77-2820

YAMASHITA, Toyoko S., 1944-
APPLICATIONS OF DISTRIBUTION THEORY IN QUANTITATIVE GENETICS.

University of Hawaii, Ph.D., 1976
Genetics

Xerox University Microfilms, Ann Arbor, Michigan 48106

APPLICATIONS OF DISTRIBUTION THEORY

IN QUANTITATIVE GENETICS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN GENETICS

AUGUST 1976

By

Toyoko S. Yamashita

Dissertation Committee:

Ming-Pi Mi, Chairman
Geoffrey C. Ashton
Chin Sik Chung
Mohamed N. Rashad
Dale W. Vogt

ACKNOWLEDGEMENT

In writing this dissertation, I have received help and guidance from several persons at the University of Hawaii and outside. While it is not possible to mention all of them, I would like to acknowledge the invaluable assistance accorded to me by Professor W. Y. Tan of Memphis State University. His distribution theory was applied in this research and his critical review of my work has been of great value. The members of the Data and Computation Unit, Cancer Center of Hawaii, have extended their helping hands whenever I needed them.

ABSTRACT

The present study was designed to develop procedures for estimating genetic parameters in a quantitative trait and detecting association between a quantitative trait and a genetic marker using a random sample of individuals from the population. An exact distribution of phenotypic values of a quantitative trait was used by which parameters such as number of loci involved, gene frequency at each locus, additive and dominant genotypic value, and the variance of environmental effects were estimated. When association between the quantitative trait and a genetic marker occurred, an additional parameter for the degree of linkage disequilibrium was introduced in the exact distribution. For testing the validity and usefulness of the procedures, individual phenotypic values were simulated by the computer with given parameters. The number of loci involved in the quantitative trait was 2, 5, and 10, and gene frequency at each locus was chosen at random. Three heritability values were investigated; namely, 0.2, 0.5, and 0.8. The genotypic values were then calculated by setting degree of the dominance either to zero (no dominance) or one (complete dominance). The environmental effects were assumed to be normally distributed; however, the phenotypic distribution of the quantitative trait under investigation was not specified. Five replicate runs were made for most tests to allow variations in gene frequencies, each sample size being set to 5000. Smaller samples of 100 and 500 individuals were also tested.

Based on the log likelihood function values, good estimations by the maximum likelihood method were obtained when heritability of the quantitative trait was moderately high ($h^2 \geq .5$). When the heritability value was low, the phenotypic distribution of the quantitative trait was mainly determined by environmental effects and, therefore, was expected to be normal by assumption. In this case, the method might not give consistent estimates of genetic parameters. As the number of loci increased, the phenotypic distribution would approach normal by the Central Limit Theorem; however, the procedure was still effective for traits with high heritability.

Several cases for the association between one of the loci involved in the quantitative trait and a genetic marker were considered, including (1) complete independence (no linkage), (2) partial linkage, and (3) complete linkage or identical locus. The approach employed was mainly based on the distributions of phenotypic values. Data were subgrouped by the genotypes of the marker locus resulting in several phenotypic distributions, one for each genotype of the genetic marker. When association occurred between a quantitative trait locus and a genetic marker locus, the subclass distributions would be expected to be different from one another. The subsequent estimates of parameters including the degree of linkage disequilibrium were then derived from the exact distributions by the maximum likelihood method. The results were generally good.

The present study using simulated data has clearly demonstrated that by employing the distribution of phenotypic values of the quantitative trait, it is possible to establish a genetic basis for the character from a random sample of individual data. When information on other genetic markers is available, the association between the quantitative trait and a genetic marker as measured by linkage disequilibrium may be estimated. However, further research would be needed to apply the procedures to real (non-simulated) data and to consider other parameters such as genotype-environmental interactions and epistatic effects in the phenotypic distribution.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

ABSTRACT iv

LIST OF TABLES ix

LIST OF FIGURES x

1. INTRODUCTION 1

2. REVIEW OF LITERATURE 3

3. THEORIES AND APPLICATIONS 9

3.1 A Model for Quantitative Traits 9

3.2 Distribution of Phenotypic Values 13

3.2.1 Single Locus (n = 1) 13

3.2.2 Two Loci (n = 2) 14

3.2.3 Generalization for n Loci ($n \geq 1$) 16

3.3 Models for Association Between a Quantitative Character Locus and a Genetic Marker Locus 19

3.3.1 Model-1: G_n-locus and M-locus are Independent. . 19

3.3.2 Model-2: G_n-locus and M-locus are Partially Linked 21

3.3.3 Model-3: G_n-locus and M-locus are Completely Linked or Identical Loci 26

3.4 Investigations of Phenotypic Distributions 27

3.5 Generation of Test Data 29

3.6 Estimation of Parameters Using Phenotypic Measurements Only 30

3.7 Estimation of Parameters Using Phenotypic Measurements and Genetic Marker Genotype Information 35

4. INTERPRETATION OF RESULTS 38

4.1 Distribution of Phenotypic Values 38

4.2 Descriptive Statistics of Simulated Data 61

4.3 Estimation of Genetic Parameters Based on Phenotypic Values Only 72

4.4 Estimation of Parameters Using Genetic Marker Information 84

4.5 General Discussion 129

APPENDIX I. Computational Formula for $W_{uv}^{(n)}$ 132

APPENDIX II. Mean and Variance of Exact Distribution 133

APPENDIX III. Estimation of a and d given σ_g^2, x and p_i for $i = 1$ to n 137

APPENDIX IV. Initial Estimate of Degree of Dominance (D) 138

APPENDIX V. Coefficients of Skewness and Kurtosis 141

LITERATURE CITED 142

LIST OF TABLES

Table		Page
3.2.2	The Genotypic Frequencies and Values for $n = 2$	15
3.2.3	Values of $W_{uv}^{(1)}$ and $W_{uv}^{(2)}$	18
3.3.1	Genotypic Values at G_n-locus Under Model-1	20
3.3.2	Expected Genotypic Frequencies Under Model-2	22
3.3.3	Genotypic Values at G_n-locus Under Model-3	26
3.5	Parameter Values for h^2, n, x, D^* and N	31
4.1.1	Parameter Values for Exact Distribution	58
4.1.2	Genotypic Values and Genotypic Frequencies	60
4.2.1	Descriptive Statistics	63
4.3.1	Initial Values	76
4.3.2	log Likelihood Function Values (log-L)	81
4.3.3	Results of Consistency Test	82
4.4.1	Statistics of Subclasses	89
4.4.2	Results of Association Test	101
4.4.3	Estimates for $\hat{D}$, $\hat{D}^*$ and $\hat{p}_n$	103

LIST OF FIGURES

Figure		Page
3.1.1	Model of a Phenotypic Value When $n = 1$	10
3.1.2	The Model of a Phenotypic Value	12
3.5	Generation for a Set of Sample Data	32
4.1.1	Phenotypic Distribution ($h^2 = .8$)	43
4.1.2	Phenotypic Distribution ($h^2 = .2$)	44
4.1.3	Phenotypic Distribution ($x = 1$).	45
4.1.4	Phenotypic Distribution ($p = .8$)	46
4.1.5	Phenotypic Distributions ($h^2 = .2$, $x = 0$).	47
4.1.6	Phenotypic Distributions ($h^2 = .2$, $x = .5$)	48
4.1.7	Phenotypic Distributions ($h^2 = .2$, $x = 1$).	49
4.1.8	Phenotypic Distributions ($h^2 = .5$, $x = 0$).	50
4.1.9	Phenotypic Distributions ($h^2 = .5$, $x = .5$)	51
4.1.10	Phenotypic Distributions ($h^2 = .5$, $x = 1$).	52
4.1.11	Phenotypic Distributions ($h^2 = .8$, $x = 0$).	53
4.1.12	Phenotypic Distributions ($h^2 = .8$, $x = .5$)	54
4.1.13	Phenotypic Distributions ($h^2 = .8$, $x = 1$).	55
4.1.14	Phenotypic Distributions ($n = 2$)	56
4.1.15	Phenotypic Distributions (different n)	57
4.2.1	Observed Phenotypic Distributions	70
4.2.2	Variations of Observed Phenotypic Distributions . . .	71
4.3	Histogram of Data 1-4.	83
4.4.1	Subclass Distributions (data 1-1).	104
4.4.2	Subclass Distributions (data 2-1).	105
4.4.3	Subclass Distributions (data 3-1).	106

Figure		Page
4.4.4	Subclass Distributions (data 4-1)	107
4.4.5	Subclass Distributions (data 5-1)	108
4.4.6	Subclass Distributions (data 6-1)	109
4.4.7	Subclass Distributions (data 7-1)	110
4.4.8	Subclass Distributions (data 8-1)	111
4.4.9	Subclass Distributions (data 9-1)	112
4.4.10	Subclass Distributions (data 10-1)	113
4.4.11	Subclass Distributions (data 11-1)	114
4.4.12	Subclass Distributions (data 12-1)	115
4.4.13	Subclass Distributions (data 13-1)	116
4.4.14	Subclass Distributions (data 14-1)	117
4.4.15	Subclass Distributions (data 15-1)	118
4.4.16	Subclass Distributions (data 16-1)	119
4.4.17	Subclass Distributions (data 17-1)	120
4.4.18	Subclass Distributions (data 18-1)	121
4.4.19	Subclass Distributions (data 19-1)	122
4.4.20	Subclass Distributions (data 20-1)	123
4.4.21	Subclass Distributions (data 21-1)	124
4.4.22	Subclass Distributions (data 22-1)	125
4.4.23	Subclass Distributions (data 23-1)	126
4.4.24	Subclass Distributions (data 24-1)	127
4.4.25	Subclass Distributions (data 9-2)	128

1. INTRODUCTION

Most studies and methodologic developments in the area of quantitative genetics, particularly in mammals, have involved the use of family data such as siblings, twins, and parent-offspring pairs, or data of several generations such as from selection experiments. There are several methods for estimating heritability of a trait, the number of loci and the genotypic values. The estimation of heritability is based on correlation or regression analysis of parent-offspring pairs, and analysis of variance of siblings. The number of loci may be estimated based on variances derived from data of more than one generation in plants. However, it is not possible to estimate number of loci and genotypic values in mammals because of unequal and unknown frequencies.

It is also known that the genetic basis for a given quantitative trait may be demonstrated by showing an association between the trait and a known genetic marker. However, methods available at present for detecting such an association require family data. The collection is usually time-consuming and sometimes difficult, particularly in man. There is so far no method available for making genetic inferences about a trait based on a random sample of individuals. The recent development of an exact distribution for phenotypic values of a quantitative trait in self-fertilized and random-mating populations, and computer simulation techniques provide an opportunity to study a quantitative trait using a random sample from a population. The exact distribution contains genetic parameters such as the number of loci, gene frequencies and genotypic values as well as the mean and the variance of environmental effects. Therefore, this distribution gives much more information than normal distribution which is frequently assumed. It is generally observed that the distribution of phenotypic values of quantitative traits are close to normal distribution, however, many depart from the normal by being asymmetric or multimodal.

The applications of simulation techniques and statistical procedures on computer have been extensively used in many research areas. It becomes feasible to generate data on a given model for testing hypotheses and estimating parameters. It is particularly useful for the exploratory study of a novel approach, the validity of which cannot be critically evaluated by using data of unknown parameters.

The present study was designed to apply the exact distribution of phenotypic values for the estimation of parameters in quantitative traits and to detect an association between a quantitative trait and a genetic marker locus in a random sample of individuals.

2. REVIEW OF LITERATURE

Mendel suggested in 1865 that an occurrence of numerous heritable grades of a single character might be due to multiple independent factors (Wright, 1952). The theoretical basis for studying quantitative characters began with Galton (1889) and Pearson (1898, 1904) who introduced the method of regression and correlation. That the possibility for the variation of a continuous trait such as human stature might be determined by a large number of genes was questioned by Bateson and Saunders (1902). Many subsequent experimental results supported the multiple factor hypothesis (Nilsson-Ehle, 1909; East and Hayes, 1911; East, 1916). Johannsen (1909) demonstrated the distinction between variation due to genetic causes and environmental effects by his pure line theory and introduced the terms *genotype* for genetic causes and *phenotype* for the individual's appearance (Whitehouse, 1965). An analysis of phenotypic variation in terms of genetic and environmental contribution was examined by Weinberg (Wright, 1952). Fisher (1918) showed that the total genetic variance could be divided into two components--the additive genetic variance and the dominance variance--and developed the fundamental theory of quantitative genetics. A concept of two kinds of genetic effect, namely major genes of large, specific effect and polygenes of smaller, less specific effect was introduced by Mather (1949). A continuous variation of phenotypic values was explained by non-heritable agencies which could blur minor discontinuities of expression of the character caused by the polygenes. The discontinuous variation was explained by major genes. The variation of phenotypic values caused by genotypes might be partly due to major gene differences and partly to a polygenic system (Mather, 1949). This concept of major gene hypothesis was tested (Morton, 1967; Elston and Campbell, 1970; Elston and Stewart, 1971; Spence et al, 1973). A theoretical distribution of phenotypic values determined by two types of random variables was derived by Tan and Chang (1972a). One was discrete nature associated with segregation of genes and the other was continuous nature associated with environmental effects which were assumed to be normally distributed. This distribution was initially derived for self-fertilized populations and extended for random mating populations (Tan, 1975). A complex segregation analysis of quantitative traits under a model which includes polygenes, a major locus, and environmental effects was considered by Morton and MacLean (1974).

Castle (1921) suggested a formula for estimating the number of genetic factors concerned in a quantitative character. Later, this formula was corrected by Wright (1921) and became known as the Castle-Wright formula. Several other methods of estimating the number of loci were proposed ("Student," 1934; Robson, 1956; Wright, 1968; O'Donald, 1971; Mather and Jinks, 1971). The formula for estimation of the number of loci was derived by using the moment method requiring several generations of experimental data. The formula did not give an integer estimate of the number of loci which had to be an integer a priori. Tan and Chang (1972a) proposed a method of estimating an integer value of the number of loci using the exact distribution of phenotypic values and the method of maximum likelihood. The method of maximum likelihood was found to be more efficient than the moment method (Kasal, 1933; Shenton, 1970; Tan and Chang, 1972b).

The whole issue of whether a characteristic is hereditary or environmental is a question of how much of the variation in that characteristic in that population is caused by differences in heredity and how much is caused by differences in the environment (Lush, 1945). The heritability was defined as the ratio of genetic variance to phenotypic variance. Various methods of estimating the heritability of a quantitative trait were developed and reviewed (Lush, 1945; Wright, 1952; Falconer, 1960).

If the heritability of a trait is low, then the phenotypic expression is mainly determined by environmental effects. Under the assumption of normality, the phenotypic values can be approximated by normal distribution regardless of the number of loci. When the heritability is high, and particularly when the number of loci is large, the phenotypic distribution also approaches normal by the Central Limit Theorem. If the phenotypic distribution is not normal, transformation may be used to convert the distribution to normal (Wright, 1952; Snedecor, 1956; Sokal and Rohlf, 1969; Falconer, 1960; Armitage, 1973; Mather and Jinks, 1971). However, the meaning of parameters are changed by the transformation employed.

Since the genes involved in a quantitative character segregate, the association between one of these genes and other loci must also be considered. It becomes extremely difficult since individual genotypes underlying the phenotypic values are unidentifiable. Evidence of association between major genes and quantitative genes using a marker to classify chromosomes was reported in plants by Sax (1923), Breese and Mather (1957, 1960) and Thoday (1961, 1967). Methods of determining genetic linkage by using human data of sib-pairs were investigated by Penrose (1935, 1938, 1946); Brues (1950) extended these methods for studying quantitative characters. These methods are based on the premise that if two siblings receive the same chromosome from a parent, there in an increased probability that the siblings will be phenotypically alike. It involved the comparison of the average phenotypic value differences squared between full sibs alike for the trait with the differences squared between unlike full sibs. Disadvantages of the sib-pair method were discussed by Lowry and Shultz (1959). An alternative method by the analysis of variance involves also the comparison of the differences between full sibs of different blood group genotypes and those between full sibs of the same genotype (Lowry and Shultz, 1959). It assumes a complete linkage between a locus affecting a quantitative trait and the genetic marker locus. This technique requires the very close linkage and a very large sample size (Neimann-Sorensen and Robertson, 1961). The assumption of complete linkage may be unrealistic and a very large sample may not be feasible in practice. Jayakar (1970) suggested a method for the detection and the estimation of linkage between a qualitative marker gene and a locus with two alleles influencing quantitative character. The method assumes that there is only one locus affecting the quantitative character and involves the pure environmental variance which is unknown and cannot be estimated.

