Corrigenda

Items will be refered to in the form page:line.

6:17 ‘affect’ not ‘effect’ (also 21:7 and 27:19).

20:24 should read

$$P(\nu_1' = n_1, \nu_2' = n_2, \ldots, \nu_N' = n_N) = \frac{cN^\alpha!}{N^{cN^\alpha} n_1! n_2! \cdots n_N!} \quad \text{where} \quad \sum_{i=1}^{N} \nu_i' = cN^\alpha,$$

22:5 $1, 2, \ldots, n$.

22:7 ‘$-s$’ not ‘$-t$’.

23:8,9 ‘relations’ not ‘classes’, ‘in case $\xi \subseteq \eta$’ not ‘$\eta \subseteq \xi$’.

23:16 ‘$p_{\xi\eta}$’ not ‘$p_{\eta\xi}$’.

24:7 ‘$\leq$’ not ‘$=$’, N^{k-1} not N_{k-1}.

25:8 ‘(2.1)’ not ‘(2.2)’.

26:15 $E(A) = 1 - \frac{cN^\alpha}{N}$.

31:20 ‘$-t$’ not ‘t’.

44:3 ‘correlated’ not ‘uncorrelated’.

49:13 ‘indefinitely’.

50:2 ‘$(T!)^2$’ not ‘$2(T!)$’.

50:15 ‘Chapter 2’ not 3.

51:21 should read

$$\pi_{\underline{x}_t} P(\underline{X}_{t+1} = \underline{x}_{t+1} \mid \underline{X}_t = \underline{x}_t) = \pi_{\underline{x}_{t+1}} P(\underline{X}_{t+1} = \underline{x}_t \mid \underline{X}_t = \underline{x}_{t+1})$$

51:23 should read

$$P(\underline{X}_{t+1} = \underline{x}_{t+1} \mid \underline{X}_t = \underline{x}_t) = P(\underline{X}_{t+1} = \underline{x}_t \mid \underline{X}_t = \underline{x}_{t+1}).$$

57:19 $= O(N^{-2m-(l-m)}) = O(N^{-(l+m)})$.

58:11 $Q_K \cap \{C_t = e\}$ where $Q_K = \{K_t = K_{t-1}\}$.

59:10 and 63:1 ‘(4.5)’ not ‘(4.6)’.

70:10 remove ‘(4.14)’.

76:3,5 ‘W’ not ‘X’.

99:10 ‘$\{b_k : k \geq 0\}$’ not ‘$(b_k : k \geq 0)$’.

105:16 should read

$$E(\Pi_{i=1}^{k}\phi_i(X_i))\,E(\Pi_{i=1}^{n}\phi_i(X_i)) \quad \leq \quad E((\Pi_{i=1}^{k+l}\phi_i(X_i)))\,E((\Pi_{i=1}^{k}\phi_i(X_i))(\Pi_{i=k+l+1}^{n}\phi_i(X_i)))$$

107:13 ‘$a, b \in \Omega$’ not ‘$a, b \in X$’.

114:5

$$C_2 = \begin{pmatrix} N-2 \\ i-k-1 \end{pmatrix} \begin{pmatrix} N-1 \\ i+k-1 \end{pmatrix}$$

Correlation Structures in Applied Probability

By: Paul Marjoram.

Thesis submitted for the degree of Doctor of Philosophy of the University of London.

University College London.

March 1992.

Abstract

This thesis examines consequences of correlation structure in three areas of applied probability: mathematical population genetics, birth processes, and "exchangeable" measures on distributive lattices. The first three chapters concern probabilistic models in genetics. Initially we generalize the Moran model to allow more than one individual to reproduce per generation, investigating the effect of this on the behaviour of the model. The conclusion is that while things apparently happen faster, the basic properties are the same. This model also serves to unify conventional neutral theory, as it links the Moran model to the Wright-Fisher model.

We then consider aspects of the neutral theory. Commonly a neutral model is supposed in which successive generations behave independently. This may well be unrealistic. Here we take the Moran model and adapt it to allow for correlations in offspring numbers between generations. An analysis of the model shows that the conditional distribution of allele frequencies is unchanged, although the expected number of alleles present decreases. Similar results are also obtained when correlation is introduced to the more general model with more than one reproducer per generation. In each case the approach involves a detailed study of the genealogy of the models.

Next we consider the effect of correlation in Markov Birth Processes. We show that if the birth rates form a super(sub)-linear sequence then the sizes of its families are positively(negatively) correlated. From this we prove a conjecture of Faddy which says that if the birth rates of a process $X(t)$ are super(sub)-linear then the variance ratio $V(t)$ (defined as $\mathrm{Var}X(t)/\left(EX(t)\left[\frac{EX(t)}{X(0)}-1\right]\right)$) is greater than (less than) 1.

Finally we study correlation inequalities. The FKG Inequality is a well known result giving sufficient conditions for positive correlations in probability measures on distributive lattices. There are few analogous results concerning negative correlation. We give sufficient conditions for a particular form of negative correlation when the underlying distributions possess a certain exchangeability property.

Acknowledgements

Much of the work for this thesis was undertaken while the author was receiving a SERC grant. It is also a pleasure to acknowledge the inspiration and assistance of Professor Peter Donnelly.

Contents

Page:

List of Figures

Page:

Chapter 1

INTRODUCTION

This thesis seemingly contains several disparate parts, the chapters concerning themselves variously with mathematical population genetics, birth processes, and inequalities on lattices. However, while it was not the motivation for studying these subjects, a linking theme does emerge: that of 'correlation'. Chambers English Dictionary defines correlation as 'the state or act of being related to one another'. The interest of course depends on what the correlation is acting upon.

In chapter 5 we find probably the most natural interpretation. We are ultimately interested in the possible correlation occurring between the sizes of families whose total number is modelled by a birth process. Can we find sets of transition rates such that the sizes of families are positively (or negatively) correlated? In chapters 2-4 the correlation is of another type. We model the evolution of populations of DNA or genes by adapting a popular standard model to allow for correlation in the choice of parents in consecutive generations. How does this effect the outcome? Finally, in chapter 6, we investigate correlation inequalities. The FKG inequality is a well known result which can be interpreted as a statement of positive correlation. There are very few analogous results for negative correlation in this context, and we present a first step in that direction.

Despite the linking theme just outlined, the subject matter of the chapters varies quite widely. Consequently, there is no attempt here to present a survey of the relevant existing theory or detailed introduction to the mathematics involved: it being felt more natural to do this separately for each subject. We therefore content ourself with an outline of each chapter and the motivation for that which is within.

As stated before, chapters 2-4 all revolve around the subject of mathematical population genetics, and as such, form something of a unit. There is a very large existing body of theory concerning the evolution through time of a population of individuals. We use the term 'individual' loosely here: commonly, what we are referring to is a population of sequences of DNA. DNA can be thought of as a long strand of data which provides essentially all the information necessary in the construction of living material. Particular sections of the DNA molecule correspond to different pieces of information. For instance, one section may define hair colour and another blood group. Alternatively it may be size and shape of petals etc. We call a sequence of DNA which defines a single piece of information (eg. one of the above) a gene, and often it is a population of such genes that is modelled.

Various simplifying assumptions are made for the purposes of such models and we interest ourself with the effects of the relaxation of some of these. Specifically, in chapter 2 we consider the two most popular mathematical models of evolution: the Moran model, in which there is only one offspring (and hence death) per generation (population size being fixed); and the Wright-Fisher model, where the entire population dies and is replaced by offspring between each generation. The conclusions following from these two models are similar when the population size is large, the differences being explained by the effects of time-scaling: in a sense, one model can be viewed as an accelerated version of the other. Consequently we present and analyse a class of models which can be seen as unifying the two apparently different models as opposite ends of a single spectrum. As one might hope, we find that all of this unifying class of models behave in the same way (modulo time-scaling). While this is not directly related to correlation it provides a useful framework for extending the results of chapter 3, as well as adding insight to the models themselves.

In chapter 3 we proceed to relax another of the common assumptions of mathematical models of genetic evolution, (one that is inherent in both the Moran and Wright-Fisher models): the assumption that offspring numbers in different generations are independent. In many circumstances it may be thought reasonable that if a particular individual has a large number of offspring in one generation, it may be likely to have a larger than expected number of offspring in the next generation. In other words, offspring numbers in different generations may be correlated. Clearly there are many possible formulations of a model to allow for such a phenomenon. We propose a simple adaptation of the Moran model (where there is only one parent per generation) which says (informally) that there is a probability p, $(0 \leq p \leq 1)$, that the parent in the next generation will be the same as the parent in the immediately previous generation. We note that this is not intended to represent a situation in which selection is acting although this may indeed have a similar effect. While our approach is clearly a simplistic one, the subsequent analysis suggests that for a wider class of models containing this sort of correlation (cf. chapter 4), the behaviour will remain fundamentally the same. The bulk of this chapter has appeared in Donnelly and Marjoram (1989).

Chapter 4 represents the final installment of mathematical genetics. Essentially it is an amalgam of the ideas in the preceding two chapters, and to some extent indicates how our conclusions about models including correlation may be dependent upon the actual form of the model chosen. We see whether correlation between parents, of the type just outlined, has any effect on some of the broader class of models presented in chapter 2. For those we are able to analyse, we find that the conclusions are substantially the same as before.

For the remaining parts of the thesis we move away from genetics. In chapter 5 we concern ourself with correlation in birth processes and relate this to a conjecture due to Faddy (1990). He notes that if the birth rates $\{\lambda_i\}$ of a birth process form a convex sequence, then we find there is more variation than there would be if the rates were linear (ie. $\lambda_i = i\lambda$ for all i), and conversely there is less variation if the $\{\lambda_i\}$ form a concave sequence. He reports that although computer simulations support this conjecture, no proof has yet been found. We first relate this condition on the variances to a statement about correlations between family sizes (viewing descendants of a particular initial individual as belonging to its 'family'), showing that it is equivalent to the existence of positive or negative correlation respectively. We then present a proof of the conjecture showing the existence of the required correlations via a consideration of the monotonicity of the birth process. Aspects of this chapter feature in Donnelly, Kurtz and Marjoram (1992).

Chapter 6 relates to correlation inequalities. There already exists a famous and much used result relating to aspects of positive correlation of measures on a lattice: namely the FKG inequality. There has been little work on the subject of negative correlation in this context however. Initially we spend a little time to determine what might represent a reasonable definition for such negative correlation by looking at various examples of situations where some sort of negative correlation is (intuitively) clearly occurring. This leads to a natural definition, and from this we proceed to prove a result which may be viewed as a partial converse to the FKG inequality. While there exist related results giving conditions on a measure μ which lead to it having negative correlations of some sort, their interpretation is far from clear. Our result has the natural interpretation that the underlying measure μ corresponds to a distribution on an exchangeable set of random variables.

Finally, some comments regarding organisation. In order to keep the amount of algebra down to manageable levels, and so improve readability, some of the more routine or lengthy mathematical details are relegated to separate appendices. There are three of them, corresponding to work in chapters 3,4 and 6. The numbering scheme for equations herein works as follows: We label definitions theorems etc. with two numbers, the first is the number of the chapter within which it is found, the second is the number of the result itself. Important formulae are numbered in the same fashion, but in a separate sequence.

Notation

$[\cdot] \equiv$ integer part

$|\cdot| \equiv$ the number of elements (or modulus)

$x(N) = O(N^\alpha) \equiv \lim_{N\to\infty} \left|\frac{x(N)}{N^\alpha}\right| \leq c$ $\quad c$ some positive constant

$x(N) = o(N^\alpha) \equiv \lim_{N\to\infty} \left|\frac{x(N)}{N^\alpha}\right| = 0$

$\subseteq \equiv$ inclusion (not strict)

$c \wedge d \equiv$ greatest lower bound of c and d

$c \vee d \equiv$ least upper bound of c and d

$\mathbb{R} \equiv$ the real numbers.

Chapter 2

POPULATION GENETICS MODELS

2.1 A Unifying Class of Genetic Models

The subject of population genetics is an old and well established one and a good tour of it is provided by Ewens (1979). During the course of its development many models of population behaviour have been proposed and investigated, but here we will consider only those models which assume a fixed population size N. This is an assumption which proves not to be as restrictive as it sounds. As long as the changes in population size are independent of genetic composition the analysis goes through with very little modification. Indeed, often we will consider effectively infinite populations anyway. Two models of a fixed size population stand out as being by far the most popular and most analysed. The first is due to S.Wright (1931) and R.A.Fisher (1930), while the second was introduced by P.Moran (1958). Here we will briefly describe both models and their properties and then move on to consider a more general class of models which may help unify their separate theories. A reader requiring a more detailed investigation of these models should consult the aforementioned Ewens (1979), or alternatively Kingman (1982a).

Firstly though, we will outline the further assumptions made by these two models and briefly discuss their import and effect. Note that because of the common applications of this theory we will use the terms ‘genes’ and ‘individuals’ interchangeably, as we will ‘alleles’ and ‘types’.

(i) Haploid/Diploid. We assume the population is haploid rather than diploid. ie. each individual has only one parent and all individuals are of the same sex. This somewhat restricts our field of attention! However things are not as bad as they sound. Many questions of genetic interest can be answered by considering the haploid gametes which transmit information between generations. Alternatively we can consider the mitochondrial DNA present in most cells which is also truly haploid. Thus we avoid overly restricting ourselves, although naturally the development of similar theory for diploid populations would still be of much interest.

(ii) Neutrality. We assume that no type of individual is at a selective advantage. In other words, no individual is more likely to reproduce than another and the distribution of offspring numbers is the same for all genetic types. This is obviously a major assumption. While almost everyone would now agree that some neutral loci exist, the extent of neutrality has been a cause of hot debate for many years now (see Ewens (1977) or Kimura (1983) for example). We have little to say on the matter here, but hope to add to the discussion in the two chapters following this one by introducing models which weaken the assumption of neutrality.

(iii) Exchangeability. This is closely related to neutrality. We assume that all individuals are probabilistically the same and that any labelling/ordering we care to give them is of no significance in the analysis. In particular, we assume

$$P(\nu_i = n_i; i = 1,2,\ldots,n) = P(\nu_i = n_i'; i = 1,2,\ldots,n)$$

where the $\{n_i'\}$ are any permutation of the $\{n_i\}$ and ν_i is the number of offspring of the ith individual in our population.

(iv) Single locus. When considering genetic populations we commonly consider only single loci. For example, if we are modelling a sequence of DNA we consider only one gene or alternatively one base (the simplest unit of genetic information). It was common to then allow this unit to be one of a finite number of types. It is easy to see that this is not very restrictive at all. If for example we model the behaviour of a population of single bases of DNA we have four possible types (corresponding to the four base types). To model a population of sequences of DNA of length l we can simply consider this to be a collection of (single locus) individuals which each has 4^l possible types and conduct the analysis as before. Clearly for large l we quite quickly begin to have an unmanageable number of possible types and this has led to the introduction of 'infinite alleles' models, which unsurprisingly allow an individual to be any one of an infinite number of types. These models have proved of much use recently.

A further consideration is that of linkage. This is the phenomenon whereby the fate of one allele is affected by that of a physically nearby allele which may be at a selective advantage say. This may cause the first allele to be more successful than might otherwise be expected and so its frequency would tend to be higher than might ordinarily have been the case. This requires separate treatment and we investigate it further in subsequent chapters.

(v) Spatial effects. We will assume there are none. So there is no tendency for related individuals to be found geographically closer to one another. This is of no consequence unless spatial considerations are likely to be particularly important, in which case the sort of models discussed here are not really the right ones to be using anyway.

So, given the assumptions just outlined, we now detail the two models most often used.

The Wright-Fisher model assumes the existence of discrete non-overlapping generations $G_0, G_1, G_2, \ldots$ which each contain a fixed number of individuals N. The members of generation G_{r+1} are the offspring of the members of G_r where, as before, the number of offspring of the jth member of G_r is a random variable ν_j, where $\sum_{j=1}^{N} \nu_j = N$. The ν_j are assumed to have a symmetric multinomial distribution with the behaviour of successive generations being independent, so

$$P(\nu_j = n_j; (j = 1, 2, \ldots, N)) = \frac{N!}{n_1! n_2! \cdots n_N! N^N} \qquad \begin{array}{l} n_1 + \cdots + n_N = N \\ n_j = 0, 1, \cdots, N \\ j = 1, 2, \cdots, N \end{array}$$

Notice that this means the model has a particularly appealing structure when looked at with time running backwards: each member of G_{r+1} simply chooses a parent at random, independently and uniformly, from the N individuals of G_r.

In contrast, the Moran model posits that we have a population of size N evolving through continuous time (say) where generational events occurs as points of a Poisson process of rate 1. At each of these generational time-points, one individual only is chosen uniformly and at random to die and simultaneously another individual (also chosen at random, and independently from the first choice) gives birth to a single offspring, (all choices in one generation being independent of events in other generations). This second (parent) individual, may or may not be allowed to be the first (dying) individual, according to the context of the model, but the subsequent analysis and results are substantially the same in either case. The Moran model also has a convenient structure when looked at backwards in time.

Both of these two models can be set in either discrete or continuous time. In the first case t simply counts the generations (which are labelled 0,1,2,...), and in the latter t is continuous and generational events happen at rate 1. The analysis is fundamentally the same in either setting.

Essentially these models are concerned simply with reproduction. However most analysis is performed when we have the additional concern of random mutations affecting offspring individuals. The early versions of such models allowed an offspring gene to be one of only a finite collection of allelic types and work was then undertaken to find the stationary or transient distribution for the frequency of a particular allele. However subsequent advances in biological theory led to the realisation that it would be more realistic to allow for an infinite number of allelic types. So now, when a parent gives birth to an offspring, there is a probability u that the offspring mutates to a completely new type never before seen in the population, otherwise with probability $(1-u)$ its type is the same as that of its parent.

These new types of model (ie. the infinite alleles models) differ markedly from the others in that eventually any particular allelic type will vanish from the population, to be replaced by new types which have arrived via mutation. Hence, although we are still interested in behaviour at stationarity, there are no longer any stationary distributions for the frequency of a particular allelic type. So instead we concentrate on investigating allelic patterns: the distribution of the number of different allelic types present at a particular moment in time, and the number of individuals of each type found.

This sort of analysis was initially performed by looking at the development of the models forward in time. Indeed the Ewens Sampling Formula was derived at length in such a manner. However analysis is much quicker and easier when one studies the structure of the family tree of the process backwards through time, and this is how much current theory is obtained. In what follows we will continually use ideas and terminology from the recently popularized genealogical approach to neutral models. For detailed background the reader is refered to Tavaré (1984), Donnelly and Tavaré (1987), or Kingman (1982a). For our purposes the following outline will suffice.

Initially we suppose no mutation is occurring and take a sample of size n say from a population at equilibrium. We arbitrarily label the generation from which we have sampled to be generation 0 (ie. G_0). We now define equivalence relations $\mathcal{R}_f$ on the set $\{1, 2, \ldots, n\}$ as follows:

$$\mathcal{R}_0 = \{(i, i); i = 1, 2, \ldots, n\}$$

and $\mathcal{R}_s$ contains the pair (i, j) if the ith and jth individuals in our sample share a common ancestor in G_{-s}. So

$$\mathcal{R}_s \subseteq \mathcal{R}_{s+1}.$$

Thus, for either of the two models mentioned previously, the sequence $(\mathcal{R}_s)$ is a Markov Chain which describes the family tree of our sample going backward through time. Each equivalence class in $\mathcal{R}_s$ corresponds to an individual in $\mathcal{G}_{-s}$ and changes of state in $(\mathcal{R}_s)$ occur when two equivalence classes are combined, ultimately terminating at time T (say) as

$$\mathcal{R}_T = \{(i, j) : \ i, j = 1, 2, \ldots, n\}$$

Thus far the description applies to either model but they now diverge slightly. For the Wright-Fisher model we define a new process R_t by changing the time-scale of $\mathcal{R}$, viz;

$$R_t = \mathcal{R}_{[Nt]},$$

(where $[\cdot]$ represents the integer part). We now find that, in the limit as $N \to \infty$, R_t converges in distribution to a process called the n-coalescent (which will be described later).

For the Moran model we need a slightly different time-scaling, so we define

$$R_t' = \mathcal{R}_{N^2 t/2}.$$

Here the result is exact: R_t' is an n-coalescent for all N.

The formal definition of an n-coalescent is as follows. An n-coalescent is a Markov chain with statespace E_n, the set of all equivalence relations on $\{1, 2, \ldots, n\}$. It starts in state $\{(i,i);\ i = 1, 2, \ldots, n\}$ and terminates in the absorbing state $\{(i,j);\ i,j = 1, 2, \ldots, n\}$. The transition rates are

$$q_{\xi\eta} = \begin{cases} -\frac{1}{2}k(k-1) & \text{if } \xi = \eta \text{ and } k = \mid \xi \mid \\ 1 & \text{if } \xi \prec \eta \\ 0 & \text{otherwise} \end{cases}$$

where $\xi \prec \eta$ denotes that η is formed from ξ by amalgamating two of its equivalence classes. Many properties of the n-coalescent are known, and for more details of these the reader is referred to Kingman (1982a) or Kingman (1982b).

If we now suppose there is a non-zero probability u that an offspring, when born, is a mutant (where u is $O(N^{-1})$ in order to achieve a balance between mutation and random genetic drift), we obtain similar behaviour. An n-coalescent still arises but now random mutations occur among the equivalence classes of the coalescent as it evolves. Results are possible for many sorts of conceivable mutation structures, but matters simplify somewhat if we suppose that each mutation is to a unique new allelic type. We proceed as follows:

As before we take a sample of n from a population of size N and define an equivalence class valued Markov process R_s in an analogous manner. Specifically we declare two individuals to be in the same *old* equivalence class at time s if they share a common ancestor s generations into the past and there have been no intervening mutations. We label the old equivalence classes as $\xi_1, \xi_2, \ldots, \xi_{D_s}$, where D_s is the number of old equivalence classes at time s. Next we declare two individuals to be in the same *new* equivalence class at time s (counting time in discrete generational units) if, for some r with $1 \leq r < s$, they have a common ancestor r generations ago, where this ancestor is itself a mutant, and there has been no subsequent intervening mutation. We label the new equivalence classes as $\eta_1, \eta_2, \ldots, \eta_{F_s}$, where F_s is the number of *new* equivalence classes at time s. This description follows Donnelly and Tavaré (1986), where they show that it is now also possible to keep track of the ages of the types of individuals with a coalescent and subsequently derive many useful results.

In fact Kingman, and Donnelly and Tavaré, show that it is true that for a broad class of exchangeable models, where family sizes in different generations are independent, we get convergence to the n-coalescent. This result does not apply to the Moran model however, but a similar result can be proved using different methods, (a different time-scaling is needed).

It is worth noting at this point that if two models for reproduction have the same genealogical structure, then the processes of *genetic* interest which count gene frequencies or partitions will behave similarly (Donnelly (1985), Donnelly and Tavaré (1987)). This is reflected in the fact that with suitable normalization and the above time scalings, both of the aforementioned models converge (forward in time) to the same diffusions (Ewens (1979)). This result and other related diffusion results are a consequence of genealogical robustness.

So we see that in some sense, in the limit, the Wright-Fisher model is like a speeded up Moran model. But in the Wright-Fisher model there are N offspring per generation, whereas in the Moran model there is only 1 . So what would happen if 2 or 3 individuals were born per generation? Or if some proportion of the population size N^α were born? Do we get effectively the same behaviour with a gradual transition of the time-scaling as the number of offspring increases, or does the population behave in an entirely different fashion? Also, will we still be able to use genealogical approaches to derive results? Finally, can we prescribe a natural class of models which includes and explains the behaviour of both the Wright-Fisher and Moran models as particular cases of a unifying whole rather than having, as at present, to consider them as two separate models which have similar properties? It is the purpose of this chapter to investigate exactly these questions.

2.2 Model Specification and Analysis

In this section we will consider models of a population of size N where between 1 and N offspring are born at a time. We could state the model in either discrete or continuous time without fundamentally changing its properties.

We will set the model in continuous time by supposing that generational events occur at rate 1 in a population of fixed size N (the individuals of which we arbitrarily label 1,2,...N for convenience). When such an event occurs cN^α, $(0 \leq \alpha \leq 1,\ c > 0)$ offspring are born, and each independently picks a parent uniformly and at random from the existing population. Hence if the number of offspring born to individual i is a random variable ν_i', $i = 1, 2, \ldots, N$ then the joint distribution of offspring numbers is as follows:

$$P(\nu_1 = n_1', \nu_2' = n_2, \ldots, \nu_N' = n_N) = \frac{cN^\alpha!}{N^{cN_\alpha} n_1! n_2! \cdots n_N!}, \text{ where } \sum_{i=1}^{N} \nu_i' = cN^\alpha,$$

and 0 otherwise. Note that we are assuming here that c and α have been chosen so that cN^α is a positive integer less than or equal to N. This is merely of notational convenience. Note also that the ν_i' are exchangeable and independent of offspring numbers in other generations. The offspring may be subject to mutation, (each offspring mutating to a unique new type, independently and with probability u), but this aspect of the model does not effect our immediate analysis. After generating the offspring, we then independently pick (uniformly and at random) cN^α members of the previously existing individuals to die. Thus the population is of a constant size.

Note that there is an alternative construction of this model, and it is this version we will use in the analysis. In this version we break down the generational event into three steps:

Step 1. The cN^α offspring choose their parents, independently, uniformly, and at random, from the original population.

Step 2. $N - cN^\alpha$ of the original population are chosen, uniformly, without replacement, to each have an additional offspring (independently of step 1), which is an identical copy of itself, (ie. not subject to mutation).

Step 3. All of the original population of N die.

This version clearly gives us a model with exactly the same behaviour as before, but the offspring number distribution has changed We denote these new offspring numbers by $\{\nu_i\}$ and note that

$$\sum_{i=1}^{N} \nu_i = N.$$

Again the ν_i are exchangeable and independent of offspring numbers in other generations.

2.2.1 The Analysis

To analyse this model we shall look at the genealogy of the process. Firstly we sample n individuals at random from the population in equilibrium. We label this generation as generation 0. We now define an equivalence relation R_s on the set of integers $1, 2, \ldots, N$ which contains the pair (i, j) if and only if the ith and jth individuals in the sample have a common ancestor in generation $-t$. So (R_s) tracks the ancestry back through time (ie. the genealogy). Note that $R_0 = \{(i,i) : i = 1, 2, \ldots, n\}$ and $R_s \subseteq R_{s+1}$. We then seek the conditions under which the Markov process $(R_s : s = 0, 1, 2, \ldots)$ converges (after time scaling) to the n-coalescent.

This approach closely follows the analysis of the general class of exchangeable models in (Kingman (1982c)), so we start by proving an extension of his Theorem 1.

Theorem 2.1 *Suppose that as the population size N tends to infinity, the variance σ^2 of ν_i (the offspring number distribution) converges in the following way*

$$\lim_{N\to\infty} \sigma^2 N^{f(\alpha)} = S^2 \ (0 \le \alpha \le 1, \ \textit{some } S > 0)$$

(where $f(\alpha)$ is some function of α) and that

$$\sup_N E(\nu_1^k) < \infty \ \textit{for all } k \ge 1$$

Then the finite dimensional distributions of the process

$$R_t = \mathcal{R}_{[N^{1+f(\alpha)}S^{-2}t]}$$

converge to those of the n-coalescent.

Note that the fact that the ν_i are exchangeable means that the variance of ν_i

is identical for all i (denote it by σ) and so we can write

$$N^{f(\alpha)}\sigma_N^2 = \lim_{N\to\infty} N^{f(\alpha)}\sigma^2 = S^2.$$

Hence we can write R_t as

$$R_t = \mathcal{R}_{[N\sigma_N^{-2}t]}.$$

Proof:

We follow the logic of Kingman's proof. Firstly we need to find

$$p_{\xi\eta} = P(R_{s+1} = \eta \mid R_s = \xi) \text{ for } \xi, \eta \in E_n$$

where E_n is the set of equivalence classes on $\{1, 2, \ldots, n\}$. This probability is zero except for the case $\eta \subseteq \xi$ when we label the equivalence classes in η as $C_\gamma\,(\gamma = 1, 2, \ldots, a)$ and those in ξ as

$$C_{\gamma\beta}\,(\gamma = 1, 2, \ldots, a;\ \beta = 1, 2, \ldots, b_\gamma)$$

where

$$C_\gamma = \bigcup_{\beta=1}^{b_\gamma} C_{\gamma\beta}$$

and the $C_{\gamma\beta}$ correspond to

$$k = \sum_{\gamma=1}^{a} b_\gamma$$

particular individuals in the offspring generation. $p_{\eta\xi}$ is the probability that if k individuals are selected at random from this generation and labelled as $C_{\gamma\beta}$, then all the $C_{\gamma\beta}$ for each fixed γ have the same parent, and that these parents for different values of γ are distinct. This leads to the conclusion that

$$p_{\xi\eta} = E[(N)_k^{-1} \textstyle\sum (\nu_{j_1})_{b_1} (\nu_{j_2})_{b_2} \cdots (\nu_{j_a})_{b_a}] \qquad (2.1)$$

when $\xi \subseteq \eta$, and zero otherwise, and we define

$$(N)_k = N(N-1)\cdots(N-k+1)$$

and the summation is over all distinct $j_1, j_2, \ldots, j_a$ in $1 \leq j_i \leq N$.

Considering first the case where η is formed from ξ by amalgamating two of its equivalence classes (written $\xi \prec \eta$) we get an upper bound on $p_{\xi\eta}$

$$\begin{aligned} p_{\xi\eta} &\leq E\left\{(N)_k^{-1} N^{k-2} \sum_{i=1}^{N} \nu_i(\nu_i - 1)\right\} \\ &= (N)_k^{-1} N^{k-1} E\{\nu_1(\nu_1 - 1)\}. \end{aligned}$$

Hence

$$\begin{aligned} p_{\xi\eta} &= \frac{S^2}{N^{f(\alpha)}} \frac{N_{k-1}}{N(N-1)\cdots(N-k+1)} \\ &= \frac{S^2}{N^{f(\alpha)}} \frac{N^{k-2}}{N^{k-1} + c_{k-2}N^{k-2} + \cdots + c_1 N + c_0} \end{aligned}$$

for constants $c_{k-2}, c_{k-3}, \ldots, c_0$

(in fact, the c_i are Stirling numbers.)

$$\begin{aligned} &= \frac{S^2/N^{f(\alpha)}}{N + c_{k-2} + c_{k-3}N^{-1} + c_{k-4}N^{-2} + \ldots} \\ &= N^{-1}\frac{S^2}{N^{f(\alpha)}} + \frac{S^2}{N^{f(\alpha)}} O(N^{-2}) \\ &= \frac{S^2}{N^{1+f(\alpha)}} + O(N^{-(2+f(\alpha))}). \qquad (2.2) \end{aligned}$$

We also get the lower bound

$$p_{\xi\eta} \geq N^{-1}E\{\nu_1(\nu_1 - 1)\} - (k-2)N^{-2}E(\nu_1^3) - \binom{k-2}{2} N^{-2}E\{\nu_1\nu_2\},$$

and so we have

$$p_{\xi\eta} \geq N^{-1}\frac{S^2}{N^{f(\alpha)}} + \frac{S^2}{N^{f(\alpha)}} O(N^{-2})$$

$$= \frac{S^2}{N^{1+f(\alpha)}} + O(N^{-(2+f(\alpha))}) \qquad (2.3)$$

asymptotically.

So combining (2.2) and (2.3) we get

$$p_{\xi\eta} = \frac{S^2}{N^{1+f(\alpha)}} + O(N^{-(2+f(\alpha))})\ ),$$

$$(\text{or alternatively, } p_{\xi\eta} = N^{-1}\sigma_N^2 + O(N^{-(2+f(\alpha))})),$$

for $\xi \prec \eta$.

Now consider the case where $\xi < \eta$ but $\xi \prec \eta$ is false. We have $a \leq k-2$ and so (2.2) gives

$$p_{\xi\eta} \leq (N)_k^{-1}(N)_a E\{\nu_1^k\}$$

as before. Thus we see that

$$p_{\xi\eta} = O(N^{-(2+f(\alpha))})$$

And so, since we can write $P_N = I + c\Omega + o(N^{1+f(\alpha)})$, we conclude in a similar manner to Kingman that the stochastic matrix $P_N = (P_{\xi\eta})$ satisfies

$$\lim_{N\to\infty} P_N^{[N\sigma_N^{-2}t]} = e^{t\Omega}$$

for all $t \geq 0$, where Ω is the infinitesimal generator of the n-coalescent. □

Note that we can now apply this result to our model with cN^α offspring per generation so conclude that it behaves as an n-coalescent in the limit provided we scale time via the map

$$t \mapsto N\sigma_N^{-2}t$$

and show that the following two properties are satisfied:

(i) $\sup_N E(\nu^K) < \infty$, for all $K \geq 1$;

(ii) $\lim_{N\to\infty} \sigma^2 N^{f(\alpha)} = S^2$, for some $S > 0$).

To show (i), let $\eta_1, \eta_2, \ldots, \eta_N$ be the offspring numbers for a Wright-Fisher model with population size N. Then it is known (see Kingman (1982c) for example), that

$$\sup_N E\eta_i^K < \infty \text{ for all } K \geq 1.$$

There is an obvious coupling of the processes for which $\nu_i - 1 \leq \eta_i$ and hence it follows that

$$E(\nu_i^K) \leq E((\eta_i + 1)^K) < \infty,$$

as required.

To show (ii), note that

$$\mathrm{Var}(\nu_i) = \mathrm{Var}(A) + \mathrm{Var}(B),$$

where A and B are the following random variables:

$$A = \begin{cases} 0 & \text{with prob. } cN^\alpha/N \\ 1 & \text{with prob. } \frac{N-cN^\alpha}{N}; \end{cases}$$

$$B \sim \mathrm{Binomial}(cN^\alpha, \frac{1}{N}).$$

Now $E(A) = cN^\alpha/N$ and $E(A^2) = (1 - \frac{cN^\alpha}{N})$ gives:

$$\begin{aligned} \mathrm{Var}(A) &= \left(1 - \frac{cN^\alpha}{N}\right) - \left(1 - \frac{cN^\alpha}{N}\right)^2 \\ &= \frac{c}{N^{1-\alpha}} - \frac{c^2}{N^{2(1-\alpha)}}. \end{aligned}$$

And

$$\mathrm{Var}(B) = cN^\alpha \frac{1}{N}\frac{N-1}{N}.$$

So

$$\mathrm{Var}(\nu_i) = \frac{2c}{N^{(1-\alpha)}} - \frac{c^2}{N^{2(1-\alpha)}} - \frac{c}{N^{2-\alpha}}.$$

Thus, if $\alpha < 1$ we define $f(\alpha) = 1 - \alpha$, to get $S^2 = 2c$, and if $\alpha = 1$ we define $f(\alpha) = 0$, (recall we must have $c \leq 1$,) to get $S^2 = c(2-c)$. In either case, we have satisfied property (ii), and so we can apply the theorem to our model to conclude that we do indeed get convergence to the n-coalescent, as $N \to \infty$.

