Analysis of some nonrandom mating models

Gauri Lall Ghai

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation

Ghai, Gauri Lall, "Analysis of some nonrandom mating models " (1971). Retrospective Theses and Dissertations. 4538.
https://lib.dr.iastate.edu/rtd/4538

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
GHAJ, Gauri Lall, 1936-
ANALYSIS OF SOME NONRANDOM MATING MODELS.
Iowa State University, Ph.D., 1971
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED
Analysis of some nonrandom mating models

by

Gauri Lall Ghai

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1971
PLEASE NOTE:

Some pages have indistinct print. Filmed as received.

UNIVERSITY MICROPILMS.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>II. MIXED RANDOM MATING AND SELFING</strong></td>
<td>9</td>
</tr>
<tr>
<td>A. Distributions of Genotypes</td>
<td>9</td>
</tr>
<tr>
<td>1. Single locus</td>
<td>10</td>
</tr>
<tr>
<td>2. Two loci</td>
<td>17</td>
</tr>
<tr>
<td>3. Three loci</td>
<td>37</td>
</tr>
<tr>
<td>4. Multiple loci</td>
<td>43</td>
</tr>
<tr>
<td>B. Loss of Heterozygosity</td>
<td>45</td>
</tr>
<tr>
<td>1. Single locus</td>
<td>46</td>
</tr>
<tr>
<td>2. Two loci</td>
<td>47</td>
</tr>
<tr>
<td>3. Three loci</td>
<td>51</td>
</tr>
<tr>
<td>4. Multiple loci</td>
<td>52</td>
</tr>
<tr>
<td>C. Genotypic Mean and Variance</td>
<td>59</td>
</tr>
<tr>
<td>1. Single locus</td>
<td>60</td>
</tr>
<tr>
<td>2. Two loci</td>
<td>66</td>
</tr>
<tr>
<td>3. Multiple loci</td>
<td>72</td>
</tr>
<tr>
<td><strong>III. MIXED RANDOM AND SIB MATING</strong></td>
<td>74</td>
</tr>
<tr>
<td>A. Mixed Random and Full-Sib Mating</td>
<td>74</td>
</tr>
<tr>
<td>B. Mixed Random and Half-Sib Mating</td>
<td>82</td>
</tr>
<tr>
<td>C. Loss of Heterozygosity</td>
<td>86</td>
</tr>
<tr>
<td>D. Genotypic Mean and Variance</td>
<td>91</td>
</tr>
<tr>
<td>1. Absence of dominance</td>
<td>93</td>
</tr>
<tr>
<td>2. Complete dominance</td>
<td>94</td>
</tr>
<tr>
<td><strong>IV. SOME OTHER MODELS</strong></td>
<td>99</td>
</tr>
<tr>
<td>A. Mixed Random and Parent-Offspring Mating</td>
<td>99</td>
</tr>
<tr>
<td>B. Mixed Random and Double First Cousin Mating</td>
<td>100</td>
</tr>
</tbody>
</table>
C. General Mixture of Consanguineous Mating Systems 101

D. Comparison of Various Models 106

V. PARTIAL ASSORTATIVE MATING 111

A. Single Locus Models 112

1. Model I: two alleles with complete dominance 113
2. Model II: multiple alleles and no dominance 115
3. Model III 130

B. Components of Variance and Covariances of Relatives 151

1. Karlin and Scudo (1969) asymmetric model 151
2. Model III(a) 155
3. Model III(b)(i) 157

C. Two Loci Model: Complete Positive Assortative Mating 159

VI. SUMMARY AND CONCLUSIONS 165

VII. BIBLIOGRAPHY 170

VIII. ACKNOWLEDGMENTS 174
I. INTRODUCTION

The foundation for the mathematical theory of population genetics was laid in numerous contributions of R. A. Fisher, J. B. S. Haldane and Sewall Wright. The mathematical theory, which is an essential part of the theory of evolution, has been largely developed for populations practicing only one system of mating, such as pure random mating or inbreeding. Inbreeding is one form of nonrandom mating where the mates are more closely related than if they were chosen at random. The most extreme example of inbreeding is that of selfing.

The theory of pure random mating and pure inbreeding is very well worked out and the results are very well documented (Fisher, 1949; Kempthorne, 1957; Wright, 1969). However, situations are not uncommon where outcrossing and inbreeding processes are working simultaneously. Some evidence is available that populations of inbreeders are not as uniform as has often been imagined. There are crops such as rice, wheat, barley, lima beans, etc., which are largely self-fertilized. On the other hand, there are crops such as cotton which are moderately self-fertilized and crops such as corn which are largely cross-fertilized. Kakizaki (1924) reported that natural crossing takes place quite frequently among egg-plants. Less than 10 percent of
natural crossing was quite general, though much higher percentages were not rare. The same author (Kakizaki, 1929) showed that the amount of natural crossing was 20 percent or less in most of the flowers in tomato. The bulk of the experimental material on cotton has shown that the amount of natural crossing ranged from less than 10 percent to approximately 50 percent in extensive areas of the Cotton Belt in the United States (Simpson, 1954). It will be therefore of very considerable theoretical as well as practical interest to study populations in which some of the matings occur at random while the remainder take place between relatives of various degrees.

The genetical structure of a population under a mixture of random mating and inbreeding has been studied by a number of authors. Haldane (1924) discussed the equilibrium distribution for one autosomal gene pair under the systems of mixed random mating and selfing, and mixed random and full-sib mating. Similar results were also discussed by Haldane and Moshinsky (1939) for equilibrium populations. Garber (1951) dealt with the approach to equilibrium with varying percentages of cross- and self-fertilization in the case of one autosomal locus for a panmictic initial population, whereas Ali and Hadley (1955) discussed the effect of such a system on the amount of heterozygosity.

Bennett and Binet (1956) considered the association
between Mendelian factors for two pairs of genes at equilibrium with mixed selfing and random mating. They concluded that when equilibrium is reached the genes are associated at random but the two factors which are segregating independently are not associated at random. Narain (1966) obtained a formula for the probability that genes are identical by descent at several loci, assuming an initial panmictic population. From the numerical results for two loci, in the absence of selection, Narain observed that in the initial stages the inbreeding process and linkage both tend to act in the same direction to increase the homozygosity at both loci but afterwards the inbreeding process tends to dampen the effect of linkage.

In the above investigations it was assumed that all genotypes are equally viable. Hayman (1953) investigated the equilibrium conditions for a single locus at which selection favors the heterozygote at the expense of the two homozygotes. Later, Page and Hayman (1960) studied a similar problem under a system of mixed sib and random mating. They observed that the approach to equilibrium is very similar to that of the system of mixed selfing and random mating.

Ghai (1964, 1965, 1966) has explored this model of mixed random mating and selfing in more detail. He has extended these results to several independently segregating loci and examined the effect of such a system on the
composition and variability of a population. More recently Ghai (1969) has considered models based on a system of mixed random and sib mating.

Another form of nonrandom mating called assortative mating is based on the phenotypic properties of mates. When the mating is between individuals who are phenotypically alike, it is called positive assortative mating, while the mating with respect to opposing characters is called negative assortative mating. This tendency to assortative mating is known to occur in certain bird, mammal, and insect populations. For example, in man, Pearson and Lee (1903) have demonstrated positive assortative mating for various physical characters such as stature and forearm length. Rendel (1944) found that in *Drosophila subobscura* yellow males tend to mate with yellow females more readily than with another phenotype. Similar examples can be found in the domestic pigeon. In human populations where homogamy prevails, conventional barriers greatly restrict the choice of a mate. Hence human parents tend to resemble one another in heritable characters, e.g., height, intelligence, etc. These observations suggest that nonrandom mating, due to assortative mating, probably is of considerable importance in determining the behavior of a number of traits.

It is commonly thought that assortative mating is in some ways rather like inbreeding. For example, we quote
from Crow and Kimura (1970, p. 142): "Since individuals with similar phenotypes will usually be somewhat similar in their genotypes, we should expect assortative mating to have generally the same consequences as inbreeding." Note that Crow and Kimura define assortative mating as mating between individuals of similar phenotypes. But there will also be differences in the effects of inbreeding and assortative mating. Mating like to like tends to bring together mates whose genes have similar effects but need not be allelic to each other. The likeness between the phenotypes can be high while the likeness between mates may be very low when compared gene by gene. This difference between mating relatives and mating phenotypic like to like becomes more extreme as more and more loci affect the characteristic.

The simplest model of complete positive assortative mating for a single locus with complete dominance, without any other disturbance, was discussed initially by Jennings (1916). These results were later extended by Wentworth and Remick (1916) who gave a more general formula. Complete positive assortative mating leads to complete homozygosis of a population, though very slowly. Consider a single locus with two alleles $A_1$ and $A_2$ with complete dominance of $A_1$ over $A_2$. Then the frequency of $A_1A_2$ individuals after $n$ generations of positive assortative
mating is given by

\[ H_n = \frac{2p_1 H_0}{2p_1 + nH_0} \]

where \( H_0 \) is the initial heterozygosity, and \( p_1 \) is the gene frequency of \( A_1 \) which remains invariant over time. If all matings are between dominants and recessives, that is negative assortative mating, then the first generation of matings \( A_1A_1 \times A_2A_2 \) and \( A_1A_2 \times A_2A_2 \) yields only genotypes \( A_1A_2 \) and \( A_2A_2 \). Thus the matings in the first generation will be only of type \( A_1A_2 \times A_2A_2 \), which lead immediately to the equilibrium state \((1/2)A_1A_2 + (1/2)A_2A_2 \) in the next generation, independent of the initial composition of the population.

Complete positive assortative mating or complete negative assortative mating may not always be found in a population. Stanton (1946) investigated the effect of partial assortative mating on the correlations between relatives when there is an autosomal locus or a sex linked locus with no dominance. He derived these results assuming a constant interparental correlation in successive generations. There have been later studies of simple patterns of mixed random and positive assortative mating (e.g., O'Donald, 1960; Parsons, 1962) and of negative assortative mating (e.g., Naylor, 1962; Workman, 1964; Karlin and Feldman, 1968a, b).
In these models negative assortative mating is described by specifying the matings which are not permitted.

More recently Scudo and Karlin (1969) and Karlin and Scudo (1969) have formulated and analyzed several one locus models of partial assortative mating based on phenotypes. It is assumed that sexual preferences are manifested only in one sex. They have taken into account such factors as different degrees of assorting in the various phenotypes, the availability of proper mates, the timing of mating, relative fertilities of matings which occur assortatively compared to those occurring at random etc.

The situation becomes very complicated in multifactorial cases. In a classic paper Fisher (1918) investigated the effects of partial assortative mating on the correlations between relatives with respect to characters assumed to be normally distributed. It is assumed that the population is in equilibrium and that the attributes are determined by a large number of factors, the genes of which have small or infinitesimal effects, and by environmental factors. Wright (1921) considered the effects of assortative mating under the simplifying assumptions that there is no dominance, the effects of different factors are equal and additive, and that all genes occur with an equal frequency of one-half. He used the method of path coefficients in his development.

The aim of the present study is to investigate the
effects of nonrandom mating due to (a) partial inbreeding, and (b) partial assortative mating. The mathematical theory for partial inbreeding has largely remained a one-locus theory for a system of mixed random mating and selfing. We will extend this to multi-locus situations. Theoretical models will be developed for populations in which some of the matings occur at random while the remainder take place between relatives of various degrees. These include mixed random and (i) selfing, (ii) sib mating, (iii) parent-offspring mating, (iv) double first cousin mating, and (v) a general mixture of consanguineous mating systems.

With regard to partial assortative mating, we will consider a number of asymmetric deterministic models for a single locus. Some results for a two loci model under complete positive assortative mating will also be presented. This model is more general than that of Wright (1921).

The models to be discussed will apply to diploid populations. We will assume that populations are large enough so that fluctuations due to sampling can be ignored, and that there are no viability or fertility differences. We will be interested in the overall dynamics of a population as well as in its equilibrium state, and in the level of heterozygosity that can be maintained under such systems of mating. The effect of these mating systems on the variability in a population will also be discussed.
II. MIXED RANDOM MATING AND SELFING

A. Distributions of Genotypes

We suppose that a fraction \(0 < \beta \leq 1\) of the population reproduces by selfing while the rest of the population mates at random. The proportion \(\beta\) is constant for all generations. \(\beta = 0\) corresponds to the case of pure random mating and \(\beta = 1\) implies pure self-fertilization. We are primarily interested in the results for \(\beta \in (0,1)\), where \(\beta\) is the parameter of partial inbreeding. The population consists of diploid individuals and is considered to be large enough so that fluctuations due to sampling can be ignored. We suppose that genotypes are all equally viable and all crosses equally fertile. The generations are taken to be nonoverlapping, that is, we consider discrete models. The generations of mixed matings are indexed by the nonnegative integer \(n\). The index 0 corresponds to the initial population before the onset of mixed mating system. In the analysis that follows it is assumed that individuals practicing selfing in each generation are a random sample of the population, that is, there is no tendency for the progeny of the self-fertilized individuals to be self-fertilized. We shall examine the simplest situation first and then proceed to the complex ones, rather than to deal with the matter in full generality from the start. We shall start with a population for which only one locus is variable and then
extend the results to multi-locus situations.

1. **Single locus**

We assume that one locus influences a character. The general case of \( 0 \leq \beta \leq 1 \) was first treated for a single locus by Haldane (1924). Ghai (1964) explored this model in more detail and discussed the effect of such a system of mating on the statistical properties of a population. All these studies related to a diploid single locus with two alleles. More recently Bennett (1968) has considered the effect of mixed self- and cross-fertilization in a tetrasomic species. We consider here a diploid single locus but with multiple alleles, say \( A_1, A_2, \ldots, A_s \).

Let \( \Omega = \{A_i A_j\} \) \( i,j = 1,2, \ldots, s \), be the sample space of genotypes associated with this locus, with the corresponding probability distribution at any time, \( \{P_{ij}\} \) \( i,j = 1,2, \ldots, s \), of genotypes with \( P_{ij} = P_{ji} \). That is,

\[
P[A_1 A_i] = P_{ii} \quad i = 1,2, \ldots, s
\]

\[
P[A_1 A_j] = 2P_{ij} \quad i \neq j = 1,2, \ldots, s
\]

such that

\[
\sum_{i=1}^{s} \sum_{j=1}^{s} P_{ij} = 1
\]

The frequencies
\[ P[A_i] = p_i = \sum_{j=1}^{s} p_{ij} \]

where \( i = 1, 2, \ldots, s \), with \( \sum p_i = 1 \) represent the gene distribution of the population. The corresponding probabilities in the next generation under a given system of mating are denoted by \( \{p'_{ij}\} \) and \( \{p'_i\} \), \( i, j = 1, 2, \ldots, s \).

Let

\[ R(P) = R' = \text{Population obtained by random mating population, } P(= \sum \sum p_{ij} A_i A_j), \text{ and} \]

\[ S(P) = S' = \text{Population obtained by selfing population, } P, \]

that is,

\[ R(P) = \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij} A_i A_j \quad (2.1) \]

\[ S(P) = \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij} (1/4)[A_i A_j + 2A_i A_j + A_i A_j] \]

\[ = (1/2)[\sum_{i=1}^{s} p_{i} A_i A_i + \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij} A_i A_j] \quad (2.2) \]

Then the genotypic array in the resulting population under the system of mixed random mating and selfing is

\[ P' = \gamma R(P) + \beta S(P) \]

\[ = \gamma \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij} A_i A_j + (\beta/2)[\sum_{i=1}^{s} p_{i} A_i A_i + \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij} A_i A_j] \]
where \( \gamma = 1 - \beta \). This gives

\[
P'_{ij} = \gamma p_i p_j + (\beta/2) p_{ij} \quad i \neq j = 1, 2, \ldots, s \quad (2.3)
\]

\[
P'_{ii} = \gamma p_i^2 + (\beta/2) p_i + (\beta/2) p_{ii} \quad i = 1, 2, \ldots, s \quad . \quad (2.4)
\]

Under the assumption that individuals practicing selfing in each generation are a random sample, the frequencies \( \{P^{(n)}_{ij}\} \) in the \( n \)th generation can be expressed in terms of those in the \( (n-1) \)th generation, and hence in terms of those in the initial population using the above transformation. It is easily seen that

\[
P'_{i} = \sum_{j=1}^{s} P'_{ij} = p_i \quad , \quad i = 1, 2, \ldots, s \quad ,
\]

so that gene frequencies remain constant. The complete time dependent behavior of the population can then be easily worked out by the set of Equations 2.3 and 2.4.

From Equation 2.3 we may write

\[
P^{(n)}_{ij} = \gamma p_i p_j + (\beta/2) P^{(n-1)}_{ij} \quad i \neq j = 1, 2, \ldots, s \quad .
\]

This linear difference equation can be easily solved yielding

\[
P^{(n)}_{ij} = \gamma p_i p_j [1 + (\beta/2)] + (\beta/2)^2 P^{(n-2)}_{ij}
\]

\[= \gamma p_i p_j [1 + (\beta/2) + \ldots + (\beta/2)^{n-1}] + (\beta/2)^n P^{(0)}_{ij}
\]

\[= \gamma p_i p_j \frac{1 - (\beta/2)^n}{1 - (\beta/2)} + (\beta/2)^n P^{(0)}_{ij} \quad . \quad (2.5)
\]
Similarly, we obtain

\[ p_{ii}^{(n)} = \gamma p_{i}^2 + (\beta/2) p_{i} + (\beta/2) p_{ii}^{(n-1)} \]

\[ = [\gamma p_{i}^2 + (\beta/2) p_{i}] \frac{[1 - (\beta/2)^n]}{1 - (\beta/2)} + (\beta/2)^n p_{ii}^{(0)} \] (2.6)

where \( i=1,2,\ldots,s \).

As \( n \to \infty \), the population reaches equilibrium with frequencies \( \{p_{ij}^{(\infty)}\}_{i,j=1}^{s} \), where

\[ p_{ij}^{(\infty)} = \lim_{n \to \infty} p_{ij}^{(n)} = \gamma p_{i} p_{j}[1 - (\beta/2)]^{-1} \quad i\neq j \] (2.7)

\[ p_{ii}^{(\infty)} = \lim_{n \to \infty} p_{ii}^{(n)} = [\gamma p_{i}^2 + (\beta/2) p_{i}][1 - (\beta/2)]^{-1} \] (2.8)

with \( i,j=1,2,\ldots,s \). Notice that when \( \beta = 1 \) we get the usual results for self-fertilization.

\( p_{ij}^{(n)} \) converges to its equilibrium value \( p_{ij}^{(\infty)} \) at the geometric rate \((\beta/2)\). It is needless to say that the system of Equations 2.5 is enough to specify the composition of the population in generation \( n \) because the gene frequencies remain invariant over time. It is interesting to observe that the distribution of genotypes in general depends upon the initial distribution but in the limiting case as \( n \to \infty \) it is independent of the initial distribution of genotypes and depends only on the initial gene frequencies and the probability of selfing.
The above results may also be derived from consideration of the probability of two genes being identical by descent (Kempthorne, 1957, p. 86). However, this method is useful in determining the distribution of genotypes only if the initial population is in Hardy-Weinberg equilibrium, because in that case (Kempthorne, 1957, p. 80) it can be expressed as

\[ F \sum p_i A_i A_i + (1-F) \sum \prod p_i p_j A_i A_j \]

Equations 2.5 and 2.6 show that if we start with an arbitrary initial population, then the resulting population under the system of mixed random mating and selfing cannot be expressed in this form. This is true even in the case of pure selfing (\( \beta = 1 \)). The point was made by Kempthorne (1957, p. 80) and we quote: "It would be erroneous to suppose that a natural population would have the genotypic array

\[ F \sum p_i A_i A_i + (1-F)(\sum p_i A_i)^2 \]

From the preceding derivations it is clear that this would happen only if the population was derived from an original population which itself arose by random mating, ...

In general, there is an additional term which depends upon the deviation of the original population from panmixia. Similar results have also been reported by Jacquard (1968) for inbreeding due to finite population size.

We have obtained the solution in a straightforward manner because the situation here is simple and further, with this approach, it is possible to give mathematical
analysis for an arbitrary initial population, and also to extend the results to populations where more than one locus is involved. We shall, however, make use of the method of "identity by descent" later in analyzing several other models describing the effect of partial inbreeding.

We can examine this model in another way as follows. Let the probability generating function of alleles $A_1, A_2, \ldots, A_s$ be defined as

$$f(t) = \sum_{i=1}^{s} p_i t^i$$  \hspace{1cm} (2.9)

where the coefficient of $t^i = p[A_i] = p_i$. Similarly, let $f(t_1, t_2)$ be the probability generating function (p.g.f.) of the ordered genotypes $\{A_i A_j\}_{i,j=1}^{s}$ where the coefficient of $t_1^i t_2^j$ is the probability of the ordered genotype $A_i A_j$, that is,

$$f(t_1, t_2) = \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij} t_1^i t_2^j$$  \hspace{1cm} (2.10)

We know that under random mating in one generation,

$$f'(t_1, t_2) = f(t_1)f(t_2)$$  \hspace{1cm} (2.11)

and under the process of selfing

$$f'(t_1, t_2) = (1/2)[f(t_1 t_2) + f(t_1, t_2)]$$  \hspace{1cm} (2.12)

These can be easily derived by considering Equations 2.1 and 2.2. Therefore, the p.g.f. in one generation of mixed random
mating and selfing is given by

\[ f'(t_1, t_2) = \gamma f(t_1) f(t_2) + (\beta/2) f(t_1 t_2) + (\beta/2) f(t_1, t_2) \]

\[ = h(t_1, t_2, \beta) + (\beta/2) f(t_1, t_2) \quad (2.13) \]

where \( h(t_1, t_2, \beta) = \gamma f(t_1) f(t_2) + (\beta/2) f(t_1 t_2) \). \( f(t) \) is a function of the gene frequencies \( \{p_i\} \) which do not change from generation to generation, therefore \( h \) will remain invariant over time. Hence, we can rewrite Equation 2.13 as

\[ f^{(n)}(t_1, t_2) = h(t_1, t_2, \beta) + (\beta/2) f^{(n-1)}(t_1, t_2) \]

\[ = h(t_1, t_2, \beta) [1 + (\beta/2)] + (\beta/2)^2 f^{(n-2)}(t_1, t_2) \]

\[ = h(t_1, t_2, \beta) [1 + (\beta/2) + \ldots + (\beta/2)^{n-1}] \]

\[ + (\beta/2)^n f^{(0)}(t_1, t_2) \]

\[ = h(t_1, t_2, \beta) \left[ \frac{1 - (\beta/2)^n}{1 - \beta/2} \right] + (\beta/2)^n f^{(0)}(t_1, t_2). \]

\[ (2.14) \]

Such a representation will be useful if the allelic and genotype frequencies have some sort of structure.

If we start with a population which was obtained by random mating i.e., \( p_{ij}^{(0)} = p_i p_j \), then the solution of the set of Equations 2.3 and 2.4 reduces to

\[ p_{ij}^{(n)} = \left\{ (\beta/2)^n + \gamma \left[ \frac{1 - (\beta/2)^n}{1 - \beta/2} \right] \right\} p_i p_j \quad (2.15) \]

\[ \text{if } i \neq j \]
with \( i, j = 1, 2, \ldots, s \). In this case the population has what may be termed generalized Hardy-Weinberg structure.

In the limiting case as \( n \to \infty \), the distribution is the same as given by Equations 2.7 and 2.8 because as we have observed before the distribution of genotypes in equilibrium is independent of the initial distribution and is completely specified by the proportion of selfing \( \beta \), and the gene frequencies \( \{p_i\} \).

2. Two loci

Let \( A_1, A_2, \ldots, A_{s_1} \) be the \( s_1 \) alleles at locus A and \( B_1, B_2, \ldots, B_{s_2} \) be the \( s_2 \) alleles at locus B. The two loci are assumed to be segregating independently. We will examine the consequences of the system of mixed mating on the joint probability distribution \( \{P_{ijk\ell}\} \) \( i, j = 1, 2, \ldots, s_1; k, \ell = 1, 2, \ldots, s_2 \), with respect to two unlinked loci defined on the sample space \( \Omega = \Omega_1 \times \Omega_2 \) where \( \Omega_1 = \{A_iA_j\} \) \( i, j = 1, 2, \ldots, s_1 \) and \( \Omega_2 = \{B_kB_\ell\} \) \( k, \ell = 1, 2, \ldots, s_2 \). Thus

\[
P_{ijk\ell} = P[A_iA_jB_kB_\ell, \text{ordered at each locus}]
\]

\[
= P_{lijk} = P_{ij\ell k} = P_{jilk}
\]  

\[\text{(2.17)}\]
such that $i, j = 1, 2, \ldots, s_1$; $k, \ell = 1, 2, \ldots, s_2$ and

$$
\sum_{i=1}^{s_1} \sum_{j=1}^{s_2} \sum_{k=1}^{s_1} \sum_{\ell=1}^{s_2} p_{ijk\ell} = 1 .
$$

Bennett and Binet (1956) considered the model for two linked loci with two alleles at each locus and discussed the genetic equilibrium. Ghai (1964) developed this model, with two alleles at each locus, in great detail and examined the complete time dependent behavior of the population and discussed several other characteristics of the population.

The analysis is carried out with the same assumptions as in a single locus model. Let

$$
g = \sum_{i=1}^{s_1} \sum_{k=1}^{s_2} g_{ik} A_i B_k
$$

with

$$
\sum_{i=1}^{s_1} \sum_{k=1}^{s_2} g_{ik} = 1
$$

be the gametic array of a population $P$. The gametic frequencies satisfy the following relations:

$$
g_{ik} = p[A_i B_k] = \sum_{j=1}^{s_1} \sum_{\ell=1}^{s_2} p_{ijk\ell} = p_{i.k} .
$$

(2.18)

$$
p_i = p[A_i] = \sum_{k=1}^{s_2} g_{ik} = p_{i.} = p_{i.} .
$$

(2.19)
\[ \mathbf{q}_k = p[B_k] = \sum_{i=1}^{s_1} g_{ik} = p_{..k} = p_{..k} \] 

\( i=1,2,\ldots,s_1; \ k=1,2,\ldots,s_2. \)

It is well known that under random mating in the population,

\[ R(P) = \left( \sum_{i=1}^{s_1} \sum_{k=1}^{s_2} g_{ik} A_i B_k \right)^2 \]

\[ = (1/2) \sum_i \sum_j \sum_k \sum_\ell \left( g_{ik} g_{j\ell} + g_{i\ell} g_{jk} \right) A_i A_j B_\ell B_k \]

because \( p'_{ijk\ell} = 2(g_{ik} g_{j\ell} + g_{i\ell} g_{jk}), \ i\neq j, \ k\neq \ell \) and so on.

Similarly, with respect to the process of selfing

\[ S(P) = \sum_{i=1}^{s_1} \sum_{j=1}^{s_1} \sum_{k=1}^{s_2} \sum_{\ell=1}^{s_2} p_{ijk\ell} \left[ \left( \frac{1}{4} A_i A_1 + \frac{1}{2} A_i A_j + \frac{1}{4} A_i A_j \right) \times \right. \]

\[ \left. \left( \frac{1}{4} B_k B_\ell + \frac{1}{2} B_k B_\ell + \frac{1}{4} B_\ell B_\ell \right) \right] \]

\[ = \frac{1}{4} \sum_{i=1}^{s_1} \sum_{k=1}^{s_1} p_{i.k} A_i A_1 B_1 B_k \]

\[ + \sum_{i=1}^{s_1} \sum_{j=1}^{s_1} \sum_{k=1}^{s_2} \sum_{\ell=1}^{s_2} p_{ijk\ell} A_i A_j B_k B_\ell \]

\[ + \sum_{i=1}^{s_1} \sum_{k=1}^{s_1} \sum_{\ell=1}^{s_1} p_{i.k\ell} A_i A_1 B_1 B_\ell \]
We can obtain the resulting population after one generation of mixed random mating and selfing by weighting $R(P)$ and $S(P)$ by their corresponding probabilities $\gamma$ and $\beta$ respectively, that is,

$$P' = \gamma R(P) + \beta S(P) \quad \text{(2.21)}$$

In the case of populations where more than one locus is involved, there are three distributions which are relevant; namely, the distribution of genotypes, the distribution of gametes, and the distribution of genes at each locus. The structure of population after one generation of mixed mating is given by Equation 2.21. We shall come back to this a little later. We shall first discuss the behavior of the gametic and gene frequencies in successive generations.

Under random mating in the population

$$g'_{ik} = \frac{1}{2} \sum_j \sum_{\ell} \left( g_{ik} g_{j\ell} + g_{i\ell} g_{jk} \right)$$

$$= \frac{1}{2} (g_{ik} + p_i g_k) \quad \text{(2.22)}$$

and with selfing

$$g'_{ik} = p_{i.k} = g_{ik} \quad \text{(2.23)}$$

where $i, j = 1, 2, \ldots, s_1$; $k, \ell = 1, 2, \ldots, s_2$. Multiplying Equation
2.22 by the factor $\gamma(=1-\beta)$ and Equation 2.23 by the factor $\beta$ which are the fractions of the population practicing random mating and selfing respectively, we obtain the recursion formula for the gametic frequency

$$g'_{ik} = (\gamma/2)(g_{ik} + p_i q_k) + \beta g_{ik}$$

$$= g_{ik} - (\gamma/2)(g_{ik} - p_i q_k)$$

(2.24)

where $i=1,2,\ldots,s_1; k=1,2,\ldots,s_2$.

It is easy to see that

$$\sum_{k=1}^{s_2} g'_{ik} = p_i$$

$$\sum_{i=1}^{s_1} g'_{ik} = q_k$$

so that gene frequencies remain the same. To discuss further the gametic frequencies, we can rewrite Equation 2.24 as

$$g^{(n)}_{ik} = g^{(n-1)}_{ik} - (\gamma/2)[g^{(n-1)}_{ik} - p_i q_k]$$

because the gene frequencies are invariant, or after subtracting $p_i q_k$ from both sides

$$g^{(n)}_{ik} - p_i q_k = [1 - (\gamma/2)][g^{(n-1)}_{ik} - p_i q_k]$$

$$= [1 - (\gamma/2)]^2[g^{(n-2)}_{ik} - p_i q_k]$$
where \( g_{ik}^{(0)} \), \( i=1,2,...,s_1; k=1,2,...,s_2 \) are the gametic frequencies in the initial population, and \( \gamma = 1 - \beta \) is the probability of random mating.

