 \quad \text{as} \quad s \to t.$$

$\underset{\sim}{V}$ is continuous in mean square.

There are not likely to be any particular times at which an animal will change velocity with non zero probability, unless there is an external stimulus at these times. Thus the assumption $P(M\{t\}=0) = 1$ is quite realistic and is assumed in the sequel.

Mean square continuity of $\underset{\sim}{V}$ implies continuity of $E\underset{\sim}{V}(t)$ and $C(\underset{\sim}{V}(s),\underset{\sim}{V}(t))$ which implies that the integral (2) is m.s. convergent (Cramér and Leadbetter, 1967, Chpt. 5). Furthermore

$$E\,\underset{\sim}{X}(t) = \int_0^t E\,\underset{\sim}{V}(u)\,du \tag{3}$$

and

$$C(\underset{\sim}{X}(s),\underset{\sim}{X}(t)) = \int_0^t \int_0^s C(\underset{\sim}{V}(u),\underset{\sim}{V}(v))\,du\,dv. \tag{4}$$

If the animal under consideration has been moving around for some time before observation begins it seems appropriate to assume that M is a stationary point process and

$\underset{\sim}{U}_0, \underset{\sim}{U}_1, \ldots$ is a stationary process. Purely as a matter of convenience assume M and $\{\underset{\sim}{U}_m\}$ are independent. In this situation $\underset{\sim}{V}$ is also stationary.

In the proof of this result it is convenient for later discussion to extend M and $\{\underset{\sim}{U}_m\}$ backward in time. From the stationarity of M and $\{\underset{\sim}{U}_m\}$ we can assume without loss of generality that they are defined as a stationary point process on $R = (-\infty,\infty)$ and as a stationary process indexed by $\{0,\pm 1,\pm 2,\ldots\}$. $\underset{\sim}{V}$ is extended to $(-\infty,\infty)$ by the definition

$$\underset{\sim}{V}(t) = \sum_{m=0}^{\infty} I_{\{m\}}(M[t,0))\underset{\sim}{U}_{-m} \quad , \quad t < 0.$$

Theorem 2.

$\underset{\sim}{V}$ is a stationary process.

Proof

Let $B_{-n},\ldots,B_n$ be Borel sets in R^d and $t_{-n} \leqslant t_{-n+1} \leqslant \ldots \leqslant t_0 = 0 \leqslant \ldots \leqslant t_n$ be fixed real numbers

$$P\big(\underset{\sim}{V}(t_i) \in B_i, i=0,\ldots,\pm n\big)$$

$$= \Sigma P\big(\underset{\sim}{U}_{m_i} \in B_i, i=0,\ldots,\pm n; M[0,t_i)=m_i, M[t_{-i},0)=-m_{-i}, i=1,\ldots,n\big)$$

where the sum is over all integers $m_{-n} \leqslant m_{-n+1} \leqslant \ldots \leqslant m_0 = 0 \leqslant \ldots \leqslant m_n$

Using independence the summand becomes

$$P(\underset{\sim}{U}_{m_i} \in B_i, i=0,\ldots,\pm n)P\big(M[0,t_i)=m_i, M[t_{-i},0)=-m_{-i}, i=1,\ldots,n\big).$$

Stationarity applied to the first factor and the random measure property to the second give

$$P(\underset{\sim}{U}_{m_i - m_{-n}} \in B_i, i=0,\ldots,\pm n)P\big(M[t_{-n},t_i)=m_i - m_{-n}, i=0,\pm 1,\ldots,\pm n\big)$$

Defining $m_i' = m_i - m_{-n}$ and using the stationarity of M

$$P(\underset{\sim}{U}_{m_i'} \in B_i, i=0,\ldots,\pm n)P\big(M[0,t_i - t_{-n})=m_i', i=0,\pm 1,\ldots,\pm n\big)$$

$$= P\big(\underset{\sim}{V}(t_i - t_{-n}) \in B_i, i=0,\ldots,\pm n, M[0,t_i - t_{-n})=m_i', i=\pm 1,\ldots,\pm n\big)$$

Summing over $0 \leqslant m'_{-n+1} \leqslant m'_{-n+2} \leqslant \ldots \leqslant m'_n$ we obtain

$$P\big(\underset{\sim}{V}(t_i - t_{-n}) \in B_i, i=0,\ldots,\pm n\big).$$

That is $P\big(\underset{\sim}{V}(t_i)\in B_i, i=0,..,\pm n\big)=P\big(\underset{\sim}{V}(t_i-t_{-n})\in B_i, i=0,..,\pm n\big)$. We are at liberty to choose $B_i = R^d$ for some i and using this device and appropriate choices of the t_i it is clear that

$$P\big(\underset{\sim}{V}(t_i+h)\in B_i, i=1,..,n\big)$$

does not depend on h for all $B_1,..,B_n \in \mathcal{B}^d$ and all finite sets $t_1,..,t_n$. Thus $\underset{\sim}{V}$ is stationary.

A more natural assumption than the independence and stationarity of M and $\{\underset{\sim}{U}_m\}$ is perhaps stationarity of $\underset{\sim}{V}$. It is clearly less restrictive. Stationarity of $\underset{\sim}{V}$ presumably implies a kind of joint stationarity for M and $\{\underset{\sim}{U}_m\}$ where these two processes are possibly dependent. The kind of dependence between $\{\underset{\sim}{U}_m\}$ amd M is of particular interest in bacterial chemotaxis and was one of the questions asked by Stroock (1975). However in the absence of a chemical attractant or repellant M and $\{\underset{\sim}{U}_m\}$ are independent.

The formulae below are calculated on the basis of independence of M and $\{\underset{\sim}{U}_m\}$ but only the covariance formula (7) depends on this assumption and it immediately generalises to cases of dependence by substituting $E(\underset{\sim}{U}_0\underset{\sim}{U}_m'-\underset{\sim}{\mu}\underset{\sim}{\mu}')\ I_{\{m\}}\big(M(t-s)\big)$ for $P\big(M(t-s)=m\big)\tilde{\Sigma}(m)$. Stroock gives one specific example, in homogeneous space, where M and $\{U_m\}$ are dependent. Other dependence models could be developed along similar lines.

Suppose $E\underset{\sim}{U}_0 = \underset{\sim}{\mu}$ and $C(\underset{\sim}{U}_0,\underset{\sim}{U}_m) = \tilde{\Sigma}(m)$ exist. The stationarity of $\underset{\sim}{V}$ implies

$$E\underset{\sim}{V}(t) = E\underset{\sim}{V}(0) = E\underset{\sim}{U}_0 = \underset{\sim}{\mu} \qquad (5)$$

and for $t \geqslant s$

$$C\big(\underset{\sim}{V}(s),\underset{\sim}{V}(t)\big) = C\big(\underset{\sim}{V}(0),\underset{\sim}{V}(t-s)\big).$$

Use the notation $\Sigma(t-s)$ for $C\left(\underset{\sim}{V}(s),\underset{\sim}{V}(t)\right)$. Note that $\Sigma'(t-s) = \Sigma(s-t)$. For $t \geqslant s$

$$\begin{aligned}\Sigma(t-s) &= E\underset{\sim}{U}_0\underset{\sim}{V}(t-s)' - \underset{\sim}{\mu}\underset{\sim}{\mu}' \\ &= E\sum_{m=0}^{\infty} \underset{\sim}{U}_0\underset{\sim}{U}_m' I_{\{m\}}\left(M(t-s)\right) - \underset{\sim}{\mu}\underset{\sim}{\mu}' \qquad (6)\end{aligned}$$

Now each element of the matrix $\underset{\sim}{U}_0\underset{\sim}{U}_m'$ is bounded in absolute value by $|\underset{\sim}{U}_0||\underset{\sim}{U}_m|$ and

$$\sum_0^{\infty}|\underset{\sim}{U}_0||\underset{\sim}{U}_m| I_{\{m\}}\left(M(t-s)\right) = |\underset{\sim}{V}(0)||\underset{\sim}{V}(t-s)|.$$

Further $E|\underset{\sim}{V}(0)||\underset{\sim}{V}(t-s)| \leqslant \left(E|\underset{\sim}{V}(0)|^2 E|\underset{\sim}{V}(t-s)|^2\right)^{\frac{1}{2}} = E|\underset{\sim}{V}(0)|^2 < \infty$. Thus the dominated convergence theorem applies to the sum (6) and so

$$\Sigma(t-s) = \sum_0^{\infty} P\left(M(t-s)=m\right) E\underset{\sim}{U}_0\underset{\sim}{U}_m' - \underset{\sim}{\mu}\underset{\sim}{\mu}'$$

or

$$\Sigma(t-s) = \sum_0^{\infty} P\left(M(t-s)=m\right)\tilde{\Sigma}(m) \qquad (7)$$

for $t \geqslant s$.

Now $V\underset{\sim}{V}(t) = \Sigma(0)$ exists and so theorem 1 applies. From (3) and (4) we conclude

$$E\underset{\sim}{X}(t) = \underset{\sim}{\mu}t \qquad (8)$$

and
$$C\left(\underset{\sim}{X}(s),\underset{\sim}{X}(t)\right) = \int_0^t \int_0^s \Sigma(v-u)\,du\,dv \qquad (9)$$

Examples

1. Let $\{\underset{\sim}{U}_n\}$ be a d dimensional stationary Markov normal process with covariance

$$\tilde{\Sigma}(m) = \rho^m\Sigma, \quad 0 \leqslant \rho < 1.$$

M is a Poisson process with parameter λ.

In this case

$$\begin{aligned}\Sigma(t) &= \sum_0^{\infty} \frac{\rho^m\lambda^n|t|^n e^{-\lambda|t|}\Sigma}{m!} \\ &= e^{-\lambda|t|(1-\rho)}\Sigma\,. \qquad (10)\end{aligned}$$

For $s \leqslant t$

$$C(\underset{\sim}{X}(s),\underset{\sim}{X}(t)) = \int_0^t \int_0^s e^{-\lambda|v-u|(1-\rho)}\,dudv\ \Sigma$$

$$= \left\{\frac{2s}{\lambda(1-\rho)} + \frac{1}{\lambda^2(1-\rho)^2} \quad e^{-\lambda(1-\rho)t} + e^{-\lambda(1-\rho)s} - e^{-\lambda(1-\rho)(t-s)} + 1)\right\} \Sigma \quad (11)$$

The means of $\underset{\sim}{V}$ and $\underset{\sim}{X}$ are given by (5) and (8).

It is interesting to note that the process $\underset{\sim}{V}$ has identical first and second order moments to a d-dimensional Ornstein-Uhlenbeck process, however there is a marked difference in the sample paths. An Ornstein-Uhlenbeck process can be chosen continuous and nowhere differentiable, but $\underset{\sim}{V}$ has many discontinuities and is differentiable almost everywhere. The process $\underset{\sim}{X}$ has the same first and second order moments as the integral of an Ornstein-Uhlenbeck process.

Generalisations of example 1. are obtained by replacing ρ by an appropriate matrix $\mathbb{P}$ and replacing the Poisson process by a compound Poisson process.

2. Suppose the animal does not orientate relative to any fixed direction in the space so that changes in direction take place relative to previous directions of movement. Consider only the case $d = 2$. Write $\underset{\sim}{U}_m = (R_m\cos\Theta_m, R_m\sin\Theta_m)$.

Let R_0 and Θ_0 be independent with Θ_0 having a uniform distribution on $[0,2\pi)$. $\Psi_1,\Psi_2\ldots$ are the successive changes in direction so that $\Theta_m = \Theta_0 + \sum_{j=1}^{m} \Psi_j$. The Ψ_j are assumed i.i.d. and independent of $R_0, R_1, \ldots$ which are also i.i.d. Under this model each Θ_m is uniform on $[0,2\pi)$ and it is easy to see that $\underset{\sim}{U}_m$ is a stationary Markov process with $E\underset{\sim}{U}_m = \underset{\sim}{0}$.

Define $\alpha_m + i\beta_m = E\exp\{i\sum_{j=1}^{m}\Psi_j\} = (Ee^{i\Psi_1})^m$, now

$$C(\underset{\sim}{U}_0, \underset{\sim}{U}_m) = (ER_0)^2 \begin{bmatrix} E\cos\Theta_0\cos\Theta_m & E\cos\Theta_0\sin\Theta_m \\ E\sin\Theta_0\cos\Theta_m & E\sin\Theta_0\sin\Theta_m \end{bmatrix}$$

$$= (ER_0)^2 \ (a_{ij}), \text{ say.}$$

Next $a_{11} - a_{22} + i(a_{12} + a_{21}) = E\,e^{i(\Theta_0+\Theta_m)} = E\,e^{2i\Theta_0}E\exp\{i\sum_1^m\Psi_j\} = 0$ since Θ_0 is uniform. On the other hand

$$a_{11} + a_{22} + i(a_{12} - a_{21}) = E\,e^{i(\Theta_m - \Theta_0)} = E\exp\{i\sum_1^m\Psi_j\} = \alpha_m + i\beta_m.$$

Hence

$$C(\underset{\sim}{U}_0, \underset{\sim}{U}_m) = \tfrac{1}{2}(ER_0)^2 \begin{bmatrix} \alpha_m & \beta_m \\ -\beta_m & \alpha_m \end{bmatrix}$$

Consideration of the isomorphism between complex numbers, $\alpha + i\beta$, and the matrices of the form

$$\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \tag{12}$$

shows that

$$\begin{bmatrix} \alpha_1 & \beta_1 \\ -\beta_1 & \alpha_1 \end{bmatrix}^m = \begin{bmatrix} \alpha_m & \beta_m \\ -\beta_m & \alpha_m \end{bmatrix} \tag{13}$$

If M is a again a Poisson process we have, for $t \geqslant 0$,

$$2(ER_0)^{-2}\ \Sigma(t) = e^{-\lambda t}\sum_{m=0}^{\infty}\frac{\lambda^m t^m}{m!}\begin{bmatrix} \alpha_1 & \beta_1 \\ -\beta_1 & \alpha_1 \end{bmatrix}$$

$$= e^{-\lambda t}\ \exp\left\{\lambda t\begin{bmatrix} \alpha_1 & \beta_1 \\ -\beta_1 & \alpha_1 \end{bmatrix}\right\}$$

$$= e^{-\lambda t}\begin{bmatrix} \cos\lambda t\alpha_1 & \sin\lambda t\beta_1 \\ -\sin\lambda t\beta_1 & \cos\lambda t\alpha_1 \end{bmatrix}. \tag{14}$$

The last step follows from the isomorphism.

For $t \geqslant s$

$$C(\underset{\sim}{X}(s), \underset{\sim}{X}(t)) = \int_0^t\int_0^s \Sigma(v-u)\,du\,dv$$

$$= \left\{\frac{2s}{\lambda(1+\alpha_1^2)} + \frac{1}{\lambda^2(1+\alpha_1^2)}\big(f_1(t)+f_1(s)-f_1(t-s)+f_1(0)\big)\right\}I_2$$

$$+ \frac{1}{\lambda^2(1+\alpha_1^2)^2}\big(f_2(t)-f_2(s)-f_2(t-s)+f_2(0)\big)J_2. \tag{15}$$

where $f_1(u) = e^{-\lambda u}[\lambda^2(1-\alpha_1^2)\cos\lambda\alpha_1 u - 2\lambda^2\alpha_1\sin\lambda\alpha_1 u]$

$f_2(u) = e^{-\lambda u}[\lambda^2(1-\alpha_1^2)\sin\lambda\alpha_1 u + 2\lambda^2\alpha_1\cos\lambda\alpha_1 u]$

and

$$J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

For this model $E\underset{\sim}{X}(t) = E\underset{\sim}{V}(t) = \underset{\sim}{0}$.

5.2 Differences between individuals and asymptotic properties.

The fine detail models of this chapter enable the comparison of several ways of modelling variation between individuals and they allow the interpretation of differences in movement process from more basic behavioural differences. Of particular interest are situations where the time transformation models of other chapters are fair approximations.

Recall the two expressions

$$\underset{\sim}{V}(t) = \sum_{m=0}^{\infty} \underset{\sim}{U}_m I_{\{m\}}(M(t)) \tag{1}$$

and

$$\underset{\sim}{X}(t) = \int_0^t \underset{\sim}{V}(u)du \ . \tag{2}$$

Consider just two animals, 1 and 2, and associated stochastic processes $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \underset{\sim}{V}_1, \underset{\sim}{V}_2$, etc. We shall look at various ways that these two animals' movement processes might differ. Note that somewhat artificially $\underset{\sim}{X}(0) = \underset{\sim}{0}$. In this chapter we are really looking at $\underset{\sim}{X}(t) - \underset{\sim}{X}(0)$ or movement relative to the initial position. Thus differences between individuals' movement processes arising from differences in the initial distributions form no part of the discussion.

1. The time transformation model is equivalent to

$\{\underset{\sim}{X}_1(\alpha t), t \geqslant 0\} \overset{d}{=} \{\underset{\sim}{X}_2(t), t \geqslant 0\}$, for some $\alpha \geqslant 0$, where "$\overset{d}{=}$" means equal in distribution. This model is implied by, and with suitable restrictions implies,

$\{M_1(\alpha t), t \geqslant 0; \alpha \underset{\sim}{U}_{m1}, m=0,1,\ldots\} \overset{d}{=} \{M_2(t), t \geqslant 0; \underset{\sim}{U}_{m2}, m=0,1,\ldots\}$

so that the spacing of the times of change of velocity and the magnitude of the velocity are altered by proportional amounts.

The relationships between the moments are

$$E\underset{\sim}{X}_2(t) = \alpha E\underset{\sim}{X}_1(t)$$

and $C\left(\underset{\sim}{X}_2(s),\underset{\sim}{X}_2(t)\right) = C\left(\underset{\sim}{X}_1(\alpha s),\underset{\sim}{X}_1(\alpha t)\right)$,

which, for both examples 1 and 2, and for large s and t, becomes $C\left(\underset{\sim}{X}_2(s),\underset{\sim}{X}_2(t)\right) \doteqdot \alpha C\left(\underset{\sim}{X}_1(s),\underset{\sim}{X}_1(t)\right)$

2. Alter the spacing between changes of velocity: $\{M_1(\alpha t), t \geq 0\} \overset{d}{=} \{M_2(t), t \geq 0\}$. This results in $\{\underset{\sim}{V}_1(\alpha t), t \geq 0\} \overset{d}{=} \{V_2(t), t \geq 0\}$. Now $\int_0^t \underset{\sim}{V}_1(\alpha u)\,du = \frac{1}{\alpha}\int_0^{\alpha t} \underset{\sim}{V}_1(u)\,du$ so that

$$\left\{\frac{1}{\alpha}\,\underset{\sim}{X}_1(\alpha t), t \geq 0\right\} \overset{d}{=} \left\{\underset{\sim}{X}_2(t), t \geq 0\right\}.$$

$$E\underset{\sim}{X}_2(t) = E\underset{\sim}{X}_1(t) \quad \text{and}$$

$$C\left(\underset{\sim}{X}_2(s),\underset{\sim}{X}_2(t)\right) = \frac{1}{\alpha^2}\, C\left(\underset{\sim}{X}_1(\alpha s),\underset{\sim}{X}_1(\alpha t)\right).$$

Both examples give

$$C\left(\underset{\sim}{X}_2(s),\underset{\sim}{X}_2(t)\right) \doteqdot \frac{1}{\alpha}\, C\left(\underset{\sim}{X}_1(s),\underset{\sim}{X}_1(t)\right),$$

for large s and t.

Now suppose $E\underset{\sim}{V}(t) = \underset{\sim}{0}$ and

$$C\left(\underset{\sim}{\gamma}'\underset{\sim}{V}_1(s),\underset{\sim}{\gamma}'\underset{\sim}{V}_1(t)\right) = \rho(|t-s|)$$

is strictly monotonic in $|s-t|$ for every $\underset{\sim}{\gamma} \neq 0$. Then if $\alpha > 1$

$$\begin{aligned} V\gamma'\left(\alpha^{-1}\underset{\sim}{X}_1(\alpha t)\right) &= \frac{1}{\alpha^2}\int_0^{\alpha t}\int_0^{\alpha t} \rho(|u-v|)\,du\,dv \\ &= \int_0^t\int_0^t \rho(\alpha|u-v|)\,du\,dv \\ &< \int_0^t\int_0^t \rho(|u-v|)\,du\,dv \\ &= V\underset{\sim}{\gamma}'\underset{\sim}{X}_1(t). \end{aligned}$$

Thus $V\underset{\sim}{X}_1(t) > V\underset{\sim}{X}_2(t)$, or $V\underset{\sim}{X}_1(t) - V\underset{\sim}{X}_2(t)$ is a positive definite matrix. It is a tenet of animal orientation theory that an increase in the rate of turning, for non directed movements, decreases the tendency to drift. An interpretation of this statement is that the variance of the movement process is decreased. This idea is verified for the case of monotonic correlation.

3. Multiply the $\underset{\sim}{U}_m$ by a constant:

$\{\alpha\underset{\sim}{U}_{m1}, m=0,1,\ldots\} \overset{d}{=} \{\underset{\sim}{U}_{m2}, m=0,1,\ldots\}$. Thus

$$\{\alpha\underset{\sim}{V}_1(t), t \geqslant 0\} \overset{d}{=} \{\underset{\sim}{V}_2(t), t \geqslant 0\}$$

and $\{\alpha\underset{\sim}{X}_1(t), t \geqslant 0\} \overset{d}{=} \{\underset{\sim}{X}_2(t), t \geqslant 0\}$. We have

$E\,\underset{\sim}{X}_2(t) = \alpha E\underset{\sim}{X}_1(t)$ and $C(\underset{\sim}{X}_2(s), \underset{\sim}{X}_2(t)) = \alpha^2 C(\underset{\sim}{X}_1(s), \underset{\sim}{X}_2(t))$.

4. Suppose that the only difference in the movements of the two animals is in the mean of the velocity process.

Put $\underset{\sim}{\mu} = E\underset{\sim}{U}_{m1}$ and

$$\{\underset{\sim}{U}_{m1} + (1-\alpha)\underset{\sim}{\mu}, m=0,1,\ldots\} \overset{d}{=} \{\underset{\sim}{U}_{m2}, m=0,1,\ldots\}$$

then $\{\underset{\sim}{X}_1(t) + \underset{\sim}{\mu}(1-\alpha)t, t \geqslant 0\} \overset{d}{=} \{\underset{\sim}{X}_2(t), t \geqslant 0\}$.

Thus $E\underset{\sim}{X}_2(t) = \alpha E\underset{\sim}{X}_1(t)$

and $C(\underset{\sim}{X}_2(s), \underset{\sim}{X}_2(t)) = C(\underset{\sim}{X}_1(s), \underset{\sim}{X}_2(t))$.

5. Now suppose that the differences occur in the covariance of the velocity:

$$\{\alpha(\underset{\sim}{U}_{m1} - \underset{\sim}{\mu}), m=0,1,\ldots\} \overset{d}{=} \{\underset{\sim}{U}_{m2} - \underset{\sim}{\mu}, m=0,1,\ldots\}.$$

$\{\alpha(\underset{\sim}{X}_1(t) - \underset{\sim}{\mu}t), m=0,1,\ldots\} \overset{d}{=} \{\underset{\sim}{X}_2(t) - \underset{\sim}{\mu}t, m=0,1,\ldots\}$ so that

$E\underset{\sim}{X}_2(t) = E\underset{\sim}{X}_1(t)$ and $C(\underset{\sim}{X}_2(s), \underset{\sim}{X}_2(t)) = \alpha^2 C(\underset{\sim}{X}_1(s), \underset{\sim}{X}_2(t))$.

There are many other ways the movement processes can differ, perhaps more subtle ways. The above ways form some basis for discussion.

Apart from the time transformation method, each method of specifying differences affects either the mean or the covariance but not both. The time transformation method is equivalent to the simultaneous application of methods 2. and 3., with the same value of α, and if consideration is restricted, to first and second moments of $\underset{\sim}{X}(t)$ in examples 1. and 2. then there is a combination of methods 4. and 5. which gives approximately the same result as method 1. There are other combinations of methods 1.-4. which give approximately the same result as method 1. It seems that method 1. is a compromise between the other methods. When the velocity process has $\underset{\sim}{0}$ mean methods 1.,2.,3. and 5. give approximately the same result in terms of first and second moments for examples 1. and 2. In this case method 4. has no effect.

Since part of the reason for the present theory is to justify the use of models specifying less detail which arise as asymptotic forms of the models given here it is of interest to see what effect these methods of modelling differences produces in the asymptotic distribution, if it exists.

The Wiener process is often argued as a model for animal movements since it arises as the limit of random walks. By the Wiener process with mean $\underset{\sim}{\mu}$ and variance Σ_0 I mean a process $\underset{\sim}{W} = \{\underset{\sim}{W}(t), t \geqslant 0\}$ such that $\underset{\sim}{W}(0) = \underset{\sim}{0}$, $\underset{\sim}{W}$ has independent increments and the distribution of $\underset{\sim}{W}(t) - \underset{\sim}{W}(s)$ is $N\left(\underset{\sim}{\mu}(t-s), \Sigma_0|t-s|\right)$, or multivariate normal with mean $\underset{\sim}{\mu}(t-s)$ and variance $\Sigma_0|t-s|$. Thus $\underset{\sim}{W}$ is a normal process. An alternative description is that $\underset{\sim}{W}$ is a normal process. $\underset{\sim}{W}(0) = \underset{\sim}{0}$, $E\underset{\sim}{W}(t) = \underset{\sim}{\mu}t$ and

$$C\left(\underset{\sim}{Y}(s), \underset{\sim}{Y}(t)\right) = \Sigma_0 \min(s,t).$$

The process $\underset{\sim}{W}$ can be chosen to have continuous sample paths; $\underset{\sim}{W}$ is a diffusion process in the sense of chapter 4 and is generally regarded as the most basic diffusion process. Other diffusion processes behave like the Wiener process locally, away from boundaries. In homogeneous space, if one wants a diffusion process model, then the Wiener process is the natural one to choose since its transition probabilities have the translation invariance property

$$P\left(\underset{\sim}{W}(t)+\underset{\sim}{x}\in B\,\middle|\,\underset{\sim}{W}(0)=\underset{\sim}{x}\right) = P\left(\underset{\sim}{W}(t)\in B\,\middle|\,\underset{\sim}{W}(0)=0\right)$$

so that future movements occur relative to the current position.

I give an example of a situation in which the Wiener process arises as a limit in distribution of fine detail models. The development is based on theory in a paper by Ibragimov (1975) which is briefly outlined below.

Let $\{\xi_m, m=0,\pm1,\ldots\}$ be a stationary sequence of real valued random variables with finite variance and zero mean. Define

$$S_m = \sum_{1}^{m} \xi_m, \quad \sigma_m^2 = V(S_m)$$

Ibragimov gives conditions under which S_m satisfies a central limit theorem in terms of a maximal correlation $\rho(m)$ defined as follows.

Let M^0 be the σ-algebra generated $\{\xi_j, j\leqslant 0\}$ and M_m the σ-algebra generated by $\{\xi_j, j\geqslant m\}$.

$$\rho(m) \overset{def}{=} \sup C(Y,Z)/\{V(Y)V(Z)\}^{\frac{1}{2}}$$

where the supremum is taken over all Y,Z with finite variance such that Y is M^0 measurable and Z is M_m measurable. We can now state theorem 3 which is a specialisation of Ibragimov's theorem 2.1.

Theorem 3

Suppose

1. $\rho(m) \to 0$ as $m \to \infty$,
2. $\sigma_m^2/m \to \sigma^2 > 0$ as $m \to \infty$,
3. and there is a $\delta > 0$ such that $E|\xi_j|^{2+\delta} < \infty$, then $m^{-\frac{1}{2}} S_m$ is asymptotically normally distributed with mean zero and variance σ^2.

This result can be used to obtain a central limit theorem for $\underset{\sim}{X}(t)$ in the situation where M is stationary point process independent of the stationary process $\{U_m\}$. The basic probability space $(\Omega, \mathcal{A}, \mathcal{P})$ can be thought of as a product space $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2, \mathcal{P}_1 \times \mathcal{P}_2)$. Using the notation $\Omega \times \mathcal{B}$ to mean $\{\Omega \times B | B \in \mathcal{B}\}$ it is assumed M is $\mathcal{A}_1 \times \Omega_2 (= \tilde{\mathcal{A}}_1)$ measurable and $\{\underset{\sim}{U}_m\}$ is $\Omega_1 \times \mathcal{A}_2$ measurable. Let $\tilde{\mathcal{L}}^t = \mathcal{L}^t \times \Omega_2$ and $\tilde{\mathcal{L}}_t = \mathcal{L}_t \times \Omega_2$ be the σ-algebras generated by M restricted to $(-\infty, t]$ and $[t, \infty)$ respectively. Similarly $\tilde{N}^0 = \Omega_1 \times N^0$ and $\tilde{N}_m = \Omega_1 \times N_m$ are generated by $\{\underset{\sim}{U}_j, j \leq 0\}$ and $\{\underset{\sim}{U}_j, j \geq m\}$.

To obtain a central limit result we make the following assumptions

1. $E|U_j|^{2+\delta} < \infty$ for some $\delta > 0$
2. $EM(0,1] > 0$
3. $\alpha(m) = \sup\{C(Z,Y)/(VZVY)^{\frac{1}{2}} | 0 < V(Z), V(Y) < \infty,$ Z is $\tilde{N}^0$ measurable, Y is $\tilde{N}_m$ measurable$\} \to 0$ as $m \to \infty$.
4. $\beta(t) = \sup\{C(Z,Y)/(V(Z)V(Y))^{\frac{1}{2}} | 0 < V(Z), V(Y) < \infty,$ Z is $\tilde{\mathcal{L}}^0$ measurable and Y is $\tilde{\mathcal{L}}_t$ measurable$\} \to 0$ as $t \to \infty$.
5. $\lim\limits_{t \to \infty} \frac{1}{t} \int_0^t \int_0^t \Sigma(u-v)\,du\,dv = \Sigma_0 > 0$.

For fixed non zero $\underset{\sim}{\alpha} \in R^d$ define $\xi_m = \underset{\sim}{\alpha}' \int_m^{m+1} \underset{\sim}{V}(u)\,du$. We want to show ξ_m satisfies theorem 3. Clearly $\{\xi_m, m = 0, \pm 1, \ldots\}$ is a stationary process. From condition 5 above it follows that $\sigma_m^2/m \to \sigma^2 > 0$. To show that $E|\xi_m|^{2+\delta} < \infty$ we need to show that $\underset{\sim}{V}$ is a $\mathcal{B}^1 \times \mathcal{A}$ measurable function of (t,ω_1,ω_2).

The random function $\underset{\sim}{V}$ is left continuous. Also if $[x]$ is the greatest integer not more than x then $(t,\omega_1,\omega_2) \mapsto \underset{\sim}{V}([tm]/m,\omega_1,\omega_2)$ is a $\mathcal{B}^1 \times \mathcal{A}$ measurable mapping and, as $\lim\limits_{m\to\infty} \underset{\sim}{V}([tm]/m) = \underset{\sim}{V}(t)$, $\underset{\sim}{V}$ is $\mathcal{B}^1 \times \mathcal{A}$ measurable. Now $|\underset{\sim}{\alpha}'\underset{\sim}{U}_m| \leqslant |\underset{\sim}{\alpha}||\underset{\sim}{U}_m|$ and so $E|\underset{\sim}{\alpha}'\underset{\sim}{U}_m|^{2+\delta} \leqslant |\underset{\sim}{\alpha}|E|\underset{\sim}{U}_m|^{2+\delta} < \infty$; but $\underset{\sim}{U}_m \overset{d}{=} \underset{\sim}{V}(t)$ so that $E|\underset{\sim}{\alpha}'\underset{\sim}{V}(t)|^{2+\delta} < \infty$. Applying Fubini's theorem $\int_0^1 E|\underset{\sim}{\alpha}'\underset{\sim}{V}(t)|^{2+\delta}dt = E\int_0^1 |\underset{\sim}{\alpha}'\underset{\sim}{V}(t)|^{2+\delta}dt \geqslant E|\underset{\sim}{\alpha}'\int_0^1 \underset{\sim}{V}(t)dt|^{2+\delta}$ by Jensen's inequality. Thus $E|\xi_0|^{2+\delta} < \infty$.

The remaining property of the sequence $\{\xi_n\}$ needed to satisfy theorem 3 is $\rho(m) \to 0$ as $m \to \infty$. This is proved by way of a lemma concerning the process $\underset{\sim}{V}$. Let $\mathcal{K}^t$ be the σ-algebra generated by $\{\underset{\sim}{V}(u), u \leqslant t\}$ and K_t the σ-algebra generated by $\{\underset{\sim}{V}(u), u \geqslant t\}$. Define

$$\gamma(t) = \sup\{C(Z,Y)/(\mathcal{V}(Z)\mathcal{V}(Y))^{1/2} \mid 0<\mathcal{V}(Z),\mathcal{V}(Y)<\infty,\ Z \text{ is } K^0 \text{ measurable and } Y \text{ is } K_t \text{ measurable.}\}$$

<u>Lemma 1</u>

$\gamma(t) \to 0$ as $t \to \infty$.

<u>Proof</u>

Let $\omega_1 \in \Omega_1$ be fixed such that $M(t,\omega_1,\omega_2) \geqslant i$. Considered as a function of ω_2, $\underset{\sim}{V}_u(\omega_1,\omega_2)$ is $\mathcal{N}^0$ measurable when $u \leqslant 0$ and $\mathcal{N}_i$ measurable when $u \geqslant t$. Let Z,Y be respectively K^0 and K_t measurable, then $Z(\omega_1,\cdot)$ is $\mathcal{N}^0$ measurable and $Y(\omega_1,\cdot)$ is $\mathcal{N}_i$ measurable. On the other hand, $E[Z|\tilde{A}_1]$ and $E[Y|\tilde{A}_1]$ are respectively $\tilde{\mathcal{L}}^0$ and $\tilde{\mathcal{L}}_t$ measurable. We have

$$C(Z,Y|\tilde{A}_1) \leqslant \alpha(i)\mathcal{V}^{1/2}(Z|\tilde{A}_1)\mathcal{V}^{1/2}(Y|\tilde{A}_1)$$

on the set $\{M(t) \geqslant i\}$ which means

$$EC(Z,Y|\tilde{A}_1) \leqslant E\{I_{\{0,\ldots,i-1\}}(M(t)) + \alpha(i)\}\; V^{1/2}(Z|\tilde{A}_1)V^{1/2}(Y|\tilde{A}_1)$$

Define $I_{ti} = I_{\{0,\ldots,i-1\}}(M[\frac{1}{3}t, \frac{2}{3}t))$, then

$$EI_{\{0,\ldots,i-1\}}(M(t))V^{1/2}(Z|\tilde{A}_1)V^{1/2}(Y|\tilde{A}_1) \leqslant EI_{ti}V^{1/2}(Z|\tilde{A}_1)V^{1/2}(Y|\tilde{A}_1).$$

Now $I_{ti}V^{1/2}(Z|\tilde{A}_1)$ is $\tilde{\mathcal{L}}^{\frac{2}{3}t}$ measurable and $V^{1/2}(Y|\tilde{A}_1)$ is $\tilde{\mathcal{L}}_t$ measurable and so

$$\left|EI_{ti}V^{1/2}(Z|\tilde{A}_1)V^{1/2}(Y|\tilde{A}_1) - EI_{ti}V^{1/2}(Z|\tilde{A}_1)EV^{1/2}(Y|\tilde{A}_1)\right|$$
$$\leqslant \beta(\tfrac{1}{3}t)E^{1/2}V(Z|\tilde{A}_1)E^{1/2}V(Y|\tilde{A}_1)\;;$$

but also $|EI_{ti}V^{1/2}(Z|\tilde{A}_1) - EI_{ti}EV^{1/2}(Z|\tilde{A}_1)| \leqslant \beta(\frac{1}{3}t)E^{1/2}V(Z|\tilde{A}_1)$.

Combining these results

$$EI_{ti}V^{1/2}(Z|\tilde{A}_1)V^{1/2}(Y|\tilde{A}_1) \leqslant (2\beta(\tfrac{1}{3}t) + EI_{ti})E^{1/2}V(Z|\tilde{A}_1)E^{1/2}V(Y|\tilde{A}_1)$$

which means

$$EC(Z,Y|\tilde{A}_1) \leqslant (\alpha(i) + P(M[\tfrac{1}{3}t,\tfrac{2}{3}t)<i) + 2\beta(\tfrac{1}{3}t))E^{1/2}V(Z|\tilde{A}_1)E^{1/2}V(Y|\tilde{A}_1).$$

On the other hand

$$C(E(Z|\tilde{A}_1),\, E(Y|\tilde{A}_1)) \leqslant \beta(t)V^{1/2}E(Z|\tilde{A}_1)V^{1/2}E(Y|\tilde{A}_1)$$

and since

$$V^{1/2}ZV^{1/2}Y \geqslant E^{1/2}V(Z|\tilde{A}_1)E^{1/2}V(Y|\tilde{A}_1) \quad \text{and} \quad V^{1/2}E[Z|\tilde{A}_1]V^{1/2}E[Y|\tilde{A}_1]$$

it follows that

$$C(Z,Y) \leqslant \{\alpha(i) + P(M[\tfrac{1}{3}t,\tfrac{2}{3}t)<i) + 2\beta(\tfrac{1}{3}t) + \beta(t)\}V^{1/2}ZV^{1/2}Y$$

or $\gamma(t) \leqslant \alpha(t) + P(M[0,\frac{1}{3}t)<i) + 2\beta(\frac{1}{3}t) + \beta(t)$ (17)

To show that $\gamma(t) \to 0$ as $t \to \infty$ we simply need to show that $P(M[0,t)<i) \to 0$ as $t \to \infty$, for each fixed i, because this means $\limsup \gamma(t) \leqslant \alpha(i)$ for every i and hence the limit must be 0.

Let $E_m = \{M(-\infty,m)>0, M[m,\infty)=0\}$.

$E_m \uparrow E = \{M(-\infty,\infty)>0, \lim\limits_{m\to\infty} M[0,m)<\infty\}$ as $m \uparrow \infty$. As $m \downarrow -\infty$ $E_m \downarrow \emptyset$. However the stationarity of M implies PE_m does not depend on m. Thus $0 = P\emptyset = PE_m = PE$ by continuity of probability measures. This result means that the events

$\{M(-\infty,\infty)>0\}$ and $\{\lim_{m\to\infty} M[0,m)=\infty\}$ differ by a set of probability 0, we shall say they are equivalent. Similar reasoning shows that the three events $\{M(-\infty,0)>0\}$, $\{M[m,\infty)>0\}$ and $\{M(-\infty,\infty)>0\}$ are equivalent. However $\mathcal{E}M[0,1)>0$ so that $\mathcal{P}\{M(-\infty,\infty)>0\}>0$. If $1>\mathcal{P}\{M(-\infty,\infty)>0\}$ then the correlation between $I_{\{0\}}\big(M(-\infty,0)\big)$ and $I_{\{0\}}\big(M[m,\infty)\big)$ is 1 contradicting assumption 3. Thus $\mathcal{P}\{M(-\infty,\infty)>0\}=1$ and so $M[0,m)\xrightarrow{a.s.}\infty$, $\mathcal{P}\big(M[0,m)<i\big)\to 0$ as $m\to\infty$. Hence $\mathcal{P}\big(M[0,t)<i\big)\to 0$ as $t\to\infty$, completing the proof of the lemma.

To see that $\rho(m)\to 0$ as $m\to\infty$ simply note that ξ_j is K^0 measurable for $j<0$ and K_t measurable for $j\geqslant t$. Thus we have shown that $\{\xi_m\}$ satisfies theorem 3 which means that $t^{-\frac{1}{2}}\underset{\sim}{\alpha}'\big(X(t)-\underset{\sim}{\mu}t\big)$ converges in distribution to $\mathcal{N}(0,\sigma^2)$, the normal distribution with mean 0 and variance σ^2, as $t\to\infty$ through integer values. To obtain the result for $t\to\infty$, continuously, the following argument is adapted from Billingsley (1968, p.179).

Let $\delta_m = \sup_{t\in[m,m+1]} \left|m^{-\frac{1}{2}}S_m - t^{-\frac{1}{2}}\underset{\sim}{\alpha}'\big(\underset{\sim}{X}(t)-\underset{\sim}{\mu}t\big)\right|$ then

$$\delta_m \leqslant m^{-\frac{1}{2}}\int_m^{m+1}\left|\underset{\sim}{\alpha}'\big(\underset{\sim}{V}(u)-\underset{\sim}{\mu}\big)\right|du \quad \text{and so}$$

$$\mathcal{V}\delta_m \leqslant \frac{1}{m}\,\mathcal{E}\left(\int_m^{m+1}\left|\underset{\sim}{\alpha}'\big(\underset{\sim}{V}(u)-\underset{\sim}{\mu}\big)\right|du\right)^2$$

$$\leqslant \frac{1}{m}\,\mathcal{E}\left|\underset{\sim}{U}_0-\underset{\sim}{\mu}\right|^2 \to 0 \quad \text{as} \quad m\to\infty .$$

Thus $\delta_m \xrightarrow{p} 0$ as $m\to\infty$ showing that the distribution of $\underset{\sim}{\alpha}'\big(\underset{\sim}{X}(t)-\underset{\sim}{\mu}t\big)$ has the same limit as $S_{[mt+1]}$ as $t\to\infty$, which is the desired result. Since $\underset{\sim}{\alpha}\neq\underset{\sim}{0}$ is arbitrary this proves theorem 4.

Theorem 4.

$t^{-\frac{1}{2}}\big(\underset{\sim}{X}(t)-\underset{\sim}{\mu}t)\big)$ converges in distribution to $N(0,\Sigma_0)$, the d dimensional normal distribution with mean 0 and variance matrix $\Sigma_0 = \lim_{t \to \infty} \int_0^t \int_0^t \Sigma(u-v)\,du\,dv$.

The stated aims were to show that the distribution of $\underset{\sim}{X}$ converges in some sense, to the distribution of a Wiener process. Since both processes have continuous sample paths the natural kind of convergence to consider is weak convergence of probability measures on a space of continuous functions; for example the space $C^d[0,1]$ which is the set of continuous functions $\underset{\sim}{f} : [0,1] \to R^d$ with metric D defined by

$$D(\underset{\sim}{f},\underset{\sim}{g}) = \sup_{t \in [0,1]} |\underset{\sim}{f}(t)-\underset{\sim}{g}(t)|.$$

The theory of weak convergence in the space $C^1[0,1]$ is well developed (e.g. Billingsley, 1968) and a corresponding theory for $C^d[0,1]$ appears to be a trivial extension of the former theory. Ibragimov's paper, cited previously, gives results on weak convergence of the distribution of a continuous interpolation of the process $\{\xi_m\}$ to a Wiener process. It appears that corresponding results should be true for the multivariate process $\underset{\sim}{X}$. However I shall be content to prove a much weaker result.

By a finite dimensional distribution of a process $\underset{\sim}{Y} = \{\underset{\sim}{Y}(t), t \geqslant 0\}$ is meant the joint distribution of $\underset{\sim}{Y}(t_1),..,\underset{\sim}{Y}(t_r)$ for some finite set $\{t_1,..,t_r\}$. A sequence of processes $\underset{\sim}{Y}_m$ are said to converge in finite dimensional distribution to a process $\underset{\sim}{Y}$ if the joint distribution of $\underset{\sim}{Y}_m(t_1),..,\underset{\sim}{Y}_m(t_r)$ converges weakly to the joint distribution of $\underset{\sim}{Y}(t_1),..,\underset{\sim}{Y}(t_r)$ for all finite sets $\{t_1,..,t_r\}$. We shall show that $h^{-\frac{1}{2}}\{\underset{\sim}{X}(ht)-\underset{\sim}{\mu}t\}$ converges in finite dimensional distribution to a Wiener process. In the following discussion it will be assumed, without loss of generality, that $\underset{\sim}{\mu} = \underset{\sim}{0}$.

Let $0 = t_0 < t_1 < .. < t_r$. From the stationarity of $\underset{\sim}{V}$ it is clear that the distribution of $\underset{\sim}{X}(t)-\underset{\sim}{X}(s)$, $t \geqslant s$ is the same as the distribution $\underset{\sim}{X}(t-s)$ and so by theorem 4 $h^{-\frac{1}{2}}\left(\underset{\sim}{X}(ht_i)-\underset{\sim}{X}(ht_{i-1})\right)$ has the asymptotic distribution $N\left(0,\Sigma_0(t_i-t_{i-1})\right)$. This distribution is the distribution of the increment $\underset{\sim}{W}(t_i)-\underset{\sim}{W}(t_{i-1})$ of a Wiener process with mean $\underset{\sim}{0}$ and variance Σ_0. If it can be shown that the increments $h^{-\frac{1}{2}}\left(\underset{\sim}{X}(ht_i)-\underset{\sim}{X}(ht_{i-1})\right)$, $i = 1,..,r$ are independent in their asymptotic distribution then it follows that $\{h^{-\frac{1}{2}}\underset{\sim}{X}(ht), t \geqslant 0\}$ converges in finite dimensional distribution to $\underset{\sim}{W}$. To prove this we need lemma 2.

Lemma 2

Let $h > 0$ and $A_i \in K_{t_{i-1}} \cap K^{t_i - h}$, $i = 1,..,r$ then $\left|P\left(\bigcap_{i=1}^{r} A_i\right) - \prod_{i=1}^{r} P(A_i)\right| \leqslant \frac{1}{4}(r-1)\gamma(h)$.

Proof

Let $A \in K_t$, $B \in K^s$, then

$$|C(I_A, I_B)| \leqslant \gamma(t-s)\, V^{\frac{1}{2}}(I_A)\, V^{\frac{1}{2}}(I_B).$$

That is $|P(A \cap B) - P(A)P(B)| \leqslant \gamma(t-s)\frac{1}{4}$.