Therefore we conclude that the genealogy of this new class of model behaves, asymptotically as $N \to \infty$, as an n-coalescent. We note that if we set $c = 1$ and $\alpha = 1$, our model is in fact the Wright-Fisher model. Conversely, if we set $c = 1$ and $\alpha = 0$ we have the Moran model, (in the form which allows the parent of two offspring to itself die), although the general result is true for the Moran model for any N. Other values of c and α give a variety of models which may be thought of as intermediate, and show that their behaviour is as one might have hoped.

Note that in most models of this type it is common to include mutation effects (cf. chapters 3 and 4). The usual form for this is to introduce a probability θ that any given offspring is of a new unique type (independently for each offspring), otherwise the offspring is of the same type as its parent. While we have not considered this here, we note that the action of mutation operates independently of the subjects studied here and so the results herein will be uneffected if mutation is introduced.

Chapter 3

MORAN MODEL WITH CORRELATION

3.1 Introduction

Two of the features common to all neutral models of reproduction are the fact that in a particular generation the number of offspring born to a given individual is independent of the genetic type of the individual, and the fact that the offspring numbers in different generations are independent. Indeed, the existence of such a complete neutral theory, and in particular the recent success in studying such models via their genealogy, depends crucially on the symmetry which results from these two assumptions. There has been considerable recent interest in comparing the predictions of neutrality with those which might be appropriate for models in which selection is acting, with a view to using data to test the applicability of the neutral theory. (See, for example, Watterson (1977) or Ewens (1990).) The introduction of selection destroys the symmetry of neutrality. Neither of the above assumptions obtain, and analysis of the model becomes substantially more difficult.

Our purpose here is to study a model in which the second assumption is relaxed while the first is retained. Imagine following the genes at a selectively neutral locus which is in linkage disequilibrium with a locus at which selection is acting. Individuals with a large number of offspring in one generation may have been so lucky because they possess a particular gene at the selected locus and consequently they (or their offspring) will have a tendency to have more offspring (than average) in the following generation. If the existence of the second (selected) locus is unknown, the effect of watching the neutral locus will be to observe correlations in offspring numbers in successive generations, while the offspring numbers themselves are still independent of the actual allele at the neutral locus. Other situations, for example varying environmental conditions acting either on the "neutral" locus itself or an associated locus, may have the same kind of effect. We must stress that we are not suggesting that exactly these scenarios will give rise to correlation in precisely the form we discuss, but rather that possible scenarios might exist, and so the phenomenon of inter-generational correlation in reproductive mechanisms warrants consideration.

Our principle interest is in examining the effect of this sort of departure from the usual assumptions of neutrality on the testing and estimation procedures which have been developed and applied in this context. The approach, as in recent work throughout the neutral setting, is to focus attention on the genealogy induced by the model. The existence of selection has proved a major stumbling block for genealogical methods. A further novelty of the current study, then, is that the use of genealogy, in spite of the relaxation of one of the key assumptions, may represent a (small) first step towards the solution of more general problems.

First of all we introduce the model we will be considering, (a variant of the

Moran model), and proceed to examine the line of descent process and the distribution of the number of alleles in a sample taken from the population at equilibrium. We then extend this to a study of the distribution of allele frequencies, in age order, in the sample. Finally we discuss the robustness of the results and the consequences for the use of statistical procedures appropriate to the neutral theory.

3.2 The Model

Consider a population of fixed size N haploid individuals (or genes) evolving through discrete generations. In each generation one individual is chosen to have a single offspring and one individual is chosen at random (from the remaining $N-1$ individuals) to die. The population in the next generation consists of the surviving individuals and the new offspring. We introduce correlation by supposing that the individual chosen to reproduce will be the reproducing individual in the previous generation with probability p, $0 \leq p < 1$, and otherwise will be chosen at random from among all the N members of the population. Formally we define a Markov chain $\{X_r : r = 0, 1, \ldots\}$, with statespace $E_N \times E_N$, where $E_N \equiv \{1, 2, \ldots, N\}$, and transition probabilities

$$P(X_{r+1} = (i', j) \mid X_r = (i, k))$$

$$= \begin{cases} (p + (1-p)/N)/(N-1) & i', j \in E_N, \quad i' = i, \quad j \neq i' \\ (1-p)/\big(N(N-1)\big) & i', j \in E_N, \quad i' \neq i, \quad j \neq i' \\ 0 & \text{otherwise} \end{cases}$$

and the initial distribution

$$P(X_0 = (i, j)) = 1/\big(N(N-1)\big), \quad i, j \in E_N, \; i \neq j.$$

Label the individuals in the initial generation from the set E_N and give offspring the label of the individual they replace. The interpretation is that if $X_r = (i, j)$, then in generation r the individual labelled i will reproduce and the individual labelled j will die. (The assumption that these two individuals are distinct is common to many versions of the model, but is of no consequence.) The offspring individual will be of the same allelic type as the parent with probability $1 - u$, or with probability u (> 0) will be of a novel type, never before seen in the population. Our process is thus a modified version of the discrete-time, infinite alleles, Moran model. (Note that the case $p = 0$ corresponds exactly to the Moran model.)

Our interest lies in studying the genetic composition of samples taken from the population at equilibrium. In this section we focus attention on $L(t)$, the number of lines of descent of the sample: $L(t)$ is the number of individuals, t generations into the past, who have descendants, without intervening mutations, in the sample. In the uncorrelated case the behaviour of $L(t)$ is well understood (Griffiths (1980); Tavaré (1984)): call the generation in which the sample was taken 0, and label the preceding generations $-1, -2, \ldots$; the process $L(t)$ behaves as a death process (set in discrete time in this context) with lines of descent being lost either because two of the individuals in generation t with non-mutant descendants in the sample (such individuals will be said to be in the line of descent at time t) share a common ancestor in generation $-t - 1$, or because one of these individuals is the mutant offspring of an individual in generation $-t - 1$.

The introduction of correlation means that the process $L(t)$ is no longer Markov: however, when it changes it will still decrease by exactly one, for one of the two reasons above. The probabilities of the above events, though, will now depend on whether or not the reproducing individual in generation $-t$

was in the line of descent. Consider instead the process

$$\Lambda(t) = (L(t), \alpha(t));\ t = 0, 1, 2, \dots$$

where $L(t)$ is the number of lines of descent of the sample at time $-t$ and $\alpha(t) = 1$ if the reproducing individual in generation $-t$ is in the line of descent and 0 otherwise.

Before proceeding we note the chain $\{X_r, r = 0, 1, 2, \dots\}$ is time reversible. It is easy to check that the equilibrium distribution places mass

$$\Pi_{(i,j)} = 1/N(N-1)$$

on each pair $(i, j) \in E_N \times E_N$ with $i \neq j$, and hence that, with the given initial distribution, the chain is stationary. Reversibility is guaranteed (Kelly (1979)) by the fact that

$$\Pi_{(i,j)} P(X_{r+1} = (k,l) \mid X_r = (i,j)) = \Pi_{(k,l)} P(X_{r+1} = (i,j) \mid X_r = (k,l))$$

for each pair $(i, j), (k, l) \in E_N \times E_N$, which follows easily in this case, since

$$P(X_{r+1} = (k,l) \mid X_r = (i,j)) = P(X_{r+1} = (i,j) \mid X_r = (k,l)).$$

As a consequence, the behaviour of the process when viewed backwards in time is (stochastically) equivalent to its behaviour forward in time. In particular, when viewed backwards in time the sequence of labels of reproducing individuals is Markov, and

$$P(i \text{ reproduces in generation } r \mid j \text{ reproduces in generation } r+1)$$

$$= \begin{cases} p + (1-p)/N, & \text{if } i = j \\ (1-p)/N, & \text{if } i \neq j. \end{cases}$$

It is thus evident that the process $\{\Lambda(t) : t = 0, 1, 2, \dots\}$ is Markov and that the transition probabilities follow from the definition of the model. Suppose that $\Lambda(t) = (k, 1)$ and recall that this has the interpretation that there are k lines of descent at t (that is, exactly k individuals in generation $-t$ with non-mutant descendants in the sample) and that one of these is the reproducing individual in that generation. With probability $p + (1 - p)/N$ this same individual will have been the one who reproduced in the previous generation $(-t - 1)$, in which case a line of descent will be lost if the individual (in generation $-t$) who is chosen to die is one of the remaining $k-1$ in the line of descent (regardless of whether or not the offspring is a mutant), and not lost if it is one of the $N-k$ individuals outside the line of descent. Conditional on this choice of individual to reproduce, these two possibilities have probability $(k-1)/(N-1)$ and $(N-k)/(N-1)$, respectively. In either case, $\alpha(t+1) = 1$. If the reproducing individual in generation $-t-1$ is different from that in generation $-t$, then we may still have $\Lambda(t+1) = (k-1, 1)$, if (and only if) the reproducing individual in generation $-t-1$ and the individual in generation $-t$ who is to die, both belong to the line of descent at t. Thus

$$\begin{aligned} &P(\Lambda(t+1) = (k-1,1) \mid \Lambda(t) = (k,1)) \\ &= (p + (1-p)/N)\frac{k-1}{N-1} + (1-p)\frac{k-1}{N}\frac{k-1}{N-1} \\ &= p\frac{k-1}{N-1} + (1-p)\frac{k}{N}\frac{k-1}{N-1}. \end{aligned}$$

With probability $(1-p)(N-k)/N$, the reproducing individual will be outside the line of descent at $t+1$ (so $\alpha(t+1) = 0$) in which case a line of descent will be lost if the individual chosen to die at $-t$ is one of the k in the line of descent at t *and* the offspring is a mutant, and not lost otherwise. Similar arguments in the case $\Lambda(t) = (k, 0)$ give the other transition probabilities, which for convenience we now collect together and rename:

$$
\begin{aligned}
p_1(k) &\equiv P(\Lambda(t+1)=(k-1,1) \mid \Lambda(t)=(k,1)) = p\frac{k-1}{N-1}+(1-p)\frac{k}{N}\frac{k-1}{N-1} \\
p_2(k) &\equiv P(\Lambda(t+1)=(k-1,0) \mid \Lambda(t)=(k,1)) = (1-p)\frac{N-k}{N}\frac{k}{N-1}u \\
p_3(k) &\equiv P(\Lambda(t+1)=(k,0) \mid \Lambda(t)=(k,1)) = (1-p)\frac{N-k}{N}\frac{N-k-1}{N-1} \\
p_4(k) &\equiv P(\Lambda(t+1)=(k,1) \mid \Lambda(t)=(k,1)) \\
&= p\frac{N-k}{N-1}+(1-p)\frac{N-k}{N}\frac{k}{N-1}(1-u)+(1-p)\frac{k}{N}\frac{N-k}{N-1} \qquad (3.1) \\
p_5(k) &\equiv P(\Lambda(t+1)=(k-1,1) \mid \Lambda(t)=(k,0)) = (1-p)\frac{k}{N}\frac{k-1}{N-1} \\
p_6(k) &\equiv P(\Lambda(t+1)=(k-1,0) \mid \Lambda(t)=(k,0)) \\
&= p\frac{k}{N-1}u+(1-p)\frac{N-k}{N}\frac{k}{N-1}u \\
p_7(k) &\equiv P(\Lambda(t+1)=(k,0) \mid \Lambda(t)=(k,0)) \\
&= p\frac{N-k-1}{N-1}+(1-p)\frac{N-k}{N}\frac{N-k-1}{N-1} \\
p_8(k) &\equiv P(\Lambda(t+1)=(k,1) \mid \Lambda(t)=(k,0)) \\
&= (1-p)\frac{k}{N}\frac{N-k}{N-1}+p\frac{k}{N-1}(1-u)+(1-p)\frac{N-k}{N}\frac{k}{N-1}(1-u).
\end{aligned}
$$

All other transition probabilities are zero, and

$$P(\Lambda(0)=(n,1)) \;=\; 1-P(\Lambda(0)=(n,0)) \;=\; n/N.$$

The number of lines of descent will decrease from $L(0)=n$ to $L(\infty)=0$, in steps of one. Each of these steps will be the result of either a coalescence (sharing of ancestors) or a mutation. Focus attention on K, the (random) number of steps which were due to a mutation. by standard genealogical arguments (Tavaré (1984) for example), K will be the number of different alleles in the original sample.

We now ask about the conditional probability that in the transition from k to $k-1$ lines of descent, the line will be lost through mutation or alternatively through coalescence. It follows from the description of the model, and in particular from arguments similar to those given above, that the state $(k-1,1)$ (rather than $(k-1,0)$) will be first entered from the states $(k,1)$ or $(k,0)$ exactly because both the individual chosen to die and the reproducing individual in the previous generation belong to the line of descent. Furthermore, the loss of a line of descent will be due to a mutation when the offspring is a mutant and to coalescence otherwise. For the state $(k-1,0)$ to be entered first, a mutation must have occurred. Thus in an obvious notation,

$$P(\text{line lost through mutation} \mid (k,1) \rightarrow (k-1,1))$$

$$= \frac{p\frac{k-1}{N-1}u + (1-p)\frac{k}{N}\frac{k-1}{N-1}u}{p\frac{k-1}{N-1} + (1-p)\frac{k}{N}\frac{k-1}{N-1}} = u \tag{3.2}$$

$$\begin{aligned} P(\text{line lost through mutation} \mid (k,0) \rightarrow (k-1,1)) &= \frac{(1-p)\frac{k}{N}\frac{k-1}{N-1}u}{(1-p)\frac{k}{N}\frac{k-1}{N-1}} \\ &= u \qquad (3.3) \end{aligned}$$

$$P(\text{line lost through mutation} \mid (k,1) \rightarrow (k-1,0)) = 1 \tag{3.4}$$

$$P(\text{line lost through mutation} \mid (k,0) \rightarrow (k-1,0)) = 1 \tag{3.5}$$

Furthermore, the Markov structure of the process $\Lambda(\cdot)$ guarantees that these conditional probabilities are independent of the values of $\Lambda(\cdot)$ other than those immediately before a line of descent is lost. In particular they are independent of the events associated with the loss of the preceding $(n-k)$ lines of descent.

Consider the embedded process $\{\alpha^*(k) : k = n, n-1, \ldots, 0\}$ defined by $\alpha^*(n) = \alpha(0)$, and for $k = n, n-1, n-2, \ldots, 0$,

$$\alpha^*(k) = \alpha(\tau_k)$$

where

$$\tau_k = \inf\{t : L(t) = k\}$$

is the time at which the number of lines of descent in the sample first becomes equal to k. The sequence $\{\alpha^*(n), \alpha^*(n-1), \ldots, \alpha^*(0)\}$ inherits the Markov property from $\Lambda(t)$, but its transition probabilities are not time-homogeneous. Denote by W the number of zeros in the sequence $\{\alpha^*(n-1), \alpha^*(n-2), \ldots, \alpha^*(0)\}$. It follows from (3.2) to (3.5) that we can write

$$K = W + Y, \tag{3.6}$$

where, conditional on W, Y has a binomial distribution with parameters $n-W$ and u. Thus we focus attention on the distribution of W.

Let

$$a_k = P(\alpha^*(k-1) = 1 \mid \alpha^*(k) = 1)$$

and

$$b_k = P(\alpha^*(k-1) = 1 \mid \alpha^*(k) = 0).$$

In diagrammatic form the transition probabilities $p_1(k), \ldots, p_8(k)$ can be represented as in Figure 3.1. From this we can see that a standard first step analysis of the Markov chain $\Lambda(t)$ gives (in the notation of (3.1))

$$a_k = p_1(k) + p_4(k)a_k + p_3(k)b_k$$

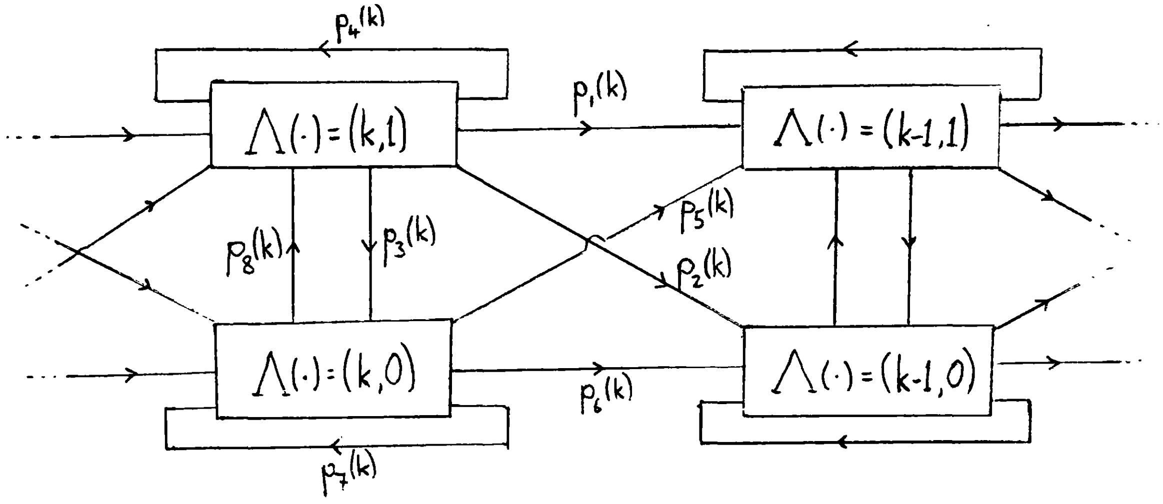


FIGURE 3·1

$$b_k = p_5(k) + p_8(k)a_k + p_7(k)b_k.$$

Hence

$$(a_k, b_k)\begin{pmatrix} 1 - p_4(k) & -p_8(k) \\ -p_3(k) & 1 - p_7(k) \end{pmatrix} = (p_1(k), p_5(k)).$$

So

$$(a_k, b_k) = \frac{(p_1(k), p_5(k))}{(1 - p_4(k))(1 - p_7(k)) - p_3(k)p_8(k)}\begin{pmatrix} 1 - p_7(k) & p_8(k) \\ p_3(k) & 1 - p_4(k) \end{pmatrix}.$$

Which gives the solution

$$a_k = [p_1(k)(1 - p_7(k)) + p_5(k)p_3(k)]/[(1 - p_4(k))(1 - p_7(k)) - p_3(k)p_8(k)]$$

$$b_k = [p_1(k)p_8(k) + p_5(k)(1 - p_4(k))]/[(1 - p_4(k))(1 - p_7(k)) - p_3(k)p_8(k)].$$

We now substitute in from (3.1) and after considerable algebraic manipulation (details of which can be found in Appendix I) we obtain

$$a_k = (k-1)(N+p-1)/[(1-p)(N-k)(N-1)u + (k-1)(N+p-1)]$$

$$b_k = (k-1)(N+p-1-pNu)/[(1-p)(N-k)(N-1)u + (k-1)(N+p-1)].$$

These forms do not lend themselves to explicit calculation. Instead we consider the usual limiting regime of fixed sample size n; large population size ($N \to \infty$); and mutation probabilities of order N^{-1}: specifically we write $u = \theta/(N+\theta)$ in the usual way. Then

$$a'_k \equiv \lim_{N\to\infty} a_k = \frac{k-1}{(1-p)\theta + (k-1)} = \lim_{N\to\infty} b_k \equiv b'_k. \tag{3.7}$$

The distribution of W may be written as a sum of (finite) products involving the a_k's and b_k's which will converge as $N \to \infty$ to the same sum of the identical products of the a_k's. In fact, since $a'_k = b'_k$,

$$\lim_{N\to\infty} P(W=m) = \sum_{\prime} a'_{i_1} a'_{i_2} \dots a'_{i_{n-m}} (1-a'_1)(1-a'_{j_1}) \dots (1-a'_{j_{m-1}}),$$

where the summation runs over all subsets $\{i_1, \dots, i_{n-m}\}$ of $\{2, \dots, n\}$ and we have written $\{j_1, \dots, j_{m-1}\}$ for $\{2, \dots, n\} \backslash \{i_1, \dots, i_{n-m}\}$. Thus from (3.7)

$$\lim_{N\to\infty} P(W=m) = \frac{((1-p)\theta)^m}{((1-p)\theta)_{(n)}} \sum_{\prime\prime} l_1 \dots l_{n-m},$$

where the summation extends over all subsets $l_1, \dots, l_{n-m}$ of $1, \dots, n-1$ and here and below we have written $x_{(n)}$ for $x(x+1)\cdots(x+n-1)$. But (Donnelly and Tavaré (1986), Eq. A.2) this summation equals $| S_n^{(m)} |$, where $S_n^{(m)}$ is a Stirling number of the first kind (Abramowitz and Stegun (1972)). So

$$\lim_{N\to\infty} P(W=m) = \frac{((1-p)\theta)^m}{((1-p)\theta)_{(n)}} \mid S_n^{(m)} \mid, \quad m = 1, 2, \dots, n.$$

Note that since W is bounded and $u \to 0$ as $N \to \infty$, $P(Y=0) \to 1$ as $N \to \infty$, so from (3.6), the (asymptotic) distribution of the number of types in the sample is

$$\lim_{N\to\infty} P(K=l) = \frac{((1-p)\theta)^l}{((1-p)\theta)_{(n)}} \mid S_n^{(l)} \mid, \quad l = 1, 2, \dots, n. \tag{3.8}$$

This has exactly the same form as the distribution of K in the uncorrelated (neutral) case (Ewens (1972)), the only difference being that in the uncorrelated case the factor $(1-p)\theta$ is replaced by θ. That is, at least under this limiting regime, the *only* effect of the correlation on the distribution of K is to change it to the distribution appropriate to an uncorrelated model with mutation rate $(1-p)\theta$ instead of θ.

In the light of the observation that as $N \to \infty$, $Y \to 0$ in probability in (3.6), lines of descent are lost through mutation (in this limit) exactly when $L(t)$ jumps down by one in such a way that the α component of $\Lambda(t)$ takes the value 0, ie. exactly when $\alpha^*(k)$ takes the value 0 $(k = n-1, n-2, \ldots, 0)$. The key to the form of the distribution (3.8) is the form of a'_k and b'_k (in fact $1-a'_k$ and $1-b'_k$) and the fact that these are equal: in the uncorrelated case, the probability that the change from k to $k-1$ lines of descent is due to a coalescence is exactly (for example Donnelly and Tavaré (1986))

$$\frac{k-1}{\theta+k-1}, \quad k = n, n-1, \ldots, 1,$$

that is (3.7) with $(1-p)\theta$ replaced by θ.

3.3 Allele Frequencies at Equilibrium

As well as studying the distribution of the number of alleles in the sample at equilibrium, it is natural to ask, in the spirit of the Ewens sampling formula (Ewens (1972)), about the distribution of allele frequencies, or in the spirit of Donnelly and Tavaré (1986), about the frequency distribution of age ordered alleles. Such questions are naturally studied in the context of an examination of the genealogy of the sample. As in Watterson (1984) or Donnelly and Tavaré (1986) one could introduce an equivalence-relation valued coalescent process. In the uncorrelated case the behaviour of this process (and certainly the fact that it gives rise to the Ewens or age-ordered sampling formulae) follows from that of the line of descent process and the fact that when two lines of descent coalesce it is equally likely to involve any of the existing lines.

We shall show that, in the limit as the population size tends to infinity, this is still true in the correlated case.

Suppose the process $\Lambda(\cdot)$ takes the value $(k, 0)$ at some time s. At some later time, t, say, the number of lines of descent will decrease to $k-1$ because

(i) an individual outside the line of descent will have a mutant offspring who is in the line of descent;

(ii) an individual in the line of descent will have a mutant offspring who is in the line of descent; or

(iii) an individual in the line of descent will have a non-mutant offspring who is also in the line of descent.

In the case of (i) it is clear that the offspring, and so the line lost, is equally likely to be any of the k available. For either (ii) or (iii) to occur, at some time r, $(s < r \leq t)$, the appropriate parent in the line of descent will have been chosen. As this parent is obviously not the parent at time s (that parent, by definition, being outside the line of descent) it will have been chosen at random from among the population, and conditional upon it being one of the individuals in the line of descent, it is equally likely to be any such individual. Conditional on its offspring also being in the line of descent the parent-offspring pair is then equally likely to be any of the $\binom{k}{2}$ possible such pairs. Thus if a coalescence (event of type (iii)) occurs, the coalescing lines are equally likely to be any of the possible pairs, while in an event of type (ii), the mutant offspring is equally likely to be any of the k possible individuals. All of the above choices are independent of the entire history of the process prior to time s, and in particular of its behaviour at each of the previous occasions on which lines of descent are lost.

Thus for some given $k \in \{1, 2, \ldots, n-1\}$, conditional on the fact that the process $\Lambda(\cdot)$ visits the state $(k, 0)$ at least once between the transition from

$L(\cdot) = k$ to $L(\cdot) = k-1$,

$$\begin{aligned} c_k &\equiv P(\text{a particular pair of lines coalesce} \mid \text{one of the } k \text{ lines is lost}) \\ &= b_k(1-u)\begin{pmatrix} k \\ 2 \end{pmatrix}^{-1} \end{aligned}$$

and

$$\begin{aligned} d_k &\equiv P(\text{a particular line of descent is lost through mutation} \\ &\qquad\qquad\qquad\qquad \mid \text{one of the } k \text{ lines is lost}) \\ &= (1 - b_k + b_k u)k^{-1}. \end{aligned}$$

Note that

$$c'_k \equiv \lim_{N\to\infty} c_k = b'_k \begin{pmatrix} k \\ 2 \end{pmatrix}^{-1} = \tfrac{2}{k((1-p)\theta+k-1)}$$

and

$$d'_k \equiv \lim_{N\to\infty} d_k = (1-b'_k)k^{-1} = \tfrac{(1-p)\theta}{k((1-p)\theta+k-1)}. \tag{3.9}$$

In view of the symmetry inherent in the initial distribution of Λ, we also have (in an obvious notation)

$$c'_n = \tfrac{2}{n((1-p)\theta+n-1)} \quad \text{and} \quad d'_n = \tfrac{(1-p)\theta}{n((1-p)\theta+n-1)} \tag{3.10}$$

regardless of whether or not the process $\Lambda(\cdot)$ ever visits $(n, 0)$. The distribution of allele frequencies at equilibrium (with or without age ordering) is completely determined by conditional probabilities of the above form.

For $k = n-1, n-2, \ldots, 1$, denote by A_k the event

$$\{\Lambda(t) \text{ visits } (k, 0) \text{ between the transition of } L(t) \text{ from } k \text{ to } k-1\}.$$

Now, trivially

$$P(A_k \mid \alpha^*(k) = 0) = 1, \tag{3.11}$$

and a standard first-step argument (using the notation of (3.1)) gives

$$P(A_k \mid \alpha^*(k) = 1) = p_3(k) + p_4(k)P(A_k \mid \alpha^*(k) = 1).$$

Thus

$$P(A_k \mid \alpha^*(k) = 1) = \frac{p_3(k)}{1 - p_4(k)}$$

and it is easily checked that

$$\lim_{N\to\infty} P(A_k \mid \alpha^*(k) = 1) = 1. \tag{3.12}$$

Now let $A = \bigcap_{k=1}^{n-1} A_k$ and we have (using B^c to denote the complement of a set B)

$$P(A^c) = P\left(\bigcup_{k=1}^{n-1} A_k^c\right) \le \sum_{k=1}^{n-1} P(A_k^c) \to 0 \tag{3.13}$$

as $N \to \infty$ from (3.11) and (3.12).

Thus if $B_{k;\mu_1,\ldots,\mu_k}$ denotes the event that the sample of n taken from the population at equilibrium contains k types, with μ_1 of the oldest type, μ_2 of the second oldest, $\ldots,\mu_k$ of the youngest type (and we note that as a consequence of the Moran model formulation, each type in the population has a unique age),

$$\begin{aligned} P(B_{k;\mu_1,\ldots,\mu_k}) &= P(B_{k;\mu_1,\ldots,\mu_k} \cap A) + P(B_{k;\mu_1,\ldots,\mu_k} \cap A^c) \\ &= P(A)P(B_{k;\mu_1,\ldots,\mu_k} \mid A) + P(A^c)P(B_{k;\mu_1,\ldots,\mu_k} \mid A^c), \end{aligned}$$

and in the light of (3.13) we have

$$\lim_{N\to\infty} P(B_{k;\mu_1,\ldots,\mu_k}) = \lim_{N\to\infty} P(B_{k;\mu_1,\ldots,\mu_k} \mid A).$$

This last, however, involves a calculation using the c'_k and d'_k, $k = 1, 2, \ldots, n$, exhibited at (3.9) and (3.10) identical to the related calculation of, for example, Donnelly and Tavaré (1986). It follows that

$$\lim_{N\to\infty} P(B_{k;\mu_1,\ldots,\mu_k}) = \frac{((1-p)\theta)^k}{((1-p)\theta)_{(n)}} \frac{n!}{\mu_k(\mu_k+\mu_{k-1})\cdots(\mu_k+\cdots+\mu_1)}, \qquad (3.14)$$

and

$$\lim_{N\to\infty} P(k \text{ types in the sample,}$$

$$\mu_1 \text{ of one type, } \mu_2 \text{ of another, } \ldots, \mu_k \text{ of the } k\text{th})$$

$$= \frac{((1-p)\theta)^k}{((1-p)\theta)_{(n)}} \frac{n!}{\mu_1\mu_2\cdots\mu_k\beta_1!\beta_2!\cdots\beta_k!} \qquad (3.15)$$

where β_j is the number of alleles represented j times, $j = 1, 2, \ldots, n$. Again, these are *exactly* the formulae appropriate to the uncorrelated neutral models with mutation rates $(1-p)\theta$: the age ordered sampling formula ((4.2) of Donnelly and Tavaré (1986) and the Ewens sampling formula Ewens (1972)) respectively.

3.4 Discussion

The motivation for this analysis stems from interest in the neutralist-selectionist controversy. Observed "departures" from neutrality in the gene frequencies might initially, without knowledge of the above results, be thought to be due not to selective forces but to a reproductive mechanism involving some form of correlation, perhaps because of linkage disequilibrium with a locus at which selection *is* operating, or a variable environment, or some other cause. While perhaps surprising, these results suggest that (for this particular model) correlation cannot explain departures from neutrality.

Tests for neutrality are usually performed *conditional* on K, the number of types in the sample (since in the uncorrelated case K is a sufficient statistic for the nuisance parameter θ). In the uncorrelated case the distribution of allele frequencies (with, and hence without, age ordering) conditional on K is *identical* to its distribution in the uncorrelated case, so in the presence of correlation the distribution of any test statistic will be identical to its distribution in the uncorrelated, neutral, model. This has two consequences. First, observed departures from neutrality cannot be attributed to this form of correlation. Second, any data which cause rejection of the neutral hypothesis at a given level would also result in rejection of the "correlated neutral" hypothesis at the same level.

If the sample data are used to *estimate* the mutation rate θ, the appropriate course in the uncorrelated neutral case is to base such estimates on K. (See Ewens (1972), for the form of the estimator.) The use of this estimator in the more general setting will give asymptotically unbiased estimation of $(1-p)\theta$, and hence underestimation of θ, if $p \geq 0$. (Of course, it may be possible to detect correlation, and possibly to estimate p separately by direct observation of the population and so correctly estimate θ.)

A priori one might expect the introduction of correlation to result in fewer alleles in the sample, with a more diverse range of frequencies, than in the neutral case: intuitively, correlation should encourage relatively more coalescences and those coalescences should affect classes which themselves were recently the result of coalescences. It seems that the correct (after the fact!) intuition is that coalescences *are* relatively more frequent and so the number of types in the sample is (stochastically) smaller (this expresses itself in a lower "mutation" rate), but that the times between losses of lines of descent are sufficiently long to ensure that the particular correlation present at one such loss will have been "forgotten" by the next loss, and so actual class sizes (conditional on K) are unaffected. On one level then, these results provide yet further evidence for the very general applicability of the Ewens sampling formula: the form of the distribution still applies, and the conditional distribution of allele frequencies is exactly as it should be. It is worth noting at this point that in this sense the results here are similar to other work (Gillespie (1977); Sawyer and Hartl (1985), for example) which shows that the Ewens sampling formula is also valid in various non-neutral situations.

It is natural to ask about the robustness of these results: that is, to what extent they are artifacts of the specific assumptions made. The assumption of a large population size (and the limiting regime studied) is common and not unreasonable. (In the uncorrelated case, with the exception of the Moran model, the Ewens sampling formula itself is an approximate result, for the validity of which the population size must be large compared to the sample size.) In fact, it is possible to be exact throughout the analysis about the effect of this assumption here, in order to show that the error in the expressions (3.8), (3.14), and (3.15) is of order N^{-1}.

What of the Moran model formulation? In the next section we investigate a model in which a fixed number $T(\geq 1)$ of individuals die and are replaced by new offspring with a similar mode of correlation, and determine what effect this has on the conclusions drawn from this section.

We also note that all the conclusions of the model remain valid if it is changed so that the reproducing individual in the "next" generation will be the reproducing individual of the current generation with probability $p/2$, its offspring with probability $p/2$, and otherwise an individual randomly chosen from the (whole) population.

It is less clear how to extend this type of correlation to more general (say

Wright-Fisher) reproductive mechanisms in a simple (and hence tractable) way (although Barton (1988) studies related issues). Some general comments are possible, however. In most formulations of a model with non-overlapping generations, correlation will not change the rate at which lines of descent are lost through mutation: in any given generation, a particular line of descent will be lost with probability u (which is usually assumed to be of order N^{-1} in this context), independently of all other events. Correlation might be expected to increase the rate at which coalescences occur, in which case, as in the model studied here, the distribution of the number of types in the sample will be stochastically smaller than in the uncorrelated case. In the model studied here, the advantage to a particular individual due to correlation lasts for a geometrically distributed number of generations, while the times between losses of lines of descent are of the order of N^{-2} generations, so that for large N, this advantage is "forgotten" between these events, and we have seen that as a consequence allele frequency distributions (conditional on K) are unaffected. In most other models, including the Wright-Fisher model (see for example Donnelly and Tavaré (1986), for a general formulation), the times between losses of lines of descent are of order N generations. (For the general model of Section 2 the times between losses are of order $N^{2-\alpha}$ where $0 \leq \alpha < 1$). If the advantage of correlation again lasted for a geometrically distributed number of generations (and this is not inconsistent with linkage to a non-neutral locus), this would also be "forgotten", and again allele frequencies, conditional on K, might be unaffected. There is some hope then that results in the same spirit as the ones discussed above might remain valid in a considerably more general setting. That is, correlation should result in a stochastically smaller number of alleles in the sample (though perhaps not with a distribution of the form (3.8)), but conditional on the number of alleles present, the distribution of allele frequencies will be as predicted by the Ewens sampling formula.