In the limit as \( n \to \infty \)

\[
g_{ik}^{(\infty)} = \lim_{n \to \infty} g_{ik}^{(n)} = p_i q_k
\]  

(2.26)

where \( i=1,2,...,s_1 \) and \( k=1,2,...,s_2 \). Thus in the limiting case the gametic frequencies are equal to the product of the corresponding gene frequencies which is also true in the case of pure random mating. However, the equilibrium value is approached only gradually at the rate of \([1 - (\gamma/2)] \) \([= (1+\beta)/2]\). Notice further from Equation 2.25 that results regarding the gametic frequencies in the case of two unlinked loci under a system of mixed random mating and selfing are the same as under pure random mating with two linked loci with recombination fraction equal to \( \gamma/2 \).

Now we come back to examine the time dependent behavior of the frequencies of different genotypes. We have the recursion formula 2.21 for the genotypic array. The contribution in this due to the random mating depends upon the gametic frequencies which, in general, vary with time. Such a general situation has been discussed by Ghai (1964) for \( s_1 = s_2 = 2 \) for an arbitrary initial population. He has
observed that when $n$ is sufficiently large, the genotypic frequencies depend, as far as initial conditions are concerned, only on the gene frequencies. If we start with an initial random mating population in equilibrium so that the gametic frequencies are products of the corresponding frequencies of genes, i.e. $q_{ik}^{(0)} = p_i q_k$, then as we can see from Equation 2.24 the gametic frequencies do not change from generation to generation. This is also true in the case of linked genes. This will simplify the situation a great deal. This implies that the contribution due to random mating in Equation 2.21 will be the same in all generations. Thus, in this case we have after $n$ generations of mixed random mating and selfing

\[ p(n) = \gamma p(0) + (\beta/4) \left[ \sum_{i=1}^{s_1} \sum_{k=1}^{s_2} p_{i,k}^{(n-1)} A_i A_k B_i B_k \right. \]

\[ + \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} \sum_{k=1}^{s_2} p_{i,j,k}^{(n-1)} A_{ij} B_i B_k B_k \]

\[ + \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} \sum_{k=1}^{s_1} \sum_{\ell=1}^{s_2} p_{i,j,k,\ell}^{(n-1)} A_{ij} B_i B_k B_\ell \]

\[ \left. \right] \] (2.27)

where $p_{ij,k}^{(0)} = p_i p_j q_k q_\ell$, $i,j=1,2,...,s_1$; $k,\ell=1,2,...,s_2$. 
This results in the following set of recursion formulas:

\[ p_{ijkl}^{(n)} = \gamma p_{ijkl}^{(0)} + (\beta/4) p_{ijkl}^{(n-1)} \quad i \neq j, \ k \neq \ell \]

\[ p_{ijkl}^{(n)} = \gamma p_{ijkl}^{(0)} + (\beta/4) [p_{ijkl}^{(n-1)} + p_{ijlk}^{(n-1)}] \quad i \neq j \]

\[ p_{ikl\ell}^{(n)} = \gamma p_{ikl\ell}^{(0)} + (\beta/4) [p_{ikl\ell}^{(n-1)} + p_{i.k\ell}^{(n-1)}] \quad k \neq \ell \]

\[ p_{iikk}^{(n)} = \gamma p_{iikk}^{(0)} + (\beta/4) [p_{iikk}^{(n-1)} + (p_{iik}^{(n-1)} + p_{ikk}^{(n-1)}) + p_{i.kk}^{(n-1)}] \]

\[ + p_{i.k}^{(n-1)} \]  \hspace{1cm} (2.28)

where \( i, j = 1, 2, \ldots, s_1; \ k, \ell = 1, 2, \ldots, s_2 \).

The situation is quite complex if the genes are linked because the recursion relations are nonlinear. However, if we have an initial population which is in linkage equilibrium under random mating, the mathematics is simplified because the gametic frequencies remain invariant over time and consequently the contribution due to random mating remains the same in successive generations. Hence, in this case it is possible to write difference equations similar to Equations 2.28 which then can be solved in a similar manner.

To solve the system of difference Equations 2.28 we note that with two loci, basically there are three different types of equations corresponding to the three kinds of genotypes namely, genotypes heterozygous at both loci, genotypes heterozygous at one locus, and genotypes heterozygous at
neither locus (completely homozygous).

If the population is segregating with respect to k loci, we can group the genotypes into (k+1) different types. For example, genotypes homozygous for all genes, genotypes heterozygous with respect to one locus, and so on. In order to solve the above set of difference equations and the equations that may arise later with more than two loci, it is convenient to introduce the following notation.

Let \( \phi_r, r=0,1,2,...,k \), where k is the number of loci, be defined as a P value in which r of the factors are in heterozygous state and the remaining (k-r) factors are in homozygous condition. That is, for a given r, \( 2^r \phi_r \) is the probability of a particular genotype heterozygous with respect to r of the k loci. Note that \( \phi_r \) depends upon the particular genotype under consideration, but as we will see the specification of such a genotype is not crucial for the solution of the recursion equations. Further, let \( \phi_{rt} \) (t=0,1,2,...,k-r) denote the sum of those P's in which t subscripts, choosing one from each of the possible (k-r) homozygous genes in all possible combinations, are replaced by a dot (.). This notation is explained for the two loci case as follows. Consider the set of Equations 2.28.

Notice that subscripts involved on the right hand side of each equation are the same as on the left hand side except
that some of the subscripts may have been replaced by a dot (.) This helps us express these equations in terms of the $\phi$'s without ambiguity and as indicated before there will be only three different types of equations for $k=2$.

These are expressed as

$$\phi_2^{(n)} = \gamma \phi_2^{(0)} + (\beta/4) \phi_{20}^{(n-1)}$$

$$\phi_1^{(n)} = \gamma \phi_1^{(0)} + (\beta/4) [\phi_{10}^{(n-1)} + \phi_{11}^{(n-1)}]$$

$$\phi_0^{(n)} = \gamma \phi_0^{(0)} + (\beta/4) [\phi_{00}^{(n-1)} + \phi_{01}^{(n-1)} + \phi_{02}^{(n-1)}]$$

Note that $\phi_{r0} = \phi_r$ for all $r$. We can derive the frequencies of all genotypes from these equations. Suppose that we are interested in the frequency of a homozygous genotype, say $A_1A_1B_2B_2$, i.e., $r=0, k-r=2$, then we will take

$$\phi_0 = p_{1122}$$

$$\phi_{00} = p_{1122} = \phi_0$$

$$\phi_{01} = p_{112} + p_{122}$$

$$\phi_{02} = p_{12}.$$

which are to be substituted in the above equations. Similarly, if we are interested in a single heterozygote, say $A_1A_1B_1B_2$, then $r=1, k-r=1$, and
\[ \phi_1 = p_{1112} \]
\[ \phi_{10} = p_{1112} = \phi_1 \]
\[ \phi_{11} = p_{1.12} \]

We also introduce some new quantities which will arise frequently in our solutions of the systems of the recursion formulas.

Let
\[ c_i = \beta/2^i \]

\[ \omega_{ij}(n) = \sum_{r=0}^{n-1} c_i^r c_{n-r-1} = \frac{c_i^n - c_j^n}{c_i - c_j} \quad i \neq j \]

\[ \omega_i(n) = \sum_{r=0}^{n-1} c_i^r = \frac{1 - c_i^n}{1 - c_i} \]

\[ a_i(n) = c_i^n + \gamma \omega_i(n) , \quad \gamma = 1 - \beta \]

\[ d_{ij}(n) = \sum_{r=0}^{n-1} c_i^r \omega_j^{(n-r-1)} \]

\[ b_{ij}(n) = \sum_{r=0}^{n-1} c_i^r a_j^{(n-r-1)} \quad . \quad (2.29) \]

Notice that \( 2^k c_{i+k} = c_i \).

Some of the above formulas can be further simplified and put in more convenient form.
\[ d_{ij}(n) = \sum_{r=0}^{n-1} c_i^r w^{(n-r-1)} \]

\[ = \sum_{r=0}^{n-1} c_i^r \left[ \frac{1 - c_i^{n-r-1}}{1 - c_j} \right] \]

\[ = \frac{1}{1 - c_j} \left[ \frac{1 - c_i^n}{1 - c_i} - \frac{c_i^n - c_j^n}{c_i - c_j} \right] \]

\[ = \frac{1}{c_i - c_j} \left[ \frac{1 - c_i^n}{1 - c_i} - \frac{1 - c_j^n}{1 - c_j} \right] \]

\[ = \frac{w_i^{(n)} - w_j^{(n)}}{c_i - c_j} \]  \hspace{1cm} (2.30)

\[ b_{ij}(n) = \sum_{r=0}^{n-1} c_i^r a_j^{(n-r-1)} \]

\[ = \sum_{r=0}^{n-1} c_i^r \left[ c_j^{n-r-1} + \gamma w_j^{(n-r-1)} \right] \]

\[ = w_{ij}(n) + \gamma d_{ij}(n) \]  \hspace{1cm} (2.31)

\[ w_{ij}(n) + \gamma d_{ij}(n) = \frac{c_i^n - c_j^n}{c_i - c_j} + \gamma \frac{w_i^{(n)} - w_j^{(n)}}{c_i - c_j} \]

\[ = \frac{1}{c_i - c_j} \left[ a_i^{(n)} - a_j^{(n)} \right] \]

Therefore,
\[ b_{ij} = w_{ij} + \gamma d_{ij} \]

\[
\frac{a_i(n) - a_j(n)}{c_i - c_j} = (2.32)
\]

Using the new notation we can solve the first equation of the set of Equations 2.28 as

\[
P_{ijkl}^{(n)} = \gamma P_{ijkl}^{(0)} + c_2 P_{ijkl}^{(n-1)} \quad i \neq j, k \neq \ell
\]

\[
= \gamma P_{ijkl}^{(0)} \left[ 1 + c_2 + c_2^2 + \ldots + c_2^{n-1} \right] + c_2 P_{ijkl}^{(0)}
\]

\[
= [c_2^n + \gamma w_2^{(n)}] P_{ijkl}^{(0)}
\]

\[
= a_2^{(n)} P_{ijkl}^{(0)}
\]

or in terms of \( \varphi \)'s

\[
\varphi_2^{(n)} = a_2^{(n)} \varphi_2^{(0)}
\]

(2.34)

To solve the other equations of the system of Equations 2.28 we need further recursion formulas for \( P_{ijk} \), \( P_{i.k\ell} \) and \( P_{i.k} \). Using the set of Equations 2.28 we can easily obtain

\[
P_{ijk}^{(n)} = \gamma P_{ijk}^{(0)} + (\beta/2) P_{ijk}^{(n-1)} \quad i \neq j
\]

\[
= \gamma P_{ijk}^{(0)} + c_1 P_{ijk}^{(n-1)}
\]

(2.35)

which is of the same form as Equation 2.33. Therefore,
\[ p_{ijk}^{(n)} = \left[ c_1^{n} + \gamma w_1 \right] p_{ijk} \quad i \neq j \]

\[ = a_{1}^{(n)} p_{ijk} \quad (2.36) \]

and hence, for \( i \neq j \),

\[ p_{ijkk}^{(n)} = \gamma p_{ijkk}^{(0)} + c_2 p_{ijkk}^{(n-1)} + c_2 a_{1}^{(n-1)} p_{ijk}^{(0)} \]

\[ = \gamma p_{ijkk}^{(0)} \left[ 1 + c_2 + c_2^2 + \ldots + c_2^{n-1} \right] \]

\[ + c_2 \left[ a_{1}^{(n-1)} + c_2 a_{1}^{(n-2)} + \ldots + c_2^{n-1} a_{1}^{(0)} \right] p_{ijk}^{(0)} \]

\[ + c_2^{n} p_{ijkk}^{(0)} \]

\[ = \gamma w_2 p_{ijkk}^{(0)} + c_2 b_{1}^{(n)} p_{ijk}^{(0)} + c_2^{n} p_{ijkk}^{(0)} \]

\[ = a_{2}^{(n)} p_{ijkk}^{(0)} + c_2 \left[ \frac{a_{1}^{(n)} - a_{2}^{(n)}}{c_1 - c_2} \right] p_{ijk}^{(0)} \]

or

\[ p_{ijkk}^{(n)} = a_{2}^{(n)} p_{ijkk}^{(0)} + [-a_{2}^{(n)} + a_{1}^{(n)}] p_{ijk}^{(0)} \quad (2.37) \]

for \( i \neq j = 1, 2, \ldots, s_1; k = 1, 2, \ldots, s_2 \).

In terms of \( \varphi \)'s this is expressible as

\[ \varphi_{1}^{(n)} = \varphi_{10}^{(n)} = a_{2}^{(n)} \varphi_{10}^{(0)} + [-a_{2}^{(n)} + a_{1}^{(n)}] \varphi_{11}^{(0)} \quad . \quad (2.38) \]

Rewriting the recursion formulas for a homozygous genotype, we get
\[ p^{(n)}_{iik} = \phi^{(n)}_0 = \gamma \phi^{(0)}_0 + c_2 [\phi^{(n-1)}_0 + \phi^{(n-1)}_{01} + \phi^{(n-1)}_{02}] \]

\[ p^{(n)}_{i.k.} + p^{(n)}_{i.kk} = \phi^{(n)}_{01} = \gamma \phi^{(0)}_{01} + c_1 [\phi^{(n-1)}_{01} + 2\phi^{(n-1)}_{02}] \]

\[ p^{(n)}_{i.k.} = \phi^{(n)}_{02} = \gamma \phi^{(0)}_{02} + c_0 \phi^{(n-1)}_{02} \]

Since \( \phi_0 = \phi_{00} \) we can express this system in the matrix form

\[
\begin{bmatrix}
\phi^{(n)}_{02} \\
\phi^{(n)}_{01} \\
\phi^{(n)}_{00}
\end{bmatrix} =
\begin{bmatrix}
c_0 & 0 & 0 \\
2c_1 & c_1 & 0 \\
c_2 & c_2 & c_2
\end{bmatrix}
\begin{bmatrix}
\phi^{(n-1)}_{02} \\
\phi^{(n-1)}_{01} \\
\phi^{(n-1)}_{00}
\end{bmatrix} +
\gamma
\begin{bmatrix}
\phi^{(0)}_{02} \\
\phi^{(0)}_{01} \\
\phi^{(0)}_{00}
\end{bmatrix} .
\]

We can find the solution by considering the eigenvalues and eigenvectors of the matrix, and the initial values. However, this system of equations can be solved directly without much difficulty. The first two equations are of the same form as equations for \( \phi_{2}^{(n)} \) and \( \phi_{1}^{(n)} \). Hence,

\[ \phi^{(n)}_{02} = a_0 \phi^{(0)}_{02} \]

\[ \phi^{(n)}_{01} = a_1 \phi^{(0)}_{01} + 2c_1 b_1 \phi^{(0)}_{02} \]

\[ = a_1 \phi^{(0)}_{01} + 2[-a_1 + a_0] \phi^{(0)}_{02} \]

The solution for \( \phi_{0}^{(n)} \) is then obtained as
The complete solution of the system of Equations 2.28 is given by Equations 2.34, 2.38, and 2.39, which can be expressed as

\[ \phi_i^{(n)} = \sum_{j=0}^{2-i} \phi_{ij}^{(0)} \sum_{x=0}^{j} (-1)^{j-x} a_{2-x}^{(n)} \quad i=0,1,2 \]

(2.40)

In the limit as \( n \to \infty \), \( a_i^{(n)} \to \frac{\gamma}{1-c_1} \). Therefore,

\[ \phi_i^{(\infty)} = \lim_{n \to \infty} \phi_i^{(n)} = \sum_{j=0}^{2-i} \phi_{ij}^{(0)} \sum_{x=0}^{j} (-1)^{j-x} \gamma (1-c_2-\gamma)^{-1} \]

(2.41)

That is, at equilibrium

\[ p_{ijkl}^{(\infty)} = \frac{\gamma}{1-c_2} p_{ijkl}^{(0)} = \frac{\gamma}{1-c_2} p_{i}^{(0)} p_{j^k}^{q_k} q_\ell \quad i \neq j, \; k \neq \ell \]

(2.42)
\[ P_{ijkk}^{(\infty)} = \frac{\gamma}{1-c_2} P_{ijkk}^{(0)} + \frac{\gamma}{1-c_1 - 1-c_2} P_{ijk}^{(0)} \quad \text{i<j} \]

\[ = \frac{\gamma}{1-c_2} P_{i} P_{j} q_{k}^2 + \frac{\gamma c_2}{(1-c_1)(1-c_2)} P_{i} P_{j} q_{k} \tag{2.43} \]

\[ P_{iikk}^{(\infty)} = \frac{\gamma}{1-c_2} P_{iikk}^{(0)} + \frac{\gamma c_2}{(1-c_1)(1-c_2)}[P_{iik}^{(0)} + P_{i.kk}^{(0)}] \]

\[ + \left( \frac{\gamma}{1-c_2} - \frac{2\gamma}{1-c_1} + \frac{\gamma}{1-c_0} \right) P_{i.k}^{(0)} \]

\[ = \frac{\gamma}{1-c_2} p_{i} q_{k}^2 + \frac{\gamma c_2}{(1-c_1)(1-c_2)}(p_{i} q_{k}^2 + p_{i} q_{k}^2) \]

\[ + \frac{\gamma c_2 (1+c_1)}{(1-c_0)(1-c_1)(1-c_2)} p_{i} q_{k}. \tag{2.44} \]

Comparing Equations 2.33 and 2.42 we note that the deviation of the frequency of a double heterozygote from its limiting value diminishes at the rate of \( (\bar{p}/4) \).

Another comparison that may be of interest is the comparison of the joint distribution versus the product of the two marginal distributions. We shall now show that the joint distribution with respect to two loci is not equal to the product of the two marginals even in the limiting case although the two loci are taken to be unlinked. In the discussion that follows we restrict ourselves to an initial panmictic population, i.e.,
\[ p_{ijkl}^{(0)} = p_i p_j q_k q_\ell, \quad i, j = 1, 2, \ldots, s_1; \ k, \ell = 1, 2, \ldots, s_2. \]

We know that

\[ p_{ijkl}^{(n)} = a_2^{(n)} p_{ijkl}^{(0)} \quad i \neq j, \ k \neq \ell \]

\[ p_{ijkl}^{(n)} = a_2^{(n)} p_{ijkl}^{(0)} + [a_1^{(n)} - a_2^{(n)}] p_{ijk}^{(0)} \quad i \neq j \]

\[ p_{iikk}^{(n)} = a_2^{(n)} p_{iikk}^{(0)} + [a_1^{(n)} - a_2^{(n)}][p_{iik}^{(0)} + p_{i.kk}^{(0)}] \]

\[ + [a_0^{(n)} - 2a_1^{(n)} + a_2^{(n)}] p_{i.k}^{(0)} \]

and for the marginal distribution at the A locus

\[ p_{ij..}^{(n)} = a_1^{(n)} p_{ij..}^{(0)} \quad i \neq j \]

\[ p_{ii..}^{(n)} = a_1^{(n)} p_{ii..}^{(0)} + [a_0^{(n)} - a_1^{(n)}] p_{i...}^{(0)} \]

with similar expressions for the marginal distribution at the B locus. Using these relations we see that

\[ p_{ijkl}^{(n)} - p_{ij..}^{(n)} p_{kk..}^{(n)} = \{a_2^{(n)} - [a_1^{(n)}]^2\} p_{ijkl}^{(0)} \quad i \neq j, \ k \neq \ell \]

\[ = \eta_n p_i p_j p_k p_\ell \]

(2.45)

where

\[ \eta_n = a_2^{(n)} - [a_1^{(n)}]^2 \]

Similarly,

\[ p_{iikk}^{(n)} - p_{ij..}^{(n)} p_{kk..}^{(n)} = -\eta_n p_i p_j q_k^2 (1 - q_\ell) \quad i \neq j \]

(2.46)
\[ p_{iikk} - p_{ii..kk} = \eta_n \ p_i (1 - p_i) q_k (1 - q_k) \quad \text{(2.47)} \]

In the limit as \( n \to \infty \), \( \eta_n \to \eta_\infty \) where

\[ \eta_\infty = a_2^{(\infty)} - [a_1^{(\infty)}]^2 \]

\[ = \frac{1 - \beta}{1 - c_2} - \left[ \frac{1 - \beta}{1 - c_1} \right]^2 \quad \text{(2.48)} \]

> 0 for \( 0 < \beta < 1 \)

We can see from Equations 2.45 - 2.47 that the relevant comparisons depend upon \( \eta_n \) apart from the gene frequencies. It will, therefore, be of interest to examine the behavior of \( \eta_n \) as a function of \( n \) and \( \beta \). \( \eta_0 \) is zero trivially for all \( \beta \). We will show that in general \( \eta_n \geq 0 \) for \( n \geq 1 \) for all \( \beta \), the equality holding only if \( \beta = 0 \) or 1.

When \( \beta = 0 \), \( a_i^{(n)} = 1 \) for all \( n \) while for \( \beta = 1 \),

\[ a_i^{(n)} = (1/2)^i n, \ i=0,1,2,.... \]

Consequently, \( \eta_n = 0 \) for \( \beta = 0 \) or 1 for all \( n \). Now we want to know what happens if \( 0 < \beta < 1 \). Since \( a_0^{(n)} \equiv 1 \) for all \( \beta \), we may write

\[ \eta_n = \eta(\beta,n) = a_2^{(n)} a_0^{(n)} - [a_1^{(n)}]^2 \]

\[ = (c_2^{n+\gamma_2^{(n)}} - c_0^{n+\gamma_0^{(n)}})^2 - [c_1^{n+\gamma_1^{(n)}}]^2 \]

\[ = \gamma D + \gamma^2 E \]

where
\[ D = c^n_2 w_0^{(n)} + c^n_0 w_2^{(n)} - 2c^n_1 w_1^{(n)} \]

\[ E = w_2^{(n)} w_0^{(n)} - [w_1^{(n)}]^2 \]

Rewriting \( D \) we have

\[ D = \sum_{r=0}^{n-1} \left[ c^n_2 c_0^r + c^n_0 c_2^r - 2c^n_1 c_1^r \right] \]

\[ = \sum_{r=0}^{n-1} \beta^{n+r} \left[ \left( \frac{1}{2} \right)^r - \left( \frac{1}{2} \right)^n \right]^2 \]

\[ > 0 \text{ for } 0 < \beta < 1, \ n \geq 1 \]

Because \( c^n_2 c_0^r - c^n_1 c_1^r = 0 \), for all \( r \), we can express \( E \) as

\[ E = w_2^{(n)} w_0^{(n)} - [w_1^{(n)}]^2 \]

\[ = \sum_{r=0}^{n-1} \sum_{s=0}^{r} \left[ c^n_2 c_0^s + c^n_0 c_2^s - 2c^n_1 c_1^s \right] \]

\[ = \sum_{r=0}^{n-1} \sum_{s=0}^{r} \beta^{r+s} \left[ \left( \frac{1}{2} \right)^s - \left( \frac{1}{2} \right)^r \right]^2 \]

\[ > 0 \text{ for } 0 < \beta < 1, \ n \geq 1 \]

with equality only if \( n = 1 \) and \( 0 < \beta < 1 \). Consequently,

\[ \eta_n = \gamma D + \gamma^2 E > 0 \]

(2.49)

for \( 0 < \beta < 1 \) and \( n \geq 1 \). Hence \( \eta_n \geq 0 \) for \( 0 \leq \beta \leq 1 \) and for all \( n \geq 1 \) where equality holds only if \( \beta = 0 \) or 1. \( \eta_n \) is
depicted graphically in Figure 1 for different values of \( n \) and \( \beta \). Since \( \eta_{\infty} > 0 \) for \( 0 < \beta < 1 \), Equations 2.45 - 2.47 show that even in equilibrium (\( n \to \infty \)) the frequencies of genotypes with respect to two unlinked loci are not the product of the corresponding genotypic frequencies at each locus. This implies that genotypes at the two loci are not independently distributed.

3. Three loci

We now extend the discussion to three unlinked loci under the same assumptions. As before the part of a population which arises from random mating depends upon the frequencies of the gametes. We have already seen that this changes with time and the mathematics gets quite complicated when more and more loci are involved even for a simpler situation of pure random mating. The treatment of changing relations among three or more loci for pure random mating population has been discussed by Geiringer (1944, 1948) and by Bennett (1954). These formulas can be usefully employed while studying the changes in the gametic frequencies under the model of mixed selfing and random mating because the frequency of a gamete under such a system can be expressed as a linear combination of the corresponding gametic frequencies under pure random mating and under pure selfing with appropriate weights. Ghai (1964) analyzed this model for an arbitrary initial population with reference to two
Figure 1. Effect of proportion of selfing on $\eta_n$.
alleles at each locus. These results can be extended to cover multiallelic systems but as pointed out before the algebra gets quite extensive. Therefore, we shall restrict ourselves to an initial random mating population in equilibrium. This will simplify the situation in that the gametic array will not change with time. With three loci there are four different types of genotypes. Following the same approach as used previously we have the following recursion formulas:

\[
\begin{align*}
\phi_3' &= \gamma \phi_3^{(0)} + c_3 \phi_3 \\
\phi_2' &= \gamma \phi_2^{(0)} + c_3 (\phi_2 + \phi_{21}) \\
\phi_1' &= \gamma \phi_1^{(0)} + c_3 (\phi_1 + \phi_{11} + \phi_{12}) \\
\phi_0' &= \gamma \phi_0^{(0)} + c_3 (\phi_0 + \phi_{01} + \phi_{02} + \phi_{03})
\end{align*}
\] (2.50)

As before these equations are solved successively. Also, many of the difference equations are similar in form to the ones already solved. The form of their solutions will be used here and later whenever needed. The solution of the first equation is easy and is given by

\[
\phi_3^{(n)} = a_3^{(n)} \phi_3^{(0)}
\] (2.51)

In order to solve the rest of the equations, we need further the recursion formulas for \(\phi_{21}^{(n)}\), \(\phi_{11}^{(n)}\), \(\phi_{12}^{(n)}\) etc. Thus,
\[ \varphi_{21}^{(n)} = \gamma \varphi_{21}^{(0)} + c_2 \varphi_{21}^{(n-1)} \]
\[ = a_2^{(n)} \varphi_{21}^{(0)} \] (2.52)

\[ \varphi_{12}^{(n)} = \gamma \varphi_{12}^{(0)} + c_1 \varphi_{12}^{(n-1)} \]
\[ = a_1^{(n)} \varphi_{12}^{(0)} \] (2.53)

\[ \varphi_{11}^{(n)} = \gamma \varphi_{11}^{(0)} + c_2 \varphi_{11}^{(n-1)} + 2c_2 \varphi_{12}^{(n-1)} \]
\[ = \gamma \varphi_{11}^{(0)} + c_2 \varphi_{11}^{(n-1)} + 2c_2 a_1^{(n-1)} \varphi_{12}^{(0)} \]
\[ = a_2^{(n)} \varphi_{11}^{(0)} + 2c_2 b_21 \varphi_{12}^{(0)} \]
\[ = a_2^{(n)} \varphi_{11}^{(0)} + 2[a_1^{(n)} - a_2^{(n)}] \varphi_{12}^{(0)} \] (2.54)

\[ \varphi_{03}^{(n)} = \gamma \varphi_{03}^{(0)} + c_0 \varphi_{03}^{(n-1)} \]
\[ = a_0^{(n)} \varphi_{03}^{(0)} \] (2.55)

\[ \varphi_{02}^{(n)} = \gamma \varphi_{02}^{(0)} + c_1 \varphi_{02}^{(n-1)} + 3c_1 \varphi_{03}^{(n-1)} \]
\[ = \gamma \varphi_{02}^{(0)} + c_1 \varphi_{02}^{(n-1)} + 3c_1 a_0^{(n-1)} \varphi_{03}^{(0)} \]
\[ = a_1^{(n)} \varphi_{02}^{(0)} + 3c_1 b_10 \varphi_{03}^{(0)} \]
\[ = a_1^{(n)} \varphi_{02}^{(0)} + 3[a_0^{(n)} - a_1^{(n)}] \varphi_{03}^{(0)} \] (2.56)
\[ \phi_{01}^{(n)} = \gamma \phi_{01}^{(0)} + c_2 \phi_{01}^{(n-1)} + 2c_2 \phi_{02}^{(n-1)} + 3c_2 \phi_{03}^{(n-1)} \]
\[ = \gamma \phi_{01}^{(0)} + c_2 \phi_{01}^{(n-1)} + 2c_2 a_1^{(n-1)} \phi_{02}^{(0)} + 3c_2 [3a_0^{(n-1)} - 2a_1^{(n-1)}] \phi_{03}^{(0)} \]
\[ = a_2^{(n)} \phi_{01}^{(0)} + 2c_2 b_{21}^{(n)} \phi_{02}^{(0)} + 3c_2 [3b_{20}^{(n)} - 2b_{21}^{(n)}] \phi_{03}^{(0)} \]
\[ = a_2^{(n)} \phi_{01}^{(0)} + 2[a_1^{(n)} - a_2^{(n)}] \phi_{02}^{(0)} + 3[a_0^{(n)} - 2a_1^{(n)} + a_2^{(n)}] \phi_{03}^{(0)} \quad (2.57) \]

Using the last three equations of 2.50 in conjunction with the solutions 2.52 - 2.57, we obtain

\[ \phi_2^{(n)} = \gamma \phi_2^{(0)} + c_3 \phi_2^{(n-1)} + c_3 a_2^{(n-1)} \phi_2^{(0)} \]
\[ = a_3^{(n)} \phi_2^{(0)} + c_3 b_{32}^{(n)} \phi_2^{(0)} \]
\[ = a_3^{(n)} \phi_2^{(0)} + [a_2^{(n)} - a_3^{(n)}] \phi_2^{(0)} \quad (2.58) \]

\[ \phi_1^{(n)} = \gamma \phi_1^{(0)} + c_3 \phi_1^{(n-1)} + c_3 a_2^{(n-1)} \phi_1^{(0)} + c_3 [3a_1^{(n-1)} - 2a_2^{(n-1)}] \phi_1^{(0)} \]
\[
\phi^{(n)} = a_3^{(n)} \phi^{(0)} + c_3 b^{(n)} \phi^{(0)} + c_3^2 \phi^{(0)} + c_3^3 \phi^{(0)} \\
+ c_3 [3b^{(n)} - 2b^{(n)}] \phi^{(0)} \\
= a_3^{(n)} \phi^{(0)} + [a_2^{(n)} - a_3^{(n)}] \phi^{(0)} \\
+ [a_1^{(n)} - 2a_2^{(n)} + a_3^{(n)}] \phi^{(0)}
\]

\[ (2.59) \]

\[
\phi^{(n)} = \gamma \phi^{(0)} + c_3 \phi^{(n-1)} + c_3 a_2^{(n-1)} \phi^{(0)} \\
+ c_3 [3a_1^{(n-1)} - 2a_2^{(n-1)}] \phi^{(0)} \\
+ c_3 [7a_0^{(n-1)} - 9a_1^{(n-1)} + 3a_2^{(n-1)}] \phi^{(0)} \\
= a_3^{(n)} \phi^{(0)} + c_3 b^{(n)} \phi^{(0)} + c_3 [3b^{(n)} - 2b^{(n)}] \phi^{(0)} \\
+ c_3 [7b^{(n)} - 9b^{(n)} + 3b^{(n)}] \phi^{(0)} \\
= a_3^{(n)} \phi^{(0)} + [a_2^{(n)} - a_3^{(n)}] \phi^{(0)} \\
+ [a_1^{(n)} - 2a_2^{(n)} + a_3^{(n)}] \phi^{(0)} \\
+ [a_0^{(n)} - 3a_1^{(n)} + 3a_2^{(n)} - a_3^{(n)}] \phi^{(0)} .
\]

\[ (2.60) \]

The frequencies of genotypes after \( n \) generations are conveniently expressed, in terms of the initial distribution, as
\[ \phi_i^{(n)} = \sum_{j=0}^{3-i} \phi_{ij}^{(0)} \sum_{x=0}^j (-1)^{j-x} a_{3-x}^{(n)} \]  
(2.61)

and in the limit as \( n \to \infty \)

\[ \phi_i^{(\infty)} = \sum_{j=0}^{3-i} \phi_{ij}^{(0)} \sum_{x=0}^j (-1)^{j-x} \gamma[1-c_{3-x}]^{-1} \]  
(2.62)

for \( i=0,1,2,3 \).