We have $\bigcap_{i=j+1}^{r} A_i \in K_{t_j}$ and $A_j \in K^{t_j - h}$

Thus $\left|P\left(\bigcap_{i=j}^{r} A_i\right) - P(A_j)P\left(\bigcap_{i=j+1}^{r} A_i\right)\right| \leqslant \frac{1}{4}\gamma(h)$

which means

$$\left|P(A_1)..P(A_{j-1})P\left(\bigcap_{i=j}^{r} A_j\right) - P(A_1)..P(A_j)P\left(\bigcap_{i=j+1}^{r} A_i\right)\right| \leqslant \tfrac{1}{4}\gamma(h).$$

Summing over j we obtain the desired result.

Define

$\underset{\sim}{Y}(i,h) = h^{-\frac{1}{2}}\{\underset{\sim}{X}(ht_i)-\underset{\sim}{X}(ht_{i-1})\}$, $\underset{\sim}{Y}^*(i,h) = h^{-\frac{1}{2}}\{\underset{\sim}{X}(ht_i - h^{\frac{1}{4}})-\underset{\sim}{X}(ht_{i-1})\}$.

$h^{-\frac{1}{2}}\underset{\sim}{X}(h^{\frac{1}{4}})$ has the same distribution as $\underset{\sim}{Y}(i,h)-\underset{\sim}{Y}^*(i,h)$, and

$$E\left|h^{-\frac{1}{2}}\underset{\sim}{X}(h^{\frac{1}{4}})\right|^2 = h^{-1}E\left|\int_0^{h^{\frac{1}{4}}}\underset{\sim}{V}(u)\,du\right|^2$$

$$= h^{-1}Eh^{\frac{1}{2}}\left|h^{-\frac{1}{4}}\int_0^{h^{\frac{1}{4}}}\underset{\sim}{V}(u)\,du\right|^2 .$$

Jensen's inequality applies giving

$$E|h^{-\frac{1}{2}}\underset{\sim}{X}(h^{\frac{1}{4}})|^2 \leqslant h^{-1}Eh^{\frac{1}{2}}\left(h^{-\frac{1}{4}}\int_0^{h^{\frac{1}{4}}} |\underset{\sim}{V}(u)|^2du\right)$$

$$= h^{-\frac{1}{2}}\left(h^{-\frac{1}{4}}\int_0^{h^{\frac{1}{4}}} E|\underset{\sim}{V}(u)|^2du\right)$$

$$= h^{-\frac{1}{2}}E|\underset{\sim}{V}(0)|^2$$

We see that $h^{-\frac{1}{2}}\underset{\sim}{X}(h^{\frac{1}{4}}t) \xrightarrow{m.s.} 0$ and the same is true for $\underset{\sim}{Y}(i,h)-\underset{\sim}{Y}^*(i,h)$. It follows that the two sets of vectors $\{\underset{\sim}{Y}(i,h), i=1,..,r\}$ and $\{\underset{\sim}{Y}^*(i,h), i=1,..,r\}$ have the same asymptotic distribution. From lemma 2 we conclude

$$|P(\underset{\sim}{Y}^*(i,h)\in B_i, i=1,..,r)-\prod_{i=1}^{r} P(\underset{\sim}{Y}^*(i,h)\in B_i, i=1,..,r)| \leqslant (r-1)\gamma(h^{\frac{1}{4}})$$

and so the asymptotic distribution of $Y^*(i,h)$, $i=1,..,r$ is a distribution of r independent random vectors.

We have proved

Theorem 5

The finite dimensional distributions of the process $\{h^{-\frac{1}{2}}(\underset{\sim}{X}(t)-\underset{\sim}{\mu}t), t\geqslant 0\}$ converge to the finite dimensional distributions of the Wiener process with mean zero and variance matrix Σ_0.

If theorem 5 is to be used as a basis of a model for animal movements, generally, it will be desirable to obtain processes with possibly non zero mean. This is obtained by considering the limit in finite dimensional distribution of the process $\{\underset{\sim}{\mu}t+h^{-1}(\underset{\sim}{X}(t)-\underset{\sim}{\mu}t), t\geqslant 0\}$. The limit is a Wiener process with mean $\underset{\sim}{\mu}$ and variance Σ_0. This procedure maintains the mean and variance at the same relative order of magnitude during the limiting process.

It is not difficult to find processes satisfying the conditions 1.,2.,3., and 4. and so satisfying theorem 5. Example 1 is such a process.

The asymptotic distribution for $\underset{\sim}{X}$ derived above depends only on $\underset{\sim}{\mu}$ and Σ_0 for $\underset{\sim}{X}$. Therefore in comparing the different methods of specifying differences between individual animals, discussed earlier, we need only look at their effect on $\underset{\sim}{\mu}$ and Σ_0 to see how they affect the asymptotic distributions. $(\underset{\sim}{\mu}_1, \Sigma_{01})$ and $(\underset{\sim}{\mu}_2, \Sigma_{02})$ refer to $\underset{\sim}{\mu}$ and Σ_0 for $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ respectively. The relationship between $(\underset{\sim}{\mu}_1, \Sigma_{01})$ and $(\underset{\sim}{\mu}_2, \Sigma_{02})$ brought about by the different methods of specifying differences are listed below.

1. $\underset{\sim}{\mu}_2 = \alpha\underset{\sim}{\mu}_1, \quad \Sigma_{02} = \alpha\Sigma_{01}$
2. $\underset{\sim}{\mu}_2 = \underset{\sim}{\mu}_1, \quad \Sigma_{02} = \frac{1}{\alpha}\Sigma_{01}$
3. $\underset{\sim}{\mu}_2 = \alpha\underset{\sim}{\mu}_1, \quad \Sigma_{02} = \alpha^2\Sigma_{01}$
4. $\underset{\sim}{\mu}_2 = \alpha\underset{\sim}{\mu}_1, \quad \Sigma_{02} = \Sigma_{01}$
5. $\underset{\sim}{\mu}_2 = \underset{\sim}{\mu}_1, \quad \Sigma_{02} = \alpha^2\Sigma_{01}$

Recall that method 1. is the time transformation method. The first thing to notice is that if $\underset{\sim}{\mu} = \underset{\sim}{0}$, that is there is no preferred direction, methods 1.,2.,3., and 5. are equivalent. Method 4. then has no effect. Secondly no single method has the same effect as the time transformation method, but there are many combinations of methods which have exactly the same effect as time transformation. All of the methods considered maintain proportionality between $\underset{\sim}{\mu}_1$ and $\underset{\sim}{\mu}_2$ and between Σ_{01} and Σ_{02}. More subtle ways of specifying differences may upset the proportionality. In modelling movements of real animals the precise method of specifying differences will depend on the simplicity demanded, the accuracy required and the behaviour of the animals.

6 APPLICATION AND INFERENCE.

The ultimate test of any theory is its application. This chapter shows how the theory of previous chapters is helpful in the investigation of the movement behaviour of the Brine Shrimp *Artemia salina*. In particular this theory enables useful questions to be posed, as anticipated in the introduction.

The experiments given should be regarded as examples of the kinds of experiments that could be done. Generally there is insufficient replication for a high degree of confidence in the conclusions but they do indicate that it is possible to do the experiments and they give some plausibility to the proposed models. Once the apparatus has been perfected and the necessary techniques mastered, replication of the experiments is not a difficult problem.

Wherever possible standard statistical procedures are adapted for our need. However for the testing of independence of different animals' movements I feel that new statistical tests are required. Methods of testing independence are the subject of the next section.

6.1 Testing independence.

Most of the theory in this thesis assumes that the movement processes $X_1, .., X_n$ of the n animals in the population are independent conditional on the environment. Thus the logical first hypothesis to test is this independence. Another reason why this hypothesis should be considered first is the obvious consequences it has with regard to sampling movements to gain further knowledge.

Thus it is envisaged that it will be desired to test independence, conditional on the environment, in the presence of very little knowledge of the stochastic processes involved. For example it is expected that it will not be possible to assume à priori that the stochastic process $(X_1,..,X_n)$ is a homogeneous Markov chain in discrete time. The presence of such assumptions may lead to considerable simplifications, for example in a discrete space setting, if we observe $\{X_1(t),..,X_n(t),t=0,1,..,m\}$, for m large, and it is known that $(X_1,..,X_n)$ is a homogeneous Markov chain then there is a simple test of independence due to Billingsley (1961, p.29).

Note that we are not concerned here with testing the stronger assumption that $X_1,..,X_n$ are independent conditional on the environment and the animals' phenotypes. In our situation of little knowledge, testing this assumption would seem to involve repeated observation of the same individual's movements, which can be avoided in a test of the assumption of independence conditional only on the environment.

For testing independence we shall generally assume that the environment is non random by experimental design so that the dependence due to the environment is not confused with dependence caused by interaction between individuals. This is likely to be impossible to achieve in the field; however in some situations randomness of the environment does not affect the test of independence that I shall give. Because of difficulties of sampling, spatial irregularities in the environmental, and temporal randomness of the environment I expect that it will be difficult if not impossible to test independence in most field situations. Thus I believe that most tests of independence will be in laboratory environments which I take as including highly modified outdoor environments, different from the "field" where the animals normally live.

Ecologists are frequently and understandably very cautious about applying results of laboratory experiments in field situations. It is to be admitted that the presence or absence of dependence may depend on environmental conditions so that laboratory experiments may not be applicable to the field but it also seems that independence, thought of as no significant interaction between individuals, is such a basic behavioural trait that it should not be too greatly affected by environmental conditions. An experimenter who wishes to extrapolate from laboratory to field needs to incorporate in laboratory experiments sufficient features of the field environment to maintain the presence or absence of dependence. To assess the chances of success in this respect the experimenter should check that readily observable qualitative features of behaviour, movement or otherwise, present in the field, are preserved in the laboratory setting. On the other hand new technology, as mentioned in the introduction, may enable some questions of dependence to be answered in the field setting by the provision of the large amounts of data I believe it requires. Sometimes animals are obviously dependent but we are not concerned with these situations.

I shall consider only the case where the animals are not tagged individually because there is no satisfactory tagging method. This is the case in my own experiments and is perhaps the only situation in which testing independence of movements is in any way special. In other situations standard tests for dependence of stationary time series, or in the absence of stationarity, tests of dependence between random vectors, can possibly be suitably adapted to the movement setting.

For simplicity it will be assumed that there is only one subpopulation.

In the situation we are considering, at best we can observe the random measures N_t, $t \in \tau$ where τ is a subset of a finite interval. We know from section 3.3 that the joint distribution of N_t, $t \in \tau$ is generally insufficient to determine independence of $\{X_\ell(t), t \in \tau\}$, $\ell = 1,..,n$. Additional hypotheses are needed, for example the hypotheses 1. and 2. of theorem 3, chapter 2, or the situation in section 3.6 so that theorem 11 of chapter 3 becomes approximately applicable. An alternative, possibly more generally useful additional hypothesis is that the joint distribution of $\{X_\ell(t), t \in \tau\}$, $\ell = 1,..,n$, is restricted to a class where independence of $X_\ell(t)$, $\ell = 1,..,n$, for each $t \in \tau$ implies independence of $\{X_\ell(t), t \in \tau\}$, $\ell = 1,..,n$. We shall refer to this as hypothesis M. In the presence of hypothesis M theorem 7 of chapter 3 can be used as the basis of a test of independence.

Belief in hypothesis M must be based on reasonable biological grounds. Perhaps the most obvious possible outcome of interaction between individuals is under or over dispersion which means that the animals tend to be aggregated more or less than expected if $X_1(t),..,X_n(t)$ are independent. We may think of under or over dispersion as resulting from the animals being attracted to each other or avoiding each other. Consider a population of free swimming aquatic organisms with limited ability to detect each other at a distance. Provided the population is not so dense that they are continually colliding there do not seem to be many ways that the animals could interact that would not lead to under or over dispersion.

The hypothesis of under or over dispersion has been of considerable interest to ecologists leading to the use of the Poisson index of dispersion test for detecting departures from a homogeneous Poisson point process for the spatial distribution of the animals (see Southwood, 1971, p. 36). However since the environment is invariably spatially heterogeneous it is rarely clear whether departures from a homogeneous Poisson process are due to interaction between individuals or environmental heterogeneity.

Whilst the assumption that the only available ways of interacting result in under or over dispersion, does lead to hypothesis M, it is apparent that some animals interact in ways that may not result in $X_1(t),..,X_n(t)$ being dependent, fox fixed t. For example, Wells and Buckley (1972) describe a tendency for individuals of the water snail *Physa* to follow each others slime trails. Suppose snail 2 follows the slime trail of snail 1 such that $X_2(t) = X_1(t-1)$. If the process X_1 has ergodicity properties resulting in $X_1(t)$ and $X_1(t-1)$ being approximately independent then

$X_1(t)$ and $X_2(t)$ are approximately independent in the presence of very high dependence between X_1 and X_2. This result depends on snail 2 generally not following too closely behind snail 1. In *Physa*, trails only remain detectable to other individuals for about 30 to 45 minutes. Thus if the unit of time is an hour and the space X, where the animals are moving, is so large that $X_1(t)$ and $X_1(t-1)$ must be relatively close, then if snail 2 does follow snail 1, it must follow close behind. If the probability that snail 2 follows snail 1 is appreciable then $X_1(t)$ and $X_2(t)$ will be dependent. If the total population is not too great, and this kind of interaction occurs generally in the population, then under dispersion will be apparent.

In the sequel we shall assume that hypothesis M is reasonable and so we concentrate on the joint distribution of $X_1(t),..,X_n(t)$. We assume that $X_1(t),..,X_n(t)$ are symmetrically distributed or "exchangeable" so that dependence between $X_1(t),..,X_n(t)$ is determined by the distribution of N_t.

If the X_ℓ are stationary processes then we have the additional simplification that the distribution of $X_\ell(t)$ is independent of t. However in the present context we want to condition on the environment and if the environment is non constant as a function of time it does not seem reasonable to postulate that X_ℓ is conditionally stationary.

If environmental effects can be specified by time transformation of a process which has a limiting distribution (c.f. model 4c or appropriate special cases of the model of

3.5) then it is reasonable to postulate that the distribution of $X_\ell(t)$, conditional on the environment, does not depend on t for t sufficiently large, furthermore this conditional distribution does not depend on the environment so that environmentally induced dependence does not become confused with interaction between individuals in the distribution of $(X_1(t),\ldots,X_n(t))$ for fixed t.

When there is no reason to believe that the effect of the environment can be modelled by time transformation methods symmetry properties of the environment may suggest that some properties of the distribution of $X_\ell(t)$ should not depend on t or the environment, provided the system has been in existence sufficiently long. For example, we may be able to partition the space into two sets A and B which are mirror images of each other and the environmental conditions for these two sets being mirror images of each other. We could then postulate that

$$P(X_\ell(t)\in A) = P(X_\ell(t)\in B) = \tfrac{1}{2}$$

and furthermore the event $\{X_\ell(t)\in A\}$ and the environmental random variable σ are independent. We can test the independence hypothesis by using the fact that $N_t(A)$ has a binomial distribution with parameters $(n,\frac{1}{2})$ when $X_1(t),\ldots,X_n(t)$ are independent. If we are interested in alternatives which lead to under or over dispersion then an appropriate test statistic is

$$T = \{(N_t(A)-n/2)^2+(N_t(B)-n/2)^2\}/(n/2).$$

Under dispersion leads to large values of T, over dispersion to small values of T and the asymptotic null distribution of T is χ^2_1. T is the Poisson index of dispersion statistic. There is no Poisson distribution here but independent Poisson random variables are jointly conditionally multinomial given their sum. The asymptotic distribution of T is seen by interpreting it as a goodness of fit statistic.

The idea of using the Poisson index of dispersion test to test independence, in the presence of environmental symmetry, was used by Naylor (1959) in laboratory experiments on dispersal of the flour beetle *Tribolium confusum*. Naylor's apparatus was circular with k chambers, $A_1,..,A_k$, $k = 48$ or 10, placed around the circumference of a circle. Beetles collected in the chambers and he postulated that a beetle would have equal chance of collecting in any of the k chambers. He used the Poisson index of dispersion statistic

$$T = \sum_1^k \left(N_t(A_i)-n/k\right)^2/(n/k)$$

and also a test of circular serial correlation between the $N_t(A_i)$, $i = 1,..,k$. If we let $k \to \infty$ in Naylor's situation we obtain a radially symmetric environment. Testing independence in radially symmetric environment is the next topic.

I hope that the development of a test of independence in a radially symmetric environment demonstrates that even when there is very little knowledge of the kind of movement processes involved it is possible to design an experiment in which a great deal can be postulated about the distribution of the $X_\ell(t)$ making the detection of dependence a much simpler task.

Testing independence in a radially symmetric environment.

Suppose the n animals inhabit a radially symmetric two dimensional environment X. For example, a circular container or enclosure with environmental conditions maintained uniform and non directional. It is envisaged that the radially symmetric environment will generally be artificially created either in the laboratory or as a modification of a natural environment. If the results of the study are needed to be applied in the field the experimenter must attempt to include the relevant features of the field environment within a radially symmetric structure. In the field a radially symmetric environment most often results from uniformity of the environment. In that situation the present theory is really only appropriate when the animals chosen for study have the same home range, then they do not wander too far away from each other and lose the opportunity of interacting, and they all recognise the same centre of the space.

If the animals move all over the space it might be expected that the joint distribution of the animals' positions approaches some distribution, after a time, which is radially symmetrical. Some results on this question are proved later. First we need a precise definition of a radially symmetric distribution. Assume the space X has a recognisable origin about which it is a radially symmetric subset of R^2. The appropriate σ-algebra is the Borel sets of R^2 which are contained in X. Denote by T_θ the transformation on R^2 which consists of rotation through the angle θ in the positive direction. For

$\underset{\sim}{x} = (x_1,\dots,x_n) \in (R^2)^n$ let $T_\theta \underset{\sim}{x}$ denote $(T_\theta x_1,\dots,T_\theta x_n)$. If $B \subset R^{2n}$, $T_\theta B = \{T_\theta \underset{\sim}{x} | x \in B\}$. Note that T_θ is a bijective transformation of R^2 and $T_\theta^{-1} = T_{-\theta}$.

Definition 1.

If $Y_1,\dots,Y_n$ are $\mathcal{X}$ valued random variables then $(Y_1,\dots,Y_n)$ has a radially symmetric distribution if $\mathcal{P}\big((Y_1,\dots,Y_n)\in B\big) = \mathcal{P}\big((Y_1,\dots,Y_n)\in T_\theta B\big)$ for all $B \in \mathcal{B}^{2n}$, the Borel sets in R^{2n}, and for all θ.

For the purposes of testing a statistical hypothesis the real nature of the space is unimportant so long as the distributions of the animals positions are radially symmetric about some point. The animals might be thought of as regarding the space radially symmetrically if

$$\mathcal{P}\big(\underset{\sim}{X}(t)\in B \,|\, \underset{\sim}{X}(0)=x\big) = \mathcal{P}\big(\underset{\sim}{X}(t)\in T_\theta B \,|\, \underset{\sim}{X}(0)=T_\theta \underset{\sim}{x}\big) \qquad (1)$$

for all $\underset{\sim}{x} \in \mathcal{X}^n$, θ, $B \in \mathcal{B}^{2n}$. When the animals regard the space radially symmetrically $\underset{\sim}{X}(t)$ can be made to have a radially symmetric distribution by giving $\underset{\sim}{X}(0)$ a radially symmetric distribution. This is theorem 1.

Theorem 1.

Suppose equation (1) holds and $\underset{\sim}{X}(0)$ has a radially symmetric distribution then $\underset{\sim}{X}(t)$ has a radially symmetric distribution.

Proof

Let $Q(B) = \mathcal{P}\big(\underset{\sim}{X}(0)\in B\big)$, then $Q(T_\theta B) = Q(B)$.

$$\begin{aligned}
\mathcal{P}\big(\underset{\sim}{X}(t)\in T_\theta B\big) &= \int \mathcal{P}\big(\underset{\sim}{X}(t)\in T_\theta B \,|\, \underset{\sim}{X}(0)=\underset{\sim}{x}\big)\,dQ(\underset{\sim}{x}) \\
&= \int \mathcal{P}\big(\underset{\sim}{X}(t)\in T_{-\theta}T_\theta B \,|\, \underset{\sim}{X}(0)=T_{-\theta}\underset{\sim}{x}\big)\,dQ(\underset{\sim}{x}) \\
&= \int \mathcal{P}\big(\underset{\sim}{X}(t)\in B \,|\, \underset{\sim}{X}(0)=T_{-\theta}\underset{\sim}{x}\big)\,dQ(\underset{\sim}{x}) \\
&= \int \mathcal{P}\big(\underset{\sim}{X}(t)\in B \,|\, \underset{\sim}{X}(0)=\underset{\sim}{x}\big)\,dQ(x)
\end{aligned}$$

The last equality results from the invariance of Q.

We have $P(\underset{\sim}{X}(t) \in T_\theta B) = P(\underset{\sim}{X}(t) \in B)$ for all $B \in \mathcal{B}^{2n}$.

Theorem 1 is particularly important when the $X_\ell(t)$ do not have a radially symmetric limiting distribution but the animals do regard the space radially symmetrically for then, in an experimental situation, one can give $\underset{\sim}{X}(0)$ a radially symmetric distribution, and if one wants to test the independence hypothesis, the $X_\ell(0)$ should be made i.i.d. A particular case where the limiting radially symmetric distribution may not apply is when each animal sets up a home distribution concentrated on part of the space. Theorem 1 may apply. In this situation, the hypothesis of independence of $X_1(t),..,X_n(t)$ includes independence of the home distributions, however it is possible that the home distributions are dependent but $X_1(t),..,X_n(t)$ are conditionally independent given the home distributions.

If an animal sets up a home distribution concentrated on part of the space, after being experimentally introduced to the space at $t = 0$, it must be able to distinguish some parts of the space from others. It may do this by changing part of the space or by learning to recognise part of the space. This is still compatible with equation (1) provided no directions are preferred in establishment of the home distribution.

When $\underset{\sim}{X}(t)$ has a radially symmetric distribution, marginally each $X_\ell(t)$ has a radially symmetric distribution on $\mathcal{X}$. The distribution of such random variables has a useful structural decomposition.

Theorem 2

Let the X valued random variable Y have a radially symmetric distribution. If (R,Θ) is the polar co-ordinates of Y then R and Θ are independent and Θ has a uniform distribution on $[0,2\pi)$.

Proof

The angle Θ has two alternative interpretations

1. as a real number belonging to $[0,2\pi)$, and
2. as a point on the unit circle.

Consider the second interpretation. Let A be a Borel set in $[0,\infty)$ and B be a Borel set on the unit circle. Then

$$P\big((R,\Theta)\in A\times B\big) = P\big(T_\theta(R,\Theta)\in A\times B\big) = P\big((R,\Theta+\theta)\in A\times B\big).$$

Keeping A fixed and defining $Q(B) = P\big((R,\Theta)\in A\times B\big)$ we see that the measure Q is invariant under the translation group on the unit circle. Thus Q is a multiple of circular Lebesgue measure.

$P\big((R,\Theta)\in A\times B\big) = S(A)\mu(B)/2\pi$ where μ is circular Lebesgue measure and $S(A)$ is some set function. Clearly R and Θ are independent and under interpretation 1. of Θ, Θ has a uniform distribution on $[0,2\pi)$.

It is easy to see that any X valued random variable Y, for which R and Θ are independent and Θ has a uniform distribution on $[0,2\pi)$, has a radially symmetric distribution. This fact can be used to find conditions under which $X_\ell(t)$ becomes approximately radially symmetrically distributed as $t \to \infty$.

Nagaev and Mukhin (1966) have derived sufficient conditions for a sum of independent angles to have a limiting uniform distribution. These help guide the intuition but

I think a different kind of result is needed here. The result I have is very simple; it is based on the following lemma given by Feller (1971, pp. 62-3).

Lemma 1

If f is unimodal probability density on R^1 then

$$\left|\sum_{m=-\infty}^{\infty} f(x+m)-1\right| \leqslant \sup_y f(y)$$

Lemma 2

Let $f_t(r,\theta)$ be a probability density on $[0,\infty) \times (-\infty,\infty)$, for fixed t. Define $g_t(r) = \int_{-\infty}^{\infty} f_t(r,\theta)d\theta$ and

$$\tilde{f}_t(r,\theta) = \sum_{m=-\infty}^{\infty} f_t(r,\theta+m).$$

Assume

1. $f_t(r,\theta)$ is unimodal in θ for fixed r,t;
2. $\int_0^{\infty} \sup_{\theta} f_t(r,\theta)dr \to 0$ as $t \to \infty$;

then

$$\int_B \tilde{f}_t(r,\theta)d\theta dr - \int_B g_t(r)d\theta dr \to 0 \quad \text{as} \quad t \to \infty$$ for every Borel set B contained in $[0,\infty) \times [0,1]$.

Proof

Let $f_t(\theta|r) = f_t(r,\theta)/g_t(r)$, $\tilde{f}_t(\theta|r) = \tilde{f}_t(r,\theta)/g_t(r)$, then condition 2. is equivalent to

$$\int \sup_{\theta} f_t(\theta|r)g_t(r)d\theta dr \to 0.$$

From lemma 1

$$\sup_{\theta} \int |\tilde{f}_t(\theta|r)-1|g_t(r)dr \leqslant \int \sup_{\theta} f_t(\theta|r)g_t(r)dr \to 0$$

as $t \to \infty$. Thus if B is any Borel set in $[0,\infty) \times [0,1]$

$$\int_B \tilde{f}_t(\theta|r)g_t(r) \quad dr - \int_B g_t(r)d\theta dr \to 0$$

as $t \to \infty$.

This result easily generalises to the case where the number of modes is bounded, below M say. The required generalisation of lemma 1 says $\left|\sum_{m=-\infty}^{\infty} f(x+m)-1\right| \leqslant M \sup_y f(y)$.

To apply lemma 2 to the movements of an animal let $\{\Theta(t), R(t)\}$ be the polar coordinates of the position of the animal at time t. Assume that $\Theta(t)$ is the projection on to a circle of a process $\Psi(t)$ on a helix, see fig. 2. $\Psi(t)$ represents the signed distance from the point 0 along the helix and

$$\Theta(t) = \Psi(t)(2\pi)$$

meaning the least non negative amount that $\Psi(t)$ exceeds an integral multiple of 2π. Intuitively, $\Psi(t)$ can be constructed from $\Theta(t)$, when motion is continuous by representing $\Theta(t)$ as a point on the boundary of a circle, breaking the circle at one point and continuing it into a helix. $\Psi(t)$ is meant to represent the animals' cumulative angular displacement. A rigorous mathematical construction can be performed when $\Theta(t)$ is a function of bounded variation by considering the signed measure induced by the interval function $\mu(s,t] = \Theta(t) - \Theta(s)$.

Let $f_t(r,\psi)$ be the joint density of $\{R(t), \Psi(t)\}$. If the animal regards the space radially symmetrically, moves about a great deal, all over the space, and never distinguishes any part of the space, then a reasonable model would seem to be that $\sup_{\theta} f_t(\theta|r) \to 0$ as $t \to \infty$ expressing the idea that, conditional on $R(t)$, the distribution of $\Psi(t)$ becomes spread out over an infinite interval of the real line. Under mild conditions

$$\int \sup_{\theta} f_t(\theta|r) g_t(r) dr \to 0 \quad \text{as} \quad t \to \infty$$

Because the animal does not distinguish any part of the space, regularly repeated modes are not expected for $f_t(\theta|r)$. Thus a reasonable model seems to be that $f_t(r,\theta)$ satisfies lemma 2. The density of $\{R(t), \Theta(t)\}$ is

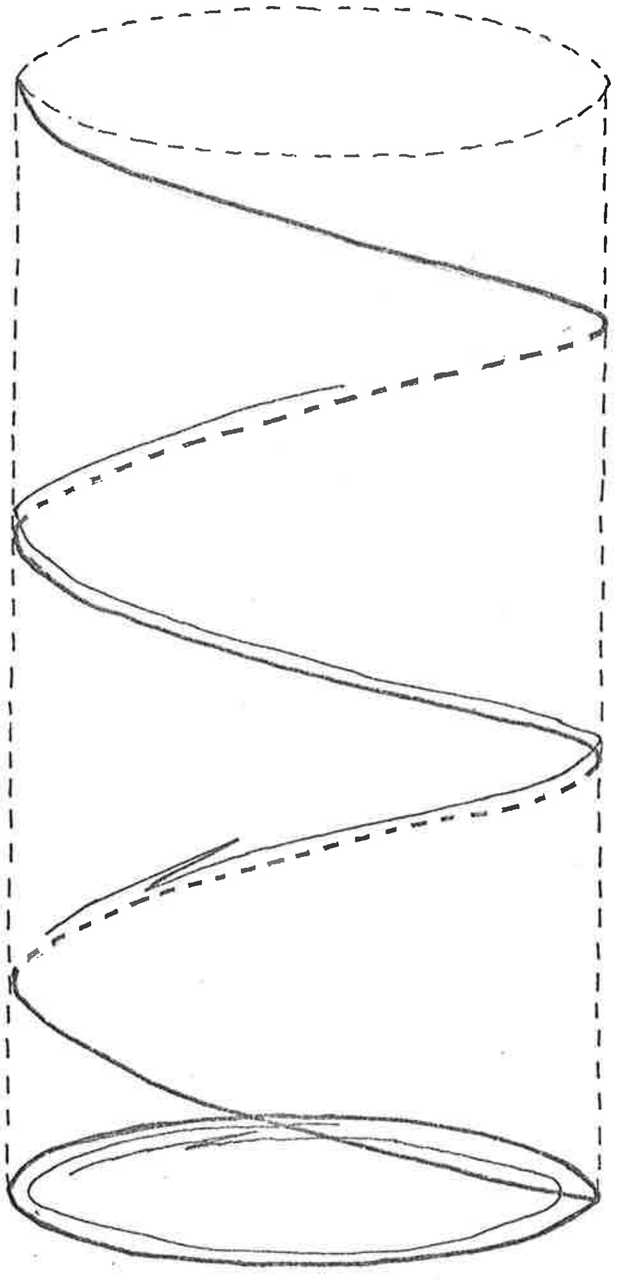


Fig. 2. Projection of helical path onto a circle.

$\sum_{m=-\infty}^{\infty} f_t(r,\theta+2\pi m)$ and so from lemma 2 we conclude that the distribution of the animal's position becomes arbitrarily close to a radially symmetric distribution as $t \to \infty$.

A test of independence

Suppose that $\underset{\sim}{X}(t)$ has a radially symmetric distribution, t is fixed. It is desired to test the hypothesis that $X_1(t),..,X_n(t)$ are independent against alternatives which lead to under or over dispersion of the population.

I have chosen to consider only the case $n = 3$ because it is a case I can handle that allows the possibility of more than two individuals interacting. It is assumed that there are m replicate populations each with 3 individuals and that if $\underset{\sim}{X}_r(t)$ is the positions of the animals for the rth replicate then $\underset{\sim}{X}_1(t),..,\underset{\sim}{X}_m(t)$ are i.i.d. Although the theory applies when X is an arbitrary radially symmetric subset of R^2, it is probably most appropriate when X is a bounded circular area with the $X_\ell(t)$ very roughly uniformly distributed over the area.

The test has an intuitive basis. Consider the triangle with $X_{1r}(t)$, $X_{2r}(t)$, $X_{3r}(t)$ as vertices. If the animals are under dispersed the area of the triangle should tend to be small. If the animals are over dispersed the area should tend to be large.

Let $X_{ir}(t)$ have the polar coordinates (R_{ir},Θ_{ir}). Since $X_{ir}(t)$ has a radially symmetric distribution R_{ir} and Θ_{ir} are independent and Θ_{ir} has a uniform distribution. However no parametric form is assumed for the distribution of R_{ir}. Therefore the approach must be non parametric. Suppose R_{ir} has a continuous distribution function and let L_{ir} be the rank of R_{ir} in the entire sample $\{R_{\ell k}, \ell=1,..,3, k=1,..,m\}$. Define

$$V_{ir} = \sqrt{\frac{L_{ir}}{3m+1}}$$

The distribution of the point with polar coordinates (V_{ir}, Θ_{ir}) approximates that of a point uniformly distributed over the interior of the unit circle. Furthermore if (R_{ir}, Θ_{ir}) is uniformly distributed over the unit circle, as we shall see later, $V_{ir} \overset{a.s.}{\rightarrow} R_{ir}$ as $m \rightarrow \infty$, i,r fixed. Thus if we regard a uniform distribution as the standard, or ideal situation, it seems quite reasonable to replace (R_{ir}, Θ_{ir}) by (V_{ir}, Θ_{ir}). If this standardisation is unacceptable there may be a permutation approach which parallels the development given here.

Define A_r to be the area of the triangle with vertices (V_{1r}, Θ_{1r}), (V_{2r}, Θ_{2r}), (V_{3r}, Θ_{3r}). I suggest that a reasonable statistic for testing independence is

$$T_m = \sum_{r=1}^{m} \alpha(A_r)$$

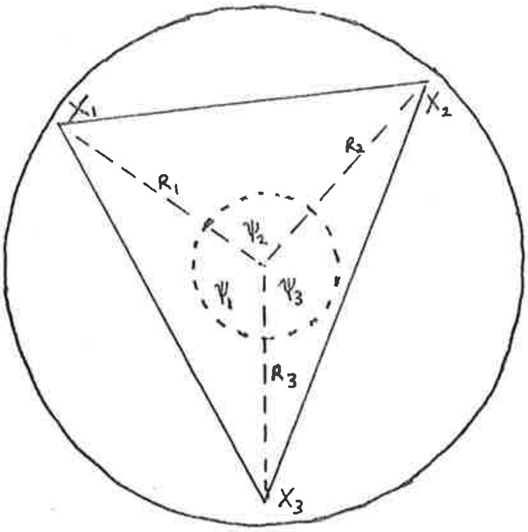
where α is some suitably chosen function. I chose $\alpha(x) = 4x^2$ on the grounds that, with this choice, I can work out the asymptotic distribution of T_m. Notice that A_r is a symmetric function of the veritices of the triangle and so it is a function of N_t.

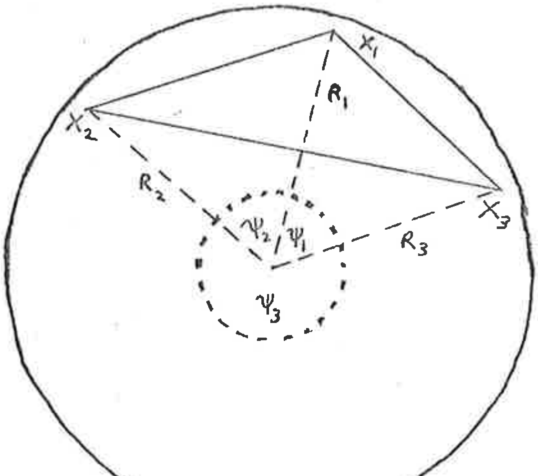
Derivation of the asymptotic distribution of T_m.

Let $X_{1r}(t)$, $X_{2r}(t)$, $X_{3r}(t)$ be i.i.d., for each r. The distribution of T_m does not depend on the distribution of the R_{ir} and so assume that R_{ir} has the distribution $P(R_{ir} < x) = x^2$ or equivalently (R_{ir}, Θ_{ir}) is uniformly distributed over the interior of the unit circle.

The angles Θ_{1r}, Θ_{2r}, Θ_{3r} partition the unit circle into 3 disjoint arcs. Let Ψ_{ir} be the angle subtended by the arc with end points Θ_{ir} and Θ_{i-1r}, subscript arithmetic being modulo 3 representing 0 by 3. It follows that

$$A_r = \tfrac{1}{2}\left|\sum_{i=1}^{3} V_{ir}V_{i-1r}\sin\Psi_{ir}\right|$$





and hence

$$T_m = \sum_{r=1}^{m}\{\sum_{i=1}^{3} V_{ir}^2 V_{i-1r}^2 \sin^2\Psi_{ir} + 2\sum_{i=1}^{3} V_{ir}^2 V_{i+1r}V_{i+2r}\sin\Psi_{1r}\sin\Psi_{i+1r}\}.$$

Note that $\{V_{ir}, i=1,2,3, r=1,..,m\}$ and $\{\Psi_{ir}, i=1,2,3, r=1,..,m\}$ are independent.

We need several lemmas concerning linear rank statistics with random coefficients. These are related to results on linear rank statistics in Hajek and Sidak (1967).

Lemma 1.

Let f be a bounded continuous function of three variables then

$$f(\underset{\sim}{V}_r) - f(\underset{\sim}{R}_r) \xrightarrow{m.s.} 0 \quad \text{as} \quad m \to \infty$$

Proof

Let i,r be fixed. Conditional on R_{ir}, the proportion, p_m, of $R_{i',r'} \leqslant R_{ir}$ converges a.s. to R^2_{ir}. Moreover $V^2_{ir} < p$ iff $p_m \leqslant p$ and $V^2_{ir} > p$ iff $p_m > p$. Thus, conditional on R_{ir}, $V_{ir} \overset{a.s.}{\to} R_{ir}$ as $m \to \infty$, that is $P(V_{ir} \to R_{ir} \mid R_{ir}) = 1$ a.s. Hence $P(V_{ir} \to R_{ir}) = 1$ meaning $V_{ir} \overset{a.s.}{\to} R_{ir}$.

Since f is bounded and continuous it follows that $E\left(f(\underset{\sim}{V}_r) - f(\underset{\sim}{R}_r)\right)^2 \to 0$ as $m \to \infty$. Furthermore the joint distribution of $\underset{\sim}{V}_r$ and $\underset{\sim}{R}_r$ does not depend on r, and so the result holds when r is a function of m.

Introduce a single subscript notation $V_{ir} = V_{i+3(r-1)}$, $R_{ir} = R_{i+3(r-1)}$. Define $\ell^{(k)} = \ell(\ell-1)\ldots(\ell-k+1)$ and

$$C_m = \frac{1}{(3m)^{(2)}} \sum_{i \neq j} R_i^2 \, R_j^2 , \qquad c_m = \frac{1}{(3m)^{(2)}} \sum_{i \neq j} V_i^2 \, V_j^2 .$$

By the strong law of large numbers $C_m \overset{a.s.}{\to} \tfrac{1}{4}$ and we shall see later that $c_m = \tfrac{1}{4} + 0\left(\frac{1}{m}\right)$ so that $C_m - c_m \overset{a.s.}{\to} 0$. Both C_m and c_m are bounded by 1 and so $C_m - c_m \overset{m.s.}{\to} 0$. Combining this result with lemma 1 we have lemma 2.

Lemma 2

$$V^2_{ir} V^2_{i-1\,r} - c_m + C_m - R^2_{ir} R^2_{i-1\,r} \overset{m.s.}{\to} 0 .$$

Lemma 3

Let $g(\underset{\sim}{\Psi}_r) = g(\Psi_{1r}, \Psi_{2r}, \Psi_{3r})$ be a function with finite variance and 0 mean, and f as in lemma 1 then

$$m^{-1/2} \sum_1^m g(\underset{\sim}{\Psi}_r)\left(f(\underset{\sim}{V}_r) - f(\underset{\sim}{R}_r)\right) \overset{m.s.}{\to} 0 .$$

Proof

$\{\underset{\sim}{\Psi}_r, r=1,2,..,m\}$ is independent of $\{\underset{\sim}{V}_r, \underset{\sim}{R}_r, r=1,..,m\}$ and the $\underset{\sim}{\Psi}_r$ are i.i.d. Hence $E[g(\underset{\sim}{\psi}_r)\left(f(\underset{\sim}{V}_r) - f(\underset{\sim}{R}_r)\right) \mid \underset{\sim}{V}_r, \underset{\sim}{R}_r] = 0$ and

$$V\Big[\sum_1^m g(\underset{\sim}{\Psi}_r)\big(f(\underset{\sim}{V}_r)-f(\underset{\sim}{R}_r)\big)\,\Big|\,\underset{\sim}{V}_r, \underset{\sim}{R}_r, r=1,..,m\Big]$$
$$= \big(Eg^2(\underset{\sim}{\Psi}_1)\big)\sum_1^m \big(f(\underset{\sim}{V}_r)-f(\underset{\sim}{R}_r)\big)^2$$

which has expected value

$$mEg^2(\underset{\sim}{\Psi}_1)E\big(f(\underset{\sim}{V}_1)-f(\underset{\sim}{R}_1)\big)^2.$$

Lemma 3 now follows from lemma 1.

Lemma 4

Let $g(\underset{\sim}{\Psi}_r)$ have finite variance and define $\bar{g} = Eg(\underset{\sim}{\Psi}_r)$ then

$$m^{-\frac{1}{2}}\sum_1^m \big(g(\underset{\sim}{\Psi}_r)-\bar{g}\big)(V_{ir}^2 V_{i-1r}^2 - c_m + C_m - R_{ir}^2 R_{i-1r}^2) \xrightarrow{m.s.} 0.$$

Proof

Similar to the proof of lemma 3 with lemma 2 replacing lemma 1.

Lemma 4 applies to $g(\underset{\sim}{\Psi}_r) = \sin^2 \Psi_{ir}$ to give

$$m^{-\frac{1}{2}}\sum_1^m V_{ir}^2 V_{i-1r}^2 \sin^2\Psi_{ir} - m^{-\frac{1}{2}}\sum_1^m (R_{ir}^2 R_{i-1r}^2 + c_m - C_m)\sin^2\Psi_{ir}$$
$$- m^{-\frac{1}{2}}(E\sin^2\Psi_{i1})\sum_1^m (V_{ir}^2 V_{i-1r}^2 - c_m + C_m - R_{ir}^2 R_{i-1r}^2) \xrightarrow{m.s.} 0. \qquad (2)$$

Lemma 5

Let $a(i,j)$ be symmetric in i and j, $|a(i,j)| \leqslant 1$,

$B = \{(i,j,k,\ell)\,|\,i,j,k,\ell$ are distinct and belong to $\{1,..,3m\}\}$,

$C = \{(i,j,k)\,|\,i,j,k$ are distinct and belong to $\{1,..,3m\}\}$,

$D = \{(i,j)\,|\,i \neq j; i,j \in \{1,..,3m\}\}$.

If $\sum_D a(i,j) = 0$ then

$$\Big|\sum_B a(i,j)a(k,\ell) + 4\sum_C a(i,j)a(j,k)\Big| \leqslant 2(3m)^{(2)}$$

Proof

$$\sum_B a(i,j)a(k,\ell) = \sum_{D^2} a(i,j)a(k,\ell) - \sum_{D^2-B} a(i,j)a(k,\ell)$$
$$= -\sum_{D^2-B} a(i,j)a(k,\ell)$$

because $\sum_{D^2} a(i,j)a(k,\ell) = \big(\sum_D a(i,j)\big)^2 = 0.$

$$_{D^2}\sum_B a(i,j)a(k,\ell)$$

$$= \sum_C a(i,j)a(j,k) + a(i,j)a(k,j) + a(i,j)a(i,k) + a(i,j)a(k,i)$$

$$+ \sum_D a(i,j)a(i,j) + a(i,j)a(j,i).$$

Using the symmetry of a this is

$$4\sum_C a(i,j)a(j,k) + 2\sum_D a^2(i,j).$$

The result now follows from

$$2\sum_D a^2(i,j) \leqslant 2\sum_D 1 = 2(3m)^{(2)}.$$

Lemma 6

$$m^{-\frac{1}{2}} \sum_{r=1}^{m} (V_{i\,r}^2 V_{i-1\,r}^2 - c_m + C_m - R_{i\,r}^2 R_{i-1\,r}^2) \xrightarrow{m.s.} 0.$$

Proof

$$E\{m^{-\frac{1}{2}} \sum_1^m (V_{i\,r}^2 V_{i-1\,r}^2 - c_m + C_m - R_{i\,r}^2 R_{i-1}^2)\}^2$$

$$= E(V_{i\,1}^2 V_{i-1\,1}^2 - c_n + C_n R_{i\,1}^2 R_{i-1\,1}^2)^2 \tag{3}$$

$$+ (m-1)E(V_{i\,1}^2 V_{i-1\,1}^2 - c_n + C_n - R_{i\,1}^2 R_{i-1\,1}^2)(V_{i\,2}^2 V_{i-1\,2}^2 - c_n + C_n - R_{i\,2}^2 R_{i-1\,2}^2).$$

The second term above can be evaluated by first conditioning on $\{\tilde{R}_i, i=1,..,3m\}$, the order statistics of $\{R_i, i=1,..,3m\}$, and using the fact that the $V_{i\,r}$ are independent of the order statistics since they are functions of the ranks. The conditional expectation is

$$\frac{m-1}{(3m)^{(4)}} \sum_B \left(\frac{ij}{(3m+1)^2} - c_m + C_m - \tilde{R}_i^2 \tilde{R}_j^2\right)\left(\frac{k\ell}{(3m+1)^2} - c_n + C_m - \tilde{R}_k^2 \tilde{R}_\ell^2\right), \tag{4}$$

with B as in lemma 5.

Defining $a(i,j) = ij/(3m+1)^2 - c_m + C_m - \tilde{R}_i^2 \tilde{R}_i^2$ lemma 5 applies and so expression (4) equals

$$-\frac{4(m-1)}{(3m)^{(4)}} \sum_C a(i,j)a(j,k) \tag{5}$$

with an error at most $2(m-1)/(3m-2)/(3m-3)$. The expected value of (5) equals

$$-4\frac{(3m)^{(3)}(m-1)}{(3m)^{(4)}} E(V_i^2 V_j^2 - c_n + C_n - R_i^2 R_j^2)(V_j^2 V_k^2 - c_n + C_m - R_j^2 R_k^2)$$

The expectation is less than $E(V_{i1}^2 V_{i-11}^2 - c_m + C_m - R_{i1}^2 R_{i-11}^2)^2$.