Chapter 4

CORRELATION IN MORE COMPLICATED MODELS

In the spirit of the previous chapter we move on to investigate how similar correlation may affect more complicated models. In particular in this section we shall introduce correlation to the model in which T individuals reproduce in each generation. We shall then see whether the effect of correlation in this model is similar to that in the conventional Moran model and try to draw some general conclusions regarding the way in which correlation effects are likely to manifest themselves in more general models. This represents an attempt to investigate to what extent the conclusions of the previous chapter are artifacts of the specific model discussed. We start by giving an exact specification of the model to be considered in this chapter.

We propose a population of N individuals evolving forward through time in discrete generations $(t = \dots, -2, -1, 0, 1, 2, \dots)$ In the general course of evolution of the population, at each generation, we choose T existing individuals, uniformly at random, to die (thus all $\begin{pmatrix} N \\ T \end{pmatrix}$ choices are equally likely) and

then we choose T other individuals to be parents, each to have one offspring. The choice of parents at time $t+1$ is according to the following regime:
(i) Label the parents at time t: $X_1, X_2, \ldots, X_T$.
(ii) Independently for each X_i ($1 \leq i \leq T$), there is a probability p that X_i is again the parent at time $t+1$. Suppose $S \leq T$ parents are chosen this way.
(iii) Choose the remaining $(T-S)$ parents uniformly at random without replacement from the $(N-S)$ individuals who are not already chosen. The surviving $(N-T)$ individuals and the T new offspring form the population in the next generation. The offspring individuals may independently, with probability u, mutate into a novel type which has not been in the population before. We will note the usual assumption that u is of order N^{-1}. We further assume that this correlated reproduction process has been going on indefinately and is therefore in equilibrium.

Note that this model is slightly different from the similar model in Chapter 2, in that here parents may have only one offspring and the reproducing individuals are distinct from the individuals that die. We do this as the algebra, which is already somewhat complicated, is a little more transparent in this case. But although we do not give the details, exactly the same results obtain for the earlier model.

More formally, we can define a Markov chain

$$Z_t = \{(\underline{X}_t, \underline{Y}_t) : t = 0, 1, 2, \ldots\} \text{ with statespace } E_N^T \times E_N^T$$

where $E_N^T \equiv \{A \subseteq \{1, 2, \ldots, N\} \mid |A| = T\}$. Here

$$\underline{X}_t = \{X_{t,1}, X_{t,2}, \ldots, X_{t,T}\} = \{\text{parents at time } t\}$$

$$\underline{Y}_t = \{Y_{t,1}, Y_{t,2}, \ldots, Y_{t,T}\} = \{\text{ dying individuals at time t}\}.$$

The initial distribution places mass

$$\frac{\binom{N}{T}\binom{N-T}{T}}{\binom{N}{2T}} = \frac{(2T)!}{\cancel{4}(T!)^2}$$

at each of the points $(\underline{x}, \underline{y}) \in E_N^T \times E_N^T$ with all components distinct (ie. $\underline{x} = (x_1, \ldots, x_T), \underline{y} = (y_1, \ldots, y_T), x_i \neq x_j, y_i \neq y_j, x_i \neq y_j,\ \forall i \neq j$)

To define the transition probabilities between $(\underline{X}_t, \underline{Y}_t)$ and $(\underline{X}_{t+1}, \underline{Y}_{t+1})$, suppose that m is the number of individuals in common between $\underline{X}_t = \underline{x}_t$ and $\underline{X}_{t+1} = \underline{x}_{t+1}$ then

$$P(\underline{X}_{t+1} = \underline{x}_{t+1}, \underline{Y}_{t+1} = \underline{y}_{t+1} \mid \underline{X}_t = \underline{x}_t, \underline{Y}_t = \underline{y}_t)$$

$$= \sum_{i=0}^{m} \binom{m}{i} p^i (1-p)^{m-i} \left\{ \frac{\binom{N-i}{T-i}\binom{N-T}{T}}{\binom{N-i}{2T-i}} \right\}.$$

We assume the population has been evolving in this way for an indefinite period of time, and take a random sample of n individuals without replacement from the population at a time which we will call $t = 0$. We then trace the composition of the sample with respect to the ancestral population at time $-t$ $(t = 0, 1, 2, \ldots)$ in a manner akin to that of the previous chapter.

We define "old" and "new" equivalence classes at time t as in Chapter 3, denoted by $\epsilon_i, (i = 1, \ldots, K_t)$ and $\eta_j, (j = 1, \ldots, F_t)$ respectively. We order the η_j by decreasing age, the age of an equivalence class being the number of generations since the mutation which led to its existence occurred, (ties are broken at random).

So we can summarize the state of the sample with respect to the ancestral population at time $-t$ by

$$R_t = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_{K_t}; \eta_1, \eta_2, \ldots, \eta_{F_t}\} = \{\underline{\epsilon}_t, \underline{\eta}_t\} \text{ say.}$$

We shall concentrate on the $\underline{X}_t$ component of the Markov chain Z_t in order to investigate the genetic composition of the population at equilibrium. In the light of this we shall study the behaviour of K_t. As in the previous chapter it is clear that $\{K_t, t = 0, 1, \ldots\}$ is no longer Markov, and again we study a more informative process.

Hence we define

$$\{\Lambda(t) = (K_t, C_t) : \ t = 0, 1, 2, \ldots\}$$

where C_t is defined to be the number of parents at time $-t$ who are ancestors of old equivalence classes in the original sample, (ie. the number of parents at time $-t$ who are in the line of descent), $0 \leq C_t \leq K_t$. It is evident that $\Lambda(t)$ is a Markov chain.

Before continuing we shall show that the process $\underline{X}_t$ is time-reversible. It is easy to check that the equilibrium distribution places mass

$$\pi_{\underline{x}} = \begin{pmatrix} N \\ T \end{pmatrix}^{-1}$$

on each $\underline{x} \in E_N^T$, and that the process is finite and irreducible. The choice of initial distribution ensures that it is also stationary.

Once again (Kelly 1979) reversibility will follow if

$$\pi_{\underline{x}_t} P(\underline{X}_{t+1} = \underline{x}_{t+1} \mid \underline{X}_t = \underline{x}_t) = \pi_{\underline{x}_{t+1}} P(\underline{X}_t = \underline{x}_t \mid \underline{X}_{t+1} = \underline{x}_{t+1})$$

for all $\underline{x}_t, \underline{x}_{t+1} \in E_N^T$. But this is clear since

$$P(\underline{X}_{t+1} = \underline{x}_{t+1} \mid \underline{X}_t = \underline{x}_t) = P(\underline{X}_t = \underline{x}_t \mid \underline{X}_{t+1} = \underline{x}_{t+1}).$$

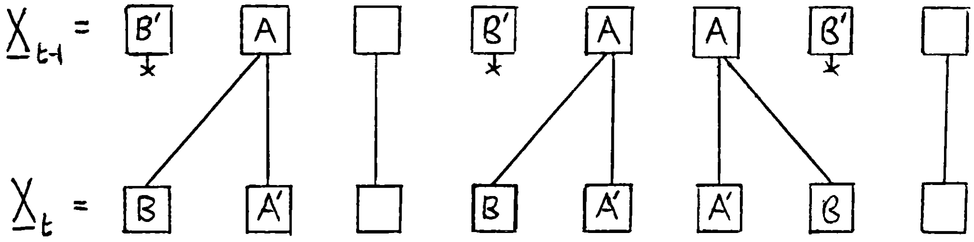


FIGURE 4·1

Therefore, as before the behaviour of the process when viewed backwards in time is (stochastically) equivalent to its behaviour going forwards. In particular, when viewed backwards $\underline{X}_t$ is Markov.

Note that henceforth we are considering the process $\underline{X}_t$ going backwards in time, so that if we are currently at time $-t$, by "next" is meant time $-(t+1)$, and so by "previous" we mean time $-(t-1)$.

For ease of determining the transition equations we now split the reproductive procedure into two steps:

1. Select the T parents in the next generation,
2. Select the T offspring of the parents just identified.

Figure 4.1 illustrates this pictorially.

Step 1 consists of selecting the individuals marked A (or equivalently those marked A').

Step 2 consists of selecting those marked B (or equivalently B').

Define the event P_d to be

$P_d = \{d$ parents are drawn at random from the line of descent (at the end of step 1)}.

$$\text{Let } C_{t-1} = c \;=\; \text{the number of parents in the line of descent in the previous generation.}$$

$$\text{And let } K_{t-1} = K \;=\; \text{the number of old equivalence classes in the previous generation.}$$

Then

$$P(P_d \mid C_{t-1} = c, K_{t-1} = K) =$$

$$\sum_{i=0}^{c\wedge d}\left(p^i(1-p)^{c-i}\binom{c}{i}\right. \tag{4.1}$$

$$\times \sum_{j=0}^{T-(c\vee d)}\left(p^j(1-p)^{T-c-j}\binom{T-c}{j}\right.$$

$$\left.\left.\frac{\binom{K-i}{d-i}\binom{N-K-j}{T-d-j}}{\binom{N-i-j}{T-i-j}}\right)\right).$$

To see this, note that we first choose i of the c parents from the line of descent again. Then we choose j of the $(T-c)$ parents not in the line of descent again. Finally we randomly choose the remaining $(T-i-j)$ parents in such a way that we finish with d parents in the line of descent.

So now we have picked the parents. It remains to select the offspring, and there are four possible ways of doing this:

(i) So that we don't lose any lines of descent (ie. $K_t = K_{t-1}$)

(ii) So that we lose a line of descent via a coalescence of two lines (ie. $K_t = K_{t-1} - 1$)

(iii) So that we lose a line of descent via a mutation (again $K_t = K_{t-1} - 1$).

(iv) So that we lose more than one line of descent (ie. $K_t = K_{t-1} - l$, $l \geq 2$).

The probabilities of these events are as follows:

(i) Let Q_K denote the event { no line of descent is lost } (ie. $K_t = K_{t-1}$). Then

$$P(Q_K \mid P_d, K_{t-1} = K) = \tag{4.2}$$

$$\frac{\begin{pmatrix} N-K-(T-d) \\ d \end{pmatrix}}{\begin{pmatrix} N-T \\ d \end{pmatrix}} \sum_{i=0}^{(T-d)\wedge(K-d)} \left[\frac{\begin{pmatrix} K-d \\ i \end{pmatrix}\begin{pmatrix} N-T-K \\ T-d-i \end{pmatrix}}{\begin{pmatrix} N-T-d \\ T-d \end{pmatrix}} (1-u)^i \right]$$

To see this, notice that the offspring of the d parents in the line of descent must not be in the line of descent themselves. Having chosen these, the other $(T-d)$ offspring are picked. We can have i $(0 \leq i \leq T-d)$ of these drawn from the line of descent, provided we ensure they do not themselves mutate and thus cause the loss of a line of descent. The remainder are picked from individuals not in the line of descent. The number of parents who are in the line of descent in the next generation is $(d+i)$, ie. $C_t = d + i$.

Note that from the preceding two equations we have

$$P(P_d \mid C_{t-1} = c, K_{t-1} = K) = \begin{cases} O(1) & \text{for } d \leq c \\ O(N^{-(d-c)}) & \text{for } d > c \end{cases}$$

and

$$P(Q_K \& \{C_t = e\} \mid P_d, K_{t-1} = K) = \begin{cases} O(N^{-(e-d)}) & \text{for } K \geq e \geq d \\ 0 & \text{otherwise.} \end{cases}$$

(ii) Now pick offspring so that we lose a line of descent via a coalescence of two lines. So:

let $\Gamma_K = \{$ lose a line of descent via coalescence $\}$.

Then

$$P(\Gamma_K \mid P_d, K_{t-1} = K)$$

$$= \frac{\binom{K-d}{1}\binom{N-T-(K-d)}{d-1}}{\binom{N-T}{d}}(1-u) \times \tag{4.3}$$

$$\sum_{i=0}^{(T-d)\wedge(K-d)} \left[\frac{\binom{K-d-1}{i}\binom{N-T-K+1}{T-d-i}}{\binom{N-T-d}{T-d}}(1-u)^i \right]$$

Here we must pick one parent in the line of descent to have an offspring in the line of descent without a mutation occurring. The remaining offspring of the $(d-1)$ parents in the line of descent must not themselves be in the line of descent. We may then pick $0 \leq i \leq T-d$ offspring from the line of descent and ensure they don't mutate. Finally, pick the rest of the offspring from individuals not in the line of descent.

Note that $C_t = d + i - 1$.

Notice also that P(lose l lines of descent via coalescence) involves choosing l offspring from the line of descent, an event of probability $O(N^{-l})$. So we have, for $l \geq 1$,

$$P(\text{lose } l \text{ lines of descent via coalescence}) = O(N^{-l}).$$

(iii) Now we pick offspring so that we lose a line of descent via mutation: Let M_K denote the event {lose a line of descent via mutation}.
Then

$$P(M_K \mid P_d, K_{t-1} = K) =$$

$$\begin{aligned}
&\left[\frac{\binom{K-d}{1}\binom{N-T-K+d}{d-1}}{\binom{N-T}{d}} u \times \right. \\
&\qquad \left. \sum_{i=0}^{(T-d)\wedge(K-d)} \left[\frac{\binom{K-d-1}{i}\binom{N-T-K+1}{T-d-i}}{\binom{N-T-d}{T-d}} (1-u)^i \right] \right] \\
&+ \frac{\binom{N-T-K+d}{d}}{\binom{N-T}{d}} \\
&\qquad \sum_{i=1}^{(T-d)\wedge(K-d)} \left[\frac{\binom{K-d}{i}\binom{N-T-K}{T-d-i}}{\binom{N-T-d}{T-d}} \binom{i}{1} u(1-u)^{i-1} \right]
\end{aligned} \tag{4.4}$$

Here we can do one of two things

(a) pick one of the offspring of the individuals in the line of descent to itself be in the line of descent, mutate it, ensure that all other such offspring are out of the line of descent, and then pick $0 \leq i \leq T-d$ other offspring in the line of descent which don't mutate,

(b) ensure that all offspring of individuals in the line of descent are themselves outside the line of descent, and then pick at least one other offspring to be in the line of descent and make sure that one of these other offspring mutates.

Note that $\mathrm{P}(M_K) = O(N^{-2})$.

Notice also that P(lose l lines of descent via mutation) will involve choosing l offspring from the line of descent and then mutating each of them. So we have, for $l \geq 1$

$$\mathrm{P}(\text{lose } l \text{ lines of descent via mutation}) = O(N^{-2l}).$$

(iv) If we are to lose more than one line of descent between two generations then a combination of coalescences and/or mutations is needed. From cases (ii) and (iii) we can see that

$$\mathrm{P}(\text{lose } l \text{ lines of descent solely via coalescence}) = O(N^{-l})$$

and

$$\mathrm{P}(\text{lose } l \text{ lines of descent solely via mutation}) = O(N^{-2l}).$$

Similarly

$$\begin{aligned} & \mathrm{P}(\text{lose } l \text{ lines of descent } (m \text{ of them via mutation})) \\ = \quad & O(N^{-2m-(l-m)}) = O(N^{l-m}). \end{aligned}$$

Since there are only a finite number of ways in which these events can occur, we see that

$$\mathrm{P}(\text{lose } l \text{ lines of descent, } l \geq 2) = O(N^{-3})$$

except for the case where we lose two lines of descent via coalescence, which if there was at least one parent in the line of descent is $O(N^{-2})$. However, it will become clear later that exact expressions for these probabilities are not required.

Now we can proceed to construct the actual transition probabilities by taking products of equation (4.1) and one of equations (4.2)-(4.4). We shall write them out in expanded form as this is necessary for later simplification. Ultimately, conditional on $\{K_{t-1} = K\}$ and $\{C_{t-1} = c\}$ it will be seen that we need only consider terms up to $O(N^{-2})$. So with $P_K, \{K_{t-1} = K\}$ and $\{C_{t-1} = c\}$ defined as before, and with

$$\begin{aligned} Q_K^e &= \{\text{no line of descent is lost this generation (ie. time } t\text{) and we finish} \\ &\qquad \text{with } e \text{ parents in the line of descent}\} \\ &= Q_K \cap \{C_{t-1} = c\} \text{ where } Q_K = \{K_{t-1} = K\} \end{aligned}$$

we have

$$P(Q_K^e \mid K_{t-1} = K, C_{t-1} = c) = \sum_{d=0}^{e} [P(P_d \mid c, K) \times P(Q_K^e \mid d)]$$

$$\begin{aligned} = \sum_{d=0}^{e} \Bigg[\Bigg\{ \sum_{i=0}^{c\wedge d} \Bigg(& p^i(1-p)^{c-i} \binom{c}{i} \sum_{j=0}^{T-(c\vee d)} \Big[p^j(1-p)^{T-c-j} \\ & \times \binom{T-c}{j} \binom{T-j-i}{d-i} \\ & \times \frac{K-i}{N-j-i} \times \frac{K-i-1}{N-j-i-1} \times \cdots \times \frac{K-i-(d-i-1)}{N-j-i-(d-i-1)} \\ & \times \frac{N-K-j}{N-d-j} \times \frac{N-K-j-1}{N-d-j-1} \times \cdots \times \frac{N-K-j-(T-d-j-1)}{N-d-j-(T-d-j-1)} \Big] \Bigg) \Bigg\} \end{aligned} \tag{4.5}$$

$$\times \frac{N-K-(T-d)}{N-T} \times \frac{N-K-(T-d)-1}{N-T-1} \times \cdots$$

$$\times \frac{N-K-(T-d)-(d-1)}{N-T-(d-1)} \times \begin{pmatrix} T-d \\ e-d \end{pmatrix} (1-u)^{e-d}$$

$$\times \frac{K-d}{N-T-d} \times \frac{K-d-1}{N-T-d-1} \times \cdots \times \frac{K-d-(e-d-1)}{N-T-d-(e-d-1)}$$

$$\times \frac{N-K-T}{N-T-e} \times \frac{N-K-T-1}{N-T-e-1} \times \cdots \times \frac{N-K-T-(T-e-1)}{N-T-e-(T-e-1)} \Bigg]$$

Note that we need only consider $e-2 \leq d \leq e$ if we are just interested in terms up to $O(N^{-2})$.

The reader is spared the details of the simplification of this equation here, but a blow-by-blow account of it, along with that of equations (4.7) and (4.8) can be found in Appendix II (sections A,B and C respectively).

From (4.6) we have

$$P(Q_K^e \mid K_{t-1} = K, C_{t-1} = c) = \begin{cases} O(1) & e \leq c \\ O(N^{-(e-c)}) & e > c. \end{cases}$$

We will also require the probabilities of losing a line of descent. So with

$$M_K = \{\text{lose a line of descent via mutation}\}$$

$$\Gamma_K = \{\text{lose a line of descent via coalescence}\}$$

we have:

$$P(M_K \mid K_{t-1} = K, C_{t-1} = c)$$

$$= \sum_{d=0}^{T \wedge K} [P(P_d \mid K_{t-1} = K, C_{t-1} = c) \times P(M_K \mid d)]$$

$$= \sum_{d=0}^{T\wedge K} \left\{ \sum_{i=0}^{c\wedge d} \left(p^{i}(1-p)^{c-i} \begin{pmatrix} c \\ i \end{pmatrix} \right.\right.$$

$$\left[\sum_{j=0}^{T-(c\vee d)} \left(p^{j}(1-p)^{T-c-j} \begin{pmatrix} T-c \\ j \end{pmatrix} \begin{pmatrix} T-j-i \\ d-i \end{pmatrix} \right.\right.$$

$$\times \frac{K-i}{N-j-i} \times \frac{K-i-1}{N-j-i-1} \times \cdots \times \frac{K-i-(d-i-1)}{N-j-i-(d-i-1)}$$

$$\left.\left.\left. \times \frac{N-K-j}{N-d-j} \times \frac{N-K-j-1}{N-d-j-1} \times \cdots \times \frac{N-K-j-(T-d-j-1)}{N-d-j-(T-d-j-1)} \right)\right]\right)$$

$$\times \left[\left\{ \begin{pmatrix} d \\ 1 \end{pmatrix} \left(\frac{K-d}{N-T} \right) u \times \frac{N-K-(T-d)}{N-T-1} \right.\right.$$

$$\times \frac{N-K-(T-d)-1}{N-T-2} \times \cdots \times \frac{N-K-(T-d)-(d-2)}{N-T-(d-1)}$$

$$\times \sum_{i=0}^{T-d} \left[\frac{K-d-1}{N-T-d} \times \frac{K-d-2}{N-T-d-1} \times \cdots \times \frac{K-d-1-(i-1)}{N-T-d-(i-1)} \right.$$

$$\times \begin{pmatrix} T-d \\ i \end{pmatrix} (1-u)^{i} \times \frac{N-K-(T-1)}{N-T-d-i} \times$$

$$\left.\left. \times \frac{N-K-(T-1)-1}{N-T-d-i-1} \times \cdots \times \frac{N-K-(T-1)-(T-d-i-1)}{N-T-d-i-(T-d-i-1)} \right] \right\}$$

$$+ \left\{ \frac{N-K-(T-d)}{N-T} \right.$$

$$\times \frac{N-K-(T-d)-1}{N-T-1} \times \cdots \times \frac{N-K-(T-d)-(d-1)}{N-T-(d-1)}$$

$$\times \sum_{i=1}^{T-d} \left[\frac{K-d}{N-T-d} \times \frac{K-d-1}{N-T-d-1} \times \cdots \times \frac{K-d-(i-1)}{N-T-d-(i-1)} \right. \tag{4.6}$$

$$\times u(1-u)^{i-1} \begin{pmatrix} i \\ 1 \end{pmatrix} \begin{pmatrix} T-d \\ i \end{pmatrix} \times \frac{N-K-T}{N-T-d-i} \times$$

$$\left. \times \frac{N-K-T-1}{N-T-d-i-1} \times \cdots \times \frac{N-K-T-(T-d-i-1)}{N-T-d-i-(T-d-i-1)} \right] \Bigg\} \Bigg] \Bigg\}.$$

After some not inconsiderable manipulation (see Appendix II, section B), (4.6) simplifies to

$$P(M_K \mid K_{t-1} = K, C_{t-1} = c) = \frac{\theta T}{N^2}[K - cp] + O(N^{-3})$$

where we define $u = \frac{\theta}{\theta + N}$ as is normal. In particular, from (4.6) we have

$$P(M_K \mid K_{t-1} = K, C_{t-1} = 0) = \frac{K\theta T}{N^2} + O(N^{-3}).$$

And in a similar manner for coalescence events we have

$$P(\Gamma_K \mid K_{t-1} = K, C_{t-1} = c)$$

$$= \sum_{d=0}^{T \wedge K} \left[P(P_d \mid K_t = K, C_t = c) \times P(\Gamma_K \mid d) \right]$$

$$= \sum_{d=0}^{T \wedge K} \left[\left\{ \sum_{i=0}^{c \wedge d} \left(p^i (1-p)^{c-i} \begin{pmatrix} c \\ i \end{pmatrix} \right. \right. \right. \tag{4.7}$$

$$\left[\sum_{j=0}^{T-(c \vee d)} (p^j (1-p)^{T-c-j} \begin{pmatrix} T-c \\ j \end{pmatrix} \begin{pmatrix} T-j-i \\ d-i \end{pmatrix} \right.$$

$$\times \left[\frac{K-i}{N-j-i} \times \frac{K-i-1}{N-j-i-1} \times \cdots \frac{K-i-(d-i-1)}{N-j-i-(d-i-1)} \right]$$

$$\times\frac{N-K-j}{N-d-j}\times\frac{N-K-j-1}{N-d-j-1}\times\cdots\times\frac{N-K-j-(T-d-j-1)}{N-d-j-(T-d-j-1)}\Bigg)\Bigg]\Bigg)\Bigg\}$$

$$\times\left\{\frac{K-d}{N-T}\begin{pmatrix} d \\ 1 \end{pmatrix}(1-u)\frac{N-T-K+d}{N-T-1}\right.$$

$$\times\frac{N-T-K+d-1}{N-T-2}\times\cdots\times\frac{N-T-K+d-(d-2)}{N-T-1-(d-2)}$$

$$\times\sum_{i=0}^{T-d}\left[\begin{pmatrix} T-d \\ i \end{pmatrix}\frac{K-d-1}{N-T-d}\right.$$

$$\times\frac{K-d-2}{N-T-d-1}\times\cdots\times\frac{K-d-1-(i-1)}{N-T-d-(i-1)}$$

$$\times(1-u)^{i}\times\frac{N-T-K+1}{N-T-d-i}$$

$$\left.\left.\left.\times\frac{N-T-K}{N-T-d-(i+1)}\times\cdots\times\frac{N-2T-K+d+i+2}{N-2T+1}\right]\right\}\right].$$

We simplify (4.7) to get

$$P(\Gamma_K \mid C_{t-1}=0, K_{t-1}=K)=\frac{K(K-1)T(1-p)}{N^2}+O(N^{-3}),$$

$$P(\Gamma_K \mid C_{t-1}=c>0, K_{t-1}=K)=\sum_{d=1}^{c}\left[p^d(1-p)^{c-d}\begin{pmatrix} c \\ d \end{pmatrix}\frac{(K-d)}{N}d\right]+O(N^{-2}).$$

In particular

$$P(\Gamma_K \mid C_{t-1}=1, K_{t-1}=K)=\frac{p(K-1)}{N}+O(N^{-2})$$

and

$$P(\Gamma_K \mid C_{t-1}=2, K_{t-1}=K)=\frac{2p(1-p)(K-1)}{N}+\frac{2p^2(K-2)}{N}+O(N^{-2}).$$

By simplifying equation (4.6) (for details see Appendix II, section C) we find the following probabilities of not losing a line of descent in varying circumstances:

$$\begin{aligned}
P(Q_K^0 \mid C_{t-1} = 2, K_{t-1} = K) &= (1-p)^2 + O(N^{-1}) \\
P(Q_K^1 \mid C_{t-1} = 2, K_{t-1} = K) &= 2p(1-p) + O(N^{-1}) \\
P(Q_K^2 \mid C_{t-1} = 2, K_{t-1} = K) &= p^2 + O(N^{-1}) \qquad (4.8) \\
P(Q_K^0 \mid C_{t-1} = 1, K_{t-1} = K) &= (1-p)\left[1 - \frac{TK(2-p)}{N} - \frac{pK}{N}\right] + O(N^{-2}) \\
P(Q_K^1 \mid C_{t-1} = 1, K_{t-1} = K) &= p + (1-p)\frac{TK}{N} + \frac{TK(1-p)^2}{N} \\
&\quad + \frac{p(1-p)K}{N} - \frac{p(K-1)T(2-p)}{N} \\
&\quad + \frac{p(1-p)(K-1)}{N} + O(N^{-2}) \\
P(Q_K^2 \mid C_{t-1} = 1, K_{t-1} = K) &= \frac{p(K-1)(T-1)(2-p)}{N} + O(N^{-2}) \\
P(Q_K^0 \mid C_{t-1} = 0, K_{t-1} = K) &= 1 - \frac{TK(2-p)}{N} + \binom{T}{2}\frac{K(K-1)(2-p)^2}{N^2} \\
&\quad + \frac{TK(K-1)(1-p)}{N^2} \\
&\quad - \frac{KT}{N^2}\left[Tp(2-p) - p(1-p)\right] + O(N^{-3}) \\
P(Q_K^1 \mid C_{t-1} = 0, K_{t-1} = K) &= \frac{(2-p)TK}{N} - \frac{TK\theta}{N^2}
\end{aligned}$$

$$-2\begin{pmatrix} T \\ 2 \end{pmatrix}\frac{K(K-1)(2-p)^2}{N^2}$$

$$-\frac{2TK(K-1)(1-p)}{N^2}$$

$$+\frac{KT}{N^2}\left[Tp(2-p)-p(1-p)\right]+O(N^{-3})$$

$$P(Q_K^2 \mid C_{t-1}=0, K_{t-1}=K) = \begin{pmatrix} T \\ 2 \end{pmatrix}\frac{K(K-1)(2-p)^2}{N^2}+O(N^{-3}).$$

These represent the relevant transition probabilities and Figure 4.2 shows the transition diagram.

For any particular state $(K_{t-1}=K, C_{t-1}=c)$ we have the situation illustrated in Figure 4.3.

Note that

$$P(Q_K^e \mid K_{t-1}=K, C_{t-1}=c) = \begin{cases} p^e(1-p)^{c-e}\begin{pmatrix} c \\ e \end{pmatrix}+O(N^{-1}) & \text{for } e \le c \\ O(N^{-(e-c)}) & \text{for } e > c. \end{cases}$$

Define

$$\Gamma_K^e = \Gamma_K \cap \{C_t = e\} \quad \text{and} \quad M_K^e = M_K \cap \{C_t = e\}.$$

(So Γ_K^e = {lose a line of descent via coalescence and finish with e parents in the line of descent} etc.). Then,

$$P(\Gamma_K^{c+d} \mid K_{t-1}=K, C_{t-1}=c) = \begin{cases} O(N^{-(d+1)}) & \text{for } d \ge 0 \\ O(N^{-1}) & \text{for } -c \le d < 0. \end{cases}$$

We now have the following result:

FIGURE 4.2

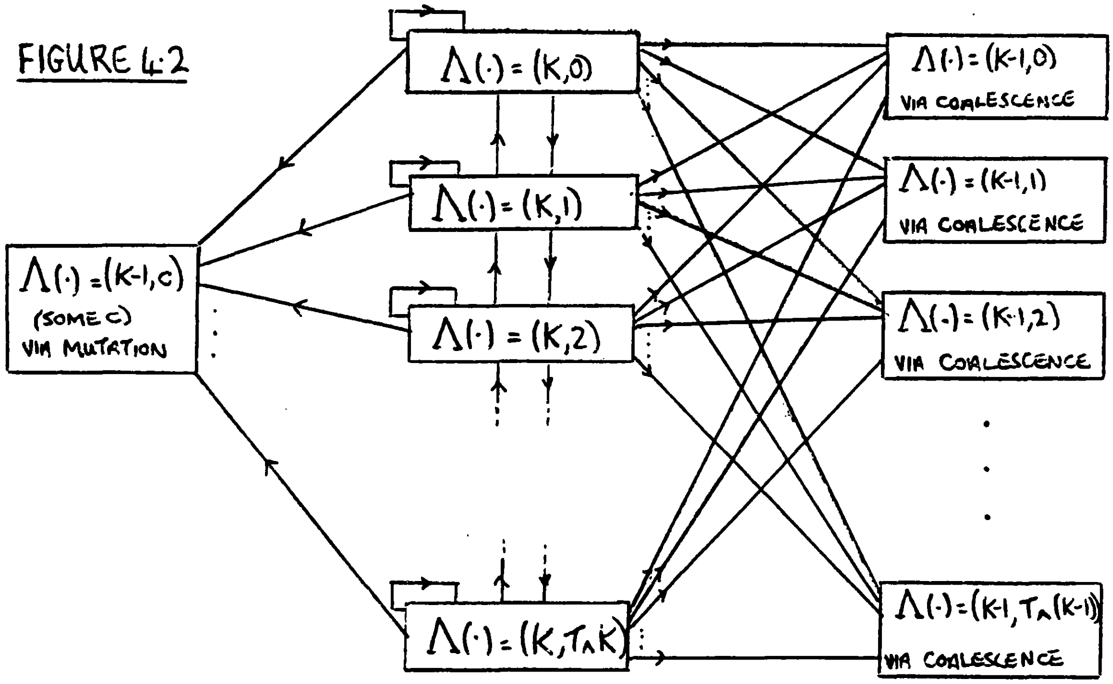
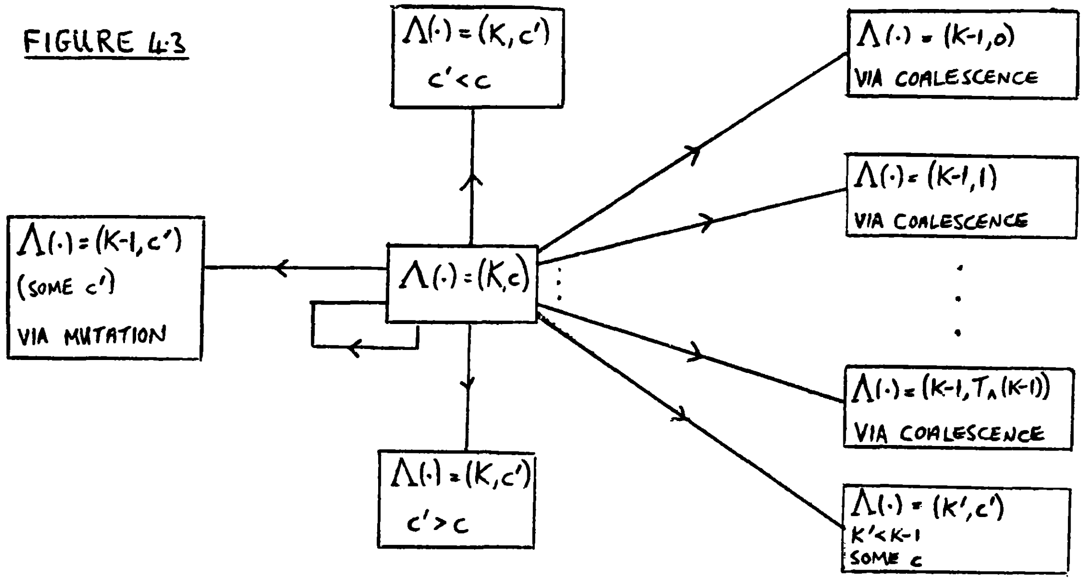


FIGURE 4.3



Lemma 4.1 *Let*

$$H = \max_{c'}\{P((M_K^{c'} \mid K_{t-1} = K, C_{t-1} = c)\} \vee \max_{c''}\{P(\Gamma_K^{c''} \mid K_{t-1} = K, C_{t-1} = c)\},$$

then $H \leq O(N^{-1})$.