Haseman and Elston (1972) provided a procedure for estimating linkage between a known m-allele locus and a hypothetical two-allele locus that governs a quantitative trait by using the sib-pair method and the method of regression with restrictions on the regression coefficient being non-positive. The power of their method was found to be extremely low (Robertson, 1973). Hill (1975) presented an approach for detecting and estimating quantitative linkage based on sibship data. By using a nested analysis of variance method, the null hypothesis of no linkage was tested comparing the two expected mean squares, one within marker genotype within sibship, and the other between genotypes within sibship. Hill assumed normality for all the variables in the nested analysis model which is not true in nature. A non-parametric test for the detection of linkage between a quantitative trait and a mendelian character was described by Smith (1975). It is not possible to detect linkage unless the quantitative locus and the marker locus are very closely linked, and it requires selection of informative families. Lange, Spence and Frank (1976) applied the *lod* method (Morton, 1955) to the detection of linkage between a quantitative trait and a qualitative marker. It was a simulation study using the known linkage between Amy_2 and *Duffy*-loci and converting Amy_2-locus to a quantitative variation by adding a normal deviate with fixed mean and variance. The simulation results were based on only one locus and small environmental variance and it is unknown in the general situations.

All the methods reviewed above for studying quantitative characters require a large amount of informative family data or data of several generations. Such data are difficult to collect, particularly in man.

Since these methods are based on unrealistic assumptions or unknown parameters, the application to real data may lead to misleading conclusions. As yet, there is no simple method that can be applied to a random sample of individuals from the population.

3. THEORIES AND APPLICATIONS

3.1 A Model for Quantitative Traits

According to Falconer (1960), quantitative traits are defined as characters that exhibit continuous variation without natural discontinuities. This continuous variation is presumably due to the simultaneous function of segregation of many genes affecting the character and the superimposition of truly continuous variation arising from non-genetic causes (environmental effects). A phenotypic value Y can be expressed as $Y = G + E$, where G is the genotypic value and E the environmental effects. The following assumptions are usually made:

(1) diploid genotype and two alleles at each locus,

(2) genotypic value being controlled by n-pairs of unlinked genes,

(3) same genotypic values for each locus,

(4) no epistatic effects among loci when $n > 1$,

(5) environmental effects being normally distributed,

(6) no genetic and environmental interactions,

(7) population at panmixia, and

(8) no selection at any locus.

Assumption (5) is expected to be satisfied by applying the Central Limit Theorem if environmental effect is considered as a sum of many independent various disturbances. Assumption (3) is based on the polygene theory for quantitative genetics (Mather, 1949). Under the formula $Y = G + E$, G and E are assumed to be uncorrelated, and the phenotypic variance σ_y^2 is the sum of the genotypic variance σ_g^2 and the environmental variance σ_e^2.

The following illustration was based on the model for a quantitative trait controlled by one locus, G_1-locus, with two alleles, A_1 and A_2, and environmental effects which were assumed to be normally distributed with mean μ_e and variance σ_e^2. This model is shown in Figure 3.1.1

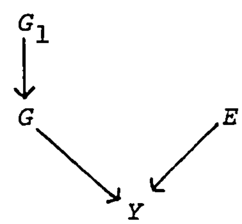


Figure 3.1.1 Model of a Phenotypic Value When $n = 1$. G_1 locus determines the genotypic value G. Then G and the environmental effects, E, determine the phenotypic value Y.

There will be three possible genotypes at the locus; namely, A_1A_1, A_1A_2, and A_2A_2 with genotypic values a, d and $-a$, respectively. Let the gene frequency of A_1 be p and the gene frequency of A_2 be $q(q = 1 - p)$. The probability that the genotypic value at G_1-locus being a, d, and $-a$ is p^2, $2pq$, and q^2, respectively.

$$P(G = a) = P(A_1A_1) = p^2$$

$$P(G = d) = P(A_1A_2) = 2pq$$

$$P(G = -a) = P(A_2A_2) = q^2$$

The degree of dominance at this locus is defined as the genotypic value of A_1A_2 divided by the genotypic value of A_1A_1, $x = \frac{d}{a}$, where x is the degree of dominance. For example, when there is no dominance, $x = 0$, and the genotypic value of heterozygote A_1A_2 is at the mid-point between two homozygotes. When the genotypic value of A_1A_2 is the same as that of A_1A_1, there is complete dominance, $x = 1$. Similarly, when $0 < x < 1$, A_1-gene is partially dominant, and when $x > 1$, A_1-gene is overdominant. The genotypic mean μ_g and genotypic variance σ_g^2 are:

$$\begin{aligned}
\mu_g &= ap^2 + 2dpq - aq^2 \\
&= a(p - q) + 2dpq \\
&= \mu_a + \mu_d \\
\sigma_g^2 &= a^2p^2 + 2d^2pq + a^2q^2 - (\mu_g)^2 \\
&= a^2(p^2 + q^2) + 2d^2pq - a^2(p - q)^2 - 4adpq(p - q) - 4d^2p^2q^2 \\
&= 2pq\ \{d^2 + a^2 + 2ad(q - p) - d^2pq\} \\
&= 2pq\ \{a^2 + 2ad(q - p) + d^2(1 - 2pq)\} \\
&= 2pq\ \{a^2 + 2ad(q - p) + d^2(p^2 + q^2)\} \\
&= 2pq\ \{a^2 + 2ad(q - p) + d^2(q - p)^2 + 2pqd^2\} \\
&= 2pq\ \{a + d(q - p)\}^2 + 4d^2p^2q^2 \\
&= \sigma_a^2 + \sigma_d^2
\end{aligned}$$

where μ_a is the additive mean, μ_d is the non-additive mean, σ_a^2 is the additive variance, and σ_d^2 is the non-additive variance. The phenotypic mean μ_y and the phenotypic variance σ_y^2 can be expressed as:

$$\mu_y = \mu_g + \mu_e$$
$$\sigma_y^2 = \sigma_g^2 + \sigma_e^2$$

In general, when the number of loci is n, the genotypic mean and the variance become the sums of means and variances due to each locus as shown by:

$$\mu_g = \sum_{i=1}^{n} \mu_{gi}$$

$$= a \sum_{i=1}^{n} (p_i - q_i) + 2d \sum_{i=1}^{n} p_i q_i$$

$$\sigma_g^2 = \sum_{i=1}^{n} \sigma_{gi}^2$$

$$= \sum_{i=1}^{n} 2p_i q_i \{a + d(q_i - p_i)\}^2 + 4d^2 \sum_{i=1}^{n} p_i^2 q_i^2$$

And the phenotypic mean and the variance are given as:

$$\mu_y = \sum_{i=1}^{n} \mu_{gi} + \mu_e$$

$$= \mu_g + \mu_e$$

$$\sigma_y^2 = \sum_{i=1}^{n} \sigma_{gi}^2 + \sigma_e^2$$

$$= \sigma_g^2 + \sigma_e^2$$

The model of a phenotypic value Y determined by n-loci, G_1, G_2, G_3, ..., G_n, and the environmental effects E is shown in Figure 3.1.2.

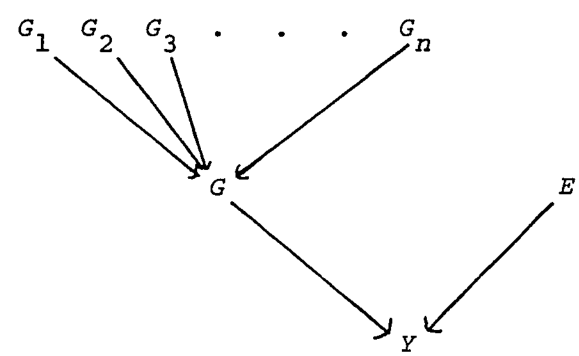


Figure 3.1.2 The Model of a Phenotypic Value
$Y = G_1 + G_2 + G_3 + . . . + G_n + E$

3.2 Distribution of Phenotypic Values

An exact distribution of the phenotypic values of quantitative characters under random mating has been derived (Tan, 1975). The parameters of this distribution are the mean μ_e and the variance σ_e^2 of environmental effects, genotypic values a and d, the number of loci n and the gene frequencies p_i for $i = 1$ to n. There will be $5 + n$ parameters when the genotypic value is controlled by n-pair of genes.

3.2.1 Single Locus ($n = 1$)

Given $n = 1$, a phenotypic value y_j follows the following distribution:

$$f(y_j) = \frac{1}{\sqrt{2\pi\sigma_e^2}} [p^2 \, exp\{\frac{-1}{2\sigma_e^2} (y_j - \mu_e - a)^2\} +$$

$$2pq \, exp\{\frac{-1}{2\sigma_e^2} (y_j - \mu_e - d)^2\} +$$

$$q^2 \, exp \, \{\frac{-1}{2\sigma_e^2} (y_j - \mu_e + a)^2 \,]$$

where a = genotypic value of A_1A_1

d = genotypic value of A_1A_2

p = gene frequency of A_1

q = gene frequency of A_2

μ_e = mean of environmental effects

σ_e^2 = variance of environmental effects and

y_j = phenotypic value of jth observation.

There are three components in this distribution because of three possible genotypes at the locus. These components are normally distributed with mean $\mu_e + a$, $\mu_e + d$, and $\mu_e - a$, and a common variance σ_e^2. The phenotypic distribution is the mixture of three normal distributions with mixing proportion p^2, $2pq$, q^2, respectively. The mean and the variance of this distribution are:

$$\mu_y = a(p - q) + 2dpq + \mu_e$$
$$\sigma_y^2 = 2pq\{a + d(q - p)\}^2 + 4d^2p^2q^2 + \sigma_e^2$$

The distribution is a function of a, d, p, μ_e and σ_e^2.

3.2.2 Two Loci ($n = 2$)

Given $n = 2$, there will be nine possible genotypes. The genotypic frequencies ϕ_i, and the phenotypic values m_i, of these nine genotypes are given in Table 3.2.2. The mean and the variance of genotypic values are:

$$\mu_g = a\sum_{i=1}^{2}(p_i - q_i) + 2d\sum_{i=1}^{2}p_iq_i$$

$$\sigma_g^2 = \sum_{i=1}^{2}2p_iq_i\{a + d(q_i - p_i)\}^2 + 4d^2\sum_{i=1}^{2} p_i{}^2q_i{}^2$$

The distribution of phenotypic values becomes:

$$f(y_j) = \frac{1}{\sqrt{2\pi\sigma_e^2}}[\sum_{j=1}^{9} \phi_j exp\{\frac{-1}{2\sigma_e^2}(y_j - \mu_e - m_{ij})^2\}]$$

The mean and the variance of this distribution are:

$$\mu_y = \mu_g + \mu_e$$
$$\sigma_y^2 = \sigma_g^2 + \sigma_e^2$$

Table 3.2.2

The Genotypic Frequencies and Values for $n = 2$

i	Genotype	Genotypic Frequency ϕ_i	Genotypic Value m_i
1	$A_1A_1B_1B_1$	$p_1^2p_2^2$	$2a$
2	$A_1A_1B_1B_2$	$2p_1^2p_2q_2$	$a + d$
3	$A_1A_1B_2B_2$	$p_1^2q_2^2$	0
4	$A_1A_2B_1B_1$	$2p_1q_1p_2^2$	$d + a$
5	$A_1A_2B_1B_2$	$4p_1q_1p_2q_2$	$2d$
6	$A_1A_2B_2B_2$	$2p_1q_1q_2^2$	$d - a$
7	$A_2A_2B_1B_1$	$p_2^2q_1^2$	0
8	$A_2A_2B_1B_2$	$2q_1^2p_2q_2$	$d - a$
9	$A_2A_2B_2B_2$	$q_1^2q_2^2$	$-2a$

3.2.3 Generalization for n Loci ($n \geq 1$)

The notations for these distributions are defined as follows:

A_{ij} = jth allele at the ith locus

p_i = gene frequency of a gene A_{i1} at ith locus

q_i = gene frequency of the other gene A_{i2} at ith locus

($q_i = 1 - p_i$)

G_i = genotypic value at ith locus

$G = \sum_{i=1}^{n} G_i = G_1 + G_2 + G_3 + \ldots + G_n$

a = genotypic value of $A_{i1}A_{i1}$, where $i = 1, 2, 3, \ldots, n$

d = genotypic value of $A_{i1}A_{i2}$, where $i = 1, 2, 3, \ldots, n$

$-a$ = genotypic value of $A_{i2}A_{i2}$, where $i = 1, 2, 3, \ldots, n$

Thus, the probability density function of G_i is:

$$P\ (G_i = a) = P\ (A_{i1}A_{i1}) = p_i^2$$

$$P\ (G_i = d) = P\ (A_{i1}A_{i2}) = 2p_iq_i$$

$$P\ (G_i = -a) = P\ (A_{i2}A_{i2}) = q_i^2$$

Let $W_{uv}^{(n)}$ be the frequency of genotypes with u homozygotes A_1A_1, v heterozygotes A_1A_2, and $n - u - v$ the other homozygotes A_2A_2, where n is the number of loci. Then, this is the probability density function of genotypic values with n loci. Values for $n = 1$ and $n = 2$ are given in Table 3.2.3. When $n = 1$, $W_{uv}^{(n)}$ is a point binomial distribution (its phenotypic distribution was discussed earlier). $W_{uv}^{(n)}$ is a multinomial distribution if $p_i = p$ for $i = 1, 2, 3, \ldots n$.

$$W_{uv}^{(n)} = \binom{n}{uv}\ (p^2)^u\ (2pq)^v (q^2)^{n-u-v}$$

When gene frequencies are different, $W_{uv}^{(n)}$ can be expressed as

$$W_{uv}^{(n)} = \Sigma\{\prod_{i=1}^{n} (p_i^2)^{\alpha_i} (2p_iq_i)^{\beta_i} (q_i^2)^{\gamma_i}$$

where Σ is overall non-negative values of α_i, β_i and γ_i satisfying $\alpha_i + \beta_i + \gamma_i = 1$ for $i = 1$ to n, $\sum_{i=1}^{n}\alpha_i = u$ and $\sum_{i=1}^{n}\beta_i = v$. Since $W_{uv}^{(n)}$ is a probability density function,

$$\sum_{u=0}^{n} \sum_{v=0}^{n-u} W_{uv}^{(n)} = 1$$

Computational formula for $W_{uv}^{(n)}$ was derived by Tan (1975) and it is given in Appendix I.

Under the assumption of normality for the environmental effects, the probability density function of environmental effects e is:

$$f(e_j) = \frac{1}{\sqrt{2\pi\sigma_e^2}} exp\{\frac{-1}{2\sigma_e^2} (e_j - \mu_e)^2\}$$

where μ_e is the mean and σ_e^2 is the variance of environmental effects.

Given the genotype with u $A_{i1}A_{i1}$, v $A_{i1}A_{i2}$ and $n-u-v$ $A_{i2}A_{i2}$, the mean of the genotypic values is:

$$ua + vd + (n-u-v)\ (-\ a)$$

and the frequency of this genotype is the proportion $W_{uv}^{(n)}$. Therefore, the probability density function of the phenotypic value (exact distribution) is:

$$f(y_j) = \sum_{u=0}^{n} \sum_{v=0}^{n-u} W_{uv}^{(n)} \frac{1}{\sqrt{2\pi\sigma_e^2}} exp\{\frac{-1}{2\sigma_e^2} (y_j - \mu_e - ua - vd + (n-u-v)a)^2\}$$

This distribution approaches normal asymptotically as n becomes large. The mean μ_y and the variance σ_y^2 of this distribution are given below and its derivations are given in Appendix II.