Thus applying lemma 2 we see that expression (2) tends to 0 as $m \to \infty$ which proves the lemma.

Applying lemma 5 and the statement (2) we obtain

$$m^{-\frac{1}{2}}\sum_1^m V_{ir}^2 V_{i-1r}^2 \sin^2\Psi_{ir} - m^{-\frac{1}{2}}\sum_1^m (R_{ir}^2 R_{i-1r}^2 + c_m - C_m)\sin^2\Psi_{ir} \xrightarrow{m.s.} 0. \tag{6}$$

Now

$$C_m = \frac{1}{(3m)^{(2)}}\left\{\left(\sum_1^{3m} R_i^2\right)^2 - \sum_1^{3m} R_i^4\right\}$$

$$= \left(\frac{1}{3m}\sum_1^{3m} R_i^2\right)^2 + \frac{1}{9m^2(3m-1)}\left(\sum_1^{3m} R_i^2\right)^2 - \frac{1}{(3m)^{(2)}}\sum_1^{3m} R_i^4$$

$$c_m = \frac{1}{(3m)^{(2)}}\left\{\left(\sum_1^{3m}\frac{i}{3m}\right)^2 - \sum_1^{3m}\left(\frac{i}{3m}\right)^2\right\}$$

$$= \tfrac{1}{4} + 0\left(\frac{1}{m}\right)$$

Thus $m^{\frac{1}{2}}(C_m - c_m) - m^{\frac{1}{2}}\left\{\left(\frac{1}{3m}\sum_1^{3m} R_i^2\right)^2 - \frac{1}{4}\right\} \xrightarrow{p} 0$; but $ER_i^2 = \frac{1}{2}$ and so by standard results

$$m^{\frac{1}{2}}\left\{\left(\frac{1}{3m}\sum_1^{3m} R_i^2\right)^2 - \frac{1}{4}\right\} - m^{\frac{1}{2}}\left(\frac{1}{3m}\sum_1^{3m} R_i^2 - \frac{1}{2}\right) \xrightarrow{p} 0$$

or $m^{\frac{1}{2}}(C_m - c_m) - m^{\frac{1}{2}}\left(\frac{1}{3m}\sum_1^{3m} R_i^2 - \frac{1}{2}\right) \xrightarrow{p} 0$. This proves that $m^{\frac{1}{2}}(C_m - c_m)$ converges in distribution. Moreover

$$m^{-\frac{1}{2}}\sum_{r=1}^m (C_m - c_m)\sin^2\Psi_{ir} = m^{\frac{1}{2}}(C_m - c_m)m^{-1}\sum_1^m \sin^2\Psi_{ir}$$

and $m^{-1}\sum_1^m \sin^2\Psi_{ir} \xrightarrow{a.s.} E\sin^2\Psi_{i1}$, which means

$$m^{-\frac{1}{2}}\sum_{r=1}^m (C_m - c_m)\sin^2\Psi_{ir} - m^{\frac{1}{2}}(C_m - c_m)E\sin^2\Psi_{i1} \xrightarrow{p} 0.$$

Putting all this together we obtain

$$m^{-\frac{1}{2}}\sum_{r=1}^m (C_m - c_m)\sin^2\Psi_{ir} - m^{\frac{1}{2}}\left(\frac{1}{3m}\sum_1^{3m} R_i^2 - \frac{1}{2}\right)E\sin^2\Psi_{i1} \xrightarrow{p} 0. \tag{7}$$

It is shown in the appendix, or is obvious by symmetry, that $E\sin^2\Psi_{i1}$ does not depend on i.

In the appendix to this chapter it is shown that $E \sin \Psi_{i\,r} \sin \Psi_{i+1\,r} = 0$. Thus lemma 3 applies to $g(\underset{\sim}{\Psi}_r) = \sin \Psi_{i\,r} \sin \Psi_{i+1\,r}$ to give

$$m^{-\frac{1}{2}} \sum_1^m V_{i\,r}^2 V_{i+1\,r} V_{i+2\,r} \sin \Psi_{i\,r} \sin \Psi_{i+1\,r} - m^{-\frac{1}{2}} \sum_1^m R_{i\,r}^2 R_{i+1\,r} R_{i+2\,r} \sin \Psi_{i\,r} \sin \Psi_{i+1\,r} \xrightarrow{m.s.} 0. \quad (8)$$

Define

$$Z_{1\,r} = \sum_{i=1}^3 R_{i\,r}^2 R_{i-1\,r}^2 \sin^2 \Psi_{i\,r}$$

$$Z_{2\,r} = \sum_{i=1}^3 (R_{i\,r}^2 - \tfrac{1}{2})$$

$$Z_{3\,r} = \sum_{i=1}^3 R_{i\,r}^2 R_{i+1\,r} R_{i+2\,r} \sin \Psi_{i\,r} \sin \Psi_{i+1\,r}.$$

Statements (6), (7) and (8) combine to give

$$m^{-\frac{1}{2}} T_m - m^{-\frac{1}{2}} \sum_{r=1}^m \{Z_{1\,r} - (E \sin^2 \Psi_{1\,1}) Z_{2\,r} + 2Z_{3\,r}\} \xrightarrow{p} 0 \quad (9)$$

as $m \to \infty$. It is clear that $m^{-\frac{1}{2}}(T_m - \text{constant} \times m)$ is asymptotically normally distributed. Using the appendix to this chapter we have proved

Theorem 3

$\left(\frac{19m}{64}\right)^{-\frac{1}{2}} \left(T_m - \frac{3}{8} m\right)$ is asymptotically distributed $N(0,1)$.

It is of interest to know how large m needs to be for this asymptotic distribution to be an adequate approximation. In this regard I have performed some simulation experiments on the University of Adelaide's CDC 6400 computer. I generated values of 1000 independent random variables each having the null distribution of T_m for $m = 60$ and $m = 120$. The value $m = 60$ was chosen because it is the value of m which I had in my laboratory experiments with *Artemia salina*. The results of these experiments, in terms of the first four non central moments of

$$Z_m = (m19/64)^{-\frac{1}{2}}(T_m - 3/8),$$

are given below where $\hat{E}$ denotes the estimated expected value based on a 1000 observations.

m	$\hat{E}\ Z_m$	$\hat{E}\ Z_m^2$	$\hat{E}\ Z_m^3$	$\hat{E}\ Z_m^4$
60	-.023	.93	.29	2.7
120	-.023	1.04	.090	2.9

Figures 3 and 4 give the function $\Phi^{-1}\hat{F}_m$ where $\hat{F}_m$ is the empirical distribution function of Z_m from these experiments and Φ^{-1} is the $N(0,1)$ distribution function. It is apparent that, for $m = 60$, the distribution of Z_m is slightly skewed.

Intuition suggest that the skewness was caused by the convexity of $\alpha(x) = 4x^2$. A concave α probably would increase the rate of convergence to normality. In this regard it seems likely that Lemmas 5 and 6 generalise in a way that allows one to prove asymptotic normality of T_m for a broad class of functions α. However the moment calculations necessary for deriving the asymptotic mean and variance of T_m will be prohibitively difficult in most cases.

With regard to the power of T_m I have simply a few heuristics to offer. Just consider alternatives leading to under dispersion and suppose that the kind of interaction present tends to cause a reduction in the A_i by the proportion ρ so that T_m has asymptotic mean $3/8m\rho^2$ and variance $19m\rho^4/64$. Z_m then has asymptotic mean and variance $-3m^{½}(1-\rho^2)/19^{½}$ and ρ^4. When $m = 60$, for a one sided 5% test with 95% power, we can have ρ no more than .73.

The test of independence based on T_m essentially assumes that the $\underset{\sim}{X}_r$ have the same σ conditional distribution for every value of the environment, σ, or altern-

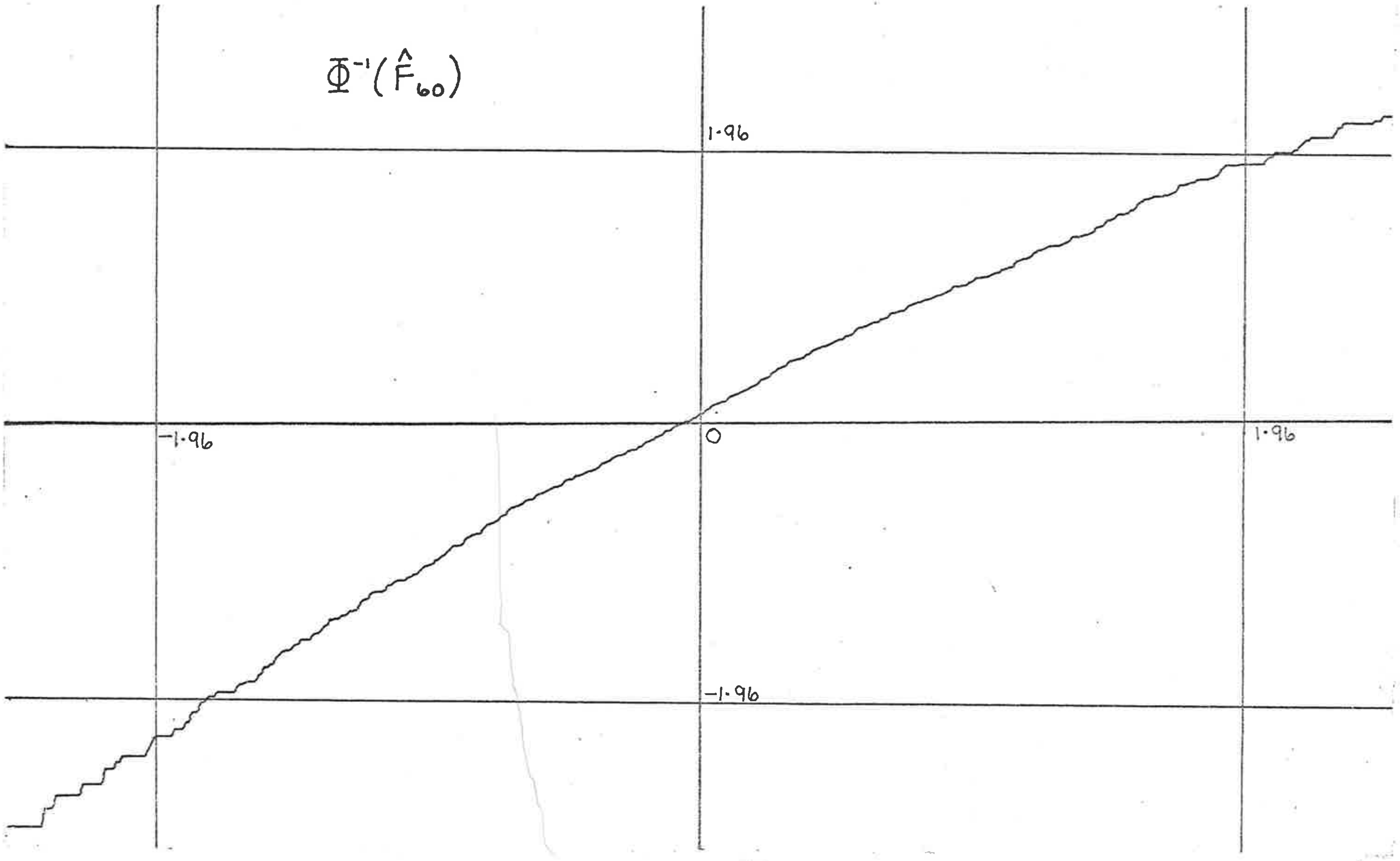


Fig.3.

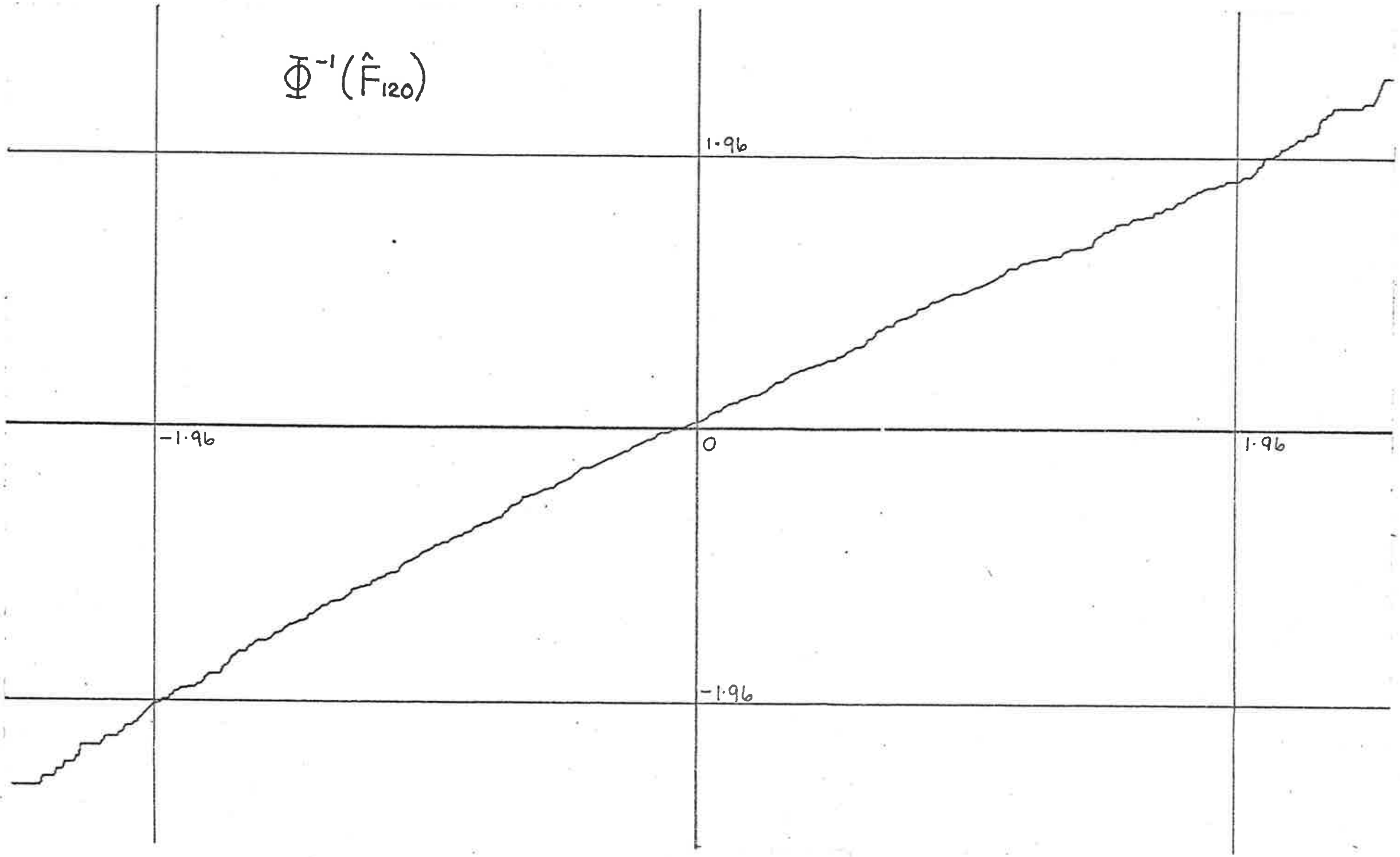


Fig.4.

atively σ takes the same value at each replicate r, $r = 1,..,m$. When this assumption is violated and the X_r, $r = 1,..,m$ are still conditionally radially symmetrically distributed, at least under the null hypothesis, it is possible to base a test of independence on the $\Theta_{i\,r}$ since under the hypothesis of independence conditional on the environment the $\Theta_{i\,r}$' are i.i.d. uniform on the unit circle. Statistics such as $\min_i \Psi_{i\,r}$ may be appropriate measures of dependence.

6.2 Observations on *Artemia salina* (L.)

The Brine Shrimp, *Artemia salina*, is a small crustacean of the order Anostraca. Adults measure about 1 cm. in total length. A concise description of *Artemia salina* is given by Lockhead (1950). Individual *Artemia* swim continuously, seemingly because this is necessary for ventilation of the gills and feeding. Kuenen (1939) and more recent workers have observed that they orientate the ventral surface towards the light so that they generally swim on their backs, and also they exhibit phototaxis which can be either positive or negative depending on the population. The direction of phototaxis can depend on environmental conditions such as temperature.

In South Australia a parthenogenitic population of *Artemia salina* lives in the evaporation ponds of the ICI salt extraction works at Dry Creek. The individuals are of course all females. My studies used a laboratory population established before June in 1975 from Dry Creek animals. The experiments reported here were conducted in 1976.

Individuals in this population showed positive phototaxis.

To improve photography I kept my animals in salinities of 220-250 ‰, as measured on a Hamon S-T bridge, because at such salinities they contain haemoglobin making them bright red rather than pale and translucent. In the sequel water with salinities in this range is referred to as concentrated sea water. To feed my animals I used a common method of adding dead yeast to the water to encourage bacterial growth on which the *Artemia* can feed. However to improve photography by reducing clouding of the water I did not add yeast directly to tanks containing *Artemia*, but to a separate aquarium from which water was transferred daily.

My experiments involved adults and the fifth instar larvae of Heath's description (1925). Note that Andersen (1967) suggests that Heath overlooked some of the very early stages.

Apparatus and Experimental conditions

All experiments were done in a 20°C constant temperature room. In this room I built the two experimental chambers A and B shown in figures 5 and 6 respectively.

In chamber A experimental vessels are placed on the base (b) and a camera is mounted above the rectangular sleeve (s). The walls (w) are constructed of white cardboard leaving a space, to allow observation, between the top of the cardboard and the top of the chamber. The top of the chamber (t) is covered with strong translucent paper leaving a hole beneath the sleeve. On each side of the sleeve (left and right) are mounted a bank of three 40 watt fluorescent tubes, (ℓ), 6 cms. above the top of the chamber. The sleeve is opaque and protects the inside of

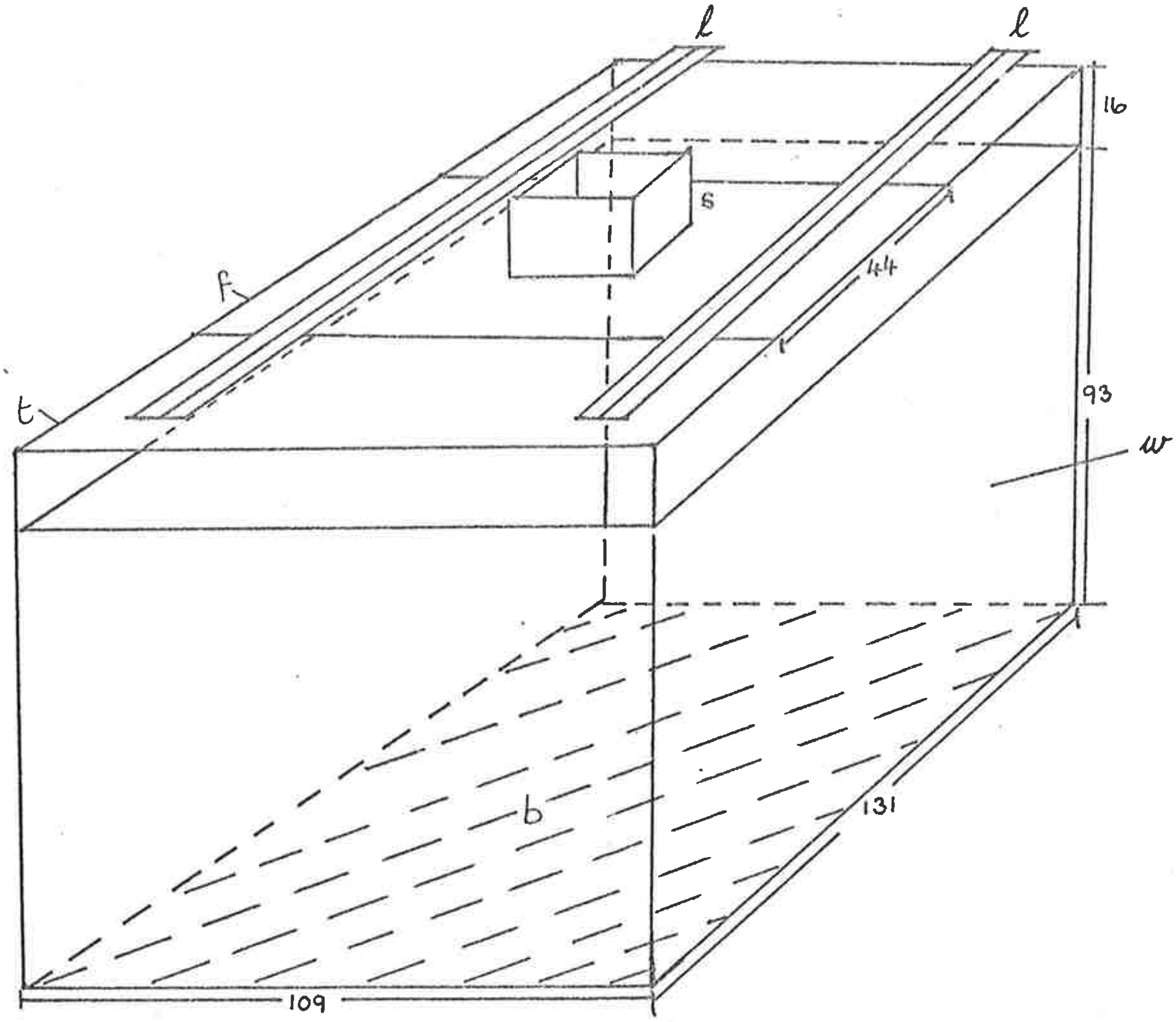


Fig.5. Chamber A (measurements in cm)

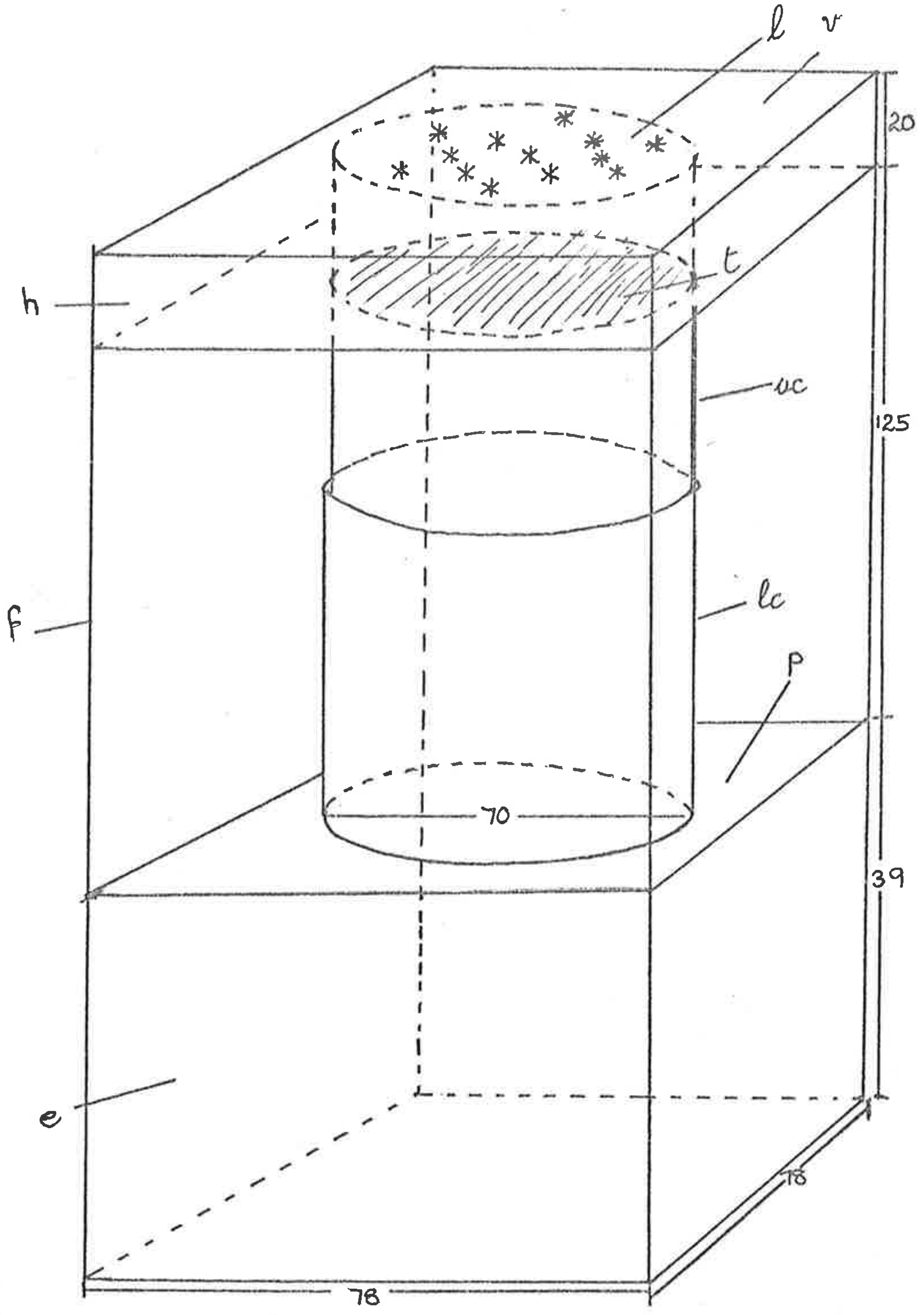


Fig.6. Chamber B (measurements in cm)

the chamber and the camera lens from direct light. To obtain a uniform light distribution over the base, a 44 cm. wide strip of aluminium foil (f) is placed transversely across the translucent paper (t) and also extra strips of translucent paper are placed beneath the fluorescent tubes (ℓ). With this arrangement the actual light intensity over the central 70 × 90 cm. portion of the base is constant at 15 lux as measured with an Eel Lightmaster photometer.

Chamber B is supported by a metal frame (f) and consists of upper and lower cylinders (uc and lc) mounted over a glass platform (p). Both cylinders are constructed of white cardboard lined internally with flat black painted paper. The lower cylinder slides up over the outside of the upper cylinder for access to the platform (p). The top of the upper cylinder is covered with translucent paper (t) and a bank of 13 75 watt Osram Filtalight incandescent light bulbs is mounted 20 cms. above the top of the upper cylinder in the arrangement indicated. The entire undersurface of the top of the frame (u) is covered with aluminium foil and a hood around the lights is formed by black painted paper (h). The frame below the glass is blackened and the space formed by the frame enclosed with black painted paper (e). This arrangement gives a radially symmetric light distribution over the platform beneath the cylinders which is constant at 33 lux inside a circle of radius 20 cm. from the centre of the platform. The intensity drops off outside this circle giving an intensity of 32 lux at 30 cm.

The camera used is a Pentax Spotmatic Motor Drive which automatically winds on film after the shutter has been released. A time clock to release the shutter at fixed intervals was used in the continuous space experiments.

Film was read in the negative form with aid of a Bucone digital film reader connected to a card punch. This instrument has a resolution of .01 mm on 35 mm film .

Testing independence

In an aquarium individual *Artemia* do not appear to interact at all except possibly for avoidance of collision. To test this hypothesis of no interaction, or independence conditional on the environment, I designed two experiments based on the theory of section 6.1. The first experiment tested the symmetry of the environment whilst the second experiment tested independence given the symmetry assumption.

For the first experiment I placed one adult *Artemia* in each of twelve white plastic tubs 12.5 cm. in diameter and 3.4 cm. deep filled with 1.9 cm. of concentrated sea water. The twelve containers were arranged in three rows of four in the central 38 × 50 cm. portion of experimental chamber A to form what I hoped would be 12 radially symmetric environments with the object of the experiment being to test this hypothesis.

After an initial wait of 44 hours for the animals to become accustomed to their homes I photographically recorded their positions at 8.30 a.m. on twenty consecutive days to give a sample $\{X_{ij}\}$ where X_{ij} is the position of the ith animal on the jth day. Under the null hypothesis of a radially symmetric environment we would expect the X_{ij} to be marginally radially symmetrically distributed so that the angles Θ_{ij} are uniformly distributed over the circumference

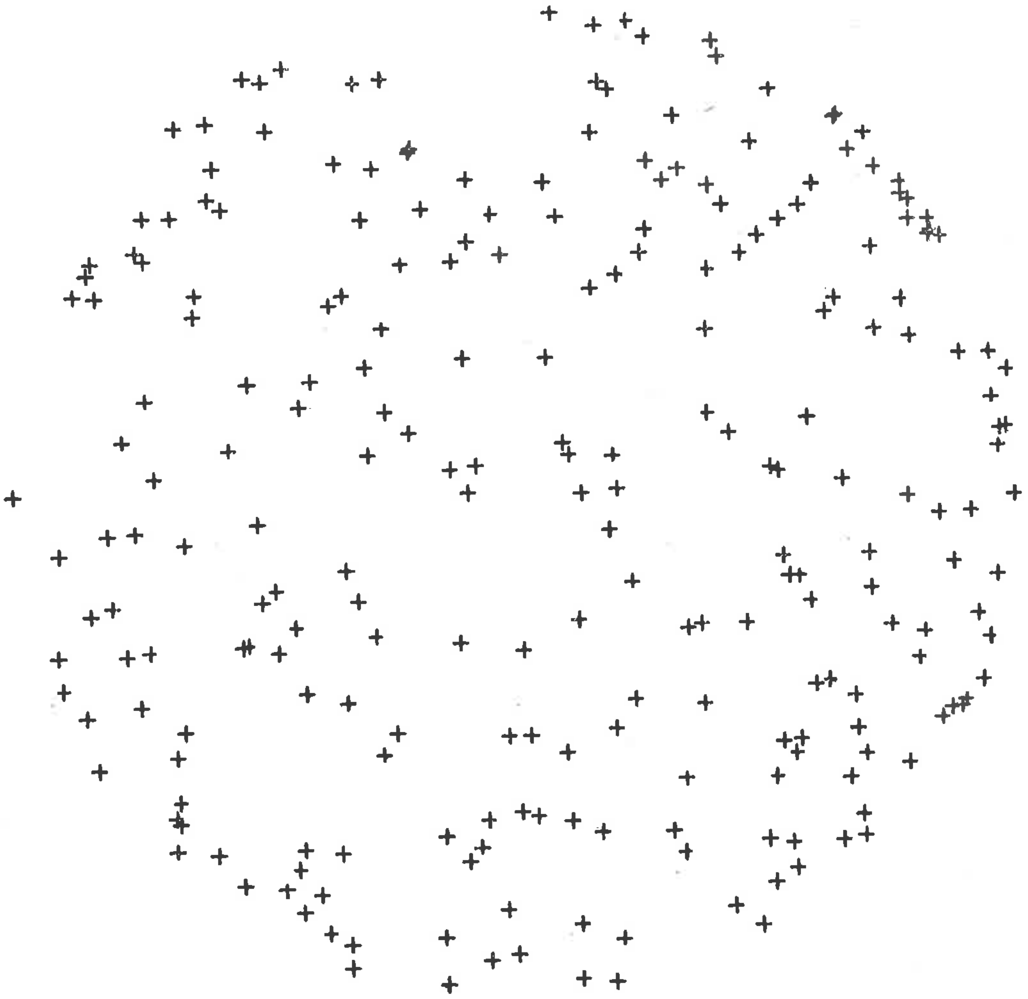


Fig. 7. Scatter plot for the test of radial symmetry: the position of all animals over all days superimposed.

of a circle. I decided to base the test of symmetry simply on the Θ_{ij} because the scale of the space compared with the magnitude of the Θ_{ij} makes it seem extremely unlikely that there could be environmental asymmetry which would not be reflected in the distribution of the Θ_{ij}. As a test statistic I used Kuiper's V_N defined by

$$V_N = \sup_\theta(\theta - F_N(\theta)) - \inf_\theta(F_N(\theta) - \theta)$$

where F_N is the sample distribution function of the Θ_{ij}. Mardia (1972) discusses the use of the statistic for detecting departures from a uniform circular distribution. The value $V_{240} = .085$ obtained is non significant at the 5% level. This significance level is used throughout without further comment.

To perform this test I assumed that since the containers only occupied a small portion of the central part of the chamber, any directional bias would be the same for all containers and also that the Θ_{ij} would be i.i.d. under the hypothesis of radial symmetry. The second assumption might be violated if the animals established asymmetrical home distributions by recognising minor imperfections in the tubs or marking them in some way - the amount of movement which occurs in twenty four hours makes it certain that conditional on the ith animal's home distribution $\Theta_{i1}, \ldots, \Theta_{i20}$ will be i.i.d. Such dependence is of a positive kind which would make the rejection of the null hypothesis more likely which means that this possibility is of no concern given the results obtained.

Having accepted that the twelve plastic tubs provided radially a symmetric environment, I proceeded to test the independence hypothesis. I had an aquarium containing more than 350 adult *Artemia* . Each day for five days I removed 36 *Artemia* from this aquarium at 8.45 a.m. and placed, at random, three in each of the twelve tubs. At 8.30 a.m. the following day their positions were recorded by a single photograph.

For this test of independence it is important that the positions of the animals $\{X_{ir}\}$, in the notation of section 6.1 "A test of independence", should be i.i.d. when no interaction exists. To use the same animals each day would run the risk of the R_{ir} being dependent because some animals may prefer the centre to the outside of the container and vice versa. Thus a new set of animals was selected each day by siphoning them from the aquarium whilst rapidly moving the siphon hose around the aquarium. After each day's sample the previous day's sample was returned to the aquarium so that if any bias was present in the sampling method the population did not become depleted of a particular kind of individual.

In this way I obtained a sample $\{X_{1r}, X_{2r}, X_{3r}\}_{r=1}^{60}$ suitable for applying the test of independence, given in section 6.1. The value of the statistic Z_{60} is .359 which is non significant for an equal tailed test.

<u>Test of discrete space models</u>

To test some of the discrete space ideas with *Artemia salina* I made a small glass bottomed three chambered container depicted in figure 8 and meant to be a discrete space environment for larval *Artemia salina* in Heath's fifth instar. I chose such fifth instar larvae about 2 mm. in length with the aid of a binocular microscope. It is

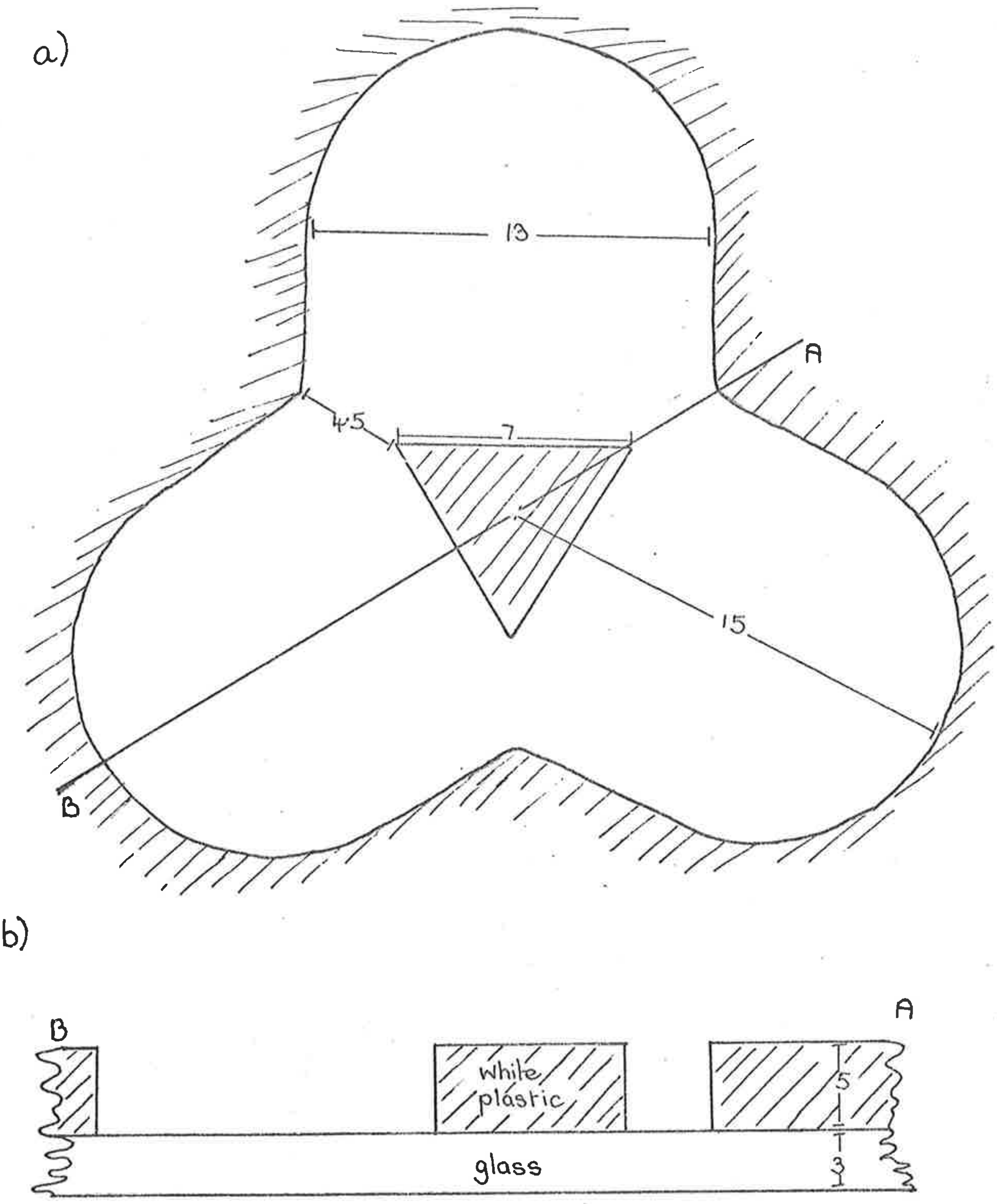


Fig.8. Discrete space

a) from above

b) section through AB

(measurements in mm)

to be admitted that such an environment seems unnatural for *Artemia salina* which is generally free swimming in lakes and ponds and this is perhaps the reason why adult *Artemia* did not appear to like a larger version of this apparatus, generally becoming slower in their movements and finally remaining (being "absorbed") in the one area. However it may not be a completely foreign kind of environment for the larval *Artemia* at Dry Creek which are often found around the gravel covered embankment surrounding the salt ponds. The spaces between the gravel may add a discrete element to the environment of the *Artemia* .

I sought to discover the nature of an individual animal's movements conditional on its phenotype and the environment, or in the notation of Chapter 3, the distribution of X_ℓ given σ and z_ℓ . If we have n animals then, assuming the general model of Chapter 3, $X_1,\ldots,X_n$ are independent but not necessarily identically distributed conditional on σ and $z_1,\ldots,z_n$. This class of joint conditional distributions is the same as the class of joint distributions in the case where σ is constant a.s. and each animal comes from a different subpopulation. Using this result we see from section 3.3 that the distribution of $\{N_t, t \in T\}$, conditional on σ and $z_1,\ldots,z_n$, does not determine the distribution of $X_1,\ldots,X_n$ given σ and $z_1,\ldots,z_n$, except possibly for special models. Thus I needed to either tag the animals individually or study them in isolation. Since one can imagine that the animals may use up some of the oxygen in the water, excrete substances into the water, or even mechanically stir up the water, the distributions of the X_ℓ may depend on n , the number of animals present, even though their movements are possibly conditionally independent given σ . Thus it would seem preferable to study several individuals together in the one container, however there did not seem to be a suitable way of tagging these animals since their cuticle is almost impermeable to most substances. I studied them in isolation.

A fifth instar larva was placed in the discrete space apparatus filled with concentrated sea water and placed in the centre of chamber B at 9 a.m. By lying on my back underneath the whole structure I could observe the animal moving about. I had a digital watch with a resolution of one second and a tape recorder with which to record observations. At 10 a.m. I began recording the time at which the animal changed area and the number of the area to which it changed until approximately 500 changes of area, or "transitions", were observed. This was repeated four times with a different animal each time. The actual number of transitions and total time taken for these four replicates were (511,4h), (410,5h), (530,4h 24 min), (509,3h 3 min). I gave up on the second animal since it did not move for more than 30 minutes - it seemed to be either dying or moulting.

For the purpose of the analysis I divided the observations into two parts, namely the sequences of times between transitions and the sequence of transitions.

In the following all statements are conditional on the phenotypes and the environment.

If for the ℓth animal the area entered at the ith transition is denoted $Y_{i\ell}$ then under the most general discrete space model (4c) the sequence $Y_{1\ell}, Y_{2\ell}, ..$ is a homogeneous discrete time Markov chain. Thus a test of this Markov hypothesis is appropriate. I assumed that $Y_{1\ell}, Y_{2\ell}, \ldots$ would in any case be a stationary sequence. Just watching the animals in this apparatus the Markov hypothesis looks quite reasonable since when an animal enters an area it generally circles around inside the area for a variable time before leaving so that the area from which the animal entered the present area does not affect the choice of the next area. However, on occasions the animal swims in through one opening to an area and passes straight out the opposite opening. On other occasions it turns around immediately it enters an area and goes back from whence it came. Asymmetry between these latter two behaviours suggests a Markov process of second order.

Another hypothesis is environmental symmetry since I went to a good deal of trouble to make the environment symmetric. However, there were seemingly minor imperfections in my discrete space. A final hypothesis is homogeneity, that is does the process $Y_{1\ell}, Y_{2\ell}, \ldots,$ have the same distribution for different animals.

Billingsley (1961) develops the necessary theory for the likelihood ratio tests of these hypotheses and Morgan (1976) discusses recent practice in the animal behaviour field. To reduce the number of test statistics I would have to look at and to increase the power, I included the four replicates in a joint analysis. The modified likelihood for the sample, assuming a Markov chain of order r, is

$$L = \prod_{\ell} \left\{ \prod_{i_0 i_1 \ldots i_r} p_{\ell}(i_0, i_1, \ldots, i_r)^{n_{\ell}(i_0, i_1, \ldots, i_r)} \right\}$$

where $n_\ell(i_0, i_i, .., i_r)$ is the number of subsequences $Y_{t\ell}, Y_{t+1\ell}, .., Y_{t+r\ell}$ equal to $i_0, i_1, .., i_r$ in the entire observation sequence $Y_{1\ell}, .., Y_{m(\ell)\ell}$ and $p_\ell(i_0, i_i, .., i_r)$ is the parameter representing

$$P(Y_{t+r\ell} = i_r | Y_{t+r-1\ell} = i_{r-1}, .., Y_{r\ell} = i_0).$$

Note that $Y_{t\ell} \neq Y_{t+1\ell}$ always. We shall use the notation P to represent the matrix of parameters $\{p_\ell(i_0, i_1, .., i_r)$, $\ell = 1, .., 4$, i_j varying over all possible values$\}$. The likelihood ratio test of the hypothesis

$$P \in \Omega_0 \quad \text{versus} \quad P \in \Omega_1, \ \Omega_0 \subset \Omega_1$$

is based on the statistic

$$\Lambda = \sup\{L|P \in \Omega_1\}/\sup\{L|P \in \Omega_0\}$$

and asymptotically, under appropriate regularity conditions given by Billingsley and trivially satisfied for all cases of interest here, $2 \log \Lambda$ has a χ^2_c distribution where c is the difference in dimension of the two sets Ω_1 and Ω_0. Note that an r-1st order Markov chain is a special case of an rth order process so that a test of Markov order r-1 versus order r is obtained by specifying $\Omega_1 = \{P \text{ is arbitrary}\}$, $\Omega_0 = \{P | p_\ell(i_0, i_1, .., i_r)$ does not depend on $i_0\}$. The hypothesis of spatial symmetry for Markov order r is specified by

$$\Omega_0 = \{P | p_\ell(i_0, i_1, .., i_r) = p_\ell(i_0+1, i_1+1, .., i_r+1)\},$$

where arithmetic is modulo 3, and the hypothesis of homogeneith at rth order is

$$\Omega_0 = \{P | p_\ell(i_0, i_1, .., i_r) \text{ does not depend on } \ell\}.$$

The values of the test statistics for Ω_0 versus Ω_1 for the above possibilities for Ω_0 and Ω_1, and for $r = 1, 2, .., 4$ are tabulated below. I did not go beyond $r = 4$ because of doubts concerning the validity of the χ_c^2 approximation.

Markov Order	χ^2 value	degrees of freedom
	Tests of homogeneity	
1	106	9
2	100 *	18
3	115 *	36
4	143 *	72
	Tests of symmetry	
1	36.0	8
2	59.8 *	16
3	79.0 *	32
4	98.2 *	64
	Markov Order	
1 vs 2	43.1 *	12
2 vs 3	25.6	24
3 vs 4	54.6	48

The results seem fairly conclusive that the animals have different parameters for their movement processes, do not regard my discrete space symmetrically and are second order Markov. In view of the inhomogeneity a decomposition of these results to give individual performances is of interest.

Animal Order	1	2	3	4	d.f.
		Symmetry			
2	7.93	29.31 *	2.61	19.99 *	4
3	8.53	29.59 *	8.58	22.21 *	8
4	10.58	38.57 *	20.88	28.19 *	16
		Markov Order			
1 vs 2	9.48 *	15.03 *	4.67	13.90 *	3
2 vs 3	6.14	5.43	6.11	6.92	6
3 vs 4	5.43	10.08	34.0 *	7.12	12

The value of 34.0 in the test of 3 vs 4 is very strange and suggests to me that either χ^2 approximation has broken down or that the stationarity assumption is false.

The second order transition matrix for animal number 4 can be found in appendix 2. Animal 4 was chosen from a somewhat arbitrary judgement of "typicalness".