Proof:

Recall $P(M_K \mid K_{t-1} = K, C_{t-1} = c) = \frac{\theta T}{N^2}[K - cp] + O(N^{-3})$.

Hence

$$P(M_K^{c'} \mid K_{t-1} = K, C_{t-1} = c) \leq \frac{\theta T}{N^2}[K - cp] + O(N^{-3})$$

and so

$$\max_{c'}\{P(M_K^{c'} \mid K_{t-1} = K, C_{t-1} = c)\} \leq \frac{\theta T}{N^2}[K - cp] + O(N^{-3}).$$

Similarly

$$P(\Gamma_K \mid K_{t-1} = K, C_{t-1} = c)$$

$$= \begin{cases} \frac{K(K-1)T(1-p)}{N^2} + O(N^{-3}) & \text{for } c = 0 \\ \sum_{d=1}^{c} \left[p^d (1-p)^{c-d} \begin{pmatrix} c \\ d \end{pmatrix} \frac{K-d}{N} d \right] + O(N^{-2}) & \text{for } c > 0. \end{cases}$$

Hence

$$P(\Gamma_K^{c''} \mid K_{t-1} = K, C_{t-1} = c) \leq \sum_{d=1}^{c} \left[p^d (1-p)^{c-d} \begin{pmatrix} c \\ d \end{pmatrix} \frac{K-d}{N} d \right] + O(N^{-2})$$

and so

$$\max_{c''}\{P(\Gamma_K^{c''} \mid K_{t-1} = K, C_{t-1} = c)\}$$

$$\leq \sum_{d=1}^{c} \left[p^d (1-p)^{c-d} \begin{pmatrix} c \\ d \end{pmatrix} \frac{K-d}{N} d \right] + O(N^{-2}).$$

Thus

$$\begin{aligned} H &\leq \left[\frac{\theta T}{N^2}[K - cp] + O(N^{-3})\right] \\ &\qquad\qquad \vee \left[\sum_{d=1}^{c}\left[p^d(1-p)^{c-d}\left(\begin{array}{c} c \\ d \end{array}\right)\frac{K-d}{N}d\right] + O(N^{-2})\right] \\ &= O(N^{-1}) \quad \square. \end{aligned}$$

Therefore, if we now define L_d to be the event {lose a line of descent before visiting (K, d), some $d \neq c$ | currently in (K, c)}, then

P(lose a line of descent before visiting (K, d) some $d < c$ | currently in (K, c))

$$= \frac{P(L_d) + O(N^{-1})}{P(L_d) + (1 - p^c) + O(N^{-1})} = O(N^{-1}),$$

Since $P(L_d) \leq 2(K-1)H = O(N^{-1})$. So now if we follow the same limiting regime as in the previous chapter and let $N \to \infty$ we get:

$$\begin{aligned} &\lim_{N\to\infty} P(\text{the process } \Lambda(t) \text{doesn't visit} (K_t, 0) \text{ before } K_t \text{ decreases} \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{from its current value}) \\ &= 0. \end{aligned}$$

In the manner of Chapter 3 we now prove the following lemma:

Lemma 4.2 *Asymptotically,* $\Lambda(t)$ *will visit* $(K, 0)$ *before* K_t *decreases, for all* $K_t = K$.

Proof:

Define

$$A_K = \{\Lambda(t) \text{ visits } (K, 0) \text{ before } K \text{ decreases}\}.$$

Set $A = \cap_{k=1}^{n-1} A_k$, and so

$$P(A^c) = P(\cup_{k=1}^{n-1} \overline{A_k}) \leq \sum_{k=1}^{n-1} P(\overline{A_k}) \to 0 \text{ as } N \to \infty \quad \square.$$

We will use this result later on to prove symmetry of equivalence classes.

We now define

$$\Pi_c = P\left(\begin{array}{l|l} \text{transition from } K \text{ to } K-1 \text{ lines} & \text{there are currently } c \text{ parents} \\ \text{of descent occurs via coalescence} & \text{in the line of descent} \end{array} \right).$$

As in the previous chapter we need to determine the probability that a particular decrease in the number of lines of descent occurs because of coalescence. So we need to evaluate the Π_c's.

A standard first-step analysis of the transition probabilities for $\Lambda(t)$ gives

$$\begin{aligned} \Pi_c &= p^c \Pi_c + \binom{c}{1} p^{c-1}(1-p)\Pi_{c-1} + \binom{c}{2} p^{c-2}(1-p)^2 \Pi_{c-2} + \cdots \\ &\quad + \binom{c}{c-1} p(1-p)^{c-1}\Pi_1 + (1-p)^c \Pi_0 + O(N^{-1}). \end{aligned} \tag{4.9}$$

Lemma 4.3

$$\Pi_c = \Pi_0 + O(N^{-1}) \text{ for all } c > 0.$$

Proof:

From (4.9) we have

$$\begin{aligned} (1-p^c)\Pi_c &= \binom{c}{1} p^{c-1}(1-p)\Pi_{c-1} + \binom{c}{2} p^{c-2}(1-p)^2 \Pi_{c-2} + \cdots \\ &\quad + (1-p)^c \Pi_0 + O(N^{-1}). \end{aligned} \tag{4.10}$$

So $\Pi_1(1-p) = (1-p)\Pi_0 + O(N^{-1}) \Rightarrow \Pi_1 = \Pi_0 + O(N^{-1})$.

We now proceed by induction. So suppose the result is true for all $d < c$, then (4.10) gives

$$\begin{aligned}
(1-p^c)\Pi_c &= \begin{pmatrix} c \\ 1 \end{pmatrix} p^{c-1}(1-p)[\Pi_0 + O(N^{-1})] \\
&\quad + \begin{pmatrix} c \\ 2 \end{pmatrix} p^{c-2}(1-p)^2[\Pi_0 + O(N^{-1})] + \\
&\quad \cdots + (1-p)^c\Pi_0 + O(N^{-1}).
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \quad (1-p^c)\Pi_c &= \Pi_0\left[\begin{pmatrix} c \\ 1 \end{pmatrix} p^{c-1}(1-p) + \begin{pmatrix} c \\ 2 \end{pmatrix} p^{c-2}(1-p)^2 + \cdots\right. \\
&\qquad\qquad \left. \cdots + (1-p)^c\right] + O(N^{-1}). \\
\Rightarrow \quad (1-p^c)\Pi_c &= \Pi_0(1-p^c) + O(N^{-1}) \\
\Rightarrow \quad \Pi_c &= \Pi_0 + O(N^{-1}) \text{ as required} \qquad \Box.
\end{aligned}$$

So it is sufficient to evaluate Π_0, but first we must obtain a more accurate expression for Π_1. Once again by a first-step argument on the transition probabilities for $\Lambda(\cdot) = (K, 1)$ we get

$$\begin{aligned}
\Pi_1 &= \frac{p(K-1)}{N} + \frac{p(K-1)(T-1)(2-p)}{N}\Pi_2 \\
&\quad + \Pi_1\left[p + (1-p)\frac{TK}{N} + \frac{TK(1-p)^2}{N} + \frac{p(1-p)K}{N}\right. \\
&\quad \left. - \frac{p(K-1)T(2-p)}{N} + \frac{p(1-p)(K-1)}{N}\right] \\
&\quad + \Pi_0(1-p)\left[1 - \frac{TK(2-p)}{N} - \frac{pK}{N}\right] + O(N^{-2}). \qquad (4.11)
\end{aligned}$$

After substituting for Π_2 and a little simplification (details of which can be found in Appendix II, section D), we find

$$\Pi_1 = \frac{p(K-1)}{(1-p)N} + \Pi_0 \left[1 + \frac{p(K-1)}{N} - \frac{p(2-p)(K-1)}{(1-p)N}\right] + O(N^{-2}). \quad (4.12)$$

Hence

$$\begin{aligned}\Pi_0 &= \frac{(1-p)K(K-1)T}{N^2} && (4.13)\\ &+\Pi_0 \left[1 - \frac{KT(2-p)}{N} + \binom{T}{2} \frac{K(K-1)(2-p)^2}{N^2}\right.\\ &\quad \left. + \frac{TK(K-1)(1-p)}{N^2} - \frac{TK}{N^2}[Tp(2-p) - p(1-p)]\right]\\ &+\Pi_1 \left[\frac{(2-p)TK}{N} - \frac{TK\theta}{N^2} - 2\binom{T}{2}\frac{K(K-1)(2-p)^2}{N^2}\right.\\ &\quad \left. - \frac{2TK(K-1)(1-p)}{N^2} + \frac{TK}{N^2}[Tp(2-p) - p(1-p)]\right]\\ &+\Pi_2 \left[\binom{T}{2}\frac{K(K-1)(2-p)^2}{N^2}\right] + O(N^{-3}). && (4.14)\end{aligned}$$

Substitution for Π_1 and Π_2 and much simplification (which again can be found in Appendix II, section D), reveals

$$\Pi_0 = \frac{K-1}{(1-p)\theta + (K-1)} + O(N^{-1}).$$

This is the same as in the uncorrelated case but with θ replaced by $(1-p)\theta$. It is also identical to the result in Chapter 3. So, informally, we have the following result:

Theorem 4.4 *As $N \to \infty$ the conditional distribution of allele frequencies for the model presented here with parameters p and θ (representing correlation between choice of parents and mutation probability respectively) is the same as for a model in which there is no such correlation (ie. $p = 0$) and where θ is replaced by $\theta' = (1-p)\theta$.*

Now recall that we have already shown that the process $\Lambda(t)$ will visit $(K_t, 0)$ before K_t decreases, for all K_t. Define B_K as the event that the sample of n taken from the population at equilibrium contains K types, with μ_1 of the oldest, μ_2 of the second oldest,$\ldots, \mu_K$ of the youngest type. Then if A denotes the event that $\Lambda(t)$ visits $(K_t, 0)$ before K_t decreases to K_{t-1} for all K_t, then we have:

$$\begin{aligned} P(B_{K;\mu_1,\ldots,\mu_K}) &= P(B_{K;\mu_1,\ldots,\mu_K} \cap A) + P(B_{K;\mu_1,\ldots,\mu_K} \cap \overline{A}) \\ &= P(A)P(B_{K;\mu_1,\ldots,\mu_K} \mid A) + P(\overline{A})P(B_{K;\mu_1,\ldots,\mu_K} \mid \overline{A}) \end{aligned}$$

and so

$$\lim_{N\to\infty} P(B_{K;\mu_1,\ldots,\mu_K}) = \lim_{N\to\infty} P(B_{K;\mu_1,\ldots,\mu_K} \mid A).$$

This implies that the distribution of $B_{K;\mu_1,\ldots,\mu_K}$ is identical to that obtained in Chapter 3 and hence it is the same as for the uncorrelated case with θ replaced by $(1-p)\theta$ (see for example Donnelly and Tavaré (1986)).

It follows that

$$\lim_{N\to\infty} P(B_{K;\mu_1,\ldots,\mu_K}) = \frac{((1-p)\theta)^K}{((1-p)\theta)_{(n)}} \frac{n!}{\mu_K(\mu_K+\mu_{K-1})\cdots(\mu_K+\cdots+\mu_1)}$$

and

$$\lim_{N\to\infty} P(K \text{ types in sample;} \mu_1 \text{ of one type,} \mu_2 \text{ of another,} \ldots, \mu_K \text{ of the } K^{th})$$

$$= \frac{((1-p)\theta)^K}{((1-p)\theta)_{(n)}} \frac{n!}{\mu_1\mu_2\cdots\mu_K\beta_1!\beta_2!\cdots\beta_K!}$$

where β_j is the number of alleles represented j times, $j = 1, 2, \cdots, K$. For a more detailed look at the consequences of this result please refer to Chapter 3.

Chapter 5

CORRELATION AND VARIABILITY IN BIRTH PROCESSES

5.1 Introduction

Definition 5.1 *A birth process* $\{X(t), t \geq 0\}$ *with birth rates* $\lambda_1, \lambda_2, \ldots$, *is a continuous time Markov chain with state space* $\{1, 2, 3, \ldots\}$ *and infinitesimal transition rates, for* $i \neq j$

$$\lim_{h \downarrow 0} P(X(t+h) = j \mid X(t) = i) = \begin{cases} \lambda_i & \textit{if } j = i+1 \\ 0 & \textit{otherwise.} \end{cases}$$

We suppose that the process starts at a value $N > 0$ at time 0. It is common to think of $X(t)$ as representing the number of individuals present at time t and so upward jumps in $X(\cdot)$ correspond to births, hence the name.

Throughout this section we will view the birth process $X(\cdot)$ in a particular way. Specifically, we regard a process which starts at a value N as consisting of N individuals. Subsequently, whenever a birth occurs we shall refer to the new individual as the 'offspring' of one of the existing individuals (the parent being chosen randomly, with the assignment of parents being independent for each birth). Consequently we refer to the family of the i^{th} individual ($i = 1, 2, \ldots, N$) as the i^{th} initial individual and all its (not necessarily direct) descendants. Birth processes arise as models in a wide variety of settings.

In the most general form of the process we allow the birth rates to take any value. However, in general, although one can write down explicit expressions for the probability distributions of $X(t)$, they fall short of being illuminating and little useful information can be obtained. In particular, expressions for the moments are far from simple. In order to make further progress one must make assumptions regarding the form of the transition rates. The simplest assumption to place is that they are constant, ie.

$$\lambda_n = \lambda \quad \text{for all } n.$$

Now things become relatively simple. Specifically, we have a Poisson process. There are many applications for such a model (we refer those interested to Karlin and Taylor (1975)).

A common way to allow the birth rates to vary is to suppose that they depend linearly on the number of individuals present, ie.

$$\lambda_n = n\lambda, \qquad n = 1, 2, \ldots \, .$$

This is usually refered to as the linear birth (or 'Yule') process. The commonest application for this formulation is to model population growth (in bacteria for instance). Here it is reasonable to suppose that each individual is having offspring at rate λ and this leads naturally to the birth rates just specified. We can write down specifically the distribution of $X(t)$ (given that we start with N individuals) as:

$$P(X(t) = n \mid X(0) = N) = \left(\begin{array}{c} n-1 \\ N-1 \end{array} \right) e^{-N\lambda t}(1 - e^{-\lambda t})^{n-N} .$$

Note that this is the Negative Binomial distribution and if we start with 1 individual we get a Geometric distribution. Other applications include the early stages of an epidemic, where $X(t)$ records the number of people who have been infected by time t.

However it is easy to see why the birth rates of a process might vary non-linearly. The environment might be such that there is a limit to available resources (food or space for instance) and so it is natural to suppose that the birth rates will begin to fall off as the population size increases. Alternatively, birth rates (per individual) might increase as the population grows due to improvements in the environment (extra warmth or more mates for instance).

For the remainder of this chapter we will explore the behaviour of non-linear birth processes. We do this in order to better understand the relationship between qualitative aspects of the process behaviour and broad features of its parameter values. In particular we will consider correlations in family sizes and use this to investigate a conjecture due to Faddy (1990) regarding the relationship between the variance of linear and non-linear birth processes.

5.2 Correlation in Birth Processes

In the remainder of this chapter we will use the concept of monotonicity. A full consideration of this can be found within Liggett (1985), so we content ourself here with a brief summary. Let a stochastic process $\{W(t),\ t \geq 0\}$ be defined on a partially ordered statespace E; for example $E = \{0, 1\}^S$ where S is a countable set, whence a natural partial order is set inclusion. A function

f is then said to be increasing (decreasing) on E if

$$e_1 \le e_2 \ \Rightarrow\ f(e_1) \le \ (\ge)\, f(e_2) \ \text{ for all } e_1 \le e_2 \text{ in } E.$$

If $W(t)$ and $Z(t)$ are E-valued random variables we then say that X is stochastically greater than Z (written $W \overset{\text{st}}{\geq} Z$) if

$$E[f(X(t))] \ge E[f(Z(t))] \quad \text{for all increasing functions } f \text{ on } E.$$

We then define W to be a monotone process if the following is true:
If $W'(\cdot)$ and $W''(\cdot)$ are two versions of $W(\cdot)$ then

$$W'(0) \overset{\text{st}}{\geq} W''(0) \ \Rightarrow\ W'(t) \overset{\text{st}}{\geq} W''(t) \ \text{ for all } t > 0.$$

The most common way of demonstrating monotonicity is via coupling.

We now return to birth processes.

Definition 5.2 *A sequence $\lambda_1, \lambda_2, \ldots$ is said to be* *superlinear* *if λ_n/n, $n = 1, 2, \ldots$, is non-decreasing and* *sublinear* *if λ_n/n, $n = 1, 2, \ldots$, is non-increasing.*

It is easy to show that convexity of the sequence $\lambda_1, \lambda_2, \ldots$ implies superlinearity and concavity implies sublinearity.

Throughout this section we will take $X(0) \equiv N > 1$, the initial value of the birth process, to be fixed. Further we will restrict our attention henceforth to non-explosive processes (so that we assume $\sum_{i=1}^{\infty} \lambda_i^{-1} = \infty$). We augment the birth process by assigning parents to newly born individuals as described earlier and label the initial individuals from $\{1, 2, \ldots, N\}$. We consider the process $\{(X_1(t), \ldots, X_N(t)),\ t > 0\}$ where $X_i(t)$ is the number of individuals present at time t who are descended, possibly via some intermediate individuals, from the ith initial individual, with the founding individuals counted as descendants of themselves. More formally, $\mathbf{X}(\cdot) = (X_1(\cdot), X_2(\cdot), \ldots, X_N(\cdot))$ is

Markov with state space $E = \{1, 2, 3, \ldots\}^N$, $X_i(0) = 1$, $i = 1, 2, \ldots, N$, and the non-zero off diagonal elements of its generator matrix are given by

$$q_{(x_1,\ldots,x_i,\ldots,x_N)(x_1,\ldots,x_i+1,\ldots,x_N)} = \lambda_{x_1+\cdots+x_N} \frac{x_i}{x_1 + \cdots + x_N}, \qquad i = 1, 2, \ldots, N. \tag{5.1}$$

We now consider the presence of correlation between the sizes of the N families. In particular we prove the following result:

Theorem 5.3 *For the family size process* $\{\mathbf{X}(t),\ t > 0\}$ *defined above, for* $i, j = 1, 2, \ldots, N$, $i \neq j$,
(i) Cov($X_i(t), X_j(t)$*)* ≥ 0 *if the birth rates are superlinear,*
(ii) Cov($X_i(t), X_j(t)$*)* ≤ 0 *if the birth rates are sublinear.*

Remark Note that since $X(0) = N$ the actual values taken by $\lambda_1, \lambda_2, \ldots, \lambda_{N-1}$ have no effect on the behaviour of $X(\cdot)$. Hence, for convenience only, we will assume that these values are defined such that the super(sub)-linearity of the sequence $\lambda_N, \lambda_{N+1}, \ldots$ is preserved.

Proof:

We first prove the superlinear case. Consider two versions $X(\cdot)$ and $\tilde{X}(\cdot)$ of the Markov chain $\mathbf{X}(\cdot)$ with transition rates as in (5.1) with $X(0) = (x_1, \ldots, x_N)$, $\tilde{X}(0) = (\tilde{x}_1, \ldots, \tilde{x}_N)$ and suppose $\tilde{x}_i \geq x_i$ for all $i = 1, \ldots, N$. We aim to demonstrate monotonicity and so we define our statespace to be $E \equiv \{1, 2, \ldots\}^N$. Note that we can define a partial order on E as follows:

$$(x_1, \ldots, x_N) \leq (y_1, \ldots, y_N) \text{ iff } x_i \leq y_i \text{ for all } i = 1, \ldots, N. \tag{5.2}$$

Returning to our two versions $X(\cdot)$ and $\tilde{X}(\cdot)$ we see that since the birth rates are superlinear the rate at which $X_i(\cdot)$ jumps is smaller than the rate at which $\tilde{X}_i(\cdot)$ jumps for all $i = 1, \ldots, N$. We can easily couple $X(\cdot)$ and $\tilde{X}(\cdot)$ so that families in the latter jump whenever corresponding families in the former do (as well as possibly jumping at other times). Thus we ensure that if $X(0) \leq \tilde{X}(0)$ then $X(t) \leq \tilde{X}(t)$ for all $t > 0$. This implies that our original process $\mathbf{X}(\cdot)$ defined in (5.1) is monotone.

By definition, all transitions of $\mathbf{X}(\cdot)$ are to comparable states in the partial order (5.2). Specifically, if the current state is $x = (x_1, \ldots, x_N)$ and the process jumps, it will be to a state x' of the form $x' = (x_1, \ldots, x_i + 1, \ldots, x_N)$. It will always be the case that $x' \geq x$ in the partial order on E, and in particular x' and x are certainly comparable. Hence it follows (as in Liggett 1985 II.2.18) that if Ω is the generator of the process $\mathbf{X}(\cdot)$ and f and g are increasing functions from E into $\mathbb{R}$ which are in the domain of Ω, then

$$\Omega f g \geq f \Omega g + g \Omega f. \tag{5.3}$$

We now appeal to a correlation inequality due to Harris (1977) (see also Liggett 1985, Theorem II.2.14) which effectively says that for a monotone process $X(\cdot)$ whose generator satisfies (5.3), we have

$$E[f(X(t))g(X(t))] \geq E[f(X(t))]\, E[g(X(t))],$$

whenever f and g are increasing functions on E.

If we could now set $f((x_1, \ldots, x_N)) = x_i$ and $g((x_1, \ldots, x_N)) = x_j$ $(i \neq j)$ we would have demonstrated the correlation required for the proof. But unfortunately we need the generator Ω to be bounded. Furthermore we also require the functions f and g to be bounded. To deal with this problem we introduce the following approximation argument.

Let $E^* \equiv \{1, 2, \ldots, \infty\}^N$ have partial order as in (5.2) but with the additional proviso that $e \leq \infty$ for all $e \in \{1, 2, \ldots\}$. Define $X^M(\cdot)$ to be the process with jump rates as in (5.1) whenever $\sum_{i=1}^N x_i < M$ and for

$(y_1, \dots, y_N) \neq (x_1, \dots, x_N)$, $\sum_{i=0}^N x_i \geq M$

$$q_{(x_1,\dots,x_N)(y_1,\dots,y_N)} = \begin{cases} \lambda_M & \text{if } y_1 = \cdots = y_N = \infty \\ 0 & \text{otherwise.} \end{cases}$$

While $\sum_{i=1}^N \tilde{x}_i < M$ (and hence $\sum_{i=1}^N x_i < M$) we may use the same coupling as before to keep the processes ordered. There are now two additional cases however. Firstly, if $\sum_{i=1}^N x_i < M \leq \sum_{i=1}^N \tilde{x}_i$, then make $\tilde{X}^M(\cdot)$ jump to $(\infty, \dots, \infty)$ no later than the first jump of $X^M(\cdot)$ (recall $\lambda_M \geq \lambda_{\sum_{i=1}^N x_i}$ by assumption). Secondly, if $M \leq \sum_{i=1}^N x_i \leq \sum_{i=1}^N \tilde{x}_i$, then make both processes jump to $(\infty, \dots, \infty)$ simultaneously. As before it follows from this that $X^M(\cdot)$ is monotone and since all its jumps are to comparable states in the partial order, the generator condition (5.3) obtains.

For $K > 0$ we now define $f_K, g_K : E^* \to \mathbb{R}$ by

$$f_K((x_1, \dots, x_N)) = x_i \wedge K, \quad g_K((x_1, \dots, x_N)) = x_j \wedge K \quad (i \neq j).$$

The process $X^M(\cdot)$ (with $X^M(0) = (1, \dots, 1)$ a.s.) now satisfies the conditions of (Liggett's version of) Harris' theorem. Thus for any $t \geq 0$ we have

$$E(f_K(X^M(t))g_K(X^M(t))) \geq E(f_K(X^M(t)))E(g_K(X^M(t))) \tag{5.4}$$

Since the original birth process $X(\cdot)$ is non-explosive,

$$\tau_M = \inf\{t : \sum_{i=1}^N X_i(t) = M\} \Rightarrow \infty$$

as $M \to \infty$ (and $\Rightarrow$ denotes weak convergence), from which it follows that $X^M(\cdot) \Rightarrow X(\cdot)$ as $M \to \infty$. Let $M \to \infty$ in (5.4) to obtain

$$E(f_K(X(t))g_K(X(t))) \geq E(f_K(X(t)))E(g_K(X(t))).$$

Now let $K \to \infty$ and use monotone convergence to write

$$E(X_i(t)X_j(t)) \;=\; E(f(X(t))g(X(t)))$$

$$\geq \; E(f(X(t)))E(g(X(t)))$$

$$= \; E(X_i(t))E(X_j(t)) \;\text{ for all } i \neq j.$$

as required.

We now consider the case of sublinear birth rates. Unfortunately this case is not so straightforward. We will again wish to demonstrate monotonicity via a coupling argument, but before we can proceed with this we need to reformulate our birth process. Firstly note that $X_1(t), \ldots, X_N(t)$ are exchangeable, so we can write

$$\mathrm{Cov}(X_i(t), X_j(t)) = \mathrm{Cov}(X_1(t), X_2(t)) \quad i, j = 1, 2, \ldots, N, \; i \neq j.$$

Next, define $\mathbf{Y}(\cdot) = (Y_1(\cdot), Y_2(\cdot))$ by

$$Y_1(t) = X_1(t), \;\; Y_2(t) = X_2(t) + \cdots + X_N(t).$$

Note that $\mathbf{Y}(\cdot)$ is a Markov process defined on a state space $E' = \{1, 2, \ldots\}^2$.

Again we now consider two versions $Y(\cdot)$ and $\tilde{Y}(\cdot)$ of $\mathbf{Y}(\cdot)$ with $Y(0) = (y_1, y_2)$, $\tilde{Y}(0) = (\tilde{y}_1, \tilde{y}_2)$ but now $\tilde{y}_1 \geq y_1$, $\tilde{y}_2 \leq y_2$. Assume further that $\tilde{y}_1 + \tilde{y}_2 \geq y_1 + y_2$. We define the non-zero off-diagonal entries of the generator matrix of the process $(Y, \tilde{Y})$ to be (writing $\Lambda_n = \lambda_n / n$):

$$\begin{aligned}
q_{(y_1,y_2)(\tilde{y}_1,\tilde{y}_2),(y_1+1,y_2)(\tilde{y}_1,\tilde{y}_2)} &= y_1 \Lambda_{y_1+y_2} \\
q_{(y_1,y_2)(\tilde{y}_1,\tilde{y}_2),(y_1,y_2)(\tilde{y}_1+1,\tilde{y}_2)} &= \tilde{y}_1 \Lambda_{\tilde{y}_1+\tilde{y}_2} \\
q_{(y_1,y_2)(\tilde{y}_1,\tilde{y}_2),(y_1,y_2+1)(\tilde{y}_1,\tilde{y}_2+1)} &= \tilde{y}_2 \Lambda_{\tilde{y}_1+\tilde{y}_2} \\
q_{(y_1,y_2)(\tilde{y}_1,\tilde{y}_2),(y_1,y_2+1)(\tilde{y}_1,\tilde{y}_2)} &= y_2 \Lambda_{y_1+y_2} - \tilde{y}_2 \Lambda_{\tilde{y}_1+\tilde{y}_2}
\end{aligned} \tag{5.5}$$

whenever $y_1 \leq \tilde{y}_1$, $y_2 \geq \tilde{y}_2$, $y_1 + y_2 < \tilde{y}_1 + \tilde{y}_2$. If $y_1 \leq \tilde{y}_1$, $y_2 \geq \tilde{y}_2$, $y_1 + y_2 =$

$\tilde{y}_1 + \tilde{y}_2 = k$ say, the generator matrix has non-zero off-diagonal entries

$$\begin{aligned} q_{(y_1,y_2)(\tilde{y}_1,\tilde{y}_2),(y_1+1,y_2)(\tilde{y}_1+1,\tilde{y}_2)} &= y_1\Lambda_k \\ q_{(y_1,y_2)(\tilde{y}_1,\tilde{y}_2),(y_1,y_2+1)(\tilde{y}_1,\tilde{y}_2+1)} &= \tilde{y}_2\Lambda_k \\ q_{(y_1,y_2)(\tilde{y}_1,\tilde{y}_2),(y_1,y_2+1)(\tilde{y}_1+1,\tilde{y}_2)} &= \Lambda_k(y_2 - \tilde{y}_2) \;=\; \Lambda_k(\tilde{y}_1 - y_1). \end{aligned} \tag{5.6}$$

Thus, when $y_1 + y_2 < \tilde{y}_1 + \tilde{y}_2$ the first components of both processes jump independently, but $Y_2(\cdot)$ jumps whenever $\tilde{Y}_2(\cdot)$ jumps, and at some additional times. If $y_1 + y_2 = \tilde{y}_1 + \tilde{y}_2$, the birth processes $Y_1(\cdot) + Y_2(\cdot)$ and $\tilde{Y}_1(\cdot) + \tilde{Y}_2(\cdot)$ jump simultaneously and if the additional individual joins the first(second) component in the $Y(\cdot)$ (respectively $\tilde{Y}(\cdot)$) process, it joins the same component in the $\tilde{Y}(\cdot)$ (respectively $Y(\cdot)$) process. We construct the coupling like this to ensure that, with the given initial conditions, we have

$$Y_1(t) \le \tilde{Y}_1(t), \quad \tilde{Y}_2(t) \le Y_2(t) \qquad \text{for all } t. \tag{5.7}$$

If $y_1 \le \tilde{y}_1,\ y_2 \ge \tilde{y}_2$ but $y_1 + y_2 > \tilde{y}_1 + \tilde{y}_2$ use the above construction (ie. (5.5)) with the role of tildes and non-tildes swapped and the roles of components 1 and 2 swapped to obtain the ordering (5.7) once more.

We now wish to invoke Harris' theorem to draw conclusions about negative correlations. As it stands this is not possible. The key is to define a partial order on E' as follows:

$$(x_1, x_2) \preceq (y_1, y_2) \ \text{ iff } \ x_1 \le y_1,\ x_2 \ge y_2.$$

The above construction then ensures that $Y(0) \preceq \tilde{Y}(0)$ implies that $Y(t) \preceq \tilde{Y}(t)$ for all t, so that with this partial order and sublinear birth rates, the process $Y(\cdot)$ with jump rates (5.5 and 5.6) is monotone. Again, all jumps are to comparable states in the partial order, so the generator condition (5.3) still obtains. (With the partial order $\preceq$, increasing functions are increasing on the first component and decreasing on the second, in the usual sense.)

We now proceed in a very similar manner to before. First we note that the generator (5.5 and 5.6) may still be unbounded, so another approximation argument is used. Denote by $Y^M(\cdot)$ the process with jump rates (5.5 and 5.6) with the birth rates λ_i, $i = 1, 2, \ldots$, replaced by $\lambda_i \wedge \lambda_M$, $i = 1, 2, \ldots$. (So Y^M has bounded generator.) This new sequence is still sublinear so that by the above construction, $Y^M(\cdot)$ is still monotone, and of course (5.3) still obtains. Now define $f_K, g_K : E \to \mathbb{R}$ by

$$f_K(y_1, y_2) = y_1 \wedge K, \quad g_K(y_1, y_2) = -(y_2 \wedge K).$$

Since f_K and g_K are both increasing in the partial order $\preceq$, Harris' theorem gives

$$E[f_K(Y^M(t))g_K(Y^M(t))] \geq E[f_K(Y^M(t))]\, E[g_K(Y^M(t))].$$

Let $M \to \infty$, multiply through by -1, and then let $K \to \infty$ to obtain

$$E(Y_1(t)Y_2(t)) \leq E(Y_1(t))E(Y_2(t))$$

as required. □

Recall that

$$Y_1(t) = X_1(t), \;\; Y_2(t) = X_2(t) + \cdots + X_N(t)$$

and the $X_i(t)$ are exchangeable random variables. Hence we have shown the existence of positive (negative) correlations in family sizes for processes with super(sub)-linear birth rates. We now aim to use this to provide a proof for a conjecture due to Faddy (1990).

5.3 Relative Variation and Faddy's Conjecture

We will now consider Faddy's conjecture for birth processes. If a birth process $X(\cdot)$ has linear birth rates ($\lambda_n = n\lambda$) then $X(t)$ has a Negative Binomial distribution and so if $X(0) = N$ we have:

$$EX(t) = Ne^{\lambda t},$$
$$\mathrm{Var}X(t) = Ne^{\lambda t}(e^{\lambda t} - 1) = EX(t)\left[\frac{EX(t)}{N} - 1\right],$$

(see for example Cox and Miller (1965) Chapter 4). As a measure of the qualitative behaviour of a birth process, Faddy (1990) defines the relative variation as follows:

$$V(t) = \frac{\mathrm{Var}X(t)}{EX(t)[\frac{EX(t)}{N} - 1]}. \tag{5.8}$$

In his paper Faddy makes some numerical calculations of this quantity for a class of piece-wise linear forms for λ_n. He finds that there is always more (relative to the linear case) variation for convex λ_n and less for concave λ_n and this leads him to conjecture that this may always be the case. However a proof of the result in this form has remained elusive and so we shall now spend a little time to obtain more insight into relative variation.

Given a birth process $X(t)$ which starts with N individuals (ie. $X(0) = N$), we write $X(\cdot)$ as above, so

$$X(t) = (X_1(t), X_2(t), \ldots, X_N(t))$$

where $X_i(\cdot)$ is the process constructed by considering the ith individual and its descendants. As before we will refer to the component $X_i(\cdot)$ as the family of the ith individual.

Faddy's conjecture involves a consideration of the relative variation. The next lemma presents an equivalent condition. (Note that for the rest of this chapter we shall drop the parameter t whenever convenient.)