4. Multiple loci

Suppose we have \( k \) unlinked loci. Consider the initial population

\[
\prod_{\text{loci}} \left[ \sum_{s=1}^k \sum_{s=1}^{s=1} p_i p_j A_i A_j \right]
\]

which has the random mating structure, the gene frequency of \( A_i \), which is \( i \)th allele at locus \( A \), being \( p_i \). From the formulas for the genotypic distributions obtained for one, two and three unlinked loci, we can easily extend the analysis to multiple loci under the assumptions stated previously. It can be easily deduced that genotypic frequency of an \( i \)-factor heterozygote \( (i=0,1,2,\ldots,k) \) in the \( n \)th generation of mixed selfing and random mating can be expressed as \( 2^i \phi_i^{(n)} \) where

\[ \phi_i^{(n)} = \sum_{j=0}^{k-i} \phi_{ij}^{(0)} \sum_{x=0}^j (-1)^{j-x} a_{k-x}^{(n)} \]  
(2.63)

for \( i=0,1,2,\ldots,k \), where
For a given $i$, there are $\binom{k}{i}$ genotypes which are heterozygous for $i$ of the loci and homozygous for the remaining loci. The value of $\phi_i^{(0)}$ to be put in the right hand side of Equation 2.63 will depend upon the particular genotype under consideration for a given $i$.

The limiting distribution is determined by

$$
\phi_i^{(\infty)} = \lim_{n \to \infty} \phi_i^{(n)} = \sum_{j=0}^{k-i} \phi_i^{(0)} \sum_{x=0}^{j} \frac{(-1)^{j-x}(1-\beta)}{1-c_{k-x}}
$$

(2.64)

for $i=0,1,2,...,k$. For example, suppose $k=3$ and we are interested in the equilibrium frequency of a double heterozygote, say $A_1A_2B_1B_2C_1C_1$. Then from Equation 2.64 this is given by $2^2\phi_2^{(\infty)}$ where

$$
\phi_2^{(\infty)} = \frac{1-\beta}{1-c_3} \phi_2^{(0)} + \frac{1-\beta}{1-c_2} \frac{1-\beta}{1-c_3} \phi_2^{(0)}
$$

$$
= \frac{\gamma}{1-c_3} \phi_2^{(0)} + \frac{\gamma C_3}{(1-c_2)(1-c_3)} \phi_2^{(0)}
$$

(2.65)
or in terms of $P$'s

$$P^{(\infty)}_{121211} = \varphi^{(\infty)}_2$$

$$= \frac{\gamma}{1-c_3} P^{(0)}_{121211} + \frac{\gamma c_3}{(1-c_2)(1-c_3)} P^{(0)}_{121211}. \quad (2.66)$$

\textbf{B. Loss of Heterozygosity}

We have considered in the previous section the consequences of the system of mixed random mating and selfing on the distribution of genotypes in the case of several unlinked loci with multiple alleles, in an initial panmictic population. Such a system of mating provides means for maintaining polymorphic variation in the absence of heterozygote advantage. We shall consider in this section the effect of such a system on heterozygosity in the population. For simplicity we shall restrict ourselves to two alleles at each locus. Let us consider a panmictic population

$$\prod_{\text{loci}}^{k} \left[ p_1^2 A_1A_1 + 2p_1p_2 A_1A_2 + p_2^2 A_2A_2 \right]$$

as an initial population where $p_1 + p_2 = 1$ and $k$ is the number of independently segregating loci.

The distribution that is more relevant here will be the probability distribution, $\{H_i^{(n)}\}, i=0,1,2,\ldots,k,$ of different types of genotypes, where $H_i^{(n)}$ is the probability in generation $n$ that an individual is heterozygous for $i$ of
the k loci and homozygous for the remainder. That is, \( H_i^{(n)} \) is the sum of the probabilities of genotypes in which i of the k loci are in heterozygous condition and the remaining are homozygous. For example, \( H_i^{(n)} \), when only two gene pairs \((A_1, A_2)\) and \((B_1, B_2)\) are involved, is the sum of frequencies of genotypes \( A_1 A_2 B_1 B_1, A_1 A_2 B_2 B_1, A_1 A_1 B_1 B_2, \) and \( A_2 A_2 B_1 B_2 \) each of which is heterozygous with respect to one locus and homozygous for the other. \( H_i^{(0)} \), \( i=0,1,2,...,k \), are the corresponding probabilities in the initial population. We shall further define, \( F_{n,k} \), the coefficient of inbreeding in generation n as the loss of heterozygosity relative to that in the initial population. When there is no ambiguity index k may be dropped.

1. Single locus

The frequency \( H_1^{(n)} \) in generation n is obtained from Equations 2.15 and 2.29 as

\[
H_1^{(n)} = 2P_{12}^{(n)} = [c_1^n + \gamma w_1^{(n)}]H_1^{(0)}
\]

\[= a_1^{(n)}H_1^{(0)} \]. \hspace{1cm} (2.67)

Thus \( F_{n,1} \) in generation n is given by

\[
F_{n,1} = 1 - H_1^{(n)}/H_1^{(0)}
\]

\[= 1 - a_1^{(n)} \]
\[(\beta/2) + (\beta/2)^2 + \ldots + (\beta/2)^n\]
\[= (\beta/2) \left[ \frac{1 - (\beta/2)^n}{1 - \beta/2} \right] \quad (2.68)\]

and in the limit as \(n \to \infty\)

\[F_{\infty,1} = \lim_{n \to \infty} F_{n,1} = \frac{\beta}{2 - \beta} \quad (2.69)\]

Equation 2.68 shows that \(F_{n,1}\) is independent of the initial distribution. The sequence \(\{F_{n,1}\}\) is monotonic increasing in both \(n\) and \(\beta\), \(0 < \beta \leq 1\). The loss in heterozygosity \((F_{n,1})\) in different generations has been examined in Table 1 for various amounts of selfing in the population. The values for \(\beta = 1.0\) correspond to the case of complete selfing. The values for \(n = \infty\) correspond to the maximum loss in heterozygosity that can occur under continued mixed random mating and selfing. It is seen from the table that loss in heterozygosity increases with the increase in the amount of selfing as it should, but is not proportionate to the increase in the amount of selfing except in the first generation. It may also be seen that for a given \(\beta\), most of the expected loss in heterozygosity takes place within two or three generations.

2. Two loci

The following relations can be easily obtained by using Equations 2.33 and 2.37 for the frequencies of double and
Table 1. Loss of heterozygosity relative to the initial heterozygosity \( (F_{n,1}) \)
under mixed random mating and selfing

<table>
<thead>
<tr>
<th>n</th>
<th>( \beta = .10 )</th>
<th>.20</th>
<th>.30</th>
<th>.40</th>
<th>.50</th>
<th>.60</th>
<th>.70</th>
<th>.80</th>
<th>.90</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.050</td>
<td>.100</td>
<td>.150</td>
<td>.200</td>
<td>.250</td>
<td>.300</td>
<td>.350</td>
<td>.400</td>
<td>.450</td>
<td>.500</td>
</tr>
<tr>
<td>2</td>
<td>.052</td>
<td>.110</td>
<td>.172</td>
<td>.240</td>
<td>.312</td>
<td>.390</td>
<td>.472</td>
<td>.560</td>
<td>.652</td>
<td>.750</td>
</tr>
<tr>
<td>3</td>
<td>.053</td>
<td>.111</td>
<td>.176</td>
<td>.248</td>
<td>.328</td>
<td>.417</td>
<td>.515</td>
<td>.624</td>
<td>.744</td>
<td>.875</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>.111</td>
<td>.176</td>
<td>.249</td>
<td>.332</td>
<td>.425</td>
<td>.530</td>
<td>.650</td>
<td>.785</td>
<td>.937</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>.250</td>
<td>.333</td>
<td>.427</td>
<td>.535</td>
<td>.660</td>
<td>.803</td>
<td>.969</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.666</td>
<td>.817</td>
<td>.999</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.818</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \infty )</td>
<td>.053</td>
<td>.111</td>
<td>.176</td>
<td>.250</td>
<td>.333</td>
<td>.429</td>
<td>.538</td>
<td>.667</td>
<td>.818</td>
<td>1.000</td>
</tr>
</tbody>
</table>

\( \beta = \) probability of selfing.
single heterozygotes in the nth generation.

\[ H_2^{(n)} = a_2^{(n)} H_2^{(0)} \]

\[ H_1^{(n)} = a_1^{(n)} H_1^{(0)} + 2[a_1^{(n)} - a_2^{(n)}] H_2^{(0)} \]  \hspace{1cm} (2.70)

Let

\[ H^{(n)} = P[\text{An individual is heterozygous at least at one locus in generation } n] \]  \hspace{1cm} (2.71)

Then

\[ H^{(n)} = H_1^{(n)} + H_2^{(n)} \]

\[ = a_1^{(n)} H_1^{(0)} + [2a_1^{(n)} - a_2^{(n)}] H_2^{(0)} \]  \hspace{1cm} (2.72)

and hence,

\[ F_{n,2} = 1 - \frac{H^{(n)}}{H^{(0)}} \]

\[ = \frac{[1 - a_1^{(n)}] - [a_1^{(n)} - a_2^{(n)}] H_2^{(0)}}{H^{(0)}} \]  \hspace{1cm} (2.73)

It may be seen that unlike in a single locus situation, \( F_{n,2} \) depends upon the constitution of the initial population in addition to the probability of selfing.

We can rewrite Equation 2.73 as

\[ F_{n,2} = \frac{[1 - 2a_1^{(n)} + a_2^{(n)}] + [a_1^{(n)} - a_2^{(n)}] H_1^{(0)}}{H^{(0)}} \]

\[ = G_{n,2} + \frac{[a_1^{(n)} - a_2^{(n)}] H_1^{(0)}}{H^{(0)}} \]  \hspace{1cm} (2.74)
where
\[ G_{n,2} = 1 - 2a_1^{(n)} + a_2^{(n)} \]
\[ = a_0^{(n)} - 2a_1^{(n)} + a_2^{(n)} \] \hspace{1cm} (2.75)

since \( a_0^{(n)} = 1 \) for all \( n \). Now \( a_i^{(n)} - a_j^{(n)} \) > 0 for \( j > i \) and \( 0 < \beta \leq 1 \). This can be seen as follows:

\[ a_i^{(n)} - a_j^{(n)} = c_i^n + \gamma[1 + c_i + c_i^2 + \ldots + c_i^{n-1}] \]
\[ - c_j^n - \gamma[1 + c_j + c_j^2 + \ldots + c_j^{n-1}] \]
\[ = (c_i^n - c_j^n) + \gamma[(c_i - c_j) + (c_i^2 - c_j^2) + \ldots \]
\[ + (c_i^{n-1} - c_j^{n-1})] \]
\[ > 0, \text{ for all } n \geq 1 \] \hspace{1cm} (2.76)

because \( c_i > c_j \) for \( j > i \) and \( 0 < \beta \leq 1 \). Since \( H_1^{(0)} \) and \( H_2^{(0)} \) are positive,

\[ G_{n,2} < F_{n,2} < 1 - a_1^{(n)} = F_{n,1} \] \hspace{1cm} (2.77)

which shows that \( F_{n,1} \) is an upper bound of \( F_{n,2} \). Also \( G_{n,2} \) which is independent of the initial distribution is a lower bound of \( F_{n,2} \).
3. Three loci

With three unlinked loci, the frequencies of various types of heterozygotes in generation $n$ can be expressed in terms of the initial values as follows:

$$H_3^{(n)} = a_3^{(n)} H_3^{(0)}$$

$$H_2^{(n)} = a_2^{(n)} H_2^{(0)} + 3[a_2^{(n)} - a_3^{(n)}]H_3^{(0)}$$

$$H_1^{(n)} = a_1^{(n)} H_1^{(0)} + 2[a_1^{(n)} - a_2^{(n)}]H_2^{(0)} + 3[a_1^{(n)} - 2a_2^{(n)} + a_3^{(n)}]H_3^{(0)}$$  \[ (2.78) \]

and hence,

$$H^{(n)} = H_1^{(n)} + H_2^{(n)} + H_3^{(n)}$$

$$= a_1^{(n)} H_1^{(0)} + [2a_1^{(n)} - a_2^{(n)}]H_2^{(0)}$$

$$+ [3a_1^{(n)} - 3a_2^{(n)} + a_3^{(n)}]H_3^{(0)}$$  \[ (2.79) \]

Therefore,

$$F_{n,3} = 1 - H^{(n)}/H^{(0)}$$

$$= [1 - a_1^{(n)}] - [a_1^{(n)} - a_2^{(n)}]H_2^{(0)}/H^{(0)}$$

$$- [2a_1^{(n)} - 3a_2^{(n)} + a_3^{(n)}]H_3^{(0)}/H^{(0)}$$

$$= G_{n,3} + I_{n,3}$$  \[ (2.80) \]
where

\[
G_{n,3} = a_0^{(n)} - 3a_1^{(n)} + 3a_2^{(n)} - a_3^{(n)}
\]

\[
I_{n,3} = [2a_1^{(n)} - 3a_2^{(n)} + a_3^{(n)}][1-H_3^{(0)}/H^{(0)}]
\]

\[
- [a_1^{(n)} - a_2^{(n)}]H_2^{(0)}/H^{(0)}
\]

\[
= [2a_1^{(n)} - 3a_2^{(n)} + a_3^{(n)}](H_1^{(0)}/H^{(0)})
\]

\[
+ [a_1^{(n)} - 2a_2^{(n)} + a_3^{(n)}](H_2^{(0)}/H^{(0)})
\]

It can be easily seen as before that \(2a_1^{(n)} - 3a_2^{(n)} + a_3^{(n)} > 0\) and \(a_1^{(n)} - 2a_2^{(n)} + a_3^{(n)} > 0\) which implies that \(I_{n,3} > 0\).

Therefore,

\[
G_{n,3} < F_{n,3} < F_{n,1}
\]

for \(n \geq 1\) and \(0 < \beta < 1\), so that \(F_{n,1}\) is an upper bound and \(G_{n,3}\) is a lower bound of \(F_{n,3}\).

4. **Multiple loci**

Let us suppose that there are \(k\) unlinked loci. Following the same procedure as in the case of one, two and three loci, we can easily extend the results to multiple loci. The previous derivations clearly suggest the structure of numerical constants involved in various terms and it is, therefore, possible to write down a general expression for
and $G_n,k$ which is a lower bound of $F_{n,k}$.

\[
H(n) = a_1^{(n)} H_1^{(0)} + [2a_1^{(n)} - a_2^{(n)}] H_2^{(0)} + [3a_1^{(n)} - 3a_2^{(n)} + a_3^{(n)}] H_3^{(0)} + ... + \left(\binom{k}{1} a_1^{(n)} - \binom{k}{2} a_2^{(n)} + \binom{k}{3} a_3^{(n)} - \ldots \pm \binom{k}{k} a_k^{(n)}\right) H_k^{(0)}
\]  

(2.84)

\[
G_{n,k} = \binom{k}{0} a_0^{(n)} - \binom{k}{1} a_1^{(n)} + \binom{k}{2} a_2^{(n)} - \ldots \pm \binom{k}{k} a_k^{(n)}
\]  

(2.85)

As indicated before the exact values of $F_{n,k}$ ($k \geq 2$) depend upon the initial distribution of the population apart from the probability of selfing. $G_{n,k}$ is only a lower bound of $F_{n,k}$. The values of $G_{n,k}$ up to 10 loci were discussed by Ghai (1965). To discuss $F_{n,k}$ in more detail we consider the following special case.

Suppose we have an initial population which has a random mating structure with equal gene frequency of one-half at each locus. Then the distribution of the number of loci in heterozygous condition is binomial $Bi(k,1/2)$. In this situation we have

\[
H^{(n)}(k=1) = (1/2)a_1^{(n)}
\]

\[
H^{(n)}(k=2) = (1/2)^2 [\binom{2}{1} a_1^{(n)} + \binom{2}{2} (2a_1^{(n)} - a_2^{(n)})]
\]
\[ H^{(n)}(k=3) = \left(\frac{1}{2}\right)^3 \left[ \binom{3}{1} a_1^{(n)} + \binom{3}{2} (2a_1^{(n)} - a_2^{(n)}) + \left(\binom{3}{3} 3a_1^{(n)} - 3a_2^{(n)} + a_3^{(n)}\right) \right] \]

and so on. In general, for \( k \) loci,

\[ H^{(n)} = \left(\frac{1}{2}\right)^k \left[ \binom{k}{1} a_1^{(n)} + \binom{k}{2} (2a_1^{(n)} - a_2^{(n)}) + \ldots \right. \]

\[ + \left. \binom{k}{k} \left\{ (\binom{k}{1}) a_1^{(n)} - (\binom{k}{2}) a_2^{(n)} + \ldots + (\binom{k}{k}) a_k^{(n)} \right\} \right] \quad (2.86) \]

because \( H^{(0)}_i = \binom{i}{0} \left(\frac{1}{2}\right)^i, \ i=0,1,2,\ldots,k. \) We notice that the coefficients of \( a_1^{(n)}, a_2^{(n)}, \ldots \) have a definite structure which is expressed as follows:

**Coeff. of** \( a_1^{(n)} = \left(\frac{1}{2}\right)^k \left[ \binom{k}{1} + \binom{k}{2} \binom{2}{1} + \binom{k}{3} \binom{3}{1} \right. \]

\[ + \ldots + \binom{k}{k} \binom{k}{1} \]

\[ = \left(\frac{1}{2}\right)^k \sum_{j=1}^{k} \binom{k}{j} \binom{j}{1} \]

\[ = \left(\frac{1}{2}\right)^k k 2^{k-1} = \frac{k}{2} \quad (2.87) \]

**Coeff. of** \( a_2^{(n)} = -\left(\frac{1}{2}\right)^k \left[ \binom{k}{2} \binom{2}{2} + \binom{k}{3} \binom{3}{2} + \ldots + \binom{k}{k} \binom{k}{2} \right] \]

\[ - \left(\frac{1}{2}\right)^k \sum_{j=2}^{k} \binom{k}{j} \binom{j}{2} \]

\[ - \left(\frac{1}{2}\right)^k \frac{k(k-1)}{2} 2^{k-2} \]
and in general

\[
\text{Coeff. of } a_i^{(n)} = (-1)^{i-1} \left( \frac{1}{2} \right)^k \left[ \binom{k}{i} \binom{i}{1} + \binom{k}{i+1} \binom{i+1}{1} \right]
\]

\[+ \ldots + \binom{k}{i} \binom{k}{i} \]

\[= (-1)^{i-1} \left( \frac{1}{2} \right)^k \sum_{j=1}^{k} \binom{k}{j} \binom{i}{j} \]

\[= (-1)^{i-1} \left( \frac{1}{2} \right)^k \sum_{j=1}^{k} \frac{k!}{j!(k-j)!} \frac{j!}{i!(j-i)!} \]

\[= (-1)^{i-1} \left( \frac{1}{2} \right)^k \binom{k}{i} \sum_{j=1}^{k} \binom{k-i}{j-i} \]

\[= (-1)^{i-1} \left( \frac{1}{2} \right)^k \binom{k}{i} 2^{-i} \quad (2.89)\]

for \(i=1,2,3,...,k\).

Therefore, we can rewrite Equation 2.86 as

\[
H^{(n)} = \sum_{i=1}^{k} (-1)^{i-1} \left( \frac{1}{2} \right)^i a_i^{(n)} \quad (2.90)
\]

Hence,

\[
F_{n,k} = \frac{(H^{(0)} - H^{(n)})}{H^{(0)}}
\]

\[= 1 - \left( \frac{1}{2} \right)^k - \sum_{i=1}^{k} (-1)^{i-1} \left( \frac{1}{2} \right)^i a_i^{(n)} \]

\[\frac{1 - \left( \frac{1}{2} \right)^k - \sum_{i=1}^{k} (-1)^{i-1} \left( \frac{1}{2} \right)^i a_i^{(n)}}{1 - \left( \frac{1}{2} \right)^k} \]
Table 2 gives the loss in heterozygosity \( F_{n,k} \) in successive generations of mixed random mating and selfing up to eight factors for different values of \( \beta \). \( F_{n,k} \) is an increasing function of \( n \) and \( \beta \). The values of \( F_{n,k} \) decrease with the increase in the number of segregating factors, \( k \). These values show that the heterozygosity is not completely lost except for complete selfing (\( \beta=1 \)). The maximum loss that can occur under continued mixed random mating and selfing in the case of a single locus (\( k=1 \)) is of the order of 5 percent when there is 10 percent selfing in the population which rises to 82 percent when there is 90 percent selfing. These values are of the order of 4 percent and 80 percent for \( k=2 \), 3 percent and 76 percent for \( k=4 \), and 1 percent and 71 percent for \( k=8 \). It may also be observed from the table that for a given \( \beta \) and \( k \), the maximum value is reached very rapidly within a few generations.

If we start with a uniform population in which
| $k=1$   | $n=1$ | .050 | .100 | .150 | .200 | .250 | .300 | .350 | .400 | .450 | .500 |
| $n=2$  | .052 | .110 | .172 | .240 | .312 | .390 | .472 | .560 | .652 | .750 |
| $n=3$  | .053 | .111 | .176 | .248 | .328 | .417 | .515 | .624 | .744 | .875 |
| $n=4$  | .053 | .111 | .176 | .250 | .332 | .425 | .530 | .650 | .785 | .937 |
| $n=\infty$ | .053 | .111 | .176 | .250 | .333 | .429 | .538 | .667 | .818 | 1.000 |
| $k=2$   | $n=1$ | .042 | .083 | .125 | .167 | .208 | .250 | .292 | .333 | .375 | .417 |
| $n=2$  | .044 | .094 | .149 | .210 | .276 | .347 | .424 | .507 | .594 | .687 |
| $n=3$  | .044 | .095 | .153 | .220 | .295 | .380 | .476 | .584 | .704 | .838 |
| $n=4$  | .044 | .095 | .154 | .222 | .300 | .390 | .495 | .616 | .757 | .918 |
| $n=\infty$ | .044 | .095 | .154 | .222 | .302 | .395 | .506 | .639 | .801 | 1.000 |
| $k=4$   | $n=1$  | .027 | .054 | .081 | .108 | .135 | .162 | .190 | .217 | .244 | .271 |
| $n=2$  | .030 | .066 | .107 | .154 | .207 | .266 | .331 | .401 | .477 | .559 |
| $n=3$  | .030 | .067 | .112 | .167 | .232 | .309 | .339 | .503 | .622 | .757 |
| $n=4$  | .030 | .067 | .113 | .170 | .239 | .324 | .426 | .550 | .697 | .873 |
| $n=\infty$ | .030 | .067 | .114 | .171 | .242 | .331 | .443 | .584 | .765 | 1.000 |
| $k=8$   | $n=1$ | .010 | .019 | .029 | .039 | .048 | .058 | .068 | .077 | .087 | .097 |
| $n=2$  | .012 | .029 | .051 | .078 | .109 | .146 | .187 | .234 | .285 | .341 |
| $n=3$  | .012 | .031 | .058 | .094 | .141 | .201 | .274 | .364 | .470 | .595 |
| $n=4$  | .012 | .031 | .059 | .099 | .152 | .244 | .318 | .437 | .588 | .775 |
| $n=\infty$ | .012 | .031 | .060 | .100 | .157 | .236 | .346 | .498 | .708 | 1.000 |

Table 2. Loss of heterozygosity relative to that in the initial population under mixed random mating and selfing up to eight factors.
individuals are completely heterozygous at all loci, for exam­
ple, a population obtained by crossing two inbred strains,
then one generation of mixed selfing and random mating will
yield a population which is in Hardy-Weinberg equilibrium
with equal gene frequency of one-half at each locus, assum­
ing that all genes are unlinked. Later generations of mixed
selfing and random mating will yield the same results as
obtained above for the special case. Therefore, starting
with a population \( H_k^{(0)} = H^{(0)} = 1 \), we shall have after \( n \)
generations

\[
H(1) = 1 - \left( \frac{1}{2} \right)^k \\
H(n) = \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \left( \frac{1}{2} \right)^i a_i^{(n-1)} , \quad n \geq 2 .
\] (2.92)

Thus, for \( n \geq 1 \)

\[
F_{n,k} = 1 - \frac{H(n)}{H^{(0)}} = 1 - H(n) \\
= 1 - \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \left( \frac{1}{2} \right)^i a_i^{(n-1)} \\
= \sum_{i=0}^{k} \binom{k}{i} \left( -\frac{1}{2} \right)^i a_i^{(n-1)}. \] (2.93)

When \( \beta = 1 \), i.e. complete selfing, Equation 2.93 reduces to
the known result
C. Genotypic Mean and Variance

In any breeding program it is important to have an idea of the variability that will be expected in the population in successive stages of breeding as well as the knowledge about the structure of the population. Means, variances, and covariances, etc., are useful tools for describing genetic populations. A vast amount of work has been done in this direction for means, variances, and covariances, regression and so on, for populations which are either purely random mated or inbred, e.g., selfing and the like. These results are very well documented in Kempthorne (1957). But there does not seem to be much work done for populations subjected to a mixture of breeding systems with respect to various statistical properties. Ghai (1964) discussed some results on this for two-allele situations. We will attempt to extend these further to multiallelic systems.

As before the general pattern will be to start with the simplest model and then to extend it to more complex
situations. Thus, we shall start with a population for which only one locus is segregating and then extend the results to multiple loci.

In the discussion that follows, in addition to the assumptions already made in Section A, it is assumed that the effect of environment may be discounted in the sense that the value observed for an individual depends solely on the genotype of the individual, and that various loci act additively. The later assumption may be justified on the consideration that actions of genes, in some situations, can be made approximately additive by suitable transformation of the scale.

1. Single locus

Let $Y$ be a random variable defined on the sample space $\Omega = \{A_iA_j\}$ of genotypes. Let $Y = y_{ij}$ be the genotypic value of $A_iA_j$ which will also be equal to $y_{ji}$, the genotypic value of $A_jA_i$. Therefore,

$$P[Y = y_{ii}] = p_{ii}$$

$$P[Y = y_{ij}] = 2p_{ij} \quad i \neq j$$

for $i, j = 1, 2, \ldots, s$.

From Section A we know that the probability distribution, $\{p_{ij}^{(n)}\}$, after $n$ generations of mixed random mating and selfing is as given in Table 3. The total genotypic
Table 3. Distribution of genotypes and their genotypic values

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Genotypic value</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_{ii}</td>
<td>Y_{ii}</td>
<td>p^{(n)}<em>{ii} = \left[ \gamma p</em>{i}^2 + (\beta/2)p_{j}\right]\left[1 - (\beta/2)^n\right]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\quad + (\beta/2)^n p^{(0)}_{ii}</td>
</tr>
<tr>
<td>A_{ij}</td>
<td>Y_{ij}</td>
<td>2p^{(n)}<em>{ij} = 2\gamma p</em>{i}p_{j}\left[1 - (\beta/2)^n\right]</td>
</tr>
<tr>
<td>(i\neq j)</td>
<td></td>
<td>\quad + 2(\beta/2)^n p^{(0)}_{ij}</td>
</tr>
</tbody>
</table>

Mean in generation \( n \) is then given by

\[ \mu_{n}(Y) = \sum_{i=1}^{s} \sum_{j=1}^{s} p^{(n)}_{ij} Y_{ij} = (\beta/2)^n \mu_0(Y) + \frac{1 - (\beta/2)^n}{1 - \beta/2} \left[ \gamma \mu_R(Y) + (\beta/2)\mu_I(Y) \right] = (1 - F_n)\mu_R(Y) + F_n \mu_I(Y) + (\beta/2)^n[\mu_0(Y) - \mu_R(Y)] \] (2.95)

where
\[ F_n = \frac{p/2}{1 - (p/2)^n} \]

\[ = \frac{p/2}{1 - p/2} + \frac{p/2}{1 - p/2}^2 + \ldots + \frac{p/2}{1 - p/2}^n, \quad (2.96) \]

\( \mu_0(Y) \) is the mean in the initial population, and \( \mu_R(Y) \) and \( \mu_I(Y) \) are the means in the random mating and completely inbred populations respectively. That is,

\[ \mu_0(Y) = \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij} y_{ij} \]

\[ \mu_R(Y) = \sum_{i=1}^{s} \sum_{j=1}^{s} p_i p_j y_{ij} \]

\[ \mu_I(Y) = \sum_{i=1}^{s} p_i y_{ii}. \quad (2.97) \]

If we start with an initial population which was obtained by random mating, then \( \mu_0(Y) = \mu_R(Y) \) and the last term in Equation 2.95 vanishes and we get the usual known result (Kempthorne, 1957, p. 377; Wright, 1969, p. 466). This will also be true in the limit as \( n \to \infty \) because as we have observed before the distribution of genotypes in equilibrium is independent of the initial distribution and depends only on the gene frequencies and \( \beta \).