Table 3.2.3

Values of $W_{uv}^{(1)}$ and $W_{uv}^{(2)}$

When $n = 1$
$W_{00}^{(1)} = q^2$
$W_{01}^{(1)} = 2pq$
$W_{10}^{(1)} = p^2$
When $n = 2$
$W_{00}^{(2)} = \phi_9^* = q_1^2 q_2^2$
$W_{01}^{(2)} = \phi_2 + \phi_4 = 2p_1^2 p_2 q_2 + 2p_1 q_1 p_2^2$
$W_{02}^{(2)} = \phi_5 = 4p_1 q_1 p_2 q_2$
$W_{10}^{(2)} = \phi_6 + \phi_8 = 2p_1 q_1 q_2^2 + 2q_1^2 p_2 q_2$
$W_{11}^{(2)} = \phi_3 + \phi_7 = p_1^2 q_2^2 + q_1^2 p_2^2$
$W_{20}^{(2)} = \phi_1 = p_1^2 p_2^2$

*ϕ_i = frequency of the ith genotype given in Table 3.2.2.

$$\mu_y = a\sum_{i=1}^{n} (p_i - q_i) + 2d\sum_{i=1}^{n} p_i q_i + \mu_e$$

$$\sigma_y^2 = \sum_{i=1}^{n} 2p_i q_i\{a + d(q_i - p_i)\}^2 + 4d^2 \sum_{i=1}^{n} p_i^2 q_i^2 + \sigma_e^2$$

These are well known equations given the model $Y = G + E$ with the assumptions stated earlier. Therefore, it confirms that the exact distribution is the distribution of phenotypic values with the given model.

3.3 Models for Association Between a Quantitative Character Locus and a Genetic Marker Locus

Three conditions for the association between one of the loci involved in the quantitative trait, e.g. the nth locus G_n, and a genetic marker locus, M-locus, may be considered; namely, (1) completely independent loci, (2) partially linked loci,[1] and (3) completely linked or identical loci. For convenience, the following assumptions were made: (1) three distinguishable genotypes at M-locus are MM, Mm and mm with genotypic frequencies of p_m^2, $2p_m q_m$ and q_m^2 and (2) three genotypes at G_n-locus are A_1A_1, A_1A_2 and A_2A_2, with genotypic frequencies of p_n^2, $2p_n q_n$, and q_n^2, respectively.

3.3.1 Model-1: G_n-locus and M-locus are Independent

This model includes the case in which linkage is at equilibrium (repulsion counter balances coupling) but not detectable. The genotypic values at G_n-locus within three genotypes of M-locus are given in Table 3.3.1.

---

[1] For a condition in which two loci were loosely linked.

Table 3.3.1

Genotypic Values at G_n-locus Under Model-1

		MM	Mm	mm
		p_m^2	$2p_mq_m$	q_m^2
A_1A_1	p_n^2	a	a	a
A_1A_2	$2p_nq_n$	d	d	d
A_2A_2	q_n^2	$-a$	$-a$	$-a$

Since the G_n-locus and the M-locus are independent, the three means and variances are expected to be identical. The expected means, μ'_{MM}, μ'_{Mm}, and μ'_{mm}, and the variances, $\sigma^{2\prime}_{MM}$, $\sigma^{2\prime}_{Mm}$, and $\sigma^{2\prime}_{mm}$, within each genotype at M-locus can be computed from Table 3.3.1.

$$\mu'_{MM} = \mu'_{Mm} = \mu'_{mm} = (p_n - q_n)\ a + 2p_nq_nd$$

$$\sigma^{2\prime}_{MM} = \sigma^{2\prime}_{Mm} = \sigma^{2\prime}_{mm} = 2p_nq_n\{a + (q_n - p_n)d\}^2 + 4p_n^2q_n^2d^2$$

Taking into consideration the effects of other loci and environmental disturbance for the quantitative trait, the means μ_{MM}, μ_{Mm}, and μ_{mm} and the variances σ^2_{MM} , σ^2_{Mm} , and σ^2_{mm} become:

$$\mu_{MM} = \mu_{Mm} = \mu_{mm} = \sum_{i=1}^{n}\mu_{gi} + \mu_e$$

$$\sigma^2_{MM} = \sigma^2_{Mm} = \sigma^2_{mm} = \sum_{i=1}^{n}\sigma^2_{gi} + \sigma^2_e$$

3.3.2 Model-2: G_n-locus and M-locus are partially linked

In random mating populations, the amount of linkage may be measured by the degree of linkage disequilibrium between the two loci, which is designated by D. For convenience, p_n is the gene frequency of A_1-gene at G_n-locus, p_m the gene frequency of M-gene at M-locus, and f_{11}, f_{12}, f_{21}, and f_{22} the expected frequencies of A_1M, A_1m, A_2M, and A_2m gamete type, respectively. Then D is measured by the difference between frequencies of coupling $\left(\frac{A_1M}{A_2m}\right)$ and repulsion $\left(\frac{A_1m}{A_2M}\right)$ phases.

$$D = f_{11}f_{22} - f_{12}f_{21}$$

The expected frequencies of gamete types can be expressed in terms of D, p_n and p_m.

$$f_{11} = p_np_m + D$$

$$f_{12} = p_nq_m - D$$

$$f_{21} = q_np_m - D$$

$$f_{22} = q_nq_m + D$$

The expected genotypic frequencies of A_1A_1, A_1A_2, and A_2A_2 at G_n-locus within the three genotypes at M-locus are given in Table 3.3.2.

Table 3.3.2

Expected Genotypic Frequencies Under Model-2

	MM	Mm	mm	Total
A_1A_1	f_{11}^2	$2f_{11}f_{12}$	f_{12}^2	p_n^2
A_1A_2	$2f_{11}f_{21}$	$2f_{11}f_{12}+2f_{12}f_{21}$	$2f_{12}f_{22}$	$2p_nq_n$
A_2A_2	f_{21}^2	$2f_{21}f_{22}$	f_{22}^2	q_n^2
Total	p_m^2	$2p_mq_m$	q_m^2	1

Under this model, the probability distributions for the three marker genotypes MM, Mm, and mm are derived as follows:

$$f_{MM}(y_j) = \sum_{u=0}^{n-1} \sum_{v=0}^{n-1-u} W_{uv}^{(n-1)} \frac{1}{\sqrt{2\pi\sigma_e^2}}$$

$$\left[\frac{f_{11}^2}{p_m^2} exp\{\frac{-1}{\sigma_e^2} (y_j - \mu_e - ua - vd + (n-1-u-v)a - a)^2\} \right.$$

$$+ \frac{2f_{11}f_{21}}{p_m^2} exp\{\frac{-1}{\sigma_e^2} (y_j - \mu_e - ua - vd + (n-1-u-v)a - d)^2\}$$

$$\left. + \frac{f_{21}^2}{p_m^2} exp\{\frac{-1}{\sigma_e^2} (y_j - \mu_e - ua - vd + (n-1-u-v)a + a)^2\} \right]$$

$$f_{Mm}(y_j) = \sum_{u=0}^{n-1} \sum_{v=0}^{n-1-u} W_{uv}^{(n-1)} \frac{1}{\sqrt{2\pi\sigma_e^2}}$$

$$\left[\frac{2f_{11}f_{12}}{2p_m q_m} exp\{\frac{-1}{\sigma_e^2} (y_j - \mu_e - ua - vd + (n-1-u-v)a - a)^2\}\right.$$

$$+ \frac{2f_{11}f_{22}+2f_{12}f_{21}}{2p_m q_m} exp\{\frac{-1}{\sigma_e^2} (y_j - \mu_e - ua - vd$$

$$+ (n-1-u-v)a - d)^2\}$$

$$+ \frac{2f_{21}f_{22}}{2p_m q_m} exp\{\frac{-1}{\sigma_e^2} (y_j - \mu_e - ua - vd$$

$$\left. + (n-1-u-v)a + a)^2\}\right]$$

$$f_{mm}(y_j) = \sum_{u=0}^{n-1} \sum_{v=0}^{n-1-u} W_{uv}^{(n-1)} \frac{1}{\sqrt{2\pi\sigma_e^2}}$$

$$\left[\frac{f_{12}^2}{q_m^2} exp\{\frac{-1}{\sigma_e^2} (y_j - \mu_e - ua - vd + (n-1-u-v)a - a)^2\}\right.$$

$$+ \frac{2f_{12}f_{22}}{q_m^2} exp\{\frac{-1}{\sigma_e^2} (y_j - \mu_e - ua - vd$$

$$+ (n-1-u-v)a - d)^2\}$$

$$\left. + \frac{f_{22}^2}{q_m^2} exp\{\frac{-1}{\sigma_e^2} (y_j - \mu_e - ua - vd + (n-1-u-v)a + a)^2\}\right]$$

When the phenotypic distribution is normally distributed, the following approximate distributions can be used for the above three distributions.

$$f_{MM}(y_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{p_m^2} [f_{11}^2 exp\{\frac{-1}{2\sigma^2} (y_j - \mu - a)^2\}]$$

$$+ 2f_{11}f_{21}\ exp\{\frac{-1}{2\sigma^2}(y_j - \mu - d)^2\}$$

$$+ f_{21}^2\ exp\{\frac{-1}{2\sigma^2}(y_j - \mu + a)^2\}]$$

$$f_{Mm}(y_j) = \frac{1}{\sqrt{2\pi\sigma^2}}\ \frac{1}{2p_m q_m}[\ 2f_{11}f_{12}\ exp\{\frac{-1}{2\sigma^2}(y_j - \mu - a)^2\}$$

$$+ 2f_{11}f_{22} + 2f_{12}f_{21}\ exp\{\frac{-1}{2\sigma^2}(y_j - \mu - d)^2\}$$

$$+ 2f_{21}f_{22}\ exp\{\frac{-1}{2\sigma^2}(y_j - \mu + a)^2\}]$$

$$f_{mm}(y_j) = \frac{1}{\sqrt{2\pi\sigma^2}}\ \frac{1}{q_m^2}[f_{12}^2\ exp\{\frac{-1}{2\sigma^2}(y_j - \mu - a)^2\}$$

$$+ 2f_{12}f_{22}\ exp\{\frac{-1}{2\sigma^2}(y_j - \mu - d)^2\}$$

$$+ f_{22}^2\ exp\{\frac{-1}{2\sigma^2}(y_j - \mu + a)^2\}]$$

These distributions are derived by assuming that the phenotypic values due to the other n - 1 loci and the environmental effect are normally distributed. Hence, the number of parameters reduce to six. These parameters are the gene frequency of A_1-gene at G_n-locus (p_n), the genotypic values (a and d), the degree of linkage disequilibrium (D), and the mean (μ) and the variance (σ^2) of the normal distribution. p_m is the gene frequency of M-gene at M-locus and is known, and $q_n = 1 - p_n$ and $q_m = 1 - p_m$.

Since f_{11}, f_{12}, f_{21}, and f_{22} are frequencies, these values range between 0 and 1, and the sum is equal to 1. Hence, there is an upper and lower bound for the values of D.

$$Min(p_n q_m\ ,\ q_n p_m) \geq D \geq -Min(p_n p_m\ ,\ q_n q_m)$$

The degree of linkage disequilibrium may conveniently be expressed as the ratio to the minimal value of two sets of the product of gene frequencies.

$$D^* = \frac{D}{Min(p_n q_m,\ q_n p_m)} \quad \text{when } D \geq 0$$

$$D^* = \frac{D}{Min(p_n p_m,\ q_n q_m)} \quad \text{when } D < 0$$

The maximum value of D^* becomes 1 and its minimum value is -1. When $D = 0$, D^* is also equal to 0 and this model reduces to Model-1.

The means for the G_n-locus, μ'_{MM}, μ'_{Mm}, and μ'_{mm}, and variances, $\sigma^{2\prime}_{MM}$, $\sigma^{2\prime}_{Mm}$, and $\sigma^{2\prime}_{mm}$ at G_n-locus within marker genotypes are:

$$\mu'_{MM} = \frac{1}{p_m^2}\,(af_{11}^2 + 2df_{11}f_{21} - af_{21}^2)$$

$$\mu'_{Mm} = \frac{1}{2p_m q_m}\,(2af_{11}f_{12} + 2df_{11}f_{22} + 2df_{12}f_{21} - 2af_{21}f_{22})$$

$$\mu'_{mm} = \frac{1}{q_m^2}\,(af_{12}^2 + 2df_{12}f_{22} - af_{22}^2)$$

$$\sigma^{2\prime}_{MM} = \frac{1}{p_m^2}\,(a^2f_{11}^2 + 2d^2f_{11}f_{21} + a^2f_{21}) - (\mu'_{MM})^2$$

$$\sigma^{2\prime}_{Mm} = \frac{1}{2p_m q_m}\,(2a^2f_{11}f_{12} + 2d^2f_{11}f_{22} + 2d^2f_{12}f_{21} + 2a^2f_{21}f_{22}) - (\mu'_{Mm})^2$$

$$\sigma^{2\prime}_{mm} = \frac{1}{q_m^2}\,(a^2f_{12}^2 + 2d^2f_{12}f_{22} + a^2f_{22}^2) - (\mu'_{mm})^2$$

Therefore, the phenotypic means μ_{MM}, μ_{Mm}, and μ_{mm}, and the variances, σ^2_{MM}, σ^2_{Mm}, and σ^2_{mm}, are:

$$\mu_{MM} = \sum_{i=1}^{n-1} \mu_{g_i} + \mu'_{MM} + \mu_e$$

$$\mu_{Mm} = \sum_{i=1}^{n-1} \mu_{g_i} + \mu'_{Mm} + \mu_e$$

$$\mu_{mm} = \sum_{i=1}^{n-1} \mu_{g_i} + \mu'_{mm} + \mu_e$$

$$\sigma^2_{MM} = \sum_{i=1}^{n-1} \sigma^2_{g_i} + \sigma^{2\prime}_{MM} + \sigma^2_e$$

$$\sigma^2_{Mm} = \sum_{i=1}^{n-1} \sigma^2_{g_i} + \sigma^{2\prime}_{Mm} + \sigma^2_e$$

$$\sigma^2_{mm} = \sum_{i=1}^{n-1} \sigma^2_{g_i} + \sigma^{2\prime}_{mm} + \sigma^2_e$$

3.3.3 Model-3: G_n-locus and M-locus are Completely Linked or Identical Loci

In this model, G_n-locus and M-locus are either very closely linked or identical so that the genotypic value due to G_n-locus is dependent upon the genotype of M-locus. This also means that the gene frequencies at M-locus and G_n-locus are the same. This is a special case of Model-2 with the value of D equal to its maximum or its minimum value and $p_n = p_m$. The genotypic values at G_n-locus are given in Table 3.3.3.