Lemma 5.4 *For a birth process $X(t)$ defined as previously, the following two conditions are equivalent:*

$$(i) \qquad V(t) \geq (\leq)\, 1$$

$$(ii) \quad \frac{E[X(t)(X(t)+1)]}{N(N+1)} \geq (\leq) \left(\frac{EX(t)}{N}\right)^2 \tag{5.9}$$

Proof:

$$V(t) \geq 1 \;\Leftrightarrow\; \frac{EX^2 - [EX]^2}{EX[\frac{EX}{N} - 1]} \geq 1$$

$$\Leftrightarrow\; E[X^2] - [EX]^2 \geq \frac{[EX]^2}{N} - EX$$

$$\Leftrightarrow\; E[X^2] + EX \geq \left(\frac{N+1}{N}\right)[EX]^2$$

$$\Leftrightarrow\; \frac{E[X(X+1)]}{N+1} \geq \frac{[EX]^2}{N}$$

$$\Leftrightarrow\; \frac{E[X(X+1)]}{N(N+1)} \geq \left(\frac{EX}{N}\right)^2$$

The reversed inequality is directly analogous. □

Clearly, the right-hand side of (5.9) (ie. $\left(\frac{EX}{N}\right)^2$) represents the square of the expected family size of a single one of the initial individuals (ie. by exchangeability $(EX_1)^2$). But is there a more natural representation of the left-hand side? The next lemma suggests that there is.

Lemma 5.5 *As before, write the birth process as $X(t) = (X_1(t), \ldots, X_N(t))$. Then*

$$\frac{E[X(X+1)]}{N(N+1)} = E(X_1X_2) \tag{5.10}$$

The interpretation of this result is illuminating. $E(X_1X_2)$ is the expected size of the product of the family size of two distinct initial individuals. Hence, the relative variation is in fact a consideration of the correlation between family sizes of initial individuals. The relative variation is greater than one exactly when the family sizes are positively correlated, and less than one when the sizes are negatively correlated.

Before we can continue with the proof of Lemma 5.5 we need the result below:

Lemma 5.6 *Let P_M denote the probability that two individuals drawn uniformly at random from the population at time t are from different families, conditional on $\{X(t) = M\}$. Then*

$$P_M = \frac{M+1}{M-1}\frac{N-1}{N+1}. \tag{5.11}$$

Proof:

The result is clear for $M = N$, since then all families have only one member and so $P_M = 1$. We proceed by induction on M.

So suppose (5.11) is true for M and note that we can get $M + 1$ individuals by adding 1 individual to one of the N existing families, (the new individual is equally likely to be an offspring of any one of the existing M individuals).

Suppose the family sizes (with M individuals present) are $F_1, F_2, \ldots, F_N$. Then if we let $\mathcal{A}$ denote the event that when we randomly select two individuals from $M+1$ we don't pick the new (ie. $M+1^{\text{th}}$) individual, and $\mathcal{B}$ be the event that when we randomly select two individuals they are from different families and one of them is the new $(M+1^{\text{th}})$ individual, we have

$$P_{M+1} \;=\; P_M \times P(\mathcal{A}) + P(\mathcal{B})$$

$$= \frac{\binom{M}{2}}{\binom{M+1}{2}} P_M + \left[1 - \frac{\binom{M}{2}}{\binom{M+1}{2}} \right] \sum_{i \neq j} P(i,j)$$

(where $P(i,j)$ denotes the probability that the first individual

is in family i and the second is in family j.)

$$= \frac{M-1}{M+1} P_M + \frac{2}{M+1} \sum_{i \neq j} P(i,j)$$

$$= \frac{M-1}{M+1} P_M + \frac{2}{M+1} \sum_{i \neq j} \sum_{F_1,\ldots,F_N} P(i,j \mid X_1 = F_1, \ldots, X_N = F_N)$$

$$\times P(X_1 = F_1, \ldots, X_N = F_N \mid M)$$

(in an obvious notation)

$$= \frac{M-1}{M+1} P_M + \frac{2}{M+1} \sum_{i \neq j} \sum_{F_1,\ldots,F_N} \frac{F_i}{M} \frac{F_j}{M} P(X_1 = F_1, \ldots, X_N = F_N \mid M)$$

$$= \frac{M-1}{M+1} P_M + \frac{2}{M+1} \sum_{i \neq j} E \left[\frac{X_i}{M} \frac{X_j}{M} \mid M \right]$$

$$= \frac{M-1}{M+1} P_M + \frac{2}{M+1} \frac{N(N-1)}{M^2} E[X_1 X_2 \mid M]$$

by exchangeability.

Now note that

$$\begin{aligned}
P_M &= \sum_{i<j} \sum_{v,w=0}^{M} \frac{vwP(X_i = v,\ X_j = w \mid X = M)}{\begin{pmatrix} M \\ 2 \end{pmatrix}} \\
&= \sum_{i<j} \frac{E[X_i X_j \mid X = M]}{\begin{pmatrix} M \\ 2 \end{pmatrix}} \\
&= \frac{\begin{pmatrix} N \\ 2 \end{pmatrix} E[X_1 X_2 \mid X = M]}{\begin{pmatrix} M \\ 2 \end{pmatrix}}.
\end{aligned}$$

So continuing from before we find

$$\begin{aligned}
P_{M+1} &= \frac{M-1}{M+1} P_M + \frac{2}{M+1} \frac{N(N-1)}{M^2} \frac{\begin{pmatrix} M \\ 2 \end{pmatrix} P_M}{\begin{pmatrix} N \\ 2 \end{pmatrix}} \\
&= P_M \left[\frac{M-1}{M+1} + \frac{2(M-1)}{M(M+1)} \right] \\
&= \frac{(M+1)+1}{(M+1)-1} \frac{N-1}{N+1} \quad \text{as required} \qquad \square.
\end{aligned}$$

We can now return to Lemma 5.5.

Proof(of Lemma 5.5):

Let $X(t) = M$.

Then, with P_M defined as before, we have

$$
\begin{aligned}
EX_1X_2 &= \sum_{M=N}^{\infty} E(X_1X_2 \mid X = M)P(X = M) \\
&= \sum_{M=N}^{\infty} \frac{\binom{M}{2}}{\binom{N}{2}} P_M P(X = M) \\
&= \sum_{M=N}^{\infty} \frac{\binom{M}{2}}{\binom{N}{2}} \frac{M+1}{M-1}\frac{N-1}{N+1} P(X = M) \\
&= \sum_{M=N}^{\infty} \frac{M(M+1)}{N(N+1)} P(X = M) \\
&= \frac{EX(X+1)}{N(N+1)} \quad \text{as required} \qquad \Box.
\end{aligned}
$$

So Faddy's conjecture about relative variation is simply a consideration of the correlation between the family sizes. If we now combine the lemmas in this section with theorem 5.3 we have shown the following:

Theorem 5.7 *Let $X(t)$ be a non-explosive birth process starting at $X(0) = N > 1$ with birth rates $\{\lambda_n : n = N, N+1, \ldots\}$, and define the relative variation to be*

$$V(t) = \frac{VarX(t)}{EX(t)\left[\frac{EX(t)}{N} - 1\right]},$$

then

(i) $\frac{\lambda_n}{n}$ increasing as n increases $\Rightarrow V(t) \geq 1$.

(ii) $\frac{\lambda_n}{n}$ *decreasing as* n *increases* $\Rightarrow V(t) \leq 1$.

For the sake of completeness we now wish to prove Faddy's conjecture for processes starting with $N = 1$ individual. Once again we will approach this via a consideration of correlations, but now we have only one family. To get round this we introduce the somewhat less natural concept of colonies.

We start with our birth process $\{X(t),\ t \geq 0\}$ with transition rates as in (5.1) and we use this to construct another Markov chain $\{(Z_1(t), Z_2(t)),\ t \geq 0\}$ with state space $\{0, 1, 2, \ldots\}^2$ and $Z_1(0) = Z_2(0) = 0$. Again we view transitions in our process as representing births and assign a parent, independently of all other events, to each new individual born into the birth process by choosing at random (uniformly) from among all the individuals present when the birth occurs. Further, we will allocate each individual born into the process into one of two colonies: each offspring of the single founding individual is assigned to a colony at random (each choice having probability $\frac{1}{2}$), independently of all other events, otherwise offspring join the colony of their parent and $Z_1(t)$, $Z_2(t)$ denote the number in each colony at time t. Thus the founding individual does not belong to either colony. Formally, the process $Z(\cdot) \equiv (Z_1(\cdot), Z_2(\cdot))$ is Markov and has generator matrix with non-zero off-diagonal elements given by

$$q_{(i_1,i_2)(j_1,j_2)} = \begin{cases} \lambda_{i_1+i_2+1} \frac{i_1+1/2}{i_1+i_2+1} & j_1 = i_1 + 1,\ j_2 = i_2, \\ \lambda_{i_1+i_2+1} \frac{i_2+1/2}{i_1+i_2+1} & j_1 = i_1,\ j_2 = i_2 + 1. \end{cases} \tag{5.12}$$

Observe that in the linear case, $Z_1(\cdot)$ and $Z_2(\cdot)$ are independent linear birth processes with immigration.

We can now prove a companion result to theorem (5.3):

Theorem 5.8 *For the process* $\{(Z_1(t), Z_2(t)),\ t \geq 0\}$ *defined above, and any* $t > 0$,

(i) $Cov(Z_1(t), Z_2(t)) \geq 0$ *if the birth rates* $\lambda_1, \lambda_2, \ldots$ *are superlinear,*
(ii) $Cov(Z_1(t), Z_2(t)) \leq 0$ *if the birth rates* $\lambda_1, \lambda_2, \ldots$ *are sublinear.*

Proof:

This is analogous to the proof of Theorem 5.3 and so we merely sketch the outline.

In the superlinear case we use a natural coupling argument and endow E with the obvious partial order to demonstrate monotonicity of $(Z_1(\cdot), Z_2(\cdot))$. Again jumps are to comparable states in the partial order so the generator condition (5.3) is satisfied. We then simply repeat the approximation argument from the proof of theorem 5.3, invoke Harris' theorem and conclude the result (i).

For the sublinear case we once again reverse part of the partial order and then a construction very similar to that in the latter half of theorem 5.3's proof shows that $(Z_1(\cdot), Z_2(\cdot))$ is monotone. After an analogous approximation argument and use of Harris' theorem we have completed the proof.

$\square$

We will now begin to relate this result to the final part of Faddy's conjecture. Recall that we have constructed our process $(Z_1(\cdot), Z_2(\cdot))$ from realisations of the original family-valued process $X(\cdot)$. A little thought reveals that conditional on the value of $X(t)$ the distribution of $(Z_1(t), Z_2(t))$ is independent of the birth rates (and of t), since it is solely dependent on the assignment of new individuals into colonies.

Lemma 5.9 *For the process* $\{(Z_1(t), Z_2(t)),\ t \geq 0\}$ *defined above, with* $X(t) = Z_1(t) + Z_2(t) + 1$ *and* $M = 1, 2, \ldots$
(i) $E(Z_1(t) \mid X(t) = M) = E(Z_2(t) \mid X(t) = M) = (M-1)/2,$
(ii) $E(Z_1(t) Z_2(t) \mid X(t) = M) = \frac{(M-1)(M-2)}{8}.$

Proof:

Part (i) follows by symmetry. The second part is clearly true for $M = 1$ and $M = 2$ since $Z_1(t)Z_2(t) = 0$ a.s. unless $X(t) \geq 3$. We will prove the result by induction and so assume the result is true for $M = m$. In what follows, consider choosing two individuals from those in the colonies at time t. Then

$$P(\text{chosen individuals are from different colonies} \mid X(t) = m)$$

$$= \frac{2E(Z_1(t)Z_2(t) \mid X(t) = m)}{(m-1)(m-2)}$$

$$= \frac{1}{4}$$

by the induction hypothesis. Now, when $X(t) = m + 1$, condition on whether or not the chosen pair contains the most recently born individual, and use the inductive hypothesis and symmetry, to obtain

$$P(\text{chosen individuals are from different colonies} \mid X(t) = m + 1)$$

$$= \binom{m-1}{2}\binom{m}{2}^{-1}\frac{1}{4} + \tag{5.13}$$

$$\left(1 - \binom{m-1}{2}\binom{m}{2}^{-1}\right)$$

$$\times 2P\left(\begin{array}{c}\text{most recently born individual is from} \\ \text{colony 1 and other individual} \\ \text{chosen is from colony 2}\end{array}\middle| X(t') = m\right)$$

(where t' is a time just before the $m+1$th individual is born). By conditioning on the values of $Z_1(t')$ and $Z_2(t')$ just before the birth of the most recent individual, and recalling that conditional on the value of $X(t')$ the Z's are independent of t', we may write this last probability as

$$E\left(\frac{Z_1(t') + \frac{1}{2}}{m}\frac{Z_2(t')}{m-1}\middle| X(t') = m\right)$$

$$= \quad m^{-1}(m-1)^{-1}\left(E(Z_1(t')Z_2(t') \mid X(t') = m) + \frac{1}{2}E(Z_2(t') \mid X(t') = m)\right)$$

$$= \quad \frac{m-2}{8m} + \frac{1}{4m} \;=\; \frac{1}{8}$$

from the inductive hypothesis and part (i). Substitute this into (5.13) to obtain

$$E(Z_1(t)Z_2(t) \mid X(t) = m+1)$$

$$= \quad \frac{m(m-1)}{2}$$

$$\times P(\text{chosen individuals are from different colonies} \mid X(t) = m+1)$$

$$= \quad \frac{1}{4}\frac{m(m-1)}{2}$$

as required. □

We may now conclude our proof of Faddy's Conjecture:

Corollary 5.10 *For a birth process* $\{X(t),\ t \geq 0\}$ *with birth rates* $\lambda_1, \lambda_2, \ldots$, *and* $X(0) = 1$,

(i) $Var(X(t)) \geq E(X(t))(E(X(t)) - 1)$ *if the birth rates are superlinear*

(ii) $Var(X(t)) \leq E(X(t))(E(X(t)) - 1)$ *if the birth rates are sublinear.*

Proof:

It follows from Lemma 5.9 that

$$E(Z_1(t)Z_2(t)) = \frac{1}{8}E((X(t)-1)(X(t)-2))$$

$$E(Z_1(t)) = E(Z_2(t)) = \frac{E(X(t)) - 1}{2}.$$

Thus

$$\begin{aligned}\mathrm{Cov}(Z_1(t), Z_2(t)) &= \frac{1}{8}E((X(t)-1)(X(t)-2)) - \frac{(E(X(t))-1)^2}{4} \\ &= \frac{1}{8}\left[\mathrm{Var}(X(t)) - E(X(t))(E(X(t))-1)\right]\end{aligned}$$

so that the result now follows from Theorem 5.8. □

It should be noted that it is possible to prove Faddy's conjecture for all $N \geq 1$ using this latter approach, but the colony process it uses seems less natural than the family process used in the earlier proof.

Having completed the proof of Faddy's conjecture we conclude by noting that it is possible to prove that, if the birth rates are super-linear, the existence of individuals in different families is positively correlated. We once again regard our process as consisting of $N > 1$ families and simply label the possible individuals within a particular family with the positive integers, so the nth individual born into family i is labelled (i, n) say. We then use a natural coupling to show that the process in this formulation is monotone. Next define $f_{i,n}$ to be 1 if the nth individual in family i is alive and 0 otherwise. Similarly define $g_{j,m}$ for the mth individual in family j. f and g are clearly increasing (and bounded) functions on our state space. It is now a straightforward application of Harris' theorem (with the usual approximation argument for the generator) to obtain the desired result.

It is clearly tempting to speculate that for sub-linear birth rates the existence of individuals in different families is negatively correlated (excluding the initial founder individuals). While the author suspects the truth of this statement, no proof has yet been found.

Chapter 6

CORRELATION INEQUALITIES

6.1 Positive Correlation Inequalities

A concept which arises naturally in many areas of applied probability is that of positive correlation. It is often intuitively reasonable to expect related events to be positively correlated. However, many of the results one expects to be straightforward are very hard to prove. A particularly useful result in this context has proved to be an inequality due to Fortuin, Kasteleyn and Ginibre (the so-called FKG Inequality). This has provided quick methods of proof for many previously very difficult results, as well as leading to the discovery of new ones. For full details of the result the interested reader is referred to Fortuin, Kasteleyn and Ginibre (1971). We shall give an explicit statement of the result in a form suitable for our future requirements.

Define a lattice Γ to be a partially ordered set in which any two elements x and y in Γ have a least upper bound $x \vee y$ and a greatest lower bound $x \wedge y$. A lattice is called distributive if the operations $\vee$ and $\wedge$ satisfy either of the following (equivalent) conditions:

$$x \wedge (y \vee z) \;=\; (x \wedge y) \vee (x \wedge z) \text{ for all } x, y, z \text{ in } \Gamma \tag{6.1}$$

$$x \vee (y \wedge z) \;=\; (x \vee y) \wedge (x \vee z) \text{ for all } x, y, z \text{ in } \Gamma. \tag{6.2}$$

We define a real-valued function f on such a set Γ to be increasing (decreasing) if for any ordered pair $x \leq y$ in Γ we have $f(x) \leq f(y)$ $(f(x) \geq f(y))$.

Now if μ is a positive finite measure on a partially ordered countable set Γ we define $\langle f \rangle$ to be

$$\langle f \rangle = \frac{\sum_{x \in \Gamma} \mu(x) f(x)}{\sum_{x \in \Gamma} \mu(x)}.$$

Then we have

Theorem 6.1 (FKG Inequality) *If Γ is a finite distributive lattice and μ is a positive measure on Γ satisfying the following condition:*
(a) For all x and y in Γ, $\mu(x \wedge y)\mu(x \vee y) \geq \mu(x)\mu(y)$,
then if f and g are both increasing (or decreasing) functions on Γ we have

$$\langle fg \rangle \geq \langle f \rangle \langle g \rangle.$$

A simple example of a finite distributive lattice is

$$\Omega = 2^X = \{\text{all subsets of } X\} \text{ where } X = X_1, X_2, \ldots, X_N \text{ (some N)}.$$

Here, for $A, B \in \Omega$ we define

$$A \wedge B = A \cap B \text{ and } A \vee B = A \cup B.$$

We shall return to this example later.

As was stated earlier, the FKG Inequality has been used to provide relatively simple proofs of results which were previously difficult. An example of this is the result first proved by Harris (1960) in the field of percolation models which now follows as an easy consequence. Percolation processes were originally introduced by Broadbent and Hammersley (1957) as a means of modelling the flow of liquid through a random medium. These models have proved useful in a wide variety of applications, such as petroleum flow in sandstone and as simple examples of critical phenomena in statistical mechanics. More details can be found in Kesten (1982) and Welsh (1986), but here we simply present a brief description of the salient features.

We start by supposing the existence of a finite regular lattice G (the infinite case is usually a straightforward extension). We then let V denote the set of vertices (or sites) of G, and E denote the set of edges (or bonds). There are two types of percolation: site percolation and bond percolation. Site percolation includes bond percolation as a special case (Fisher (1961)), and so we shall only consider the former here. In this model sites are either open or closed, and it is defined by introducing a random field μ on V where $\mu(A)$ is the probability that A is exactly the set of open sites. Hence we can define the distribution function F by $F(A) = \sum_{Y \supseteq A} \mu(Y)$, and so $F(A)$ is simply the probability that the set of sites A is open.

For events $\{A \text{ open}\}$ and $\{B \text{ open}\}$ to be positively correlated we would expect a statement like the following to be true:

$$P(A \text{ open} \mid B \text{ open}) \geq P(A \text{ open}) \quad \text{where } A, B \subseteq V,$$

or more generally

$$P(A \text{ open} \mid B \cup C \text{ open}) \geq P(A \text{ open} \mid B \text{ open}) \quad A, B, C \subseteq V.$$

Note that the above results are trivial in the case of classical percolation (which defines $\mu(A) = p^{|A|}(1-p)^{|V \setminus A|}$), but we will move on to give more interesting examples. Indeed these results will be true of any probability measure μ which obeys the conditions of the FKG Inequality, ie.

$$\mu(A \cup B)\mu(A \cap B) \geq \mu(A)\mu(B) \quad A, B \subseteq V.$$

To see this choose

$$f(X) = \begin{cases} 1 & A \subseteq X \\ 0 & \text{otherwise} \end{cases}$$

$$g(X) = \begin{cases} 1 & B \subseteq X \\ 0 & \text{otherwise.} \end{cases}$$

Note that in this case $\langle f \rangle = P(A \text{ open})$, similarly $\langle g \rangle = P(B \text{ open})$ and $\langle fg \rangle = P(A \cup B \text{ open})$. Thus the conclusion of the FKG Inequality gives us the required behaviour.

Moving on, if we now define

$$P(a \rightarrow b) = P(\exists \text{ a path of open sites from } a \text{ to } b)$$

and similarly $P(c \rightarrow d)$, for $a, b, c, d \in V$, we can write

$$P(a \rightarrow b) = \sum_{X \subseteq V} \mu(X) f_{a,b}(X)$$

where

$$f_{a,b}(X) = \begin{cases} 1 & \text{if } X \text{ contains a set of sites forming a} \\ & \text{path from } a \text{ to } b \\ 0 & \text{otherwise} \end{cases}$$

Clearly $f_{a,b}$ and $f_{c,d}$ are increasing functions (since $f_{a,b}(X) = 1$ and $X \subseteq Y$ implies $f_{a,b}(Y) = 1$), so we apply the FKG Inequality to deduce that $P(a \rightarrow b \mid c \rightarrow d) \geq P(a \rightarrow b)$. In other words, the probability that there exists a path of open sites from a to b, given that a similar path exists from c to d, is greater than the probability of existence of such a path given no prior knowledge. This is intuitively what one would expect, since the existence of a path from c to d tells us that more vertices than might be expected are open and this increases the likelihood of there being a similar path from a to b. This is just one example of the sort of information which can be deduced using the FKG Inequality and it is the result proved at some length directly by Harris (1960). There are obviously many others.

Many models in statistical mechanics are concerned with ferromagnetism. Ferromagnetism derives from the quantum mechanical spinning of electrons. The "spin" (and hence the magnetic moment) can be represented by an arrow which points up or down and which flips between the two orientations. So each site of the lattice G has a "spin" of either +1 or -1 and we interpret $\mu(A)$ as the probability that the set of sites with positive spins is exactly A. The most common such model is the classical Ising model (see Ellis (1985) Section IV for instance). It has μ given by

$$\mu(X) = e^{-\alpha|X|} - \frac{e^{-\alpha|X|-\beta e(X)}}{Z} \quad X \subseteq V,$$

where α, β, Z are constants and $e(X)$ is the number of edges having only one endpoint in X. Obviously we can view this as a percolation (with the two different spins corresponding to open and closed sites). It is easy to check that $e(X \cup Y) + e(X \cap Y) \leq e(X) + e(Y)$, and this implies that μ satisfies the conditions of the FKG Inequality, which can hence be applied. For instance, it can be used to prove the existence of infinite volume Gibbs states (for details see Ellis (1985) Section IV).

Another common area of application for the FKG Inequality is that of particle systems. An easy example is the voter process. Informally, in this process sites of a graph wait independently for an exponentially distributed period of time until their associated "bell" rings. A site can have one of two possible colours (commonly black and white). When its bell rings a site chooses a neighbour at random and assumes the colour of that neighbour. Should the neighbour be of the same colour then no change occurs. We follow the evolution of the process through time and investigate its behaviour. Amongst other things the FKG Inequality can be used to show that the invariant measures for the distribution of vertex colours is positively correlated. For a more rigorous and in-depth look at these processes the reader is referred to Liggett (1985).

Finally we give an application in the field of combinatorics. Define an infinite sequence $\{a_k : k \geq 0\}$ to be log convex if $a_k^2 \leq a_{k-1}a_{k+1}$ $(1 \leq k \leq \infty)$. A sequence $(b_k : k \geq 0)$ is log concave if $\{b^{-1} : k \geq 0\}$ is log convex. Commonly occurring examples of log convex sequences include the binomial coefficients and Stirling numbers of both kinds (Abramowitz and Stegun (1972)). The FKG Inequality is used to prove the following result:

Theorem 6.2 *If* $(a_k : 0 \leq k \leq n)$ *is log convex and positive and* $(b_i : 0 \leq i \leq n), (c_i : 0 \leq i \leq n)$ *are both increasing (or both decreasing) sequences then*

$$\sum_{k=0}^{n} a_k \sum_{k=0}^{n} a_k b_k c_k \geq \sum_{k=0}^{n} a_k b_k \sum_{k=0}^{n} a_k c_k.$$

The proof proceeds by defining $d_k = a_k/\binom{n}{k}$ for $0 \leq k \leq n$, $S = \{1, 2, \ldots, n\}$ and $\mu(A) = d_{|A|}$ for all $A \subseteq S$. Then if we set $f(A) = b_{|A|}$, $g(A) = c_{|A|}$ $(A \subseteq S)$ we can, after a little work, apply the FKG Inequality to conclude the result. For more details refer to Seymour and Welsh (1975).

Notice that if we put $a_k = 1$ for all k we get:

Theorem 6.3 (Chebyshev) *If* $b_0 \leq \ldots \leq b_n$ *and* $c_0 \leq \ldots \leq c_n$, *or* $b_0 \geq \ldots \geq b_n$ *and* $c_0 \geq \ldots \geq c_n$ *then*

$$\left(\frac{1}{n}\sum_{i=0}^{n} b_i\right)\left(\frac{1}{n}\sum_{i=0}^{n} c_i\right) \leq \left(\frac{1}{n}\sum_{i=0}^{n} b_i c_i\right).$$

6.2 Negative Correlation

We now want to turn our attention to the subject of negative correlation, but first we must give a little thought to how we will define the concept. It is, of course, no problem to define positive correlation. If one looks at the conclusion of the FKG Inequality one has an excellent definition, ie. the system (or underlying measure μ) is positively correlated if $E(fg) \geq E(f)E(g)$ for f, g both increasing (or both decreasing) functions. But unfortunately it is not so simple where negative correlation is concerned. Our first thought might be to simply reverse the direction of the FKG Inequality to get $E(fg) \leq E(f)E(g)$. But if we put $f = g$ we contradict the Cauchy-Schwartz Inequality and so clearly this is inappropriate. So what kind of definition is appropriate? To help determine this we will now give a few examples of situations in which some sort of negative correlation seems to be present.

(i) Death Processes: These occur widely throughout statistical literature and can arise naturally in many ways, (for instance, many occur as components of compartmental systems). An elementary account of death processes can be found in Taylor and Karlin (1984), but for our purposes the following definition will suffice.

A death process $\{X(t) : t > 0\}$ is a continuous time Markov chain with $X(0) = N$ and infinitesimal transition probabilities ("death rates")

$$\lim_{h \downarrow 0} h^{-1} P(X(t+h) = j \mid X(t) = k) = \begin{cases} \mu_k & \text{if } j = k-1 \\ 0 & \text{if } j \neq k, k-1. \end{cases}$$

The state 0 is an absorbing state. In Ball and Donnelly (1987) the authors prove a conjecture of Faddy concerning the relative variation of death processes for concave or convex death rates. In the course of this they prove that if we label the individuals initially present in the process $1, 2, \ldots, N$ and define indicator random variables

$$I_n(t) = \begin{cases} 1 & \text{if } n \text{ is alive at time } t \\ 0 & \text{otherwise} \end{cases} \qquad n = 1, \ldots, N$$

then if the death rates $\mu_N, \mu_{N-1}, \ldots, \mu_1$ form a concave sequence we have

$$E(I_m(t)I_n(t)) \geq E(I_m(t))E(I_n(t))$$

and if the death rates form a convex sequence we have

$$E(I_m(t)I_n(t)) \leq E(I_m(t))E(I_n(t)).$$

The result for a convex sequence of death rates can be interpreted in the following way

$$P(\text{individual } m \text{ present at } t \mid \text{individual } n \text{ present at } t)$$

$$\leq P(\text{individual } m \text{ present at } t)$$

In other words, the presence of particles is negatively correlated.

(ii) The Anti-Voter Process: We formally define the anti-voter process η_t^A by supposing the existence of a finite connected graph G with vertex set V and edge set E. Initially we colour the vertices in A black and those in $V \backslash A$ white. Then we associate a "random clock" with each vertex of G which rings, independently for each vertex, at the instances of a Poisson process of rate 1. When a clock rings the associated vertex chooses a neighbour at random and adopts the opposite colour to the chosen neighbour. The set of vertices which are black at time t is denoted by η_t^A. Clearly, since the statespace of the process is finite, it must have an equilibrium distribution, and one would expect because of the nature of the model some form of negative correlation in the equilibrium distribution. In fact it is possible to prove that for graphs with enough symmetry, the equilibrium distribution is negatively correlated in the sense that if x and y are neighbouring vertices, then

$$P(x \text{ black } \mid y \text{ black}) \leq P(x \text{ black})$$

and

$$P(x \text{ white } \mid y \text{ white}) \leq P(x \text{ white})$$

For the exact result and further details see Donnelly and Welsh (1984).

(iii) Anti-ferromagnetic Models: If one considers the Ising model as stated in section 6.1 for $\beta > 0$ we have an example of positive correlation. In some sense a positive β encourages neighbouring vertices to have the same spin. However, if we allow $\beta < 0$ we get a form of negative correlation. At present it has proved impossible to treat this model in any generality and indeed it is not clear how such an analysis should proceed. But these models do exhibit negative correlation and they are discussed in more detail in Griffiths (1972 Section V.C.1.).

(iv) A Simple Infection Model: Suppose the existence of a finite population of individuals $x_1, x_2, \ldots, x_N$, and introduce a model of infection where the probability that a particular subset $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$ is infected is simply a function of k. Then, given that there are exactly M infected individuals, we have

$$P(x_i \text{ infected}) = \frac{M}{N} \quad \text{for } 1 \leq i \leq N,$$

$$P(x_i \text{ infected} \mid x_j \text{ infected}) = \frac{M-1}{N-1} < \frac{M}{N} \quad \text{for } 1 \leq i, j \leq N, \ \ i \neq j.$$

So the status of individuals is negatively correlated.

6.3 A Negative Correlation Inequality

Before we do anything it is necessary to spend a little time proving a lemma which will be useful later on:

Lemma 6.4 *Given sequences of positive numbers* $(a_k)_{k=0}^n, (b_k)_{k=0}^n$ *and* $(x_k)_{k=0}^n$ *such that the following conditions are satisfied:*

(i) $a_0 - b_0 \leq 0$

(ii) $a_k - b_k$ *increases as* k *increases,*

(iii) $\sum_{k=0}^n a_k - \sum_{k=0}^n b_k = 0,$

(iv) x_k *decreases as* k *increases*
(iii) $\sum_{k=0}^n a_k - \sum_{k=0}^n b_k = 0,$

(iv) x_k *decreases as* k *increases.*

$$\sum_{k=0} a_k x_k \leq \sum_{k=0}^{n} b_k x_k.$$

(Note that condition (i) is implied by (ii) and (iii). It is included merely for clarity.)

Proof:

Define $c_k = a_k - b_k, \quad k = 1, \ldots, n$. Then c_k increases as k increases and x_k decreases as k increases. Thus since $(-x_k)$ increases as k increases we can apply Chebyshev's inequality (theorem (6.3)) to get

$$\left(\frac{1}{n}\sum_{k=0}^{n} c_k\right)\left(\frac{1}{n}\sum_{k=0}^{n}(-x_k)\right) \leq \frac{1}{n}\sum_{k=0}^{n} c_k(-x_k).$$

Thus

$$\left(\frac{1}{n}\sum_{k=0}^{n} c_k\right)\left(\frac{1}{n}\sum_{k=0}^{n} x_k\right) \geq \frac{1}{n}\sum_{k=0}^{n} c_k x_k.$$

But $\frac{1}{n}\sum_{k=0}^n c_k = 0$ by assumption (iii), so

$$\sum_{k=0}^{n} c_k x_k \leq 0,$$

ie.

$$\sum_{k=0}^{n} a_k x_k \leq \sum_{k=0}^{n} b_k x_k. \quad \text{as required.} \qquad \square$$

Having completed our brief diversion we now proceed as before and suppose the existence of a set

$$X = \{X_1, X_2, \ldots, X_N\}$$

and let

$$\Omega = 2^X = \{\text{all subsets of } X\}.$$

Hence $| \Omega |= 2^N$.

Let μ be a measure, $\mu : \Omega \rightarrow [0,1]$ such that $\sum_{A \in \Omega} \mu(A) = 1$. Finally, define

$$Q(A) = \sum_{Y \in \Omega, Y \supseteq A} \mu(Y)$$

We now return to the problem of defining what we mean by negative correlation. Recall that

$$\langle fg \rangle \leq \langle f \rangle \langle g \rangle \quad \text{for all increasing functions } f, g$$

is not possible. Restricting this to functions f, g which are the indicators of disjoint sets gives

$$Q(A \cup B) \leq Q(A)Q(B)$$

or alternatively

$$P(A \mid B) \leq P(B)$$

as is suggested by the examples of negative correlation in the previous section. However, we can't extend this to all indicators (because of the Cauchy-Schwartz Inequality again), but we may hope to "correct" the left-hand side to get the following

$$Q(A \cup B)Q(A \cap B) \leq Q(A)Q(B) \quad \text{for all } A, B.$$

It is in fact in this form that we will prove a negative correlation inequality:

So we aim to prove a result analogous to the FKG Inequality, but for a negatively correlated measure μ. Specifically, we will prove the following result:

Theorem 6.5 *Suppose we have a measure μ and a function Q both acting on $\Omega = 2^X$ where $X = \{X_1, X_2, \ldots, X_N\}$ and $Q(A) = \sum_{Y \in \Omega, Y \supseteq A} \mu(Y)$. Suppose further that $\mu(\cdot)$ depends only on $|\cdot|$ and that $\mu(A \cup B)\mu(A \cap B) \leq \mu(A)\mu(B)$ for all $A, B \in \Omega$. Then*

$$Q(A \cup B)Q(A \cap B) \leq Q(A)Q(B) \quad \textit{for all } A, B \in \Omega.$$

Before beginning the formal proof we have a few general comments. Firstly we should point out that this work is closely related to that given by Karlin and Rinott (1980). In particular the theorem on page 501 of that paper. Their work is in a continuous setting however. They define a measure $\mu(\cdot)$ as the density of a continuous random vector $(X_1, X_2, \ldots, X_n)$ and show that for a $\mu(\cdot)$ satisfying certain conditions we can make the following statement:

$$E(\Pi_{i=1}^{k} \phi_i(X_i)) \leq E((\Pi_{i=1}^{k+l} \phi_i(X_i))(\Pi_{i=k+l+1}^{n} \phi_i(X_i)))$$

for $1 \leq k \leq k + l \leq n$, for a general class of functions $\{\phi_i\}_1^n$, (which includes the indicator functions). This is a more general result, but as several authors have pointed out (eg. Block, Savits, and Shaked (1982), and Joag-Dev and Proschan (1983)) the condition on μ (ie. that μ is strongly multivariate reverse rule of order 2 (see Karlin and Rinott (1980) for definition)) is very hard to interpret and difficult to verify. Additionally, although the authors briefly mention the case where the X_i are discrete (as they are for our result), and say that some of the results carry over, they give no details. The advantages of the approach presented here are that the condition on the underlying measure

μ is more natural, easier to check, and that the proof of our result is more direct. It is worth noting that if the set of random variables $(X_1, \ldots, X_n)$ are exchangeable then the condition that $\mu(A) = \mu(B)$ whenever $\mid A \mid = \mid B \mid$ is immediately satisfied. We also note without proof the result

Lemma 6.6 *Suppose $\mu(A) = \mu(B)$ whenever $\mid A \mid = \mid B \mid$ and denote the value taken by μ on sets of size i by μ_i. Then*

$\mu_{i-1}\mu_{i+1} \geq (\leq)\, \mu_i^2$ *for all i*

$$\Rightarrow \mu(A \cap B)\mu(A \cup B) \geq (\leq)\, \mu(A)\mu(B) \text{ for all } A, B.$$

So it is sufficient to check $\mu_{i-1}\mu_{i+1} \geq (\leq)\, \mu_i^2$ for all i in order to satisfy the first condition of our result.