The genotypic variance for the population in generation \( n \) is denoted by \( V_n(Y) \), where
\[ V_n(Y) = \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij}^{(n)} y_{ij}^2 - \left[ \mu_n(Y) \right]^2 \]

\[ = \sum_{i=1}^{s} \sum_{j=1}^{s} (1-F_n)p_{ij}p_j y_{ij}^2 + \sum_{i=1}^{s} F_n p_{ii} y_{ii}^2 \]

\[ + (\beta/2)^n \sum_{i=1}^{s} \sum_{j=1}^{s} (p_{ij}^{(0)} - p_{ij}p_j) y_{ij}^2 - \mu_n^2(Y) \]

\[ = (1-F_n) V_R(Y) + F_n V_I(Y) + F_n (1-F_n) \left[ \mu_R(Y) - \mu_I(Y) \right]^2 \]

\[ + (\beta/2)^n [V_0(Y) - V_R(Y)] + (\beta/2)^n [1 - (\beta/2)^n] \]

\[ \times \left[ \mu_0(Y) - \mu_R(Y) \right]^2 \]

\[ + 2F_n (\beta/2)^n [\mu_0(Y) - \mu_R(Y)] \left[ \mu_R(Y) - \mu_I(Y) \right] \quad (2.98) \]

where \( V_0(Y) \) is the variance in the initial population, and \( V_R(Y) \) and \( V_I(Y) \) are the variances in the completely random mated and completely inbred populations, i.e.,

\[ V_R(Y) = \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij}p_j y_{ij}^2 - \mu_R^2(Y) \]

\[ V_I(Y) = \sum_{i=1}^{s} p_{ii} y_{ii}^2 - \mu_I^2(Y) \quad (2.99) \]

Again if \( p_{ij}^{(0)} = p_{ij}p_j \), the last three terms in Equation 2.98 vanish and it reduces to the known formula (Kempthorne, 1957, p. 274; Wright, 1969, p. 466).
Of particular interest is the case in which the effects of alleles are additive i.e. the absence of dominance. Letting \( y_i \) to be the contribution of allele \( A_i \), \( i=1,2,\ldots,s \), then the contribution of genotype \( A_iA_j \) is \( y_{ij} = y_i + y_j \). The mean is unchanged by the system of mixed mating in this case.

\[
\mu_R(Y) = \sum_{i=1}^{s} \sum_{j=1}^{s} p_i p_j (y_i + y_j) = 2 \sum_{i=1}^{s} p_i y_i
\]

\[
\mu_I(Y) = 2 \sum_{i=1}^{s} p_i y_i
\]

\[
\mu_0(Y) = \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij} (y_i + y_j) = 2 \sum_{i=1}^{s} p_i y_i
\]

and hence from Equation 2.95

\[
\mu_n(Y) = \mu_0(Y) = \mu_R(Y) = \mu_I(Y) \quad (2.100)
\]

With respect to the variance, we have

\[
V_R(Y) = \sum_{i=1}^{s} \sum_{j=1}^{s} p_i p_j (y_i + y_j)^2 - 4 \left( \sum_{i=1}^{s} p_i y_i \right)^2
\]

\[
= 2 \left[ \sum_{i=1}^{s} p_i y_i^2 - (\sum_{i=1}^{s} p_i y_i)^2 \right]
\]

\[
V_I(Y) = \sum_{i=1}^{s} p_i (2y_i)^2 - 4 \left( \sum_{i=1}^{s} p_i y_i \right)^2
\]

\[
= 2V_R(Y)
\]
\[ V_0(Y) = \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij}^{(0)} (y_i + y_j)^2 - 4 \sum_{i=1}^{s} p_i y_i^2 \]

\[ = 2 \sum_{i=1}^{s} p_i y_i^2 + 2 \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij}^{(0)} y_i y_j - 4 \sum_{i=1}^{s} p_i y_i^2 \]

\[ = V_R(Y) + 2 \sum_{i=1}^{s} \sum_{j=1}^{s} (p_{ij}^{(0)} - p_i p_j) y_i y_j \] (2.102)

This expression for the variance contains an additional term because of the deviation of the initial distribution from panmixia. The contribution due to this term becomes smaller and smaller as \( n \) increases and vanishes in the limiting case as \( n \to \infty \). In the limit as \( n \to \infty \), the variance becomes independent of the initial distribution of genotypes.

\[ V_\infty(Y) = \lim_{n \to \infty} V_n(Y) = (1 + F_\infty) V_R(Y) \]

\[ = \frac{2}{2 - \beta} V_R(Y) \] (2.102)

If \( \theta_{ij} = 0 \), that is we start with a random mating population in equilibrium, the second term in Equation
2.101 vanishes and the variance in generation \( n \) reduces to 
\[(1+F_n)V_{R}(Y)\]
which is the classical result of Wright (1921) for the case of purely additive genes. Since \( F_n \) is an increasing function of \( \beta \) and \( n \), the variance will increase from generation to generation for a given amount of selfing and also over the range of \( \beta \) for a given \( n \). The increase in variance with respect to \( \beta \) is linear for \( n = 1 \).

Now we consider a special case of \( s = 2 \) and \( \theta_{ij} = 0 \), \( i,j=1,2 \). This was first discussed by Ghai (1964). In the case of complete dominance of allele \( A_1 \) over allele \( A_2 \) he has observed that for values of the frequency of the recessive allele, \( p_2 \), less than or equal to one-half, the variance increases with \( n \) and \( \beta \). For \( p_2 > 0.7 \), the variance decreases. For \( p_2 = 0.6 \) and \( 0.7 \), the variance increases up to a certain stage and decreases thereafter. The relative variance \((V_n/V_0)\) for different values of the gene frequency, \( p_2 \), and the amount of selfing, \( \beta \), is given in Table 4.

2. Two loci

The structure of population with respect to two loci with alleles \( \{A_i\} \) and \( \{B_k\} \) under the system of mating being considered is given earlier in Section A. Let \((Y,Z)\) be a random variable defined on the sample space \( \Omega = \Omega_1 \times \Omega_2 \). Let \( y_{ij} \) be the contribution of \( A_iA_j \) and \( z_{kl} \) the contribution of \( B_kB_l \) to the character under consideration. Under the
Table 4. Effect of mixed random mating and selfing on the genotypic variance relative to the initial variance \((V_n/V_0)\) in the case of complete dominance

<table>
<thead>
<tr>
<th>(p_2)</th>
<th>(n)</th>
<th>(\beta=0.10)</th>
<th>(0.20)</th>
<th>(0.30)</th>
<th>(0.50)</th>
<th>(0.80)</th>
<th>(0.90)</th>
<th>(1.00)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1</td>
<td>1.190</td>
<td>1.377</td>
<td>1.560</td>
<td>1.917</td>
<td>2.427</td>
<td>2.590</td>
<td>2.750</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.199</td>
<td>1.414</td>
<td>1.641</td>
<td>2.133</td>
<td>2.938</td>
<td>3.218</td>
<td>3.500</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.200</td>
<td>1.417</td>
<td>1.654</td>
<td>2.186</td>
<td>3.132</td>
<td>3.482</td>
<td>3.844</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1.032</td>
<td>1.063</td>
<td>1.092</td>
<td>1.145</td>
<td>1.213</td>
<td>1.232</td>
<td>1.250</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.034</td>
<td>1.069</td>
<td>1.105</td>
<td>1.176</td>
<td>1.267</td>
<td>1.293</td>
<td>1.312</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.034</td>
<td>1.070</td>
<td>1.107</td>
<td>1.183</td>
<td>1.286</td>
<td>1.311</td>
<td>1.328</td>
</tr>
<tr>
<td>0.6</td>
<td>1</td>
<td>1.014</td>
<td>1.027</td>
<td>1.038</td>
<td>1.057</td>
<td>1.077</td>
<td>1.081</td>
<td>1.083</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.015</td>
<td>1.029</td>
<td>1.043</td>
<td>1.067</td>
<td>1.085</td>
<td>1.084</td>
<td>1.078</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.015</td>
<td>1.029</td>
<td>1.044</td>
<td>1.069</td>
<td>1.085</td>
<td>1.079</td>
<td>1.064</td>
</tr>
<tr>
<td>0.7</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>.999</td>
<td>.993</td>
<td>.978</td>
<td>.972</td>
<td>.964</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.000</td>
<td>.999</td>
<td>.998</td>
<td>.988</td>
<td>.954</td>
<td>.936</td>
<td>.913</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.000</td>
<td>.999</td>
<td>.997</td>
<td>.986</td>
<td>.942</td>
<td>.915</td>
<td>.880</td>
</tr>
<tr>
<td>0.8</td>
<td>1</td>
<td>.990</td>
<td>.979</td>
<td>.968</td>
<td>.944</td>
<td>.904</td>
<td>.890</td>
<td>.875</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.989</td>
<td>.977</td>
<td>.963</td>
<td>.928</td>
<td>.856</td>
<td>.826</td>
<td>.792</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.989</td>
<td>.977</td>
<td>.962</td>
<td>.924</td>
<td>.835</td>
<td>.794</td>
<td>.745</td>
</tr>
<tr>
<td>0.9</td>
<td>1</td>
<td>.982</td>
<td>.963</td>
<td>.944</td>
<td>.906</td>
<td>.846</td>
<td>.826</td>
<td>.806</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.981</td>
<td>.959</td>
<td>.936</td>
<td>.882</td>
<td>.780</td>
<td>.741</td>
<td>.698</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.981</td>
<td>.959</td>
<td>.935</td>
<td>.875</td>
<td>.753</td>
<td>.701</td>
<td>.692</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>.981</td>
<td>.959</td>
<td>.934</td>
<td>.873</td>
<td>.735</td>
<td>.668</td>
<td>.585</td>
</tr>
</tbody>
</table>
assumption of additive effects of the two loci, the con-
tribution by the genotype \( A_i A_j B_k B_\ell \) is \( y_{ij} + z_{k\ell} \). The mean
of the population in generation \( n \) is given by \( \mu_n(Y+Z) \) where

\[
\mu_n(Y+Z) = \sum_{i=1}^{s_1} \sum_{j=1}^{s_1} \sum_{k=1}^{s_2} \sum_{\ell=1}^{s_2} p_{ijkl}^{(n)} (y_{ij} + z_{k\ell})
\]

\[
= \mu_n(Y) + \mu_n(Z)
\]  

(2.103)

where \( \mu_n(Y) \) and \( \mu_n(Z) \) are the means of the marginal dis-
tributions for each individual locus.

The genotypic variance, \( V_n(Y+Z) \), in generation \( n \), is
given by

\[
V_n(Y+Z) = \sum_{i=1}^{s_1} \sum_{j=1}^{s_1} \sum_{k=1}^{s_2} \sum_{\ell=1}^{s_2} p_{ijkl}^{(n)} (y_{ij} + z_{k\ell})^2 - \mu_n^2(Y+Z)
\]

\[
= V_n(Y) + V_n(Z) + 2C_n(Y,Z)
\]  

(2.104)

where

\[
C_n(Y,Z) = \sum_{ij} \sum_{k\ell} p_{ijkl}^{(n)} y_{ij} z_{k\ell} - \mu_n(Y)\mu_n(Z)
\]

which is actually the covariance of the \( Y \) and \( Z \) values.

This covariance term arises because as we have seen earlier
in Section A, the joint distribution of genotypes with
respect to two loci is not equal to the product of the two
marginals even in the limiting case, although the two loci
are taken to be unlinked. In the discussion that follows
we restrict ourselves to an initial population which has
the structure of a random mating population in equilibrium
that is, \( p_{ijk\ell}^{(0)} = p_i p_j q_k q_{\ell} \), \( i,j=1,2,\ldots,s_1; k,\ell=1,2,\ldots,s_2 \).

Therefore, using Equations 2.45 - 2.47, we have

\[
C_n(Y,Z) = \sum_{i,j,k,l} \left[ \sum_{\ell} \left( p_{ijk\ell}^{(n)} - p_{ij}^{(n)} \right) \right] y_{ij} z_{k\ell}
\]

\[
= \sum_{i,k} \left[ p_{iik\ell}^{(n)} - p_{iik}^{(n)} \right] y_{ii} z_{kk}
\]

\[
+ \sum_{i,j\neq i} \left[ \sum_{k} \left( p_{ijk\ell}^{(n)} - p_{ij}^{(n)} \right) \right] y_{ij} z_{kk}
\]

\[
+ \sum_{i,k} \left[ \left( p_{iik\ell}^{(n)} - p_{iik}^{(n)} \right) \right] y_{ii} z_{kl}
\]

\[
+ \sum_{i,j\neq i} \left[ \sum_{k,l\neq k} \left( p_{ijk\ell}^{(n)} - p_{ij}^{(n)} \right) \right] y_{ij} z_{kl}
\]

\[
= \eta_n \left[ \sum_{i,k} \left( p_i(1-p_i)q_k(1-q_k) \right) y_{ii} z_{kk} \right]
\]

\[
- \sum_{i,j\neq i} \left[ p_ip_j q_k (1-q_k) y_{ij} z_{kk} \right]
\]

\[
- \sum_{i,k} \left[ p_i(1-p_i)q_k q_{\ell} y_{ii} z_{kk} \right]
\]

\[
+ \sum_{i,j\neq i} \left[ \sum_{k,l\neq k} p_i p_j q_k q_{\ell} y_{ij} z_{kk} \right]
\]

\[
= \eta_n \left[ \sum_{i,k} \left( p_i q_k (1-q_k) y_{ii} z_{kk} \right) \right]
\]

\[
- \sum_{i,j,k} \left[ p_i p_j q_k (1-q_k) y_{ij} z_{kk} \right]
\]
- \sum_{i} \sum_{j} \sum_{k} \sum_{\ell} p_i p_j q_k q_\ell y_{ij} z_{k\ell}

+ \sum_{i} \sum_{j} \sum_{k} \sum_{\ell} p_i p_j q_k q_\ell y_{ij} z_{k\ell}

= \eta_n \left[ \sum_{i} \sum_{j} \sum_{k} \sum_{\ell} p_i p_j q_k q_\ell y_{ij} z_{k\ell} \right]

- \sum_{i} \sum_{j} \sum_{k} p_i p_j y_{ij} z_{kk} - \sum_{i} \sum_{k} p_i q_k y_{ii} z_{k\ell}

+ \sum_{i} \sum_{j} \sum_{k} p_i q_k y_{ii} z_{kk}

= \eta_n \left[ \sum_{i} \sum_{j} p_i p_j y_{ij} - \sum_{i} p_i y_{ii} \right] \left[ \sum_{k} \sum_{\ell} q_k q_\ell z_{k\ell} - \sum \sum_{k} q_k z_{kk} \right]

= \eta_n \left[ \sum_{i} \sum_{j} p_i p_j \left( y_{ij} - \frac{y_{ii} + y_{jj}}{2} \right) \right]

\times \left[ \sum_{k} \sum_{\ell} q_k q_\ell \left( z_{k\ell} - \frac{z_{kk} + z_{\ell \ell}}{2} \right) \right]

= \frac{1}{4} \eta_n \left[ \sum_{i} \sum_{j} p_i p_j \left( 2y_{ij} - y_{ii} - y_{jj} \right) \right]

\times \left[ \sum_{k} \sum_{\ell} q_k q_\ell \left( 2z_{k\ell} - z_{kk} - z_{\ell \ell} \right) \right]

(2.105)

where \( \eta_n = a_2^{(n)} - [a_1^{(n)}]^2 \) and is nonnegative. This shows that \( \text{Cov}_n(Y, Z) \) depends upon the dominance deviations of the two loci, and also on \( \eta_n \) which is a nonnegative quantity. \( \eta_n \) is zero only for random mating (\( \beta=0 \)) or pure selfing (\( \beta=1 \)).

It is obvious from the expressions for the genotypic
mean and variance that the mean value due to two loci segregating independently in a population undergoing mixed selfing and random mating is equal to the sum of the means due to each locus separately, whereas in the case of variance in addition to the sum of variances due to each factor separately, there appears another term which vanishes in the case of complete random mating, complete selfing or in the absence of dominance. Thus the usual assumption that the variance due to independently segregating loci is equal to the sum of variances due to each locus separately is subject to limitations, that is, it holds good only when the population is practicing only one system of breeding or when there is no dominance at least in one locus. It is, therefore, evident that trend of changes in the genotypic variance due to several factors, under the given mating system of mixed selfing and random mating, cannot be predicted from the behavior of a single factor because of the occurrence of an additional term in the formula for variance.

In the absence of dominance

\[ Y_{ij} = \frac{Y_{ii} + Y_{jj}}{2} \quad \text{and} \quad z_{k\ell} = \frac{z_{kk} + z_{\ell\ell}}{2}, \]

then we shall have

\[ \mu_n(Y + Z) = \mu_n(Y) + \mu_n(Z) \]
\[ V_n(Y+Z) = V_n(Y) + V_n(Z) \]
\[ = (1 + F_n)[V_0(Y) + V_0(Z)] \]
\[ = \frac{1 - (\beta/2)^{n+1}}{1 - \beta/2} V_0(Y+Z) \]  
(2.107)

where \( V_0(Y+Z) = V_0(Y) + V_0(Z) = V_R(Y) + V_R(Z). \)

In the limit as \( n \to \infty \)

\[ V_\infty(Y+Z) = \frac{2}{2 - \beta} [V_R(Y) + V_R(Z)] \]  
(2.108)

Therefore, in the absence of dominance, the mean is unaffected but the variance, for a given \( \beta \), increases monotonically with \( n \) until it reaches its limiting value \( V_\infty \) given in Equation 2.108. It also increases with \( \beta \) for a given \( n \).

3. **Multiple loci**

The above results now can be easily extended to cover multiple loci situations (\( k \geq 2 \)) when the initial population is taken to be panmictic. To derive the formula for the genotypic mean and variance of a population under the assumption that the effects of loci are additive, we need the marginal distributions of individual loci and the marginal
distributions in pairs. This will yield, in the case of \( k \) unlinked loci,

\[
\mu_n(Y+Z+\ldots) = \sum \mu_n(Y)
\]  \hspace{1cm} (2.109)

\[
V_n(Y+Z+\ldots) = \sum V_n(Y) + 2\Sigma'\text{Cov}_n(Y,Z)
\]  \hspace{1cm} (2.110)

where \( \Sigma \) denotes the summation over all loci and \( \Sigma' \) is the summation over all possible pairs. Formulas for \( V_n(Y) \) and \( \text{Cov}_n(Y,Z) \) for a given pair are obtained in Equations 2.98 and 2.105.

It may be of interest to examine the relative contribution of the second term in Equation 2.110 to the total genotypic variance. This depends upon the dominance deviations, the gene frequencies, the number of generations, the probability of selfing and the number of factors. Ghai (1964) studied this problem with respect to two alleles at each locus under the assumptions of (a) complete dominance at each locus, (b) equal and additive effects of different factors and (c) equal frequency of all recessive alleles, say \( p_2 \). He has observed that the relative contribution is higher for intermediate gene frequencies and also for intermediate proportions of selfing than for extreme values. In equilibrium, the maximum relative contribution for \( k=6 \) is of the order of 18 percent of the total genotypic variance. This value is attained when \( p_2 = 0.3 \) and \( \beta = .60 \), approximately.
III. MIXED RANDOM AND SIB MATING

We have in the previous chapter developed models for the system of mixed random mating and selfing. Selfing is the most intensive inbreeding in plant populations whereas in animals the most intensive inbreeding is full-sib mating. Further important inbreeding systems are parent-offspring, half-sib mating and so on. We shall presently develop theoretical models for partial inbreeding which include mixed random and (a) full-sib mating, and (b) half-sib mating, assuming that a single locus is segregating.

A. Mixed Random and Full-Sib Mating

Let a proportion \( \beta (0 \leq \beta \leq 1) \) of a population be mated to full-sibs and the remaining proportion \( \gamma = (1 - \beta) \) mated at random in each generation. This model seems to have been first examined by Haldane (1924) who derived its equilibrium distribution. This was later discussed by Page and Hayman (1960). Wright (1951) discussed this model for a finite population using the method of path coefficients. More recently Ghai (1969) has discussed the dynamics of a population and its equilibrium behavior and the effect of such a mating system on various other characteristics of the population like its mean and variance. The model is developed under the same assumptions as in Chapter II, namely, the population is considered to be large enough so that
sampling fluctuations may be ignored and there are no viability or fertility differences. The mathematical analysis is similar to that of the previous chapter.

As the mating system involves sib mating the distribution of genotypes in one generation alone does not help to determine the distribution in the next generation. It is the frequencies of the various types of mating that determine the composition of the population in the next generation. With respect to a single locus with two alleles, say $A_1$ and $A_2$, let the initial frequencies (indexed by $-1$) of six possible types of mating associated with three genotypes $A_1A_1$, $A_1A_2$, and $A_2A_2$ be

<table>
<thead>
<tr>
<th>Mating type</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1 \times A_1A_1$</td>
<td>$f_1^{(-1)} = p_1^4$</td>
</tr>
<tr>
<td>$A_2A_2 \times A_2A_2$</td>
<td>$f_2^{(-1)} = p_2^4$</td>
</tr>
<tr>
<td>$A_1A_1 \times A_2A_2$</td>
<td>$f_3^{(-1)} = 2p_1^2p_2^2$</td>
</tr>
<tr>
<td>$A_1A_1 \times A_1A_2$</td>
<td>$f_4^{(-1)} = 4p_1^3p_2$</td>
</tr>
<tr>
<td>$A_2A_2 \times A_1A_2$</td>
<td>$f_5^{(-1)} = 4p_1p_2^3$</td>
</tr>
<tr>
<td>$A_1A_2 \times A_1A_2$</td>
<td>$f_6^{(-1)} = 4p_1^2p_2^2$</td>
</tr>
</tbody>
</table>

(3.1)

giving the initial population (generation 0) as
\[ p_1^2 A_1 A_1 + 2p_1 p_2 A_1 A_2 + p_2^2 A_2 A_2 \]  

(3.2)

where \( p_1 \) and \( p_2, p_1 + p_2 = 1, \) are frequencies of the two alleles \( A_1 \) and \( A_2, \) respectively.

Let \( f_1^{(n)}, f_2^{(n)}, ..., f_6^{(n)} \) be the corresponding frequencies of the six types of mating in the \( n \)th generation. The relations between the frequencies of these mating types in two successive generations of the mixed mating system may be stated in the form of the following difference equations:

\[
\begin{align*}
    f_1^{(n+1)} &= \gamma [p_{11}^{(n)}]^2 + \beta [f_1^{(n)} + (1/4)f_4^{(n)} + (1/16)f_6^{(n)}] \\
    f_2^{(n+1)} &= \gamma [p_{22}^{(n)}]^2 + \beta [f_2^{(n)} + (1/4)f_5^{(n)} + (1/16)f_6^{(n)}] \\
    f_3^{(n+1)} &= 2\gamma p_{11}^{(n)} p_{22}^{(n)} + (\beta/8)f_6^{(n)} \\
    f_4^{(n+1)} &= 4\gamma p_{11}^{(n)} p_{12}^{(n)} + \beta [(1/2)f_4^{(n)} + (1/4)f_6^{(n)}] \\
    f_5^{(n+1)} &= 4\gamma p_{22}^{(n)} p_{12}^{(n)} + \beta [(1/2)f_5^{(n)} + (1/4)f_6^{(n)}] \\
    f_6^{(n+1)} &= 4\gamma [p_{12}^{(n)}]^2 + \beta [f_3^{(n)} + (1/4)f_4^{(n)} + (1/4)f_5^{(n)}} \\
    &\quad \quad \quad \quad \quad \quad + (1/4)f_6^{(n)}] .
\end{align*}
\]

(3.3)

It is not difficult to see that the frequency, \( H_{n+1}, \) of the \( A_1 A_2 \) genotype in the \( (n+1) \)th generation can be written as
\[ H_{n+1} = 2p_1^{(n+1)} = f_3^{(n)} + (1/2)f_4^{(n)} + (1/2)f_5^{(n)} + (1/2)f_6^{(n)} \]  

(3.4)

and similarly for the frequencies of the other genotypes.

It can be verified directly then that \( p_1^{(n+1)} = p_1^{(n+1)} + p_{12}^{(n+1)} = p_1^{(n)} \) which implies that gene frequencies remain unchanged.

Substituting the values of \( f \)'s in the right hand side of Equation 3.4 from Equations 3.3, we get in particular

\[ H_1 = [1 - (\beta/4)]H_0 \]  

(3.5)

\[ H_2 = [1 - (\beta/2)][1 + (\beta/4)]H_0 \]  

(3.6)

where \( H_0 = 2p_1p_2 \).

It is possible to follow the changes of \( H_n \) over successive generations by considering Equations 3.3 and 3.4. Thus referring to these equations we obtain, after some simplifications

\[ H_{n+3} - (1 + \beta/2)H_{n+2} + (\beta/4)H_{n+1} + (\beta/4)H_n = 0 \]  

(3.7)

The solution of the difference Equation 3.7 can be written down as

\[ H_n = A \lambda_1^n + B \lambda_2^n + C \lambda_3^n \]

where \( \lambda_1, \lambda_2, \) and \( \lambda_3 \) are the roots of the characteristic equation

\[ \lambda^3 - (1 + \beta/2)\lambda^2 + (\beta/4)\lambda + (\beta/4) = 0 \]  

(3.8)
associated with the difference Equation 3.7, and A, B, and C are constants to be determined by the initial conditions.

The roots of the characteristic Equation 3.8 are

\[
\lambda_1 = \frac{\beta + (\beta^2 + 4\beta)^{1/2}}{4},
\]

\[
\lambda_2 = \frac{\beta - (\beta^2 + 4\beta)^{1/2}}{4},
\]

\[
\lambda_3 = 1
\]

with \(\lambda_1\) and \(\lambda_2\) being less than unity in absolute value and \(\lambda_1 > \lambda_2\). Solving for A, B, and C from the initial conditions, we have

\[
A = \frac{(\beta/4)\lambda_1 H_0}{(1 - \lambda_1)(\lambda_1 - \lambda_2)}
\]

\[
B = \frac{-((\beta/4)\lambda_2 H_0}{(1 - \lambda_2)(\lambda_1 - \lambda_2)}
\]

and

\[
C = \frac{\gamma^H_0}{(1 - \lambda_1)(1 - \lambda_2)}.
\]

Hence,

\[
H_n = 2P_{12}^{(n)} = \left[ \frac{\gamma}{(1 - \lambda_1)(1 - \lambda_2)} + \frac{(\beta/4)\lambda_1^{n+1}}{(1 - \lambda_1)(\lambda_1 - \lambda_2)} \right]H_0
\]

\[= \frac{(\beta/4)\lambda_2^{n+1}}{(1 - \lambda_2)(\lambda_1 - \lambda_2)} \right]H_0
\]

(3.10)
which gives the expected proportion of genotype $A_1A_2$ after $n$ generations of mixed random and full-sib mating. The expected proportions of other genotypes $A_1A_1$ and $A_2A_2$ after $n$ generations can be derived from

$$P_{11}^{(n)} = P_1 - P_{12}^{(n)}$$ (3.11)

$$P_{22}^{(n)} = P_2 - P_{22}^{(n)}$$ (3.12)

In the limiting case

$$H_n = \lim_{n \to \infty} H_n = \frac{\gamma H_0}{(1 - \lambda_1)(1 - \lambda_2)}$$

$$= 2p_1p_2[1 - \frac{\beta}{4 - 3\beta}]$$ (3.13)

The approach of $H_n$ to its equilibrium value $H_\infty$ is governed by the largest characteristic root, $\lambda_1$, which is less than unity in absolute value. It is evident that the smaller the value of $\beta$, the more rapidly the population will reach equilibrium.

Using Equation 3.13 the equilibrium frequency of other genotypes can be easily determined.

$$P_{11}^{(\infty)} = P_1 - P_{12}^{(\infty)} = P_1 + \frac{\beta}{4 - 3\beta} p_1p_2$$ (3.14)

$$P_{22}^{(\infty)} = P_2 - P_{12}^{(\infty)} = P_2 + \frac{\beta}{4 - 3\beta} p_1p_2$$ (3.15)
where \( 2P_{12}^{(\infty)} = H_\infty \).

The above results in the case of a single locus can also be developed by using the concept of "identity by descent." This involves two indices, \( r_{XY} \), the "coefficient de parentè", or the coefficient of parentage as translated by Kempthorne, of two individuals \( X \) and \( Y \), and \( F_X \), the coefficient of inbreeding of an individual \( X \), the terms introduced by Malécot. We shall also denote these by \( r_n \) and \( F_n \) when the individual(s) belong to generation \( n \). These are defined as follows. Consider two individuals \( X \) and \( Y \) with genotypes \( (ab) \) and \( (cd) \). Then \( r_{XY} \) is defined to be the probability that a random gene from \( X \) is identical by descent with a random gene from \( Y \).

\[
\begin{align*}
    r_{XY} &= \frac{1}{4} \left[ P(a=c) + P(a=d) + P(b=c) + P(b=d) \right] \\
\end{align*}
\]

where \( P(a=c) \) denotes the probability that genes \( a \) and \( c \) are identical by descent.

The coefficient of inbreeding of an individual \( X = (ab) \) is defined to be the probability that the two genes possessed by that individual at a locus are identical by descent, i.e.,

\[
    F_X = P(a=b) \quad .
\]

These two indices are obviously related. Let \( X \) and \( Y \) be the parents of an individual \( Z \). It is sometimes
convenient to denote the offspring by $XxY$. Then the coefficient of inbreeding of the offspring $Z$ is given by (Kempthorne, 1957, p. 75)

$$F_Z = F_{XxY} = r_{XY}$$

(3.18)

which means that the coefficient of inbreeding of an offspring of the mating $XxY$ is equal to the coefficient of parentage $r_{XY}$ of its parents.

Using these concepts we can develop a formula for $r_{XY}$ or $F_X$ for the given system of mixed random mating and full-sibbing.

$$r_{XY} = \beta \left[ r_{XY} \mid F.S. \right]$$

(3.19)

where $r_{XY} \mid F.S.$ denotes the coefficient of parentage given $X$ and $Y$ are full-sibs. The recurrence formulas for $r_n$ and $F_n$ for a given system of inbreeding are well known (Kempthorne, 1957, Ch. 5). We shall frequently use these formulas here and in later chapters. Therefore, using Equation 3.19 in conjunction with Equation 3.18 we can write

$$F_{n+2} = \beta \left[ F_{n+2} \mid F.S. \right]$$

$$= (\beta/4) \left[ 1 + F_n + 2F_{n+1} \right]$$

(3.20)

This will lead to the difference Equation 3.7 in $H_n$ by using the transformation $H_n = (1 - F_n)H_0$. 