From the qualitative behaviour of the animals it was possible to predict a first or second order Markov behaviour with second order Markov holding if there is a significant imbalance between the two behaviours occasionally observed when an animal enters a new area. To test this latter idea I reasoned that elimination of transitions to areas occupied for very small times should result in first order Markov behaviour. Thus I transofmred the data by the following process (the subscript ℓ is sometimes omitted for simplicity of notation).

Let $X_\ell(t)$ be the position (number of the area) occupied by animal ℓ at real time t (t is in seconds) and let T_i be the waiting time for X_ℓ to change value after the ith transition so that $X_\ell(\sum_1^{i-1} T_j) = Y_i$ and $X_\ell(\sum_1^{i} T_j) = Y_{i+1}$ and we regard X_ℓ as right continuous. The number t^* is a fixed constant. Define

$$Y_1^* = Y_1 \text{ , } m^*(1) = 1$$

$$m^*(2) = \text{the first } i \text{ such that } T_{i-1} > t^*$$

$$Y_2^* = Y_{m^*(2)} \text{ , } T_1^* = \sum_1^{m^*(2)-1} T_i$$

The remainders of these sequences are defined recursively

$$m^*(s) = \text{first } i > m^*(s-1) \text{ such that } T_{i-1} > t^*$$

$$Y_s^* = Y_{m^*(s)} \text{ , } T_{s-1}^* = \sum_{m^*(s-1)}^{m^*(s)-1} T_i \text{ .}$$

This transformation treats short visits to areas as though they did not happen - you don't see them if you blink - and essentially treats a short passage from area i to j to k as an alternative route from area i to area k. Real time is maintained.

The above transformation was applied for $t^* = 5,4,3,2,1$ seconds and the statistics for testing Markov order, taking all animals jointly are tabulated below.

Test of Markov order for transformed data

t* order	5	4	3	2	1	d.f.
1 vs 2	11.3	8.57	12.1	29.5 *	40.3 *	12
2 vs 3	16.5	18.0	26.2	30.2	25.4	24
3 vs 4	44.9	37.0	35.4	51.0	51.6	48

The only significant values are for 1 vs 2 for $t^* = 2$ and 1 and the sudden jump in χ^2 value suggests that the second order Markov behaviour is associated with waiting times less than or equal to 3 seconds.

The remaining analyses are all in relation to the transformed data with $t^* = 3$. The numbers of transitions for the four animals in these transformed data are 468, 391, 406 and 450 and the total times taken are the same as for the untransformed data. ~~The tests of homogeneity and symmetry persist.~~

"The tests of homogeneity and symmetry for this data, under the assumption of a first-order Markov process, gave the χ^2 values 59.5 and 54.2 on 9 and 8 degrees of freedom. We see that inhomogeneity and asymmetry persist."

In the transformed data we have agreement with model 4c when the time between transitions is eliminated. Many models apart from 4c are consistent with this result, for example the semi Markov model (Cane, 1959, suggests application of semi Markov models to animal behaviour). To gain some idea of the further applicability of 4c I suggest the following model for the data, which I believe is a close approximation to 4c amenable to analysis. Conditional on $Y_1^*, Y_2^*, \ldots$ the waiting times $T_1^*, T_2^*, \ldots$ are a sequence of independent exponential random variables with

$$E[T_i^* | Y_i^*] = \lambda(Y_i^*) f(i)$$

where f is some function depending on the animal and λ is a function of the area occupied - the waiting time refers to leaving this area - and λ also depends on the animal. Note that we are conditioning on random activity as well as the environment and the phenotype. The chief departure from 4c is that f is a function of the number of transitions that have taken place rather than real time.

Clearly not much can be learnt if f is very irregular. We might expect (hope) the environment to be almost constant since I designed it that way! However a gradual change can be expected mainly as an increase in salinity as the water evaporates through the day. Simply watching the animals move about in my discrete space strongly suggests periods of activity and inactivity but we can not tell what is imagination and what else is going on that is imperceptible.

I divided the analysis of the waiting times into two stages. First testing whether the sequences $\{T_i^*, Y_i^* = j\}$, $j = 1,2,3$, $\ell = 1,.,4$, that is the sequences of waiting times for departure from a particular area and for a particular animal, are sequences of i.i.d. random variables. The null hypothesis in this case is consistent with both a stationarity semi Markov and the model proposed here with f constant.

I chose a test based on the total number of runs above or below the median (e.g. Gibbons, 1971, pp. 51-58). This test has its main advantage in simplicity. It is also allegedly sensitive to the kinds of alternatives of interest here. Other tests are more sensitive in particular situations. To perform the test I extracted the sequences

$$\{T^*_{1j}, T^*_{2j}, \ldots\}, \; j = 1,2,3,$$

for each animal, where, for a particular animal T^*_{ij} is the waiting time to leave state j for the ith time the animal occupies state j. If M^*_j is the median of $T^*_{1j}, T^*_{2j}, \ldots$ then a run is a subsequence $T^*_{sj}, T^*_{s+1j}, \ldots, T^*_{s+rj}$ of maximal length such that the elements of the subsequence are either all above the median or all below. Define L^*_j = total number of runs in $T^*_{1j}, T^*_{2j}, \ldots$ and $L^* = L^*_1 + L^*_2 + L^*_3$. Under the null hypothesis the L^*_i are independent and L^* is asymptotically normally distributed with mean and variance to be found in Gibbons. The values of L^* for the different animals, standardised to zero mean and unit variance are given below

animal	1	2	3	4
Standardised L*	-5.58	-1.73	-6.59	-6.45

The clear impression is non randomness and in all cases a deficiency of runs, that is a tendency to remain on one side or other of the median for long periods. We can not accept a stationary semi Markov model or the model proposed here with f constant. The next stage of the investigation is to see if the exponential model for the distribution of the T_i^* is appropriate. Of course the transformation ensures that $T_i^* > 3$ which must cause a departure from exponentiality at the origin. One can argue a case for subtraction of 3 from each of the T_i^* but I have not made this adjustment.

Specify $\Sigma f(i)$ equal to the number of transitions so that f and λ become identifiable. Assuming f changes slowly we can use the sample mean, $\hat{\lambda}(j)$, of $T_{1j}^*, T_{2j}^*, \ldots$ to estimate $\lambda(j)$. The sizes of these samples are more than 100 in all cases. Making the transformation $T_i^{**} = T_i^*/\hat{\lambda}(j)$ when $Y_i^* = j$ we obtain a sequence $T_1^{**}, T_2^{**}, \ldots$ such that, under the model, the T_i^{**} are approximately independent with exponential distributions and $E[T_i^{**}] = f(i)$. The values of $\hat{\lambda}(j)$ obtained are tabulated below

	$\hat{\lambda}(j)$			
Animal Area	1	2	3	4
1	28.5	52.7	22.5	26.95
2	36.2	33.8	16.4	25.9
3	27.7	38.0	18.0	19.7

We can gain some idea of the function f, when it is reasonably smooth, from moving averages applied to the T_i^{**}. Also, if $U_1, U_2, \ldots, U_s$ are i.i.d. and marginally exponential then $U = U_i / \sum_1^s U_j$ has the distribution

$$P(U<u) = 1 - (1-u)^{s-1} .$$

Thus if f is reasonably smooth we can expect

$$T_{i+r}^{**} / \sum_{j=i}^{i+s-1} T_j^{**}$$

to have the above distribution approximately when $0 \leqslant r \leqslant s - 1$ suggesting yet another transformation of the data.

Define $T_i^{***} = T_{i+r}^{**} / \sum_{j=i}^{i+s-1} T_j^{**}$ with $r = [\frac{1}{2}s]$.

The sample distribution function for the sequence

$$\{1-(1-T_i^{**})^{s-1}\}$$

was calculated for $s = 2,5,10,20,40,80$ and for each animal. Under the model the sample distribution should be approximately a straight line through the origin for s not too large. How big s can be for this result to hold depends on the smoothness of f. The sample distribution functions for $s = 2$, 10 and 20 are given in appendix 2. The other plots are little different except that the flat portion at the top of graphs for $s = 2$ are hardly apparent at all for $s > 2$. Note that the flat portions near the origin and also at the top of the graphs for $s = 2$ are a necessary consequence of the transformation resulting in $T_i^* > 3$. Otherwise the fit to straight line through the origin is remarkably good. The sample distribution function for the sequences

$$\{1-\exp(-T_i^{**})\}$$

are also given in appendix 2. The deviation from an exponential is not as great as I expected. It seems that the average of the distributions of the T_i^{**} is not too far from exponential.

If s is reasonably large and for this s

$$\hat{f}(i+[\tfrac{1}{2}s]) = \frac{1}{s} \sum_{j=i}^{s-1} T_j^{**}$$

is not too biassed an estimator of $f(i+[\frac{1}{2}s])$ then we expect under our model that the T_i^{***} will be approximately i.i.d. Note that if s is small the T_i^{***} and T_j^{***} will tend to be negatively dependent for i close to

j so that if we calculate the total number of runs L^{***} for the sequence $\{T_i^{***}\}$ a surplus is expected over the number of runs from an i.i.d. sequence of the same length. For s large a deficit is expected as $\hat{f}(i)$ becomes a poor estimator of $f(i)$. This behaviour is observed but I do not think that this is necessarily evidence for the model, for this behaviour might occur in other situations as well. Simulation studies might be helpful in this area. For $s = 20$, L^{***} becomes near that expected. The table below gives the observed values standardised to unit mean and variance.

Animal	1	2	3	4
Standardised L^{***}	-.756	.623	.408	0.00

The function $\hat{f}$ for $s = 20$ is given in appendix II.

I have not been able to find any serious deficiencies in the model so far. As well as the simulation studies mentioned above, a final check on the model might be to compare the empirical distribution of the T_i^{**} with the average of the distribution function $1 - \exp\{-\hat{f}(i)x\}$ over all i. I have not done this.

The analysis of the discrete space experiments given above show reasonable agreement with the model 4c when appropriately transformed, and in all aspects investigated. Part of the model that was not investigated was independence between the $\{T_i^{**}\}$ or $\{T_i^{***}\}$ and the $\{Y_i^*\}$ which seems to be a difficult problem. A limitation on the investigation was the fact that individuals were studied in isolation. Now that a model has been obtained for the behaviour of an individual there is more scope for investigating the effect of the environment and the variation in behaviour from individual to individual. Theorem 9 of chapter 3 suggests that, if we condition on the environment we may be able to obtain the desired information by studying non tagged individuals in a common environment.

Continuous space experiments

Artemia salina lives in continuous space, in fact in large open areas. Therefore a natural study of the movements should be in continuous space. In designing an experiment one meets various physical limitations for example inside a building one tends to be limited by floor to ceiling height. Then there are factors such as resolution of photography, and the amount of film the camera can hold. Many such limitations can be overcome if one has sufficient time and/or money. From a consideration of such factors I chose a circular glass bottomed dish illustrated in figure 9. In the sequel this is referred to as the continuous space. The continuous space was meant to be a radially symmetric environment in order to simplify the analysis. The continuous space was filled with concentrated sea water and placed in the centre of Chamber B. The camera was mounted beneath, loaded with more than 250 frames of black and white 35 mm film.

The object of the investigation was the conditional distribution of X_ℓ given z_ℓ and for the same reasons as in the discrete space experiments I studied single individ-

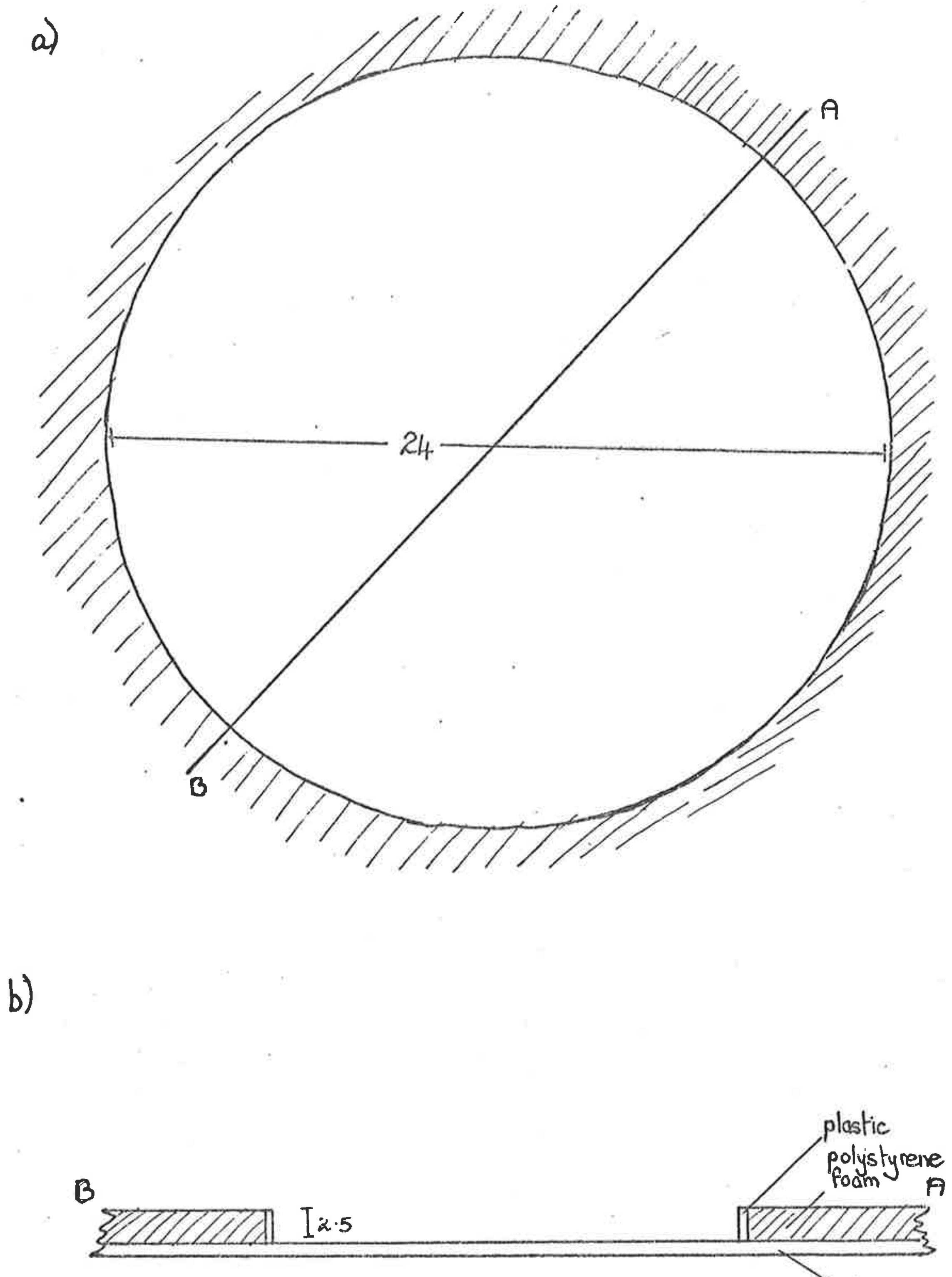


Fig.9. Continuous space

a) from above

b) section through AB

(measurements in cm)

uals in isolation. The basic hypothesis was that the distribution of X_ℓ given z_ℓ conforms to the model in section 3.5.

The analysis that I present for the data described below is far from complete, in particular it does not make full use of available statistical techniques. It is intended primarily to indicate the kinds of hypotheses that I think are important and to illustrate a way in which the theory in this thesis is helpful in studying animal movements.

There were four replicates of the experiment with the following method.

A single adult *Artemia salina* was placed in the continuous space at 9.00 a.m. It was hoped that by 4.00 p.m. she would be accustomed to this environment and her behaviour would have stabilised and for a short period, say less than 30 minutes after that time, we may be able to assume that X_ℓ is a stationary process. The environment was constant as far as I could tell. At 4.00 p.m. the camera was started. It took one photograph every 6 seconds until the film ran out. The actual numbers of observations obtained were 266, 237, 267 and 274. The 237 is a result of camera failure. The number of observations for the ℓth experiment will be denoted by $m(\ell)$. Defining $m'(\ell) = m(\ell) - 1$ the sequence of observations for the ℓth animal is

$$X_\ell(0), X_\ell(1), \ldots, X_\ell\big(m'(\ell)\big).$$

An isolated *Artemia* in my continuous space tends to move in small circles, with a certain amount of error. Sometimes the circling is entirely clockwise or anticlockwise, sometimes a mixture of both and sometimes there can be spurts of activity where the circling breaks down. The overall effect seems to be an undirected wandering about. This suggests that zero drift Brownian motion might be a reasonable underlying model for application of the ideas in section 3.5 and chapter 4. Zero drift Brownian motion, then, will take the place of the process Y of section 3.5. The transition density function for zero drift reflecting Brownian motion in the unit circle satisfies the differential equation

$$\frac{\partial}{\partial t}\ \hat{p}(t,\underset{\sim}{x},\underset{\sim}{y}) = \tfrac{1}{2}\sigma^2\left(\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}\right)\hat{p}(t,\underset{\sim}{x},\underset{\sim}{y})$$

with the boundary conditions $\frac{\partial}{\partial n}\hat{p}(t,\underset{\sim}{x},\underset{\sim}{y}) = 0$ at $|\underset{\sim}{x}| = 1$ where $\frac{\partial}{\partial n}$ means the derivative with respect to $\underset{\sim}{x}$ in the direction of the normal to the boundary at $\underset{\sim}{x}$.

The solution to this equation, obtained for an equivalent problem in heat flow, is given in Gray and MacRobert (1922, pp.91-94). Writing this solution in terms of the transition operator P^t in the notation of chapter 4 we have

$$P^t f = \sum_{m=0}^{\infty} e^{-\lambda_n t} <f,\xi_m> \xi_n$$

where the measure π is $\frac{1}{\pi}\times$Lebesgue measure (the second π is the number π), and in terms of polar coordinates the sequence $\{\xi_m(r,\theta)\}$ equals the set

$$\{h_{pq}J_p(\alpha_{pq}r)\cos p\theta;\ p=0,1,\ldots;q=1,2,\ldots\}$$
$$\cup\ \{h_{pq}J_p(\alpha_{pz}r)\sin p\theta;\ p=1,2,\ldots;q=1,2,\ldots\}$$

J_p being the Bessel function of the first kind of order p α_{pq} is the qth non negative zero of the derivative of J_p. $h_{pq} = |J_p(\alpha_{pq})|\{\tfrac{1}{2}(1-p^2/\alpha_{pq}^2)\}$ is the constant giving ξ_n unit norm (Watson, 1966 p.577). Finally $\lambda_m = \sigma^2\alpha_{pq}$ when $\xi_m(r,\theta) = h_{pq}J_p(\alpha_{pq}r)\cos p\theta$ or $h_{pq}J_p(\alpha_{pq}r)\sin p\theta$.

Spiegel (1968) tabulates α_{pq}.

Scatter plots for the positions of the animals are given in appendix 3. It is immediately apparent that the animals are not uniformly distributed over the area. They keep away from the boundary. For this reason a basic process of zero drift Brownian motion may not be very suitable. Later I shall suggest a possibly better process on which to base the analysis. To achieve a more uniform spatial distribution I transformed the data replacing $X_\ell(t) = (R_\ell(t), \Theta_\ell(t))$ with $X_\ell^*(t) = (V_\ell(t)/(m(\ell)+1))^{1/2}, \Theta_\ell(t))$ $= (R_\ell^*(t), \Theta_\ell(t))$ where $V_\ell(t)$ is the rank of $R_\ell(t)$ in the entire sequence $R_\ell(0), R_\ell(1), \ldots$. A second feature of the plots is radial asymmetry. One is led to the suspicion that the space is not regarded radially symmetrically. Certainly if the points were i.i.d. in each plot one would come to this conclusion. However there is a fair degree of dependence present and so this impression might be wrong. It seems appropriate to try to use a test which is likely to be sensitive to departures most likely to upset the model we are going to try to impose based on Brownian motion.

Now if the marginal density of $X_\ell^*(t)$, with X_ℓ assumed to be a stationary process, is absolutely continuous with respect to π, and has a square summable Radon Nikodym derivative with respect to π then since the sequence $\{\xi_n\}$ is a basis for $\mathcal{L}^2(\pi)$ (Watson, 1966, section 18.3 proves the required completeness of the Bessel functions) $E\xi_n(X_\ell^*(t)) = 0$ for all $n \geq 1$ implies that the distribution of $X_\ell^*(t)$ is π. Strictly speaking the distribution of $X_\ell^*(t)$ is singular with respect to π because $R_\ell^*(t)$

is discrete. However the distribution of $X_\ell^*(t)$ surely closely approximates a distribution which is absolutely continuous relative to π with a square summable Radon-Nikodym derivative for which radial symmetry is exactly equivalent to equality with π. Thus I suggest as a test of radial symmetry comparing $\bar{\xi}_m = \frac{1}{m(\ell)} \Sigma \xi_m(X_\ell^*(t))$ with zero for a finite number of the ξ_m. In future we shall write $\xi_m(t)$ for $\xi_m(X_\ell^*(t))$. Four of the functions ξ_m which seem most suitable for testing radial symmetry are
$\xi_1(r,\theta) = h_{11} J_1(\alpha_{11} r) \sin\theta$, $\xi_2(r,\theta) = h_{11} J_1(\alpha_{11} r) \cos\theta$
$\xi_2(r,\theta) = h_{21} J_2(\alpha_{21} r) \sin 2\theta$, $\xi_4(r,\theta) = h_{21} J_2(\alpha_{21} r) \cos 2\theta$
(the labelling of the ξ_m is in increasing order of magnitude of α_{pq}). The sample means for these functions are tabulated below

function animal	$\bar{\xi}_1$	$\bar{\xi}_2$	$\bar{\xi}_3$	$\bar{\xi}_4$
1	-.189	.043	.246	-.081
2	.289	-.200	-.135	-.046
3	.203	-.022	-.069	.122
4	.001	-.122	.180	.003

We need some way of assessing the significance of the observed deviations from 0. First of all, under the null hypothesis we might expect that the vector

$$\sqrt{m(\ell)}\ (\bar{\xi}_1, \ldots, \bar{\xi}_4)' = \sqrt{m(\ell)}\ \underset{\sim}{\bar{\xi}},$$

say, is asymptotically multinormal. Hannan (1970) gives a central limit result in the case of generalised moving averages and Heyde (1974) gives a quite general central limit result for stationary processes which includes Hannan's result. A result more obviously of relevance to our case is a central limit theorem for Markov processes given by Doob (1953, pp.221-232). Zero drift relecting Brownian motion in a circle satisfies Doob's theorem. However a more appropriate central limit result for our use would be for a process obtained by random time transformation of a Markov process.

If we assume that $\sqrt{m(\ell)}\ \bar{\underset{\sim}{\xi}}$ is approximately multinormal we are left with the problem of consistently estimating its variance matrix. The theoretical variance of $\sqrt{m(\ell)}\ \bar{\underset{\sim}{\xi}}$ is

$$V\underset{\sim}{\xi}(0) + \sum_{t=1}^{m'(\ell)}\left(1-\frac{t}{m(\ell)}\right)\{C\left(\underset{\sim}{\xi}(0),\underset{\sim}{\xi}(t)\right) + C\left(\underset{\sim}{\xi}(t),\underset{\sim}{\xi}(0)\right)\} \qquad (7)$$

However substituting the sample estimates of $C\left(\underset{\sim}{\xi}(0),\underset{\sim}{\xi}(t)\right)$ in this formula does not give a consistent estimator. The observed sample estimates of the $C\left(\underset{\sim}{\xi}(0),\underset{\sim}{\xi}(t)\right)$ fluctuate about 0 for $t > 10$ with most elements of the matrix less than $.1$ in absolute value. Assuming that the impression that the covariances are near enough to 0 for $t > 10$ is correct, then truncation of the sum (7) at 10 gives a consistent estimator $\hat{\Sigma}_{10}$, say, of a matrix close to $V\sqrt{m(\ell)}\ \bar{\underset{\sim}{\xi}}$ and we might expect $m(\ell)\ \bar{\underset{\sim}{\xi}}\ \hat{\Sigma}_{10}^{-1}\ \bar{\underset{\sim}{\xi}}$ to be an approximate χ_4^2 random variable. We may be able to do much better in the estimation of $V\sqrt{m(\ell)}\ \bar{\underset{\sim}{\xi}}$ by using the theory in Hannan (1970) particularly using the idea of smoothed spectral estimation. However a statement at the beginning of Chapter 5 section 4 of Hannan's book suggests that we may not have sufficient observations to justify the use of these methods. In any case I have simply used $\hat{\Sigma}_{10}$ and the results are given below. The calculation for $\hat{\Sigma}_{15}$ and $\hat{\Sigma}_5$ is included for comparison

Animal	1	2	3	4	
$m(\ell)\bar{\underset{\sim}{\xi}}\ \hat{\Sigma}_{10}^{-1}\bar{\underset{\sim}{\xi}}$	23.0	7.10	9.90	33.4	χ_4^2 5% point
$m(\ell)\underset{\sim}{\xi}\ \hat{\Sigma}_{15}^{-1}\underset{\sim}{\xi}$	55.9	7.49	10.9	94.8	9.49
$m(\ell)\underset{\sim}{\xi}\ \hat{\Sigma}_{5}^{-1}\underset{\sim}{\xi}$	16.4	7.18	8.45	49.4	

The variation in values between different truncation points gives some reason for suspicion but I will not assume that $X_\ell^*(t)$ has the measure π as an approximate marginal distribution in the sequel. Note that the animals appear to have different home distributions.

The main hypothesis we want to test is whether the distribution of X_ℓ can be obtained by an independent random time transformation of a Markov process. Suppose the transition operator for this Markov process Y, say, can be expressed in the form

$$P^t f = \sum_0^\infty e^{-\lambda_n t} \langle f, \xi_n^* \rangle\ \xi_n^* \tag{8}$$

We want to know if $X_\ell(t) = Y(A(t))$ where A is a process with stationary increments independent of Y.

We shall see later that a qualitative examination of the observed movements in these experiments throws doubt on the self adjoint nature of the representation (8). However we can still see what the theory we have developed is able to tell us, in particular, we can see if this doubt is supported by our analysis.

Proceeding under the assumption that $X_\ell(t) = Y(A(t))$ with Y having transition operator (8) we have

$$E\xi_n^*(X(s))\,\xi_m^*(X(t)) = \varphi_{t-s}(\lambda_m)\,\delta_{mn}.$$

φ_t being the Laplace transform of $A(t)$. We shall write

$\xi_m^*(t)$ for $\xi_m^*(X_\ell(t))$ in the sequel. It should be possible to approximate $\xi_m^*(t)$ by a finite linear combination of the $\xi_m(t) - E\xi_m(t)$ in a similar manner to the approximation behind the Rayleigh-Ritz method discussed in Chapter 4. We use a modification of the suggestion in section 4.3, Bivariate distributions to take account of the fact that $\xi_1(t), \xi_2(t), \ldots$ is not an orthonormal sequence. Letting $\underset{\sim}{\xi}(t)$ represent a vector consisting of a fixed finite set, d in number say, of the $\xi_m(t) - E\xi_m(t)$ we want to find $\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, \ldots$ such that, writing $\xi_m^*(t) = \underset{\sim}{\alpha}_m'\underset{\sim}{\xi}(t)$, we have $V\xi_m^*(t) = 1$ and $\underset{\sim}{\alpha}_m$ is a value of $\underset{\sim}{\alpha}$ maximising

$$C\left(\underset{\sim}{\alpha}'\underset{\sim}{\xi}(0), \underset{\sim}{\alpha}'\underset{\sim}{\xi}(1)\right)$$

over the set of $\underset{\sim}{\alpha}$ such that $C\left(\underset{\sim}{\alpha}'\underset{\sim}{\xi}(0), \xi_n^*(0)\right) = 0$ for $n < m$. The process is initialised by putting $\xi_0^* = 1$. We have a classical canonical correlation problem of multivariate analysis. To estimate the $\underset{\sim}{\alpha}_m$ we simply obtain a symmetric estimate of $C\left(\underset{\sim}{\xi}(0), \underset{\sim}{\xi}(1)\right)$ from the formulae $\hat{\Sigma}_s(1) = \frac{1}{2}\{\hat{\Sigma}(1) + \hat{\Sigma}(1)'\}$,

$$\hat{\Sigma}(1) = \frac{1}{m(\ell)} \sum_{t=0}^{m'(\ell)-1} \left(\underset{\sim}{\xi}(t) - \bar{\underset{\sim}{\xi}}\right)\left(\underset{\sim}{\xi}(t+1) - \bar{\underset{\sim}{\xi}}\right)'$$

where $\bar{\underset{\sim}{\xi}}$ is the sample mean of $\underset{\sim}{\xi}(t)$. If $\hat{\Sigma}(0)$ is the ordinary sample variance of $\underset{\sim}{\xi}(0), \underset{\sim}{\xi}(1), \ldots$ then we seek the canonical variables for the first and second sets of size d for random variables with the variance matrix

$$\left[\begin{array}{c:c} \hat{\Sigma}(0) & \hat{\Sigma}_s(1) \\ \hdashline \hat{\Sigma}_s(1) & \hat{\Sigma}(0) \end{array}\right]$$

By standard methods we can find $A = [\underset{\sim}{\alpha}_1, \ldots, \underset{\sim}{\alpha}_d]$ such that $A\hat{\Sigma}(0)A' = I$ and $A\hat{\Sigma}_s(1)A'$ is diagonal. This is the solution to our problem.

To make up $\underset{\sim}{\xi}(t)$ I chose $\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6, \xi_7$ and

ξ_{16} where $\xi_5(r,\theta) = h_{02}J_0(\alpha_{02}r)$, $\xi_6(r,\theta) = h_{31}J_3(\alpha_{31}r)\sin 3\theta$, $\xi_7(r,\theta) = h_{31}J_3(\alpha_{31}r)\cos 3\theta$, $\xi_{16}(r,\theta) = h_{03}J_0(\alpha_{03}r)$. This seemed to be the smallest set which would have a reasonable chance of approximating the first few of the ξ_m^* when our model is correct. The values of A, for each animal are given in appendix 3.

For the purpose of judging the adequacy of this model the sample estimates $\hat{\Sigma}^*(t)$ of $C\left(\underset{\sim}{\xi}^*(0), \underset{\sim}{\xi}^*(t)\right)$, $\underset{\sim}{\xi}^*(t) = A\underset{\sim}{\xi}(t)$, are given in appendix 3 for $t = 1,2,3$. In judging the fit it should be remembered that the accuracy should decrease towards the right and towards the bottom of these matrices. The theoretical expectation is a diagonal matrix and they do appear roughly diagonal. I can suggest no test of fit but a test may be suggested by the asymptotic theory for estimated covariances and spectra in Hannan's book. An alternative method of assessing the model would be to calculate A form the matrix

$$\left[\begin{array}{c:c} \hat{\Sigma}\ (0) & \hat{\Sigma}_\xi(t) \\ \hdashline \hat{\Sigma}_s(t) & \hat{\Sigma}\ (0) \end{array}\right]$$

for several values of t. A should be approximately independent of t.

Recall that the diagonal elements of $\hat{\Sigma}^*(t)$ estimate $\varphi_t(\lambda_m)$ so that if we accept the model proposed we can start thinking about a model for U_t.

On the other hand we may not accept the model proposed. There are two ways the model can be false. First X_ℓ may not have the representation $X_\ell(t) = Y\left(A(t)\right)$ or the transition operator for Y may not have a representation of the form (8). Therefore we can not reject the model $X_\ell(t) = Y\big(A(t)\big)$ on the basis of the results given above which assumes that (8) holds.

In appendix 3 there is a plot of X_2^* with the positions at successive points in time joined to show the path followed. From this plot there appears to be a tendency for moving around the continuous space in a clockwise direction. This impression is supported by the plots for the other animals. A crude model for this sort of behaviour might be to model $Y(t) = \big(\tilde{R}(t), \tilde{\Theta}(t)\big)$, in polar coordinates, by the equation $\tilde{R}(t) = R(t)$, $\tilde{\Theta}(t) = \Theta(t) + t\theta$ where the process $\big(R(t), \Theta(t)\big)'$ is zero drift reflecting Brownian motion in the circle with initial distribution π (see earlier) so that the process is stationary. In this case defining

$$\xi_{m_1}(r,\theta) = \alpha_{pq} J_p(\alpha_{pq} r)\sin p\theta$$

$$\xi_{m_2}(r,\theta) = \alpha_{pq} J_p(\alpha_{pq} r)\cos p\theta$$

we have

$$E\xi_{m_1}\big(Y(0)\big)\xi_{m_2}\big(Y(t)\big) = -e^{-Y\sigma^2\alpha_{pq}}\sin t\theta$$

and

$$E\xi_{m_2}\big(Y(0)\big)\xi_{m_1}\big(Y(t)\big) = e^{-t\sigma^2\alpha_{pq}}\sin t\theta$$

The anti symmetry shows that Y does not have a transition operator of the form (8). Also, using the notation of chapter 4, if we model Y as a diffusion process with $\Sigma(x) \equiv \sigma^2 I$, σ constant, and $\mu\big(\alpha(t)\big)' \frac{d}{dt}\alpha(t) > 0$ for some closed curve $\alpha(t)$ we see that the vector A can not be irrotational.

For a test of the representation $X_\ell(t) = Y\big(A(t)\big)$ which does not assume (8) we may use asymmetric canonical correlation procedures. Let $P = P^{\frac{1}{s}}$ be the $\frac{1}{s}$ step transition operator for Y; s is a positive integer.

Let us assume that P is compact, we know from Ringrose (1971 p.58) that P can be approximated by an operator of finite rank. The relevant Hilbert space is $\mathcal{L}^2(\pi^*)$, π^* being the common marginal distribution of $Y(t)$ for each t, and it is now appropriate to regard $\mathcal{L}^2(\pi^*)$ as a complex Hilbert space. It will be possible to approximate P on the subspace orthogonal to the constant space by the operator P_d defined by

$$P_d f = \underset{\sim}{\xi}' B \underset{\sim}{a}$$

where $f = \underset{\sim}{\xi}' a$, the components of $\underset{\sim}{\xi}$ are a finite subset of the functions $\xi_1 - E\xi_1(t)$, $\xi_2 - E\xi_2(t), \ldots$ and B is some matrix. Assuming that A is semi simple we have $A = G\,\Lambda\,G^{-1}$ where Λ is diagonal and $P_d^r = G\,\Lambda^r G^{-1}$. Writing $\underset{\sim}{\xi}(t)$ to be the vector $\underset{\sim}{\xi}\big(X^*(t)\big)$ as before we expect from the model that

$$E[\underset{\sim}{\xi}(t)\,|\,\underset{\sim}{\xi}(0)] \doteq G(E\Lambda^{s\,A(t)})G^{-1}\underset{\sim}{\xi}(0).$$

We can estimate G and $E\,\Lambda^{s\,A(t)}$ from the equation $G(E\Lambda^{s\,A(t)})G^{-1} = \hat{\Sigma}^{-1}(0)\hat{\Sigma}(t)$ so that a test of the model can come from a comparison of the estimates of G for different values of t or alternatively from the approximation or by comparing

$$\hat{G}^{-1}\hat{\Sigma}^{-1}(0)\hat{\Sigma}(t)\hat{G}$$

to a diagonal matrix when $\hat{G}$ is the estimate of G obtained from $\hat{\Sigma}^{-1}(0)\hat{\Sigma}(1)$. I have not performed the necessary calculations for this analysis.

For the analyses I have presented above the functions ξ_m were obtained from a Brownian motion model. Considering the preference of the animals for the centre of the continuous space it may have been better to obtain the sequence $\{\xi_m\}$ from another diffusion type process in which such a central tendency exists, for example the process obtained from a d-dimensional Ornstein-Uhlenbeck process discussed in section 4.3, Bivariate distributions.

As has been pointed out Markov processes arise as models of animal movements from the asymptotic behaviour of random walks and some other processes. An underlying Markov model for *Artemia salina* was suggested by its habit of moving in small circles. At intervals of six seconds not much information is available about the behaviour. This was deliberate because the Markov approximation would certainly break down if such information were available. My personal suspicion is that the circles were too large relative to the size of the continuous space for the underlying Markov model to be a good one. Probably the best way of finding this out is to examine the movements in finer detail.

Conclusion

This thesis is about the development of models for the scientific investigation of animal movements. One can try to investigate movements theoretically but at some stage a calibration of theory with practice must occur. This chapter is about that calibration process. In particular it aims to show how the theory of previous chapters is applied and is useful in learning about animal movements.

Artemia salina was chosen as a convenient animal for study in the laboratory. Because the study of *Artemia salina* is primarily exemplary we are not too concerned that some parts of the study are based on observations of a few individuals and that experiments are done in rather artificial circumstances. We are more concerned that the right kinds of observations are obtained for the stated purpose. The theory of chapter 3 provided guidance in this direction.

In spite of these reservations it has been possible to find out a good deal more about the few individuals studied than has come from the majority of animal movement studies. I hope that it will be recognised that this has been possible because the experiments are designed and analysed in relation to models.

The analysis of the continuous space experiment is left incomplete but even so our models have enabled the identification of useful questions and allowed us to think about ways of testing the relevant hypotheses. The analysis of the discrete space experiments is more complete and provides a reasonable basis for the next stage in the investigation, that is separating the three different kinds of variation discussed in section 2.4. These are environmental, between individual and within individual.

Above all I hope that it will be seen that mathematical models are useful ~~Subjects~~ OBJECTS.

APPENDIX 1 : FORMULAE FOR CHAPTER 6

Joint distribution of $(\Psi_{1r}, \Psi_{2r}, \Psi_{3r})$

To derive the joint distribution of $(\Psi_{1r}, \Psi_{2r}, \Psi_{3r})$ for the test of independence we use a familiar argument conditioning on Θ_{1r}. The subscript r is omitted in the sequel. Since $\Theta_1, \Theta_2, \Theta_3$ i.i.d. uniform, Θ_2 and Θ_3 are uniformly distributed on $[0, 2\pi)$ conditional on Θ_1. Putting $\Phi_2 = \Theta_2 - \Theta_1 (2\pi)$, $\Phi_3 = \Theta_3 - \Theta_1 (2\pi)$ we see that Φ_2 and Φ_3 are i.i.d. uniform on $[0, 2\pi)$, conditional on Θ_1. This conditional distribution is the same for all values of Θ_1 and so it is also the marginal distribution. It can now be seen that (Ψ_1, Ψ_2, Ψ_3) has the same distribution as the spacings between uniform order statistics and so from Feller (1971, p.75) Ψ_1 and Ψ_2 have the joint density $2!/(2\pi)^2$ on the set $\{(\varphi_1, \varphi_2) \mid \varphi_1, \varphi_2 \geqslant 0, \varphi_1 + \varphi_2 \leqslant 2\pi\}$. Note that Ψ_1, Ψ_2, Ψ_3 are exchangeable random variables.

Some moment formulae

We need the integral formulae

$$\int_0^{\pi/2} \sin^{p-1} x dx = 2^{p-2} \frac{\left(\Gamma\left(\frac{p}{2}\right)\right)^2}{\Gamma(p)} \quad , \quad p > 0$$

and

$$\int_0^{\pi} x \sin^q x dx = \frac{\pi^2 \Gamma(q+1)}{2^{q+1} \left(\Gamma(\tfrac{1}{2}q+1)\right)^2} \quad , \quad q > -1$$

given by Ryshik and Gradstein (1963, pp.157 and 167).

The marginal density of Ψ_1 is

$$\frac{1}{2\pi^2} \int_0^{2\pi-\varphi} d\theta = \frac{2\pi-\varphi}{2\pi^2} \quad , \quad 0 \leqslant \varphi \leqslant 2\pi$$

$$2\pi^2 \; E \sin^2 \Psi_1 = \int_0^{2\pi} (2\pi-\varphi) \sin^2 \varphi d\varphi$$

$$= \int_0^{2\pi} \varphi \sin^2 \varphi d\varphi$$

$$= \int_0^{\pi} \varphi \sin^2 \varphi d\varphi + \int_0^{\pi} (\varphi+\pi) \sin^2 \varphi d\varphi$$

$$= 2\int_0^{\pi} \varphi \sin^2 \varphi d\varphi + \pi\int_0^{\pi} \sin^2 \varphi d\varphi$$

$$= 2\pi^2\Gamma(3)/\left(2^3\Gamma^2(2)\right) + \pi^2/2$$

$$= \pi^2$$

Hence $E \sin^2 \Psi_1 = \frac{1}{2}$ (1)

$$2\pi^2 E \sin^4 \Psi_1 = \int_0^{2\pi} \varphi \sin^4 \varphi d\varphi$$

$$= 2\int_0^{\pi} \varphi \sin^4 \varphi d\varphi + \pi\int_0^{\pi} \sin^4 \varphi d\varphi$$

$$= 2\pi^2 4!\left(2^5\Gamma^2(3)\right)^{-1} + \pi 2^4\Gamma^2(2\tfrac{1}{2})/4!$$

$$= \frac{3}{4}\pi^2$$

Hence $E \sin^4 \Psi_1 = \frac{3}{8}$

$$2\pi^2 E \sin \Psi_1 \sin \Psi_2 = \int_0^{2\pi} \int_0^{2\pi-\varphi} \sin \varphi \sin \theta d\theta d\varphi$$

$$= \int_0^{2\pi} \sin \varphi (1-\cos \varphi) d\varphi$$

$$= 0$$

Hence $E \sin \Psi_1 \sin \Psi_2 = 0$ (2)

$2\pi^2 E \sin^2 \Psi_1 \sin \Psi_2 \sin \Psi_3$

$$= \int_0^{2\pi} \int_0^{2\pi-\varphi} \sin^2 \varphi \sin \theta \sin (2\pi-\theta-\varphi) d\theta d\varphi$$

$$= -\int_0^{2\pi} \sin^2 \varphi \int_0^{2\pi-\varphi} \sin \theta \sin (\theta+\varphi) d\theta d\varphi$$

$$= -\tfrac{1}{2}\int_0^{2\pi} \sin^2 \varphi \int_0^{2\pi-\varphi} \cos \varphi - \cos(\varphi+2\theta) d\theta d\varphi$$

$$= -\tfrac{1}{2}\int_0^{2\pi} \sin^2 \varphi \{(2\pi-\varphi) \cos \varphi - \sin(2\pi-\varphi)\} d\varphi$$

Since the integrand is odd about π the value of the integral is 0.

Hence $E \sin^2 \Psi_1 \sin \Psi_2 \sin \Psi_3 = 0$ (3)

$$2\pi^2 \; E \sin^2\Psi_1 \sin^2\Psi_2 = \int_0^{2\pi} \int_0^{2\pi-\varphi} \sin^2\theta \sin^2\varphi \, d\theta d\varphi$$

$$= \tfrac{1}{4}\int_0^{2\pi} \int_0^{2\pi-\varphi} \{\cos(\varphi-\theta) - \cos(\varphi+\theta)\}^2 d\theta d\varphi$$

Make the substitution $\alpha = \varphi-\theta$, $\beta = \varphi+\theta$ to obtain

$$\tfrac{1}{8}\int_0^{2\pi} \int_{-\beta}^{\beta} \{\cos \alpha - \cos \beta\}^2 \, d\alpha d\beta$$

$$= \tfrac{1}{4}\int_0^{2\pi} \int_0^{\beta} \{\cos \alpha - \cos \beta\}^2 \, d\alpha d\beta$$

$$= \tfrac{1}{4}\int_0^{2\pi} \int_0^{\beta} 1 - \sin^2\alpha + 1 - \sin^2\beta - 2 \cos \alpha \cos \beta \, d\alpha d\beta$$

$$= \pi^2 - \tfrac{1}{4}\int_0^{2\pi} \beta \sin^2 \beta d\beta - \tfrac{1}{4}\int_0^{2\pi} \int_0^{\beta} \sin^2\alpha d\alpha d\beta - \tfrac{1}{2}\int_0^{2\pi} \sin \beta \cos \beta d\beta$$

$$= \pi^2 - \tfrac{1}{4}\pi^2 - \tfrac{1}{4}\int_0^{2\pi} \sin^2 \alpha \int_{\alpha}^{2\pi} d\beta d\alpha$$

$$= \tfrac{3}{4}\pi^2 - \tfrac{1}{4}\int_0^{2\pi} \alpha \sin^2 \alpha d\alpha$$

$$= \tfrac{1}{2}\pi^2$$

Hence $E \sin^2\Psi_1 \sin^2\Psi_2 = \tfrac{1}{4}$ (4)

$$2\pi^2 \; E \sin^3\Psi_1 \sin \Psi_2 = \int_0^{2\pi} \sin^3\varphi \int^{2\pi-\varphi} \sin \theta d\theta$$

$$= \int_0^{2\pi} \sin^3\varphi \; (1-\cos \varphi) d\varphi$$

Since the integrand is odd about π the integral is 0.

Hence $E \sin^3\Psi_1 \sin \Psi_2 = 0$ (5)

The above formulae apply to the evaluation of the mean and variance of $Z_{1r} - (E \sin^2\Psi_{11}) Z_{2r} + 2Z_{3r}$. Note that the distribution of R_{ir}^2 is uniform on $(0,1)$, the R_{ir} are i.i.d., and are independent of the Ψ_{ir}.