We shall prove theorem 6.5 in stages. In outline we will proceed as follows. Initially we prove the result for singleton sets A, B. There are then two cases to consider for more general A, B:

(i) $A \cap B \neq \emptyset$

(ii) $A \cap B = \emptyset$.

In the first case we suppose $A \cap B = C$ and define

$$\Omega_C = \{D \in \Omega \mid C \subseteq D\} \text{ and } Q_C(E) = \sum_{Y \in \Omega_C, Y \supseteq E} \mu(Y).$$

Then for a natural choice of measure μ_C on Ω_C we apply induction on $N = \mid \Omega \mid$ to get

$$Q_C(E \cup F)Q_C(E \cap F) \leq Q_C(E)Q_C(F) \text{ for all } E < F \in \Omega_C$$

and so deduce that

$$Q(E \cup F)Q(E \cap F) \leq Q(E)Q(F) \text{ for all } E, F \in \Omega$$

as required. In the second case (ie. $A \cap B = \emptyset$) we define $C = A \cup \{b\}$ where $b \in B$ and so

$$Q(B \cap C)Q(B \cup C) \leq Q(B)Q(C)$$

by case (i). Then, by induction on $\mid A \cup B \mid$, we see that

$$Q(A \cup b)Q(A \cap b) \leq Q(A)Q(b).$$

Combining the two we get the required conclusion.

We now begin the formal proof of theorem 6.5 with the following proposition:

Proposition 6.7 *Suppose*

$$\mu(A \cap B)\mu(A \cup B) \leq \mu(A)\mu(B) \textit{ for all } A < B \in \Omega$$

and that $\mu(A) = \mu(B)$ *whenever* $\mid A \mid = \mid B \mid$. *Then*

$$Q(\{a\} \cap \{b\})Q(\{a\} \cup \{b\}) \leq Q(\{a\})Q(\{b\})$$

for all $a, b \in X$ *with* $a \neq b$.
(Clearly if $a = b$ *we simply have equality.)*

[Note that in future we shall abuse notation by writing $Q(a)$ instead of Q({a}) etc.].

Proof:

As $\mu(Y)$ depends only upon $\mid Y \mid$ denote by μ_i $(i = 0, 1, \ldots, N)$ the value taken by μ on sets of size i, (ie. $\mu_i = \mu(Y)$ where $\mid Y \mid = i$). Hence the conditions $\mu(A \cap B)\mu(A \cup B) \leq \mu(A)\mu(B)$ for all $A, B \in \Omega$ are equivalent to

$$\{\mu_{i-k}\mu_{j+k} \leq \mu_i\mu_j\} \quad \text{for all} \quad \begin{array}{l} i, j = 0, 1, \ldots N \\ i \leq j \\ k = 0, 1, \ldots, (i \wedge j). \end{array} \tag{6.3}$$

In our case we have $| a | = | b | = 1$ and so $\mu(a) = \mu(b) = \mu_1$.

Since $Q(a) = \sum_{Y \in \Omega, Y \ni a} \mu(Y)$ etc. we have

$$Q(a) = \sum_{i=1}^{N} \mu_i \begin{pmatrix} N-1 \\ i-1 \end{pmatrix} = Q(b).$$

Similarly

$$Q(a \cup b) = \sum_{i=2}^{N} \mu_i \begin{pmatrix} N-2 \\ i-2 \end{pmatrix}$$

and

$$Q(a \cap b) = Q(\emptyset) = 1 = \sum_{i=0}^{N} \mu_i \begin{pmatrix} N \\ i \end{pmatrix}.$$

Now let $n_z = Q(a \cup b)$. Then

$$\begin{aligned}
Q(a) = Q(b) &= \sum_{i=1}^{N} \mu_i \begin{pmatrix} N-1 \\ i-1 \end{pmatrix} \\
&= \mu_1 \begin{pmatrix} N-1 \\ 0 \end{pmatrix} + \mu_N \begin{pmatrix} N-1 \\ N-1 \end{pmatrix} + \sum_{i=2}^{N-1} \mu_i \begin{pmatrix} N-1 \\ i-1 \end{pmatrix} \\
&= \mu_1 + \mu_N + \sum_{i=2}^{N-1} \mu_i \left[\begin{pmatrix} N-2 \\ i-1 \end{pmatrix} + \begin{pmatrix} N-2 \\ i-2 \end{pmatrix} \right] \\
&= \sum_{i=2}^{N} \mu_i \begin{pmatrix} N-2 \\ i-2 \end{pmatrix} + \sum_{i=1}^{N-1} \mu_i \begin{pmatrix} N-2 \\ i-1 \end{pmatrix} \\
&= n_z + n_x
\end{aligned}$$

where n_x is defined as $\sum_{i=1}^{N-1} \mu_i \begin{pmatrix} N-2 \\ i-1 \end{pmatrix}$. Now define n_y by

$$n_y = 1 - (2n_x + n_z)$$

$$= \sum_{i=0}^{N} \mu_i \binom{N}{i} - \left[2 \sum_{i=1}^{N-1} \mu_i \binom{N-2}{i-1} + \sum_{i=2}^{N} \mu_i \binom{N-2}{i-2} \right]$$

$$= \mu_0 + \mu_1(N-2) + \mu_{N-1}[N-(2+N-2)] + \mu_N(1-1)$$

$$+ \sum_{i=2}^{N-2} \mu_i \left[\binom{N}{i} - 2\binom{N-2}{i-1} - \binom{N-2}{i-2} \right].$$

Now for $i = 2, 3, \ldots, N-2$ we have

$$\binom{N}{i} = \binom{N-1}{i} + \binom{N-1}{i-1}$$

$$= \binom{N-2}{i} + \binom{N-2}{i-1} + \binom{N-2}{i-1} + \binom{N-2}{i-2}.$$

So, on simplification,

$$n_y = \sum_{i=0}^{N-2} \mu_i \binom{N-2}{i}.$$

Note that by definition of n_y,

$$Q(\emptyset) = 1 = n_y + 2n_x + n_z$$

Now we have

$$Q(a \cup b) = n_z,$$

$$Q(a) = Q(b) = n_z + n_x,$$

and

$$Q(a \cap b) = Q(\emptyset) = n_y + 2n_x + n_z.$$

We want to show $Q(a \cap b)Q(a \cup b) \leq Q(a)Q(b)$, that is

$$n_z(n_y + 2n_x + n_z) \leq (n_z + n_x)^2 = n_z^2 + n_x^2 + 2n_z n_x,$$

which is equivalent to

$$n_z n_y \leq n_x^2.$$

So it is sufficient to show that

$$\left[\sum_{i=2}^{N} \mu_i \binom{N-2}{i-2}\right]\left[\sum_{i=0}^{N-2} \mu_i \binom{N-2}{i}\right] \leq \left[\sum_{i=1}^{N-1} \mu_i \binom{N-2}{i-1}\right]^2$$

that is,

$$\sum_{j=2}^{N}\sum_{i=0}^{N-2} \mu_i\mu_j \binom{N-2}{j-2}\binom{N-2}{i} \leq \sum_{i=1}^{N-1}\sum_{j=1}^{N-1} \mu_i\mu_j \binom{N-2}{i-1}\binom{N-2}{j-1}. \tag{6.4}$$

We can rewrite (6.4) as

$$\sum_{i=1}^{N-1}\sum_{k=0}^{i\wedge(N-i)} \mu_{i-k}\mu_{i+k}a_{i,k} + \sum_{i=2}^{N-1}\sum_{k=0}^{(i-1)\wedge(N-i)} \mu_{i-1-k}\mu_{i+k}a'_{i,k} \tag{6.5}$$

$$\leq \sum_{i=1}^{N-1}\sum_{k=0}^{(i-1)\wedge(N-i-1)} \mu_{i-k}\mu_{i+k}b_{i,k} + \sum_{i=2}^{N-1}\sum_{k=0}^{(i-2)\wedge(N-i-1)} \mu_{i-1-k}\mu_{i+k}b'_{i,k}$$

for a sequence of coefficients $\{a_{i,k}\}, \{a'_{i,k}\}, \{b_{i,k}\}, \{b'_{i,k}\}$. In other words

$a_{i,k}$ = the coefficient of $\mu_{i-k}\mu_{i+k}$ in the left-hand side of (6.4)

$a'_{i,k}$ = the coefficient of $\mu_{i-1-k}\mu_{i+k}$ in the left-hand side of (6.4)

$b_{i,k}$ = the coefficient of $\mu_{i-k}\mu_{i+k}$ in the right-hand side of (6.4)

$b'_{i,k}$ = the coefficient of $\mu_{i-i-k}\mu_{i+k}$ in the right-hand side of (6.4).

In some cases some of these coefficients will be zero. Their actual values will be found later.

Hence to prove (6.4) (and hence the proposition), it is sufficient to show that the following two conditions are satisfied:

$$\sum_{i=1}^{N-1}\sum_{k=0}^{i\wedge(N-i)} \mu_{i-k}\mu_{i+k}a_{i,k} \leq \sum_{i=1}^{N-1}\sum_{k=0}^{(i-1)\wedge(N-i-1)} \mu_{i-k}\mu_{i+k}b_{i,k} \tag{6.6}$$

$$\sum_{i=2}^{N-1} \sum_{k=0}^{(i-1)\wedge(N-i)} \mu_{i-1-k}\mu_{i+k}a'_{i,k} \leq \sum_{i=2}^{N-1} \sum_{k=0}^{(i-2)\wedge(N-i-1)} \mu_{i-1-k}\mu_{i+k}b'_{i,k}. \tag{6.7}$$

Note that since $\mu_i = \begin{pmatrix} N \\ i \end{pmatrix}$ we have that $\mu_i = \mu_{N-i}$ for all i. Hence it is sufficient to show that (6.6) is true for $i \leq \frac{N}{2}$ and that (6.7) is true for $i \leq \frac{N}{2}+1$.

We start by showing (6.6):
It is sufficient to show it for each $i = 0, 1, \ldots, \left[\frac{N}{2}\right]$. That is, to show

$$\sum_{k=0}^{i\wedge(N-i)} \mu_{i-k}\mu_{i+k}a_{i,k} \leq \sum_{k=0}^{(i-1)\wedge(N-i-1)} \mu_{i-k}\mu_{i+k}b_{i,k} \tag{6.8}$$

for each $i = 1, 2, \ldots, \left[\frac{N}{2}\right]$.
Note that if we now define

$$x_k = \mu_{i+k}\mu_{i-k}, \quad a_k = a_{i,k}, \quad \text{and} \quad b_k = b_{i,k},$$

we have sequences as in the statement of Lemma 6.4. Clearly the x_k are decreasing as k increases (by the conditions for the proposition), so it remains to check conditions (i), (ii) and (iii). If these are satisfied then the lemma implies that (6.6) is true. So we need to determine the exact values of the a_k's and b_k's.

Now the a_k's are as follows:

$$a_k = \begin{cases} \binom{N-2}{i-k}\binom{N-2}{i+k-2} + \binom{N-2}{i+k}\binom{N-2}{i-k-2} & \begin{array}{l} k > 0 \\ i-k \geq 2 \\ i+k \leq N-2 \end{array} \\ \binom{N-2}{i}\binom{N-2}{i-2} & \begin{array}{l} k = 0 \\ 2 \leq i \leq N-2 \end{array} \\ \binom{N-2}{i-k}\binom{N-2}{i+k-2} & \begin{array}{l} 0 \leq i-k < 2 \\ k > 0, \\ \underline{\text{or}} \\ N-2 < i+k \leq N \\ k > 0 \end{array} \\ 0 & \text{otherwise.} \end{cases}$$

And the b_k's are given by:

$$b_k = \begin{cases} 2\binom{N-2}{i-k-1}\binom{N-2}{i+k-1} & \begin{array}{l} k > 0 \\ i-k \geq 1 \\ i+k \leq N-1 \end{array} \\ \binom{N-2}{i-1}\binom{N-2}{i-1} & \begin{array}{l} k = 0 \\ 1 \leq i \leq N-1 \end{array} \\ 0 & \text{otherwise.} \end{cases}$$

By symmetry we are assuming $i \leq \frac{N}{2}$, so for $i = 0$ we have no terms for either a_k or b_k. For $i = 1$ we have to show

$$\binom{N-2}{i-1}\binom{N-2}{i+1-2} \leq \binom{N-2}{i-1}\binom{N-2}{i-1},$$

which is true since both sides equal 1.

For $i = 2, 3, \ldots, \left[\frac{N}{2}\right]$ and $k > 0$ we have

$$a_k - b_k = \left[\binom{N-2}{i-k}\binom{N-2}{i+k-2} + \binom{N-2}{i+k}\binom{N-2}{i-k-2}\right]$$

$$-2\binom{N-2}{i-k-1}\binom{N-2}{i+k-1}. \qquad (6.9)$$

Note that for $k = 0$ we have

$$\begin{aligned} a_k - b_k &= \binom{N-2}{i}\binom{N-2}{i-2} - \binom{N-2}{i-1}^2 \\ &= \frac{(N-2)!(N-2)!}{(i-1)!(i-2)!(N-i-1)!(N-i-2)!} \\ &\qquad \left[\frac{1}{i(N-i)} - \frac{1}{(i-1)(N-i-1)}\right] \\ &< 0 \text{ as required.} \end{aligned}$$

So condition (i) is satisfied. Note also that $\binom{N}{M} = \binom{N}{M-1}\frac{N-M+1}{M}$ implies that

$$\begin{aligned} &\binom{N-2}{i-k}\binom{N-2}{i+k-2} \\ &= \binom{N-2}{i-k-1}\binom{N-2}{i+k-1}\frac{N-i+k-1}{i-k}\frac{i+k-1}{N-i-k} \end{aligned}$$

$$\binom{N-2}{i+k}\binom{N-2}{i-k-2}$$

$$= \begin{pmatrix} N-2 \\ i+k-1 \end{pmatrix} \begin{pmatrix} N-2 \\ i-k-1 \end{pmatrix} \frac{N-i-k-1}{i+k} \frac{i-k-1}{N-i+k}.$$

Hence, if we write

$$C_1 = \begin{pmatrix} N-2 \\ i-k-1 \end{pmatrix} \begin{pmatrix} N-2 \\ i+k-1 \end{pmatrix}$$

and

$$C_2 = \begin{pmatrix} N-2 \\ i-k-1 \end{pmatrix} \begin{pmatrix} N-2 \\ i+k-1 \end{pmatrix}$$

then for $k > 0$, we may re-write (6.9) as

$$\begin{aligned}
a_k - b_k &= C_1 \left[\left(\frac{(N-i+k-1)(i+k-1)}{(i-k)(N-i-k)} - 1 \right) \right. \\
&\qquad \left. + \left(\frac{N-i-k-1}{i+k} \frac{i-k-1}{N-i+k} - 1 \right) \right] \\
&= C_1 \left[\frac{(N-i+k-1)(i+k-1) - (i-k)(N-i-k)}{(i-k)(N-i-k)} \right. \\
&\qquad \left. + \frac{(N-i-k-1)(i-k-1) - (i+k)(N-i+k)}{(i+k)(N-i+k)} \right] \\
&= C_1 \left[\frac{1 + 2kN - N - 2k}{(i-k)(N-i-k)} + \frac{1 + 2k - 2kN - N}{(i+k)(N-i+k)} \right] \\
&= C_1 \left[\frac{(1-N)(1-2k)}{(i-k)(N-i-k)} + \frac{(1-N)(1+2k)}{(i+k)(N-i+k)} \right] \\
&= C_1(1-N) \left[\frac{(1-2k)(i+k)(N-i+k)}{(i-k)(i+k)(N-i-k)(N-i+k)} \right. \\
&\qquad \left. + \frac{(1+2k)(i-k)(N-i-k)}{(i-k)(i+k)(N-i-k)(N-i+k)} \right]
\end{aligned}$$

$$
\begin{aligned}
&= C_1\frac{2(1-N)(iN-i^2+k^2-2k^2N)}{(i-k)(i+k)(N-i-k)(N-i+k)} \\
&= C_2\frac{2(2k^2N+i^2-k^2-iN)}{(i-k)(i+k)(N-i+k)} \\
&= C_2\frac{2[k^2(2N-1)-i(N-i)]}{(i-k)(i+k)(N-i+k)}.
\end{aligned}
$$

This clearly starts negative (for $k = 0$) and becomes positive as k increases (since $N > i \geq k$). So part (ii) is satisfied. It remains to show part (iii). Hence

$$
\begin{aligned}
\sum_{k=0}^{i} a_k &= \binom{N-2}{i}\binom{N-2}{i-2} + \binom{N-2}{1}\binom{N-2}{2i-3} \\
&\quad + \binom{N-2}{0}\binom{N-2}{2i-2} \qquad (6.10) \\
&\quad + \sum_{k=1}^{i-2}\left[\binom{N-2}{i-k}\binom{N-2}{i+k-2} + \binom{N-2}{i+k}\binom{N-2}{i-k-2}\right]
\end{aligned}
$$

where the first component corresponds to the $(k = 0)$ term, the second corresponds to $(k = i-1)$, the third to $(k = i)$, and the fourth and fifth components correspond to $(k = 1, 2, \ldots, i-2)$. Also

$$
\sum_{k=0}^{i} b_k = \binom{N-2}{i-1}\binom{N-2}{i-1} + \sum_{k=1}^{i-1} 2\binom{N-2}{i-k-1}\binom{N-2}{i+k-1}. \qquad (6.11)
$$

(Note that $b_i = 0$.) Now note that the $(k = l)$ term in the fourth component of the right-hand side of (6.10) equals half the $(k = l-1)$ term in the final component of the right-hand side of (6.11). Similarly the $(k = l)$ term in the final component of (6.10) equals half the $(k = l+1)$ term in the final component of (6.11). Hence we write

$$C = \sum_{k=2}^{i-2} \left[\binom{N-2}{i-k} \binom{N-2}{i+k-2} + \binom{N-2}{i+k} \binom{N-2}{i-k-2} \right]$$

to get

$$\begin{aligned} \sum_{k=0}^{i} a_k &= C + \binom{N-2}{i}\binom{N-2}{i-2} + \binom{N-2}{1}\binom{N-2}{2i-3} \\ &\quad + \binom{N-2}{0}\binom{N-2}{2i-2} + \binom{N-2}{i-1}\binom{N-2}{i-1} \\ &\quad + \binom{N-2}{i+1}\binom{N-2}{i-3} \end{aligned}$$

$$\begin{aligned} \sum_{k=0}^{i} b_k &= C + \binom{N-2}{i-1}\binom{N-2}{i-1} + \binom{N-2}{i-2}\binom{N-2}{i} \\ &\quad + \binom{N-2}{i-3}\binom{N-2}{i+1} + \binom{N-2}{1}\binom{N-2}{2i-3} \\ &\quad + \binom{N-2}{0}\binom{N-2}{2i-2} \end{aligned}$$

and so we have

$$\sum_{k=0}^{i} a_k - \sum_{k=0}^{i} b_k = 0$$

as required. So condition (iii) is satisfied and hence (6.6) is shown. It now remains to show (6.7). This proceeds in an exactly analogous manner to the proof of (6.6) and the interested reader can find the details in Appendix III.

Having demonstrated that conditions (6.6) and (6.7) are satisfied, we have completed the proof of Proposition 6.7. □

We now move on to the proof of the main result, Theorem 6.5.

Proof (of theorem 6.5):

We proceed by induction on N ($= \mid X \mid$).

The result is clearly true for $N = 1$ as either $A = \emptyset$ or $B = \emptyset$ or $A = B$. For $N = 2$ we have either of the following two cases:

(i) $A = \emptyset$ or $B = \emptyset$ or $A \subseteq B$ or $B \subseteq A$, in which case the result is obvious.

(ii) $A = \{a\}, B = \{b\}$ some $a \in X, b \in X, a \neq b$.

In Case (ii) Proposition 6.7 obtains. So suppose the result is true for all Ω' such that $\mid \Omega' \mid = 2^M$ ($M < N$).

Now, given $A, B \in \Omega$ there are two cases:

Case 1: $A \cap B \neq \emptyset$

Case 2: $A \cap B = \emptyset$.

We start with Case 1.

Suppose $A \cap B = C \neq \emptyset$. Now define a new set Ω_C as follows:

$$\Omega_C = \{D \in \Omega \mid C \subseteq D\}$$

Define $\mu_C(E) = \frac{\mu(E)}{T}$ where $T = \sum_{E \in \Omega_C} \mu(E)$. Hence $\sum_{E \in \Omega_C} \mu_C(E) = 1$ and

$$\begin{aligned}
\mu_C(E)\mu_C(F) &= \frac{\mu(E)}{T}\frac{\mu(F)}{T} \\
&\geq \frac{1}{T^2}\mu(E \cap F)\mu(E \cup F) \\
&= \mu_C(E \cap F)\mu_C(E \cup F).
\end{aligned}$$

Clearly μ_C depends only on $\mid E \mid$ (by definition of μ), so μ_C satisfies the conditions in Theorem 6.5. But since $\mid \Omega_C \mid < \mid \Omega \mid$ we can use the inductive hypothesis. So we define

$$Q_C(E) = \sum_{Y \in \Omega_C, Y \supseteq E} \mu_C(Y)$$

to get

$$Q_C(E \cap F) Q_C(E \cup F) \leq Q_C(E) Q_C(F).$$

But $Q_C(E) = \frac{Q(E)}{T}$, and so we have

$$Q(E \cap F) Q(E \cup F) \leq Q(E) Q(F) \quad \text{for all } E, F \in \Omega.$$

In particular

$$Q(A \cap B) Q(A \cup B) \leq Q(A) Q(B)$$

and so Case 1 is proven by induction.

<u>Case 2</u>: $A \cap B = \emptyset$.

Let

$$M = \mid A \cup B \mid = \mid A \mid + \mid B \mid \leq N.$$

Without loss of generality suppose that $\mid A \mid \leq \mid B \mid$

Define $C = A \cup \{b\}$ where $b \in B$.

Then $B \cap C = \{b\} \neq \emptyset$, and

$$\mid B \cup C \mid = \mid B \cup A \mid = M \leq N.$$

So by Case 1 we have

$$Q(B \cap C) Q(B \cup C) \leq Q(B) Q(C),$$

ie.

$$Q(b) Q(A \cup B) \leq Q(B) Q(A \cup b),$$

which implies

$$Q(A \cup B) \leq \frac{Q(B) Q(A \cup b)}{Q(b)}. \tag{6.12}$$

We now start another induction:
So suppose by induction, that for a set $Z = \{Z_1, Z_2, \ldots, Z_N\}$ of size N, Theorem 6.5 is true for all sets $D, E \in Z$ such that $\mid D \cup E \mid < M$. (The first step of this induction, that is $\mid D \cup E \mid =1$ or 2, again follows from Proposition 6.7).

Then without loss of generality we have picked A such that $\mid A \mid \leq \frac{M}{2}$. Thus

$$\mid A \cup \{b\} \mid \leq \frac{M}{2} + 1 < M \quad (\text{for } M > 2) \text{ where } b \in B,$$

and so by the second induction we have

$$Q(A \cup b)Q(A \cap b) \leq Q(A)Q(b)$$

ie.

$$Q(A \cup b)Q(\emptyset) \leq Q(A)Q(b)$$

Thus,

$$Q(A \cup b) \leq Q(A)Q(b). \tag{6.13}$$

Hence by applying (6.12) followed by (6.13) we have

$$\begin{aligned} Q(A \cap B)Q(A \cup B) &= 1 \times Q(A \cup B) \\ &\leq \frac{Q(B)Q(A \cup b)}{Q(b)} \\ &\leq \frac{Q(B)[Q(A)Q(b)]}{Q(b)} \\ &= Q(A)Q(B). \end{aligned}$$

In other words $Q(A \cap B)Q(A \cup B) \leq Q(A)Q(B)$ as required. So both inductions go through, and the theorem is proven.

□

.1 Appendix I

We present here details of the derivation of a_k and b_k from Chapter 3 page 37. We start with

$$a_k = [p_1(k)(1-p_7(k)) + p_5(k)p_3(k)]/[(1-p_4(k))(1-p_7(k)) - p_3(k)p_8(k)]$$

$$b_k = [p_1(k)p_8(k) + p_5(k)(1-p_4(k))]/[(1-p_4(k))(1-p_7(k)) - p_3(k)p_8(k)]$$

Where the $p_i(k)$ are as defined in (3.1). Let us begin by considering the numerator of a_k:

$$\begin{aligned}
& p_1(k)(1-p_7(k)) + p_5(k)p_3(k) \\
= & \left(p\frac{k-1}{N-1} + (1-p)\frac{k}{N}\frac{k-1}{N-1}\right) \\
& \quad \times \left(1 - p\frac{N-k-1}{N-1} - (1-p)\frac{N-k}{N}\frac{N-k-1}{N-1}\right) \\
& \quad +(1-p)\frac{k}{N}\frac{k-1}{N-1}(1-p)\frac{N-k}{N}\frac{N-k-1}{N-1} \\
= & (1-p)^2\left[\frac{k}{N}\frac{k-1}{N-1}\frac{N-k}{N}\frac{N-k-1}{N-1} - \frac{k}{N}\frac{k-1}{N-1}\frac{N-k}{N}\frac{N-k-1}{N-1}\right] \\
& \quad -p(1-p)\left[\frac{k-1}{N-1}\frac{N-k}{N}\frac{N-k-1}{N-1} + \frac{k}{N}\frac{k-1}{N-1}\frac{N-k-1}{N-1}\right] \\
& \quad +(1-p)\frac{k}{N}\frac{k-1}{N-1} + p\frac{k-1}{N-1} - p^2\frac{k-1}{N-1}\frac{N-k-1}{N-1} \\
= & \frac{k-1}{N-1}\left[p - p^2\frac{N-k-1}{N-1} + (1-p)\frac{k}{N}\right. \\
& \qquad\qquad\qquad \left. -p(1-p)\frac{N-k-1}{N-1}\left(\frac{N-k}{N} + \frac{k}{N}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{k-1}{N-1}\left[p+(1-p)\frac{k}{N}-p^2\frac{N-k-1}{N-1}-p\frac{N-k-1}{N-1}+p^2\frac{N-k-1}{N-1}\right] \\
&= \frac{k-1}{N-1}\left[p+(1-p)\frac{k}{N}-p\frac{N-k-1}{N-1}\right] \\
&= \frac{k}{N-1}\left[p\frac{k-1}{N-1}+\frac{(1-p)(k-1)}{N}\right].
\end{aligned}
$$

Next the numerator of b_k:

$$
\begin{aligned}
&p_1(k)p_8(k)+p_5(k)(1-p_4(k)) \\
&= \left(p\frac{k-1}{N-1}+(1-p)\frac{k}{N}\frac{k-1}{N-1}\right) \\
&\qquad \times\left(p\frac{k}{N-1}(1-u)+(1-p)\frac{k}{N}\frac{N-k}{N-1}(2-u)\right) \\
&\qquad +(1-p)\frac{k}{N}\frac{k-1}{N-1}\left(1-p\frac{N-k}{N-1}-(1-p)\frac{k}{N}\frac{N-k}{N-1}(2-u)\right) \\
&= (1-p)^2\left[\frac{k}{N}\frac{k-1}{N-1}\frac{k}{N}\frac{N-k}{N-1}(2-u)-\frac{k}{N}\frac{k-1}{N-1}\frac{k}{N}\frac{N-k}{N-1}(2-u)\right] \\
&\qquad +p(1-p)\left[\frac{k-1}{N-1}\frac{k}{N}\frac{N-k}{N-1}(2-u)\right. \\
&\qquad\qquad \left.+\frac{k}{N}\frac{k-1}{N-1}\frac{k}{N-1}(1-u)-\frac{k}{N}\frac{k-1}{N-1}\frac{N-k}{N-1}\right] \\
&\qquad +p^2\frac{k-1}{N-1}\frac{k}{N-1}(1-u)+(1-p)\frac{k}{N}\frac{k-1}{N-1} \\
&= p(1-p)\left[\frac{k-1}{N-1}\frac{k}{N}\frac{N-k}{N-1}(1-u)+\frac{k}{N}\frac{k-1}{N-1}\frac{k}{N-1}(1-u)\right]
\end{aligned}
$$

$$+p^2\frac{k-1}{N-1}\frac{k}{N-1}(1-u)+(1-p)\frac{k}{N}\frac{k-1}{N-1}$$

$$= p(1-p)(1-u)\frac{k-1}{N-1}\frac{k}{N}\left[\frac{N-k+k}{N-1}\right]$$

$$+p^2\frac{k-1}{N-1}\frac{k}{N-1}(1-u)+(1-p)\frac{k}{N}\frac{k-1}{N-1}$$

$$= (1-u)\frac{k-1}{N-1}\frac{k}{N-1}(p-p^2)+p^2\left(\frac{k-1}{N-1}\frac{k}{N-1}(1-u)\right)$$

$$+(1-p)\frac{k}{N}\frac{k-1}{N-1}$$

$$= p(1-u)\frac{k-1}{N-1}\frac{k}{N-1}+(1-p)\frac{k}{N}\frac{k-1}{N-1}$$

$$= \frac{k}{N-1}\left[p(1-u)\frac{k-1}{N-1}+(1-p)\frac{k-1}{N}.\right]$$

Finally we move on to the denominator:

$$(1-p_4(k))(1-p_7(k))-p_3(k)p_8(k)]$$

$$= \left(1-p\frac{N-k}{N-1}-(1-p)\frac{k}{N}\frac{N-k}{N-1}(2-u)\right)$$

$$\times\left(1-p\frac{N-k-1}{N-1}-(1-p)\frac{N-k}{N}\frac{N-k-1}{N-1}\right)$$

$$-(1-p)\frac{N-k}{N}\frac{N-k-1}{N-1}\left(p\frac{k}{N-1}(1-u)+(1-p)\frac{k}{N}\frac{N-k}{N-1}(2-u)\right)$$

$$= (1-p)^2\left[\frac{k}{N}\frac{N-k}{N-1}(2-u)\frac{N-k}{N}\frac{N-k-1}{N-1}\right.$$

$$
\begin{aligned}
& \qquad\qquad\qquad\qquad\qquad \left.-\frac{N-k}{N}\frac{N-k-1}{N-1}\frac{k}{N}\frac{N-k}{N-1}(2-u)\right] \\
& +p(1-p)\left[\frac{N-k}{N-1}\frac{N-k}{N}\frac{N-k-1}{N-1}+\frac{k}{N}\frac{N-k}{N-1}(2-u)\frac{N-k-1}{N-1}\right. \\
& \qquad\qquad\qquad\qquad\qquad \left.-\frac{N-k}{N}\frac{N-k-1}{N-1}\frac{k}{N-1}(1-u)\right] \\
& +p^2\left[\frac{N-k}{N-1}\frac{N-k-1}{N-1}\right]-(1-p)\left[\frac{k}{N}\frac{N-k}{N-1}(2-u)+\frac{N-k}{N}\frac{N-k-1}{N-1}\right] \\
& -p\left[\frac{N-k}{N-1}+\frac{N-k-1}{N-1}\right]+1 \\
= {} & p(1-p)\left[\frac{N-k}{N-1}\frac{N-k}{N}\frac{N-k-1}{N-1}+\frac{k}{N}\frac{N-k}{N-1}\frac{N-k-1}{N-1}\right] \\
& +p^2\left[\frac{N-k}{N-1}\frac{N-k-1}{N-1}\right]-(1-p)\left[\frac{k}{N}\frac{N-k}{N-1}(2-u)+\frac{N-k}{N}\frac{N-k-1}{N-1}\right] \\
& -p\left[\frac{N-k}{N-1}+\frac{N-k-1}{N-1}\right]+1 \\
= {} & (p-p^2)\left[\frac{N-k}{N-1}\frac{N-k-1}{N-1}\left(\frac{N-k}{N}+\frac{k}{N}\right)\right]+p^2\left[\frac{N-k}{N-1}\frac{N-k-1}{N-1}\right] \\
& -(1-p)\left[\frac{k}{N}\frac{N-k}{N-1}(2-u)+\frac{N-k}{N}\frac{N-k-1}{N-1}\right] \\
& -p\left[\frac{N-k}{N-1}+\frac{N-k-1}{N-1}\right]+1 \\
= {} & p\frac{N-k}{N-1}\frac{N-k-1}{N-1}-(1-p)\left[\frac{k}{N}\frac{N-k}{N-1}(2-u)+\frac{N-k}{N}\frac{N-k-1}{N-1}\right]
\end{aligned}
$$

$$
\begin{aligned}
& -p\left[\frac{N-k}{N-1}+\frac{N-k-1}{N-1}\right]+1 \\
= \; & p\left[\frac{N-k}{N-1}\frac{N-k-1}{N-1}-\frac{N-k}{N-1}-\frac{N-k-1}{N-1}\right] \\
& -(1-p)\left[\frac{k}{N}\frac{N-k}{N-1}(1-u)+\frac{k}{N}\frac{N-k}{N-1}+\frac{N-k}{N}\frac{N-k-1}{N-1}\right]+1 \\
= \; & p\left[\frac{N-k}{N-1}\frac{N-k-1}{N-1}-\frac{N-k}{N-1}-\frac{N-k-1}{N-1}\right] \\
& -(1-p)\left[\frac{k}{N}\frac{N-k}{N-1}(1-u)+\frac{N-k}{N}\right]+1 \\
= \; & p\left[\frac{N-k}{N-1}\frac{N-k-1}{N-1}-\frac{N-k}{N-1}-\frac{N-k-1}{N-1}\right] \\
& -(1-p)\left[\frac{k}{N}\frac{N-k}{N-1}(1-u)+\frac{N-k}{N}-1\right]+p \\
= \; & p\left[1+\frac{N-k}{N-1}\frac{N-k-1}{N-1}-\frac{N-k}{N-1}-\frac{N-k-1}{N-1}\right] \\
& +\frac{(1-p)}{N(N-1)}\left[N(N-1)-k(N-k)(1-u)-(N-k)(N-1)\right] \\
= \; & \frac{p}{N-1}\left[(N-1)+\frac{N-k}{N-1}(N-k-1)-(N-k)-(N-k-1)\right] \\
& +\frac{(1-p)k}{N(N-1)}\left[k(1-u)-1+Nu\right] \\
= \; & \frac{p}{N-1}\left[2k-N+\frac{N^2-N}{N-1}+\frac{k^2+k-2kN}{N-1}\right]
\end{aligned}
$$

$$+\frac{(1-p)k}{N(N-1)}[k(1-u)-1+Nu]$$

$$= \frac{p}{N-1}\left[2k+\frac{k^2+k-2kN}{N-1}\right]+\frac{(1-p)k}{N(N-1)}[k(1-u)-1+Nu]$$

$$= \frac{p}{N-1}\left[\frac{k^2-2k+2kN+k-2kN}{N-1}\right]+\frac{(1-p)k}{N(N-1)}[k(1-u)-1+Nu]$$

$$= p\frac{k}{N-1}\frac{k-1}{N-1}+\frac{(1-p)k}{N(N-1)}[k(1-u)-1+Nu]$$

$$= \frac{k}{N-1}\left[p\frac{k-1}{N-1}+(1-p)\frac{N-k}{N}u+(1-p)\frac{k-1}{N}\right].$$

Hence

$$a_k = \frac{p\frac{k-1}{N-1}+\frac{(1-p)(k-1)}{N}}{p\frac{k-1}{N-1}+(1-p)\frac{N-k}{N}u+(1-p)\frac{k-1}{N}}$$

$$b_k = \frac{p(1-u)\frac{k-1}{N-1}+\frac{(1-p)(k-1)}{N}}{p\frac{k-1}{N-1}+(1-p)\frac{N-k}{N}u+(1-p)\frac{k-1}{N}}$$

and this simplifies routinely to give

$$a_k = \frac{(k-1)(N+p-1)}{(1-p)(N-k)(N-1)u+(k-1)(N+p-1)}$$

$$b_k = \frac{(k-1)(N+p-1-pNu)}{(1-p)(N-k)(N-1)u+(k-1)(N+p-1)}$$

as required.