Table 3.3.3

Genotypic Values at G_n-locus Under Model-3

		MM	Mm	mm
		p_m^2	$2p_mq_m$	q_m^2
A_1A_1	p_n^2	a	-	-
A_1A_2	$2p_nq_n$	-	d	-
A_2A_2	q_n^2	-	-	$-a$

The expected means, μ'_{MM}, μ'_{Mm}, and μ'_{mm} of genotypic values at G_n-locus are a, d and $-a$, respectively. The variances within subclasses are zero. Therefore, the phenotypic means of subclass also differ by a, d, and $-a$, and the variances are the same.

$$\mu_{MM} = \sum_{i=1}^{n-1} \mu_{gi} + a + \mu_e$$

$$\mu_{Mm} = \sum_{i=1}^{n-1} \mu_{gi} + d + \mu_e$$

$$\mu_{mm} = \sum_{i=1}^{n-1} \mu_{gi} - a + \mu_e$$

$$\sigma^2_{MM} = \sigma^2_{Mm} = \sigma^2_{mm} = \sum_{i=1}^{n-1} \sigma^2_{gi} + \sigma^2_e$$

Probability distributions of these subclasses are:

$$f_{MM}(y_j) = \sum_{u=0}^{n-1} \sum_{v=0}^{n-1-u} W_{uv}^{(n-1)} \frac{1}{\sqrt{2\pi\sigma_e^2}} exp \{\frac{-1}{2\sigma_e^2} (y_j - \mu_e - ua - vd - (n - 1 - u - v)a + a)^2\}$$

$$f_{Mm}(y_j) = \sum_{u=0}^{n-1} \sum_{v=0}^{n-1-u} W_{uv}^{(n-1)} \frac{1}{\sqrt{2\pi\sigma_e^2}} exp \{\frac{-1}{2\sigma_e^2} (y_j - \mu_e - ua - vd - (n - 1 - u - v)a + d)^2\}$$

$$f_{mm}(y_j) = \sum_{u=0}^{n-1} \sum_{v=0}^{n-1-u} W_{uv}^{(n-1)} \frac{1}{\sqrt{2\pi\sigma_e^2}} exp\{\frac{-1}{2\sigma_e^2} (y_j - \mu_e - ua - vd - (n -1 - u - v)a - a)^2\}$$

3.4 Investigations of Phenotypic Distributions

The phenotypic distribution is the probability distribution of the phenotypes, thus taking into consideration both genetic segregation and environmental disturbances. Considering different genotypes as centers with normal environmental disturbances superimposed upon them, the phenotypic distribution is a mixture of normal distributions. The mixing proportion is $W_{uv}^{(n)}$ for a genotype with u $A_{i_1}A_{i_1}$, v $A_{i_1}A_{i_2}$, and $(n - u - v)$ $A_{i_2}A_{i_2}$, which is a function of the gene frequencies. When we are dealing with a quantitative trait, we can only measure phenotypic values and study the phenotypic distributions. The phenotypic distributions cannot be normal in general, especially when n is small. Also, the phenotypic distribution depends on simultaneous effect of all the parameter values (the number of loci n, gene frequencies p_i for $i = 1$ to n, genotypic values a and d, and the mean μ_e, and the variance σ_e^2, of the environmental disturbances).

Phenotypic distributions as a mixture of three distributions each with means a, d and $-a$, and the variance σ_e^2 were investigated. It was assumed that: a quantitative trait was controlled by one locus with two alleles (A_1 and A_2); and the environmental effects were normally distributed. The following sets of parameter values were employed. In all cases, the mean μ_y, and the variance σ_y^2, of phenotypic distribution were set to be constants.

(1) $p = .5$, $h^2 = .8$, and $x = 0$

(2) $p = .5$, $h^2 = .2$, and $x = 0$

(3) $p = .5$, $h^2 = .8$, and $x = 1$

(4) $p = .8$, $h^2 = .8$, and $x = 0$

Assuming $\mu_y = 0$ and $\sigma_y^2 = 1$, for a given set of heritability (h^2), degree of dominance (x) and gene frequency (p), values a, d, μ_e and σ_e^2 were calculated as:

$$a = \frac{h}{\sqrt{2pq\,\{1 + x(q - p)\}^2 + 4x^2p^2q^2}}$$

$$d = ax$$

$$\mu_e = \mu_y - \mu_g \text{ where } \mu_g = a(p - q) + 2dpq$$

$$\sigma_e^2 = \sigma_y^2 - \sigma_y^2 h^2 = \sigma_y^2\,(1 - h^2) = 1 - h^2$$

The phenotypic distributions as affected by heritability, degree of dominance, and gene frequency were investigated by assuming $n = 1$. A total of twenty-seven sets of parameters each with three different values was considered.

(1) Heritability: $h^2 = .2$, $.5$ and $.8$

(2) Degree of dominance: $x = 0$, $.5$ and 1

(3) Gene frequency: $p = .1$, $.3$ and $.5$

Assuming $n = 2$, $h^2 = .5$ and $x = 0$, effects of different pairs of gene frequency on the phenotypic distribution were studied by considering three pairs of gene frequencies: $p_1 = .2$ and $p_2 = .8$, $p_1 = p_2 = .5$, and $p_1 = p_2 = .8$. The effects of number of loci on the phenotypic distribution were also investigated by comparing different number of loci, $n = 1$, 5 and 10, holding other parameter constants ($h^2 = .8$, $x = 0$, and $p_i = .5$ for $i = 1$ to n).

3.5 Generation of Test Data

Given the following parameters: mean of phenotypic values, $\mu_y = 0$; variance of phenotypic values, $\sigma_y^2 = 1$; heritability of the trait, h^2; number of loci, n; degree of dominance, x; degree of linkage disequilibrium, D^*; gene frequency of M-gene at M-locus, $p_m = .5$; and sample size, N, phenotypic values of N individuals were simulated based on the models discussed previously. The values chosen for h^2, n, x, D^* and N are given in Table 3.5. Parameters a, d, and p_i for $i = 1$ to n, μ_e, and σ_e^2 were computed as:

$$a = \frac{h}{\sqrt{\sum_{i=1}^{n} 2p_i q_i \{1 + x(q_i - p_i)\}^2 + x^2 \sum_{i=1}^{n} (2p_i q_i)^2}}$$

$$d = ax$$

$$\mu_g = a\sum_{i=1}^{n} (p_i - q_i) + d\sum_{i=1}^{n} 2p_i q_i$$

$$\mu_e = \mu_y - \mu_g = -\mu_g$$

$$\sigma_e^2 = \sigma_y^2 - \sigma_y^2 h^2 = 1 - h^2$$

The derivation of these formulas is given in Appendix III. Five independent runs for given set of parameter values were made to allow variations in gene frequencies. A flow chart of generation of sample data is given in Figure 3.5.

3.6 Estimation of Parameters Using Phenotypic Measurements Only

The procedure used was outlined as follows:

(1) Scale the data of individual phenotypes to find the maximum and minimum values.

(2) Construct a frequency distribution table using equal intervals between the two extreme values.

(3) Compute cumulative frequency distribution for testing normality by the Kolomogrov-Smirnov one-sample test (Siegel, 1956). If the distribution was not significantly different from normal distribution, the estimation of parameters was not continued for most

Table 3.5

Parameter Values for h^2, n, x, D^* and N

Data Set	h^2	n	x	D^*	N
1	.8	2	0	0	5000
2	.5	2	0	0	5000
3	.2	2	0	0	5000
4	.8	2	1	0	5000
5	.8	5	0	0	5000
6	.8	10	0	0	5000
7	.8	2	0	0	500
8	.8	2	0	0	100
9	.8	2	0	.5	5000
10	.5	2	0	.5	5000
11	.2	2	0	.5	5000
12	.8	2	1	.5	5000
13	.8	5	0	.5	5000
14	.8	10	0	.5	5000
15	.8	2	0	.5	500
16	.8	2	0	.5	100
17	.8	2	0	1	5000
18	.5	2	0	1	5000
19	.2	2	0	1	5000
20	.8	2	1	1	5000
21	.8	5	0	1	5000
22	.8	10	0	1	5000
23	.8	2	0	1	500
24	.8	2	0	1	100

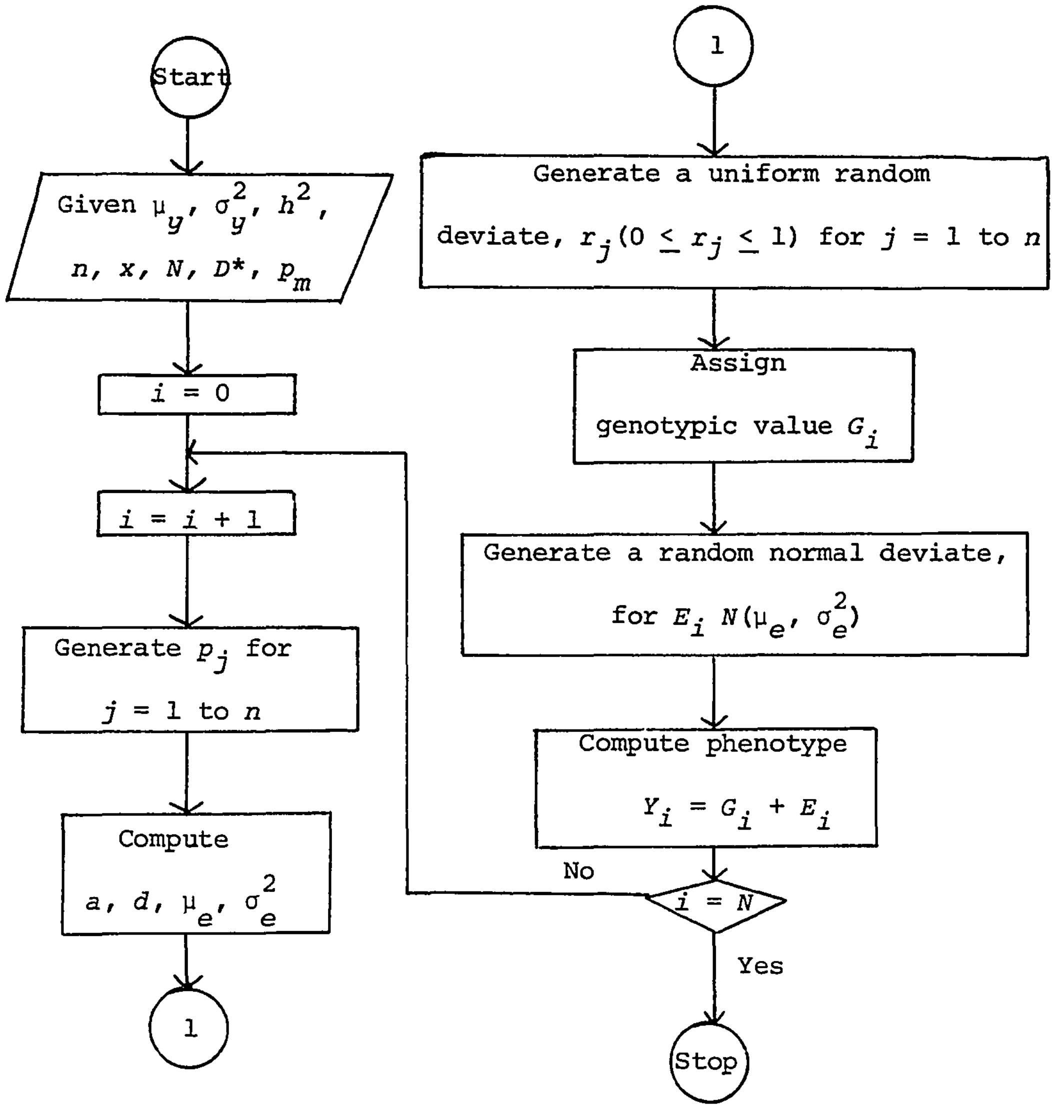


Figure 3.5 Generation for a Set of Sample Data

cases so that the number of computations was reduced.

(4) Compute the sample phenotypic mean, $\hat{\mu}_y$, and phenotypic variance, $\hat{\sigma}^2_y$, as:

$$\hat{\mu}_y = \frac{1}{N} \sum_{j=1}^{NN} y_j f_j$$

$$\hat{\sigma}^2_y = \frac{1}{N-1} \sum_{j=1}^{NN} (y_j - \hat{\mu}_y)^2 f_j$$

where y_j = the mid-point phenotypic value of jth interval, f_j = its frequency, NN = the total number of intervals and N = the total number of individual measurements.

(5) Find the initial values for parameter estimation.

i) Degree of dominance, x: select 0 for codominant system and 1 for complete dominance.

ii) Heritability, h^2: choose .2, .5, .6, .7 and .8.

iii) Number of loci, n: select 1, 2, 3, 4, and 5.

(6) Compute genetic variance $\hat{\sigma}^2_g$, and environmental variance $\hat{\sigma}^2_e$ with given h^2 and $\hat{\sigma}^2_y$ calculated from the sample.

$$\hat{\sigma}^2_g = \hat{\sigma}^2_y h^2$$

$$\hat{\sigma}^2_e = \hat{\sigma}^2_y - \hat{\sigma}^2_g$$

(7) Set gene frequency p_i for $i = 1, 2, \ldots n$ based on random deviates between zero and one by the congruential generation method (MacLaren and Marsaglia, 1965; Naylor et al, 1966; Downham and Roberts, 1967; Lewis et al, 1969).

(8) Estimate genotypic values $\hat{a}$ and $\hat{d}$, the mean of genotypic values $\hat{\mu}_g$, and the mean of environmental effects $\hat{\mu}_e$, as:

$$\hat{a} = \frac{\hat{\sigma}_g}{\sqrt{\sum_{i=1}^{n} 2p_iq_i\{1 + x(q_i - p_i)\}^2 + x^2 \sum_{i=1}^{n} (2p_iq_i)^2}}$$

$$\hat{d} = \hat{a}x$$

$$\hat{\mu}_g = \hat{a} \sum_{i=1}^{n} (p_i - q_i) + \hat{d} \sum_{i=1}^{n} 2p_iq_i$$

$$\hat{\mu}_e = \hat{\mu}_y - \hat{\mu}_g$$

(9) Compute log likelihood function value, Log L_{jklm}, for each set of selected initial values x_j, n_k, h_l^2, where x_j = the jth value of degree of dominance, e.g. $x_1 = 0$, $x_2 = 1$; n_k = the kth value of number of loci, e.g. $n_1 = 2$, $n_2 = 4$, $n_3 = 6$, $n_4 = 8$; h_l^2 = the lth heritability value selected, e.g. $h_1^2 = .2$, $h_2^2 = .5$, $h_3^2 = .8$, and δ_m = the mth vector of n_k gene frequencies for m = 1 to 10.

$$\log L_{jklm} = \sum_{i=1}^{NN} \log L(y_i)f_i$$

where

$$L(y_i) = \sum_{u=0}^{n_k} \sum_{v=0}^{n_k-u} W_{uv}^{(n_k)} \frac{1}{\sqrt{2\pi\sigma_e^2}} exp\{\frac{-1}{2\sigma_e^2} (y_i - \hat{\mu}_e - \hat{a}u - \hat{d}v + (n_k-u-v)\hat{a})^2\}$$

(10) Compare the log likelihood function values for all selected sets of initial values and choose the set with the maximum log likelihood function value.

(11) Estimate parameters by the method of maximum likelihood with the best set of initial values obtained in Step (9). A subroutine package for the maximum likelihood estimation, MAXLIK (Kaplan and Elston, 1972) was used. Since all data were simulated with known parameters in the present study, this step was used only for illustration.

3.7 Estimation of Parameters Using Phenotypic Measurements and Genetic Marker Genotype Information

The method of estimating the degree of linkage disequilibrium, D, was similar to the one discussed above for estimating parameters from phenotypic measurements only. However, the probability distributions for the three genetic marker genotype classes discussed in Section 3.3.2 were used instead of overall phenotypic distribution discussed in Section 3.2.

(1) Classify data into subclasses according to the marker genotypes (*MM*-, *Mm*- and *mm*-class) and construct three frequency distribution tables using the same intervals.

(2) Compute cumulative frequency distribution for each subclass and compare these distributions pairwise by the Kolomogrov-Smirnov two-sample test (Siegel, 1956). If M-locus and G_n-locus were independent, the three sample distributions would be expected to be identical, inasmuch as they showed only random deviations; therefore, testing and estimation would not be continued. On the other hand, if they were not independent, these three distributions of phenotypic values should be different. However, when the significant difference was found only by one pair of distributions, the estimation was not continued since this difference might be due to chance.

(3) Compute the expected means of the three subclasses $\hat{\mu}_i$, i = 1, 2, 3, and the variances $\hat{\sigma}_i^2$, i = 1, 2, 3, as:

$$\hat{\mu}_i = \frac{1}{N_i} \sum_{j=1}^{NN} y_j f_{ij}$$

$$\hat{\sigma}_i^2 = \frac{1}{N_i - 1} \sum_{j=1}^{NN} (y_i - \hat{\mu}_i)^2 f_{ij}$$

where y_j= the mid-point phenotypic value of the jth interval and f_{ij} = the ijth frequency, N_i = the ith subclass sample size, and NN = the total number of intervals.