B. Mixed Random and Half-Sib Mating

Consider an infinite population where a fraction \( \beta (0 \leq \beta \leq 1) \) of the offspring mate with their half-sibs and the remaining fraction mates at random. We shall consider a system of half-sib mating as shown in Figure 2, where one male is mated with an indefinite number of his half-sisters who are also half-sisters of each other (Wright, 1969, p. 191).

\[
\begin{array}{c}
\text{Generation} \\
n-1 \\
n \\
n+1 \\
n+2 \\
\end{array}
\begin{array}{c}
\text{A} \\
\text{B} \\
\text{C} \\
\text{D} \\
\text{I} \\
\text{J} \\
\text{K} \\
\text{X} \\
\text{Y} \\
\text{Z} \\
\end{array}
\]

Figure 2. Half-sib inbreeding

Under half-sib inbreeding alone in successive generations, we have

\[
F_Z = F_{n+2}, \quad r_{n+1} = r_{XY} = r_{IJ}, r_{IK}
\]

Then
because I, J; I, K and J, K are all half-sibs. Therefore,

\[ r_{n+1} = (1/4)[1 + F_I + F_J + 3r_{IJ}] \]

Hence,

\[ r_{n+1} = (1/4)[1/2 + 3r_{IJ}] \]

This yields

\[ r_{XY} = \beta[r_{XY}|H.S.] \]  \hspace{1cm} (3.22)\]

This yields

\[ F_{n+2} = \beta[F_{n+2}|H.S.] \]

\[ = (3\beta/4)F_{n+1} + (\beta/8)F_n + (\beta/8) \]  \hspace{1cm} (3.23)\]

In the limit as \( n \to \infty \)
The genotypic distribution is completely determined once we know \( F \) provided we start with an initial population which has a random mating structure. The value of \( F \) in equilibrium is given explicitly by Equation 3.24. However, to discuss the consequences of this system of mating on the dynamics of the population and to specify the structure at any given time, say after \( n \) generations of mixed mating, we have to use Equation 3.23 along with the transformation

\[
H_n = (1 - F_n)H_0
\]

where \( H_0 = 2p_1p_2 \) is the frequency of the heterozygote \( A_1A_2 \) in the initial panmictic population for a single locus with two alleles \( A_1 \) and \( A_2 \). We can make the substitution \( H_n = (1 - F_n)H_0 \) in Equation 3.23 and write a recursion formula for heterozygosity. This yields

\[
H_{n+2} - \left( \frac{3\beta}{4} \right)H_{n+1} - \left( \frac{\beta}{8} \right)H_n - (1 - \beta)H_0 = 0 \quad (3.25)
\]

This difference equation can be readily solved yielding

\[
H_n - H_\infty = B \lambda_1^n + C \lambda_2^n \quad (3.26)
\]

where

\[
H_\infty = (1 - F_\infty)H_0 = \frac{16(1 - \beta)}{8 - 7\beta} p_1p_2 \quad (3.27)
\]

is the limiting value of \( H_n \), and \( \lambda_1 \) and \( \lambda_2 \) are the roots of the characteristic equation
\[ \lambda^2 - (3\beta/4)\lambda - (\beta/8) = 0 \]  

(3.28)

associated with the difference Equation 3.25. B and C in Equation 3.26 are constants which are determined from the initial conditions \( H_0 = 2p_1p_2 \) and \( H_1 = (1 - \beta/8)H_0 \).

The roots of the characteristic equation are

\[ \lambda_1 = \frac{1}{8}[3\beta + (9\beta^2 + 8\beta)^{1/2}] \]

\[ \lambda_2 = \frac{1}{8}[3\beta - (9\beta^2 + 8\beta)^{1/2}] \]  

(3.29)

and \(-1 < \lambda_2 < 0 < \lambda_1 < 1\) for \(0 < \beta \leq 1\). The approach of \( H_n \) to \( H_\infty \) is governed by the dominant root \( \lambda_1 \) i.e.,

\[ H_n - H_\infty \approx B\lambda_1^n \]  

(3.30)

when \( n \) is large.

Using the initial conditions we can determine the constants \( B \) and \( C \) and then we have the explicit solution for \( H_n \) given in Equation 3.31.

\[ H_n = 2p_{12}^{(n)} = \left[ (1 - F_\infty) + \frac{(\beta/8)\lambda_1^{n+1}}{(1 - \lambda_1)(\lambda_1 - \lambda_2)} \right. \]

\[ \left. - \frac{(\beta/8)\lambda_2^{n+1}}{(1 - \lambda_2)(\lambda_1 - \lambda_2)} \right] H_0 \]  

(3.31)

This gives the expected frequency of \( \lambda_1\lambda_2 \) genotype after \( n \) generations of mixed random and half-sib mating. \( H_\infty \) given by Equation 3.27 is the expected proportion of heterozygotes.
that will be present in the population in equilibrium state. The frequencies of other genotypes in generation $n$ can be easily derived to give $p_{11}^{(n)} = p_1 - p_{12}^{(n)}$ and $p_{22}^{(n)} = p_2 - p_{12}^{(n)}$ because the gene frequencies remain invariant over time.

In equilibrium these will reduce to

$$\lim_{n \to \infty} p_{11}^{(n)} = p_{11}^{(\infty)} = p_1 - p_{12}^{(\infty)}$$

(3.32)

and

$$\lim_{n \to \infty} p_{22}^{(n)} = p_{22}^{(\infty)} = p_2 - p_{12}^{(\infty)}$$

(3.33)

where $2p_{12}^{(\infty)} = H_{\infty}$.

C. Loss of Heterozygosity

Under the mating systems considered two opposing forces are acting on the population, inbreeding acting to reduce the heterozygosity and random mating acting to restore the initial value. Ultimately a stage will be reached where the loss due to one is balanced by the gain due to other and the population eventually reaches an equilibrium state which has a certain amount of heterozygosis depending upon the amounts of inbreeding in the population. The equilibrium will, however, be reached asymptotically.

Defining as before the coefficient of inbreeding, $F_n$, in the $n$th generation as the loss in heterozygosity relative to that in the initial population, we have $F_n = 1 - H_n / H_0$. 
Thus under the system of mixed random and full-sib mating, after simplification, we have from Equation 3.10

\[ F_n = \frac{(\beta/4)}{\lambda_1 - \lambda_2} \left[ \frac{1}{1-\lambda_1} \frac{\lambda_1^{n+1}}{\lambda_2^{n+1}} - \frac{1}{1-\lambda_2} \frac{\lambda_2^{n+1}}{\lambda_1^{n+1}} \right] \]  

(3.34)

where \( \lambda_1 \) and \( \lambda_2 \) are given by Equations 3.9 which in the limit as \( n \to \infty \) reduces to

\[ F_\infty = 1 - \frac{\gamma}{(1-\lambda_1)(1-\lambda_2)} \]

\[ = \frac{\beta}{4 - 3\beta} \]  

(3.35)

Similarly, under the process of mixed random and half-sib mating

\[ F_n = F_\infty - \frac{(\beta/8)}{\lambda_1 - \lambda_2} \left[ \frac{\lambda_1^{n+1}}{1-\lambda_1} - \frac{\lambda_2^{n+1}}{1-\lambda_2} \right] \]  

(3.36)

\[ F_\infty = \frac{\beta}{8 - 7\beta} \]  

(3.37)

where \( \lambda_1 \) and \( \lambda_2 \) in this case are given by Equation 3.29.

We now give an alternative expression for \( F_n \) given by Equation 3.36 which may be more convenient.

\[ F_n = F_\infty + \frac{(\beta/8)}{\lambda_1 - \lambda_2} \left[ \frac{1 - \lambda_1^{n+1}}{1-\lambda_1} - \frac{1 - \lambda_2^{n+1}}{1-\lambda_2} \right] \]

\[ = F_\infty + \frac{(\beta/8)}{\lambda_1 - \lambda_2} \left[ \frac{1 - \lambda_1^{n+1}}{1-\lambda_1} - \frac{1 - \lambda_2^{n+1}}{1-\lambda_2} \right] \]
\[
- \frac{(\beta/\theta)}{\lambda_1 - \lambda_2} \left[ \frac{1}{1 - \lambda_1} - \frac{1}{1 - \lambda_2} \right]
\]

\[
= \frac{(\beta/\theta)}{\lambda_1 - \lambda_2} \left[ \frac{1 - \lambda_1^{n+1}}{1 - \lambda_1} - \frac{1 - \lambda_2^{n+1}}{1 - \lambda_2} \right]
\]

\[
= \frac{(\beta/\theta)}{\lambda_1 - \lambda_2} \left[ (\lambda_1 + \lambda_1^2 + \ldots + \lambda_1^n) - (\lambda_2 + \lambda_2^2 + \ldots + \lambda_2^n) \right]
\]

(3.38)

These results show that under the two systems of mating heterozygosity is not completely lost except for \( \beta = 1 \) i.e., complete inbreeding.

The loss in heterozygosity in successive generations up to 20 generations and in the limiting case is examined in Tables 5 and 6 for the two systems of mating for various values of \( \beta \). The values for \( \beta = 1.0 \) correspond to the case of pure inbreeding by the respective processes. The values for \( n = 0 \) correspond to the maximum loss in heterozygosity that can occur under such mixed mating systems. It may be seen that loss in heterozygosity increases with increasing amounts of inbreeding as it should, but is not proportionate to the increase in the amount of inbreeding for \( n > 1 \). When \( n = 1 \), the increase is linear in \( \beta \). There is practically no loss in heterozygosity when the system of mating deviates from complete random mating and the population in successive generations includes individuals obtained
Table 5. Loss of heterozygosity relative to the initial heterozygosity ($F_n$) under mixed random and full-sib mating

<table>
<thead>
<tr>
<th>n</th>
<th>$\beta = .10$</th>
<th>.20</th>
<th>.30</th>
<th>.40</th>
<th>.50</th>
<th>.60</th>
<th>.70</th>
<th>.80</th>
<th>.90</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.025</td>
<td>.050</td>
<td>.075</td>
<td>.100</td>
<td>.125</td>
<td>.150</td>
<td>.175</td>
<td>.200</td>
<td>.225</td>
<td>.250</td>
</tr>
<tr>
<td>2</td>
<td>.026</td>
<td>.055</td>
<td>.086</td>
<td>.120</td>
<td>.156</td>
<td>.195</td>
<td>.236</td>
<td>.280</td>
<td>.326</td>
<td>.375</td>
</tr>
<tr>
<td>3</td>
<td>.027</td>
<td>.058</td>
<td>.094</td>
<td>.134</td>
<td>.180</td>
<td>.231</td>
<td>.288</td>
<td>.352</td>
<td>.422</td>
<td>.500</td>
</tr>
<tr>
<td>4</td>
<td>.059</td>
<td>.095</td>
<td>.139</td>
<td>.189</td>
<td>.249</td>
<td>.317</td>
<td>.397</td>
<td>.488</td>
<td>.594</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.096</td>
<td>.141</td>
<td>.195</td>
<td>.259</td>
<td>.336</td>
<td>.429</td>
<td>.540</td>
<td>.672</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>.143</td>
<td>.200</td>
<td>.272</td>
<td>.365</td>
<td>.489</td>
<td>.656</td>
<td>.886</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.273</td>
<td>.368</td>
<td>.498</td>
<td>.683</td>
<td>.961</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.273</td>
<td>.368</td>
<td>.500</td>
<td>.690</td>
<td>.986</td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>.027</td>
<td>.059</td>
<td>.097</td>
<td>.143</td>
<td>.200</td>
<td>.273</td>
<td>.368</td>
<td>.500</td>
<td>.692</td>
</tr>
</tbody>
</table>
Table 6. Loss of heterozygosity relative to the initial heterozygosity ($F_n$) under mixed random and half-sib mating

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\beta = .10$</th>
<th>.20</th>
<th>.30</th>
<th>.40</th>
<th>.50</th>
<th>.60</th>
<th>.70</th>
<th>.80</th>
<th>.90</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.012</td>
<td>.025</td>
<td>.037</td>
<td>.050</td>
<td>.062</td>
<td>.075</td>
<td>.087</td>
<td>.100</td>
<td>.112</td>
<td>.125</td>
</tr>
<tr>
<td>2</td>
<td>.013</td>
<td>.029</td>
<td>.046</td>
<td>.065</td>
<td>.086</td>
<td>.109</td>
<td>.133</td>
<td>.160</td>
<td>.188</td>
<td>.219</td>
</tr>
<tr>
<td>3</td>
<td>.014</td>
<td>.030</td>
<td>.049</td>
<td>.072</td>
<td>.099</td>
<td>.130</td>
<td>.165</td>
<td>.206</td>
<td>.252</td>
<td>.305</td>
</tr>
<tr>
<td>4</td>
<td>.030</td>
<td>.050</td>
<td>.075</td>
<td>.105</td>
<td>.141</td>
<td>.186</td>
<td>.240</td>
<td>.304</td>
<td>.381</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.051</td>
<td>.076</td>
<td>.108</td>
<td>.148</td>
<td>.200</td>
<td>.264</td>
<td>.346</td>
<td>.449</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>.077</td>
<td>.111</td>
<td>.157</td>
<td>.224</td>
<td>.318</td>
<td>.464</td>
<td>.691</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>.158</td>
<td>.225</td>
<td>.330</td>
<td>.506</td>
<td>.827</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>.158</td>
<td>.226</td>
<td>.333</td>
<td>.521</td>
<td>.903</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\infty$</td>
<td>.014</td>
<td>.030</td>
<td>.051</td>
<td>.077</td>
<td>.111</td>
<td>.158</td>
<td>.226</td>
<td>.333</td>
<td>.529</td>
<td>1.000</td>
</tr>
</tbody>
</table>


by inbreeding up to $\beta = 0.20$. Maximum loss in heterozygosity in this case is of the order of 6 percent if inbreeding is by full-sib mating, and 3 percent if there is half-sib mating. As the amount of inbreeding in the population increases the maximum loss in heterozygosity also rises and in the case of 90 percent full-sib mating and 10 percent random mating it is of the order of 69 percent. The corresponding value for a mixture of half-sib and random mating is 53 percent. Thus, in a population mating at random if there is also inbreeding in small amounts, its effect on the population will be negligible, whereas in the case of inbred populations the effect of even small amounts of random mating in the population will be substantial.

D. Genotypic Mean and Variance

If we let $y_{11}$, $y_{12}$, $y_{22}$ be the genotypic values of the three genotypes $A_1A_1$, $A_1A_2$, and $A_2A_2$ respectively, the results can be put in a simple form in terms of $y_{11} - y_{22}$ and $y_{11} - 2y_{12} + y_{22}$. We, therefore, use coded genotypic values as follows:
The genotypic mean ($\mu_n$) and variance ($V_n$) in generation $n$ are given by

$$
\mu_n = p_1u - \frac{1}{2}H_nv 
$$

(3.39)

$$
V_n = (4p_1p_2 - H_n)^{\frac{u^2}{4}} + \frac{p_1 - p_2}{2}H_nuv + H_n(1-H_n)\frac{v^2}{4}
$$

(3.40)

with

$$
\mu_0 = p_1u - p_1p_2v
$$

and

$$
V_0 = \frac{P_1P_2}{2}[u^2 + 2(p_1 - p_2)uv + (1 - 2p_1p_2)v^2]
$$
These are general formulas for the genotypic mean and variance for any population which is completely specified by two parameters $H_n$ and $p_1$ ($p_2 = 1 - p_1$) in the case of two alleles at a locus. The value of $H_n$ to be used in the above formulas depends upon the system of mating. This is given by Equation 3.10 for the system of mixed random and full-sib mating and by Equation 3.31 in the case of mixed random and half-sib mating. The effects of the mating systems considered on the first two moments of the population are discussed in detail for two cases (1) absence of dominance, and (2) complete dominance.

1. Absence of dominance ($\nu = 0$)

When the average effect of the heterozygote is midway between the two parents, the mean does not change in successive generations and retains its initial value, that is,

$$\mu_n = p_1u = \mu_0 \quad .$$

The genotypic variance in this case reduces to the known form

$$V_n = \frac{(4p_1p_2 - H_n)u^2}{4}$$

$$= (1 + F_n)V_0 \quad (3.42)$$

where $V_0$ ($= \frac{p_1p_2}{2} \cdot u^2$) is the variance in the initial population. The values of $F_n$ are given in Tables 5 and 6 for
the two systems of mating. Since $F_n$ is an increasing function of $n$, genotypic variance also increases with $n$ for a given $\beta$. The genotypic variance also increases with the increasing amounts of inbreeding, the increase being relatively higher for higher amounts of inbreeding in the population. Since $F_n$ is a bounded function, $V_n \to V_\infty$ as $n \to \infty$

where

$$V_\infty = \frac{2(2-\beta)}{4-3\beta} V_0$$

(3.43)

if the partial inbreeding is by full-sibbing and

$$V_\infty = \frac{2(4-3\beta)}{8-7\beta} V_0$$

(3.44)

if the partial inbreeding is due to half-sib mating.

2. Complete dominance ($v = -u$)

When allele $A_1$ is completely dominant over allele $A_2$, Equations 3.39 and 3.40 giving the mean and variance respectively reduce to

$$\mu_n = (p_1 + \frac{1}{2} H_n)u$$

(3.45)

$$V_n = \frac{1}{4} \left[ 4p_1p_2 - 2(p_1 - p_2)H_n - H_n^2 \right] u^2$$

$$= V_0 + \frac{1}{4} p_1p_2F_n \left[ 1 - 2p_2^2 - p_1p_2F_n \right] u^2$$

(3.46)

with $V_0 = p_2^2(1-p_2^2)u^2$. 
Now

\[ \mu_{n+1} - \mu_n = -\frac{1}{2}(H_n - H_{n+1})u < 0 \]  \hspace{1cm} (3.47)

because \( \{H_n\} \) is a decreasing sequence of \( n \) for all \( \beta \), \( 0 < \beta < 1 \). This shows that mean decreases in successive generations for all values of \( p_2 \) and \( \beta \). Such a result, of course, holds for any mating system involving inbreeding. The only relevant thing in such situations is the rate at which such a decrease takes place. We have already discussed the behavior of \( H_n \) or alternatively of \( F_n \) in Section C of this chapter for the two mating systems.

The comparison of variance in two successive generation shows that

\[ V_{n+1} - V_n = (H_n - H_{n+1})[2(p_1 - p_2) + H_n + H_{n+1}]\frac{u^2}{4}. \]  \hspace{1cm} (3.48)

When \( p_2 < 0.5 \), \( V_{n+1} - V_n > 0 \) for all \( \beta \), \( 0 < \beta < 1 \). For different values of the gene frequency, \( p_2 \), and the amount of inbreeding, \( \beta \), the variance \( (V_n) \) relative to the initial variance \( (V_0) \), for the two mating systems, is given in Tables 7 and 8 for the first three generations and in the limiting case. It may be seen that the limiting values are approached very rapidly within a few generations.

For values of the frequency of the recessive allele less than or equal to one-half \( (p_2 \leq 0.5) \), the variance increases in successive generations and also with increasing
amount of inbreeding. For $p_2 > 0.7$, the variance decreases. For $p_2 = 0.6$ and 0.7, there is an increase in the variance up to a certain stage and decrease thereafter. This is true both with respect to $n$ and $\beta$ ($0 < \beta \leq 1$).

This shows that in the case of complete dominance, the variance increases with $n$ as well as $\beta$ ($0 < \beta \leq 1$) when the frequency of recessive individuals is less than or equal to one quarter ($p_2^2 \leq 1/4$) in the initial population but it decreases when it is at least as high as that of the dominants ($p_2^2 \geq 1/2$).
Table 7. Effect of mixed random and full-sib mating on the genotypic variance relative to the initial variance \((V_n/V_0)\) in the case of complete dominance

<table>
<thead>
<tr>
<th>(p_2)</th>
<th>(n)</th>
<th>(\beta=.10)</th>
<th>(\beta=.20)</th>
<th>(\beta=.30)</th>
<th>(\beta=.50)</th>
<th>(\beta=.80)</th>
<th>(\beta=.90)</th>
<th>(\beta=1.00)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.2</td>
<td>1</td>
<td>1.095</td>
<td>1.190</td>
<td>1.284</td>
<td>1.469</td>
<td>1.740</td>
<td>1.829</td>
<td>1.917</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.100</td>
<td>1.209</td>
<td>1.326</td>
<td>1.583</td>
<td>2.021</td>
<td>2.180</td>
<td>2.344</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.103</td>
<td>1.220</td>
<td>1.353</td>
<td>1.667</td>
<td>2.267</td>
<td>2.500</td>
<td>2.750</td>
</tr>
<tr>
<td></td>
<td>(\infty)</td>
<td>1.103</td>
<td>1.224</td>
<td>1.365</td>
<td>1.740</td>
<td>2.750</td>
<td>3.334</td>
<td>4.167</td>
</tr>
<tr>
<td>.5</td>
<td>1</td>
<td>1.016</td>
<td>1.032</td>
<td>1.048</td>
<td>1.078</td>
<td>1.120</td>
<td>1.133</td>
<td>1.146</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.017</td>
<td>1.036</td>
<td>1.055</td>
<td>1.096</td>
<td>1.160</td>
<td>1.182</td>
<td>1.203</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.018</td>
<td>1.037</td>
<td>1.059</td>
<td>1.109</td>
<td>1.193</td>
<td>1.222</td>
<td>1.250</td>
</tr>
<tr>
<td></td>
<td>(\infty)</td>
<td>1.018</td>
<td>1.038</td>
<td>1.061</td>
<td>1.120</td>
<td>1.250</td>
<td>1.302</td>
<td>1.333</td>
</tr>
<tr>
<td>.6</td>
<td>1</td>
<td>1.007</td>
<td>1.014</td>
<td>1.020</td>
<td>1.033</td>
<td>1.048</td>
<td>1.053</td>
<td>1.057</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.007</td>
<td>1.015</td>
<td>1.023</td>
<td>1.040</td>
<td>1.062</td>
<td>1.068</td>
<td>1.074</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.008</td>
<td>1.016</td>
<td>1.025</td>
<td>1.044</td>
<td>1.072</td>
<td>1.079</td>
<td>1.083</td>
</tr>
<tr>
<td></td>
<td>(\infty)</td>
<td>1.008</td>
<td>1.016</td>
<td>1.026</td>
<td>1.048</td>
<td>1.083</td>
<td>1.082</td>
<td>1.042</td>
</tr>
<tr>
<td>.7</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>.999</td>
<td>.996</td>
<td>.995</td>
<td>.993</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>.998</td>
<td>.991</td>
<td>.987</td>
<td>.981</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>.997</td>
<td>.984</td>
<td>.975</td>
<td>.964</td>
</tr>
<tr>
<td></td>
<td>(\infty)</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>.996</td>
<td>.964</td>
<td>.927</td>
<td>.840</td>
</tr>
<tr>
<td>.8</td>
<td>1</td>
<td>.995</td>
<td>.990</td>
<td>.985</td>
<td>.974</td>
<td>.957</td>
<td>.951</td>
<td>.944</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.995</td>
<td>.989</td>
<td>.982</td>
<td>.967</td>
<td>.937</td>
<td>.925</td>
<td>.911</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.995</td>
<td>.988</td>
<td>.981</td>
<td>.961</td>
<td>.918</td>
<td>.898</td>
<td>.875</td>
</tr>
<tr>
<td></td>
<td>(\infty)</td>
<td>.995</td>
<td>.988</td>
<td>.980</td>
<td>.957</td>
<td>.875</td>
<td>.812</td>
<td>.694</td>
</tr>
<tr>
<td>.9</td>
<td>1</td>
<td>.991</td>
<td>.982</td>
<td>.972</td>
<td>.954</td>
<td>.925</td>
<td>.916</td>
<td>.906</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.990</td>
<td>.980</td>
<td>.968</td>
<td>.942</td>
<td>.894</td>
<td>.876</td>
<td>.856</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.990</td>
<td>.979</td>
<td>.966</td>
<td>.933</td>
<td>.866</td>
<td>.837</td>
<td>.805</td>
</tr>
<tr>
<td></td>
<td>(\infty)</td>
<td>.990</td>
<td>.978</td>
<td>.964</td>
<td>.925</td>
<td>.805</td>
<td>.724</td>
<td>.585</td>
</tr>
</tbody>
</table>
Table 8. Effect of mixed random and half-sib mating on the genotypic variance relative to the initial variance \((V_n/V_0)\) in the case of complete dominance

<table>
<thead>
<tr>
<th>(p_2)</th>
<th>(n)</th>
<th>(\beta=0.10)</th>
<th>(\beta=0.20)</th>
<th>(\beta=0.30)</th>
<th>(\beta=0.50)</th>
<th>(\beta=0.80)</th>
<th>(\beta=0.90)</th>
<th>(\beta=1.00)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.2)</td>
<td>1</td>
<td>1.048</td>
<td>1.095</td>
<td>1.143</td>
<td>1.237</td>
<td>1.377</td>
<td>1.423</td>
<td>1.469</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.051</td>
<td>1.110</td>
<td>1.175</td>
<td>1.324</td>
<td>1.596</td>
<td>1.699</td>
<td>1.807</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.052</td>
<td>1.114</td>
<td>1.187</td>
<td>1.372</td>
<td>1.761</td>
<td>1.925</td>
<td>2.106</td>
</tr>
<tr>
<td></td>
<td>(\infty)</td>
<td>1.052</td>
<td>1.115</td>
<td>1.193</td>
<td>1.418</td>
<td>2.204</td>
<td>2.842</td>
<td>4.167</td>
</tr>
<tr>
<td>(0.5)</td>
<td>1</td>
<td>1.008</td>
<td>1.016</td>
<td>1.024</td>
<td>1.040</td>
<td>1.063</td>
<td>1.071</td>
<td>1.078</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.009</td>
<td>1.019</td>
<td>1.030</td>
<td>1.055</td>
<td>1.098</td>
<td>1.114</td>
<td>1.130</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.009</td>
<td>1.020</td>
<td>1.032</td>
<td>1.062</td>
<td>1.123</td>
<td>1.147</td>
<td>1.172</td>
</tr>
<tr>
<td></td>
<td>(\infty)</td>
<td>1.009</td>
<td>1.020</td>
<td>1.033</td>
<td>1.070</td>
<td>1.185</td>
<td>1.259</td>
<td>1.333</td>
</tr>
<tr>
<td>(0.6)</td>
<td>1</td>
<td>1.004</td>
<td>1.007</td>
<td>1.010</td>
<td>1.017</td>
<td>1.027</td>
<td>1.030</td>
<td>1.032</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.004</td>
<td>1.008</td>
<td>1.013</td>
<td>1.023</td>
<td>1.040</td>
<td>1.046</td>
<td>1.052</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.004</td>
<td>1.008</td>
<td>1.014</td>
<td>1.026</td>
<td>1.049</td>
<td>1.058</td>
<td>1.066</td>
</tr>
<tr>
<td></td>
<td>(\infty)</td>
<td>1.004</td>
<td>1.009</td>
<td>1.014</td>
<td>1.029</td>
<td>1.069</td>
<td>1.084</td>
<td>1.042</td>
</tr>
<tr>
<td>(0.7)</td>
<td>1</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
<td>0.997</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.996</td>
<td>0.993</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td>(\infty)</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.986</td>
<td>0.959</td>
<td>0.840</td>
</tr>
<tr>
<td>(0.8)</td>
<td>1</td>
<td>0.997</td>
<td>0.995</td>
<td>0.992</td>
<td>0.987</td>
<td>0.979</td>
<td>0.977</td>
<td>0.974</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.997</td>
<td>0.994</td>
<td>0.991</td>
<td>0.982</td>
<td>0.966</td>
<td>0.959</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.997</td>
<td>0.994</td>
<td>0.990</td>
<td>0.980</td>
<td>0.955</td>
<td>0.944</td>
<td>0.930</td>
</tr>
<tr>
<td></td>
<td>(\infty)</td>
<td>0.977</td>
<td>0.994</td>
<td>0.990</td>
<td>0.977</td>
<td>0.923</td>
<td>0.866</td>
<td>0.694</td>
</tr>
<tr>
<td>(0.9)</td>
<td>1</td>
<td>0.995</td>
<td>0.991</td>
<td>0.986</td>
<td>0.977</td>
<td>0.963</td>
<td>0.958</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.995</td>
<td>0.989</td>
<td>0.983</td>
<td>0.968</td>
<td>0.941</td>
<td>0.930</td>
<td>0.918</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.995</td>
<td>0.989</td>
<td>0.982</td>
<td>0.964</td>
<td>0.923</td>
<td>0.905</td>
<td>0.884</td>
</tr>
<tr>
<td></td>
<td>(\infty)</td>
<td>0.995</td>
<td>0.989</td>
<td>0.981</td>
<td>0.959</td>
<td>0.873</td>
<td>0.793</td>
<td>0.585</td>
</tr>
</tbody>
</table>
IV. SOME OTHER MODELS

A. Mixed Random and Parent-Offspring Mating

A further mixed mating system is that of a parent-offspring and random mating, which corresponds fairly closely to the back-crossing system practiced in cattle breeding. In an infinite population with mating of offspring with younger parent in proportion \( \beta \), and random mating in proportion \( 1 - \beta \), we have as before

\[
F_{n+2} = \beta [F_{n+2} | P.O.]
\]

\[
= (\beta/4) [1 + F_n + 2F_{n+1}]
\]

and in the limit

\[
F_\infty = \lim_{n \to \infty} F_n = \frac{\beta}{4 - 3\beta}
\]

The recurrence relation 4.1 is identical with that for a system of mixed random and full-sib mating discussed in the previous chapter. Therefore, the system of mixed random and parent-offspring mating would yield the same results as obtained for the process of mixed random and full-sib mating which we have already discussed in the previous chapter.