.2 Appendix II

Here are the details of the simplification of equations (4.5), (4.6) and (4.7).

.2.1 Section A

Firstly we give the derivation of the equations (4.8) from equation (4.5).

$$
\begin{aligned}
&P(Q_K^0 \mid C_{t-1} = 2, K_{t-1} = K) \\
&= \sum_{d=0}^{0} \left[\left\{ \sum_{i=0}^{0} p^i (1-p)^{2-i} \begin{pmatrix} 2 \\ i \end{pmatrix} \sum_{j=0}^{T-2} p^j (1-p)^{T-2-j} \begin{pmatrix} T-2 \\ j \end{pmatrix} \right\} \right] \\
&\quad + O(N^{-1}) \\
&= (1-p)^2 + O(N^{-1})
\end{aligned}
$$

$$
\begin{aligned}
&P(Q_K^1 \mid C_{t-1} = 2, K_{t-1} = K) \\
&= p(1-p) \begin{pmatrix} 2 \\ 1 \end{pmatrix} \sum_{j=0}^{T-2} p^j (1-p)^{T-2-j} \begin{pmatrix} T-2 \\ j \end{pmatrix} + O(N^{-1}) \\
&= 2p(1-p) + O(N^{-1}).
\end{aligned}
$$

$$
\begin{aligned}
P(Q_K^2 \mid C_{t-1} = 2, K_{t-1} = K) &= p^2 \sum_{j=0}^{T-2} p^j (1-p)^{T-2-j} \begin{pmatrix} T-2 \\ j \end{pmatrix} + O(N^{-1}) \\
&= p^2 + O(N^{-1}).
\end{aligned}
$$

All these are very easy since for them we ignore terms $O(N^{-1})$. For the next three however we can only ignore terms $O(N^{-2})$, so they take a little more care.

$$P(Q_K^0 \mid C_{t-1} = 1, K_{t-1} = K)$$

$$= \sum_{d=0}^{0} \left[\sum_{i=0}^{0} \left(p^i (1-p)^{1-i} \binom{1}{i} \sum_{j=0}^{T-1} p^j (1-p)^{T-1-j} \binom{T-1}{j} \binom{T-j}{d} \right. \right.$$

$$\left. \times \frac{N-K-j}{N-d-j} \times \cdots \times \frac{N-K-j-(T-0-j-1)}{N-d-j-(T-0-j-1)} \right)$$

$$\left. \times 1 \times 1 \times \frac{N-K-T}{N-T} \times \cdots \times \frac{N-K-T-(T-1)}{N-T-(T-1)} \right]$$

$$= (1-p) \sum_{j=0}^{T-1} \left(p^j (1-p)^{T-1-j} \binom{T-1}{j} \right.$$

$$\times \left(\frac{N-K-j}{N-j} \times \cdots \times \frac{N-K-j-(T-j-1)}{N-j-(T-j-1)} \right)$$

$$\left. \times \left(\frac{N-K-T}{N-T} \times \cdots \times \frac{N-K-T-(T-1)}{N-T-(T-1)} \right) \right) + O(N^{-2})$$

$$= (1-p) \sum_{j=0}^{T-1} \left(p^j (1-p)^{T-1-j} \binom{T-1}{j} \left(\frac{N-K}{N} \right)^{2T-j} \right) + O(N^{-2})$$

$$\left(\text{since } \frac{N-K-A}{N-A} = 1 - \frac{K}{N} + O(N^{-2}) \right)$$

$$= (1-p) \sum_{j=0}^{T-1} \left(p^j (1-p)^{T-1-j} \binom{T-1}{j} \left(1 - \frac{K}{N} \binom{2T-j}{1} \right) \right)$$

$$+O(N^{-2})$$

$$= (1-p)\left[1-\frac{K}{N}(2T-E[\mathrm{Bin}(T-1,p)])\right]+O(N^{-2})$$

$$= (1-p)\left[1-\frac{K}{N}(2T-(T-1)p)\right]+O(N^{-2})$$

$$= (1-p)\left[1-\frac{TK(2-p)}{N}-\frac{pK}{N}\right]+O(N^{-2}).$$

We now move on to:

$$P(Q_K^1 \mid C_{t-1}=1, K_{t-1}=K)=\sum_{d=0}^{1}\sum_{i=0}^{1\wedge d} c_{d,i} \text{ say (cf. (4.5))},$$

ie. $P(Q_K^1 \mid C_{t-1}=1, K_{t-1}=K)=c_{0,0}+c_{1,0}+c_{1,1}$.

So dealing with this in parts we have

$$\begin{aligned}
c_{0,0} &= \Bigg\{(1-p)\Bigg(\sum_{j=0}^{T-1}\Bigg[p^j(1-p)^{T-1-j}\binom{T-1}{j} \\
&\qquad \times\frac{N-K-j}{N-j}\times\frac{N-K-j-1}{N-j-1}\times\cdots\times\frac{N-K-j-(T-j-1)}{N-j-(T-j-1)}\Bigg] \\
&\qquad \times\binom{T}{1}\frac{N}{N+\theta}\frac{K}{N-T}\times\frac{N-K-T}{N-T-1}\times\frac{N-K-T-1}{N-T-2}\times\cdots \\
&\qquad\qquad \times\frac{N-K-T-(T-2)}{N-T-(T-1)}\Bigg)\Bigg\} \\
&= (1-p)\sum_{j=0}^{T-1}\Bigg[p^j(1-p)^{T-1-j}\binom{T-1}{j}\left(1-\frac{K}{N}\right)^{T-j}\left(1-\frac{\theta}{N}\right) \\
&\qquad \times\binom{T}{1}\frac{K}{N-T}\left(1-\frac{K}{N}\right)^{T-1}+O(N^{-2})\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= (1-p)\sum_{j=0}^{T-1}\left[p^j(1-p)^{T-1-j}\binom{T-1}{j}\right.\\
&\qquad\qquad \left.\times\left(1-\frac{K}{N}(2T-j-1)\right)\binom{T}{1}\frac{K}{N-T}\right]+O(N^{-2})\\
&= (1-p)\sum_{j=0}^{T-1}\left[p^j(1-p)^{T-1-j}\binom{T-1}{j}\frac{KT}{N}\right]+O(N^{-2})\\
&= \frac{(1-p)KT}{N}+O(N^{-2}).
\end{aligned}
$$

Next we have

$$
\begin{aligned}
c_{1,0} &= (1-p)\sum_{j=0}^{T-1}\left[p^j(1-p)^{T-1-j}\binom{T-1}{j}\binom{T-j}{1}\frac{K}{N-j}\right.\\
&\quad \left.\times\frac{N-K-j}{N-1-j}\times\frac{N-K-j-1}{N-1-j-1}\times\cdots\times\frac{N-K-j-(T-1-j-1)}{N-1-j-(T-1-j-1)}\right]\\
&\quad \times\frac{N-K-(T-1)}{N-T-1}\times\frac{N-K-(T-1)-1}{N-T-2}\times\cdots\\
&\qquad\qquad\qquad\qquad \times\frac{N-K-(T-1)-(T-2)}{N-T-1-(T-2)}\\
&\quad +O(N^{-2})\\
&= (1-p)\sum_{j=0}^{T-1}\left[p^j(1-p)^{T-1-j}\binom{T-1}{j}\frac{K(T-j)}{N}\right]+O(N^{-2})\\
&= (1-p)\frac{K}{N}(T-E[\mathrm{Bin}(T-1,p)])+O(N^{-2})
\end{aligned}
$$

$$= (1-p)\frac{K}{N}(T-(T-1)p)+O(N^{-2})$$

$$= (1-p)\frac{K}{N}(T(1-p)+p)+O(N^{-2})$$

$$= \frac{K}{N}T(1-p)^2+\frac{p(1-p)K}{N}+O(N^{-2}).$$

Finally

$$\begin{aligned} c_{1,1} &= p\sum_{j=0}^{T-1}\left[p^j(1-p)^{T-1-j}\begin{pmatrix} T-1 \\ j \end{pmatrix}\right. \\ &\quad \left.\times\frac{N-K-j}{N-1-j}\times\frac{N-K-j-1}{N-1-j-1}\times\cdots\times\frac{N-K-j-(T-1-j-1)}{N-1-j-(T-1-j-1)}\right] \\ &\quad \times\frac{N-K-(T-1)}{N-T} \\ &\quad \times\frac{N-K-T}{N-T-1}\times\frac{N-K-T-1}{N-T-2}\times\cdots\times\frac{N-K-T-(T-2)}{N-T-1-(T-2)} \\ &\quad +O(N^{-2}) \\ &= p\sum_{j=0}^{T-1}\left[p^j(1-p)^{T-1-j}\begin{pmatrix} T-1 \\ j \end{pmatrix}\left(1-\frac{K-1}{N}\right)^{2T-2-j+1}\right]+O(N^{-2}) \\ &= p\sum_{j=0}^{T-1}\left[p^j(1-p)^{T-1-j}\begin{pmatrix} T-1 \\ j \end{pmatrix}\left(1-\frac{(K-1)(2T-1-j)}{N}\right)\right] \\ &\quad +O(N^{-2}) \\ &= p\left(1-\frac{(K-1)(2T-1)}{N}+\sum_{j=0}^{T-1}\left[p^j(1-p)^{T-1-j}\begin{pmatrix} T-1 \\ j \end{pmatrix}\frac{(K-1)j}{N}\right]\right) \end{aligned}$$

$$+O(N^{-2})$$

$$= p\left(1-\frac{(K-1)(2T-1)}{N}+\frac{K-1}{N}E[\mathrm{Bin}(T-1,p)]\right)+O(N^{-2})$$

$$= p\left(1-\frac{(K-1)(2T-1)}{N}+\frac{(K-1)(T-1)p}{N}\right)+O(N^{-2}).$$

Hence

$$P(Q_K^1 \mid C_{t-1}=1, K_{t-1}=K) = p+(1-p)\frac{TK}{N}+\frac{KT(1-p)^2}{N}$$

$$+\frac{p(1-p)K}{N}-\frac{p(K-1)T(2-p)}{N}$$

$$+\frac{p(1-p)(K-1)}{N}+O(N^{-2}).$$

Next

$$P(Q_K^2 \mid C_{t-1}=1, K_{t-1}=K)$$

$$= \sum_{d=1}^{2}\left[\left\{\sum_{i=0}^{1} p^i(1-p)^{1-i}\binom{1}{i}\right.\right.$$

$$\times \sum_{j=0}^{T-(1\vee d)}\left(p^j(1-p)^{T-1-j}\binom{T-1}{j}\binom{T-j-i}{d-1}\right.$$

$$\left.\left.\times\frac{K-i}{N-j-i}\times\cdots\times\frac{K-i-(d-i-1)}{N-j-i-(d-i-1)}\right)\right\}$$

$$\times\binom{T-d}{2-d}\times\frac{K-d}{N-T-d}\times\frac{K-d-1}{N-T-d-1}\times\cdots$$

$$\left.\times\frac{K-d-(2-d-1)}{N-T-d-(2-d-1)}\right]$$

$$+O(N^{-2})$$

$$= \sum_{d=1}^{2}\Bigg[p\Bigg[\sum_{j=0}^{T-d}\Bigg(p^j(1-p)^{T-1-j}\binom{T-1}{j}\binom{T-j-i}{d-1}$$

$$\times\frac{K-1}{N}\times\frac{K-1-1}{N}\times\cdots\times\frac{K-1-(d-1-1)}{N}\Bigg)\Bigg]$$

$$\times\binom{T-d}{2-d}\times\frac{K-d}{N}\times\frac{K-d-1}{N}\times\cdots\times\frac{K-d-(2-d-1)}{N}\Bigg]$$

$$+O(N^{-2})$$

$$= p\Bigg[\sum_{j=0}^{T-1}\Bigg(p^j(1-p)^{T-1-j}\binom{T-1}{j}\times 1\times 1\Bigg)\binom{T-1}{1}\frac{K-1}{N}\Bigg]$$

$$+p\Bigg[\sum_{j=0}^{T-2}p^j(1-p)^{T-1-j}\binom{T-1}{j}\binom{T-j-1}{1}\frac{K-1}{N}\times 1\Bigg]+O(N^{-2})$$

$$= p\Bigg[\frac{(T-1)(K-1)}{N}$$

$$+\frac{K-1}{N}\sum_{j=0}^{T-2}\Bigg(p^j(1-p)^{T-1-j}\binom{T-1}{j}\binom{T-j-1}{1}\Bigg)\Bigg]+O(N^{-2})$$

$$= p\Bigg[\frac{(T-1)(K-1)}{N}$$

$$+\frac{K-1}{N}(1-p)(T-1)\sum_{j=0}^{T-2}\Bigg(p^j(1-p)^{(T-2)-j}\binom{T-2}{j}\Bigg)\Bigg]+O(N^{-2})$$

$$\left(\text{since } \binom{T-1}{j}\binom{T-j-1}{1} = \binom{T-2}{j}(T-1)\right)$$

$$= p\left[\frac{(T-1)(K-1)}{N} + \frac{(K-1)(1-p)(T-1)}{N}\right] + O(N^{-2})$$

$$= \frac{p(K-1)(T-1)(2-p)}{N} + O(N^{-2}).$$

Finally, for the last three equations we can only ignore terms $O(N^{-3})$, so these again take a little more effort.

$$P(Q_K^0 \mid C_{t-1} = 0, K_{t-1} = K)$$

$$= \sum_{d=0}^{0}\left[\sum_{j=0}^{T}\left(p^j(1-p)^{T-j}\binom{T}{j}\right.\right.$$

$$\left.\times\frac{N-K-j}{N-j}\times\frac{N-K-j-1}{N-j-1}\times\cdots\times\frac{N-K-j-(T-j-1)}{N-j-(T-j-1)}\right)$$

$$\left.\times\frac{N-K-T}{N-T}\times\frac{N-K-T-1}{N-T-1}\times\cdots\times\frac{N-K-T-(T-1)}{N-T-(T-1)}\right]$$

$$+O(N^{-3})$$

$$= \sum_{j=0}^{T}\left[p^j(1-p)^{T-j}\binom{T}{j}\frac{N-K-j}{N-j}\times\frac{N-K-j-1}{N-j-1}\times\right.$$

$$\left.\cdots\times\frac{N-K-(2T-1)}{N-(2T-1)}\right] + O(N^{-3})$$

$$= \sum_{j=0}^{T}\left[p^j(1-p)^{T-j}\binom{T}{j}\left[1 - \frac{K}{N} - \frac{Kj}{N^2}\right]\right.$$

$$\times\left[1-\frac{K}{N}-\frac{K(j+1)}{N^2}\right]\times\cdots\times\left[1-\frac{K}{N}-\frac{K(2T-1)}{N^2}\right]\Bigg]+O(N^{-3})$$

$$\left(\text{since } \frac{N-A}{N-B}=(N-A)\left[\frac{1}{N}+\frac{B}{N^2}+\frac{B^2}{N^3}\right]+O(N^{-3})\right.$$

$$\left.=1-\frac{(A-B)}{N}-\frac{B(A-B)}{N^2}+O(N^{-3})\right)$$

$$= \sum_{j=0}^{T}\left[p^j(1-p)^{T-j}\begin{pmatrix} T \\ j \end{pmatrix}\right.$$

$$\left.\times\left[1-\frac{K}{N}\begin{pmatrix} 2T-j \\ 1 \end{pmatrix}+\frac{K^2}{N^2}\begin{pmatrix} 2T-j \\ 2 \end{pmatrix}-\frac{K}{N^2}\sum_{i=j}^{2T-1} i\right]\right]+O(N^{-3})$$

$$= \sum_{j=0}^{T}\left[p^j(1-p)^{T-j}\begin{pmatrix} T \\ j \end{pmatrix}\right.$$

$$\left.\left[1-\frac{K}{N}(2T-j)+\frac{K^2}{N^2}\begin{pmatrix} 2T-j \\ 2 \end{pmatrix}-\frac{K}{N^2}\left(\frac{(2T-1)2T}{2}-\frac{j(j-1)}{2}\right)\right]\right]$$

$$+O(N^{-3})$$

$$\left(\text{since } \sum_{i=j}^{2T-1} i=\sum_{i=1}^{2T-1} i-\sum_{i=1}^{j-1} i=\frac{(2T-1)2T}{2}-\frac{j(j-1)}{2}\right)$$

$$= 1-\frac{K}{N}(2T-E(\mathrm{Bin}(T,p)))-\frac{K}{N^2}\frac{(2T-1)2T}{2}$$

$$+\sum_{j=0}^{T}\left(p^j(1-p)^{T-j}\begin{pmatrix} T \\ j \end{pmatrix}\left[\frac{K^2}{N^2}\frac{(2T-j)(2T-j-1)}{2}+\frac{K}{N^2}\frac{j(j-1)}{2}\right]\right)$$

$$+O(N^{-3})$$

$$= \quad 1-\frac{K}{N}(2T-Tp)-\frac{K}{2N^2}(2T-1)2T$$

$$+\sum_{j=0}^{T} p^j(1-p)^{T-j}\left(\begin{array}{c} T \\ j \end{array}\right)\left[\frac{K^2}{2N^2}(4T^2-2T-4Tj+j^2+j)+\frac{K}{2N^2}(j^2-j)\right]$$

$$+O(N^{-3})$$

$$= \quad 1-\frac{KT}{N}(2-p)-\frac{K(2T-1)2T}{2N^2}$$

$$+\frac{K^2}{2N^2}(4T^2-2T-4T^2p+(Tp)^2+Tp(1-p)+Tp)$$

$$+\frac{K}{2N^2}[(Tp)^2+Tp(1-p)-Tp]+O(N^{-3})$$

$$\Big(\text{since } J \sim \text{Bin}(T,p) \Rightarrow E(J^2) = (\text{mean})^2 + (\text{variance})^2$$

$$= (Tp)^2+Tp(1-p)\Big)$$

$$= \quad 1-\frac{KT}{N}(2-p)-\frac{K}{2N^2}[(2T-1)2T+4T^2K-2TK-4T^2pK$$

$$+(Tp)^2K+Tp(1-p)K+TpK+(Tp)^2+Tp(1-p)-Tp]+O(N^{-3})$$

$$= \quad 1-\frac{KT}{N}(2-p)+\frac{KT}{2N^2}[TK(2-p)^2-T(4-p^2)-K(2-2p+p^2)+2-p^2]$$

$$+O(N^{-3})$$

$$= \quad 1-\frac{KT}{N}(2-p)+\frac{KT}{2N^2}[TK(2-p)^2-T(2-p)^2-K(2-p)^2+(2-p)^2$$

$$-T(4p-2p^2)+K(2-2p)-(2-4p+2p^2)]+O(N^{-3})$$

$$= 1-\frac{KT(2-p)}{N}+\begin{pmatrix} T \\ 2 \end{pmatrix}\frac{K(K-1)(2-p)^2}{N^2}$$

$$-\frac{KT}{N^2}[T(2p-p^2)-K(1-p)+(1-2p+p^2)]+O(N^{-3})$$

$$= 1-\frac{KT(2-p)}{N}+\begin{pmatrix} T \\ 2 \end{pmatrix}\frac{K(K-1)(2-p)^2}{N^2}+\frac{TK(K-1)(1-p)}{N^2}$$

$$-\frac{KT}{N^2}[(K-1)(1-p)+T(2p-p^2)-K(1-p)+(1-2p+p^2)]+O(N^{-3})$$

$$= 1-\frac{KT(2-p)}{N}+\begin{pmatrix} T \\ 2 \end{pmatrix}\frac{K(K-1)(2-p)^2}{N^2}+\frac{TK(K-1)(1-p)}{N^2}$$

$$-\frac{KT}{N^2}[Tp(2-p)-p(1-p)]+O(N^{-3})$$

as required.

Next we have

$$P(Q_K^1 \mid C_{t-1}=0, K_{t-1}=K) = d_0+d_1 \text{ say,}$$

where d_0 denotes the terms in (4.5) with $d=0$, and d_1 represents the terms with $d=1$. So:

$$d_0 = \sum_{j=0}^{T}\left(p^j(1-p)^{T-j}\begin{pmatrix} T \\ j \end{pmatrix}\times\frac{N-K-j}{N-j}\times\frac{N-K-j-1}{N-j-1}\times\cdots\right.$$

$$\left.\times\frac{N-K-(T-1)}{N-(T-1)}\right)\times\begin{pmatrix} T \\ 1 \end{pmatrix}\frac{K}{N-T}\frac{N}{N+\theta}$$

$$
\begin{aligned}
&\times\left(\frac{N-K-T}{N-(T+1)}\times\frac{N-K-T-1}{N-(T+2)}\times\cdots\times\frac{N-K-T-(T-2)}{N-T-(T-1)}\right)\\
=\ &\sum_{j=0}^{T}\left(p^j(1-p)^{T-j}\begin{pmatrix}T\\ j\end{pmatrix}\begin{pmatrix}T\\ 1\end{pmatrix}\frac{N}{N+\theta}\frac{K}{N-(2T-1)}\right.\\
&\qquad\left.\times\left[\frac{N-K-j}{N-j}\times\frac{N-K-j-1}{N-j-1}\times\cdots\times\frac{N-K-T-(T-2)}{N-T-(T-2)}\right]\right)\\
=\ &\sum_{j=0}^{T}\left(p^j(1-p)^{T-j}\begin{pmatrix}T\\ j\end{pmatrix}\begin{pmatrix}T\\ 1\end{pmatrix}\left[1-\frac{\theta}{N}\right]\right.\\
&\qquad\left.\times\left[\frac{K}{N}+\frac{K(2T-1)}{N^2}\right]\left[1-\frac{K}{N}\right]^{2T-2-j+1}\right)+O(N^{-3})\\
=\ &\sum_{j=0}^{T}\left(p^j(1-p)^{T-j}\begin{pmatrix}T\\ j\end{pmatrix}\begin{pmatrix}T\\ 1\end{pmatrix}\right.\\
&\qquad\left.\times\left[\frac{K}{N}-\frac{K\theta}{N^2}+\frac{K(2T-1)}{N^2}\right]\left[1-\frac{K}{N}\begin{pmatrix}2T-j-1\\ 1\end{pmatrix}\right]\right)+O(N^{-3})\\
=\ &\sum_{j=0}^{T}\left(p^j(1-p)^{T-j}\begin{pmatrix}T\\ j\end{pmatrix}\begin{pmatrix}T\\ 1\end{pmatrix}\right.\\
&\qquad\left.\times\left[\frac{K}{N}-\frac{K(\theta-2T+1)}{N^2}-\frac{K^2(2T-1-j)}{N^2}\right]\right)+O(N^{-3})\\
=\ &\frac{TK}{N}-\frac{TK\theta}{N^2}+\frac{TK(2T-1)}{N^2}-\frac{TK^2}{N^2}(2T-1-E(\mathrm{Bin}(T,p))+O(N^{-3})\\
=\ &\frac{TK}{N}-\frac{TK\theta}{N^2}+\frac{TK(2T-1)}{N^2}-\frac{T^2K^2}{N^2}(2-p)+\frac{TK^2}{N^2}+O(N^{-3}).
\end{aligned}
$$

And

$$\begin{aligned}
d_1 &= \sum_{j=0}^{T-1}\left[p^j(1-p)^{T-j}\begin{pmatrix} T \\ j \end{pmatrix}\begin{pmatrix} T-j \\ 1 \end{pmatrix}\frac{K}{N-j}\right. \\
&\qquad \left.\times\frac{N-K-j}{N-j-1}\times\frac{N-K-j-1}{N-j-2}\times\cdots\times\frac{N-K-(2T-2)}{N-(2T-1)}\right] \\
&= \sum_{j=0}^{T-1}\left[p^j(1-p)^{T-j}\begin{pmatrix} T \\ j \end{pmatrix}\begin{pmatrix} T-j \\ 1 \end{pmatrix}\frac{K}{N-(2T-1)}\right. \\
&\qquad \left.\times\frac{N-K-j}{N-j}\times\frac{N-K-j-1}{N-j-1}\times\cdots\times\frac{N-K-(2T-2)}{N-(2T-2)}\right] \\
&= \sum_{j=0}^{T-1}\left[p^j(1-p)^{T-j}\begin{pmatrix} T \\ j \end{pmatrix}\begin{pmatrix} T-j \\ 1 \end{pmatrix}\right. \\
&\qquad\qquad \left.\times\left[\frac{K}{N}+\frac{K(2T-1)}{N^2}\right]\left[1-\frac{K}{N}\right]^{2T-2-j+1}\right]+O(N^{-3}) \\
&= (1-p)T\sum_{j=0}^{T-1}p^j(1-p)^{T-1-j}\begin{pmatrix} T-1 \\ j \end{pmatrix} \\
&\qquad\qquad \times\left[\frac{K}{N}+\frac{K(2T-1)}{N^2}-\frac{K^2}{N^2}(2T-1-j)\right]+O(N^{-3}) \\
&= (1-p)T\left[\frac{K}{N}+\frac{K(2T-1)}{N^2}-\frac{K^2}{N^2}(2T-1-E(\mathrm{Bin}(T-1,p)))\right]+O(N^{-3}) \\
&= (1-p)T\left[\frac{K}{N}+\frac{K(2T-1)}{N^2}-\frac{K^2}{N^2}(2T-1-(T-1)p)\right]+O(N^{-3})
\end{aligned}$$

Hence

$$
\begin{aligned}
&P(Q_K^1 \mid C_{t-1} = 0, K_{t-1} = K) \\
= \;& (2-p)\frac{TK}{N} - \frac{TK\theta}{N^2} + \frac{TK(2T-1)}{N^2} - \frac{T^2K^2(2-p)}{N^2} + \frac{TK^2}{N^2} \\
& + \frac{TK(2T-1)(1-p)}{N^2} - \frac{(1-p)TK^2}{N^2}(T(2-p) - 1 + p) + O(N^{-3})
\end{aligned}
$$

(what follows is simply rearrangement...)

$$
\begin{aligned}
= \;& (2-p)\frac{TK}{N} - \frac{TK\theta}{N^2} + \frac{TK}{N^2}[(2T-1) - TK(2-p) + K + (2T-1)(1-p) \\
& \qquad -(1-p)KT(2-p) + (1-p)^2K] + O(N^{-3}) \\
= \;& (2-p)\frac{TK}{N} - \frac{TK\theta}{N^2} + \frac{TK}{N^2}[-TK[(2-p) + (1-p)(2-p)] \\
& \qquad +K[(1-p)^2 + 1] + T[2 + 2(1-p)] - (2-p)] + O(N^{-3}) \\
= \;& (2-p)\frac{TK}{N} - \frac{TK\theta}{N^2} - \frac{2T^2K^2(2-p)^2}{2N^2} \\
& + \frac{TK}{N^2}\left[K\left[\frac{(2-p)^2}{2} + \frac{p^2}{2}\right] + 2T(2-p) - (2-p)\right] + O(N^{-3}) \\
= \;& (2-p)\frac{TK}{N} - \frac{TK\theta}{N^2} - \frac{2T^2K^2(2-p)^2}{2N^2} + \frac{TK^2(2-p)^2}{2N^2} \\
& + \frac{TK}{N^2}\left[\frac{Kp^2}{2} + T\left[\frac{(2-p)^2}{2} + 2 - \frac{p^2}{2}\right] - (2-p)\right] + O(N^{-3}) \\
= \;& (2-p)\frac{TK}{N} - \frac{TK\theta}{N^2} - \frac{2T^2K^2(2-p)^2}{2N^2} + \frac{TK^2(2-p)^2}{2N^2} + \frac{T^2K(2-p)^2}{2N^2}
\end{aligned}
$$

$$+\frac{TK}{N^2}\left[-\frac{(2-p)^2}{2}-p+\frac{p^2}{2}+\frac{Kp^2}{2}+T\left(2-\frac{p^2}{2}\right)\right]+O(N^{-3})$$

$$= (2-p)\frac{TK}{N}-\frac{TK\theta}{N^2}-\frac{2T^2K^2(2-p)^2}{2N^2}+\frac{TK^2(2-p)^2}{2N^2}+\frac{T^2K(2-p)^2}{2N^2}$$

$$-\frac{TK(2-p)^2}{2N^2}+\frac{TK}{N^2}\left[-p+\frac{p^2}{2}+\frac{Kp^2}{2}+T\left(2-\frac{p^2}{2}\right)\right]+O(N^{-3})$$

$$= (2-p)\frac{TK}{N}-\frac{TK\theta}{N^2}-2\begin{pmatrix} T \\ 2 \end{pmatrix}\frac{K(K-1)(2-p)^2}{N^2}$$

$$+\frac{TK}{N^2}\left[-\frac{K(2-p)^2}{2}-\frac{T(2-p)^2}{2}+\frac{(2-p)^2}{2}\right.$$

$$\left.-p+\frac{p^2}{2}+\frac{Kp^2}{2}+T\left(2-\frac{p^2}{2}\right)\right]+O(N^{-3})$$

$$= (2-p)\frac{TK}{N}-\frac{TK\theta}{N^2}-2\begin{pmatrix} T \\ 2 \end{pmatrix}\frac{K(K-1)(2-p)^2}{N^2}-\frac{2TK(K-1)(1-p)}{N^2}$$

$$+\frac{TK}{N^2}\left[2(K-1)(1-p)-\frac{K(2-p)^2}{2}-\frac{T(2-p)^2}{2}\right.$$

$$\left.+\frac{(2-p)^2}{2}-p+\frac{p^2}{2}+\frac{Kp^2}{2}+T\left(2-\frac{p^2}{2}\right)\right]+O(N^{-3})$$

$$= (2-p)\frac{TK}{N}-\frac{TK\theta}{N^2}-2\begin{pmatrix} T \\ 2 \end{pmatrix}\frac{K(K-1)(2-p)^2}{N^2}-\frac{2TK(K-1)(1-p)}{N^2}$$

$$+\frac{TK}{2N^2}[K(4-4p-(2-p)^2+p^2)+T(4-p^2-(2-p)^2)$$

$$-4(1-p)+(2-p)^2-2p+p^2]+O(N^{-3})$$

$$= (2-p)\frac{TK}{N}-\frac{TK\theta}{N^2}-2\begin{pmatrix} T \\ 2 \end{pmatrix}\frac{K(K-1)(2-p)^2}{N^2}-\frac{2TK(K-1)(1-p)}{N^2}$$

$$+\frac{TK}{2N^2}\left[T(4p-2p^2)-2p+2p^2\right]+O(N^{-3})$$

$$= (2-p)\frac{TK}{N}-\frac{TK\theta}{N^2}-2\begin{pmatrix} T \\ 2 \end{pmatrix}\frac{K(K-1)(2-p)^2}{N^2}-\frac{2TK(K-1)(1-p)}{N^2}$$

$$+\frac{TK}{N^2}[Tp(2-p)-p(1-p)]+O(N^{-3})$$

as required.