Reiterating

$$Z_{1r} = \sum_{i=1}^{3} R_{ir}^2 R_{i-1\,r}^2 \sin^2\Psi_{ir}$$

$$Z_{2r} = \sum_{i=1}^{3} (R_{ir}^2 - \tfrac{1}{2})$$

$$Z_{3r} = \sum_{i=1}^{3} R_{ir}^2 R_{i+1\,r} R_{i+2\,r} \sin \Psi_{ir} \sin \Psi_{i+1\,r}.$$

Now $$E\,Z_{1\,r} = 3(ER_{i\,r})^2\ E\sin^2\Psi_{i\,r}$$

$$= \tfrac{3}{8}$$

$$E\,Z_{2\,r} = 0$$

$$E\,Z_{3\,r} = 0$$

$$V\,Z_{1\,r} = 3VR_1^2R_2^2\sin^2\Psi_1 + 6C(R_1^2R_2^2\sin^2\Psi_2\,,R_1^2R_3^2\sin^2\Psi_1)$$

$E\ R_1^4R_2^4\sin^4\Psi_1 = \left(\tfrac{1}{3}\right)^2\tfrac{3}{8} = 1/24$, therefore $VR_1^2R_2^2\sin^2\Psi_1 = 1/24 - 1/64$

$E\ R_1^4R_2^2R_3^2\sin^2\Psi_1\sin^2\Psi_2 = 1/48$,

therefore $C(R_1^2R_2^2\sin^2\Psi_2\,,R_1^2R_2^2\sin^2\Psi_1) = 1/48 - 1/64$.

Hence $$V\,Z_{1\,r} = \tfrac{1}{4} - 9/64.$$

$$V\,Z_{2\,r} = 3VR_1^2 = \tfrac{1}{4}$$

$$V\,Z_{3\,r} = 3ER_1^4R_2^2R_3^2\sin^2\Psi_1\sin^2\Psi_2 + 6ER_1^3R_2^3R_3\sin\Psi_1\sin^2\Psi_2\sin\Psi_3$$

$$= 1/16$$

$$C(Z_{1\,r}\,,Z_{2\,r}) = 6C(R_1^2\,,R_1^2R_2^2\sin^2\Psi_2)$$

$$= 6ER_1^4R_2^2\sin^2\Psi_2 - \tfrac{1}{2}\cdot\tfrac{1}{8}$$

$$= \tfrac{1}{8}$$

$$C(Z_{1\,r}\,,Z_{3\,r}) = 3E\{R_1^2R_1^3R_3^3\sin^3\Psi_2\sin\Psi_1 + R_1^4R_2^3R_3\sin^3\Psi_1\sin\Psi_2 + R_1^3R_2^2R_3^3\sin^2\Psi_1\sin\Psi_2\sin\Psi_3\}$$

$$= 0$$

$$C(Z_{2\,r}\,,Z_{3\,r}) = 0$$

Thus $$V\left(Z_{1\,r} - (E\sin^2\Psi_{1\,1})Z_{2\,r} + 2Z_{3\,r}\right) = \tfrac{1}{4} - 9/64 + 1/16 + \tfrac{1}{4} - \tfrac{1}{8}$$

$$= 19/64$$

Also $$E\left(Z_{1\,r} - (E\sin^2\Psi_{1\,1})Z_{2\,r} + 2Z_{3\,r}\right) = \tfrac{3}{8}$$

This completes the proof that

$$(m\,19/64)^{-\frac{1}{2}}(T_m - \tfrac{3}{8})$$

is asymptotically distributed as $N(0,1)$.

APPENDIX 2 : DISCRETE SPACE EXPERIMENTS

Second order transition probabilities ($p_{i\,j\,k}$)
for animal 4.

		k		
i = 1		–	–	–
	j	0.403	0.000	0.597
		0.553	0.447	0.000
i = 2		0.000	0.671	0.329
		–	–	–
		0.768	0.232	0.000
i = 3		0.000	0.794	0.206
		0.568	0.000	0.432
		–	–	–

first order transition probabilities

for Y^*_ℓ, $t^* = 3$ seconds

animal 1

0.000	0.470	0.530
0.582	0.000	0.418
0.490	0.510	0.000

animal 2

0.000	0.703	0.300
0.600	0.000	0.400
0.584	0.416	0.000

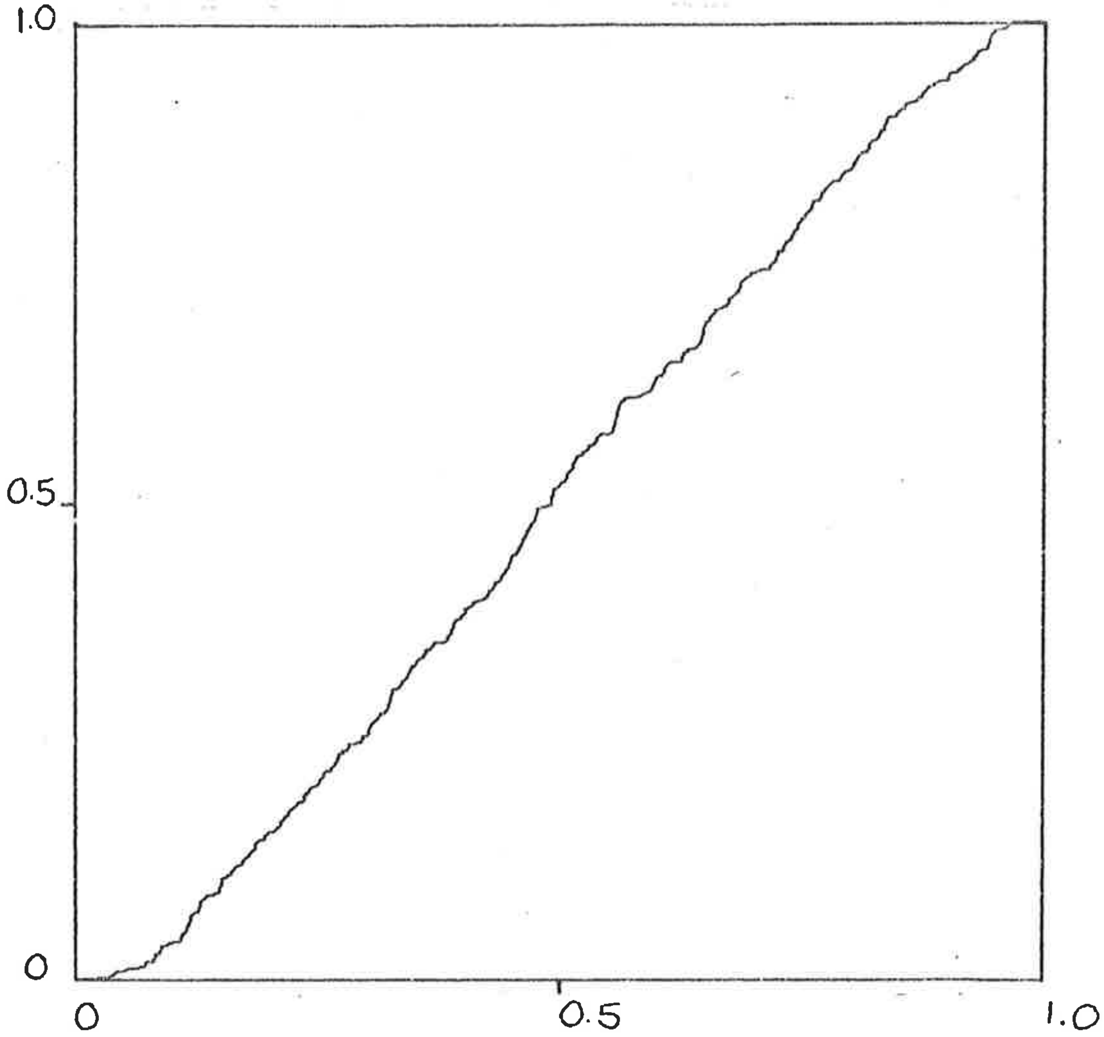
animal 3

0.000	0.585	0.415
0.500	0.000	0.500
0.630	0.370	0.000

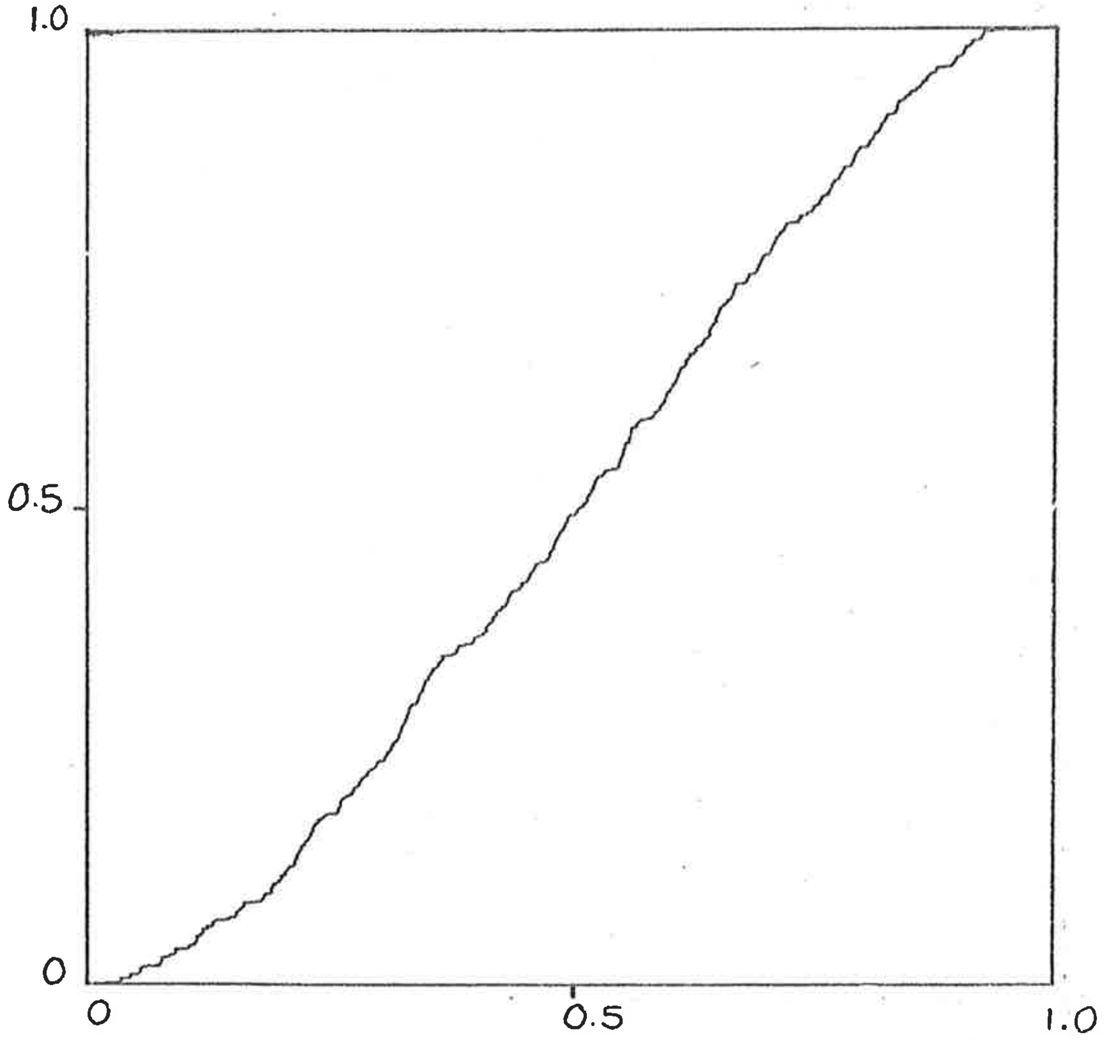
animal 4

0.000	0.767	0.233
0.460	0.000	0.540
0.714	0.286	0.000

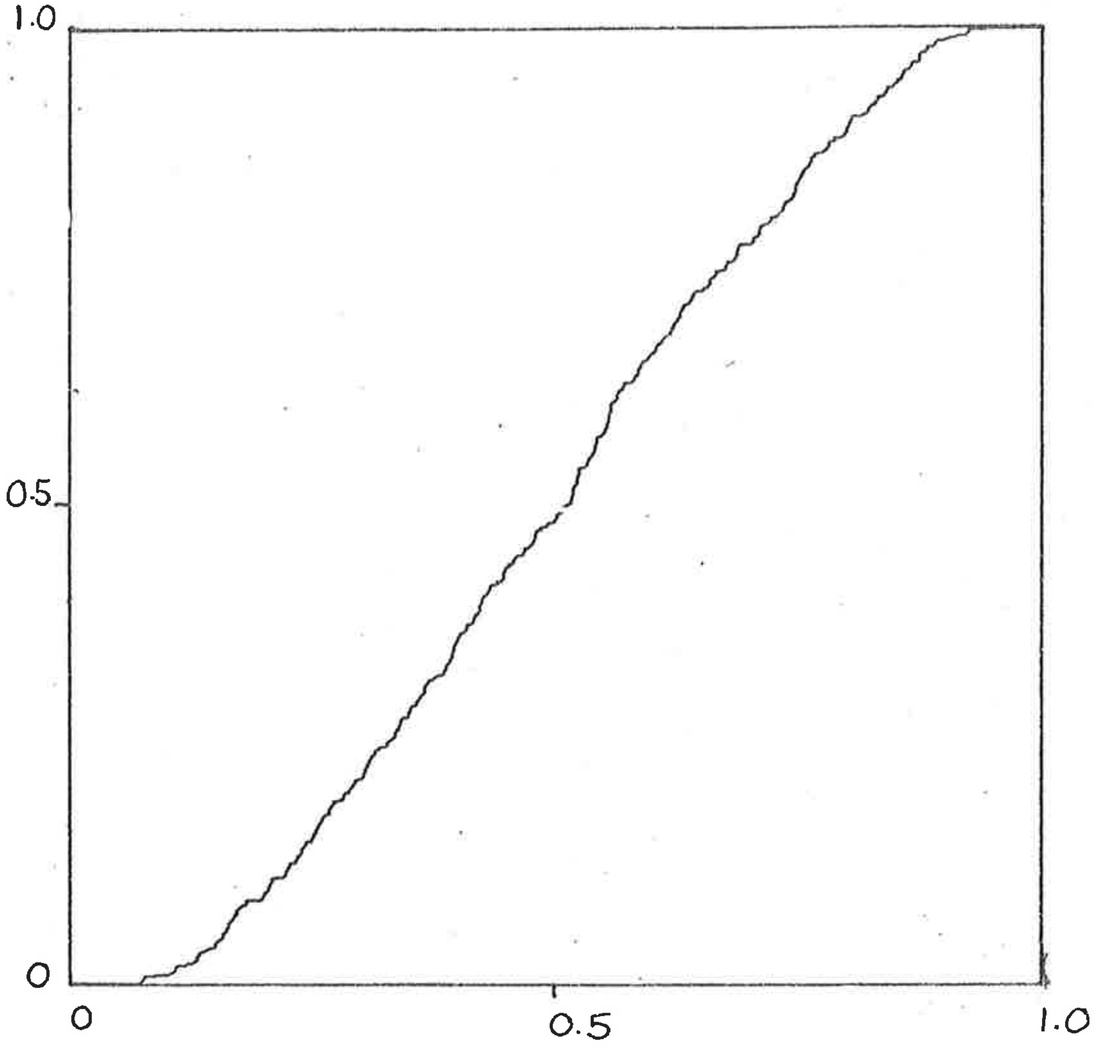
On the following pages are given the sample distribution functions of $\{1-(1-T_i^{**})^{s-1}\}$ for $s = 2$, 10 and 20 and for each animal $\ell = 1,..,4$.