The amount of heterozygosity that will be present in a population in equilibrium under such a system is given by Equation 3.13. Similar results have also been reported by Karlin (1968) by following the generation matrix approach.
B. Mixed Random and Double First Cousin Mating

Under this model with double first cousin mating in a fraction \( p \), \( 0 \leq p \leq 1 \), of the population while rest of the population mates at random, we have under the previous assumptions,

\[
F_{n+3} = \left(\frac{p}{8}\right)[1 + F_n + 2F_{n+1} + 4F_{n+2}] \tag{4.3}
\]

which gives at equilibrium

\[
F_\infty = \lim_{n \to \infty} F_n = \frac{\beta}{8 - 7\beta} \cdot \tag{4.4}
\]

From this we can easily derive the recurrence relation for \( H_n \) by using the transformation \( H_n = H_0(1 - F_n) \).

\[
H_{n+3} - \left(\frac{p}{2}\right)H_{n+2} - \left(\frac{p}{4}\right)H_{n+1} - \left(\frac{p}{8}\right)H_n - (1 - p)H_0 = 0 \tag{4.5}
\]

This will yield in the limit

\[
H_\infty = \lim_{n \to \infty} H_n = \frac{16(1 - \beta)}{8 - 7\beta} P_1 P_2 \tag{4.6}
\]

where \( H_0 = 2P_1 P_2 \) is the frequency of heterozygotes in the initial panmictic population with respect to a single locus with two alleles. Equation 4.6 gives the expected proportion of \( A_1 A_2 \) genotype in the equilibrium population under the system of mixed random and double first cousin mating. This is identical with Equation 3.27 which gives the corresponding frequency in equilibrium under the model of mixed random
and half-sib mating. The recurrence relations for the two systems are quite different and hence, in general, will generate different genotypic distributions after \( n \) generations of mixed mating. But when \( n \) is sufficiently large, i.e. at equilibrium, the genotypic distributions of the two systems coincide.

C. General Mixture of Consanguineous Mating Systems

We now consider a general situation where the mating system involves mating at random and mating among relatives with varying degrees of relationship. The mating system to be considered here is characterized by the following probabilities:

\[
\begin{align*}
P[\text{Random mating}] &= \beta_0 \\
P[\text{Selfing}] &= \beta_1 \\
P[\text{Full-sib mating}] &= \beta_2 \\
P[\text{Half-sib mating}] &= \beta_3
\end{align*}
\]

with \( \beta_0 + \beta_1 + \beta_2 + \beta_3 = 1 \). Matings between parents and offspring, and those among double first cousins are not included because as we have seen earlier, the effect of these matings on the population structure is the same, at least in equilibrium, as that of full-sib and half-sib matings, respectively.

The situation may not seem to be as general and
exhaustive to include all kinds of relatives but it does
take account of some important ones.

We have in this case the following recurrence relation:

\[
F_{n+2} = (\beta_1/2)(1+F_{n+1}) + (\beta_2/4)(1+F_n + 2F_{n+1})
\]

\[
+ (\beta_3/8)(1+F_n + 6F_{n+1}) .
\]  

This relationship can be easily verified directly and we
illustrate this for a system of mixed selfing and full-sib
mating.

Suppose there is selfing with probability \(x\) and full-
sib mating with probability \(y(=1-x)\). Then we have

\[
F_{n+2} = (x/2)(1+F_{n+1}) + (y/4)(1+F_n + 2F_{n+1}) .
\]

This can be seen by considering the following probabilistic
argument:

\[
\begin{array}{c|ccc}
\text{Generation} & Z_n & X_n & Y_n \\
n & & & \\
n+1 & & & \\
n+2 & & & \\
\end{array}
\]

We know that

\[
F_{n+2} = r_{n+1} = r_{X_{n+1}} Y_n + Y_{n+1}
\]
and
\[ r_{X_n+1} Y_{n+1} = x[r_{X_n+1} Y_{n+1} | Y = X] + y[r_{X_n+1} Y_{n+1} | X, Y \text{ F.S.}] \]

where \( Y = X \) denotes selfing and F.S. denotes full-sibbing.

Now
\[ [r_{X_n+1} Y_{n+1} | Y = X] = \frac{1 + F_{n+1}}{2} \] (4.8)
\[ [r_{X_n+1} Y_{n+1} | X, Y \text{ F.S.}] = \frac{1}{4}[r_{X_n} X_n + r_{Y_n} Y_n + 2r_{X_n} Y_n] \]
\[ = \frac{1}{4}[\frac{1 + F_{X_n}}{2} + \frac{1 + F_{Y_n}}{2} + 2F_{n+1}] \]
\[ = \frac{1}{4}(1 + F_{n} + 2F_{n+1}) \] (4.9)

Therefore,
\[ F_{n+2} = \frac{x}{2}(1 + F_{n+1}) + \frac{y}{4}(1 + F_{n} + 2F_{n+1}) \] (4.10)

The formula 4.10 can also be verified as follows:

\begin{center}
\begin{tabular}{c|c|c|c|c}
\hline
Generation & 0 & 1 & 2 & 3 \\
\hline
S & F.S. & S & F.S. & (a) \\
F.S. & S & S & F.S. & (b) \\
F.S. & S & F.S. & S & (c) \\
F.S. & S & F.S. & S & (d) \\
\hline
\end{tabular}
\end{center}

S = Selfing \quad F.S. = Full-sibbing
As we progress from zeroth generation \((F_0 = 0)\), there are various ways in which the individuals arise to form successive generations. This is shown in the above diagram. In generation 1, \(F = \frac{1}{2}\) if selfed or \(F = \frac{1}{4}\) if there is full-sib mating. This will give an average coefficient of inbreeding in generation 1 as

\[
F_1 = \frac{x}{2} + \frac{y}{4} = \frac{1 + x}{4} \quad .
\]  
(4.11)

In generation 2 an individual can arise in the following four ways:

\[
\begin{array}{ccc}
\text{Mating in Generation} & F \\
1 & 2 \\
(a) & S & S & \frac{3}{4} \\
(b) & F.S. & S & \frac{5}{8} \\
(c) & S & F.S. & \frac{1}{2} \\
(d) & F.S. & F.S. & \frac{3}{8} \\
\end{array}
\]

Therefore,

\[
F_2 = \frac{x^2}{4} + \frac{xy}{5} + \frac{xy}{2} + \frac{y^2}{3} \\
= \frac{1}{8} \left(6x^2 + 9xy + 3y^2\right) \\
= \frac{3}{8} (x + y)(2x + y) = \frac{3}{8}(1 + x) \quad .
\]  
(4.12)

which is in agreement if we use the general formula 4.10.

We can now extend the above argument to the individuals of generation 3 which can arise in the following eight ways:
Mating in Generation

\[
\begin{array}{ccc}
  & 1 & 2 & 3 \\
(a) & S & S & S & 7/8 \\
(b) & S & S & F.S. & 3/4 \\
(c) & S & F.S. & S & 3/4 \\
(d) & F.S. & S & S & 13/16 \\
(e) & S & F.S. & F.S. & 5/8 \\
(f) & F.S. & S & F.S. & 5/8 \\
(g) & F.S. & F.S. & S & 11/16 \\
(h) & F.S. & F.S. & F.S. & 1/2 \\
\end{array}
\]

Thus,

\[
F_3 = x^3(7/8) + x^2y(3/4 + 3/4 + 13/16) + xy^2(5/8 + 5/8 + 11/16) + y^3(1/2) \\
= (1/16) (14x^3 + 37x^2y + 31xy^2 + 8y^3) \\
= (1/16) (14x^2 + 8y^2 + 23xy)(x + y) \\
= (1/16) \{8(x + y)^2 + 6x(x + y) + xy\} \\
= (1/16) (9 + 6x + xy) \\
\] \hspace{1cm} (4.13)

Again this can be easily seen to be in agreement with the results that would be obtained by using the general formula 4.10.

In the limit as \( n \to \infty \), we have from Equation 4.7

\[
F_{\infty} = \lim_{n \to \infty} F_n = \frac{4\beta_1 + 2\beta_2 + \beta_3}{8 - 4\beta_1 - 6\beta_2 - 7\beta_3} \\
\] \hspace{1cm} (4.14)
This yields the frequency of heterozygotes at equilibrium $H_\infty$ where

$$H_\infty = \frac{16P_0}{8P_0 + 4P_1 + 2P_2 + P_3} \cdot P_1 P_2 .$$  \hspace{0.5cm} (4.15)

Recently in a review article Karlin (1968) has given some results of mixed imprinting, full-sib mating, random mating and selfing. He has derived these results following the generation matrix approach. The situation discussed here is slightly different. If we make the necessary modifications, the results at equilibrium are found to be in agreement with those of Karlin.

D. Comparison of Various Models

The results obtained have bearing on populations that reproduce by a mixture of cross-fertilization and inbreeding. The results could also be interpreted as describing the effect of departure from a breeding process of complete random mating or of departure from complete inbreeding. It is assumed that the population is large so that the effect of sampling fluctuation can be ignored, and there are no viability or fertility differences.

The mathematical analysis shows that the systems of mixed random and parent-offspring mating, and mixed random and full-sib mating yield identical recurrence relations. Therefore, these two systems will lead to the same genotypic
distribution. The systems of mixed random and half-sib mating and mixed random and double first cousin mating result in different genotypic distributions in a dynamic population. These distributions, however, coincide when \( n \) is sufficiently large i.e., at equilibrium.

Under such mating systems even with a high degree of inbreeding, there is a considerable amount of heterozygosity in the population at equilibrium. The level of heterozygosity depends upon the system of mixed mating, the amount of inbreeding and the initial heterozygosity. It may be of interest to compare the relative effects of these mixed mating systems on the maintenance of heterozygosity. The results are discussed with respect to a single locus with two alleles at that locus. The expected heterozygosity relative to the initial heterozygosity in the population at equilibrium under the three systems is depicted graphically in Figure 3. There is practically no loss in heterozygosity when the system of mating deviates from complete random mating by small amounts of inbreeding. In a highly inbred population with say 90-95 percent of inbreeding, the heterozygosity at equilibrium is to the order of 18-9.5 percent of the initial heterozygosity when inbreeding is by selfing. There is about one and a half times as much heterozygosity when inbreeding is by full-sib or parent-offspring mating, and three times as much when inbreeding
Figure 3. Amount of heterozygosity relative to initial heterozygosity expected in populations at equilibrium under mixed random mating and inbreeding
S = Selfing
F.S. = Full-sib mating
H.S. = Half-sib mating
is by half-sib mating or double first cousin mating.

If $x_1, x_2, x_3$ are the equivalent amounts of selfing, full-sib or parent-offspring mating, and half-sib or double first cousin mating respectively in the system of mixed mating, which yield the same amount of heterozygosis in the population at equilibrium, then we have

$$\frac{x_1}{2 - x_1} = \frac{x_2}{4 - 3x_2} = \frac{x_3}{8 - 7x_3}.$$  

This relationship yields

$$x_2 = \frac{2x_1}{1 + x_1}$$

and

$$x_3 = \frac{2x_2}{1 + x_2}$$

showing that $x_3$ bears the same relationship with $x_2$ as $x_2$ bears with $x_1$. Thus, for example, 20 percent of selfing is equivalent to about 33 percent of full-sib or parent-offspring mating which is equivalent to about 50 percent of half-sib or double first cousin mating in the population in the sense that they would give the same final genotypic proportions in the population. Such comparisons, however, are meaningful when inbreeding procedures are compared in pairs.
Assortative mating means that mated individuals are more similar or dissimilar for some phenotypic traits than if they were chosen at random. There are great difficulties in the theoretical development with regard to the population behavior because this tendency to mate assortatively which depends upon the phenotypic expression of the mates, can be formulated in several ways. Unlike with inbreeding, the choice of a mathematical model is not so obvious with assortative mating and different mating patterns are likely to have different consequences. In this chapter we will consider a few asymmetric deterministic models for a single locus. Some results for a two loci model which is more general than that of Wright (1921) will also be discussed.

The single locus models are developed by specifying the intensities for preference for mating for various phenotypes. It is assumed that preference to choose a mate lies only in one sex, say females. These models allow for an excess of males so that assorting does not decrease the chance of a female to be fertilized. These may be appropriate for populations in which one of the sexes is either in excess, as in plant producing excess pollen, or is able to mate several times. One example of such a system is provided by those insect populations in
which the male is mobile and polygamous and the female is essentially sessile and monogamous and mating in part depends on female preference.

A. Single Locus Models

Consider a single locus with s alleles $A_1, A_2, \ldots, A_s$. We denote, as before, the genotypic distribution at any time by

$$\sum_{i=1}^{s} \sum_{j=1}^{s} P_{ij} A_i A_j,$$

with $P_{ij} = P_{ji}$, and the genic distribution by

$$\sum_{i=1}^{s} p_i A_i.$$

The corresponding arrays in the next generation will be denoted by

$$\sum_{i=1}^{s} \sum_{j=1}^{s} P'_{ij} A_i A_j$$

and

$$\sum_{i=1}^{s} p'_i A_i.$$

The distribution in equilibrium is determined by setting $P'_{ij} = P_{ij}$ for all $(ij)$ and solving the resulting system of equations. In the case of two alleles and complete dominance we shall take $A_1$ to be dominant over $A_2$. The degrees of
assortative mating are described by specifying the fraction of females preferring to mate with individuals of a certain kind.

1. **Model I: two alleles with complete dominance** ($s=2$)

Let $a$ ($0 < a < 1$) be a fraction of the dominant females which prefers to mate with their kind and $\beta$ ($0 < \beta < 1$) that of the recessive females which prefers to mate with their own type. The remaining fraction of females prefers to mate with dissimilar types. The frequencies of various types of mating are given in Table 9. This yields the following recurrence relations for the genotypic frequencies:

\[ P'_{11} = \frac{\alpha P^2_1}{(P_{11} + 2P_{12})} \]

\[ 2P'_{12} = (1 - a)P_1 + \frac{2\alpha P_1 P_{12}}{P_{11} + 2P_{12}} + \frac{(1 - \beta) P_1 P_{22}}{P_{11} + 2P_{12}} \]

\[ P'_{22} = \beta P_{22} + (1 - a)P_{12} + \frac{(1 - \beta) P_{12} P_{22}}{P_{11} + 2P_{12}} + \frac{\alpha P^2_{12}}{P_{11} + 2P_{12}} \]  \hspace{1cm} (5.1)

We know that, in the case of two alleles, two variables, say $P_1$ and $P_{22}'$, are enough to describe the whole population structure at any given time. Thus after simplification, we have

\[ P_1' = P'_{11} + P'_{12} \]

\[ = \frac{P_1}{2(1 - P_{22})[1 + \alpha - P_{22}(\alpha + \beta)]} \]  \hspace{1cm} (5.2)
and

\[ P'_{22} = 1 - \frac{P_1}{1 - P_{22}} \left[ 1 + \alpha - \alpha P_1 - P_{22} (\alpha + \beta) \right]. \]  (5.3)

The change in gene frequency, \( p_1 \), in one generation is described by \( \Delta p_1 \), where

\[ \Delta p_1 = p'_1 - p_1 = \frac{1}{2} \frac{(1 - \alpha)}{P_{22}} + \frac{(1 - \beta) P_{22}}{2 (1 - P_{22})} . \]  (5.4)

Table 9. Mating types and their frequencies

<table>
<thead>
<tr>
<th>Mating type</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1A_1 \times A_1A_1 )</td>
<td>( \alpha P_{11}/(P_{11} + 2P_{12}) )</td>
</tr>
<tr>
<td>( A_1A_1 \times A_1A_2 )</td>
<td>( 4\alpha P_{11}P_{12}/(P_{11} + 2P_{12}) )</td>
</tr>
<tr>
<td>( A_1A_2 \times A_1A_2 )</td>
<td>( 4\alpha P_{12}^2/(P_{11} + 2P_{12}) )</td>
</tr>
<tr>
<td>( A_1A_1 \times A_2A_2 )</td>
<td>( (1 - \alpha)P_{11} + (1 - \beta)P_{22}P_{11}/(P_{11} + 2P_{12}) )</td>
</tr>
<tr>
<td>( A_1A_2 \times A_2A_2 )</td>
<td>( 2(1 - \alpha)P_{12} + 2(1 - \beta)P_{22}P_{12}/(P_{11} + 2P_{12}) )</td>
</tr>
<tr>
<td>( A_2A_2 \times A_2A_2 )</td>
<td>( \beta P_{22} )</td>
</tr>
</tbody>
</table>
The distribution in equilibrium is obtained by setting $P_1' = P_1$ and $P_{22}' = P_{22}$ in Equations 5.2 and 5.3. From Equation 5.2 we see that the nontrivial equilibrium value of $P_{22}$ is $P_{22}^{(\infty)}$ where

$$P_{22}^{(\infty)} = \frac{1 - \alpha}{2 - \alpha - \beta}.$$ \hspace{1cm} (5.5)

$P_{22}^{(\infty)}$, the equilibrium gene frequency of $A_1$, is the admissible solution in $(0,1)$ of the quadratic Equation 5.6 obtained by setting $P_{22}' = P_{22} = P_{22}^{(\infty)}$ in Equation 5.3.

$$\alpha P_1^2 - 2P_1(1 - P_{22}^{(\infty)}) + (1 - P_{22}^{(\infty)})^2 = 0.$$ \hspace{1cm} (5.6)

The admissible root of the quadratic is

$$P_1^{(\infty)} = \frac{(1 - P_{22}^{(\infty)})(1 - (1 - \alpha)^{1/2})}{\alpha}.$$ \hspace{1cm} (5.7)

This model was discussed by Karlin (1968). He has shown that such an interior equilibrium is stable.

2. **Model II: multiple alleles and no dominance**

We express the degree of assortative mating by the following probabilities. Suppose

$$P[A_iA_j \dagger \text{ mates with its own type}] = \alpha_{ij}$$

$$P[A_iA_j \dagger \text{ mates at random}] = 1 - \alpha_{ij}$$

$0 < \alpha_{ij} < 1$, $i,j=1,2,\ldots,s$. As indicated before, it is
assumed that preference is exercised only by one sex, say females, and all females are fertilized. We can easily determine the probabilities of various types of mating under such a model. For example,

\[ P[A_i A_j \times A_i A_j] = \alpha_{ij} p_{ij} + 2(1 - \alpha_{ij}) p_{ij}^2 \]  

for \( i \neq j = 1, 2, \ldots, s \), and

\[ P[A_i A_j \times A_k A_\ell] = (2 - \alpha_{ij} - \alpha_{k\ell}) p_{ij} p_{k\ell} \]

where \((ij) \neq (k\ell)\), and \(i, j, k, \ell = 1, 2, \ldots, s\). We can easily generate the distribution of genotypes in the next generation. An offspring \(A_i A_i\) arises by the matings of the type given in Table 10. This yields

\[ P'_{ii} = \alpha_{ii} p_{ii} + (1 - \alpha_{ii}) p_{ii}^2 + \sum_{j \neq i} (1 - \alpha_{ij}) p_{ij} p_{ii} \]

\[ + \sum_{j \neq i} (1 - \alpha_{ij}) p_{ij} p_{ii} + \frac{1}{2} \sum_{j \neq i} \alpha_{ij} p_{ij} \]

\[ + \sum_{j \neq i} (1 - \alpha_{ij}) p_{ij}^2 + \sum_{j \neq i} \sum_{k \neq j \neq i} (1 - \alpha_{ij}) p_{ij} p_{ik} \]

\[ = \alpha_{ii} p_{ii} + (1 - \alpha_{ii}) p_{ii} p_i + \frac{1}{2} \sum_{j \neq i} \alpha_{ij} p_{ij} \]

\[ + \sum_{j \neq i} (1 - \alpha_{ij}) p_{ij} p_i \]

\[ = \frac{1}{2} \alpha_{ii} p_{ii} + p_i^2 (1 - \alpha_i) + \frac{1}{2} p_i \alpha_i. \]  

(5.10)
where

\[ p_{i}^{a_{i}} = \sum_{j=1}^{s} p_{ij}^{a_{ij}}, \quad i=1,2,...,s. \]

By a similar argument we obtain

\[ 2p'_{ik} = \alpha_{ik} p_{ik} + 2p_{i} p_{k} (1 - \frac{a_{i} + a_{k}}{2}) \]

or

\[ p'_{ik} = \frac{\alpha_{ik} p_{ik}}{2} + p_{i} p_{k} (1 - \frac{a_{i} + a_{k}}{2}) \quad (5.11) \]

\[ i \neq k = 1,2,...,s. \]

Table 10. Distribution of parental matings which give rise to \( A_{i}A_{i} \) offspring

<table>
<thead>
<tr>
<th>Mating Type</th>
<th>Probability</th>
<th>Conditional ( P[A_{i}A_{i}] ) in Progeny</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_{i}A_{i} \times A_{i}A_{i} )</td>
<td>( \alpha_{ii} p_{ii} + (1 - \alpha_{ii}) p_{ii}^2 )</td>
<td>1</td>
</tr>
<tr>
<td>( A_{i}A_{i} \times A_{i}A_{j} )</td>
<td>( 2(1 - \alpha_{ii}) p_{ii} p_{ij} ), ( j \neq i )</td>
<td>1/2</td>
</tr>
<tr>
<td>( A_{i}A_{j} \times A_{i}A_{i} )</td>
<td>( 2(1 - \alpha_{ij}) p_{ij} p_{ii} ), ( j \neq i )</td>
<td>1/2</td>
</tr>
<tr>
<td>( A_{i}A_{j} \times A_{i}A_{j} )</td>
<td>( 2\alpha_{ij} p_{ij} + 4(1 - \alpha_{ij}) p_{ij}^2 ), ( j \neq i )</td>
<td>1/4</td>
</tr>
<tr>
<td>( A_{i}A_{j} \times A_{i}A_{k} )</td>
<td>( 4(1 - \alpha_{ij}) p_{ij} p_{ik} ), ( k \neq j \neq i )</td>
<td>1/4</td>
</tr>
</tbody>
</table>
From the set of Equations 5.10 and 5.11 we can write

\[ p'_i = \sum_{k=1}^{s} p'_{ik} \]

\[ = \sum_{k=1}^{s} p_i p_k (1 - \frac{\alpha_i + \alpha_k}{2}) + \frac{1}{2} \sum_{k=1}^{s} \alpha_{ik} p_{ik} + \frac{1}{2} p_i \alpha_i. \]

\[ = p_i [1 - \frac{\alpha_i + \alpha_i}{2}] + p_i \alpha_i. \]

\[ = p_i [1 + \frac{\alpha_i - \alpha_i}{2}] \]

(5.12)

where \( i = 1, 2, \ldots, s \) and

\[ \alpha_i = \sum_{i=1}^{s} p_i \alpha_i = \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij} \alpha_{ij}. \]

The change in gene frequency in one generation is expressed by

\[ \Delta p_i = p'_i - p_i = p_i \frac{(\alpha_i - \alpha_i)}{2} \]

(5.13)

\[ i = 1, 2, \ldots, s, \]

and

\[ \sum_{i=1}^{s} \Delta p_i = 0 \]

At equilibrium \( \Delta p_i = 0, i = 1, 2, \ldots, s \), which will yield, for nontrivial equilibrium,

\[ \alpha_i = \alpha_i \]

(5.14)

for all \( i \), and there is one dependency among these \( s \)
equations, namely

\[ \sum_{i=1}^{s} p_i a_{i1} = a_{22} . \]

Also, at equilibrium \( P'_{ik} = P_{ik} \) which will yield, incorporating Equation 5.14,

\[
P_{ik}(1 - \frac{a_{ik}}{2}) = P_i P_k (1 - a_{22}) \quad (5.15)
\]

\( i \neq k = 1, 2, \ldots, s \). The admissible solution for \( \{p_i\} \) and \( \{P_{ik}\}_{i \neq k} \) from these equations will determine the distribution in equilibrium. Karlin (1968) discussed this model for \( s = 2 \) and obtained the explicit solution. If \( a_{12} \) is intermediate between \( a_{11} \) and \( a_{22} \), then one or the other of the two homozygous types gets fixed and there is no internal equilibrium.

We now consider a special case of \( a_{12} = a_{11} + a_{22} \) with \( 0 < a_{12} < 1 \). As pointed out before, there is no interior equilibrium for \( a_{12} = (a_{11} + a_{22})/2 \). In this case we obtain

\[
P_{1a_{11}} = p_{11} a_{11} + p_{12} a_{12}
\]

\[= p_{11} a_{11} + p_{12} (a_{11} + a_{22})
\]

\[= p_{1} a_{11} + p_{12} a_{22} \cdot \]

Similarly,

\[
P_{2a_{22}} = p_{22} a_{22} + p_{12} a_{11} \cdot \]
Therefore,

\[ \Delta p_1 = p_1 \frac{(\alpha_1 - \alpha_2)}{2} \]

\[ = \frac{p_1 p_2}{2} (\alpha_1 - \alpha_2) \]

\[ = \frac{p_1 p_2}{2} (\alpha_{11} - \alpha_{22}) + \frac{p_{12}}{2} (p_2 \alpha_{22} - p_1 \alpha_{11}) \quad (5.16) \]

and

\[ p_{12}' = \frac{p_{12}}{2} \alpha_{12} + p_1 p_2 (1 - \frac{\alpha_1 + \alpha_2}{2}) \]

\[ = \frac{p_{12}}{2} (p_2 \alpha_{11} + p_1 \alpha_{22}) + \frac{p_1 p_2}{2} (2 - \alpha_{11} - \alpha_{22}) \quad (5.17) \]

Using the conditions that at equilibrium \( \Delta p_1 = 0 \) and \( p_{12}' = p_{12} \) we get the equilibrium values \( p_1^{(\infty)} \) and \( p_{12}^{(\infty)} \) which are the solution of

\[ p_{12} (p_2 \alpha_{22} - p_1 \alpha_{11}) + p_1 p_2 (\alpha_{11} - \alpha_{22}) = 0 \]

\[ p_{12} (2 - p_2 \alpha_{11} - p_1 \alpha_{22}) - p_1 p_2 (2 - \alpha_{11} - \alpha_{22}) = 0 \quad (5.18) \]

This will yield

\[ p_1^{(\infty)} = \frac{2 \alpha_{11} - (\alpha_{11}^2 + \alpha_{22}^2)}{2[ (\alpha_{11} + \alpha_{22}) - (\alpha_{11}^2 + \alpha_{22}^2) ]} \]

\[ p_2^{(\infty)} = 1 - p_1^{(\infty)} \]
\[ p_{12}^{(\infty)} = \frac{p_1^{(\infty)} p_2^{(\infty)} (\alpha_{11} - \alpha_{22})}{p_1^{(\infty)} \alpha_{11} - p_2^{(\infty)} \alpha_{22}}. \]  

(5.19)

We can derive the frequencies of other genotypes easily by considering

\[ p_{11}^{(\infty)} = p_1^{(\infty)} - p_{12}^{(\infty)} \]
\[ = \frac{p_1^{(\infty)} [p_1^{(\infty)} - p_2^{(\infty)}] \alpha_{11}}{p_1^{(\infty)} \alpha_{11} - p_2^{(\infty)} \alpha_{22}}. \]

(5.20)

\[ p_{22}^{(\infty)} = p_2^{(\infty)} - p_{12}^{(\infty)} \]
\[ = \frac{p_2^{(\infty)} [p_1^{(\infty)} - p_2^{(\infty)}] \alpha_{22}}{p_1^{(\infty)} \alpha_{11} - p_2^{(\infty)} \alpha_{22}}. \]

This internal equilibrium will be admissible if and only if \( \alpha_{11} \) and \( \alpha_{22} \) are such that

\[ \alpha_{11} + \alpha_{22} \leq 1 \]  

(5.21)

and

\[ \alpha_{11}^2 + \alpha_{22}^2 - 2\alpha_{11} \leq 0 \]
\[ \alpha_{11}^2 + \alpha_{22}^2 - 2\alpha_{22} \leq 0 \]  

(5.22)

These conditions arise by considering \( 0 < p_1^{(\infty)} < 1 \) and
\[ 0 < u_{12} \leq 1. \]  

Also, we can show that...
\[ [\alpha_{11} - \alpha_{22}] / [p_1^{(\infty)}\alpha_{11} - p_2^{(\infty)}\alpha_{22}] \]

which is the coefficient of \( p_1^{(\infty)} \) \( p_2^{(\infty)} \) in Equation 5.19 is positive and less than two. We can rewrite the inequalities 5.21 and 5.22 as

\[ \alpha_{11} + \alpha_{22} \leq 1 \]

and

\[ (1 - \alpha_{11})^2 + \alpha_{22}^2 < 1 \]

\[ \alpha_{11}^2 + (1 - \alpha_{22})^2 < 1 \]  \hspace{1cm} (5.23)

This equilibrium will be stable because this satisfies the conditions obtained by Karlin (1968) which in this case are the same as the inequalities 5.22.

The analysis of the model, in its general form, for an explicit formula, gets complicated and this necessitates the consideration of some simplifying assumptions. Therefore, in the mathematical analysis that follows we confine ourselves to the following special cases:

(a) \( \alpha_{ii} = \alpha \), \( \alpha_{ik} = \beta \neq \alpha \), \( i \neq k \), \( i,k = 1,2,\ldots,s \).

We know from Equations 5.14 that at equilibrium \( \alpha_{i.} = \alpha_{..} \) for all \( i \). This implies that for \( i=1,2,\ldots,s \)

\[ p_{i.} \alpha + \beta \sum_{k \neq i} p_{i.} = p_{i.} \alpha_{i.} = p_{i.} \alpha_{..} \]
or
\[ p_{ii} \alpha + \beta(p_i - p_{ii}) = p_i \alpha \]
or
\[ p_{ii}(\alpha - \beta) = p_i(\alpha_{..} - \beta) \quad (5.24) \]

which shows that \( p_{ii} \prec p_i \).

Setting \( P'_{ii} = P_{ii} \) and \( P'_{ik} = P_{ik} \) in Equations 5.10 and 5.11, and incorporating \( \alpha_{i.} = \alpha_{..} \), \( i = 1, 2, \ldots, s \), we have at equilibrium

\[ P_{ii}(1 - \alpha/2) = p_i[p_i(1 - \alpha_{..}) + \alpha_{..}/2] \quad (5.25) \]

and

\[ P_{ik}(1 - \beta/2) = p_i p_k (1 - \alpha_{..}) \quad (5.26) \]

Using the result that \( P_{ii} \) is proportional to \( p_i \), we get from Equation 5.25

\[ p_i (1 - \alpha_{..}) + \alpha_{..}/2 = \text{constant} \]

\( i = 1, 2, \ldots, s \), which shows that \( p_i \) is a constant, independent of \( i \). Thus at equilibrium

\[ p_i^{(\infty)} = \text{constant} = \frac{1}{s} \quad (5.27) \]

\( i = 1, 2, \ldots, s \), because \( \sum p_i^{(\infty)} = 1 \).

Now

\[ \alpha_{..} = \sum_{i=1}^{s} \sum_{k=1}^{s} P_{ik} g_{ik} \]

where

\[ H = \sum_{i} \sum_{k \neq i} P_{ik} \] = Proportion of all heterozygotes.