(4) Compute gene frequency $(\hat{p}_m)$ at genetic marker locus by:

$$p_m = \frac{2N_1 - N_2}{2N}$$

where N_1 and N_2 are the total number of samples in MM and Mm-class, respectively.

(5) Select gene frequency of A_1 gene at G_n-locus, $\hat{p}_n$, $(\hat{p}_n = .1, .2, \ldots, .9)$ and other $(n - 1)$ gene frequencies as outlined in Section 3.6. The gene frequency of A_1, $(\hat{p}_n)$ can be selected from the estimates obtained. Since the locus associated with M-locus is unknown, all the values ($\hat{p}_i$ for $i = 1$ to n) must be considered as a possible value for $\hat{p}_n$.

(6) Solve for $\hat{D}$ given $\hat{p}_n$ and other parameter estimates. The derivation is given in Appendix IV.

(7) Compute the log likelihood function value for each set of initial values and select the set with the maximum likelihood function value. This set provides the estimates of the degree of linkage disequilibrium $(\hat{D})$ and the gene frequency of A_1 at G_n-locus $(\hat{p}_n)$.

(8) Calculate $\hat{D}^*$ as:

$$\hat{D}^* = \frac{\hat{D}}{Min(\hat{p}_n\hat{q}_m,\ \hat{q}_n\hat{p}_m)} \quad \text{when } \hat{D} \geq 0$$

$$\hat{D}^* = \frac{\hat{D}}{Min(\hat{p}_n\hat{p}_m,\ \hat{q}_n\hat{q}_m)} \quad \text{when } \hat{D} < 0$$

Computer programs for simulating data sets and estimating parameters were written in FORTRAN. Most of the numerical computations were done by using IBM 370/158 at the University of Hawaii.

4. INTERPRETATION OF RESULTS

4.1 Distribution of Phenotypic Values

The phenotypic distribution for a trait with $n = 1$, $p = .5$, $h^2 = .8$ and $x = 0$ is shown in Figure 4.1.1. This distribution was tri-modal and symmetric about its mean. Because of $p = .5$ and $x = 0$, symmetry of the phenotypic distribution was expected as the genotypic frequencies and genotypic values were:

Genotype	A_1A_1	A_1A_2	A_2A_2
Genotypic Frequency	.25	.5	.25
Genotypic Value	a	0	$-a$

Since heritability of the trait was high ($h^2 = .8$), the phenotypic variation was mainly determined by the genotypic values. When the number of locus was one ($n = 1$), heritability was .2 ($h^2 = .2$) and there was no dominance ($x = 0$), the phenotypic distribution was unimodal and close to a normal distribution (Figure 4.1.2). This distribution was also symmetric as gene frequency was .5 and $x = 0$ was assumed. Since heritability was low ($h^2 = .2$), the environmental effects which were assumed to be normally distributed became dominant in determining the phenotypic distribution. Under complete dominance ($x = 1$), the genotypic values for A_1A_1 and A_1A_2 were identical.

Genotype	A_1A_1, A_1A_2	A_2A_2
Genotypic Frequency	.75	.25
Genotypic Value	a	$-a$

The phenotypic distribution was bimodal and asymmetric as shown in Figure 4.1.3. When gene frequency deviated from .5 (p = .8), the phenotypic values gave a non-symmetric distribution due to different frequencies of three genotypes assuming the degree of dominance was zero (x = 0).

Genotype	A_1A_1	A_1A_2	A_2A_2
Genotypic Frequency	.64	.32	.04
Genotypic Value	a	0	$-a$

Figure 4.1.4 shows a distribution with p= .8 and h^2 = .8 in which expected frequency for genotypes A_1A_1, A_1A_2 and A_2A_2 were .64, .32 and .04, respectively, giving a trimodality.

The above examples demonstrated that gene frequency, heritability and degree of dominance did affect the phenotypic distribution of a quantitative trait. These three factors of phenotypic variations were further investigated by comparing distributions with various combinations of parameter values. The values considered were: .2, .5 and .8 for the heritability; .0, .5 and 1 for the degree of dominance; and .1, .3, and .5 for gene frequency. A total of twenty-seven distributions (Figure 4.1.5 through Figure 4.1.13) were compared. It was assumed that the phenotypic mean was one and the variance was unity. The expected frequencies of the three genotypes, A_1A_1, A_1A_2 and A_2A_2 as well as the mean of environmental effects (μ_e), genotypic values (a and d), and coefficients of skewness (λ_1) and kurtosis (λ_2) of each distribution are shown in Table 4.1.1. When heritability of a quantitative trait was low ($h^2 = .2$), the exact distribution of phenotypic values was almost normal regardless of parameter values for the gene frequency and the degree of dominance (Figure 4.1.5 to 4.1.7). This was expected since 80 percent of phenotypic variation was due to environmental effects which were assumed to be normally distributed. Given a model $P = G + E$ and assumptions stated in 3.1, the phenotypic variance σ_y^2 was the sum of the genetic variance σ_g^2 and the variance of environmental effects σ_e^2.

$$\sigma_y^2 = \sigma_g^2 + \sigma_e^2$$

As heritability ($h^2 = .5$) increased, the deviation of phenotypic distribution from normality was observed (Figure 4.1.8 to 4.1.10). The deviations of degree of dominance (x) from zero and gene frequency (p) from .5 caused asymmetry of the distribution and changed the height of the distribution. The departure from normality was measured by coefficients of skewness and kurtosis. Since these coefficients were expected to be zero for a normal distribution, the deviation from zero indicated the degree of non-normality. The formula for the coefficients of skewness and kurtosis are given in Appendix V. When the heritability of a trait was high ($h^2 = .8$), genotypic distribution was observed on the phenotypic distribution since the higher the heritability, the more weight the genetic effects were imposed on the phenotypic expression than the environmental effects (Figure 4.1.11 to 4.1.13). Three modes were apparent when the degree of dominance was zero ($x = 0$) and only two were observed when $x \geq .5$. The proportions of these modes and asymmetry of distribution were expected since these were the function of genotypic frequencies and genotypic values. By studying these twenty-seven distributions, the following observations were made:

(1) When the heritability was high, more than one mode presented in the phenotypic distribution was observable.

(2) When the gene frequency was .5 and the degree of dominance was zero, the distribution was symmetric.

(3) The deviation of gene frequency from .5 caused asymmetry of distribution.

(4) The deviation of degree of dominance from zero also caused asymmetry of distribution.

The phenotypic distributions for traits with frequency pairs (n=2) $p_1 = .2$ and $p_2 = .8$, $p_1 = p_2 = .5$ and $p_1 = p_2 = .8$ are shown in Figure 4.1.14. When $p_1 = p_2 = .5$, the distribution was almost normal; when $p_1 = .2$ and $p_2 = .8$, the distribution was symmetrical but it was leptokurtic; and when $p_1 = p_2 = .8$, the distribution was asymmetric, due basically to the asymmetry of gene frequencies. The expected genotypic frequencies and genotypic values given these three pairs of gene frequencies are shown in Table 4.1. 2.

The phenotypic distributions for traits with the number of loci $n = 1$, 5, and 10 assuming $p_i = .5$, $h^2 = .8$, and $x = 0$, are shown in

Figure 4.1.15. It appeared that, when the number of loci increased, the phenotypic distribution approached normal. Since a phenotype was a sum of random variables, the results were expected by the applications of the Central Limit Theorem.