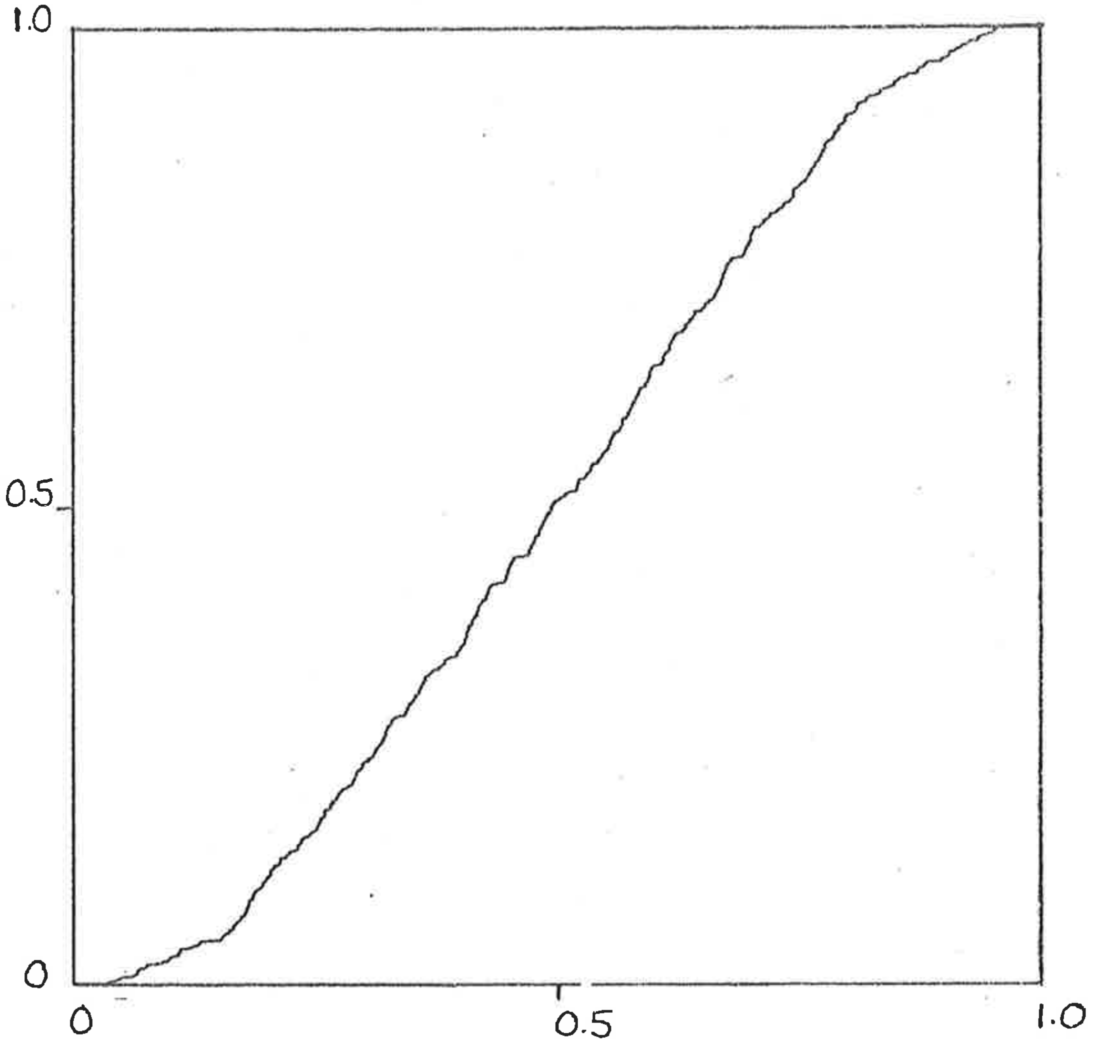
S=2
ℓ=1



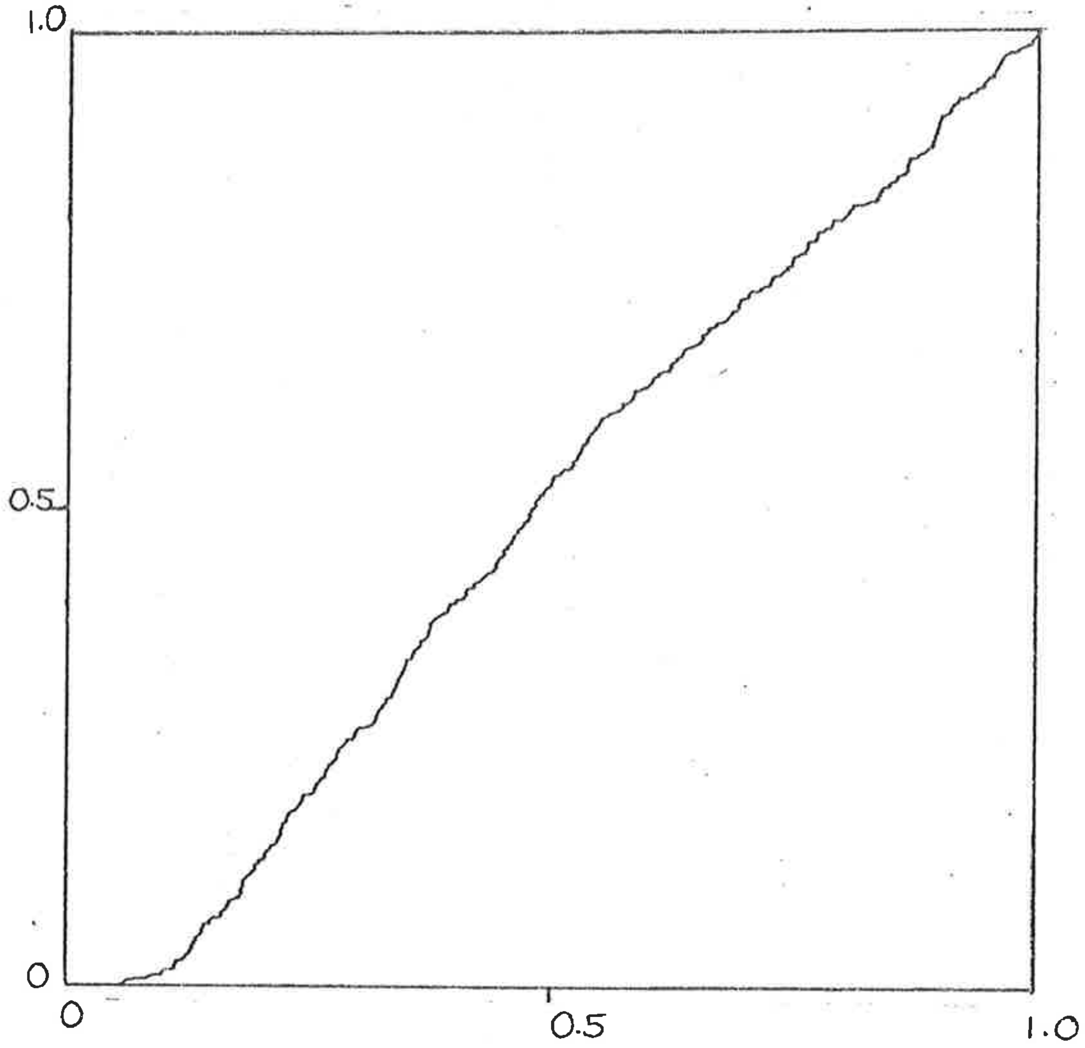
s=2
ℓ=2



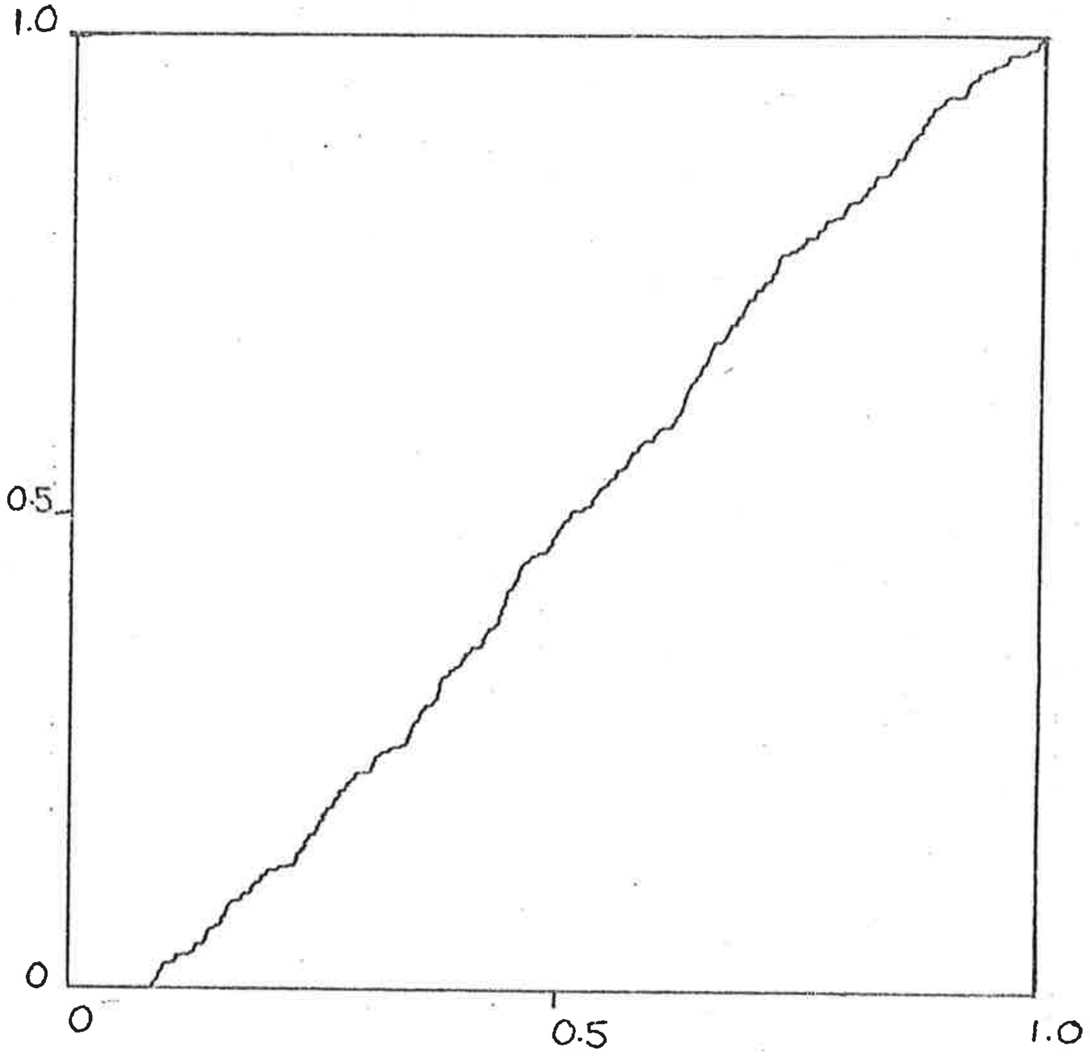
$s=2$
$l=3$



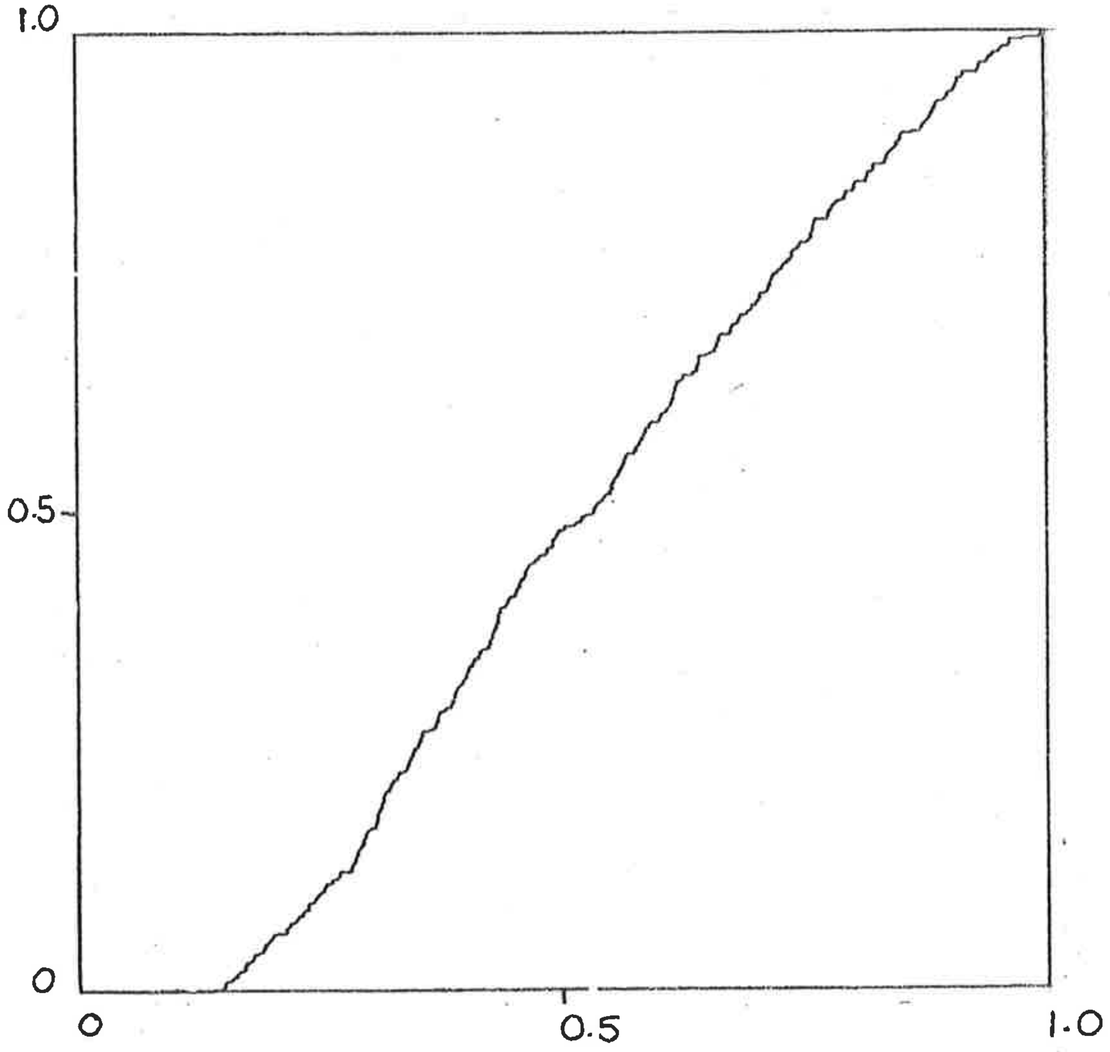
s = 2
$l = 4$



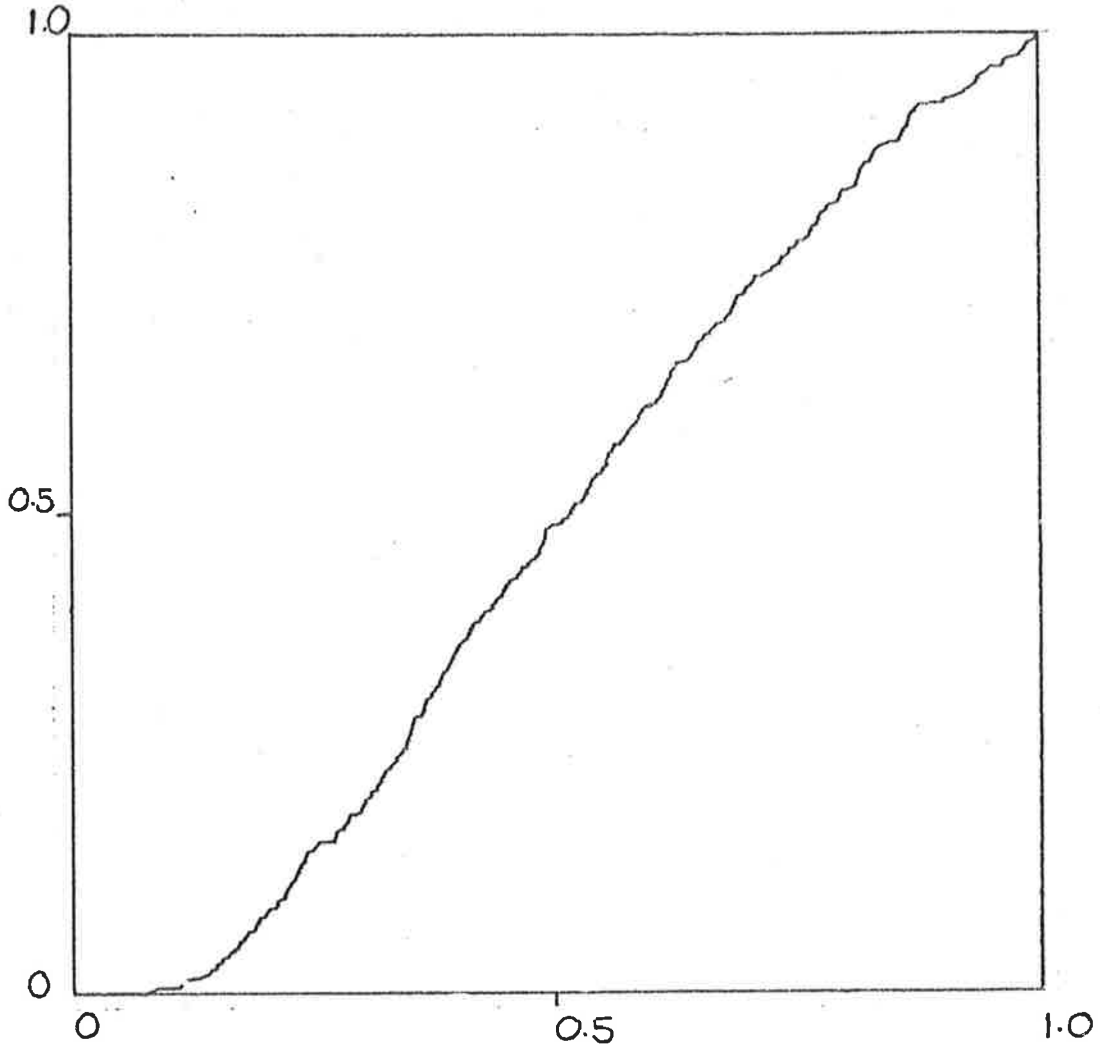
S = 10
l = 1



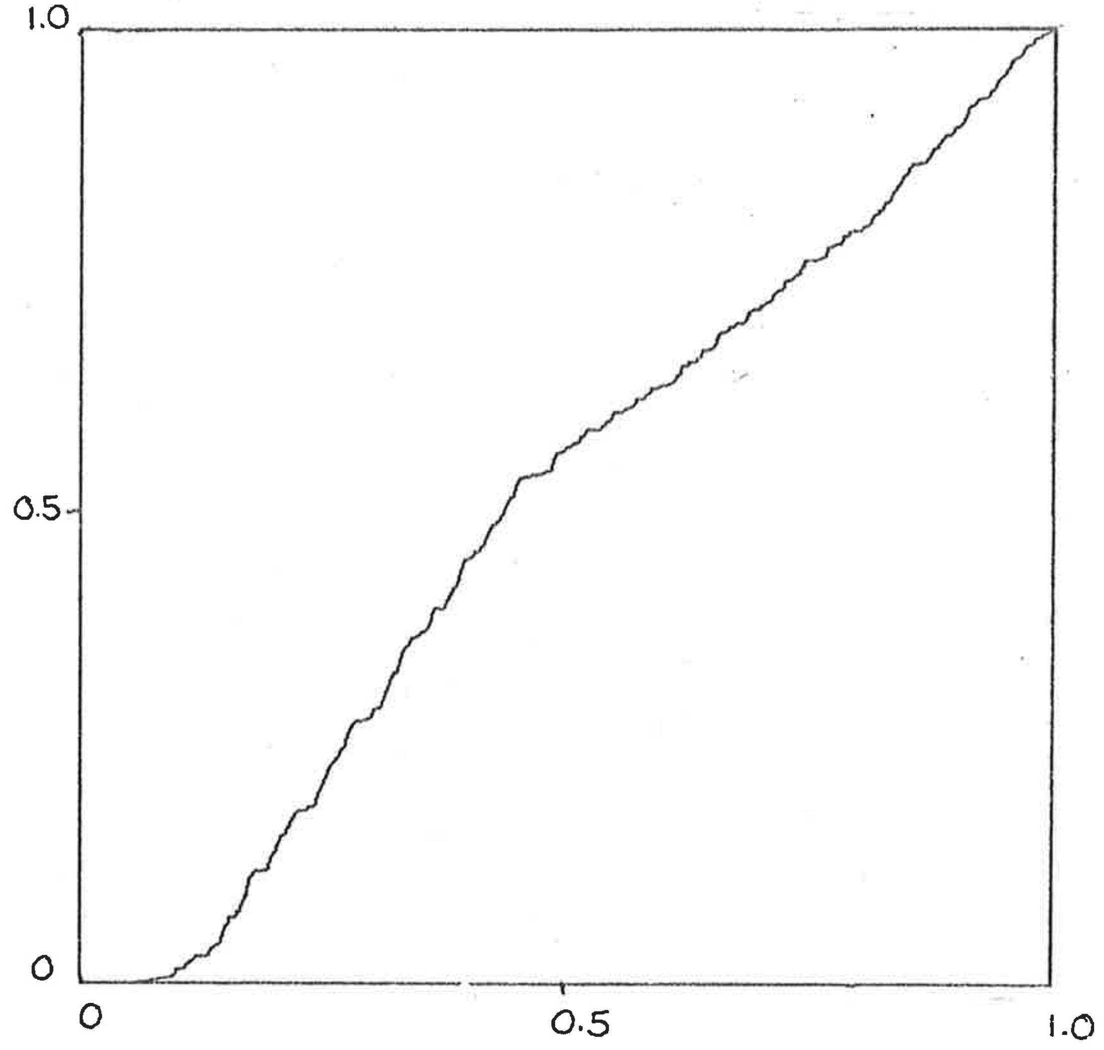
S = 10
$l = 2$



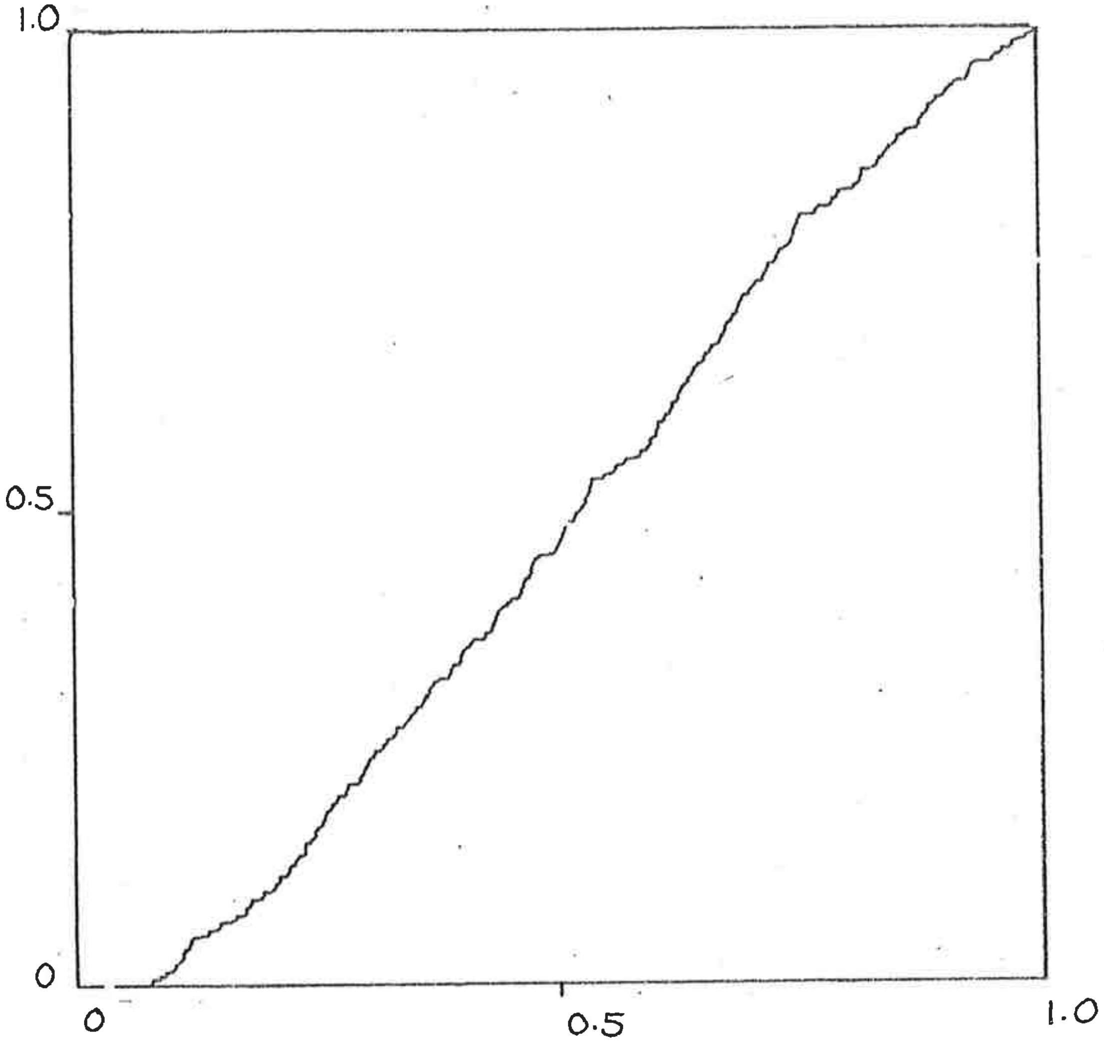
S= 10
ℓ= 3



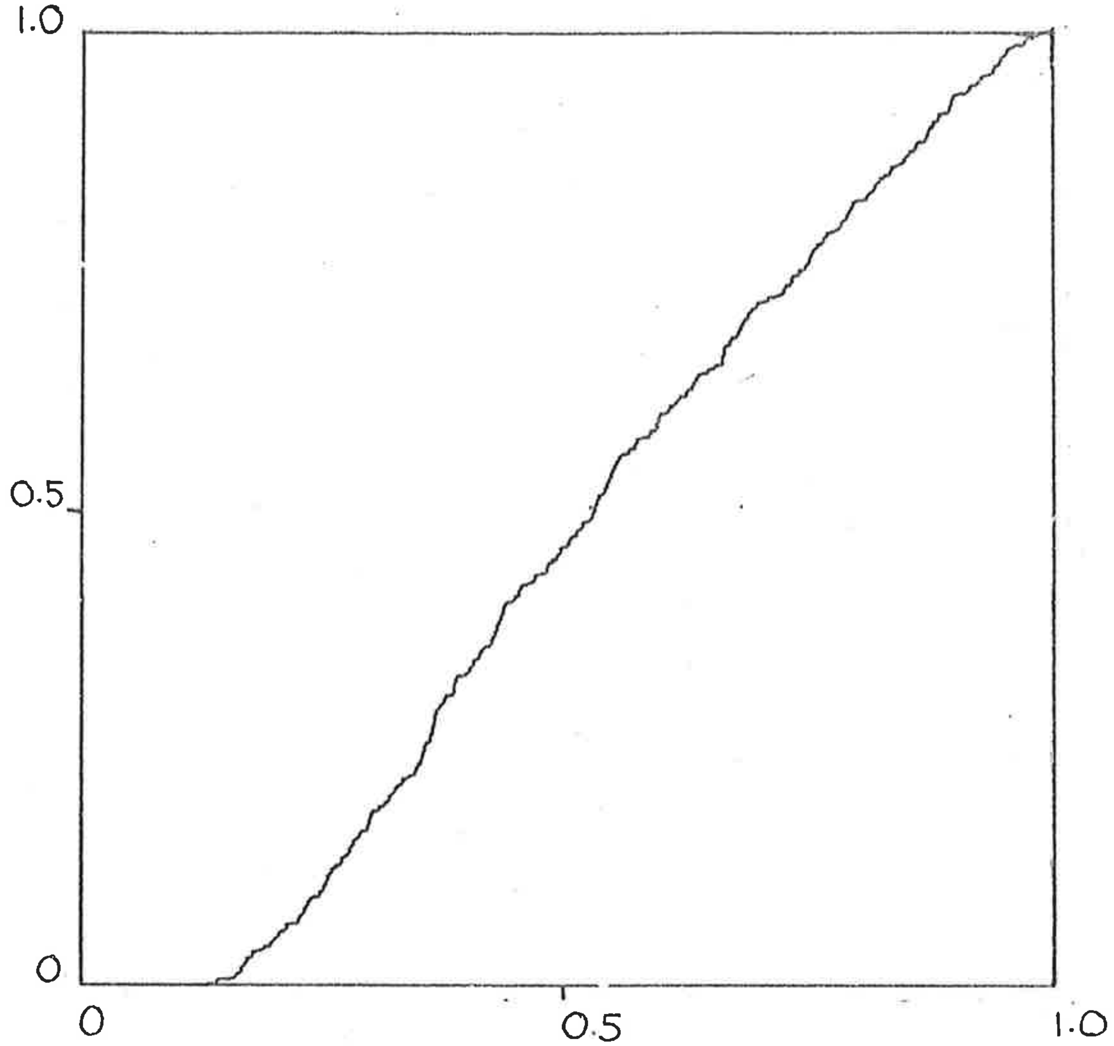
S = 10
$\ell = 4$



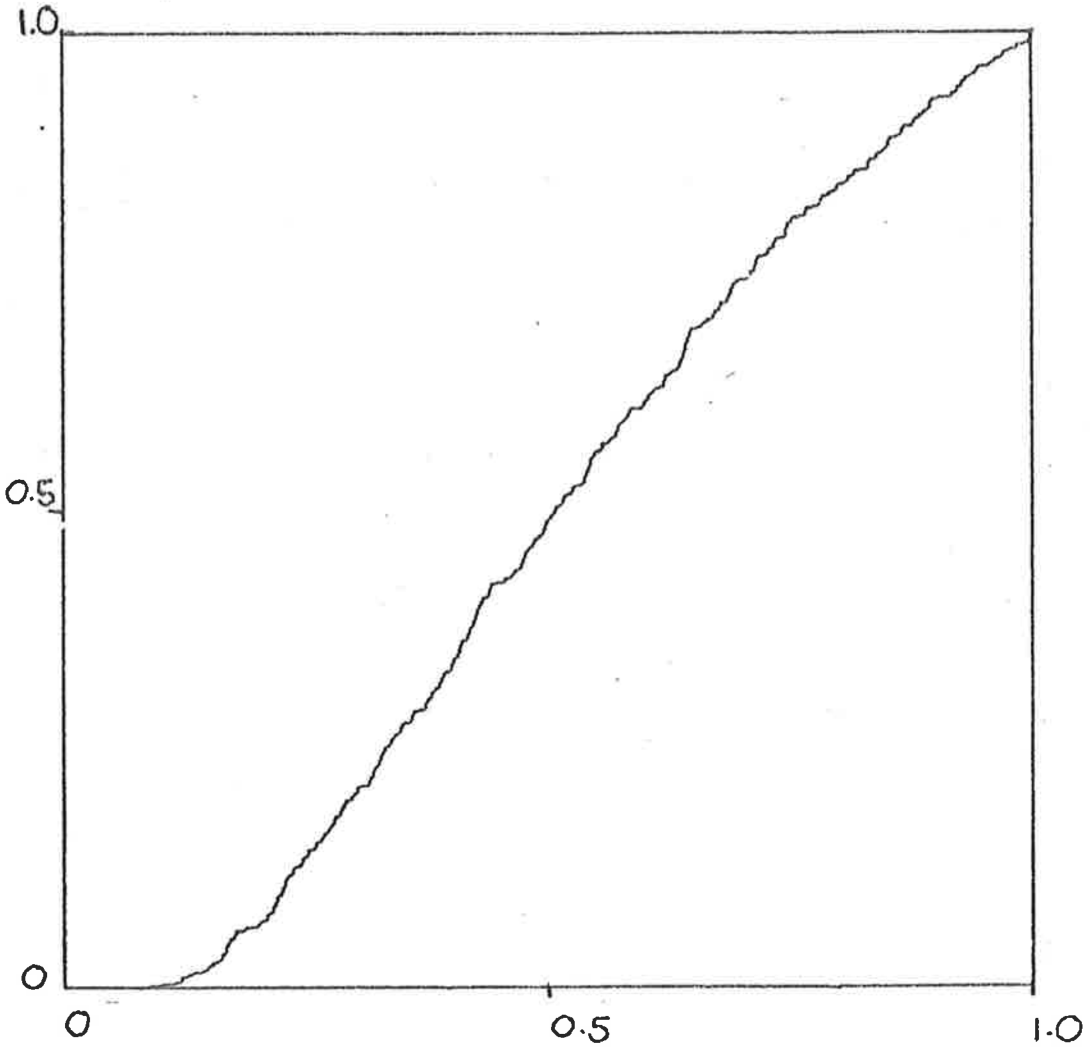
S=20
$\ell=1$



S = 20
$\ell = 2$

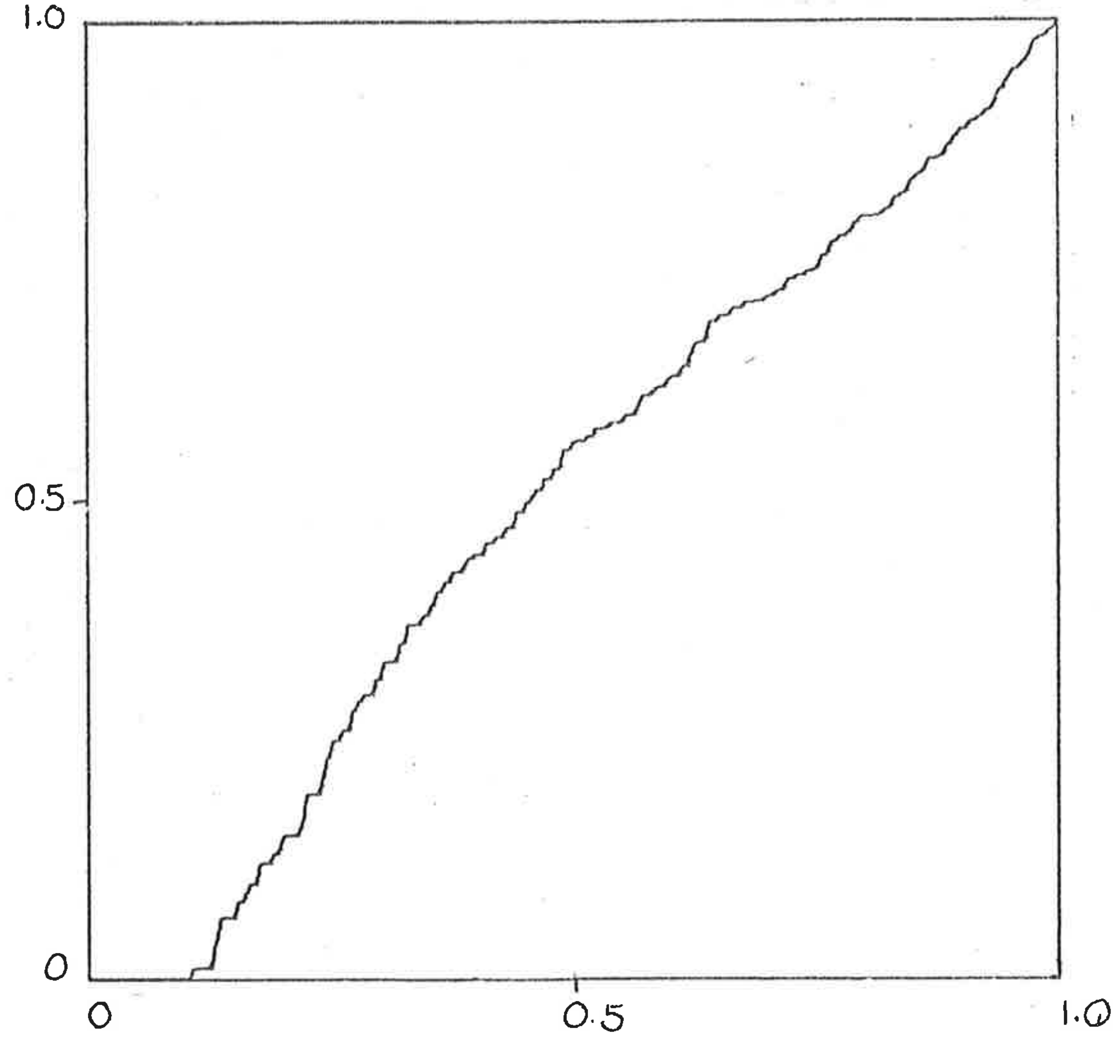


$s = 20$
$l = 3$

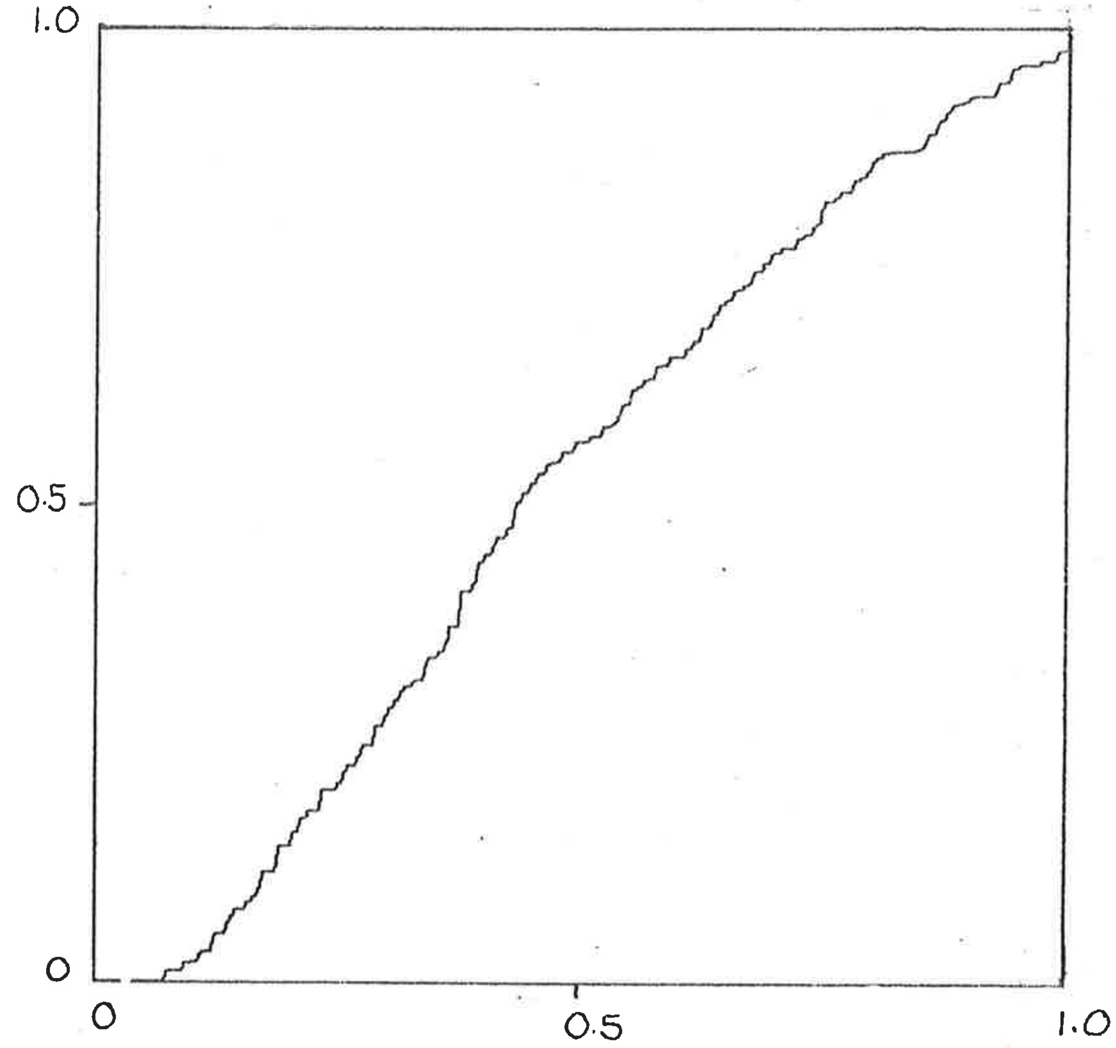


S = 20
ℓ = 4

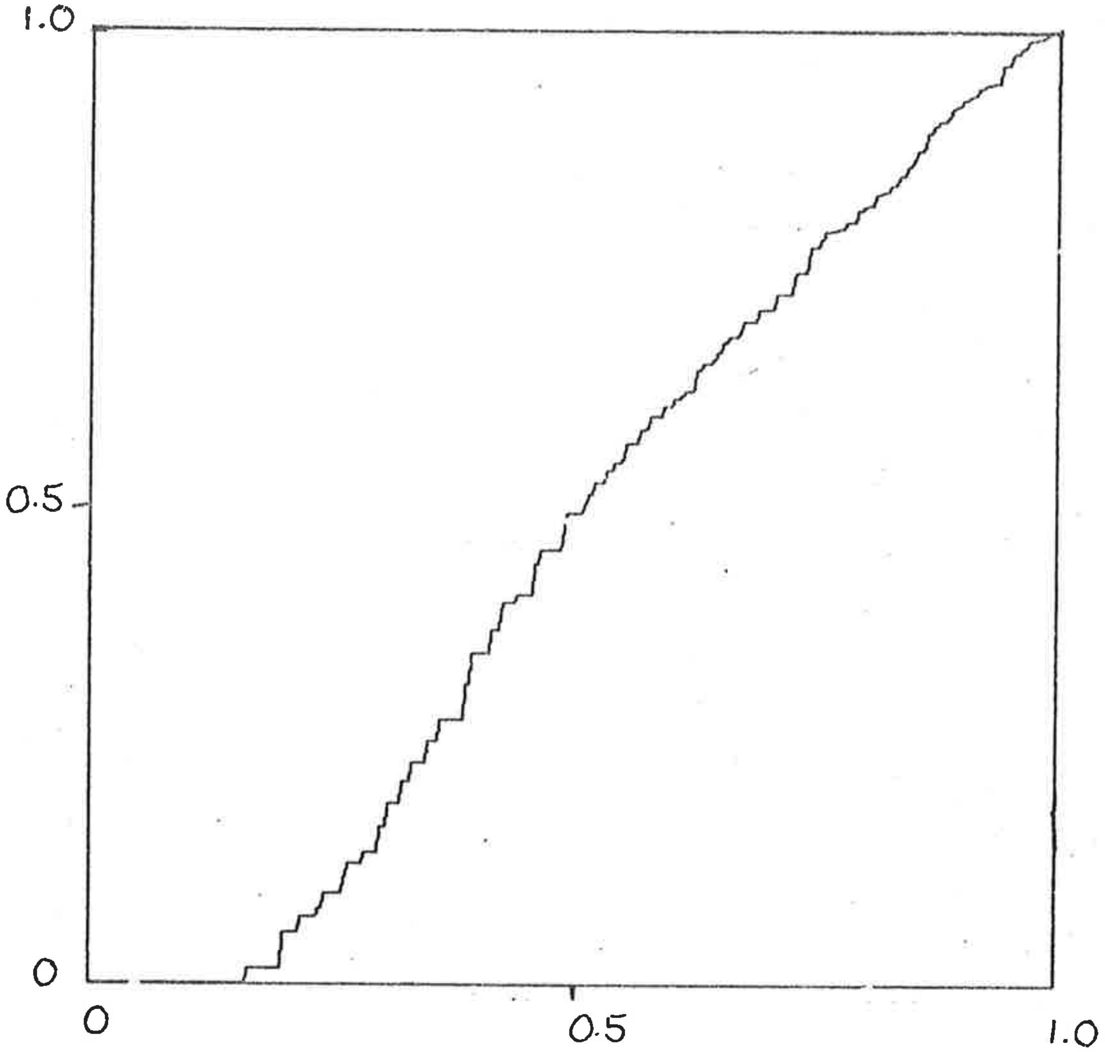
Sample distribution functions for $\{1-\exp(-T_1{**})\}$ for each animal.



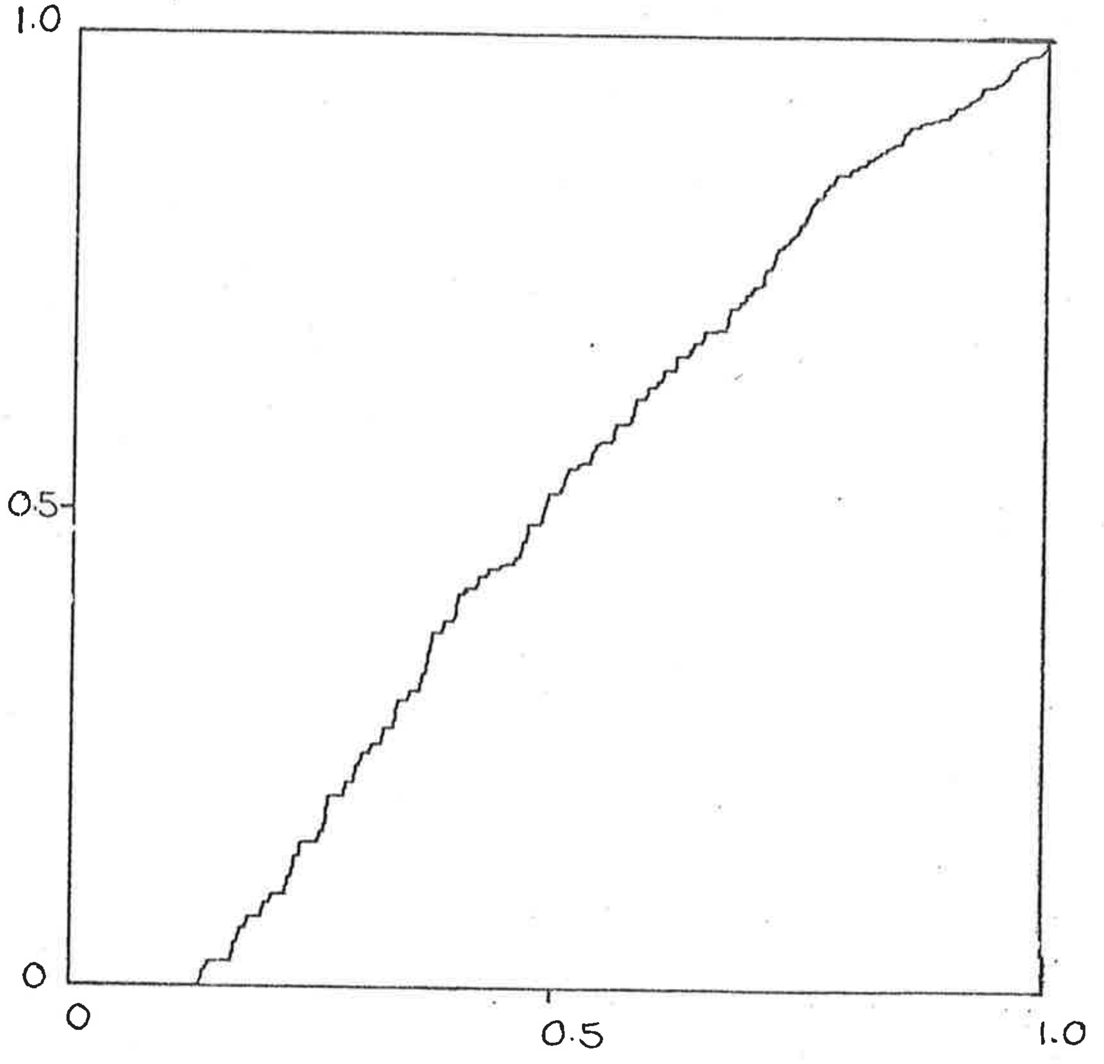
$l=1$



$l = 2$

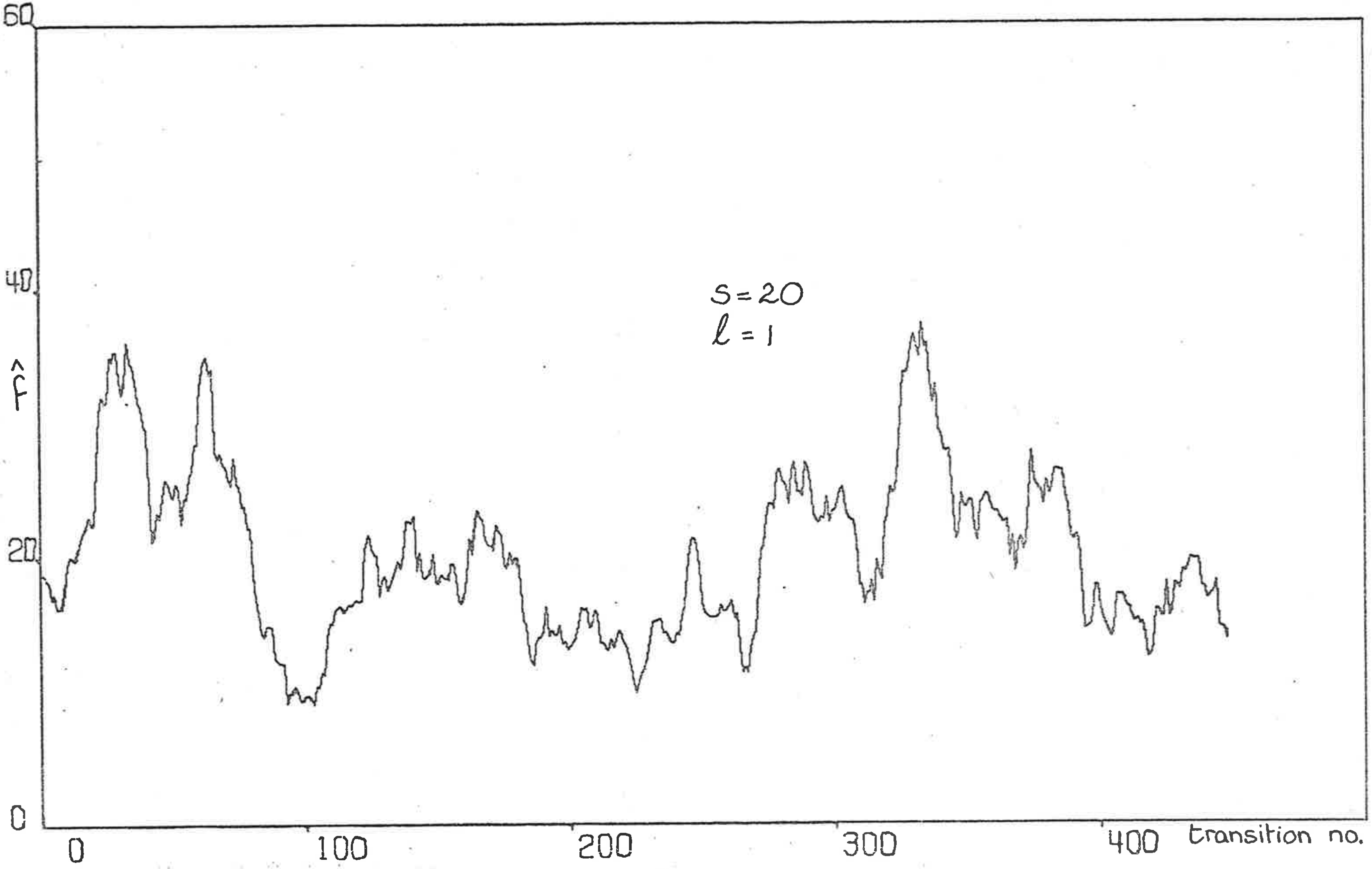


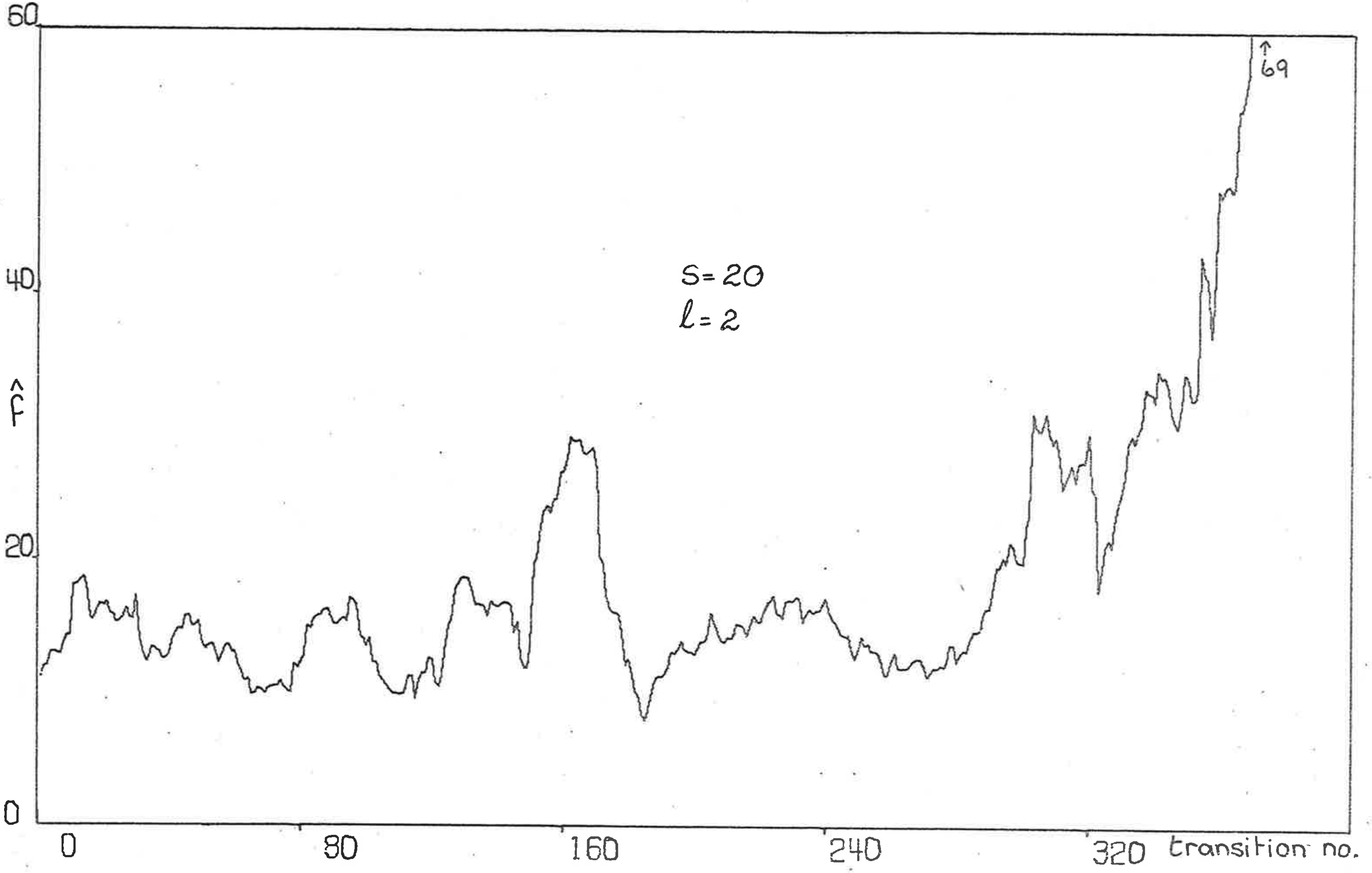
$l = 3$

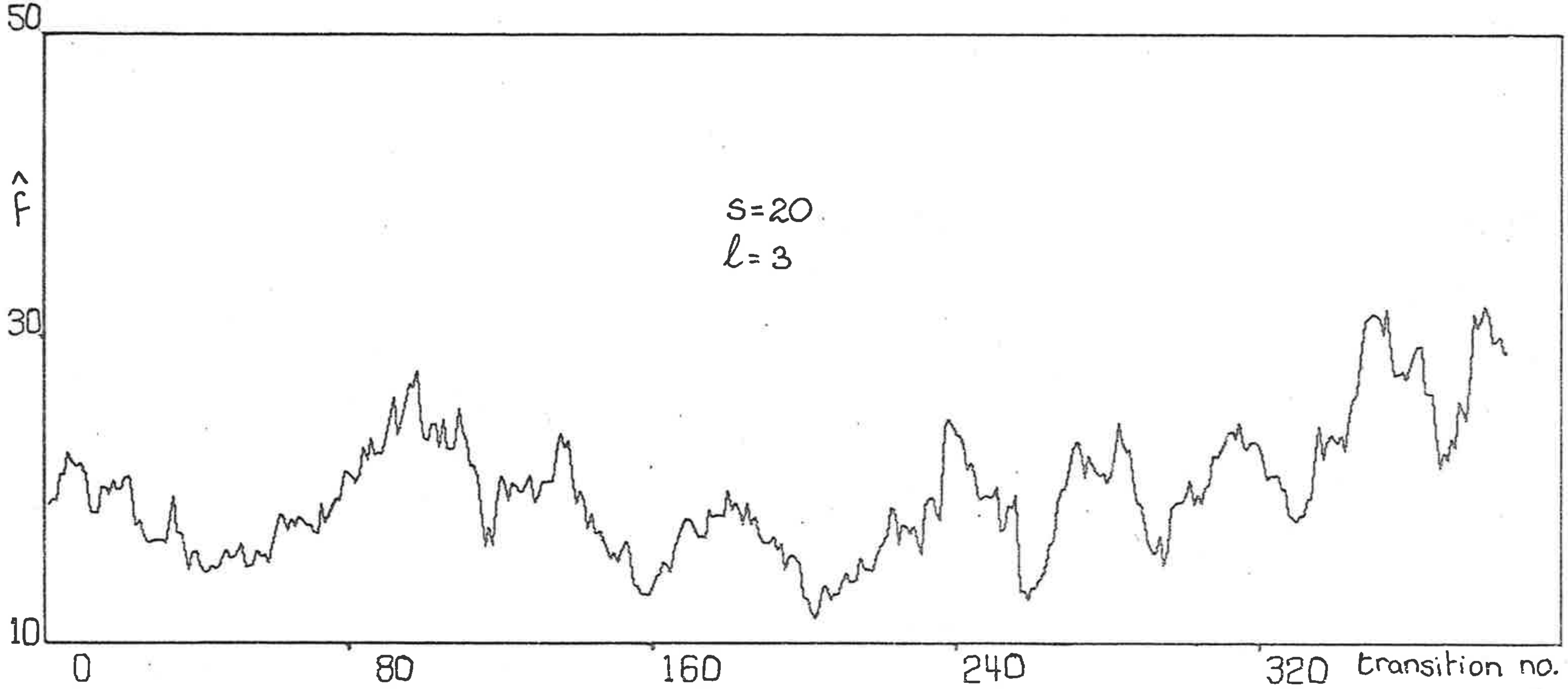


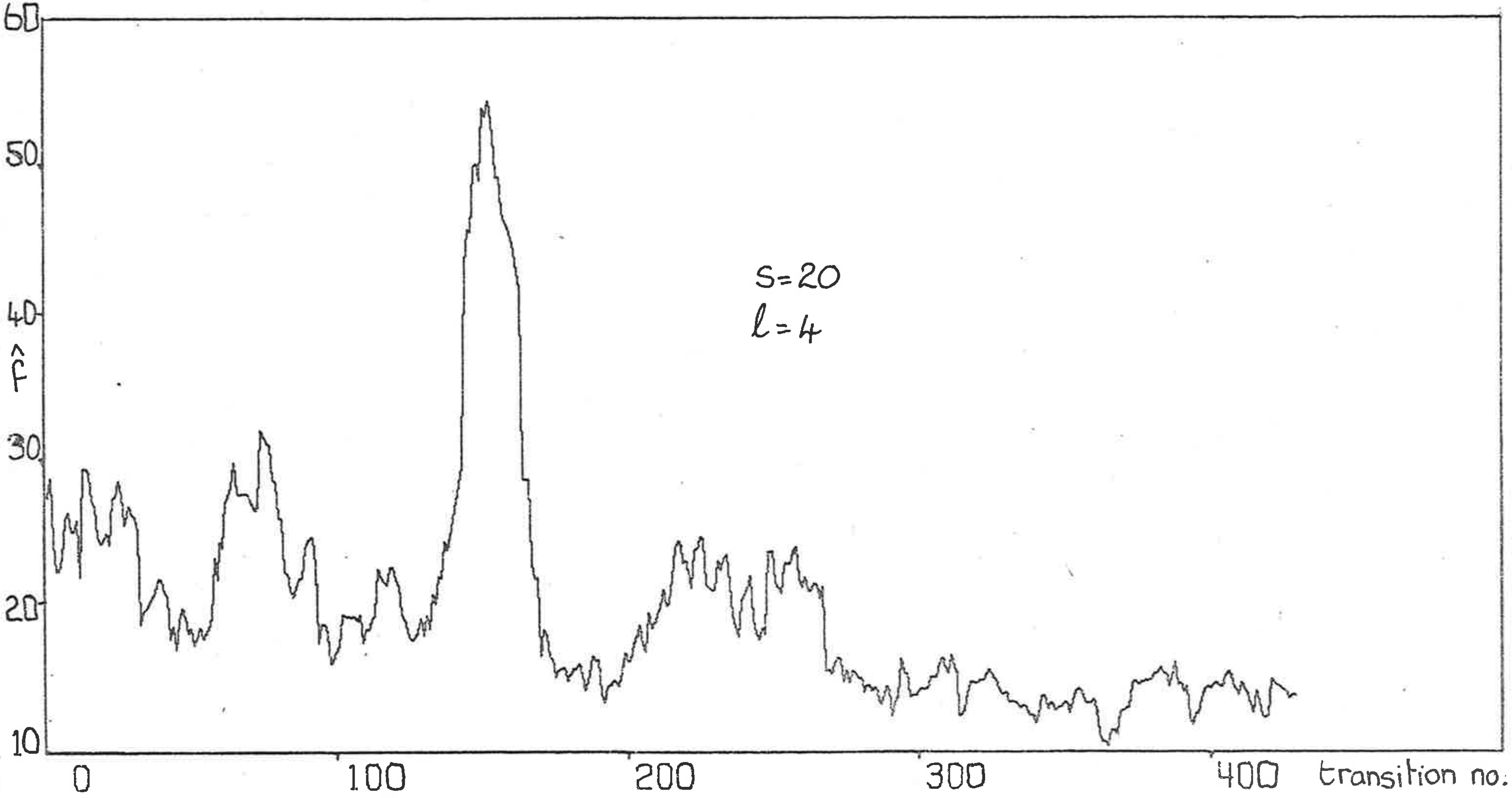
$l = 4$

The functions $\hat{f}$ for each animal.


60
40
20
0
F̂
s = 20
ℓ = 1
0
100
200
300
400
transition no.

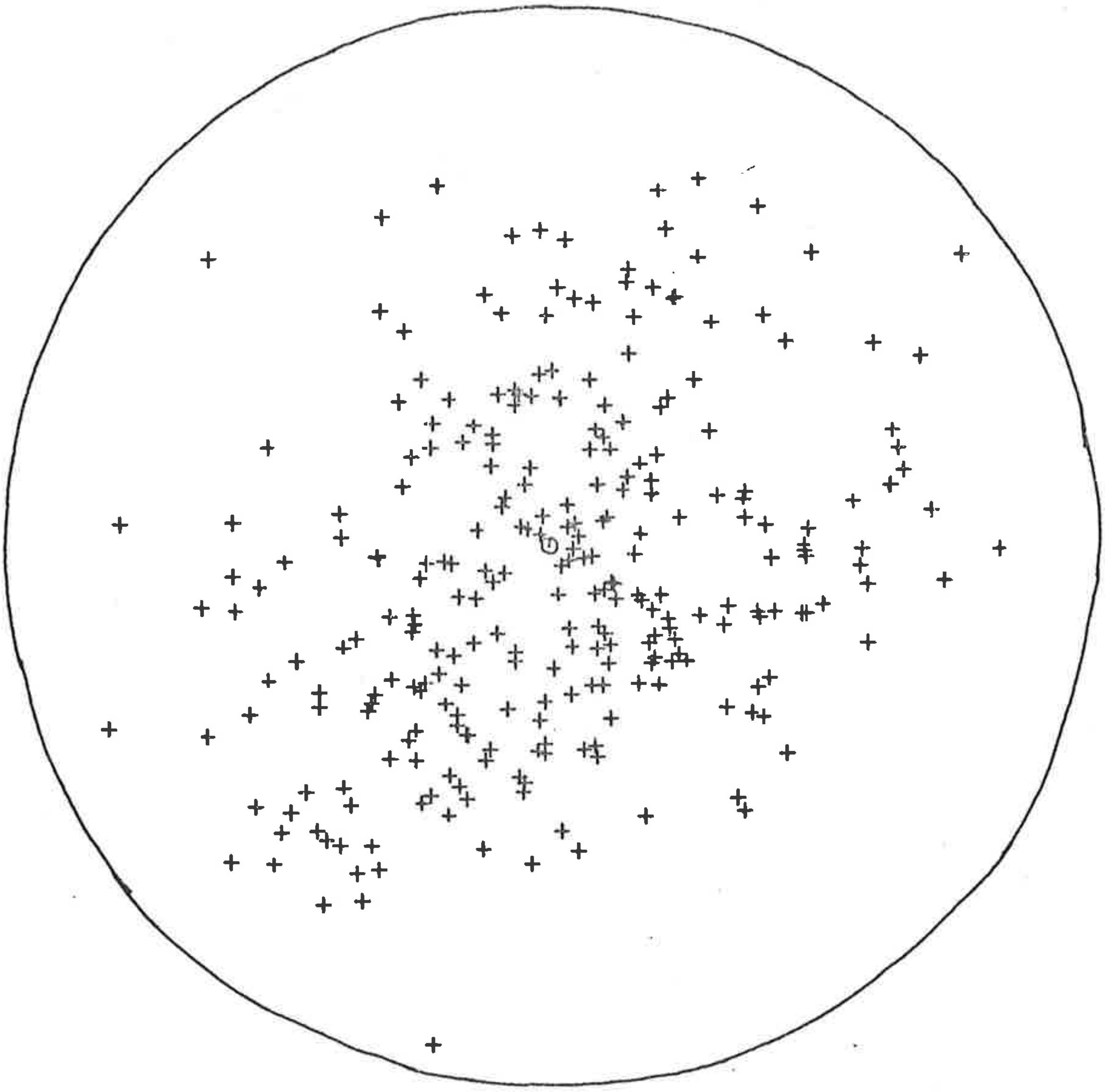

60
40
20
0
F̂
S=20
l=2
↑ 69
0
80
160
240
320
transition no.


50
30
10
$\hat{F}$
s=20
ℓ=3
0
80
160
240
320
transition no.

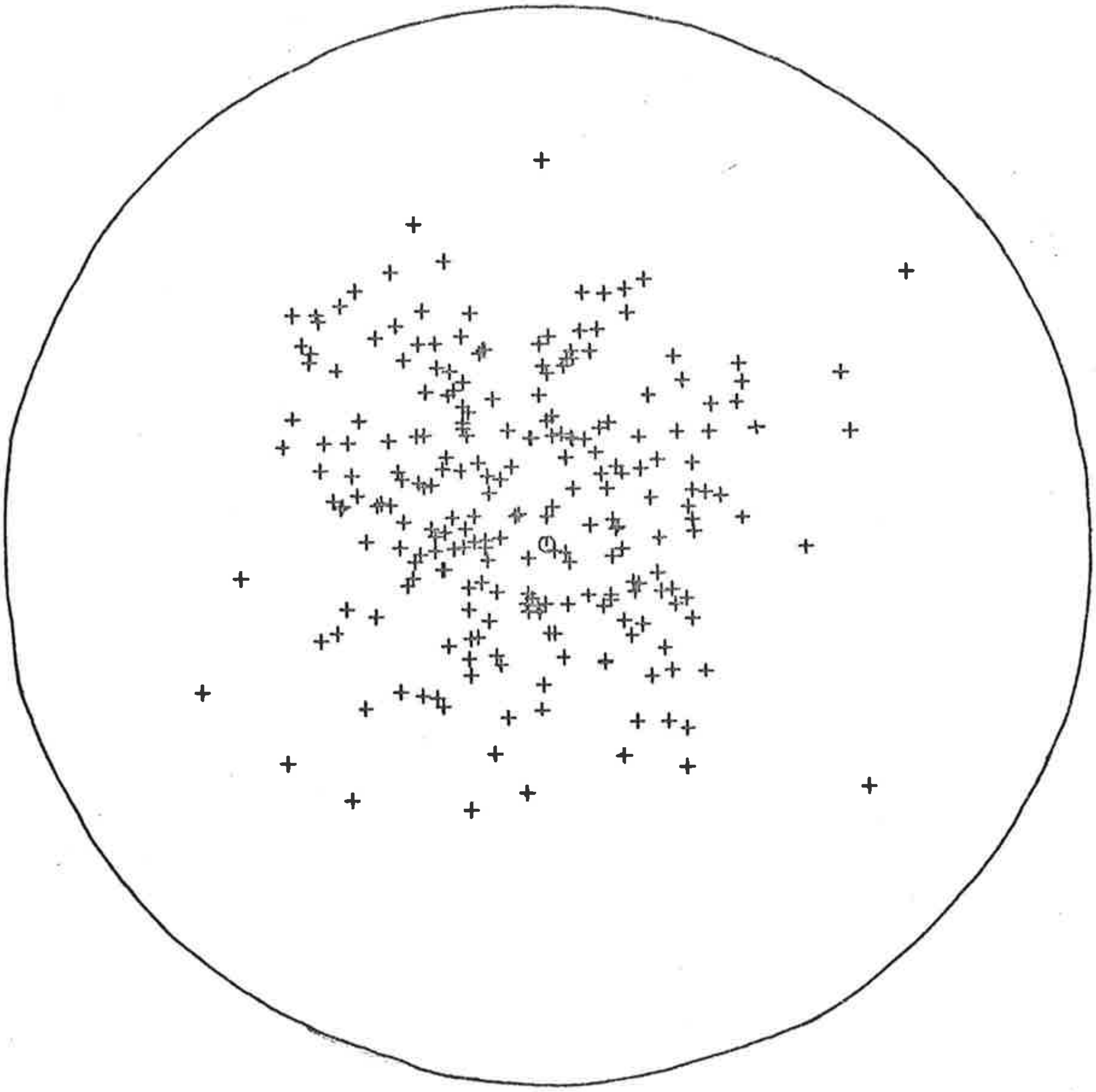

60
50
40
30
20
10
$\hat{f}$
0
100
200
300
400
transition no.
s=20
l=4

APPENDIX 3 : CONTINUOUS SPACE EXPERIMENTS

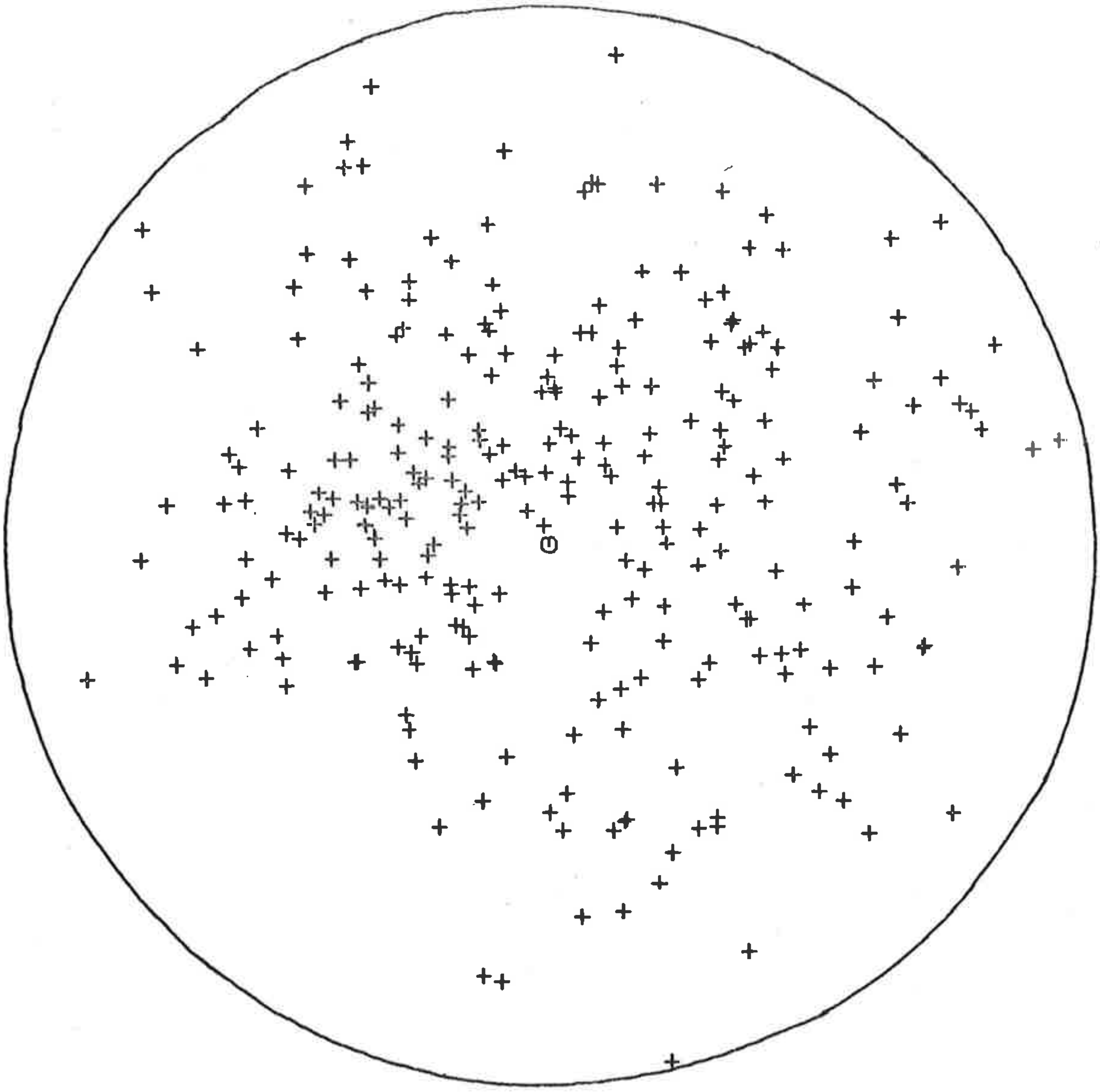
Scatter of the positions of each animal.