Substituting the values of \( p_{i} \) and \( \alpha \) in Equation 5.24, we get at equilibrium

\[ P_{ii} = p_{i}(1-H) \]

\[ = \frac{1}{s}(1-H) \] \hspace{1cm} (5.29)

We see from Equation 5.26 that, at equilibrium, \( P_{ik} \) = constant, because \( p_{i} \)'s are constant. This will yield, at equilibrium

\[ P_{ik} = \text{constant} \quad i \neq k \]

\[ = \frac{H}{s(s-1)} \] \hspace{1cm} (5.30)

where \( H \) is the total heterozygosity.

Equations 5.26 in conjunction with Equations 5.28 and 5.29 yields at internal equilibrium

\[ (2-\beta) \sum_{k \neq i} P_{ik} = 2p_{i}(1-\alpha..) \sum_{k \neq i} P_{k} \]

or

\[ (2-\beta)(p_{i} - p_{ii}) = 2p_{i}(1-p_{i})(1-\alpha..) \]
or

\[(2 - \beta)p_i H = 2p_i (1 - p_i) (1 - \alpha)\]

or

\[(2 - \beta)H = \frac{2(s - 1)}{s} \left[ (1 - \alpha) + H(\alpha - \beta) \right]\]

which gives

\[H = \frac{2(s - 1)(1 - \alpha)}{2s - 2\alpha(s - 1) + \beta(s - 2)} \quad (5.31)\]

Thus, the mathematical analysis yields the nontrivial equilibrium given by

\[P_{ii}^{(\infty)} = \frac{(1 - H)}{s}, \quad i=1,2,...,s\]

\[P_{ik}^{(\infty)} = \frac{H}{s(s - 1)}, \quad i\neq k=1,2,...,s\]

\[P_{i}^{(\infty)} = \frac{1}{s}, \quad i=1,2,...,s \quad (5.32)\]

where \(H\) is the total heterozygosity and is determined by Equation 5.31.

Karlin and Scudo (1969) discussed this model for two alleles \((s = 2)\) and observed that the equilibrium distribution is independent of the parameter \(\beta\). These results of Karlin and Scudo are a particular case of a more general model that we have discussed here with multiple alleles. As we can see from Equations 5.31 and 5.32, the equilibrium distribution is no longer independent of the parameter \(\beta\) when more than
two alleles are involved.

(b) \( a_{ik} = \beta \), \( i \neq k \)

This is a less restrictive assumption than in (a) in that we now permit the homozygous females to exercise different preference in assortment but we still assume that heterozygous females have the same degree of assortment. Proceeding in the same manner as we did in (a) we get, from Equation 5.15, at equilibrium

\[
P_{ik} = \frac{2p_ip_k(1-\alpha)}{2-\beta} \quad (5.33)
\]

\( i \neq k = 1, 2, \ldots, s \). We rewrite this as

\[
P_{ik} = p_ip_k(1-F) \quad (5.34)
\]

where \( 1 - F = \frac{2(1-\alpha)}{2-\beta} \).

With this we can write

\[
P_{ii} = p_iF + (1-F)p_i^2 \quad (5.35)
\]

The restriction \( \alpha_i = \alpha \) for all \( i \) (Equation 5.14) at equilibrium results in the following set of equations:

\[
[p_iF + p_F]a_{ii} + \beta \sum_{k \neq i} p_{ik} = p_i\alpha
\]

or

\[
[p_i^2(1-F) + p_iF]a_{ii} + \beta \sum_{k \neq i} p_ip_k(1-F) = p_i\alpha
\]
or

\[ p_i(1-F) + F\alpha_i + \beta(1-F)(1-p_i) = \alpha_{..} \]

or

\[ p_i(1-F)(\alpha_i - \beta) + F(\alpha_i - \beta) = \alpha_{..} - \beta \]  \hspace{1cm} (5.36)

for \( i=1,2,...,s \). Thus we have the following set of \( s \) equations:

\[ p_1(1-F)a_1 + Fa_1 = \alpha_{..} - \beta \]
\[ p_2(1-F)a_2 + Fa_2 = \alpha_{..} - \beta \]
\[ \ldots \]
\[ p_s(1-F)a_s + Fa_s = \alpha_{..} - \beta \]  \hspace{1cm} (5.37)

where \( a_i = \alpha_i - \beta \), \( i=1,2,...,s \). There are \( (s-1) \) independent equations because of the restriction

\[ \sum_{i=1}^{s} p_i a_i = \alpha_{..} \]

and we have \( s-1 \) unknown \( p_i \)'s. Subtracting the last equation from each of the first \( (s-1) \) equations, we have

\[ (1-F)(p_1 a_1 - p_s a_s) + F(a_1 - a_s) = 0 \]
\[ (1-F)(p_2 a_2 - p_s a_s) + F(a_2 - a_s) = 0 \]
\[ \ldots \]
\[ (1-F)(p_{s-1} a_{s-1} - p_s a_s) + F(a_{s-1} - a_s) = 0 \]  \hspace{1cm} (5.38)
These equations will yield

\[ p_i = \frac{p_s a_s}{a_i} - \frac{F}{1-F} \frac{(a_i - a_s)}{a_i} \]  \hspace{1cm} (5.39)

\( i=1,2,\ldots,s-1. \)

\[ 1-p_s = \frac{s-1}{\sum_{i=1}^{s-1} p_i} = p_s a_s \frac{s-1}{\sum_{i=1}^{s-1} \frac{1}{a_i}} - \frac{F}{1-F} \left[ s-1 - a_s \sum_{i=1}^{s-1} \frac{1}{a_i} \right] \]

or

\[ \frac{1-p_s}{a_s} = p_s \frac{s-1}{\sum_{i=1}^{s-1} \frac{1}{a_i}} - \frac{F}{1-F} \left[ s - \sum_{i=1}^{s-1} \frac{1}{a_i} \right] \]

or

\[ \frac{1}{a_s} = p_s \frac{s}{\sum_{i=1}^{s} \frac{1}{a_i}} - \frac{F}{1-F} \left[ s - \sum_{i=1}^{s} \frac{1}{a_i} \right] \]

or

\[ p_s \left( \sum_{i=1}^{s} \frac{1}{a_i} \right) = \frac{1}{a_s} + \frac{F}{1-F} \left[ s - \sum_{i=1}^{s} \frac{1}{a_i} \right] \] .  \hspace{1cm} (5.40)

This will give

\[ \frac{sp_s}{a} = \frac{1}{a_s} \left[ \frac{1 + (s-1)F}{1-F} \right] - \frac{F}{1-F} \frac{s}{a} \]  \hspace{1cm} (5.41)

where

\[ \frac{1}{a} = \frac{1}{s} \sum_{i=1}^{s} \frac{1}{a_i} \]

Thus, we have at equilibrium
\[
\Pi_i^{(\infty)} = \frac{1}{s} \sum a_i \left[ \frac{1 + (s-1)F}{1-F} \right] - \frac{F}{1-F}
\]

(5.42)

\[i=1,2,...,s.\] The other unknown quantity to be determined is \(F\) in order that the equilibrium distribution is completely specified.

We know from Equation 5.36 that

\[\alpha_i = p_i (1-F)a_i + Fa_i + \beta\]

Substituting \(2(1-\alpha_i) = (2-\beta)(1-F)\) from the defining equation for \(F\), and the value of \(p_i\) from Equation 5.42, we obtain

\[(2-\beta)(1-F) = 2[1-p_i a_i (1-F) - Fa_i - \beta]\]

or

\[2p_i a_i (1-F) + 2Fa_i + \beta F - 2F + \beta = 0\]

or

\[2a_i \left[ \frac{1}{s} \sum a_i \left[ 1 + (s-1)F \right] - F \right] + 2Fa_i + \beta F - 2F = 0\]

or

\[F \left[ \frac{2a(s-1)}{s} - (2-\beta) \right] + \beta + \frac{2a}{s} = 0\]

This will give

\[F = \frac{2a + \beta s}{s(2-\beta) - 2a(s-1)}\]

(5.43)

Thus, we will have the nontrivial equilibrium
\[ p_{ik}^{(\infty)} = p_i^{(\infty)} p_k^{(\infty)} (1-F) , \quad i \neq k = 1, 2, \ldots, s \]

\[ p_i^{(\infty)} = \frac{1}{s} \frac{\alpha_i [1 + (s-1)F]}{1-F} - \frac{F}{1-F} , \quad i = 1, 2, \ldots, s \]  

(5.44)

where \( F \) is given by Equation 5.43, \( \alpha_i = \alpha_{ii} - \beta, i = 1, 2, \ldots, s \); and \( \alpha \) is the harmonic mean of the \( \alpha_i \)'s. This nontrivial equilibrium will be admissible provided the assortment parameters \( \alpha_{ii}, i = 1, 2, \ldots, s \), and \( \beta \) are such that \( 1 - F > 0 \), and \( 0 < p_i^{(\infty)} < 1 \) for all \( i \).

3. Model III

We discuss now another class of models, where a female has a specified probability of mating with a particular male type. These conditional probabilities are likely to be different for different types of females. We assume, however, that these conditional probabilities are frequency independent and remain the same from generation to generation. Thus, the model is specified by

\[ P[\text{Mate is } A_r A_r | O A_i A_j] = P_{rt}(ij) \]  

(5.45)

where \( P_{rt}(ij) = P_{rt}(ji) = P_{tr}(ij) = P_{tr}(ji) \), and

\[ \sum_{r=1}^{s} \sum_{t=1}^{s} P_{rt}(ij) = 1 \]

Note that these probabilities are for ordered genotypes.

Let \( p_r(i, j) \) be the conditional probability of the male
gamete being \( A_r \) given that the female parent is \( A_1A_j \), that is,

\[
\Pr(ij) = \sum_{t=1}^{s} \Pr_t(ij) = P[\text{gamete } A_r|O A_1A_j]
\]

for \( r=1,2,\ldots,s \).

Similarly we define \( q_k \), \( k=1,2,\ldots,s \), as the average frequency of the male gamete \( A_k \), averaged over all females, that is,

\[
q_k = \frac{1}{s} \sum_{i=1}^{s} \Pr[ij] P[k(ij)]
\]

for \( k=1,2,\ldots,s \).

Let \( \Delta_k = q_k - p_k \), \( k=1,2,\ldots,s \), where \( p_k \) is the frequency of the allele \( A_k \) in females and

\[
\sum_{k=1}^{s} \Delta_k = 0
\]

The quantity \( \Delta_k \) is the deviation of the frequency of \( A_k \) in females from the average frequency of \( A_k \) among the male gametes.

The whole model is represented by the distribution of mates given in Table 11 with

\[
q_k = \sum_{i=1}^{s} \sum_{j=1}^{s} \Pr[ij] P[k(ij)]
\]

where

\[
L_i = \frac{1}{s} \sum_{j=1}^{s} \Pr[ij]
\]

and

\[
P_{k(ij)} = \frac{1}{s} \sum_{i=1}^{s} \Pr[ij]
\]
<table>
<thead>
<tr>
<th>Females</th>
<th>Conditional probability of mates</th>
<th>Males</th>
<th>gamete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype Probability</td>
<td>$A_1^1A_1^1$</td>
<td>$A_1^1A_2$</td>
<td>$...$</td>
</tr>
<tr>
<td>$A_1^1A_1^1$</td>
<td>$P_{11}$</td>
<td>$P_{11}(11)$</td>
<td>$2P_{12}(11)$</td>
</tr>
<tr>
<td>$A_1^1A_2$</td>
<td>$2P_{12}$</td>
<td>$P_{11}(12)$</td>
<td>$2P_{12}(12)$</td>
</tr>
<tr>
<td>$A_2^2A_2$</td>
<td>$P_{22}$</td>
<td>$P_{11}(22)$</td>
<td>$2P_{12}(22)$</td>
</tr>
<tr>
<td>$...$</td>
<td>$...$</td>
<td>$...$</td>
<td>$...$</td>
</tr>
<tr>
<td>$A_s^sA_s$</td>
<td>$P_{ss}$</td>
<td>$P_{11}(ss)$</td>
<td>$2P_{12}(ss)$</td>
</tr>
</tbody>
</table>

This results in the following probabilities of mating types among the ordered genotypes:

$$
P[\varnothing A_i^1A_j^j \times \varnothing A_r^rA_t^t] = P[\varnothing A_i^1A_j^j]P[\varnothing A_r^rA_t^t|\varnothing A_i^1A_j^j] = P_{ij}P_{rt}(ij)
$$

(5.49)
i,j,r,t=1,2,...,s.

These will result in the progeny array

$$
\sum_i \sum_j \sum_r \sum_t P_{ij}P_{rt}(ij) (\frac{1}{2}A_i^1 + \frac{1}{2}A_j^j)(\frac{1}{2}A_r^r + \frac{1}{2}A_t^t) = \frac{1}{2} \sum_i \sum_r \sum_j \sum_{ij} P_{ij}P_{rt}(ij)A_i^1A_r^r
$$
\[ + \frac{1}{2} \sum_{j=1}^{s} \sum_{r=1}^{s} \left[ \sum_{i=1}^{s} P_{ij} P_{r}(ij) \right] A_{j} A_{r} \]

\[ = \sum_{i=1}^{s} \sum_{r=1}^{s} \left( \sum_{j=1}^{s} P_{ij} P_{r}(ij) \right) A_{i} A_{r} \]

Therefore,

\[ P'_{ii} = P[A_{i}i \text{ in offspring generation}] \]

\[ = \sum_{j=1}^{s} P_{ij} P_{i}(ij) \] (5.50)

and

\[ 2P'_{ik} = P[A_{i}A_{k} \text{ in offspring generation, } i \neq k] \]

\[ = \sum_{j=1}^{s} P_{ij} P_{k}(ij) + \sum_{j=1}^{s} P_{kj} P_{i}(kj) \] (5.51)

The frequency of gene \( A_{k} \) in the offspring generation is

\[ P'_{k} = \sum_{i=1}^{s} P'_{ik} \]

\[ = \frac{1}{2} \sum_{i=1}^{s} \sum_{j=1}^{s} P_{ij} P_{k}(ij) + \frac{1}{2} \sum_{i=1}^{s} \sum_{j=1}^{s} P_{kj} P_{i}(kj) \]

\[ = \frac{1}{2} \alpha_{k} + \frac{1}{2} \sum_{j=1}^{s} P_{kj} \]

\[ = \frac{1}{2} \alpha_{k} + \frac{1}{2} P_{k} \]

\[ = P_{k} + \frac{1}{2} \Delta_{k} \] (5.52)
As should be obvious for this kind of development with regard to genotypic distribution, we do not need to know the conditional distributions of mates. All that we need to know are the conditional probabilities of male gametes for each female type. However, the specification of such conditional distributions will be needed when we are investigating some other characteristics of the population, e.g., covariances between relatives.

The above recursion formulas are quite complicated and the complete analysis of the model in its general form may not be possible. In order to have some insight of the effect of this model, we shall in the first instance discuss only the following special cases. However, we shall discuss the behavior in equilibrium for the general model. We shall come back to it later in the section.

(a) $P_k(ij) = q_k$ for all $(ij)$

This implies that assortment parameters are such that the conditional probabilities of male gametes are the same for all female types. This further implies that $q_k$, which in general varies with time, will remain constant over generations under the above assumption. This restriction, however, does not necessarily mean that the conditional probabilities of mates of different females are the same. For example, in the case of two alleles, we may have the information contained in Table 12, where the frequency of male
Table 12. Conditional probability distributions of mates

<table>
<thead>
<tr>
<th>♀ genotype</th>
<th>Conditional probability of mates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A_1A_1$</td>
</tr>
<tr>
<td>$A_1A_1$</td>
<td>0.3</td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td>0.2</td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Gamete $A_1$ is 0.4 for all female types but the conditional distributions are different.

This assumption leads to a great simplification in the mathematical analysis, and the complete time dependent behavior of $\{P_i^{(n)}\}$ and $\{P_{ij}^{(n)}\}$ can be easily worked out. We know that $P_k' = P_k + \frac{1}{2} \Delta_k$ which yields

$$\Delta_k' = q_k - P_k' = \frac{1}{2} \Delta_k$$

because $q_k$ is now time invariant. Thus,

$$\Delta_k^{(n)} = \frac{1}{2} \Delta_k^{(n-1)}$$

$$= \left(\frac{1}{2}\right)^n \Delta_k^{(0)} \quad \text{(5.53)}$$

which goes to zero as $n \to \infty$. Hence,
As \(n \to \infty\), \(p_k^{(n)} \to p_k^{(\infty)} = q_k\).

From Equations 5.50 and 5.51 we can write for all \(i, k = 1, 2, \ldots, s\), in generation \(n\),

\[
p_{ik}^{(n)} = \frac{1}{2} q_k \sum_{j=1}^{s} p_{ij}^{(n-1)} + \frac{1}{2} q_i \sum_{j=1}^{s} p_{kj}^{(n-1)}
\]

\[
= \frac{1}{2} [q_k p_i^{(n-1)} + q_i p_k^{(n-1)}]
\]

\[
= \frac{1}{2} [q_k (\Delta_i^{(n-1)}) + q_i (\Delta_k^{(n-1)})]
\]

\[
= q_i q_k - \frac{1}{2} [q_k \Delta_i^{(n-1)} + q_i \Delta_k^{(n-1)}]
\]

\[
= q_i q_k - \left( \frac{1}{2} \right)^n [q_k \Delta_i^{(0)} + q_i \Delta_k^{(0)}]
\]

(5.55)

where \(\Delta_k^{(0)} = q_k^{(0)} - p_k^{(0)}\). In the limit as \(n \to \infty\)

\[
p_{ik}^{(\infty)} = \lim_{n \to \infty} p_{ik}^{(n)} = q_i q_k
\]

\[
p_k^{(\infty)} = \lim_{n \to \infty} p_k^{(n)} = q_k
\]

(5.56)

for \(i, k = 1, 2, \ldots, s\).

This shows that at equilibrium, the population has Hardy-Weinberg structure. The equilibrium is reached when the frequencies of female gametes become the same as the
conditional probabilities of male gametes. Such an equilibrium will be stable and is approached fairly rapidly, the difference $|p_k - q_k|$ falling off at the rate of one-half in each generation.

(b) Two alleles ($s=2$)

We shall in the rest of our discussion restrict to the two-allele case. We shall examine the model under several simplifying assumptions which may not be as severe as in (a). In the case of $s=2$ we essentially need two independent quantities to specify a population, which from Equation 5.51 and 5.52 are

\[ 2p'_{12} = p_{11}p_{2}(11) + p_{12}p_{2}(12) + p_{21}p_{1}(21) + p_{22}p_{1}(22) \]

\[ = p_{11}p_{2}(11) + p_{12} + p_{22}p_{1}(22) \]

and

\[ p'_1 = p_1 + \frac{1}{2} \Delta_1 \]

where

\[ \Delta_1 = q_1 - p_1 \]

\[ = p_{11}p_{1}(11) + 2p_{12}p_{1}(12) + p_{22}p_{1}(22) - p_1 \]

\[ = [p_{2}p_{1}(22) - p_{1}p_{2}(11)] - p_{12}[p_{1}(11) - 2p_{1}(12) + p_{1}(22)]. \]
(i) Suppose

\[ p_1(11) - 2p_1(12) + p_1(22) = 0 \]

and

\[ \frac{p_1(11) + p_2(22)}{2} < 1 \]

Then from Equation 5.59

\[ \Delta_1 = p_2 p_1(22) - p_1 p_2(11) \]

and hence,

\[ \Delta'_1 = p'_2 p_1(22) - p'_1 p_2(11) \]

\[ = \left[ \frac{p_1(11) + p_2(22)}{2} \right] \Delta_1 \]

because the conditional probabilities \{p_{x(ij)}\} are time invariant. This will give in the nth generation

\[ \Delta_1^{(n)} = p_2^{(n)} p_1(22) - p_1^{(n)} p_2(11) \]

\[ = \left[ \frac{p_1(11) + p_2(22)}{2} \right] \Delta_1^{(n-1)} \]

\[ = \left[ \frac{p_1(11) + p_2(22)}{2} \right]^n \Delta_1^{(0)} \rightarrow 0 \text{ as } n \rightarrow \infty \]

(5.60)

because \([p_1(11) + p_2(22)]/2\) is assumed to be less than unity.

From Equation 5.58 we can easily write that
\[ p_1^{(n)} = p_1^{(0)} + \frac{1}{2} \sum_{r=0}^{n-1} \Delta_1(r) \]

\[ = p_1^{(0)} + \frac{1}{2} \frac{1 - \left[ \frac{P_1(11) + P_2(22)}{2} \right]^n}{1 - \frac{P_1(11) + P_2(22)}{2}} \Delta_1(0) \]

\[ = p_1^{(0)} + \frac{1 - \left[ \frac{P_1(11) + P_2(22)}{2} \right]^n}{P_2(11) + P_1(22)} \Delta_1(0) \]  

(5.61)

As \( n \to \infty \)

\[ p_1^{(n)} \to p_1^{(\infty)} = p_1^{(0)} + \frac{1}{P_2(11) + P_1(22)} \Delta_1(0) \]

\[ = \frac{P_1(22)}{P_1(22) + P_2(11)} \]  

(5.62)

which is dependent only on the assortment parameters. It is very easy to see that this would be a stable equilibrium position because we can write

\[ p_1^{(n)} - p_1^{(\infty)} = \left[ p_1^{(0)} - p_1^{(\infty)} \right] \left[ \frac{P_1(11) + P_2(22)}{2} \right]^n \]  

If \( p_1^{(0)} < p_1^{(\infty)} \), then \( p_1^{(n)} \) increases to \( p_1^{(\infty)} \) and if \( p_1^{(0)} > p_1^{(\infty)} \), then \( p_1^{(n)} \) decreases to \( p_1^{(\infty)} \) and the rate of approach is

\[ \frac{P_1(11) + P_2(22)}{2} \]  

.
From Equation 5.57

\[ p_{12}^{(n)} = \frac{1}{2} [p_{2(11)}^{(n-1)} p_{11}^{(n-1)} + p_{12}^{(n-1)} + p_{1(22)}^{(n-1)}] \]

\[ = \frac{1}{2} [p_1^{(n-1)} p_{2(11)} + p_2^{(n-1)} p_{1(22)}] \]

\[ - \frac{1}{2} p_{12}^{(n-1)} [p_{2(11)} - 1 + p_{1(22)}] \]

\[ = \frac{1}{2} \Delta_1^{(n-1)} + p_{2(11)} p_1^{(n-1)} + \frac{1}{2} p_{12}^{(n-1)} [p_{1(11)} - p_{1(22)}] \]

\[ = \frac{1}{2} \delta_1^{(n-1)} + \frac{p_{1(11)} - p_{1(22)}}{2} p_{12}^{(n-1)} \]  

(5.63)

where

\[ \delta_1^{(n)} = \Delta_1^{(n)} + 2p_{2(11)} p_1^{(n)} \]

Substituting the values of \( p_{12}^{(n-1)} \), \( p_{12}^{(n-2)} \), ..., successively in the right hand side of Equation 5.63, we obtain

\[ p_{12}^{(n)} = \frac{1}{2} \left[ \delta_1^{(n-1)} + \frac{p_{1(11)} - p_{1(22)}}{2} \delta_1^{(n-2)} \right] \]

\[ + \left[ \frac{p_{1(11)} - p_{1(22)}}{2} \right]^{n-3} \delta_1^{(n-3)} + ... \]

\[ + \left[ \frac{p_{1(11)} - p_{1(22)}}{2} \right]^{n-1} \delta_1^{(0)} \}

\[ + \left[ \frac{p_{1(11)} - p_{1(22)}}{2} \right]^n p_{12}^{(0)} \]
\[
\delta_1^{(n)} = \sum_{r=0}^{n-1} \left[ \frac{p_1(11) - p_1(22)}{2} \right]^r \delta_1^{(n-r-1)} + \left[ \frac{p_1(11) - p_1(22)}{2} \right]^n p_1(0).
\]  
(5.64)

Now we need to evaluate
\[
\sum_{r=0}^{n-1} \left[ \frac{p_1(11) - p_1(22)}{2} \right]^r \delta_1^{(n-r-1)}.
\]

Thus we have
\[
\begin{align*}
\delta_1^{(n)} &= \Delta_1^{(n)} + 2p_2(11)p_1^{(n)} \\
&= p_1^{(n)} p_2(11) + p_2^{(n)} p_1(22) \\
&= p_1^{(n-1)} p_2(11) + p_2^{(n-1)} p_1(22) + \frac{1}{2} \Delta_1^{(n-1)} (p_2(11) - p_1(22)) \\
&= \delta_1^{(n-1)} + \frac{1}{2} [p_2(11) - p_1(22)] \Delta_1^{(n-1)}.
\end{align*}
\]

Thus we have
\[
\begin{align*}
\delta_1^{(n)} - \delta_1^{(n-1)} &= \frac{1}{2} [p_2(11) - p_1(22)] \Delta_1^{(n-1)} \\
\delta_1^{(n-1)} - \delta_1^{(n-2)} &= \frac{1}{2} [p_2(11) - p_1(22)] \Delta_1^{(n-2)} \\
&\ldots \ldots \ldots \ldots \ldots \\
\delta_1^{(1)} - \delta_1^{(0)} &= \frac{1}{2} [p_2(11) - p_1(22)] \Delta_1^{(0)}.
\end{align*}
\]  
(5.65)

Adding all these equations, we have
\[
\delta_1^{(n)} = \delta_1^{(0)} + \frac{1}{2} [p_2(11) - p_1(22)] [\Delta_1^{(0)} + \Delta_1^{(1)} + \ldots + \Delta_1^{(n-1)}].
\]
\[ \delta_1^{(0)} + \frac{1}{2} [P_2(11) - P_1(22)] \left( 1 - \frac{P_1(11) + P_2(22)}{2} \right)^n \Delta_1^{(0)} = \delta_1^{(0)} + \frac{P_2(11) - P_1(22)}{P_2(11) + P_1(22)} \left( 1 - \frac{P_1(11) + P_2(22)}{2} \right)^n \Delta_1^{(0)} \]

\[ \delta_1^{(n)} = 2P_1^{(\infty)} P_2(11) - \frac{P_2(11) - P_1(22)}{P_2(11) + P_1(22)} \left( \frac{P_1(11) + P_2(22)}{2} \right)^n \Delta_1^{(0)} \]

\[ = 2P_2(11) P_1^{(\infty)} \quad \text{as } n \to \infty \quad \text{(5.66)} \]

Therefore,

\[ \delta_1^{(n-1)} + \left[ \frac{P_1(11) - P_1(22)}{2} \right] \delta_1^{(n-2)} + \left[ \frac{P_1(11) - P_1(22)}{2} \right]^2 \delta_1^{(n-3)} + \ldots + \left[ \frac{P_1(11) - P_1(22)}{2} \right]^{n-1} \delta_1^{(0)} = 2P_2(11) P_1^{(\infty)} \sum_{r=0}^{n-1} \left[ \frac{P_1(11) - P_1(22)}{2} \right]^r \]

\[ - \frac{P_2(11) - P_1(22)}{P_2(11) + P_1(22)} \Delta_1^{(0)} \sum_{r=0}^{n-1} \left[ \frac{P_1(11) - P_1(22)}{2} \right]^r x \left[ \frac{P_1(11) + P_2(22)}{2} \right]^{n-r-1} \]
Combining Equation 5.67 with Equation 5.64, we obtain the frequency of heterozygotes $2p_{12}^{(n)}$ in generation $n$ in terms of the initial values, where

$$p_{12}^{(n)} = \left[ \frac{p_{11}^{(1)} - p_{22}^{(2)}}{2} \right]^n p_{12}^{(0)}$$

$$+ \frac{2p_{21}^{(1)}p_{12}^{(0)}}{1 + p_{21}^{(1)} + p_{12}^{(2)}} \left[ 1 \right. - \left. \left\{ \frac{p_{11}^{(1)} - p_{22}^{(2)}}{2} \right\}^n \right]$$

$$- \frac{p_{21}^{(1)} - p_{12}^{(2)}}{p_{21}^{(1)} + p_{12}^{(2)}} \left[ \left\{ \frac{p_{11}^{(1)} + p_{22}^{(2)}}{2} \right\}^n \right.$$

$$- \left. \left\{ \frac{p_{11}^{(1)} - p_{22}^{(2)}}{2} \right\}^n \right] \Delta_1^{(0)} . \quad (5.68)$$
At equilibrium the heterozygotes are expected to be in the proportion $2P_{12}^{(\infty)}$ where

$$P_{12}^{(\infty)} = \frac{2P_{2}(11)P_{1}(12)}{1 + P_{2}(11) + P_{1}(22)}$$

$$= \frac{2P_{2}(11)P_{1}(22)}{[1 + P_{2}(11) + P_{1}(22)][P_{2}(11) + P_{1}(22)]} . \quad (5.69)$$

These results show that at equilibrium, the genic and genotypic distributions are entirely determined by the assortment parameters in terms of the conditional probabilities of male gametes for different types of females.

(ii) Suppose

$$P_{1}(11) - 2P_{1}(12) + P_{1}(22) = 0$$

and

$$\frac{P_{1}(11) + P_{2}(22)}{2} = 1 .$$

The second condition implies that $P_{1}(22) = 0$, and $P_{2}(11) = 0$. This together with the first condition leads to $P_{1}(12) = 1/2$ and $A_{1} = 0$. These conditions make the model very restrictive in that homozygous females ($A_{1}A_{1}$, $A_{2}A_{2}$) mate with their own type of males, whereas $A_{1}A_{2}$ females prefer to mate with all the three types of males subject to $P_{1}(12) = P_{2}(12) = 1/2$.