Lastly

$P(Q_K^2 \mid C_{t-1}=0, K_{t-1}=K)$

$$= \sum_{d=0}^{2}\left\{\left[\sum_{j=0}^{T-d} p^j(1-p)^{T-j}\begin{pmatrix} T \\ j \end{pmatrix}\begin{pmatrix} T-j \\ d \end{pmatrix}\right.\right.$$

$$\times\frac{K}{N-j}\times\frac{K-1}{N-j-1}\times\cdots\times\frac{K-(d-1)}{N-j-(d-1)}$$

$$\left.\times\frac{N-K-j}{N-d-j}\times\frac{N-K-j-1}{N-d-j-1}\times\cdots\times\frac{N-K-(2T-e-1)}{N-T-(T-1)}\right]$$

$$\times\begin{pmatrix} T-d \\ 2-d \end{pmatrix}\frac{K-d}{N-T-d}\times\frac{K-d-1}{N-T-d-1}\times\cdots$$

$$\left.\times\frac{K-d-(2-d-1)}{N-T-d-(2-d-1)}\times(1-u)^{2-d}\right\}+O(N^{-3})$$

$$\begin{aligned} = & \sum_{d=0}^{2}\Biggl\{\Biggl[\sum_{j=0}^{T-d}\Biggl(p^j(1-p)^{T-j}\binom{T}{j}\binom{T-j}{d} \\ & \qquad \times\frac{K}{N-j}\times\frac{K-1}{N-j-1}\times\cdots\times\frac{K-(d-1)}{N-j-(d-1)}\Biggr)\Biggr] \\ & \times\binom{T-d}{2-d}\frac{K-d}{N-T-d}\times\frac{K-d-1}{N-T-d-1}\times\cdots\times\frac{K-d-(2-d-1)}{N-T-d-(2-d-1)}\Biggr\} \\ & +O(N^{-3}) \\ = & \sum_{d=0}^{2}\Biggl\{\Biggl[\sum_{j=0}^{T-d}\Biggl(p^j(1-p)^{T-j}\binom{T}{j}\binom{T-j}{d} \\ & \qquad \times\frac{K}{N}\times\frac{K-1}{N}\times\cdots\times\frac{K-(d-1)}{N}\Biggr)\Biggr] \\ & \times\binom{T-d}{2-d}\frac{K-d}{N}\times\frac{K-d-1}{N}\times\cdots\times\frac{K-d-(2-d-1)}{N}\Biggr\} \\ & +O(N^{-3}) \end{aligned}$$

$= d_0 + d_1 + d_2$ say, where d_i represents the term for $d = i$ in the equation immediately above ($i = 0, 1, 2$). Now

$$d_0 = \binom{T}{2}\frac{K}{N}\frac{K-1}{N} + O(N^{-3})$$

$$d_1 = \left(\sum_{j=0}^{T-1}\left(p^j(1-p)^{T-j}\binom{T}{j}\binom{T-j}{1}\frac{K}{N}\right)\binom{T-1}{1}\frac{K-1}{N}\right) + O(N^{-3})$$

$$d_2 = \sum_{j=0}^{T-2} \left(p^j (1-p)^{T-j} \binom{T}{j} \binom{T-j}{2} \frac{K}{N}\frac{K-1}{N} \right) + O(N^{-3}).$$

Hence

$$P(Q_K^2 \mid C_{t-1} = 0, K_{t-1} = K)$$

$$\begin{aligned}
&= \frac{K}{N}\frac{K-1}{N} \Bigg[\binom{T}{2} + \binom{T-1}{1} \sum_{j=0}^{T-1} \left(p^j (1-p)^{T-j} \binom{T}{j} \binom{T-j}{1} \right) \\
&\qquad + \sum_{j=0}^{T-2} \left(p^j (1-p)^{T-j} \binom{T}{j} \binom{T-j}{2} \right) \Bigg] + O(N^{-3}) \\
&= \frac{K(K-1)}{N^2} \Bigg[\binom{T}{2} + \binom{T-1}{1} \sum_{j=0}^{T-1} \left(p^j (1-p)^{T-j} \binom{T-1}{j} \binom{T}{1} \right) \\
&\qquad + \sum_{j=0}^{T-2} \left(p^j (1-p)^{T-j} \binom{T}{j} \binom{T-j}{2} \right) \Bigg] + O(N^{-3}) \\
&= \frac{K(K-1)}{N^2} \Bigg[\binom{T}{2} \\
&\qquad + \binom{T-1}{1} T(1-p) \sum_{j=0}^{T-1} \left(p^j (1-p)^{(T-1)-j} \binom{T-1}{j} \right) \\
&\qquad + \sum_{j=0}^{T-2} \left(p^j (1-p)^{T-j} \frac{T(T-1)}{2} \binom{T-2}{j} \right) \Bigg] + O(N^{-3}) \\
&= \frac{K(K-1)}{N^2} \Bigg[\binom{T}{2} + T(T-1)(1-p)
\end{aligned}$$

$$+\binom{T}{2}(1-p)^2\sum_{j=0}^{T-2}\left(p^j(1-p)^{T-2-j}\binom{T-2}{j}\right)\Bigg)\Bigg]+O(N^{-3})$$

$$=\frac{K(K-1)}{N^2}\left[\binom{T}{2}+\binom{T}{2}2(1-p)+\binom{T}{2}(1-p)^2\right]+O(N^{-3})$$

$$=\frac{K(K-1)}{N^2}\binom{T}{2}(2-p)^2+O(N^{-3})$$

as required.

.2.2 Section B

Next we proceed to the simplification of equation(4.6):

$P(M_K \mid K_{t-1}=K, C_{t-1}=c)$

$$=\sum_{d=0}^{T\wedge K}\left\{\sum_{i=0}^{c\wedge d}(\cdots)\right\}=\sum_{d=0}^{T\wedge K}\left\{\sum_{i=c\wedge d}^{c\wedge d}(\cdots)\right\}+O(N^{-3})$$

$$=\sum_{d=0}^{c}\Bigg\{\Bigg(p^d(1-p)^{c-d}\binom{c}{d}\Bigg)$$

$$\times\left[\sum_{j=0}^{T-c}\left(p^j(1-p)^{T-c-j}\binom{T-c}{j}\binom{T-j-d}{0}\right)\right]\Bigg]\Bigg)$$

$$\times\Bigg[\Bigg\{\binom{d}{1}\frac{K-d}{N-T}\frac{\theta}{N+\theta}$$

$$
\begin{aligned}
& \times \sum_{i=0}^{T-d} \frac{K-d-1}{N-T-d} \times \cdots \times \frac{K-d-1-(i-1)}{N-T-d-(i-1)} \begin{pmatrix} T-d \\ i \end{pmatrix} \Bigg\} \\
& + \Bigg\{ \sum_{i=1}^{T-d} \Bigg[\frac{K-d}{N-T-d} \times \cdots \times \frac{K-d-(i-1)}{N-T-d-(i-1)} \\
& \qquad \times \frac{\theta}{N+\theta} \begin{pmatrix} i \\ 1 \end{pmatrix} \begin{pmatrix} T-d \\ i \end{pmatrix} \Bigg] \Bigg\} \Bigg] \Bigg\} + O(N^{-3}) \\
= {} & \sum_{d=0}^{c} \Bigg\{ \Bigg(p^d (1-p)^{c-d} \begin{pmatrix} c \\ d \end{pmatrix} \Bigg) \\
& \times \Bigg[\Bigg\{ \begin{pmatrix} d \\ 1 \end{pmatrix} \frac{K-d}{N-T} \frac{\theta}{N+\theta} \Bigg\} + \Bigg\{ \frac{K-d}{N-T-d} \frac{\theta}{N+\theta} \begin{pmatrix} T-d \\ 1 \end{pmatrix} \Bigg\} \Bigg] \Bigg\} \\
& + O(N^{-3}) \\
= {} & \sum_{d=0}^{c} \Bigg\{ \Bigg(p^d (1-p)^{c-d} \begin{pmatrix} c \\ d \end{pmatrix} \Bigg) \\
& \times \Bigg[\Bigg\{ \begin{pmatrix} d \\ 1 \end{pmatrix} \frac{K-d}{N} \frac{\theta}{N} \Bigg\} + \Bigg\{ \frac{K-d}{N} \frac{\theta}{N} \begin{pmatrix} T-d \\ 1 \end{pmatrix} \Bigg\} \Bigg] \Bigg\} + O(N^{-3}) \\
= {} & \sum_{d=0}^{c} \Bigg\{ p^d (1-p)^{c-d} \begin{pmatrix} c \\ d \end{pmatrix} \frac{K-d}{N} \frac{\theta}{N} \Bigg[\begin{pmatrix} d \\ 1 \end{pmatrix} + \begin{pmatrix} T-d \\ 1 \end{pmatrix} \Bigg] \Bigg\} + O(N^{-3}) \\
= {} & \sum_{d=0}^{c} \Bigg\{ p^d (1-p)^{c-d} \begin{pmatrix} c \\ d \end{pmatrix} \frac{\theta T}{N^2} (K-d) \Bigg\} + O(N^{-3})
\end{aligned}
$$

$$= \frac{\theta T}{N^2}\sum_{d=0}^{c}\left\{p^d(1-p)^{c-d}\binom{c}{d}(K-d)\right\}+O(N^{-3})$$

$$= \frac{\theta T}{N^2}[K-E(\mathrm{Bin}(c,p))]+O(N^{-3})$$

$$= \frac{\theta T}{N^2}[K-cp]+O(N^{-3}).$$

.2.3 Section C

Lastly we turn to the simplification of equation(4.7) which we break into two cases, $(c > 0)$ and $(c = 0)$:

$$P(\Gamma_K \mid K_{t-1}=K, C_{t-1}=c>0)$$

$$= \sum_{d=0}^{T\wedge K}\Bigg[\Bigg\{\sum_{i=0}^{c\wedge d}\Bigg(p^i(1-p)^{c-i}\binom{c}{i}$$

$$\times\Bigg[\sum_{j=0}^{T-(c\vee d)}\Bigg(p^j(1-p)^{T-c-j}\binom{T-c}{j}\binom{T-j-i}{d-i}$$

$$\times\left[\frac{K-i}{N-j-i}\times\cdots\times\frac{K-i-(d-i-1)}{N-j-i-(d-i-1)}\right]\Bigg)\Bigg]\Bigg)\Bigg\}$$

$$\times\frac{K-d}{N-T}\binom{d}{1}\frac{N}{N+\theta}\Bigg]\Bigg\}\Bigg]+O(N^{-2})$$

$$= \sum_{d=0}^{c}\Bigg[\Bigg\{\sum_{i=0}^{c\wedge d}\Bigg(p^i(1-p)^{c-i}\binom{c}{i}$$

$$\times\left[\sum_{j=0}^{T-c}\left(p^j(1-p)^{T-c-j}\binom{T-c}{j}\binom{T-j-i}{d-i}\right.\right.$$

$$\left.\left.\times\frac{K-i}{N}\times\cdots\times\frac{K-i-(d-i-1)}{N}\right)\right]\Big)\Big\}$$

$$\times\left\{\frac{K-d}{N-T}\binom{d}{1}\frac{N}{N+\theta}\right\}\Bigg]+O(N^{-2})$$

$$= \sum_{d=0}^{c}\left[\{(i=d\text{ term})\}\times\left\{\frac{K-d}{N-T}\binom{d}{1}\right\}\right]+O(N^{-2})$$

$$= \sum_{d=0}^{c}\Bigg[\Big\{p^d(1-p)^{c-d}$$

$$\times\binom{c}{d}\left[\sum_{j=0}^{T-c}\left(p^j(1-p)^{T-c-j}\binom{T-c}{j}\right)\right]\Big\}\times\left\{\frac{K-d}{N-T}\binom{d}{1}\right\}\Bigg]$$

$$+O(N^{-2})$$

$$= \sum_{d=1}^{c}\left[p^d(1-p)^{c-d}\binom{c}{d}\times\frac{K-d}{N}d\right]+O(N^{-2}).$$

Now for $c=0$ we have $P(\Gamma_K \mid K_{t-1}=K, C_{t-1}=0)=\sum_{d=0}^{T\wedge K} D_d$ say, where D_l represents the $d=l$ term from (4.7) $(i=0,1,\ldots,c)$. Hence

$$P(\Gamma_K \mid K_{t-1}=K, C_{t-1}=0)$$

$$= D_1+O(N^{-3})$$

$$= \left\{ \sum_{j=0}^{T-1} \left(p^j (1-p)^{T-j} \begin{pmatrix} T \\ j \end{pmatrix} \begin{pmatrix} T-j \\ 1 \end{pmatrix} \frac{K}{N-j} \right) \right\} \frac{K-1}{N-T}(1-u)$$

$$+O(N^{-3})$$

$$= \frac{K(K-1)}{N^2} \left\{ \sum_{j=0}^{T-1} p^j (1-p)^{T-j} \begin{pmatrix} T \\ j \end{pmatrix} \begin{pmatrix} T-j \\ 1 \end{pmatrix} \right\} + O(N^{-3})$$

$$= \frac{TK(K-1)(1-p)}{N^2} \sum_{j=0}^{T-1} \left\{ p^j (1-p)^{(T-1)-j} \begin{pmatrix} T-1 \\ j \end{pmatrix} \right\} + O(N^{-3})$$

$$= \frac{TK(K-1)(1-p)}{N^2} + O(N^{-3}).$$

.2.4 Section D

Finally we give the details of the simplification of the first-step analysis of the transition equations (ie. equations (4.11) and (4.13)). We start with (4.11):

$$\Pi_1 = \frac{p(K-1)}{N} + \frac{p(K-1)(T-1)(2-p)\Pi_2}{N}$$

$$+\Pi_1 \left[p + (1-p)\frac{TK}{N} + \frac{TK(1-p)^2}{N} + \frac{p(1-p)K}{N} \right.$$

$$\left. - \frac{p(K-1)T(2-p)}{N} + \frac{p(1-p)(K-1)}{N} \right]$$

$$+\Pi_0(1-p) \left[1 - \frac{TK(2-p)}{N} - \frac{pK}{N} \right] + O(N^{-2})$$

Hence since $\Pi_2 = \Pi_1 + O(N^{-1})$ we have

$$\Pi_1 \left[(1-p) - (1-p)\frac{TK}{N} - \frac{TK(1-p)^2}{N} - \frac{p(1-p)K}{N} \right.$$

$$+\frac{p(K-1)T(2-p)}{N}-\frac{p(1-p)(K-1)}{N}-\frac{p(K-1)(T-1)(2-p)}{N}\Bigg]$$

$$= \frac{p(K-1)}{N}+\Pi_0(1-p)\left[1-\frac{TK(2-p)}{N}-\frac{pK}{N}\right]+O(N^{-2})$$

So

$$\Pi_1\left[(1-p)-(1-p)\frac{TK(2-p)}{N}-\frac{p(1-p)(2K-1)}{N}+\frac{p(K-1)(2-p)}{N}\right]$$

$$= \frac{p(K-1)}{N}+(1-p)\left[1-\frac{TK(2-p)}{N}-\frac{pK}{N}\right]\Pi_0+O(N^{-2})$$

and so

$$\begin{aligned}\Pi_1 &= \left[\frac{N}{(1-p)N-(1-p)TK(2-p)-p(1-p)(2K-1)+p(K-1)(2-p)}\right]\\ &\quad\times\left[\frac{p(K-1)}{N}+(1-p)\left[1-\frac{TK(2-p)}{N}-\frac{pK}{N}\right]\Pi_0\right]+O(N^{-2})\\ &= \frac{1}{1-p}\left[\frac{N}{N-[TK(2-p)+p(2K-1)-\frac{p(2-p)}{(1-p)}(K-1)]}\right]\\ &\quad\times\left[\frac{p(K-1)}{N}+(1-p)\left[1-\frac{TK(2-p)}{N}-\frac{pK}{N}\right]\Pi_0\right]+O(N^{-2})\\ &= \frac{1}{1-p}\left[1+\frac{TK(2-p)+p(2K-1)-\frac{p(2-p)}{(1-p)}(K-1)}{N}\right]\\ &\quad\times\left[\frac{p(K-1)}{N}+(1-p)\left[1-\frac{TK(2-p)}{N}-\frac{pK}{N}\right]\Pi_0\right]+O(N^{-2})\end{aligned}$$

$$\left(\text{since } \frac{N}{N-A}=1+\frac{A}{N}+O(N^{-2})\right)$$

$$
\begin{aligned}
&= \frac{p(K-1)}{(1-p)N} + \Pi_0 \left[1 - \frac{TK(2-p)}{N} - \frac{pK}{N} \right. \\
&\qquad \left. + \frac{TK(2-p) + p(2K-1) - \frac{p(2-p)}{(1-p)}(K-1)}{N} \right] + O(N^{-2}) \\
&= \frac{p(K-1)}{(1-p)N} + \Pi_0 \left[1 + \frac{p(K-1)}{N} - \frac{p(2-p)(K-1)}{(1-p)N} \right] + O(N^{-2}).
\end{aligned}
$$

Now for (4.13):

$$
\begin{aligned}
\Pi_0 &= \frac{(1-p)K(K-1)T}{N^2} \\
&\quad + \Pi_0 \left[1 - \frac{KT(2-p)}{N} + \binom{T}{2} \frac{K(K-1)(2-p)^2}{N^2} \right. \\
&\qquad \left. + \frac{TK(K-1)(1-p)}{N^2} - \frac{KT}{N^2}[Tp(2-p) - p(1-p)] \right] \\
&\quad + \Pi_1 \left[\frac{(2-p)TK}{N} - \frac{TK\theta}{N^2} - 2\binom{T}{2} \frac{K(K-1)(2-p)^2}{N^2} \right. \\
&\qquad \left. - \frac{2TK(K-1)(1-p)}{N^2} + \frac{KT}{N^2}[Tp(2-p) - p(1-p)] \right] \\
&\quad + \Pi_2 \left[\binom{T}{2} \frac{K(K-1)(2-p)^2}{N^2} \right] + O(N^{-3}) \\
&= \frac{(1-p)K(K-1)T}{N^2} \\
&\quad + \Pi_0 \left[1 - \frac{KT(2-p)}{N} + \binom{T}{2} \frac{K(K-1)(2-p)^2}{N^2} \right.
\end{aligned}
$$

$$+\frac{TK(K-1)(1-p)}{N^2}-\frac{KT}{N^2}[Tp(2-p)-p(1-p)]\Bigg]$$

$$+\Pi_1\left[\frac{(2-p)TK}{N}-\frac{TK\theta}{N^2}\right]$$

$$-\left[\Pi_0+O(N^{-1})\right]\left[2\begin{pmatrix}T\\2\end{pmatrix}\frac{K(K-1)(2-p)^2}{N^2}+\frac{2TK(K-1)(1-p)}{N^2}\right.$$

$$\left.-\frac{KT}{N^2}[Tp(2-p)-p(1-p)]\right]$$

$$+\left[\Pi_0+O(N^{-1})\right]\left[\begin{pmatrix}T\\2\end{pmatrix}\frac{K(K-1)(2-p)^2}{N^2}\right]+O(N^{-3})$$

$$=\frac{(1-p)K(K-1)T}{N^2}+\Pi_0\left[1-\frac{TK(2-p)}{N}-\frac{TK(K-1)(1-p)}{N^2}\right]$$

$$+\Pi_1\left[\frac{(2-p)TK}{N}-\frac{TK\theta}{N^2}\right]+O(N^{-3}).$$

Hence

$$\Pi_0\left[\frac{TK(2-p)}{N}+\frac{TK(K-1)(1-p)}{N^2}\right]=\frac{(1-p)K(K-1)T}{N^2}+\Pi_1\left[\frac{(2-p)TK}{N}\right]$$

$$-\left[\Pi_0+O(N^{-1})\right]\left[\frac{TK\theta}{N^2}\right]+O(N^{-3}).$$

And so

$$\Pi_0\left[\frac{TK(2-p)}{N}+\frac{TK\theta}{N^2}+\frac{TK(K-1)(1-p)}{N^2}\right]=$$

$$\frac{(1-p)K(K-1)T}{N^2}+\Pi_1\left[\frac{(2-p)TK}{N}\right]+O(N^{-3}).$$

We now substitute in for Π_1 using (4.12) to get:

$$\Pi_0 \left[(2-p) + \frac{\theta}{N} + \frac{(K-1)(1-p)}{N} \right] =$$

$$\frac{(1-p)(K-1)}{N} + (2-p)\left[\frac{p(K-1)}{(1-p)N} + \Pi_0 \left[1 + \frac{p(K-1)}{N} - \frac{p(2-p)(K-1)}{(1-p)N} \right] \right]$$

$$+O(N^{-2})$$

Therefore

Π_0

$$= \frac{\frac{(1-p)(K-1)}{N} + \frac{p(2-p)(K-1)}{(1-p)N} + O(N^{-2})}{(2-p) + \frac{\theta}{N} + \frac{(K-1)(1-p)}{N} - (2-p) - \frac{(2-p)p(K-1)}{N} + \frac{(2-p)^2 p(K-1)}{(1-p)N}}$$

$$= \frac{(1-p)^2(K-1) + p(2-p)(K-1) + O(N^{-1})}{(1-p)\theta + (K-1)(1-p)^2 - (2-p)p(1-p)(K-1) + (2-p)^2 p(K-1)}$$

$$= \frac{(K-1)[(1-p)^2 + p(2-p)] + O(N^{-1})}{(1-p)\theta + (K-1)[(1-p)^2 - (2-p)p(1-p) + (2-p)^2 p]}$$

$$= \frac{(K-1) + O(N^{-1})}{(1-p)\theta + (K-1)[(1-p)^2 + (2-p)p]}$$

(since $(1-p)^2 + p(2-p) = 1$)

$$= \frac{(K-1) + O(N^{-1})}{(1-p)\theta + (K-1)}$$

ie. $$\Pi_0 = \frac{K-1}{(1-p)\theta + (K-1)} + O(N^{-1}) \text{ as required.}$$

.3 Appendix III

We need to show

$$\sum_{i=2}^{N-1}\sum_{k=0}^{(i-1)\wedge(N-i)} \mu_{i-1-k}\mu_{i+k}a'_{i,k} \leq \sum_{i=2}^{N-1}\sum_{k=0}^{(i-2)\wedge(N-i-1)} \mu_{i-1-k}\mu_{i+k}b'_{i,k} \qquad (.14)$$

Again it is sufficient to show it for each $i = 0, 1, \ldots, [\frac{N}{2}+1]$. That is, to show

$$\sum_{k=0}^{(i-1)\wedge(N-i)} \mu_{i-1-k}\mu_{i+k}a'_{i,k} - \sum_{k=0}^{(i-2)\wedge(N-i-1)} \mu_{i-1-k}\mu_{i+k}b'_{i,k} \qquad (.15)$$

If we now define

$$x_k = \mu_{i+k}\mu_{i-1-k}, \quad a_k = a'_{i,k} \text{ and } b_k = b'_{i,k}$$

we again have sequences as in the statement of Lemma 6.4. So if the four conditions of that Lemma are satisfied then its conclusion gives us the desired inequality (.14). Condition (iv) is clear since the x_k are decreasing by supposition. To check conditions (i), (ii) and (iii) we need to determine the exact values of the a_k's and b_k's.

The a_k's are as follows:

$$a_k = \begin{cases} \binom{N-2}{i-1-k}\binom{N-2}{i+k-2} + \binom{N-2}{i+k}\binom{N-2}{i-k-3} & i-1-k \geq 2 \\ & i+k \leq N-2 \\ \binom{N-2}{i-1-k}\binom{N-2}{i+k-2} & 0 \leq i-1-k < 2 \\ & \underline{or} \\ & N-2 < i+k \leq N \\ 0 & \text{otherwise.} \end{cases}$$

And the b_k's are :

$$b_k = \begin{cases} 2\binom{N-2}{i-k-2}\binom{N-2}{i+k-1} & \begin{matrix} i-1-k \geq 1 \\ i+k \leq N-1 \end{matrix} \\ 0 & \text{otherwise} \end{cases}$$

Note that for $i = 2$ we have to show

$$\sum_{k=0}^{1} \binom{N-2}{2-1-k}\binom{N-2}{2+k-2} \leq 2\binom{N-2}{2-2-k}\binom{N-2}{2+k-1},$$

which is true since both sides equal $2(N-2)$. Now note that

$$\binom{N}{M} = \binom{N}{M-1}\frac{N-M+1}{M} \Rightarrow \binom{N-2}{i-k-1} = \binom{N-2}{i-k-2}\frac{N-i+k}{i-k-1}$$

and

$$\binom{N-2}{i+k-2} = \binom{N-2}{i+k-1}\frac{i+k-1}{N-i-k}$$

and

$$\binom{N-2}{i+k} = \binom{N-2}{i+k-1}\frac{N-i-k-1}{i+k}$$

and

$$\binom{N-3}{i-k-3} = \binom{N-2}{i-k-2}\frac{i-k-2}{N-i+k-1}.$$

Hence if we write

$$C_{i,k} = \binom{N-2}{i-k-2}\binom{N-2}{i+k-1}$$

(and note that $C_{i,k} \geq 0$), then we have

$$a_k - b_k$$

$$= \left[\binom{N-2}{i-k-1}\binom{N-2}{i+k-2} + \binom{N-2}{i+k}\binom{N-2}{i-k-3}\right]$$

$$-2\begin{pmatrix} N-2 \\ i-k-2 \end{pmatrix}\begin{pmatrix} N-2 \\ i+k-1 \end{pmatrix}$$

$$= C_{i,k}\left[\frac{(N-i+k)(i+k-1)}{(i-k-1)(N-i-k)} + \frac{(N-i-k-1)(i-k-2)}{(i+k)(N-i+k+1)} - 2\right]$$

$$= C_{i,k}\left[\frac{(N-i+k)(i+k-1)-(i-k-1)(N-i-k)}{(i-k-1)(N-i-k)}\right.$$

$$\left. + \frac{(N-i-k-1)(i-k-2)-(i+k)(N-i+k+1)}{(i+k)(N-i+k+1)}\right]$$

$$= 2C_{i,k}\left[\frac{k(N-1)}{(i-k-1)(N-i-k)} - \frac{k(N-1)+N-1}{(i+k)(N-i+k+1)}\right]$$

$$= 2C_{i,k}(N-1)\left[\frac{k(i+k)(N-i+k+1)-(k+1)(i-k-1)(N-i-k)}{(i-k-1)(N-i-k)(i+k)(N-i+k+1)}\right]$$

$$= 2C_{i,k}(N-1)\left[\frac{(2N-1)k(k+1)-(N-i)(i-1)}{(i-k-1)(N-i-k)(i+k)(N-i+k+1)}\right]$$

So since we have $k \le i-1$ and $i+k \le N$, we can see that the numerator starts negative (at $k=0$) and increases (eventually becoming positive) as k increases. The denominator is always positive. So conditions (i) and (ii) are satisfied. Now for condition (iii) we have:

$$\sum_{k=0}^{i-1} a_k = \sum_{k=0}^{i-3}\left[\begin{pmatrix} N-2 \\ i-k-1 \end{pmatrix}\begin{pmatrix} N-2 \\ i+k-2 \end{pmatrix} + \begin{pmatrix} N-2 \\ i+k \end{pmatrix}\begin{pmatrix} N-2 \\ i-k-3 \end{pmatrix}\right]$$

$$+\begin{pmatrix} N-2 \\ 1 \end{pmatrix}\begin{pmatrix} N-2 \\ 2i-4 \end{pmatrix} + \begin{pmatrix} N-2 \\ 0 \end{pmatrix}\begin{pmatrix} N-2 \\ 2i-3 \end{pmatrix} \qquad (.16)$$

$$\sum_{k=0}^{i-2} b_k = \sum_{k=0}^{i-2} 2 \binom{N-2}{i-k-2}\binom{N-2}{i+k-1} \tag{.17}$$

Now consider the two terms inside the summation on the right-hand side of (.16) and note that the value of the first of these terms corresponding to $k = l$ equals half the term on the right-hand side of (.17) corresponding to $k = l-1$. Similarly the value of the second term in (.16) corresponding to $k = l$ equals half the term in (.17) corresponding to $k = l+1$. We therefore define

$$D = \sum_{k=1}^{i-3} \left[\binom{N-2}{i-k-1}\binom{N-2}{i+k-2} + \binom{N-2}{i+k}\binom{N-2}{i-k-3} \right].$$

So

$$\begin{aligned} \sum_{k=0}^{i-1} a_k &= D + \binom{N-2}{i-1}\binom{N-2}{i-2} + \binom{N-2}{i}\binom{N-2}{i-3} \\ &\quad + \binom{N-2}{1}\binom{N-2}{2i-4} + \binom{N-2}{0}\binom{N-2}{2i-3} \end{aligned}$$

and

$$\begin{aligned} \sum_{k=0}^{i-2} b_k &= D + \binom{N-2}{i-3}\binom{N-2}{i} + \binom{N-2}{i-1}\binom{N-2}{i-2} \\ &\quad + \binom{N-2}{2i-4}\binom{N-2}{1} + \binom{N-2}{2i-3}\binom{N-2}{0}. \end{aligned}$$

Hence

$$\sum_{k=0}^{i-1} a_k - \sum_{k=0}^{i-1} b_k = 0$$

as required.

So all the conditions of Lemma 6.4 are satisfied and we thus have the required conclusion.

REFERENCES

ABRAMOWITZ, M., AND STEGUN, L.A., 1972. "Handbook of Mathematical Functions," Dover, New York.

ANDERSON, I., 1987. "Combinatorics of Finite Sets," Oxford Science Publications.

BALL, F., AND DONNELLY, P., 1987. Inter particle correlation in death processes with application to variability in compartmental models, *Adv. Appl. Prob.* **19**, 755-766.

BARTON, N., 1988. Random drift due to inherited variation in fitness: A generalization of the hitch-hiking effect, (preprint).

BLOCK, H., SAVITS, T., AND SHAKED, M., 1982. Some concepts of negative dependence, *Ann. Prob.* **10**, 765-772.

BROADBENT, S.R., AND HAMMERSLEY, J.M., 1957. Percolation processes, I. Crystals and mazes, *Proc. Camb. Phil. Soc.* **53**, 629-641.

COX, D.R., AND MILLER, H.D., 1965. "The Theory of Stochastic Processes," Wiley, New York.

DONNELLY, P., 1985. Dual processes and an invariance result for exchangeable models in population genetics, *J. Math. Biol.* **23**, 103-118.

DONNELLY, P., KURTZ, T., AND MARJORAM, P., 1992. Correlation and Variability in Birth Processes, (submitted).

DONNELLY, P., AND MARJORAM, P., 1989. The Effect on Genetic Sampling Distributions of Correlations in Reproduction, *Theor. Pop. Biol.* **35**, 22-35.

DONNELLY, P., AND TAVARÉ, S., 1986. The ages of alleles and a coalescent, *Adv. Appl. Prob.* **18**, 1-19.

DONNELLY, P., AND TAVARÉ, S., 1987. The population genealogy of the infinitely-many neutral alleles model, *J. Math. Biol.* **25**, 381-391.

DONNELLY, P., AND WELSH, D., 1984. The anti-voter problem: random 2-colourings of graphs, *in "Graph Theory and Combinatorics," (Bollobas, B., ed.)*, Academic Press.

ELLIS, R.S., 1985. "Entropy, Large Deviations and Statistical Mechanics," Springer-Verlag.

EWENS, W.J., 1972. The sampling theory of selectively neutral alleles, *Theor. Pop. Biol.* **3**, 87-112.

EWENS, W.J., 1979. "Mathematical Population Genetics," Springer-Verlag.

EWENS, W.J., 1978. The neutralist-selectionist controversy, *in "Studies in Mathematical Biology" (Levin, S., ed.),* Vol. 16, pp477-502, Mathematical Association of America.

FADDY, M.J., 1985. Non-linear stochastic compartmental models, *IMA. J. Math. Appl. Med. Biol.* **2**, 287-297.

FADDY, M.J., 1990. On variation in birth processes, *Adv. Appl. Prob.* **22**, 480-484.

FISHER, M.E., 1961. Critical probabilities for cluster size and percolation problems, *J. Math. Phys.* **2**, 620-627.

FISHER, R.A., 1930. "The Genetical Theory of Natural Selection," Clarendon Press, Oxford.

FORTUIN, C.M., KASTELEYN, P.W., AND GINIBRE, J., 1971. Correlation inequalities on some partially ordered sets, *Comm. Math. Phys.* **22**, 89-103.

GILLESPIE, J.H., 1977. Sampling theory for alleles in a random environment, *Nature (London)* **266**, 443-445.

GRIFFITHS, R.B., 1972. Rigorous results and theorems, *in "Phase Tran-*

sitions and Critical Phenomena," (Domb, C., and Green, M.S., ed.), Vol. **1**, 7-109, Academic Press, New York.

GRIFFITHS, R.C., 1980. Lines of descent in the diffusion approximation of neutral Wright-Fisher models, *Theor. Pop. Biol.* **17**, 37-50.

HARRIS, T.E., 1960. A lower bound for the critical probabilty in a certain percolation process, *Proc. Camb. Phil. Soc.* **56**, 13-20.

JOAG-DEV, K., AND PROSCHAN, F., 1983. Negative association of random variables, with applications, *Ann. Stats.* **11**, 286-295.

KARLIN, S., AND RINOTT, Y., 1980. Classes of orderings of measures and related correlation inequalities. II Multivariate reverse rule distributions, *J. Mult. Anal.* **10**, 499-516.

KELLY, F., 1979. "Reversibility and Stochastic Networks," Wiley, New York.

KESTEN, H., 1982. "Percolation for Mathematicians," Birkhauser, Boston.

KIMURA, M., 1983. "The Neutral Theory of Molecular Evolution," Cambridge University Press.

KINGMAN, J.F.C., 1982a. On the genealogy of large populations, *J. Appl. Prob.* **19A**, 27-43.

KINGMAN, J.F.C., 1982b. The coalescent, *Stoch. Proc. Appl.* **13**, 235-248.

KINGMAN, J.F.C., 1982c. Exchangeability and the evolution of large populations, *in "Exchangeability in Probability and Statistics," (Koch, G., and Spizzichino, F., ed.)*, pp 97-112, North Holland, Amsterdam.

LIGGETT, T.M., 1985. "Interacting Particle Systems," Springer, Verlag.

MORAN, P.A.P., 1958. Random processes in genetics, *Proc. Camb. Phil. Soc.* **54**, 60-71.

SAWYER, S., AND HARTL, D., 1985. A sampling theory for local selection, *J. Genet.* **64**, 21-29.

SEYMOUR, P.D., AND WELSH, D.J.A., 1975. Combinatorical applications of an inequality from statistical mechanics, *Math. Proc. Camb. Phil. Soc.* **77**, 485-495.

TAVARÉ, S., 1984. Line-of-descent and genealogical processes and their applications in population genetics models, *Theor. Pop. Biol.* **26**, 119-164.

TAYLOR, H.M., AND KARLIN, S., 1984. "An Introduction to Stochastic Modelling," Academic Press.

WATTERSON, G.A., 1977. Heterosis or neutrality?, *Genetics* **85**, 789-814.

WATTERSON, G.A., 1984. Lines of descent and the coalescent, *Theor. Pop. Biol.* **26**, 77-92.

WELSH, D.J.A., 1986. Correlated percolation and repulsive particle systems, *in "Stochastic Spatial Processes," (Tanta, P., ed.), Lecture notes in Mathematics* **1212**, 300-311, Springer-Verlag, Berlin.

WRIGHT, S., 1931. Evolution in mendelian populations, *Genetics* **16**, 97-159.