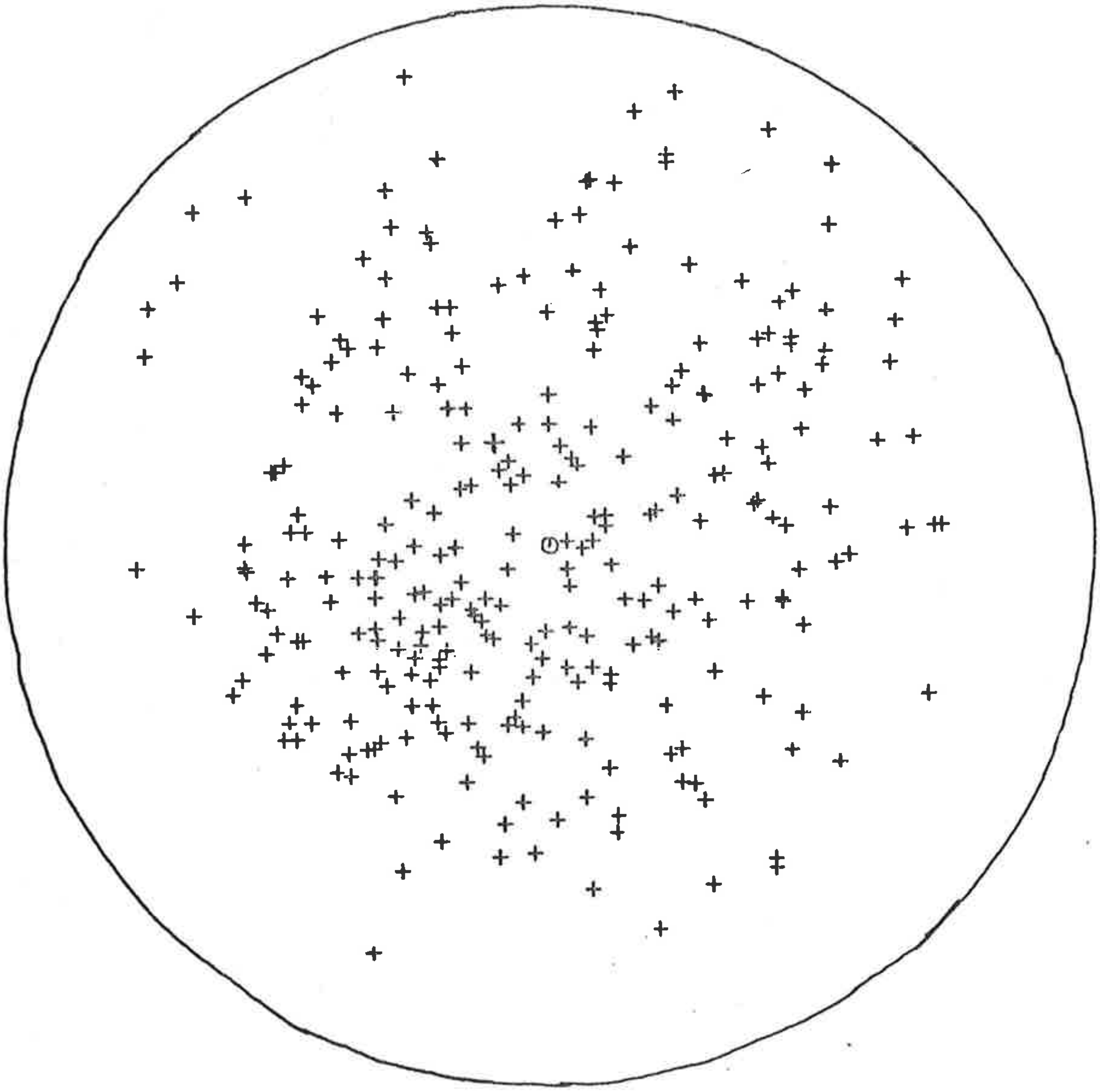
$l=1$



l=2



l=3



l=4

The next eight pages give the matrices A, $\hat{\Sigma}^*(1)$, $\hat{\Sigma}^*(2)$ and $\hat{\Sigma}^*(3)$ for each animal.

animal 1

$$
A \quad \begin{array}{rrrrrrrr}
.243 & -.104 & -.567 & -.168 & -.293 & -.229 & .016 & .007 \\
.006 & .148 & .150 & .111 & -.071 & -.354 & -.555 & .209 \\
.471 & .793 & .105 & .088 & .091 & -.119 & .451 & .081 \\
.566 & -.584 & .287 & .378 & -.519 & .327 & -.043 & .177 \\
-.194 & .236 & -.253 & .253 & -.917 & .424 & .124 & -.143 \\
.037 & .162 & .365 & -.714 & -.260 & -.042 & -.039 & -.607 \\
.005 & .103 & .029 & .496 & .220 & -.358 & -.029 & -.913 \\
-.075 & -.129 & .202 & -.038 & -.221 & -.643 & .662 & .268
\end{array}
$$

$$
\hat{\Sigma}^*(1) \quad \begin{array}{rrrrrrrr}
.832 & .036 & .017 & .015 & .018 & .013 & .001 & -.016 \\
-.036 & .714 & -.085 & -.036 & .019 & .018 & .002 & .010 \\
-.017 & .085 & .643 & .058 & -.051 & .020 & .024 & -.097 \\
-.015 & .036 & -.058 & .567 & .146 & .022 & .004 & .027 \\
-.018 & -.019 & .051 & -.146 & .472 & -.021 & -.053 & -.080 \\
-.013 & -.018 & -.020 & -.022 & -.021 & .385 & .059 & .174 \\
-.001 & -.002 & -.024 & -.004 & .053 & -.059 & .302 & -.190 \\
.016 & -.010 & .010 & -.027 & .080 & -.174 & .190 & .162
\end{array}
$$

animal 1 (continued)

$\hat{\Sigma}^*(2)$	.694	.063	.002	.014	.003	-.027	.076	-.048
	-.082	.530	-.067	-.026	-.010	-.011	-.040	.115
	-.081	.151	.443	.047	-.042	.072	-.114	-.052
	.013	.115	-.042	.265	.201	.083	-.042	.084
	-.067	-.088	.056	-.222	.138	-.003	-.059	.001
	.035	-.043	.046	.061	.067	.185	.212	.139
	.026	.060	.059	.044	.029	-.024	.020	-.117
	-.031	.032	.051	-.116	.112	-.141	-.036	-.087

$\hat{\Sigma}^*(3)$	.560	.079	-.048	.018	.030	-.016	.102	-.027
	-.063	.342	-.060	.049	.027	-.068	-.031	.082
	-.081	.177	.357	.015	-.015	.035	-.082	.021
	-.013	.134	-.052	.080	.253	-.011	-.022	.185
	-.121	-.121	.090	-.133	.027	-.022	-.048	.012
	.064	.033	-.011	.031	.163	-.085	.127	.051
	-.019	.045	-.015	.123	.045	.066	-.022	-.088
	-.070	.065	.081	-.073	.072	-.121	-.067	-.056

animal 2

A	.154	-.228	.077	.555	-.249	.267	-.108	-.118
	-.004	.108	-.184	-.232	-.008	.599	.212	-.146
	-.389	.723	-.106	.706	-.096	.032	.065	.097
	.795	.576	-.129	-.221	-.415	-.115	-.029	.114
	.097	.159	.100	.228	.875	.247	-.337	.110
	.184	.000	-.860	.249	.084	-.022	.446	.689
	.015	.005	-.338	-.006	.095	-.212	-.612	-.825
	-.039	-.016	-.240	-.125	-.170	.262	-.696	.697

$\hat{\Sigma}^*(1)$	.642	-.008	.047	.036	-.016	.005	-.031	-.050
	.008	.572	.013	-.029	.024	-.045	-.026	-.012
	-.047	-.013	.404	-.026	-.014	-.004	-.029	.004
	-.036	.029	.026	.282	-.036	-.061	-.005	-.025
	.016	-.024	.014	.036	.252	.006	.017	.034
	-.005	.045	.004	.061	-.006	.103	-.055	.039
	.031	.026	.029	.005	-.017	.055	-.031	.066
	.050	.012	-.004	.025	-.034	-.039	-.066	.007

animal 2 (continued)

$$\hat{\Sigma}^*(2) \quad \begin{matrix}
.482 & -.069 & .017 & .088 & .077 & -.038 & -.043 & -.113 \\
-.040 & .422 & -.027 & -.104 & -.042 & -.054 & -.014 & -.050 \\
-.022 & .002 & .199 & .070 & .110 & .051 & .044 & -.036 \\
-.036 & .081 & -.002 & .114 & -.074 & -.109 & .028 & -.139 \\
-.003 & -.062 & -.133 & .020 & .232 & .075 & -.072 & .127 \\
-.024 & .086 & .000 & .079 & .055 & .023 & -.032 & -.063 \\
.053 & .151 & -.073 & .106 & -.189 & -.034 & .046 & -.058 \\
.028 & -.006 & .122 & .078 & -.071 & .059 & .013 & -.060
\end{matrix}$$

$$\hat{\Sigma}^*(3) \quad \begin{matrix}
.360 & -.073 & .114 & .037 & .057 & -.008 & .060 & -.089 \\
-.003 & .386 & -.021 & -.136 & .032 & .017 & -.011 & -.022 \\
-.007 & .003 & .050 & .051 & .126 & .019 & .040 & .149 \\
.022 & .127 & .050 & -.003 & -.154 & .045 & .022 & -.001 \\
.047 & .022 & -.111 & .013 & .079 & -.060 & -.041 & .029 \\
-.002 & .010 & -.029 & .114 & .103 & -.039 & .041 & .007 \\
-.008 & .095 & -.033 & .063 & -.048 & .092 & -.017 & -.097 \\
.030 & -.079 & .076 & -.060 & .009 & -.011 & -.023 & -.108
\end{matrix}$$

animal 3

$$\mathbf{A} \quad \begin{matrix} .030 & .188 & -.229 & .623 & .078 & .113 & -.167 & .015 \\ -.069 & .023 & -.053 & -.152 & .044 & .051 & -.342 & .608 \\ -.121 & .987 & -.074 & -.177 & -.146 & -.422 & -.131 & -.136 \\ -.883 & -.117 & -.174 & .229 & -.105 & -.270 & -.080 & -.244 \\ -.136 & .149 & .663 & .069 & .429 & .386 & .523 & .358 \\ .015 & .050 & .527 & .301 & -.744 & -.204 & -.242 & .071 \\ -.140 & -.081 & -.350 & -.179 & -.067 & .719 & -.506 & -.444 \\ -.032 & .022 & -.222 & -.085 & -.457 & .562 & .670 & .333 \end{matrix}$$

$$\hat{\Sigma}^*(1) \quad \begin{matrix} .798 & -.072 & -.005 & .031 & -.042 & .027 & -.008 & .062 \\ .072 & .768 & -.019 & .037 & .054 & .029 & -.014 & .008 \\ .005 & .019 & .582 & -.042 & .102 & .017 & .005 & .023 \\ -.031 & -.037 & .042 & .477 & .008 & .022 & .041 & .064 \\ .042 & -.054 & -.102 & -.008 & .439 & -.053 & .052 & -.048 \\ -.027 & -.029 & -.017 & -.022 & .053 & .250 & -.051 & -.065 \\ .008 & .014 & -.005 & -.041 & -.052 & .051 & .226 & .011 \\ -.062 & -.008 & -.023 & -.064 & .048 & .065 & -.011 & .084 \end{matrix}$$

animal 3 (continued)

$\hat{\Sigma}^*(2)$	.619	-.170	-.053	.079	-.086	-.038	-.064	.106
	.099	.572	.019	.011	.052	.056	.046	.053
	-.025	.065	.385	.055	.010	-.050	-.101	-.012
	-.023	.005	.022	.238	.030	.041	.081	-.000
	.078	-.098	-.146	.065	.167	-.144	.073	-.119
	-.036	.017	.052	-.001	-.050	.070	-.019	-.006
	-.014	-.004	-.034	-.052	-.047	-.016	-.052	.014
	-.084	.030	.005	-.022	-.004	-.124	.063	-.040

$\hat{\Sigma}^*(3)$	.504	-.193	.018	.119	-.068	-.045	-.104	.045
	.147	.456	.006	-.004	.027	.020	.073	.071
	-.032	.089	.233	.008	-.057	-.043	-.020	-.070
	-.036	.077	.068	.067	-.040	.055	.010	.035
	.011	-.064	-.131	.062	.070	-.112	-.001	-.070
	-.038	-.013	.108	.032	-.041	.087	.012	-.057
	-.008	.018	.082	-.045	-.019	.082	.012	.022
	-.034	.093	.030	-.075	.029	.008	.006	.068

animal 4

$$A \quad \begin{matrix} .030 & .188 & -.229 & .623 & .078 & .113 & -.167 & .015 \\ -.069 & .023 & -.053 & -.152 & .044 & .051 & -.342 & .608 \\ -.121 & .987 & -.074 & -.177 & -.146 & -.422 & -.131 & -.136 \\ -.883 & -.117 & -.174 & .229 & -.105 & -.270 & -.030 & -.244 \\ -.136 & .149 & .663 & .069 & .429 & .386 & .523 & .358 \\ .015 & .050 & .527 & .301 & -.744 & -.204 & -.242 & -.071 \\ -.140 & .081 & .035 & -.179 & -.067 & .719 & -.506 & -.444 \\ -.032 & .022 & -.222 & -.085 & -.457 & .562 & .670 & .333 \end{matrix}$$

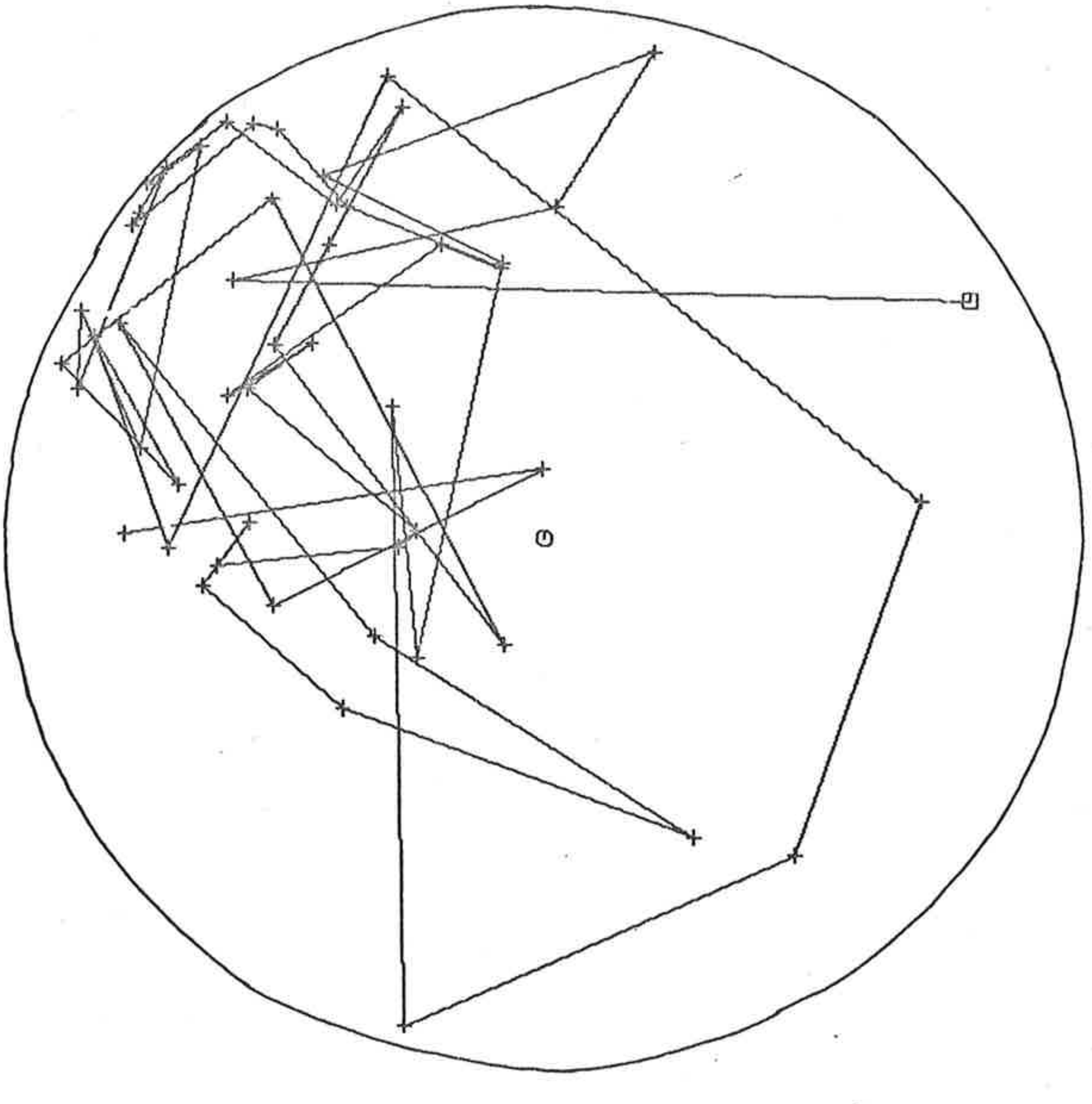
$$\hat{\Sigma}^*(1) \quad \begin{matrix} .798 & -.072 & -.005 & .031 & -.042 & .027 & -.008 & .062 \\ .072 & .769 & -.019 & .037 & .054 & .029 & -.014 & .008 \\ .005 & .019 & .582 & -.042 & .102 & .017 & .005 & .023 \\ -.031 & -.037 & .042 & .477 & .008 & .022 & .041 & .064 \\ .042 & -.054 & -.102 & -.008 & .439 & -.053 & .052 & -.048 \\ -.027 & -.029 & -.017 & -.022 & .053 & .250 & -.051 & -.065 \\ .008 & .014 & -.005 & -.041 & -.052 & .051 & .226 & .011 \\ -.062 & -.008 & -.023 & -.064 & .048 & .065 & -.011 & .084 \end{matrix}$$

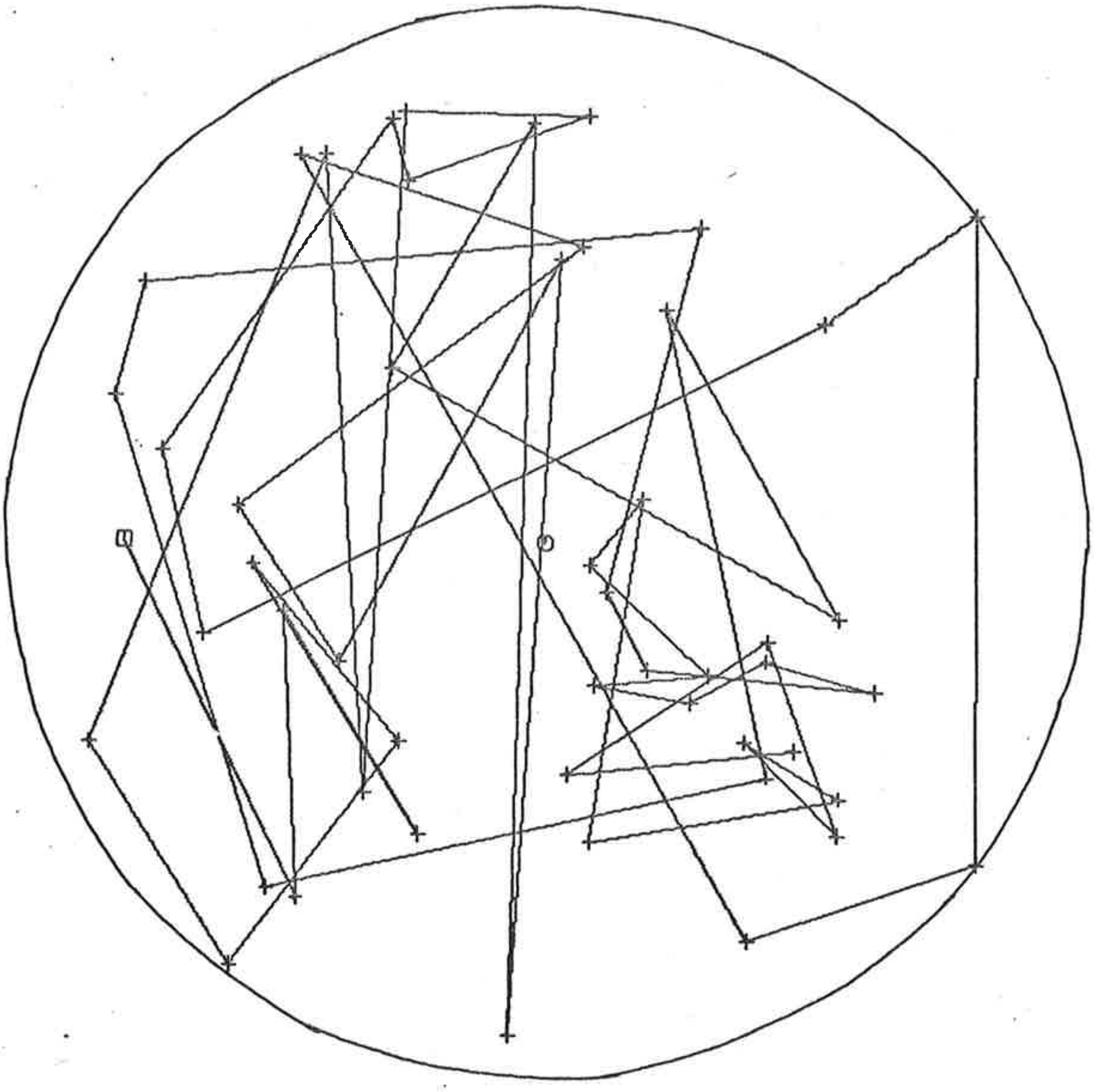
animal 4 (continued)

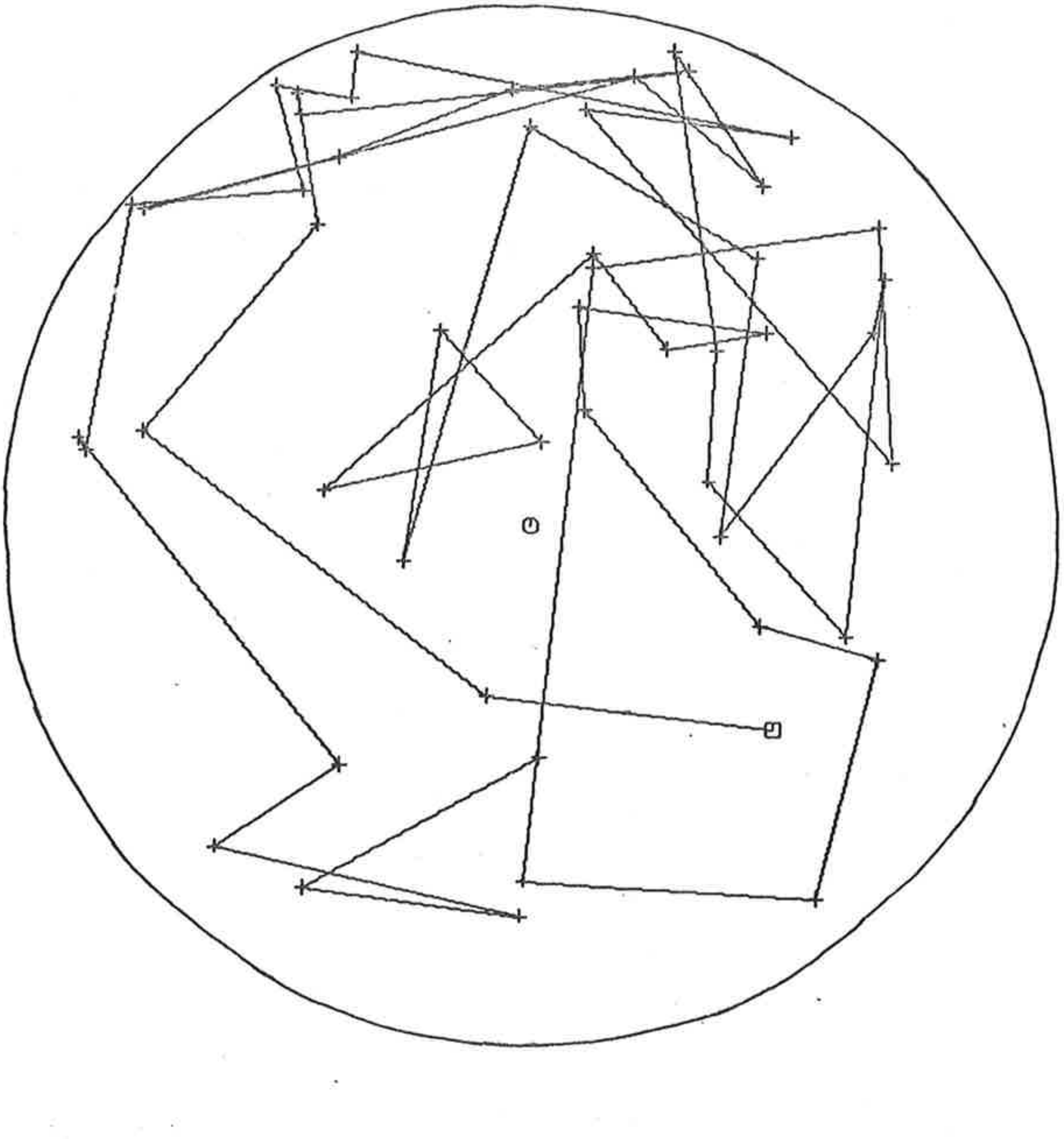
$$\hat{\Sigma}^*(2) \quad \begin{array}{rrrrrrrr}
.619 & -.170 & -.053 & .079 & -.086 & -.038 & -.064 & .106 \\
.099 & .572 & .019 & .011 & .052 & .056 & .046 & .053 \\
-.025 & .065 & .385 & .055 & .010 & -.050 & -.101 & -.012 \\
-.023 & .005 & .022 & .238 & .030 & .041 & .081 & -.000 \\
.028 & -.098 & -.146 & .065 & .167 & -.144 & .073 & -.119 \\
-.036 & .017 & .052 & -.001 & -.050 & .070 & -.019 & -.006 \\
-.014 & -.004 & -.034 & -.052 & -.047 & -.016 & -.052 & .014 \\
-.084 & .030 & .005 & -.022 & -.004 & -.124 & .063 & -.040
\end{array}$$

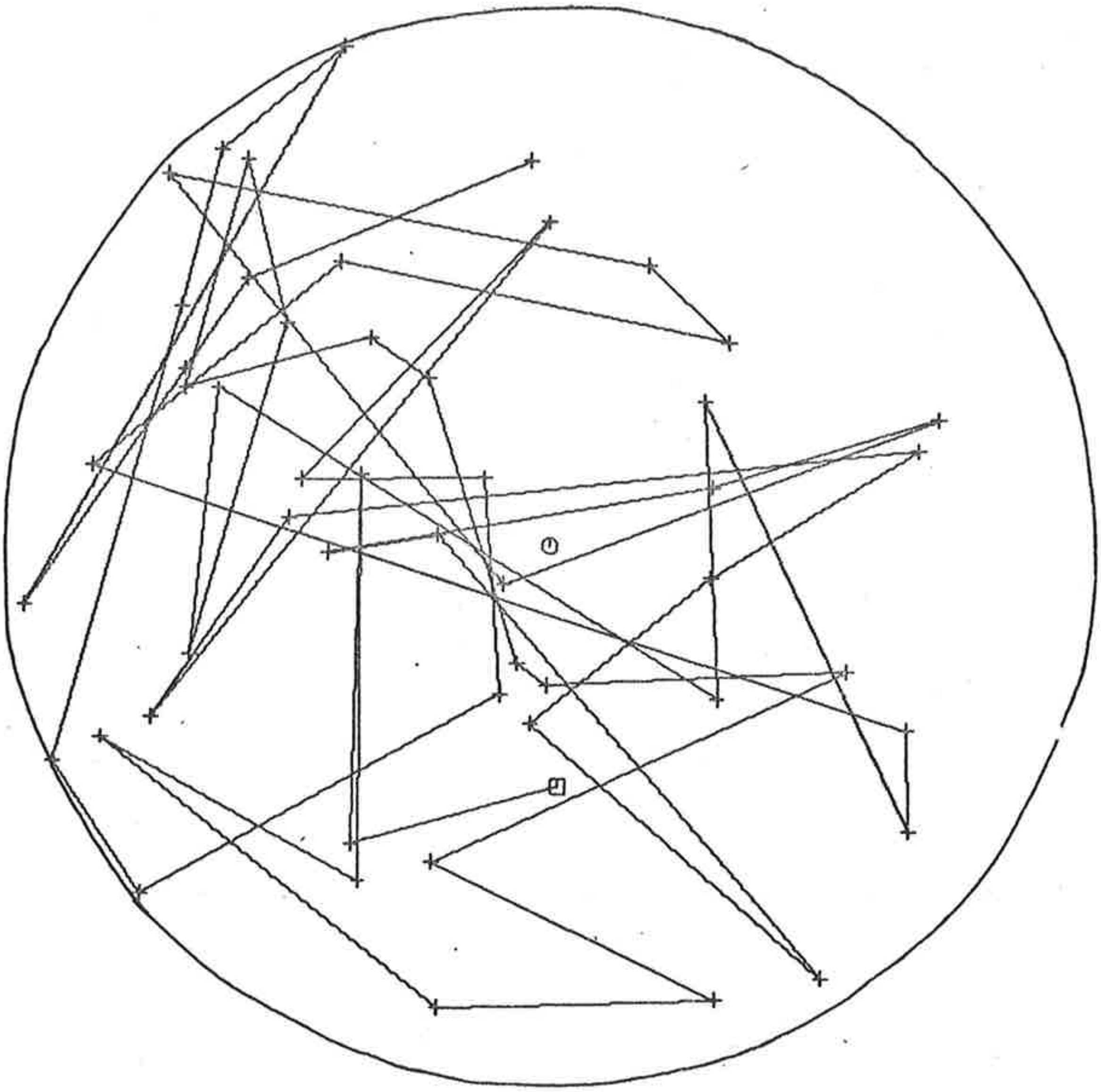
$$\hat{\Sigma}^*(3) \quad \begin{array}{rrrrrrrr}
.504 & -.193 & .018 & .119 & -.068 & -.045 & -.104 & .045 \\
.147 & .456 & .006 & -.004 & .027 & .020 & .073 & .071 \\
-.032 & .089 & .233 & .008 & -.057 & -.043 & -.020 & -.070 \\
-.036 & .077 & .068 & .067 & -.040 & .055 & .010 & .035 \\
.011 & -.064 & -.131 & .062 & .070 & -.112 & -.001 & -.070 \\
-.038 & -.013 & .108 & .032 & -.041 & .087 & .012 & -.057 \\
-.008 & .018 & .082 & -.045 & -.019 & .082 & .012 & .022 \\
-.034 & .093 & .030 & -.075 & .029 & .008 & .006 & .078
\end{array}$$

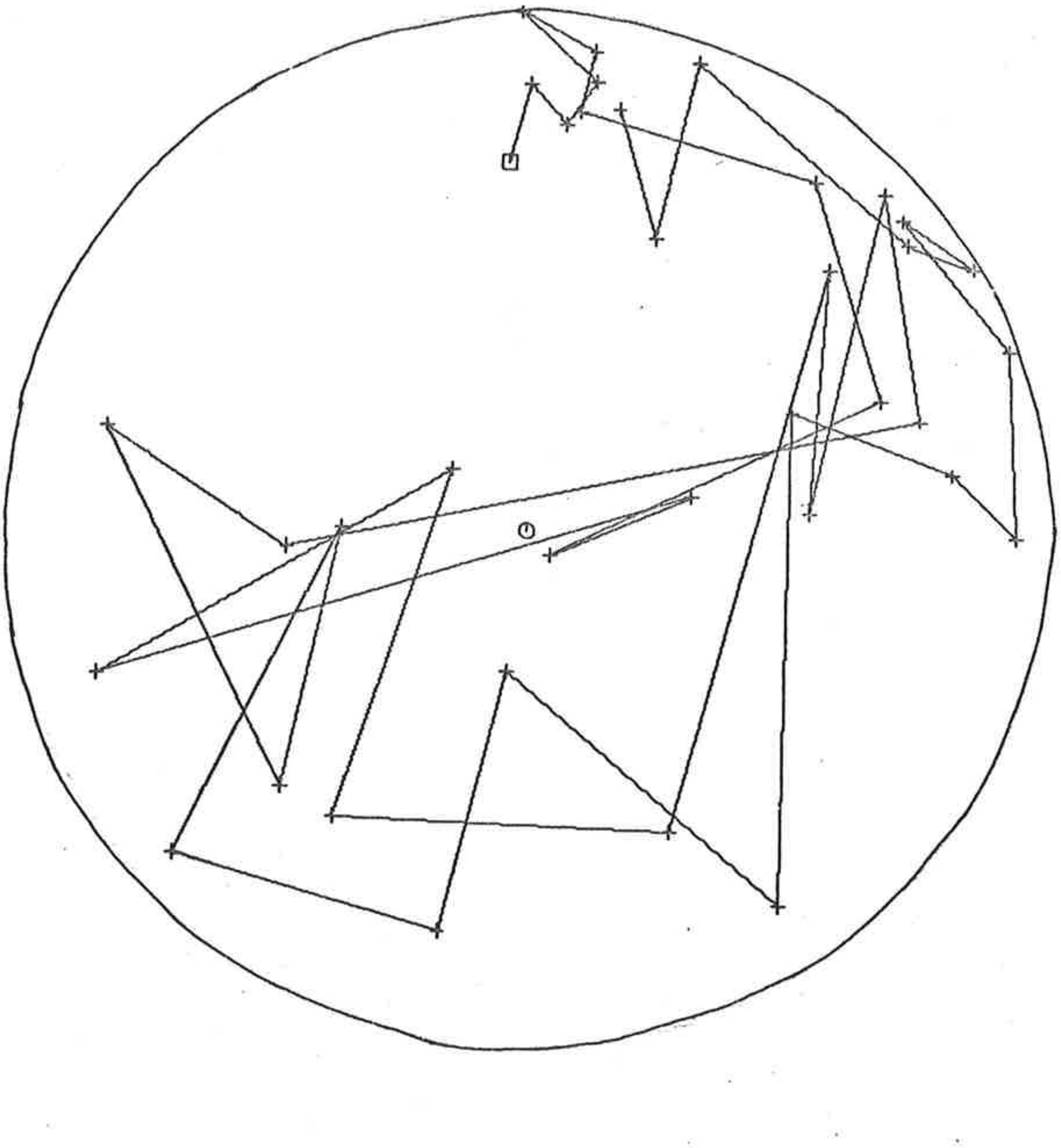
On the following pages the movements for animal number 2 are given. Consecutive positions are joined by straight lines. For clarity the movements are given in lots of 50. The □ in each plot marks the first position and this matches with the last + of the previous plot. The O marks the origin.


O
□









APPENDIX 4 : THE CANONICAL DECOMPOSITION OF BIVARIATE DISTRIBUTIONS.

THE CANONICAL DECOMPOSITION OF BIVARIATE DISTRIBUTIONS

P.L. Chesson *

Department of Statistics, University of
Adelaide, South Australia.

To appear in

Journal of Multivariate Analysis 6(4)

December 1976

---

* This work was supported by a C.S.I.R.O. Postgraduate Studentship.

ABSTRACT

The ordinary notion of a bivariate distribution has a natural generalisation. For this generalisation it is shown that a bivariate distribution can be characterised by a Hilbert space H and a family M_ρ, $0 \leqslant \rho \leqslant 1$, of subspaces of H. H specifies the marginal distributions whilst M_ρ is a summary of the dependence structure.

This characterisation extends existing ideas on canonical correlation.

Key words and phrases : Bivariate distribution, canonical correlation, spectral theorem.

AMS 1970 subject classification : 60E05, 62H05

1. Introduction

The general theory of canonical correlation in bivariate distributions can be said to have originated with the paper of Lancaster (1958) although previous workers considered the special case of a finite probability space. An extensive bibliography can be found in Lancaster (1969). Lancaster was chiefly concerned with finding expansions of the Radon-Nikodym derivative of a bivariate distribution relative to the product of its margins. However in more general situations than Lancaster considered such expansions may not exist.

It is the purpose of this paper to generalise the notion of a bivariate distribution and to obtain decompositions of such distributions which are general analogues of Lancaster's results.

It will be shown that any bivariate distribution can be characterised in a natural way by a Hilbert space H and a family $\{M_\rho, \rho \in [0,1]\}$ of subspaces of H. H depends only on the two marginal distributions and specifies them completely whilst M_ρ summarises the dependence structure of the bivariate distribution.

The main tool in the analysis is the spectral theorem for self-adjoint operators on real Hilbert spaces. Most discussions of spectral theory deal exclusively with complex Hilbert spaces. However all of the results needed here are easily proved from their complex counterparts.

Hannan (1961) was the first to apply general Hilbert space operator theory to canonical correlation. The approach given here is closely related to Hannan's but the use of the polar decomposition of a bounded linear operator is deliberately avoided because of the asymmetry inherent in that approach; however the polar decomposition does provide a useful alternative method for deriving the results given in this paper.

By a subspace of a Hilbert space H we will always mean a subset which is a Hilbert space with the same inner product. If M and N are two subspaces of H, $M \oplus N$ is the orthogonal sum of M and N, and $M \ominus N$ is the orthogonal complement of N in M.

2. The notion of a bivariate distribution

Let F and G be σ-fields of subsets of some set Ω. Define $F \vee G$ as the σ-field generated by the set $S = \{A \cap B | A \in F, B \in G\}$.

Definition. A probability measure P is a bivariate distribution for F and G if P is a probability measure on $F \vee G$.

To see the analogy with the usual concept let X and Y be random variables on some probability space (Ω, A, Q). The bivariate distribution of X and Y is generally thought of as the measure P' on B^2, the Borel sets in R^2, defined by $P'(B) = Q\big((X,Y) \in B\big)$, for $B \in B^2$. In terms of the definition above P' is a bivariate distribution for the σ-fields $\{B \times R^1 | B \in B\}$ and $\{R^1 \times B | B \in B\}$ where B is the Borel sets in R^1. More naturally we could consider the corresponding measure in the original space, that is, P defined as the restriction of Q to the sets $\{\omega \in \Omega | \big(X(\omega), Y(\omega)\big) \in B\}$ for $B \in B^2$. Clearly if F and G are the σ-subfields of A generated by X and Y respectively then P is a bivariate distribution for F and G. We will sometimes refer to a bivariate distribution arising in this way as a bivariate distribution for X and Y.

The general definition allows one to consider the joint distributions of pairs of random variables taking values in arbitrary measurable spaces, and includes the case where these random variables are stochastic processes such as considered by Hannan (loc. cit.).

Given a bivariate distribution P for the σ-fields $\mathcal{F}$ and $\mathcal{G}$ of subsets of Ω, let F and G be the restrictions of P to $\mathcal{F}$ and $\mathcal{G}$ respectively. The probability spaces $(\Omega,\mathcal{F},F)$ and $(\Omega,\mathcal{G},G)$ will be referred to as the margin spaces. Define $\mathcal{H}$ to be the set of ordered pairs, (f,g), of real valued functions on Ω such that f is $\mathcal{F}$-measurable, g is $\mathcal{G}$-measurable and $E[f^2+g^2] < \infty$. $\mathcal{H}$ is a real Hilbert space under the inner product

$$<(f_1,g_1), (f_2,g_2)> = \tfrac{1}{2} E[f_1 f_2 + g_1 g_2]$$

since $\mathcal{H}$ is essentially the direct sum of the spaces $L^2(\Omega,\mathcal{F},F)$ and $L^2(\Omega,\mathcal{G},G)$. Write $\mathcal{H} = L^2(\Omega,\mathcal{G},G) \dotplus L^2(\Omega,\mathcal{F},F)$.

Defined in this way $\mathcal{H}$ determines the margin spaces but tells us nothing about the dependence between them. To investigate this dependence note that $\mathcal{S}$ is a semi-ring generating $\mathcal{F} \vee \mathcal{G}$. Thus P is determined on $\mathcal{F} \vee \mathcal{G}$ by $E[fg]$ for $(f,g) \in \mathcal{H}$.

Lancaster (loc. cit.) was concerned with the space $(R^2,\mathcal{B}^2,P')$ induced by two random variables X and Y as discussed above. Let $F'(B) = Q(X \in B)$, $G'(B) = Q(Y \in B)$ for $B \in \mathcal{B}$, then Lancaster's result can be stated as follows:

If P' has a square summable Radon-Nikodym derivative, p, with respect to $F' \times G'$ then p has the representation

$$p(x,y) = \sum_{n=0}^{\ell} \rho_n \xi_n(x) \eta_n(y) \qquad (1)$$

where the series is mean square convergent with respect to $F' \times G'$,

$\{\xi_n\}$ and $\{\eta_n\}$ are orthonormal sequences of real valued functions relative to F' and G' respectively, $\rho_n \geqslant \rho_{n+1} > 0$, $\sum_0^{\ell} \rho_n^2 < \infty$, and $\ell \in \{0,1,..,\infty\}$.

In expression (1) the (ξ_n,η_n) are unique up to a change of sign when the ρ_n are distinct. If, however, $\rho_k = \rho_{k+1} = \dots = \rho_m$ then uniqueness holds only up to an equivalence. Specifically $\{(\xi_n,\eta_n)\}_{n=k}^{m}$ can be replaced by any set $\{\sum_{r=k}^{m} h_{nr}(\xi_r,\eta_r)\}_{n=k}^{m}$ where (h_{nr}), n, r=k,..,m is an orthogonal matrix. Later we shall find a convenient way to summarise this uniqueness.

Lancaster referred to the condition that p is square summable relative to $F' \times G'$ as "φ^2-boundedness".

In the general setting p may not even exist but from (1) we can easily obtain an expression (2) which does not involve p and so is more suitable for generalisation:

$$E[f(X)g(Y)] = \sum_{n=0}^{\ell} \rho_n E[f(X)\xi_n(X)]E[g(Y)\eta_n(Y)] \qquad (2)$$

for all f and g such that $E[f^2(X)+g^2(Y)] < \infty$.

Note that the requirement $\sum_{n=0}^{\ell} \rho_n^2 < \infty$ in (2) implies the existence and square summability of p and (1) follows from this. Thus (1) and (2) are equivalent.

To see how a family of subspaces summarises the dependence structure define M_ρ to be the subspace generated by the set $\{(\xi_n(X),\eta_n(Y)) \mid \rho_n \leqslant \rho\}$ where $\{(\xi_n(X),\eta_n(Y))\}_{n=0}^{\ell}$ is any particular sequence for which (2) (or (1)) holds. Expression (2) is true when and only when $\{(\xi_n(X),\eta_n(Y))\}_{n=0}^{\ell}$ is some orthonormal basis of M_1 with $(\xi_n(X),\eta_n(Y)) \in M_{\rho_n} \ominus M_\rho$, for $\rho_n > \rho$. The sequence $\{\rho_n\}$ is determined by the jumps in M_ρ.

3. <u>Main results.</u>

Let P be a bivariate distribution for the σ-fields

$\mathcal{F}$ and $\mathcal{G}$ of subsets of some set Ω. Using the definitions of $\mathcal{H}'$, F and G given in section 2 we have the following theorem.

Theorem

There exists a unique family of subspaces $\mathcal{M}_\rho$ of $\mathcal{H}$, $0 \leqslant \rho \leqslant 1$, such that

(i) $\bigcap_{\rho > \rho'} \mathcal{M}_\rho = \mathcal{M}_{\rho'}$, $0 \leqslant \rho' < 1$, and $\mathcal{M}_0 = \{0\}$.

(ii) If $\{(\xi_t^\rho, \eta_t^\rho)\}_{t \in T_\rho}$ is an orthonormal basis for $\mathcal{M}_\rho$ then $\{\xi_t^\rho\}_{t\in T_\rho}$ and $\{\eta_t^\rho\}_{t \in T_\rho}$ are orthonormal families of functions on the margin spaces $(\Omega, \mathcal{F}, F)$ and $(\Omega, \mathcal{G}, G)$ respectively.

(iii) For $(f,g) \in \mathcal{H}$ and $\{(\xi_t^\rho, \eta_t^\rho)\}_{t \in T_\rho}$ as in (ii)

$$E[fg] = \int_{(0,1]} \rho \, dQ(\rho)$$

where
$$Q(\rho) = \sum_{t \in T_\rho} \left(\int f\xi_t^\rho dF\right)\left(\int g\eta_t^\rho dG\right)$$
$$= \sum_{t \in T_\rho} E[f\xi_t^\rho]E[g\eta_t^\rho].$$

Proof Define the linear operator $\underset{\sim}{A}$ on $\mathcal{H}$ by $\underset{\sim}{A}(f,g) = (E[g|\mathcal{F}], E[f|\mathcal{G}])$. To check the boundedness of $\underset{\sim}{A}$ note that

$$E\{E[g|\mathcal{F}]\}^2 + E\{E[f|\mathcal{G}]\}^2 \leqslant E[f^2+g^2] = 2|(f,g)|^2.$$

This means $|\underset{\sim}{A}(f,g)|^2 \leqslant |(f,g)|^2$ so that $|\underset{\sim}{A}| \leqslant 1$ and from the fact that $\underset{\sim}{A}(1,1) = (1,1)$ we see that $|\underset{\sim}{A}| = 1$. Finally the symmetry of the expression $\langle \underset{\sim}{A}(f_1,g_1), (f_2,g_2)\rangle = \frac{1}{2} E[f_2 g_1 + f_1 g_2]$ indicates that $\underset{\sim}{A}$ is self-adjoint.

Now $\langle \underset{\sim}{A}(f,g), (f,g)\rangle = E[fg]$ and it is the spectral decomposition of $\underset{\sim}{A}$ that will be used to give the desired decomposition of $\mathcal{H}$.

From the above properties of $\underset{\sim}{A}$ there is a spectral measure $\underset{\sim}{P}$ on $\mathcal{B}$ such that $\underset{\sim}{A} = \int_{[-1,1]} \rho \, d\underset{\sim}{P}(\rho)$ and so

$$E[fg] = \int_{[-1,1]}\rho d\langle \underset{\sim}{P}(\rho)(f,g),(f,g)\rangle \tag{3}$$

Consider the self-adjoint isometry $\underset{\sim}{L}$ defined by $\underset{\sim}{L}(f,g) = (f,-g)$. Now $\underset{\sim}{L}\underset{\sim}{A}\underset{\sim}{L} = -\underset{\sim}{A}$ so that $-\underset{\sim}{A} = \int_{[-1,1]}\rho d\underset{\sim}{L}\underset{\sim}{P}(\rho)\underset{\sim}{L} = \int_{[-1,1]}\rho d\underset{\sim}{P}'(\rho)$ where $\underset{\sim}{P}'(B) = \underset{\sim}{P}(-B)$ for $B \in \mathcal{B}$. Since both $\underset{\sim}{L}\underset{\sim}{P}(\cdot)\underset{\sim}{L}$ and $\underset{\sim}{P}'(\cdot)$ are spectral measures it follows from the spectral theorem that they are equal. Hence (3) can be written as

$$E[fg] = \int_{(0,1]}\rho d\left\langle\left(\underset{\sim}{P}(\rho) - \underset{\sim}{L}\underset{\sim}{P}(\rho)\underset{\sim}{L}\right)(f,g),(f,g)\right\rangle$$

and the integral can be interpreted as a Lebesgue-Stieltjes integral with respect to the function of bounded variation

$$Q(\rho) = \left\langle\left(\underset{\sim}{P}(0,\rho] - \underset{\sim}{L}\underset{\sim}{P}(0,\rho]\underset{\sim}{L}\right)(f,g),(f,g)\right\rangle.$$

Define $M_\rho = \underset{\sim}{P}(0,\rho]H$. $\underset{\sim}{P}(0,\rho]$ can be written

$$\sum_{t\in T_\rho} \langle (f,g),(\xi_t^\rho,\eta_t^\rho)\rangle\,(\xi_t^\rho,\eta_t^\rho) = \sum_{t\in T_\rho} \tfrac{1}{2}E[f\xi_t^\rho+g\eta_t^\rho]\,(\xi_t^\rho,\eta_t^\rho)$$

where $\{(\xi_t^\rho,\eta_t^\rho)\}_{t\in T_\rho}$ is a basis for M_ρ.