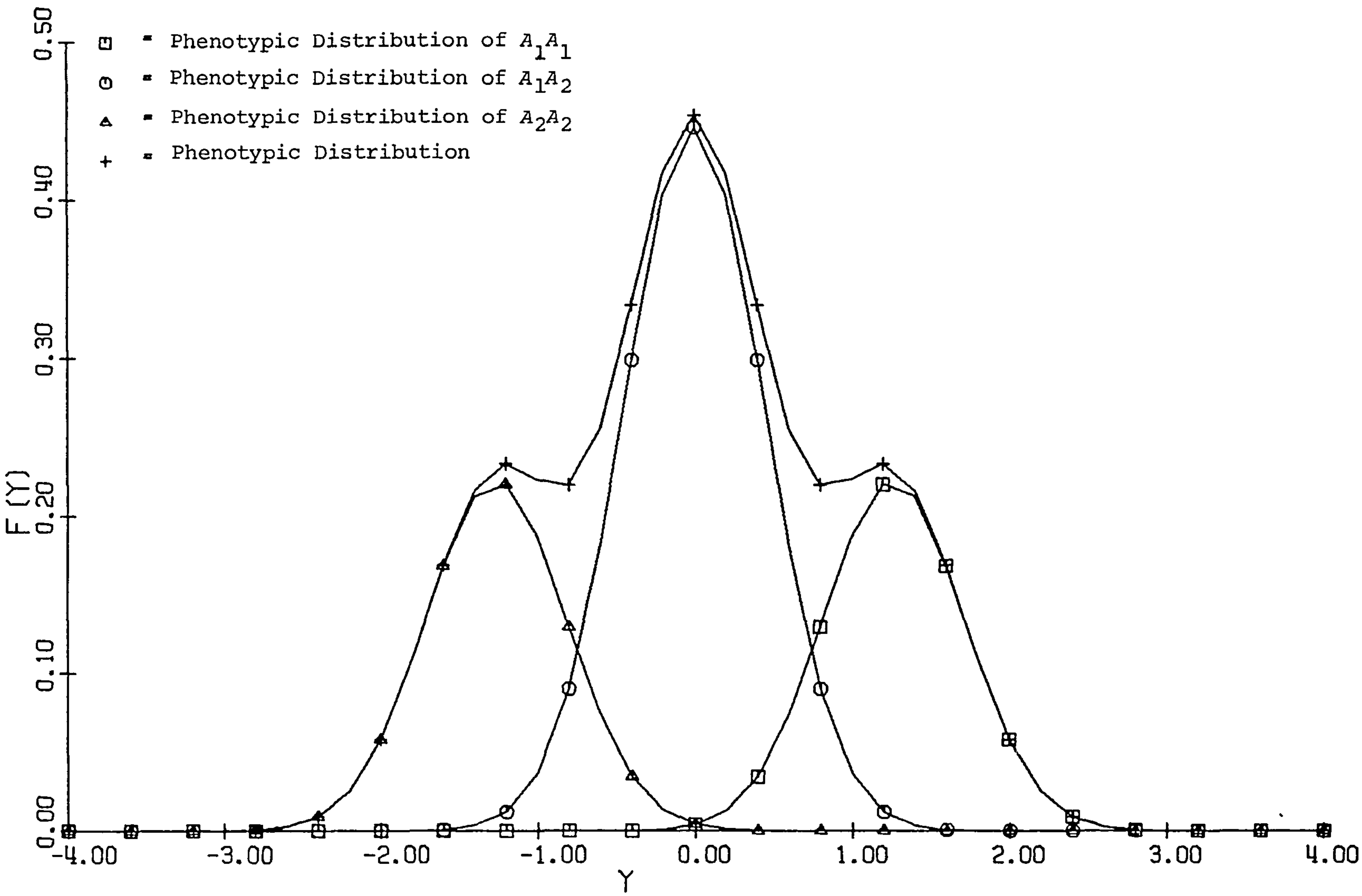


Figure 4.1.1 Phenotypic Distribution

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=1$, $p=.5$

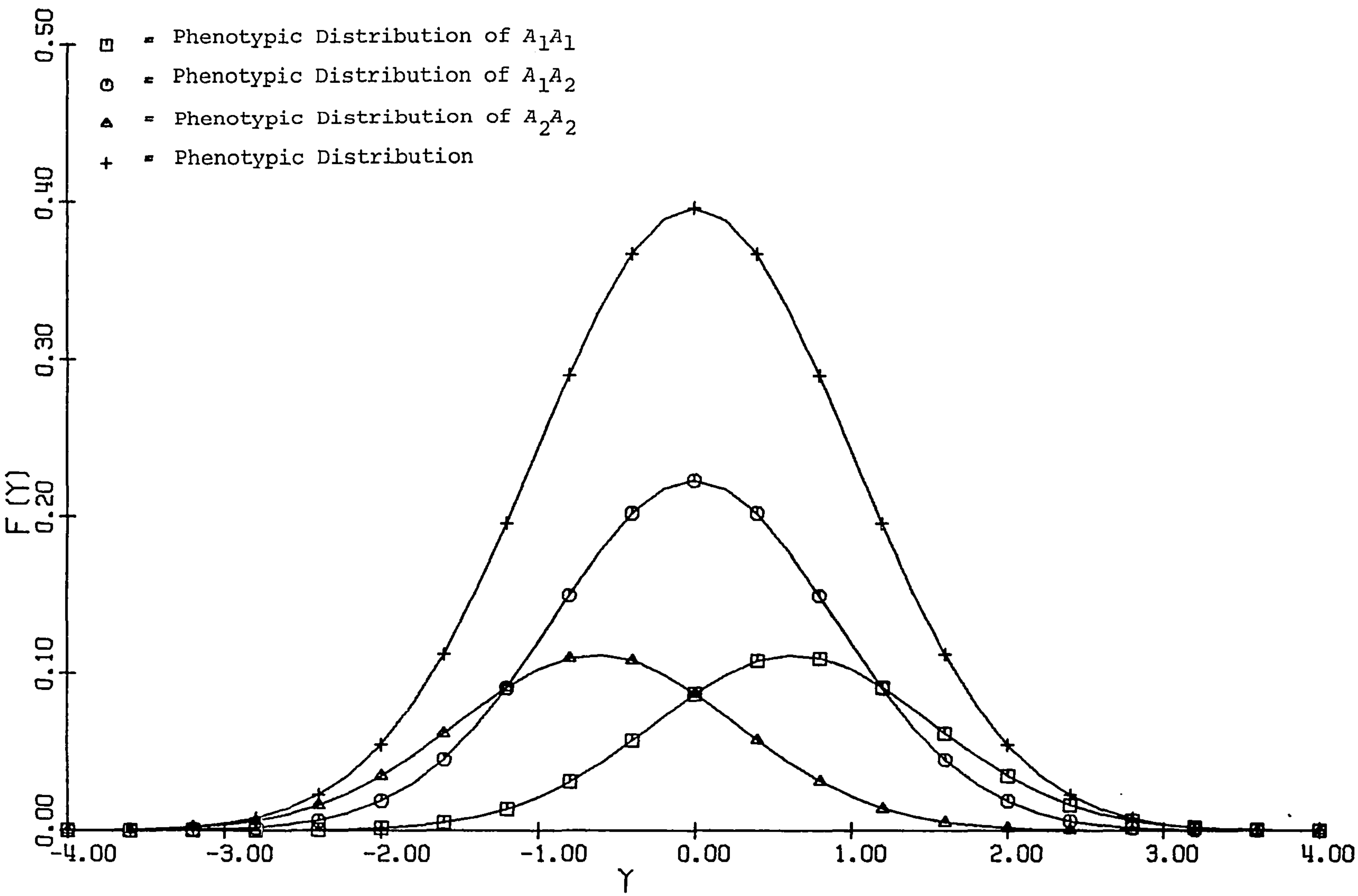


Figure 4.1.2 Phenotypic Distribution

Parameters: μ_y=0, σ_y^2=1, h^2=.2, x=0, n=1, p=.5

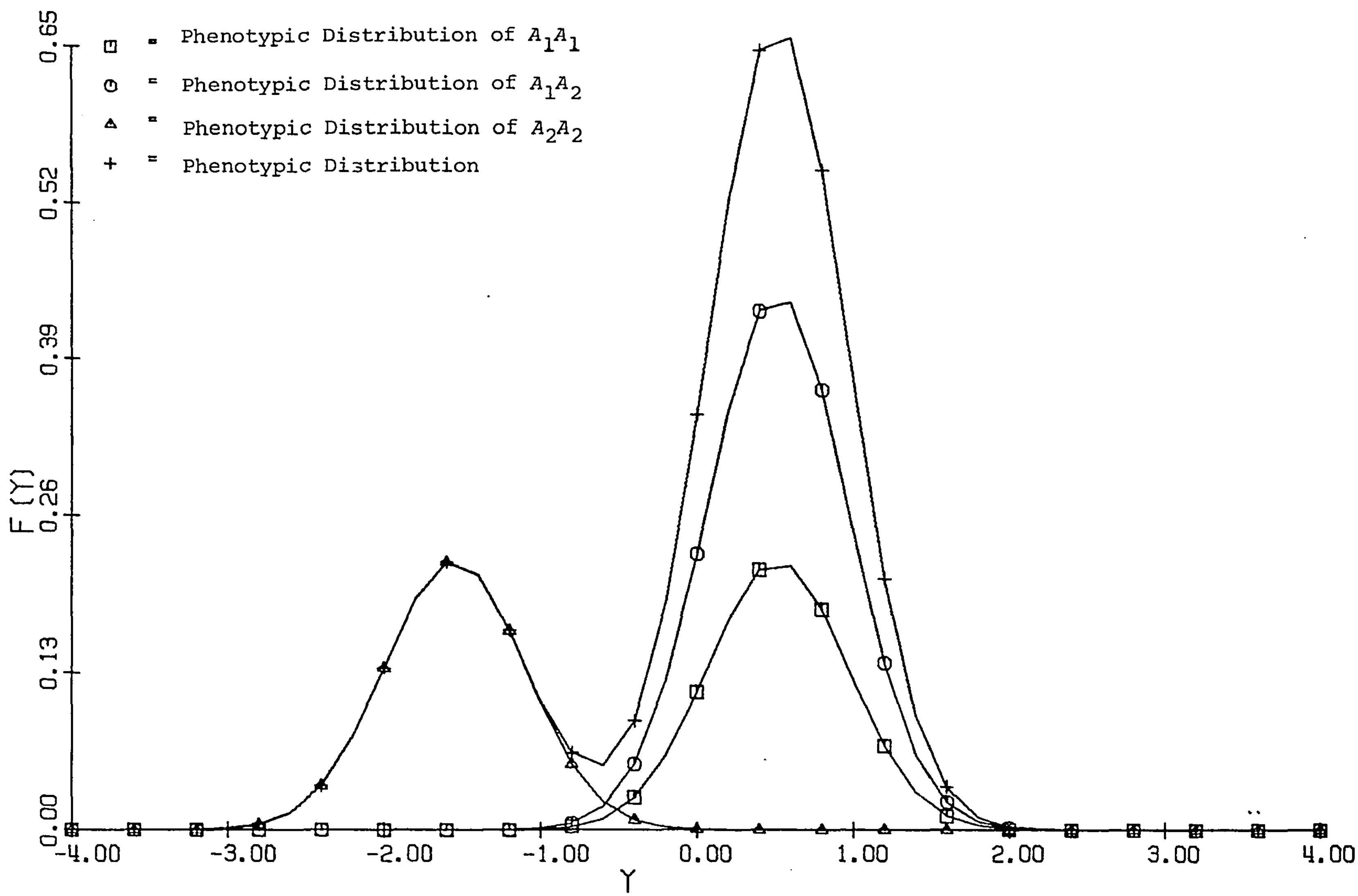


Figure 4.1.3 Phenotypic Distribution

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=1.0$, $n=1$, $p=.5$

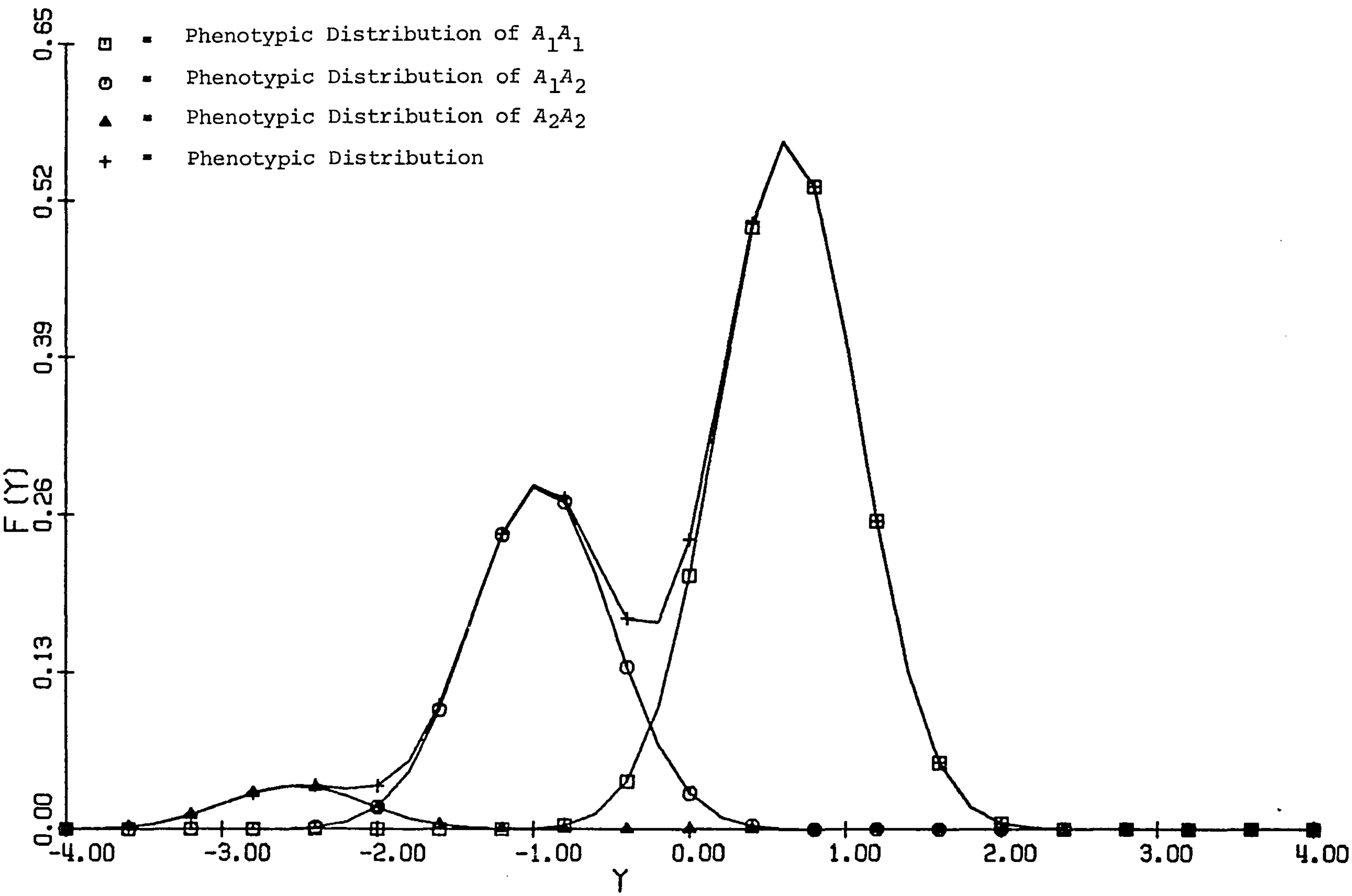


Figure 4.1.4 Phenotypic Distribution

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=1$, $p=.8$

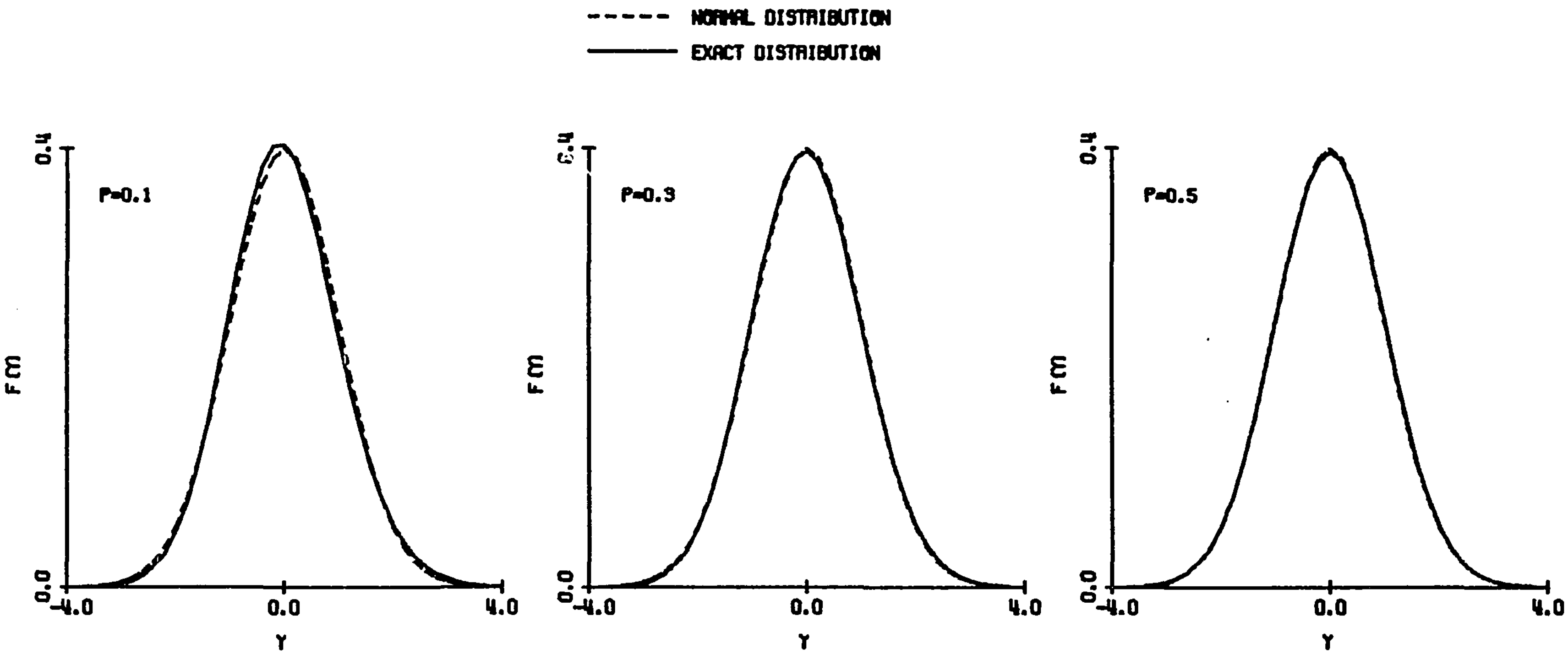


Figure 4.1.5 Phenotypic Distributions

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.2$, $x=0$, $n=1$

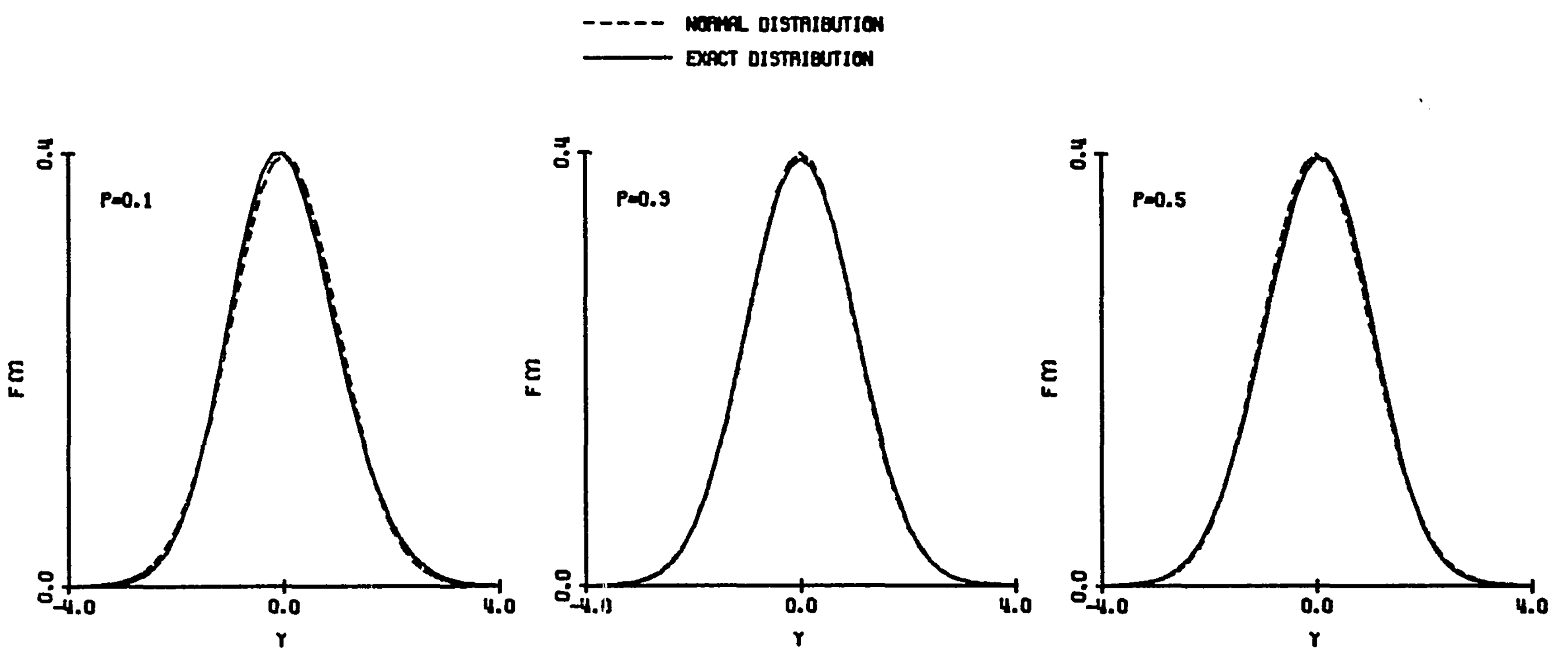


Figure 4.1.6 Phenotypic Distributions

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.2$, $x=.5$, $n=1$

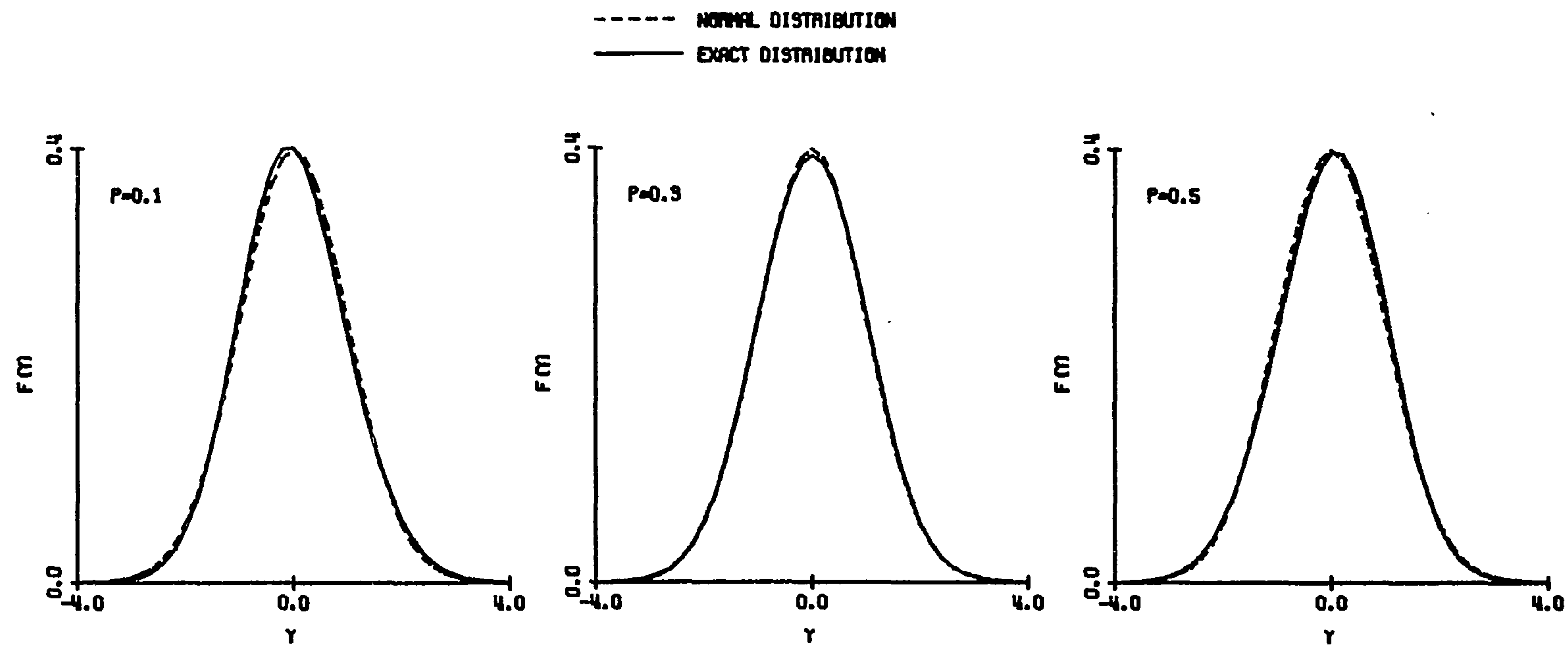


Figure 4.1.7 Phenotypic Distributions

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.2$, $x=1$, $n=1$

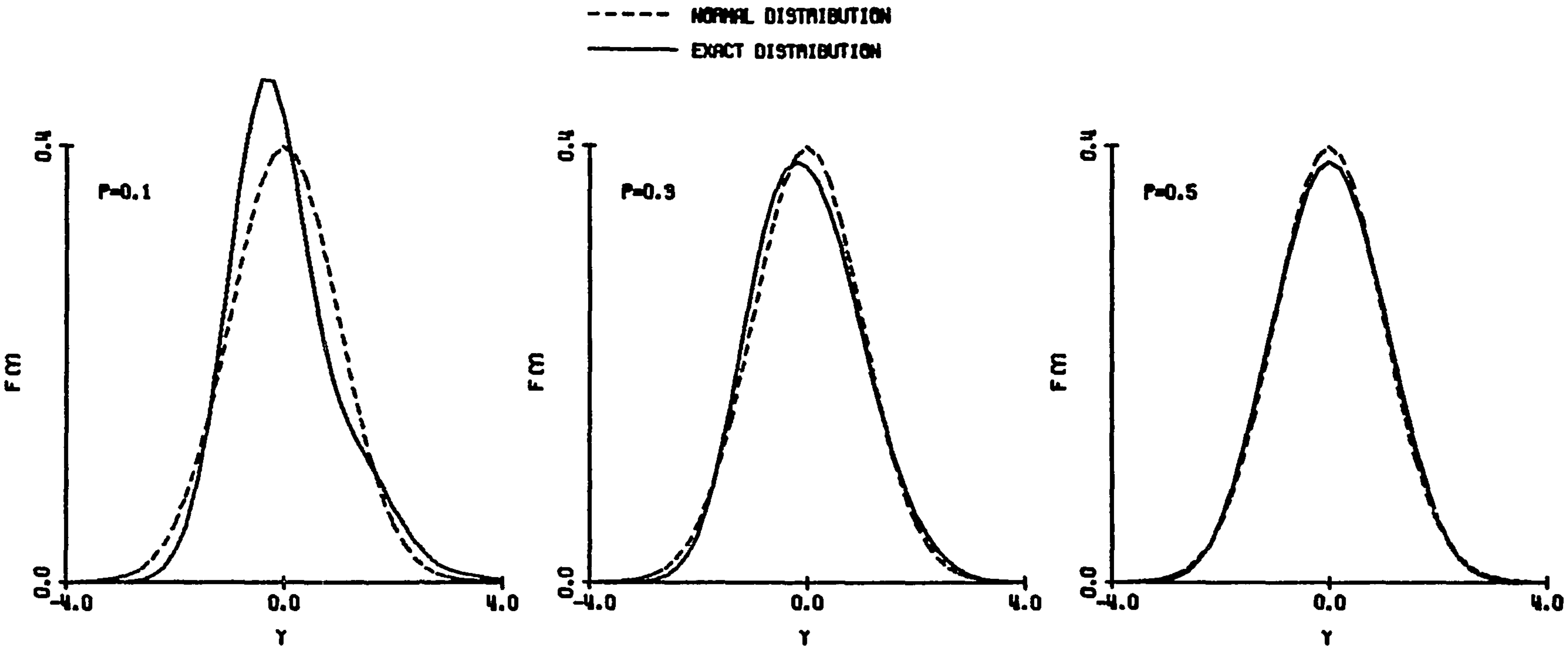