With the above restrictions, it can be easily seen
from Equations 5.57 and 5.58 that the gene frequency is invariant over time or equivalently
\[ p_1^{(n)} = p_1^{(n-1)} = \ldots = p_1^{(0)} \]
and
\[ p_{12}^{(n)} = (1/2)p_{12}^{(n-1)} \]
\[ = (1/2)^n p_{12}^{(0)} \to 0 \text{ as } n \to \infty . \] (5.70)

These results are similar to what we would expect under selfing or complete positive assortative mating without dominance. This development, however, covers more general situations in that, unlike under selfing or complete positive assortative mating, heterozygotes need not necessarily mate with heterozygotes alone as long as \( p_{12} = p_{21} = 1/2 \). For example, we may have the following conditional probabilities of mating:

<table>
<thead>
<tr>
<th>( \Phi ) genotype</th>
<th>( \mathbf{A}_1 \mathbf{A}_1 )</th>
<th>( \mathbf{A}_1 \mathbf{A}_2 )</th>
<th>( \mathbf{A}_2 \mathbf{A}_2 )</th>
<th>( p_1(ij) )</th>
<th>( p_2(ij) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbf{A}_1 \mathbf{A}_1 )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( \mathbf{A}_1 \mathbf{A}_2 )</td>
<td>.2</td>
<td>.6</td>
<td>.2</td>
<td>1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>( \mathbf{A}_2 \mathbf{A}_2 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

(iii) We now discuss the model in its general form without the restrictions that we have imposed in (b)(i), and (b)(ii). That is, we suppose
We shall investigate the behavior in equilibria which is obtained by setting \( P'_{12} = P_{12} \) and \( t' = t \) in Equations 5.57 and 5.58. This will lead to \( \Delta_1 = 0 \), at equilibrium, which gives

\[
P_{12} = \frac{P_2 P_{1(22)} - P_1 P_{2(11)}}{P_{1(11)} - 2P_{1(12)} + P_{1(22)}}.
\]  

(5.71)

Also from Equation 5.57

\[
P_{12} = P_{11} P_{2(11)} + P_{22} P_{1(22)}
\]

\[
= P_{2(11)} (P_1 - P_{12}) + P_{1(22)} (P_2 - P_{12})
\]

This will yield

\[
P_{12} [1 + P_{2(11)} + P_{1(22)}] = P_1 P_{2(11)} + P_2 P_{1(22)}
\]  

(5.72)

We have two equations in two unknowns, which determine the structure of the population. After some algebraic simplifications, this results in the nontrivial equilibrium which can be expressed as

\[
P_1^{(\infty)} = \frac{P_{1(22)} [P_{2(11)} + P_{1(12)}]}{2P_{1(22)} P_{2(11)} + P_{2(11)} P_{2(12)} + P_{1(12)} P_{1(22)}}
\]  

(5.73)

and
\[ 2p_{12}^{(\infty)} = 2p_1^{(\infty)} p_2^{(\infty)} (1 - \varnothing) \]  

(5.74)

where

\[ p_2^{(\infty)} = 1 - p_1^{(\infty)} \]

and

\[ (1 - \varnothing) = \frac{p_2(11)}{p_2(11) + p_1(12)} + \frac{p_1(22)}{p_2(12) + p_1(22)} \]  

(5.75)

This is an admissible equilibrium. We can see easily that 
\[ 0 < p_1^{(\infty)} < 1 \] and because the maximum value of \( 2p_1^{(\infty)} p_2^{(\infty)} \) is 
\[ 1/2 \] and \( 0 < 1 - \varnothing < 2 \), then \( 2p_{12}^{(\infty)} = 2p_1^{(\infty)} p_2^{(\infty)} (1 - \varnothing) \leq \frac{1}{2} (1 - \varnothing) < 1 \). This interior equilibrium will exist if \( p_1(22) \neq 0 \), 
and \( p_2(11) \neq 0 \), that is, if \( A_1A_1 \) and \( A_2A_2 \) females do not 
mate completely assortatively.

If \( p_1(12) = 0 \), i.e. if \( A_1A_2 \) females mate only with 
\( A_2A_2 \) males, the equilibrium will not depend upon the way 
in which \( A_1A_1 \) females mate provided \( p_2(11) \neq 0 \), because 
in this case we will have, at equilibrium

\[ p_1^{(\infty)} = \frac{p_1(22)}{1 + 2p_1(22)} \]  

(5.76)

\[ 2p_{12}^{(\infty)} = \frac{2p_1(22)}{1 + 2p_1(22)} = 2p_1^{(\infty)} \]  

(5.77)

This also shows that in equilibrium the population will con­
sist of only two types of individuals \( A_1A_2 \) and \( A_2A_2 \) in
proportions $2p_1^{(\infty)}$ and $1 - 2p_1^{(\infty)}$ respectively.

We consider now some special cases.

Case 1: Let $p_{1(22)} = 0$ and $p_{2(11)} \neq 0$. This results in the recursion relations

$$P'_{11} = p_{1(11)}P_{11} + p_{1(12)}P_{12}$$

$$2P'_{12} = p_{2(11)}P_{11} + P_{12}$$

$$P'_{22} = p_{2(12)}P_{12} + P_{22}$$

The equilibrium distribution is obtained as before by setting $P'_{ij} = P_{ij}$ for all $(ij)$ which yields

$$p_{2(11)}P_{11} = p_{1(12)}P_{12}$$

$$P_{12} = p_{2(11)}P_{11}$$

$$p_{2(12)}P_{12} = 0$$

If $p_{2(12)} \neq 0$, then $p_{12}^{(\infty)} = 0$, $p_{11}^{(\infty)} = 0$, and $p_{22}^{(\infty)} = 1$. Thus, in this case there is only the trivial equilibrium which is the fixation of allele $A_2$.

If $p_{2(12)} = 0$, then $P'_{22} = P_{22}'$, that is, the frequency of $A_2A_2$ is invariant. Therefore, at equilibrium

$$p_{22}^{(\infty)} = p_{22}$$

$$p_{12}^{(\infty)} = p_{2(11)}p_{11}^{(\infty)}$$
This gives

\[ p_{12}^{(\infty)} = \frac{p_{2}(11)[1 - p_{22}^{(0)}]}{1 + 2p_{2}(11)} \]  

(5.81)

This polymorphic equilibrium is dependent on the initial structure of the population in addition to the assortment parameters.

Case 2: Let \( p_{2}(11) = 0 \), \( p_{1}(22) = 0 \), that is homozygous females mate only with their own kind. In this case the recursion formulas are greatly simplified and we can discuss the complete time dependent behavior. The Equations 5.57 and 5.59 yield

\[ p_{12}^{(n)} = \left(\frac{1}{2}\right) p_{12}^{(n-1)} \]

\[ = \left(\frac{1}{2}\right)^n p_{12}^{(0)} \]  

(5.82)

\[ \Delta_{1}^{(n)} = - [1 - 2p_{1}(12)] p_{12}^{(n)} \]

\[ = \left(\frac{1}{2}\right) \Delta_{1}^{(n-1)} \]

\[ = \left(\frac{1}{2}\right)^n \Delta_{1}^{(0)} \]  

(5.83)
where

\[ \Delta_1^{(0)} = - [1 - 2p_{1(12)}] p_{12}^{(0)} . \]

As \( n \to \infty \), \( p_{12}^{(n)} \to 0 \), and \( \Delta_1^{(n)} \to 0 \) at the geometric rate of 1/2.

From Equation 5.58 we can write

\[ p_1^{(n)} - p_1^{(n-1)} = \frac{1}{2} \Delta_1^{(n-1)} \]

\[ p_1^{(n-1)} - p_1^{(n-2)} = \frac{1}{2} \Delta_1^{(n-2)} \]

\[ \vdots \]

\[ p_1^{(1)} - p_1^{(0)} = \frac{1}{2} \Delta_1^{(0)} . \]

By adding all these equations, we obtain

\[ p_1^{(n)} = p_1^{(0)} + \frac{1}{2} \sum_{r=0}^{n-1} \Delta_1^{(r)} \]

\[ = p_1^{(0)} + [1 - (\frac{1}{2})^n] \Delta_1^{(0)} . \] (5.84)

This shows that the gene frequency changes from generation to generation and \( p_1^{(n)} \to p_1^{(0)} + \Delta_1^{(0)} \) as \( n \to \infty \), that is

\[ p_1^{(\infty)} = p_1^{(0)} + \Delta_1^{(0)} \]

\[ = p_1^{(0)} + [p_{1(12)} - p_{2(12)}] p_{12}^{(0)} . \] (5.85)

Thus, we have in generation \( n \)
\[ p_{12}^{(n)} = \left( \frac{1}{2} \right)^n p_{12} \to 0 \quad \text{as } n \to \infty \]

\[ p_{11}^{(n)} = p_1 - p_{12}^{(n)} \to p_1^{(\infty)} \quad \text{as } n \to \infty \]

\[ p_{22}^{(n)} = p_2 - p_{12}^{(n)} \to p_2^{(\infty)} \quad \text{as } n \to \infty \]  \hspace{1cm} (5.86)

where \( p_1^{(n)} \) and \( p_1^{(\infty)} \) are given by Equations 5.84 and 5.85.

B. Components of Variance and Covariances of Relatives

Another important feature of the composition of the population resulting from a given system of mating is its variability. It is meaningful sometimes to further partition the variance into components like additive variance and dominance variance, and also to look at the covariances of relatives. We shall discuss such properties in this section for populations in equilibrium for some of the models. Since we shall be dealing with equilibrium populations, we denote, for convenience, the frequency of an ordered genotype \( A_i A_j \) by \( p_{ij} \) and drop the superscript \( \infty \) used earlier. We treat the gene frequencies in the similar way.

1. Karlin and Scudo (1969) asymmetric model

We consider here the effect of assortment based on a pair of alleles \( A_1 \) and \( A_2 \) where the heterozygote can be distinguished from both homozygotes, that is there is no dominance. Let the degrees of assortment for the three
genotypes $A_1A_1$, $A_1A_2$, and $A_2A_2$ be described by $\alpha$, $\beta$, and $\gamma$ respectively. That is, a fraction $\alpha$ of the females of genotypes $A_1A_1$ tends to mate with their own type while the remaining fraction does not have any preference and mates at random. $\beta$ and $\gamma$ are defined in a similar way.

In the symmetric case $\alpha = \gamma < \beta$, the interior equilibrium obtained by Karlin and Scudo (1969) simplifies to be

$$P_{11} = P_{22} = \frac{1}{2(2 - \alpha)}$$

$$2P_{12} = \frac{1 - \alpha}{2 - \alpha}$$

with equal gene frequencies of one-half.

Let the genotypic values of the three genotypes $A_1A_1$, $A_1A_2$, and $A_2A_2$ be 2, 1, and 0 respectively. Then the genotypic mean ($\mu$) and variance ($\nu$) are 1 and $1/(2 - \alpha)$ respectively.

We now turn to the calculation of covariances between relatives. First we construct an inter-parental array as in Table 13. We find

$$\text{Cov}(P_{F}, O) = P_{11}(3 + \alpha) + 2P_{12}\mu^2$$

$$= \frac{1 + \alpha}{2(2 - \alpha)}$$

$$= \frac{1}{2}(1 + \alpha)\nu$$

(5.87)

where $P_{F}$ stands for the female parent. Since any between-
Table 13. Distribution of mating types

<table>
<thead>
<tr>
<th>$\varphi$</th>
<th>$O^*$</th>
<th>$A_1A_1$</th>
<th>$A_1A_2$</th>
<th>$A_2A_2$</th>
<th>Progeny mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1$</td>
<td></td>
<td>$\alpha P_{11} + (1 - \alpha)P_{12}^2$</td>
<td>$2(1 - \alpha)P_{11}P_{12}$</td>
<td>$(1 - \alpha)P_{11}P_{22}$</td>
<td>$\frac{3 + \alpha}{2}$</td>
</tr>
<tr>
<td>$A_1A_2$</td>
<td></td>
<td>$2(1 - \beta)P_{11}P_{12}$</td>
<td>$2\beta P_{12} + 4(1 - \beta)P_{12}^2$</td>
<td>$2(1 - \beta)P_{12}P_{22}$</td>
<td>1</td>
</tr>
<tr>
<td>$A_2A_2$</td>
<td></td>
<td>$(1 - \alpha)P_{11}P_{22}$</td>
<td>$2(1 - \alpha)P_{12}P_{22}$</td>
<td>$\alpha P_{22} + (1 - \alpha)P_{22}^2$</td>
<td>$\frac{1 - \alpha}{2}$</td>
</tr>
</tbody>
</table>
group variance component is equal to the covariance of the members of the group, it follows that

\[ \text{Cov(F.S.)} = \left[ aP_{11} + (1 - a)P_{22}^2 \right] 4 + 2(1 - a)P_{11}P_{12}(9/4) \]
\[ + (1 - a)P_{11}P_{22} + 2(1 - \beta)P_{11}P_{12}(9/4) \]
\[ + \left[ 2\beta P_{12} + 4(1 - \beta)P_{12}^2 \right] + 2(1 - \beta)P_{12}P_{22}(1/4) \]
\[ + (1 - a)P_{11}P_{22} + 2(1 - a)P_{12}P_{22}(1/4) - \mu^2 . \]

This, after simplification, can be expressed as

\[ \text{Cov(F.S.)} = \frac{1}{4(2 - \alpha)^2} \left[ 4 + (1 - a)(3\alpha - \beta) \right] \]
\[ = \left[ \frac{1}{2 - \alpha} + \frac{(1 - a)(3\alpha - \beta)}{4(2 - \alpha)} \right] \nu . \quad (5.88) \]

Similarly,

\[ \text{Cov(Maternal H.S.)} \]
\[ = P_{11}\left( \frac{3 + a}{2} \right)^2 + 2P_{12}\left( \frac{1 - a}{2} \right)^2 - 1^2 \]
\[ = \frac{(1 + a)^2}{4(2 - \alpha)} \]
\[ = (\frac{1 + a}{2})^2 \nu . \quad (5.89) \]

Further

\[ \text{Cov(Maternal H.S.)}/\text{Cov(P,F,0)} = (1 + a)/2 . \quad (5.90) \]

The results show that with this system of mating, the
equilibrium distribution is independent of the parameter $\beta$ and so are covariances between parent-offspring and between maternal half-sibs. But the covariance of full-sibs depends upon $\beta$.

2. Model III(a)

The equilibrium frequencies in this case are given from Equations 5.56 to be

$$p_{ij} = q_i q_j \quad \text{for } i,j=1,2,\ldots,s$$

This has Hardy-Weinberg structure. Therefore, the additive and dominance components of variance are the same as in a random mating population. Let $y_{ij}$ be the genotypic value of the genotype $A_iA_j$ measured from the mean. Then we may write

$$y_{ij} = a_i + a_j + \delta_{ij}$$

such that

$$\sum_i \sum_j p_{ij} y_{ij} = 0$$

and $a_i = \text{effect of gene } A_i = \sum_j q_j y_{ij}$. Therefore,

$$v = \sum_i \sum_j q_i q_j y_{ij}^2$$

$$v_A = 2 \sum_i q_i a_i^2$$
where \( V \) is the total genotypic variance, and \( V_A \) and \( V_D \) are the additive and dominance components of variance (Kempthorne, 1957, p. 319). Although the structure of the population is panmictic, the covariances between relatives need not be the same as in a random mating population because the mating structure here is different. These are obtained as follows:

\[
\text{Cov}(P_{ij}, 0) = \sum_{ij} P_{ij} \gamma_{ij} \sum_{rt} P_{rt(ij)} [(1/4)(\gamma_{ir} + \gamma_{it} + \gamma_{jr} + \gamma_{jt})]
\]

\[
= (1/2) \sum_{ij} q_i q_j \gamma_{ij} \sum_{r} q_r (\gamma_{ir} + \gamma_{jr})
\]

\[
= (1/2) \sum_{ij} q_i q_j \gamma_{ij} (a_i + a_j)
\]

\[
= (1/2)V_A
\]

(5.92)

because we have assumed in this model that \( P_{r(ij)} = q_r \) for all \((ij)\).

\[
\text{Cov}(F.S.) = \sum_{ij} \sum_{rt} P_{ij} P_{rt(ij)} [(1/4)(\gamma_{ir} + \gamma_{jr} + \gamma_{it} + \gamma_{jt})]^2
\]

\[
= \sum_{ij} \sum_{rt} P_{ij} P_{rt(ij)} [(1/2)(a_i + a_j + a_r + a_t) + (1/4)(\delta_{ir} + \delta_{jr} + \delta_{it} + \delta_{jt})]^2
\]

After some simplification we can write this as

\[
\text{Cov}(F.S.) = \frac{1}{2} V_A + \frac{1}{4} V_D + \frac{1}{4} \sum_{ij} q_i q_j \sum_{rt} P_{rt(ij)}
\]
\[ x \left[ 2a_r(a_t + \delta_{it} + \delta_{jt}) + \delta_{ir}(\delta_{it} + \delta_{jt}) \right]. \]

(5.93)

In the simplification of the above results we have used the facts that \( E(a_i) = 0, E(\delta_{ij}) = 0, \sum_i q_i \delta_{ij} = 0, E(a_i \delta_{ij}) = 0 \) which are well known for the random mating populations.

Similarly,

\[
\text{Cov(Maternal H.S.)} = \sum_{ij} p_{ij} \left[ \frac{1}{8} \sum_{rt} p_{rt(ij)}(y_{ir} + y_{jr} + y_{it} + y_{jt}) \right]^2
\]

\[
= \sum_{ij} q_i q_j \left[ \frac{a_i + a_j}{2} \right]^2
\]

\[
= \frac{1}{4} V_A. \quad (5.94)
\]

These results show that with this system of mating the covariances of parent-offspring and of maternal half-sibs are the same as would be expected for a random mating population which is not surprising. However, in the case of covariance of full-sibs there arises an additional term besides the usual expression for such a covariance under pure random mating.

3. Model III(b)(i)

This model is developed for \( s = 2 \) under the assumptions that

\[ P_1(11) - 2P_1(12) + P_1(22) = 0 \]
The equilibrium frequencies in this case are given by Equations 5.62 and 5.69 and are functions of the assortment parameters only. These can be expressed as

\[
\begin{align*}
\bar{P}_{11} &= \bar{p}_1^2 + \bar{p}_1 \bar{p}_2 \varnothing \\
\bar{P}_{12} &= \bar{p}_1 \bar{p}_2 (1 - \varnothing) \\
\bar{P}_{22} &= \bar{p}_2^2 + \bar{p}_1 \bar{p}_2 \varnothing
\end{align*}
\]  

where \( \bar{p}_1, \bar{p}_2, \) and \( \varnothing \) are all functions of the assortment parameters only and are given by

\[
\begin{align*}
\bar{p}_1 &= \bar{p}_1(22)/[\bar{p}_1(22) + \bar{p}_2(11)] \\
\bar{p}_2 &= 1 - \bar{p}_1 \\
\varnothing &= [\bar{p}_1(11) - \bar{p}_1(22)]/[1 + \bar{p}_2(11) + \bar{p}_1(22)]
\end{align*}
\]  

Notice that \( \varnothing \) could be positive or negative.

The additive and dominance variances can be obtained by using the known formula given by Kempthorne (1957, p. 367) since the population structure is similar to that of an inbred population. Hence, using the coded genotypic values, \( u, (u-v)/2, \) and \( 0 \) for the three genotypes \( A_1A_1, A_1A_2, \) and \( A_2A_2 \) respectively, we get
\[ V_A = \frac{p_1 p_2}{1 + p_2(11) + p_1(22)} \left[ u + 2(p_1 - p_2)(p_2(11) + p_1(22)v) \right]^2 \]

\[ V_D = \frac{4p_1 p_2 p_1(12)p_2(12)[p_2(11) + p_1(22)]}{[1 + p_2(11) + p_1(22)]^2} v^2 \]

where \( p_1 \) and \( p_2 \) are the equilibrium gene frequencies and are entirely dependent upon the mating parameters as can be seen from Equations 5.96, and \( u = y_{11} - y_{22} \) and \( v = y_{11} - 2y_{12} + y_{22} \).

The progeny means of the female parents \( A_1A_1, A_1A_2, \) and \( A_2A_2 \) are \( p_1(11)u + p_2(11)(u-v)/2, p_1(12)u/2 + (u-v)/4, \) and \( p_1(22)(u-v)/2 \) respectively. Therefore, the parent-offspring covariance can be expressed, after simplification, as

\[ \text{Cov}(P_F, O) = \frac{p_1(11) + p_2(22)}{2} V_A + \frac{p_1(11) - p_1(22)}{2} V_D \]  

(5.97)

We can similarly develop formulas for \( \text{Cov}(F.S.) \) and \( \text{Cov}(H.S.) \) but there do not arise any nice expressions.

C. Two Loci Model: Complete Positive Assortative Mating

Consider the case of complete positive assortative mating with respect to a character which depends upon two factors. We shall assume that there is no dominance at either locus and that the effects of the two factors are equal and additive. Let us further assume that the population is in
Hardy-Weinberg equilibrium before the beginning of assortative mating. Wright (1921) discussed this problem with equal gene frequency of one-half for each allele. He concluded that assortative mating based on resemblance leads to a composition of the population very different from that reached by inbreeding. With perfect assortative mating, a two-factor population is converted ultimately into only two extreme types $A_1A_1B_1B_1$ and $A_2A_2B_2B_2$. With inbreeding, all the four homozygous types tend equally toward fixation.

Obviously, Wright's conclusion was based on the assumption of equal gene frequencies. But this result is often quoted without mentioning this assumption (Li, 1955, p. 237; Crow and Felsenstein, 1968) and without realizing that this result may not be valid in a more general situation where this assumption of equal gene frequency is not fulfilled. For example, Li (1955, p. 237) states that "...for a metric character dependent on two pairs of genes with additive and equal effects (Ch. 8, Ex. 5) complete assortative mating within each of the five phenotypes would ultimately lead to a population consisting of the two extreme types...." That this result is not true in a general situation and in particular in the case of Ex. 5 (Ch. 8) of Li, can be easily detected if one keeps in mind the fact that with this type of mating system gene frequencies remain invariant over time. However, as we shall show, such a result will hold
if \( P(B_1) = P(A_1) \), not necessarily one-half.

With two alleles at each locus, there are five phenotypes on the basis of which assortative mating is to be made. Their distribution is given in Table 14 where \( p_1 \) and \( q_1 \) are the frequencies of genes \( A_1 \) and \( B_1 \) respectively, and \( p_1 + p_2 = 1 \) and \( q_1 + q_2 = 1 \).

With this system of mating, heterozygosity will be lost ultimately. Let us suppose that all the four homozygous types are present in the population in equilibrium. Let \( X, Y, Z \) and \( W \) be the frequencies of \( A_1A_1B_1B_1 \), \( A_1A_1B_2B_2 \), \( A_2A_2B_1B_1 \), and \( A_2A_2B_2B_2 \) respectively such that \( X + Y + Z + W = 1 \). There will be three phenotypic classes \( (A_1A_1B_1B_1) \), \( (A_1A_1B_2B_2) \), \( (A_2A_2B_1B_1) \) and \( (A_2A_2B_2B_2) \) with corresponding frequencies of \( X, Y + Z, \) and \( W \). With complete assortative mating within each of these phenotypic classes we have in the next generation,

\[
\begin{align*}
X' &= X \\
Y' &= Y^2/(Y + Z) \\
Z' &= Z^2/(Y + Z) \\
W' &= W 
\end{align*}
\]

and

\[
P'[A_1A_2B_1B_2] = 2YZ/(Y + Z) \quad .
\]
Table 14. Distribution of phenotypes in the initial population

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Value</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1B_1B_1$</td>
<td>4</td>
<td>$p_1^2q_1^2$</td>
</tr>
<tr>
<td>$A_1A_2B_1B_1, A_1A_1B_1B_2$</td>
<td>3</td>
<td>$2p_1q_1(p_2q_1 + p_1q_2)$</td>
</tr>
<tr>
<td>$A_1A_2B_2B_2, A_1A_2B_1B_2, A_2A_2B_1B_1$</td>
<td>2</td>
<td>$p_1^2q_2^2 + 4p_1p_2q_1q_2$ + $p_2^2q_1^2$</td>
</tr>
<tr>
<td>$A_1A_2B_2B_2, A_2A_2B_1B_2$</td>
<td>1</td>
<td>$2p_2q_2(p_1q_2 + p_2q_1)$</td>
</tr>
<tr>
<td>$A_2A_2B_2B_2$</td>
<td>0</td>
<td>$p_2^2q_2^2$</td>
</tr>
</tbody>
</table>

Since the population is in equilibrium, we have

$$YZ = 0 \quad .$$  \hspace{1cm} (5.99)

Also, we have

$$X + Y = p_1 \quad \quad X + Z = q_1$$
$$Z + W = p_2 \quad \quad Y + W = q_2$$  \hspace{1cm} (5.100)

which do not change with time. From Equations 5.100 we get

$$Y - Z = p_1 - q_1 \neq 0 \quad .$$  \hspace{1cm} (5.101)
Equations 5.99 and 5.101 imply that either \( Y = 0 \) or \( Z = 0 \) but not both zero because \( Y - Z \neq 0 \). This shows that at equilibrium only three homozygous types can be present in the population. In addition it also shows that the population is not composed of only two homozygous genotypes as is usually thought, because \( Y \) and \( Z \) cannot both be zero simultaneously. Now consider Equation 5.101 again. If \( p_1 > q_1 \), then Equation 5.99 yields that \( Z = 0 \) because \( Y > Z \). Similarly, if \( p_1 < q_1 \), then \( Y = 0 \). We can determine the frequencies of other genotypes by using Equations 5.100. For example, if \( p_1 > q_1 \), then \( Z = 0, X = q_1, W = p_2 \) and \( Y = q_2 - p_2 \). Therefore, the three types of phenotypes or genotypes that will be present in an equilibrium population will depend upon the initial gene frequencies. Thus, if \( p_1 \neq q_1 \), this mating system will result in three homozygous genotypes in a population in equilibrium instead of the two extreme types with the frequencies as given in Table 15.

It can also be seen that if \( p_1 = q_1, p_2 = q_2 \), there would be only two extreme types with frequencies \( p_1 \) and \( p_2 \).
Table 15. Genotypic distribution in equilibrium

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>$p_1 &lt; q_1$</th>
<th>$p_1 &gt; q_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1A_1B_1B_1$</td>
<td>$p_1$</td>
<td>$q_1$</td>
<td></td>
</tr>
<tr>
<td>$A_1A_2B_2B_2$</td>
<td>0</td>
<td>$p_1 - q_1$</td>
<td></td>
</tr>
<tr>
<td>$A_2A_2B_1B_1$</td>
<td>$q_1 - p_1$</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$A_2A_2B_2B_2$</td>
<td>$q_2$</td>
<td>$p_2$</td>
<td></td>
</tr>
</tbody>
</table>
VI. SUMMARY AND CONCLUSIONS

The present study deals with the effect of nonrandom mating due to (a) partial inbreeding, and (b) partial assortative mating, on various statistical properties of a population. These systems of mating are important because it has been observed that situations are not uncommon where outcrossing and inbreeding processes are working simultaneously in varying degrees, or where there is a deviation from complete positive or negative assortative mating. In developing the mathematical theory we have assumed that the populations are large enough so that fluctuations due to sampling can be ignored, and that there are no viability or fertility differences.

The mathematical theory of partial inbreeding has largely remained a one locus theory for a system of mixed random mating and selfing. We have developed theoretical models for populations in which some of the matings occur at random while the remainder take place between relatives of various degrees. These include mixed random mating and (i) selfing, (ii) full- and half-sib mating, (iii) parent-offspring mating, (iv) double first cousin mating, and (v) a general mixture of consanguineous mating systems.

For each system of mating, general expressions are given from which it is possible to calculate the frequency
of any genotype in any generation, and at equilibrium, in terms of the initial frequencies and the parameter of partial inbreeding. These results are derived assuming an initial panmictic population for a single locus. In the case of mixed random mating and selfing some results are also obtained for an arbitrary initial population.

The mathematical analysis shows that the systems of mixed random and parent-offspring mating, and mixed random and full-sib mating yield identical recurrence relations. Therefore, these two systems would lead to the same genotypic distributions. The systems of mixed random and half-sib mating and mixed random and double first cousin mating result in different genotypic distributions in a dynamic population. But these distributions coincide at equilibrium.

An important feature of the results is the considerable amount of heterozygosity that these systems can maintain in predominantly inbred populations. This is because there are two opposing forces acting on the population, inbreeding acting to reduce the heterozygosity and random mating acting to restore the initial values. Ultimately a stage is reached where loss due to one is balanced by the gain due to the other and the population reaches an equilibrium state with a certain amount of heterozygosity. The level of heterozygosity in a population depends upon the system of mixed mating, amount of inbreeding and the initial
heterozygosity. The relative effects of these systems on the maintenance of heterozygosity at equilibrium are discussed.

The effects of such systems on the genotypic mean and variance of a population are also discussed. In the case of complete dominance, the genotypic variance increases when the frequency of recessive individuals in an initial population is less than or equal to one quarter \( (p_2^2 \leq 1/4) \) but it decreases when it is at least as high as that of the dominants \( (p_2^2 \geq 1/2) \). For other values there is no specific trend. The change in variance also depends upon the form of inbreeding. The more intense the form of inbreeding the larger is the increase or decrease in the variance.

The system of mixed random mating and selfing has been examined in great detail. The results are extended to several unlinked loci with multiple alleles in an initial panmictic population. Expressions are obtained for the distribution of genotypes, and the genotypic mean and variance, under the assumption that the effects of different loci are additive. The genotypes at different loci are not distributed independently. Therefore, in the formula for the variance, in addition to the sum of variances due to each factor separately, there appears another term which corresponds to the sum of covariances between factors taken in pairs. This covariance term depends upon the dominance
deviations, probability of selfing, $\beta$, and the gene frequencies, and therefore vanishes in the absence of dominance or when $\beta = 0$ or $1$. It is, therefore, evident that the trend of changes in the variance due to several factors cannot be predicted from the behavior of a single factor.

The situation with assortative mating is very complex because this tendency to mate assortatively can be expressed in several ways. We have considered a few asymmetric deterministic models for a single locus with two alleles and in some cases with the multiple alleles. The models are developed by specifying the intensities for preference for mating for various phenotypes. It is assumed that preference to choose a mate lies only in one sex, say females. These models allow for an excess of males so that assorting does not decrease the chance of a female to be fertilized; and also take into account that females may have different preferences for different types of males.

The models are analyzed with respect to the equilibrium behavior of a population under some simplifying assumptions. In some cases, the complete time dependent behavior of a population is also discussed. The gene frequency usually changes with time and the equilibrium distribution depends only on the assortment parameters. Expressions are obtained giving the additive and dominance components of variance, and covariances between relatives in equilibrium populations.
for some of the models.

A two loci model for complete positive assortative mating is also discussed. This is more general than that of Wright (1921). If the gene frequencies for the two loci are different, as they are likely to be, it is shown that in equilibrium the population is not composed of only two homozygous types as is usually thought. There will be three homozygous genotypes present in an equilibrium population depending upon the initial gene frequencies.
VII. BIBLIOGRAPHY


VIII. ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to Professor Oscar Kempthorne for his constant encouragement, understanding and guidance throughout.

Thanks are due to Dr. Edward Pollak who read parts of the preliminary draft of the manuscript. The author also wishes to express his appreciation to Drs. T. A. Bancroft, L. N. Hazel and J. C. Kegley, the other members of his graduate committee, for their encouragement and help.

The assistance of Mr. T. Emigh in computer programming and preparing some of the tables is gratefully acknowledged.

The author wishes to thank Mrs. Marlys Phipps for her skillful preparation of the typewritten manuscript.

The research was supported by the National Institute of Health grant number GM 13827. This is gratefully acknowledged.