$\underset{\sim}{L}\underset{\sim}{P}(0,\rho]\,\underset{\sim}{L}(f,g)$ is then $\sum_{t\in T_\rho} \tfrac{1}{2}E[f\xi_t^\rho-g\eta_t^\rho](\xi_t^\rho,-\eta_t^\rho)$ and $\left(\underset{\sim}{P}(0,\rho]-\underset{\sim}{L}\underset{\sim}{P}(0,\rho]\underset{\sim}{L}\right)(f,g) = \sum_{t\in T_\rho}\left(E[g\eta_t^\rho]\xi_t^\rho,\ E[f\xi_t^\rho]\eta_t^\rho\right)$ from which we conclude $Q(\rho) = \sum_{t\in T_\rho} E[f\xi_t^\rho]E[g\eta_t^\rho]$. This proves (iii).

M_ρ clearly has property (i); to verify (ii) suppose that (f_1,g_1), (f_2,g_2) belong to M_1 and are orthogonal, that is $E[f_1f_2+g_1g_2]=0$. $(f_1,-g_1)$ is an element of $\underset{\sim}{L}M_1 = \underset{\sim}{P}[-1,0)H$ and so is orthogonal to both (f_1,g_1) and (f_2,g_2) giving the additional equations $E[f_1^2] - E[g_1^2] = 0$, $E[f_1f_2] - E[g_1g_2] = 0$. Hence $E[f_1^2] = E[g_1^2]$ and $E[f_1f_2] = E[g_1g_2] = 0$ from which (ii) follows.

To see that M_ρ is unique assume M_ρ' is another family with properties (i) and (iii) and let $\underset{\sim}{P}_\rho'$ be the projection on M_ρ'. Define $\underset{\sim}{A}_1 = \int_{(0,1]}\rho d\underset{\sim}{P}_\rho'$ and $\underset{\sim}{A}_2 = \underset{\sim}{A}_1 - \underset{\sim}{L}\underset{\sim}{A}_1\underset{\sim}{L}$. With this construction $\langle\underset{\sim}{A}_2(f,g),(f,g)\rangle = E[fg] = \langle\underset{\sim}{A}(f,g),(f,g)\rangle$ and since both operators are self-adjoint they must be equal. Hence $M'_\rho = M_\rho$.

The origin of M_ρ from the spectral decomposition of $\underset{\sim}{A}$ immediately yields a number of interesting properties.

(iv) If $(f,g) \in M_\beta \ominus M_\alpha$ and $E[f^2] = 1$ $(= E[g^2])$ then $\alpha < E[fg] \leqslant \beta$.

(v) H can be decomposed as $H = M_1 \oplus O \oplus M_{-1}$ where, for $(f,g) \neq 0$, $E[fg]$ is positive, zero, or negative according to whether (f,g) belongs to M_1, O or M_{-1} respectively. Furthermore O is the unique subspace with the property $E[fg'] = E[f'g] = 0$ for all $(f,g) \in O$ and $(f',g') \in H$.

(vi) If $(f,g) \in H$ is written as the sum $\sum_{i=1}^{4}(f_i,g_i)$, $(f_1,g_1) \in M_1 \ominus M_\rho$, $(f_2,g_2) \in M_\rho$, $(f_3,g_3) \in O$, $(f_4,g_4) \in M_{-1}$ then $E[fg] = \sum_{i=1}^{4} E[f_i g_i]$.

(vii) The properties $E[f_1 g_1] > \rho\,|(f_1 g_1)|^2$ for $0 \neq (f_1,g_1) \in M_1 \ominus M_\rho$, $E[f_2 g_2] \leqslant \rho\,|(f_2,g_2)|^2$ for $(f_2,g_2) \in (M_1 \ominus M_\rho)^\perp$, and $E[(f_1+f_2)(g_1+g_2)] = E[f_1 g_1 + f_2 g_2]$ characterise $M_1 \ominus M_\rho$.

Proof

(iv) If $(f,g) \in M_\beta \ominus M_\alpha$ with $|(f,g)|^2 = 1$, or equivalently $E[f^2] = E[g^2] = 1$, the monotonic function $<\underset{\sim}{P}(\rho)(f,g),(f,g)>$ has the value 1 for $\rho \geqslant \beta$ and 0 for $\rho \leqslant \alpha$. Hence $\alpha < \int \rho\, d <\underset{\sim}{P}(\rho)(f,g),(f,g)> \leqslant \beta$ and so by (3) $\alpha < E[fg] \leqslant \beta$.

(v) Define M_{-1} as $\underset{\sim}{L}M_1$ and O as $H \ominus (M_1 \oplus M_{-1})$. Clearly $E[fg] > 0$ for $0 \neq (f,g) \in M_1$, and $E[fg] < 0$ for $0 \neq (f,g) \in M_{-1}$. O is the null space of $\underset{\sim}{A}$ so that if $(f,g) \in O$, $E[fg] = <\underset{\sim}{A}(f,g),(f,g)> = 0$. Also $<\underset{\sim}{A}(f,g),(f',g')> = 0$ for any $(f',g') \in H$ yielding $E[fg'] + E[f'g] = 0$. Since this is also true for $(f',-g')$ it follows that $E[fg'] = E[f'g] = 0$.

On the other hand if (f,g) has the property $E[fg'] = E[f'g] = 0$ for all $(f',g') \in H$ then clearly $\underset{\sim}{A}(f,g) = 0$ or $(f,g) \in O$.

(vi) The four subspaces $M_1 \ominus M_\rho$, M_ρ, O and M_{-1} are mutually orthogonal invariant subspaces of $\underset{\sim}{A}$.

(vii) Let L be another subspace with these properties. The equation $E[(f_1+f_2)(g_1+g_2)] = E[f_1g_1+f_2g_2]$ for $(f_1,g_1) \in L$, $(f_2,g_2) \in L^\perp$, entails that $E[f_2g_1+f_1g_2] = 0$ or $<\underset{\sim}{A}(f_1,g_1),(f_2,g_2)> = 0$ and hence that $\underset{\sim}{A}(f_1,g_1) \in L$ for $(f_1,g_1) \in L$. Thus L is an invariant subspace of $\underset{\sim}{A}$ and so

$$L = L \cap (M_1 \ominus M_\rho) \oplus L \cap (M_1 \ominus M_\rho)^\perp,$$
$$L^\perp = L^\perp \cap (M_1 \ominus M_\rho) \oplus L^\perp \cap (M_1 \ominus M_\rho)^\perp.$$

Considering the remaining common properties of L and $M_1 \ominus M_\rho$ it is clear that $L = M_1 \ominus M_\rho$.

4. The φ^2-bounded case

In this section an analogous condition to Lancaster's φ^2-boundedness is investigated.

We need to relate the bivariate distribution P, for F and G, to a measure which is in some sense the product of F and G. There are two approaches that could be taken to do this:

A. Attempt to construct a measure F · G, say, on $F \vee G$ such that F and G are independent under F · G, and the restrictions of F · G to F and G are respectively F and G.

B. Redefine P on $F \times G$, a σ-field for which we know a suitable product measure exists.

Since approach A is more natural and more informative it will be adopted here. On the other hand approach B is not devoid of interest and so a brief outline of that approach will be given at the end of the section.

We want to define $F \cdot G\ (A \cap B)$ as $F(A)G(B)$ for $A \in \mathcal{F}$, $B \in \mathcal{G}$. In order for $F \cdot G$ to be well defined we need to impose the condition:

I $A \cap B = \emptyset$ implies $F(A)G(B) = 0$ for all $A \in \mathcal{F}$, $B \in \mathcal{G}$.

However it has only been possible to prove finite additivity of $F \cdot G$ without further restrictions. To prove countable additivity it seems necessary to require:

II $A \cap B = \emptyset$ implies $A = \emptyset$ or $B = \emptyset$ for all $A \in \mathcal{F}$, $B \in \mathcal{G}$.

Renyi (1970, page 115) gives the following lemma which we prove here by a simpler method.

Lemma. Let condition II hold and $F \cdot G\ (A \cap B) = F(A)G(B)$, $A \in \mathcal{F}$, $B \in \mathcal{G}$. $F \cdot G$ is countably additive on $\mathcal{S}$ and so has a unique extension to a probability measure on $\mathcal{F} \vee \mathcal{G}$

Proof. Let $I : \Omega \to \Omega^2$ be defined by $I(\omega) = (\omega, \omega)$. The inverse image, I^{-1}, maps the semi-ring $\mathcal{T} = \{A \times B \mid A \in \mathcal{F}, B \in \mathcal{G}\}$ onto $\mathcal{S}$. Condition II entails that $I^{-1}(C) = \emptyset$ implies $C = \emptyset$ for $C \in \mathcal{T}$. This fact will be used to infer the countable additivity of $F \cdot G$ on $\mathcal{S}$ from that of $F \times G$ on $\mathcal{T}$.

Suppose $A \cap B = \bigcup_{n=0}^{\infty} A_n \cap B_n$; $A, A_n \in \mathcal{F}$; $B, B_n \in \mathcal{G}$; and the union is disjoint. From condition II it follows that $\bigcup_{n=0}^{\infty} A_n \times B_n$ is a disjoint union. Now

$$\begin{aligned} \emptyset &= A \cap B - \bigcup_{n=0}^{\infty} A_n \cap B_n \\ &= I^{-1}\left(A \times B - \bigcup_{n=0}^{\infty} A_n \times B_n\right) \\ &= I^{-1}\left(\bigcap_{m=0}^{\infty} A \times B - \bigcup_{n=0}^{m} A_n \times B_n\right) \end{aligned}$$

Using the semi-ring properties of $\mathcal{T}$ this can be written

$$\begin{aligned} & I^{-1}\left(\bigcap_{m=0}^{\infty} \bigcup_{n=0}^{r_m} A_{nm} \times B_{nm}\right), \quad \text{for } A_{nm} \times B_{nm} \in \mathcal{T}, \\ &= I^{-1}\left(\bigcup_{f} \bigcap_{m=0}^{\infty} A_{f(m)m} \times B_{f(m)m}\right) \end{aligned}$$

$$= I^{-1}\Big(\bigcup_{f}\ \big(\bigcap_{m=0}^{\infty} A_{f(m)m}\big)\times\big(\bigcap_{m=0}^{\infty} B_{f(m)m}\big)\Big)$$

Where the union is over all functions $f : \{0,1,\ldots\} \to \{0,1,\ldots\}$ satisfying $f(m) \leqslant r_m$. Hence

$$\emptyset = \bigcup_{f} I^{-1}\Big(\bigcap_{m=0}^{\infty} A_{f(m)m} \times \bigcap_{m=0}^{\infty} B_{f(m)m}\Big)$$

and so

$$\bigcap_{m=0}^{\infty} A_{f(m)m} \times \bigcap_{m=0}^{\infty} B_{f(m)m} = \emptyset$$

which means

$$A \times B - \bigcup_{n=0}^{\infty} A_n \times B_n = \emptyset .$$

It is easy to see that $\bigcup_{n=0}^{\infty} A_n \times B_n - A \times B = \emptyset$ giving $A \times B = \bigcup_{n=0}^{\infty} A_n \times B_n$. Finally

$$F \cdot G(A \cap B) = F \times G(A \times B) = \sum_{n=0}^{\infty} F \times G(A_n \times B_n) = \sum_{n=0}^{\infty} F \cdot G(A_n \cap B_n)$$

Hence $F \cdot G$ is countably additive on $\mathcal{S}$.

The extension of $F \cdot G$ to $\mathcal{F} \vee \mathcal{G}$ will be referred to as $F \cdot G$.

We now impose a φ^2-boundedness condition.

III Suppose that P is absolutely continuous with respect to $F \cdot G$ and the Radon-Nikodym derivative, p, is square summable relative to $F \cdot G$.

Let $\{\varphi_s\}_{s \in S}$ be an orthonormal basis for $L^2(\Omega, \mathcal{F}, F)$ and $\{\psi_t\}_{t \in T}$ an orthonormal basis for $L^2(\Omega, \mathcal{G}, G)$. $\{\varphi_s \psi_t\}_{(s,t) \in S \times T}$ is an orthonormal set (indeed a basis) for $L^2(\Omega, \mathcal{F} \vee \mathcal{G}, F \cdot G)$. Condition III entails that $p \in L^2(\Omega, \mathcal{F} \vee \mathcal{G}, F \cdot G)$ so that $c_{st} = \int p \varphi_s \psi_t \, d\, F \cdot G$ satisfies $\sum_{s,t} c_{st}^2 < \infty$.

Now $\{x_u\}_{u \in U} = \{2^{\frac{1}{2}}(\varphi_s, 0)\}_{s \in S} \cup \{2^{\frac{1}{2}}(0, \psi_t)\}_{t \in T}$ is an orthonormal basis for $\mathcal{H}$.

$c_{st} = E[\varphi_s\psi_t] = 2 <\underset{\sim}{A}(\varphi_s, 0), (0, \psi_t)> = 2<\underset{\sim}{A}(0, \psi_t), (\varphi_s, 0)>$ and $<\underset{\sim}{A}(\varphi_t, 0), (\varphi_{t'}, 0)> = <\underset{\sim}{A}(0, \psi_s), (0, \psi_{s'})> = 0$ so that $\sum_{u,v} |<\underset{\sim}{A}x_u, x_v>|^2 = 8 \sum_{s,t} c_{st}^2 = < \infty$. Hence $\underset{\sim}{A}$ is a Hilbert-Schmidt operator (see Dunford & Schwartz, 1963). Thus the only element of the continuous spectrum of $\underset{\sim}{A}$ is the point 0. The discrete spectrum consists of the set $\{\rho_n\}_{n=0}^{\infty} \cup \{-\rho_n\}_{n=0}^{\infty}$, $\rho_n \geqslant 0$ and $\rho_n \downarrow 0$. $\mathcal{M}_\rho$ is the Hilbert space generated by the eigenvectors of $\underset{\sim}{A}$ corresponding to positive eigenvalues $\leqslant \rho$. Finally, if $\ell + 1 = \dim \mathcal{M}_1$, 3(iii) reduces to

$$E[fg] = \sum_{n=0}^{\ell} \rho_n E[f\xi_n]\, E[g\eta_n]$$

for all $(f,g) \in \mathcal{H}$ and $\{(\xi_n, \eta_n)\}_{n=0}^{\ell}$ is a basis for $\mathcal{M}$, such that $(\xi_n, \eta_n) \in \mathcal{M}_{\rho_n} \ominus \mathcal{M}_\rho$, for $\rho_n > \rho$.

To proceed by approach B note that $I^{-1}(\mathcal{F} \times \mathcal{G}) = \mathcal{F} \vee \mathcal{G}$ and so define $P'(C)$ as $P(I^{-1}(C))$ for $C \in \mathcal{F} \times \mathcal{G}$. Since $\mathcal{F} \times \mathcal{G} = \mathcal{F}' \vee \mathcal{G}'$, where $\mathcal{F}' = \{A \times \Omega | A \in \mathcal{F}\}$, $\mathcal{G}' = \{\Omega \times B | B \in \mathcal{G}\}$, P' is a bivariate distribution for $\mathcal{F}'$ and $\mathcal{G}'$ and the derivation can proceed as a special case of approach A by passing the lemma since $F \times G$ replaces $F \cdot G$. At the final stage it is necessary to relate the $\mathcal{F}'$ measurable functions on $\Omega \times \Omega$ to the $\mathcal{F}$ measurable functions on Ω, etc. to obtain a decomposition on the original space.

It is interesting to note that the absolute continuity of P' relative to $F \times G$ implies condition I. Further if P' is absolutely continuous relative to $F \times G$ for every bivariate distribution P, for $\mathcal{F}$ and $\mathcal{G}$, then condition II holds.

5. Examples of φ^2-unbounded bivariate distributions

(i) Mixtures of bivariate normals.

Let P_ρ be the standardised bivariate normal measure on $(R^2,\mathcal{B}^2)$ with correlation $\rho \geqslant 0$, and Φ the standard normal measure on $(R,\mathcal{B})$.

Let X and Y be the functions from R^2 to R defined by $X(x,y) = x$ and $Y(x,y) = y$. It is well known that P_ρ has the decomposition

$$E[f(X)g(Y)] = \sum_{n=0}^{\infty} \rho^n E[H_n(X)f(X)]\ E[H_n(Y)g(Y)]$$

for $\big(f(X),g(Y)\big) \in L^2(R,\mathcal{B},\Phi) \dot{+} L^2(R,\mathcal{B},\Phi) = \mathcal{H}$. H_n is the n^{th} Hermite polynomial normalised to yield $\int H_n^2(x)\ d\Phi(x) = 1$.

Let Q be a probability measure on the Borel sets in $[0,1]$ then $P = \int P_\rho dQ(\rho)$ has the expansion

$$E[f(X)g(Y)] = \sum_{n=0}^{\infty} \rho_n E[H_n(X)f(X)]\ E[H_n(Y)g(Y)] \qquad (4)$$

where $\rho_n = \int \rho^n dQ(\rho)$. $\mathcal{H}$ is unchanged but M_ρ is the subspace spanned by $\bigcup_{\rho_n \leqslant \rho} \{\big(H_n(X),H_n(Y)\big)\}$. Many probability measures Q satisfy $\sum_{n=0}^{\infty} \rho_n^2 = \infty$, that is, P is not φ^2-bounded; but, as is shown above, a useful decomposition of P is still available. For a discussion of the φ^2-bounded case see Sarmanov and Bratoeva (1967).

Other examples of φ^2-unbounded bivariate distributions having a discrete canonical decomposition can be found in Eagleson (1964) and Griffiths (1970) although the decompositions are not pointed out explicitly. Unfortunately such distributions have frequently been excluded from discussion, because they are not φ^2-bounded, when an expression such as (4) was all that was required.

(ii) A bivariate distribution with continuous canonical decomposition

Hannan (loc. cit.) gives a number of examples of bivariate distributions, for pairs of stochastic processes, which have continuous canonical decompositions. A simple example of a bivariate distribution on $\mathcal{B}[0,1]^2$, the Borel sets in $[0,1]^2$, having a continuous canonical decomposition is given here.

Let P be defined by

$$P(C) = \int_{\{x \mid (x,x) \in C\}}^{\rho} d\rho + \int_0^1 \frac{1}{\rho}\Big(\int_{C \cap [0,\rho]^2} d(u,v)\Big) d\rho$$

for $C \in \mathcal{B}[0,1]^2$. P is a bivariate distribution for the random variables X and Y defined in example (i). The margin spaces are both equivalent to Lebesgue measure restricted to the Borel sets in $[0,1]$.

For $(f(X), g(Y)) \in \mathcal{H}$

$$E[f(X)g(Y)] = \int_0^1 \rho f(\rho) g(\rho) d\rho + \int_0^1 \rho\Big(\frac{1}{\rho}\int_0^\rho f(u)du\Big)\Big(\frac{1}{\rho}\int_0^\rho g(u)du\Big)d\rho$$

$\mathcal{M}_\rho$ is simply the space

$$\mathcal{L}_\rho = \{(f(X), f(Y)) \in \mathcal{H} \mid \int_0^1 f(x)dx = 0 \text{ and } f(x) = 0 \text{ for } x > \rho\},$$

if $\rho < 1$, and $\mathcal{M}_1 = \mathcal{L}_1 = \{(f(X), f(Y)) \in \mathcal{H}\}$. The proof of this illustrates the application of the characterisation 2(vii) of $\mathcal{M}_1 \ominus \mathcal{M}_\rho$.

Now $\mathcal{H} = (\mathcal{L}_1 \ominus \mathcal{L}_\rho) \oplus \mathcal{L}_\rho \oplus \mathcal{L}_1^\perp$ so that $(\mathcal{L}_1 \ominus \mathcal{L}_\rho)^\perp = \mathcal{L}_\rho \oplus \mathcal{L}_1^\perp$. $\mathcal{L}_1 \ominus \mathcal{L}_\rho$ is the space $\{(f(X), f(Y)) \in \mathcal{H} \mid f \text{ is constant on } [0,\rho]\}$. Let $(f_1(X), g_1(Y)) \in \mathcal{L}_1 \ominus \mathcal{L}_\rho$, $(f_2(X), g_2(Y)) \in \mathcal{L}_\rho$, $(f_3(X), g_3(Y)) \in \mathcal{L}_1^\perp$. It is easy to see that $E[f_1(X)g_2(Y)] = E[f_2(X)g_1(Y)] = 0$ and, noting that $g_3 = -f_3$ and $g_1 = f_1$, $E[f_1(X)g_3(Y) + f_3(X)g_1(Y)] = 0$. Hence

$$E\Big[\Big(\sum_1^3 f_i(X)\Big)\Big(\sum_1^3 g_i(Y)\Big)\Big] = \sum_1^3 E[f_i(X)g_i(Y)] \tag{5}$$

$$\begin{aligned} E[f_1(X)g_1(Y)] &= \int_0^1 u f_1^2(u)du + \int_0^1 u\Big(\frac{1}{u}\int_0^u f_1(v)dv\Big)^2 du \\ &> \int_0^\rho uc^2 du + \rho\int_\rho^1 f_1^2(u)du + \int_0^\rho uc^2 du, \end{aligned}$$

provided $f_1 \not\equiv 0$ and where $f_1^{-1}\{c\} = [0,\rho]$.

Thus $E[f_1(X)g_1(Y)] > \rho\int_0^1 f_1^2(u)du = \rho\|(f_1(X), g_1(Y))\|^2$

$$\begin{aligned} E[f_2(X)g_2(Y)] &= \int_0^1 uf_2^2(u)\,du + \int_0^1 u\Big(\frac{1}{u}\int_0^u f_2(v)\,dv\Big)^2 du \\ &\leqslant \int_0^1 u\,d\{\int_0^u f_2^2(v)\,dv\} + \int_0^1 \{\int_0^u f_2^2(v)\,dv\}du \\ &= \rho\int_0^1 f_2^2(u)\,dv = \rho\big|\big(f_2(X),g_2(Y)\big)\big|^2 \end{aligned}$$

Considering (5) and the fact that $E[f_3(X)g_3(Y)] \leqslant 0$ it follows that $E[f(X)g(Y)] \leqslant \rho|(f(X),g(X))|^2$ when $(f(X),g(Y)) \in (L_1 \ominus L_\rho)^\perp$. This completes the verification that $L_1 \ominus L_\rho$ has the properties 2(vii) of $M_1 \ominus M_\rho$ for all ρ in $[0,1]$. Thus we see that $M_\rho = L_\rho$.

ACKNOWLEDGEMENT

I would like to thank Dr. G.M. Tallis for introducing the theory of canonical correlation to me and for encouraging me to continue with this work.

REFERENCES

[1] Dunford, N. and Schwartz, J.T. (1963). Linear operators part II : Spectral Theory. Interscience, New York.

[2] Eagleson, G.K. (1964). Polynomial expansions of bivariate distributions. Ann. Math. Statist. 35 1208-1215.

[3] Griffiths, R.C. (1970). Positive definite sequences and canonical correlation coefficients. Austral. J. Statist. 12 162-165.

[4] Hannan, E.J. (1961). The general theory of canonical correlation and its relation to functional analysis. J. Austral. Math. Soc. 2 229-242.

[5] Lancaster, H.O. (1958). The structure of bivariate distributions. Ann. Math. Statist. 29 719-736.

[6] Lancaster, H.O. (1969). The chi-squared distribution. Wiley, New York.

[7] Renyi, A. (1970). Foundations of probability. Holden-Day, San Francisco.

[8] Sarmanov, I.O. and Bratoeva, Z.N. (1967). Probabilistic properties of bilinear expansions of Hermite polynomials. Theo. Probability Appl. 12 470-481.

[9] Tyan, S. and Thomas, J.B. (1975). Characterisation of a class of bivariate distribution functions. J. Multivariate Anal. 5 227-235.

[10] Helson, H. (1948). Remark on measures in almost independent fields. Studia Math. 10, 182-183.

Additions to the Bibliography

Beverton, R.J.H. and S.J. Holt (1957). On the dynamics of exploited fish populations. Fish Invest. Series II, Vol. XIX.

Freedman, D. (1971). Markov chains. Holdenday, San Francisco.

Heyde, C.C. (1974). On the central limit theorem for stationary processes. Z. Wahr. 30, 315-320.

Lipton, G.R. and D.J. Sutherland (1970). Activity rhythms in the American cockroach, *Periplaneta americana*. J. Insect Physiol. 16, 1555-1566.

BIBLIOGRAPHY

Adke, S.R. (1969). A birth, death and migration process. J. Appl. Prob. 6, 687-691.

Ames, W.F. (1969). Numerical methods for partial differential equations. Nelson, London.

Andersen, D.T. (1967). Larval development and segment formation in the branchiopod crustaceans *Limnadia stanleyana* King (Conchostraca) and *Artemia salina* (L) (Anostraca). Aust. J. Zool. 15, 47-91.

Andrewartha, H.G. (1970). Introduction to the study of animal populations. 2nd ed. Methuen, London.

Andrewartha, H.G. and L.C. Birch (1954). The distribution and abundance of animals. University of Chicago Press, Chicago.

Arnason, A.N. (1970). Migration models for animal population Ph.D. Thesis, Edinburgh.

Bailey, N.T.J. (1968). Stochastic birth, death and migration processes for spatially distributed populations. Biometrika 55, 189-198.

Baggett, L. and D. Stroock (1974). An ergodic theorem for Poisson processes on a compact group with applications to random evolutions. J. Functional Analysis 16. 405-414.

Bartlett, M.S. and D.G. Kendall (1951). On the use of the characteristic functional in the analysis of some stochastic processes occurring in physics and biology. Proc. Camb. Phil. Soc. 47, 65-76.

Bartlett, M.S. (1949). Some evolutionary stochastic processes. J.R. Statist. Soc. B, 11, 211-229.

Bellman, R.E. (1970). Introduction to matrix analysis. 2nd ed. McGraw-Hill, New York.

Berg, H.C. and D.A. Brown (1972). Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. Nature 239, 500-504.

Billingsley, P. (1968). Convergence of probability measures. Wiley, New York.

Billingsley, P. (1961). Statistical inference for Markov processes. Univ. of Chicago Press, Chicago.

Blumenthal, R.M. and R.K. Getoor (1968). Markov processes and potential theory. Academic Press, New York.

Bochner, S. (1949). Diffusion equation and stochastic processes. Proc. Nat. Acad. Sci. USA 35, 368-370.

Bradford, E. and J.R. Philip (1970). Stability of steady distributions of asocial populations dispersing in one dimension. J. Theor. Biol. 29, 13-26.

Breiman, L. (1968). Probability, Addison-Wesley.

Brown, S.L. and G.H. Orians (1970). Spacing patterns in mobile animals. Ann. Rev. Ecol. Syst. 1, 239-257.

Brown, L.E. (1966). Home range and movement of small mammals. Symp. zool. Soc. Lond. 18, 111-142.

Bryant, E. (1969). A model of plant-to-plant movement of aphids: A new approach: Res. Popul. Ecol. 11, 34-39.

Burt, W.H. (1943). Territoriality and home range concepts as applied to mammals. J. Mamm. 24, 346-352.

Cane, V.R. (1959). Behaviour sequences as semi-Markov chains. J.R. Statist. Soc. B. 21, 36-58.

Chesson, P.L. (1976). The canonical decomposition of bivariate distributions. To appear in J. Multivariate Anal. 6.

Chewning, W.C. (1975). Migratory effects in predator-prey models. Math. Bioscience 23, 253-262.

Chung, K.L. (1967). Markov chains with stationary transition probabilities. Springer-Verlag, Berlin.

Cohen, D. (1967). Optimisation of seasonal migratory behaviour. Amer. Nat. 101, 1-17.

Comins, H.N. and D.W.E. Blatt (1974). Prey-predator models in spatially heterogeneous environments. J. Theor. Biol. 48, 75-83.

Cox, D.R. and W.L. Smith (1957). On the distribution of *Tribolium confusum* in a container. Biometrika 44, 328-335.

Cramér, H. and M.R. Leadbetter (1967). Stationary and related stochastic processes. Wiley, New York.

Croll, N.A. and A. Blair (1973). Inherent movement patterns of larval nematodes, with a stochastic model to simulate movement of infective hookworm larvae. Parasitology 67, 53-66.

Crumpacker, D.W. and J.S. Williams (1973). Density, dispersion and population structure in *Drosophila pseudoobscura*. Ecol. Monog. 43, 499-538.

Davis, A.W. (1970). Some generalisations of Bailey's birth, death and migration model. Adv. Appl. Prob. 2, 83-109.

Dempster, J.P. (1957). The population dynamics of the Moroccan locust in Cyprus. Anti-Locust Bull. 27.

Dobzhansky, T. and S. Wright (1947). Genetics of natural populations. XV. Rate of diffusion of a mutant gene through a population of *Drosophila pseudoobscura*. Genetics 32, 303-324.

Dobzhansky, T. and S. Wright (1943). Genetics of natural populations X. Dispersion rates in *Drosophila pseudoobscura*. Genetics 28, 304-340.

Doob, J.L. (1953). Stochastic processes. Wiley, New York.

Doyle, R.W. (1975). Settlement of planktonic larvae; A theory of habitat selection in varying environments. Am. Nat. 109, 113-126.

Dudzinski, M.L.. P.J. Pahl and G.W. Arnold (1969). Quantitative assessment of grazing behaviour of sheep in arid areas. J. Range Mgmt. 22, 230-5.

Dunford, N. and J.T. Schwartz (1971). Linear operators Part III. Spectral operators. Wiley-Interscience, New York.

Dynkin, E.B. (1965). Markov Processes. Springer-Verlag, Berlin.

Feller, W. (1971). An introduction to probability theory and its applications, Vol. II. 2nd ed. Wiley, New York.

Fleming, W.H. (1975). A selection migration model in population genetics. J. Math. Biol. 2, 219-233.

Fleming, W.H. and C.H. Su (1974). Some one dimensional migration models in population genetics theory. Theor. Pop. Biol. 5, 431-449.

Foguel, S.R. (1969). The ergodic theory of Markov processes. Van-Nostrand-Reinhold, New York.

Fraenkel, G.S. and D.L. Gunn (1940). Orientation of animals: kineses, taxes and compass reactions. Oxf. Clar. Press.

Gadgil, M. (1971). Dispersal: Population consequences and evolution. Ecology 52, 253-261.

Gibbons, J.D. (1971). Nonparametric statistical inference. McGraw-Hill, New York.

Gillespie, J.H. (1975). The role of migration in the genetic structure of populations in temporarily and spatially varying environments I. Conditions for polymorphisms. Am. Nat. 109, 127-135.

Goodman, G.S. (1970). An intrinsic time for non-stationary finite Markov chains. Z. Wahr. 16, 165-180.

Goodrich, R.K. (1972). The spectral theorem for real Hilbert space. Acta Sci. Math. 33, 123-127.

Gray, A. and T.M. MacRobert (1922). Treatise on Bessel functions and their applications to physics. Macmillan, London.

Griffiths, R.C. (1970). Positive definite sequences and canonical correlation coefficients. Austral. J. Statist. 12, 162-165.

Gurney, W.S.C. and R.M. Nisbet (1975). The regulation of inhomogeneous populations. J. Theor. Biol. 52, 441-457.

Gutierrez, A.P. et al (1974). The ecology of *Aphis craccivora* KOCH and subterranean clover stunt virus in south-east. Australia II. A model of cowpea aphid populations in temperate pastures. J. Appl. Ecol. 11, 1-20.

Hadeler, K.P., U. an der Heiden and F. Rothe (1974). Non-homogeneous spatial distributions of populations. J. Math. Biol. 1, 165-176.

Hajek, J. and Z. Sidak (1967). Theory of Rank Tests. Academic Press, New York.

Halmos, P.R. (1951). An introduction to Hilbert space and the theory of spectral multiplicity. Chelsea Pub. Co. New York.

Hannan, E.J. (1970). Multiple time series. J. Wiley, New York.

Harris, T.E. (1971). Random measure and point processes. Z. Wahr. 18, 85-115.

Harris, T.E. (1968). Counting measures, monotone random set functions. Z. Wahr. 10, 102-119.

Heath, M. (1924). The external morphology of certain phyllopods. J. Morph. 38, 453-483.

Helmberg, G. (1969). Introduction to spectral theory in Hilbert space. North Holland Pub. Co., Amsterdam.

Hewitt, E. and L.J. Savage, (1955). Symmetric measures on cartesian products. Trans. Amer. Math. Soc. 80, 470-501.

Hindley, J.P.R. (1975). Effects of endogenous and some exogenous factors on the activity of the juvenile banana prawn *Penaeus mirguiensis*. Marine Biol. 29, 1-8.

Hughes, R.D. and N. Gilbert (1968). A model of an aphid population - a general statement. J. Anim. Ecol. 37, 553-563.

Hutchinson, G.E. (1951). Copepodology for the ornithologist. Ecology 32, 571-577.

Ibragimov, I.A. (1975). A note on the central limit theorem for dependent random variables. Theory Prob. Applications 20, 135-141.

Ionescu Tulcea, C.T. (1949). Mesures dans les espaces produits. Atti. Accad. Naz. Lincei. Rend. 7, 208-211.

Jagers, P. (1974). Aspects of random measures and point processes, in P. Ney, And S. Port (eds.) Advances in probability and related topics. Vol. 3. Dekker, New York.

Jewell, P.A. and C. Loizos (eds.) (1966). Play, exploration and territory in mammals. Symp. Zool. Soc. Lond. No. 18.

Jirina, M. (1972). Convergence in distribution of random measures. Ann. Math. Statist. 43, 1727-1731.

Johansen, S. (1973a). A central limit theorem for finite semi-groups and its application to the imbedding problem for finite state space Markov chains. Z. Wahr. 26, 171-190.

Johansen, S. (1973b). The bang-bang problem for stochastic matrices. Z. Wahr. 26, 191-195.

Johnson, C.G. (1969). Migration and dispersal of insects by flight. Methuen, London.

Jones, R. (1959). A method of analyses of some tagged haddock returns. J. Cons. Perm. Int. Explor. Mer. 25, 58-72.

Kallenberg, O. (1975). Random measures. Schriftenreihe des ZI für Math. und Mech. 23, 1-104.

Kelley, J.L. (1955). General topology. van Nostrand, New York.

Kendall, D.G. (1959a). Unitary dilations of Markov transition operators, and the corresponding integral representations for transition probability matrices. Probability and statistics (The Harold Cramér Volume) 139-161. John Wiley, New York.

Kendall, D.G. (1959b). Unitary dilations of one-parameter semigroups of Markov transition operators, and the corresponding integral representations for Markov processes with a countable infinity of states. Proc. London Math. Soc. 9, 417-431.

Kielson, J. (1965). A review of transient behaviour in regular diffusion and birth-death processes. Part II. J. Appl. Prob. 2, 405-428.

Kielson, J. (1964). A review of transient behaviour in regular diffusion and birth-death processes. J. Appl. Prob. 1, 247-266.

Kiester, A.R. and M. Slatkin (1974). A strategy of movement and resource utilisation. Theor. Popul. Biol. 6, 1-20.

Kimura, M. and G.H. Weiss (1964). The stepping stone model of population structure and the decline of genetic correlation with distance. Genetics 49, 561-576.

Kingman, J.F.C. (1962). The imbedding problem for finite Markov chains. Z. Wahr. 1, 14-24.

Kitching, R. (1971). A simple simulation model of dispersal of animals among units of discrete habitat. Oecologia 7, 95-116.

Krebs, C.J. (1972). Ecology. Harper and Row, New York.

Kuenen, D.J. (1939). Systematical and physiological notes on the brine shrimp *Artemia*. Ach. neér Zool. 3, 365-449.

Lancaster, H.O. (1969). The chi-squared distribution. J. Wiley, New York.

Lapidus, I.R. and R. Schiller (1975). Bacterial chemotaxis in a fixed attractant gradient. J. Theor. Biol. 53, 215-222.

Ledermann, W. (1950). On the asymptotic probability distribution for certain Markov processes. Proc. Camb. Phil. Soc. 46, 581-594.

Lefkovitch, L.P. (1967). Larval movement of *Lasioderma serricorne* (F.) (Coleoptera Anobiidae) in a plane. Res. Popul. Ecol. 9, 45-52.

Levin, S.A. (1974). Dispersion and population interactions. Amer. Nat. 108, 207-228.

Lockhead, J.H. (1950). *Artemia* in F.A. Brown (ed.) Selected invertebrate types. Wiley, New York. 394-399.

Lockie, J.D. (1966). Territory in small carnivores. Symp. Zool. Soc. Lond. 18, 143-165.

Lovely, P.S. and F.W. Dahlquist (1975). Statistical measures of bacterial mobility and chemotaxis. J. Theor. Biol. 50, 477-496.

Mardia, K.V. (1972). Statistics of directional data. Academic Press, London.

Marten, G.G. (1973). The patterns of *Peromyscus* activity and their correlations with weather. J. Mamm. 54, 169-188.

Marten, G.G. (1972). The remote sensing approach to censusing. Res. Popul. Ecol. 14, 36-57.

May, R.M. (1973). Stability and complexity in model ecosystems. Princeton Univ. Press, Princeton.

Maynard Smith, J. (1974). Models in ecology. C.U.P., Cambridge.

Morgan, B.J.T. (1976). Markov properties of sequences of behaviours. Appl. Statist. 25, 31-36.

Mosimann, J.E. (1958). The evolutionary significance of rare matings in animal populations. Evolution 12, 246-261

Mossing, T. (1975). Measuring small mammal activity with passage counters. Oikos 26, 237-239.

Moyal, J.E. (1962). The general theory of stochastic population processes. Acta Math. 108, 1-31.

Murdie, G. and M.D. Hassel (1973). Food distribution, search success and predator-prey models in M.S. Bartlett and R.W. Hiorns(eds.). The mathematical theory of the dynamics of biological populations. Academic Press, New York.

Murray, J.D. (1975). Non-existence of wave solutions for the class of reaction-diffusion equations given by the Volterra interacting population equation with diffusion. J. Theor. Biol. 52, 459-469.

Nagaev, S.V. and A.B. Mukhin (1966). On a case of convergence to a uniform distribution on an interval. In S.H. Sirazudinov (ed.). Limit theorems and statistical inference. Tashkent, FAN.

Naylor, A.F. (1959). Experimental analysis of dispersal in flour beetle *Tribolium confusum*. Ecology 40, 453-465.

Needham, J.G. et al (eds.) (1959). Culture methods for invertebrate animals. Dover Pub. Inc., New York.

Nelson, E. (1958). A functional calculus using singular Laplace integrals. Trans. Amer. Math. Soc. 88, 400-413.

Newman, M.A. (1956). Social behaviour and interspecific competition in two trout species. Physiol. Zool. 29, 64-81.

Nossal, R. and G.H. Weiss (1974). A descriptive theory of cell migration on surfaces. J. Theor. Biol. 47, 103-113.

Novak, B. and F.F. Seelig (1976). A model for aggregation of amoebae: A computer study. J. Theor. Biol. 56, 301-327.

Paloheimo, J.E. (1971). A stochastic theory of search; implications for predator-prey situations. Math. Biosci. 12, 105-132.

Parthasarathy, K.R. (1967). Probability measures on metric spaces. Academic Press, New York.

Philip, J.R. (1957). Sociality and sparse populations. Ecology 38, 107-111.

Radcliff, J. and P.J. Staff (1970). Immigration-migration-death processes with multiple latent roots. Math. Bioscience 8, 279-290.

Revuz, D. (1975). Markov chains. North Holland, Amsterdam.

Richardson, R.H. (1970). Models and analyses of dispersal patterns. In Kenichi Kojima (ed.) Mathematical topics in population genetics. Springer-Verlag, N.Y.

Ringrose, J.R. (1971). Compact non-self adjoint operators. Reinhold, London.

Roff, D.A. (1974a). Spatial heterogeneity and the persistence of populations. Oecologia 15, 245-258.

Roff, D.A. (1974b). The analysis of a population model demonstrating the importance of dispersal in a heterogeneous environment. Oecologia 15, 259-275.

Rosen, G. (1973). Fundamental theoretical aspects of bacterial chemotaxis. J. theor. Biol. 41, 201-208.

Roughgarden, J. (1974). Population dynamics in a spatially varying environment. How population size 'tracks' spatial variation in carrying capacity. Amer. Nat. 108, 649-664.

Ryshik, I.M. and I.S. Gradstein (1963). Tables of series, products and integrals. Veb. Deutscher, Berlin.

Saila, S.B. (1961). A study of flounder movements. Limmnol. Oceanogr. 6, 292-298.

Saila, S.B. and R.A. Shappy (1963). Random movement and orientation in salmon migration. J. Cons. perm. Exp. Mer. 28, 153-166.

Sanderson, G.C. (1966). The study of mammal movements - a review. J. Wildlf. Mgt. 30, 215-235.

Sarmanov, O.V. (1961a). The properties of a two dimensional density defining a stationary Markov process. Dokl. Akad. Nauk. SSSR. 136, 200-201.

Sarmanov, O.V. (1961b). Investigation of stationary Markov processes by the method of eigenfunction expansion. Trudy Mat. Inst. Steklov, 60, 238-261.

Sarmanov, O.V. (1960). Characteristic correlation functions and their applications in the theory of stationary Markov process. Dokl. Akad. Nauk SSSR 132, 564-567.

Saunders, R. (1975). Conservative processes with stochastic rates. J. Appl. Prob. 12, 447-456.

Seckel, R.G. (1972). Hawaiian-caught skipjack tuna and their physical environment. Fishery Bulletin 70, 763-787.

Shiyomi, M. (1967). A model of plant-to-plant movement of aphids. I. Description of the model. Res. Popul. Ecol. 9, 53-61.

Siniff, D.B. and C.R. Jessen (1969). A simulation model of animal movement patterns. Adv. in Ecol. Res. 6, 185-219.

Skellam, J.G. (1951). Random dispersal in theoretical populations. Biometrika 38, 196-218.

Slatkin, M. (1973). Gene flow and selection in a cline. Genetics 75, 733-756.

Southwood, T.R.E. (1971). Ecological methods, with particular reference to the study of insect populations. Chapman and Hall, London.

Southwood, T.R.E. (1962). Migration of terrestial arthropods in relation to habitat. Biol. Rev. (1962) 37, 171-214.

Spiegel, M.R. (1968). Mathematical handbook of formulae and tables. Schaum's outline series. McGraw-Hill, new York.

Strang, G. and G.J. Fix (1973). An analysis of the finite element method. Prentice Hall, Englewood Cliffs, New Jersey.

Stroock, D.W. and S.R.S. Varadhan (1971). Diffusion processes with boundary conditions. Comm. Pure Appl. Math. 24, 147-225.

Stroock, D.W. (1974). Some stochastic processes which arise from a model of the motion of a bacterium. A. Wahr. 28, 305-315.

Tallis, G.M. (1966). A migration model. Biometrics 22, 409-410.

Tyan, S. and J.B. Thomas (1975). Characterisation of a class of bivariate distribution functions. J. Multivariate Anal. 5, 227-235.

Usher, M.B. and M.H. Williamson(1970). A deterministic matrix model for handling the birth, death and migration processes of spatially distributed populations. Biometrics 26, 1-12.

Watson, G.N. (1966). A treatise on the theory of Bessel functions. Cambridge Univ. Press, Cambridge.

Weinberger, H.F. (1974). Variational methods for eigenvalue approximation. Regional conference series in applied mathematics 15, SIAM.

Wells, M.J. and S.K.J. Buckley (1972). Snails and trails. Anim. Beh. 20, 345-355.

Whittle, P. (1968). Equilibrium distributions for an open migration process. J. Appl. Prob. 5, 567-571.

Wierzbowska, Teresa (1975). Review of methods for estimating the parameters of the home range of small forest rodents from the aspect of sample size. Acta theoriol. 20, 3-22.

Wierzbowska, Teresa (1972). Statistical estimation of home range size of small rodents. Ekol. Pol. 20, 781-831.

Wilkinson, D.H. (1952). The random element in bird navigation. J. Exp. Biol. 29, 532-560.

Wilson, A.G.L., R.D. Hughes and N. Gilbert (1972). The response of cotton to pest attack. Bull. Ent. Res. 61, 405-414.

Wilson, D.E. and J.D. ~~Findly~~ FINDLEY (1972). Randomness in bat homing. Amer. Nat. 106, 418-424.

Wright, S. (1969). Evolution and the genetics of populations. Volume II. The theory of gene frequencies. University of Chicago Press, Chicago.

Wright, S. (1968). Evolution and the genetics of populations. Volume I. Genetic and biometrical foundations. University of Chicago Press, Chicago.

Young, E. (1972). Observations on the movement patterns and daily home range size of impala, *Aepyceros melampus* (Lichtenstein) in the Kruger National Park. Zool. Afr. 7, 187-195.