Figure 4.1.8 Phenotypic Distributions

Parameters: μ_y=0, σ_y^2=1, h^2=.5, x=0, n=1

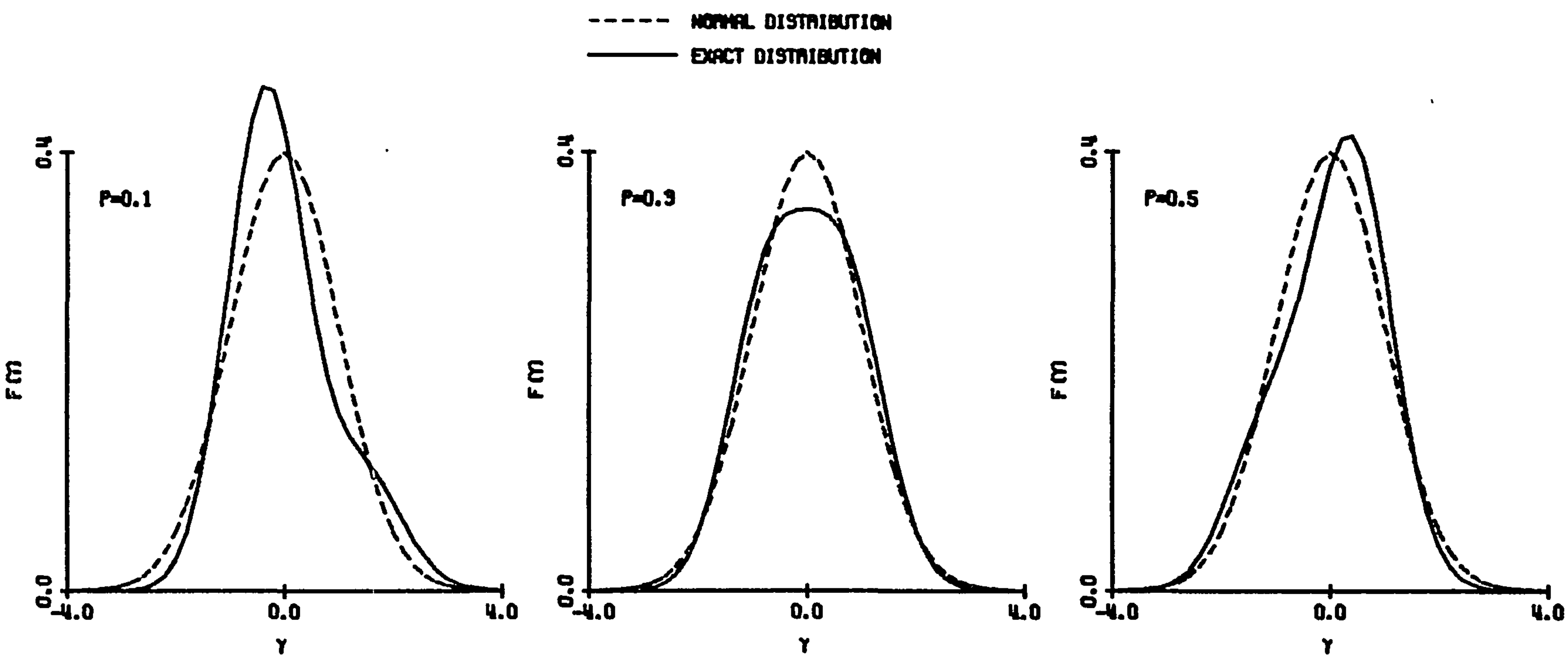


Figure 4.1.9 Phenotypic Distributions

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.5$, $x=.5$, $n=1$

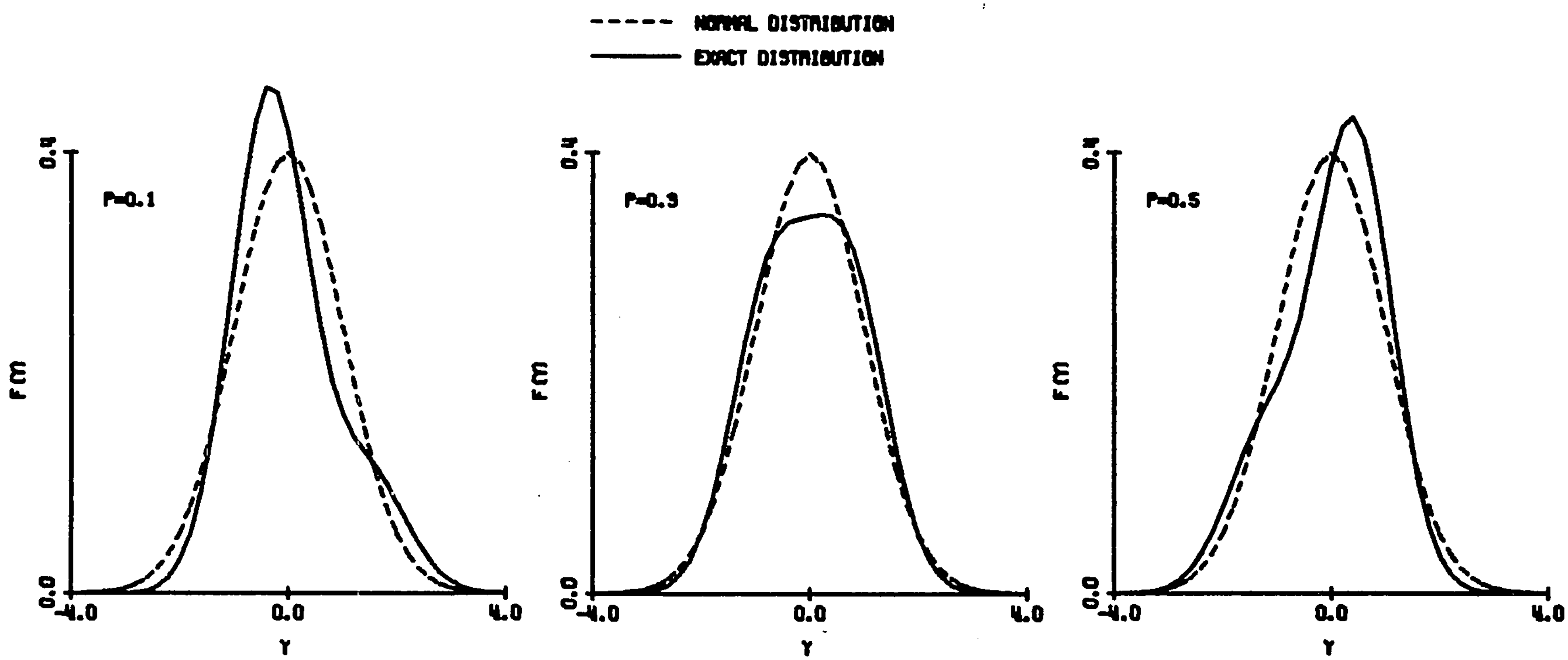


Figure 4.1.10 Phenotypic Distributions

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.5$, $x=1$, $n=1$

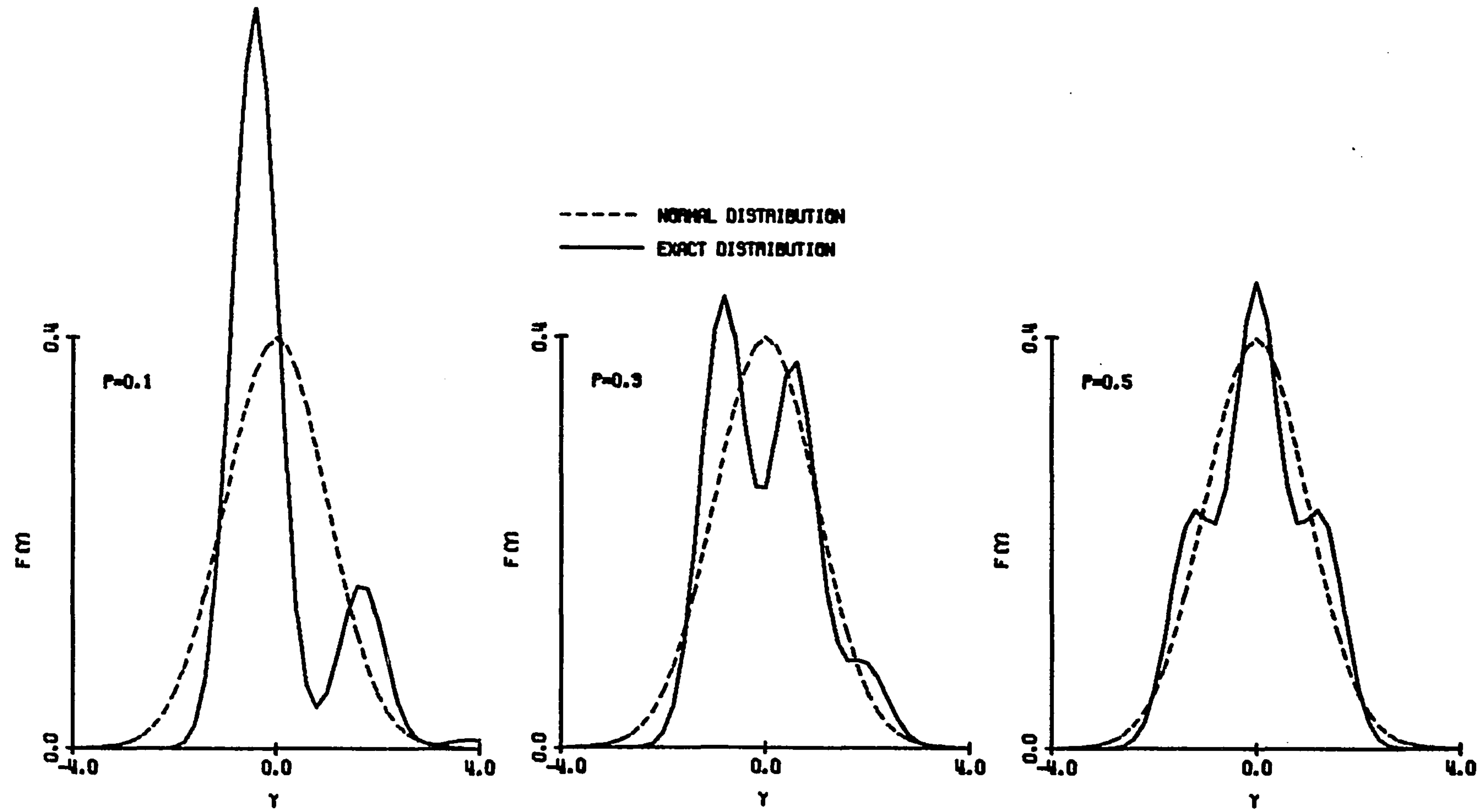


Figure 4.1.11 Phenotypic Distributions

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=1$

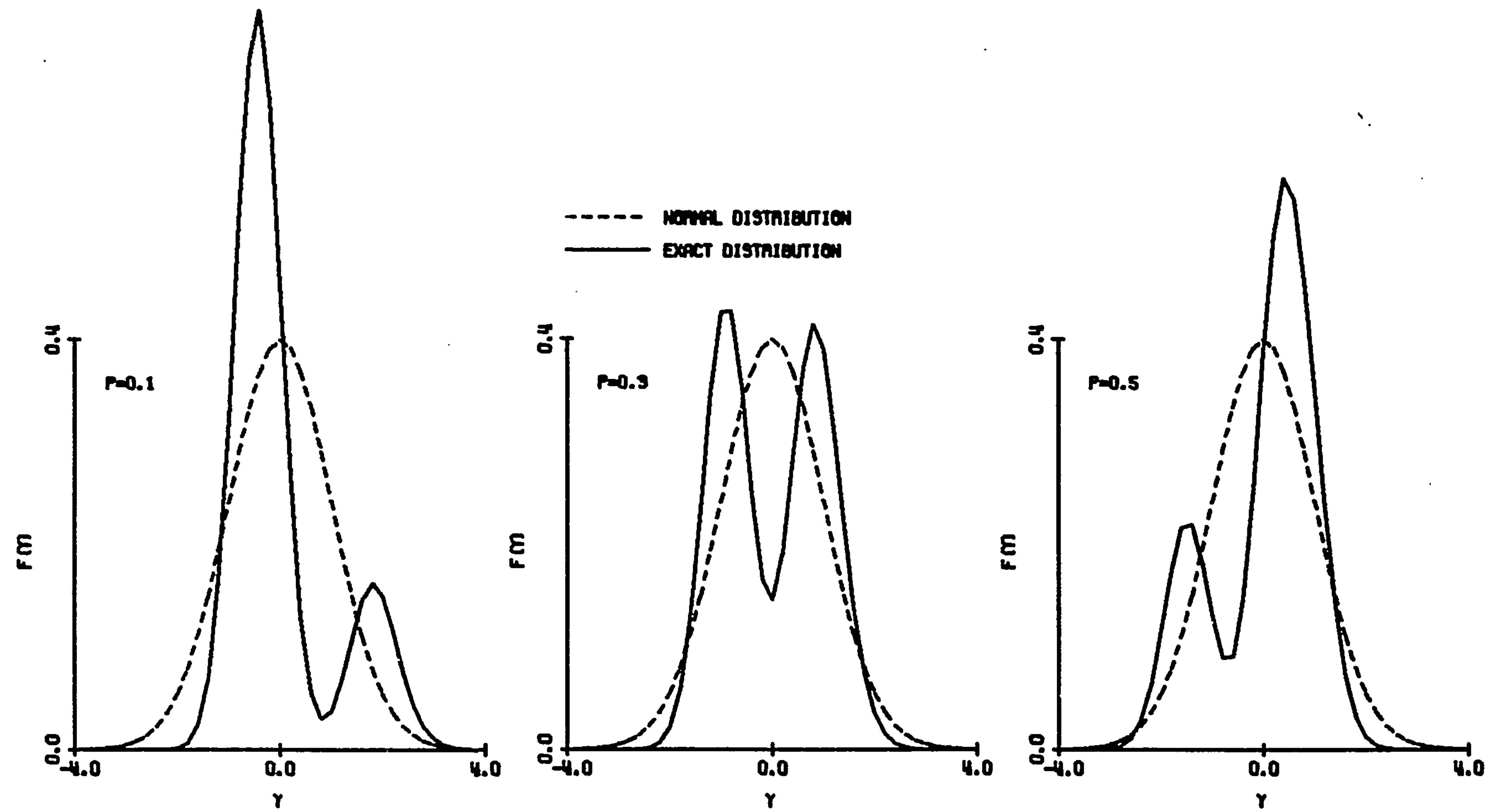


Figure 4.1.12 Phenotypic Distributions

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=.5$, $n=1$

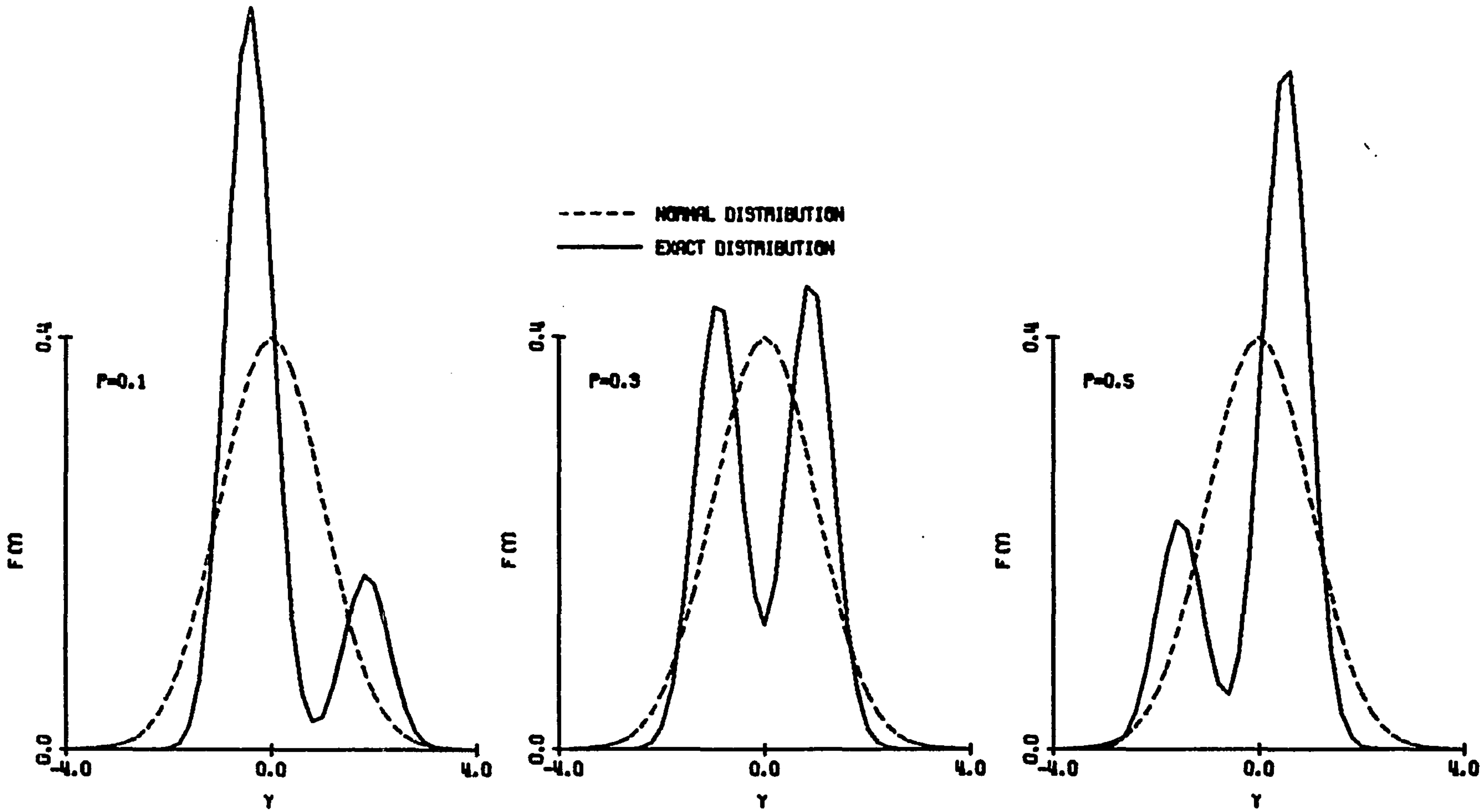


Figure 4.1.13 Phenotypic Distributions

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=1$, $n=1$

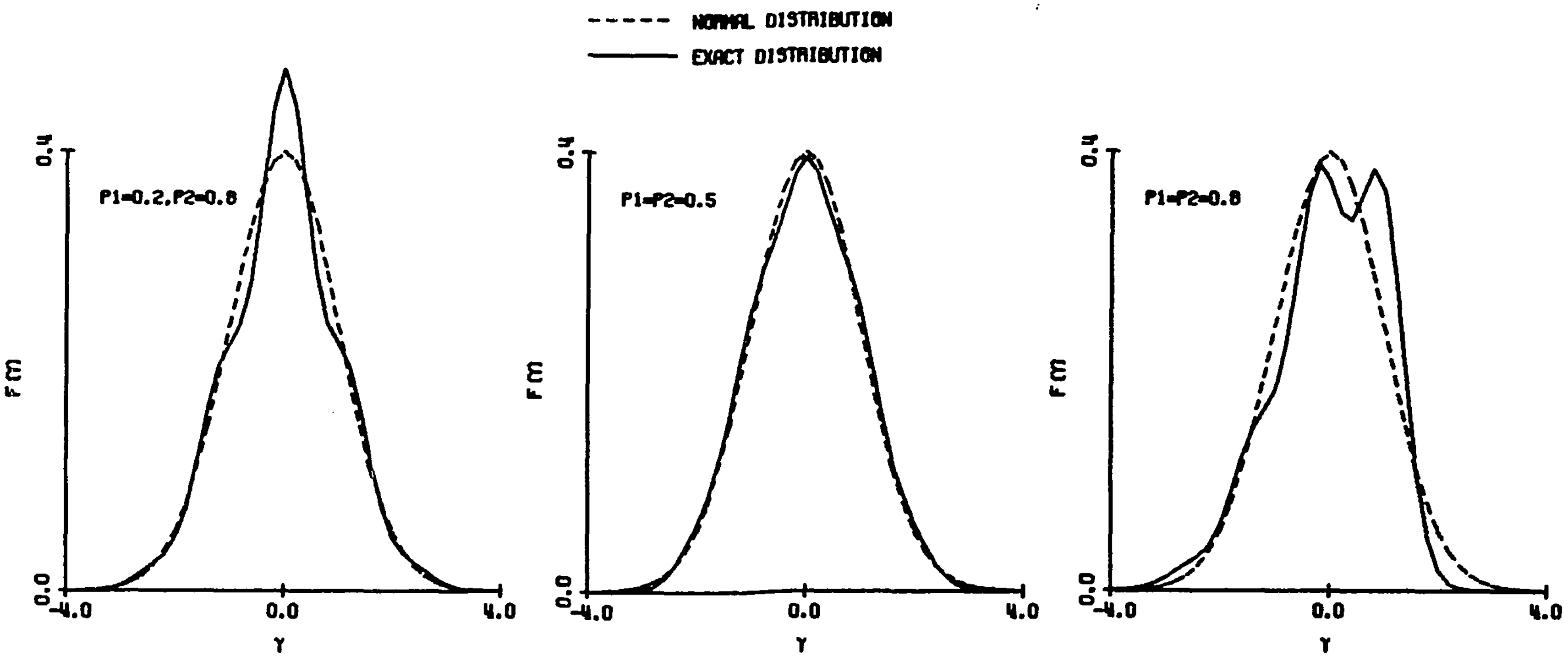


Figure 4.1.14 Phenotypic Distributions

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=2$

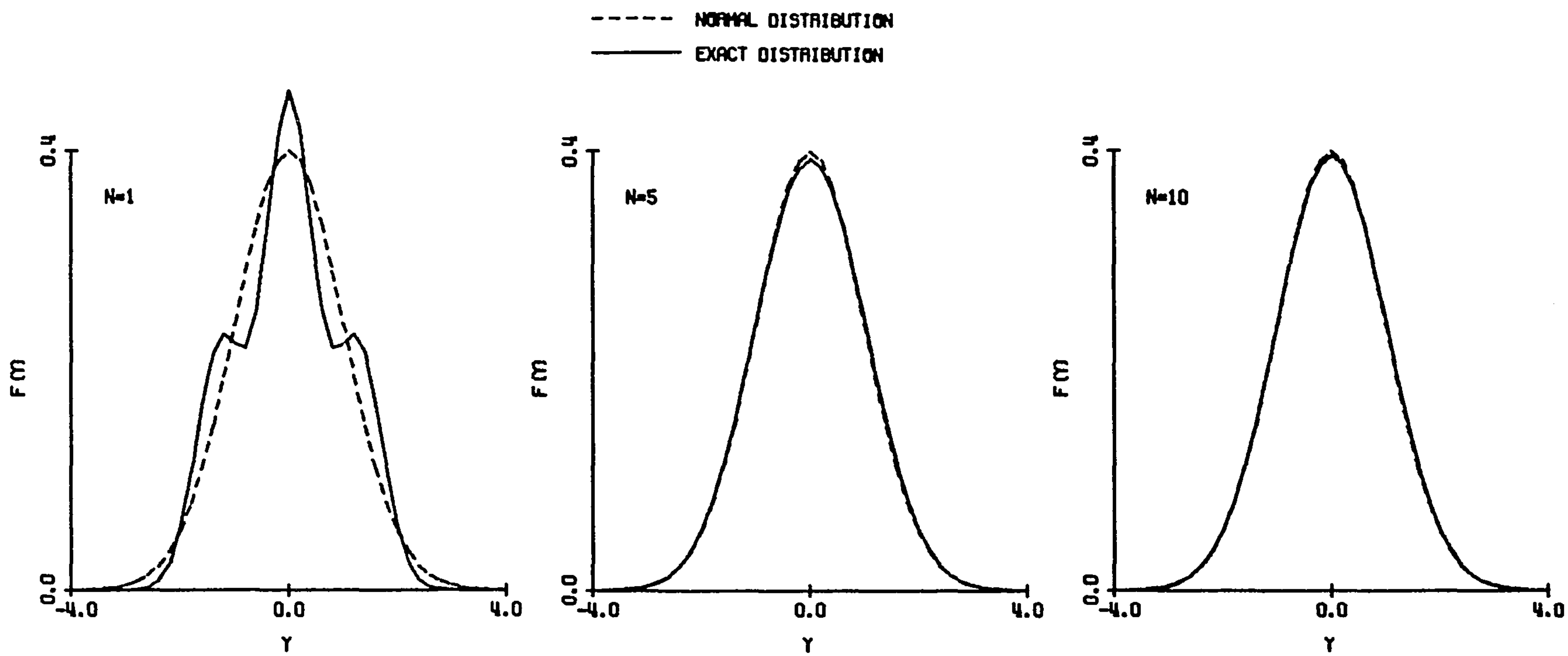


Figure 4.1.15 Phenotypic Distributions

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $p_i=.5$ for i = 1 to n

Table 4.1.1

Parameter Values for Exact Distribution

Parameters								Genotypic Frequency		
h^2	x	p	μ_e	a	d	λ_1	λ_2	A_1A_1	A_1A_2	A_2A_2
.2	.0	.1	.843	1.054	.0	.168	.102	.01	.18	.81
.2	.0	.3	.276	.690	.0	.05	-.024	.09	.42	.49
.2	.0	.5	.0	.632	.0	.0	-.040	.25	.50	.25
.2	.5	.1	.528	.744	.372	.144	.026	.01	.18	.81
.2	.5	.3	.105	.555	.277	.005	-.073	.09	.42	.49
.2	.5	.5	-.149	.596	.298	-.079	-.032	.25	.50	.25
.2	1.0	.1	.353	.569	.569	.141	.019	.01	.18	.81
.2	1.0	.3	-.008	.447	.447	-.003	-.079	.09	.42	.49
.2	1.0	.5	-.258	.516	.516	-.103	-.026	.25	.50	.25
.5	.0	.1	1.333	1.666	.0	.666	.638	.01	.18	.81
.5	.0	.3	.436	1.091	.0	.218	-.154	.09	.42	.49
.5	.0	.5	.0	1.0	.0	.0	-.250	.25	.50	.25

Table 4.1.1--Continued

				Parameters				Genotypic Frequency		
h^2	x	p	μ_e	a	d	λ_1	λ_2	A_1A_1	A_1A_2	A_2A_2
.5	.5	.1	.835	1.177	.588	.569	.165	.01	.18	.81
.5	.5	.3	.166	.877	.438	.019	-.461	.09	.42	.49
.5	.5	.5	-.235	.942	.471	-.314	-.203	.25	.50	.25
.5	1.0	.1	.558	.901	.901	.558	.124	.01	.18	.81
.5	1.0	.3	-.014	.707	.707	-.014	-.499	.09	.42	.49
.5	1.0	.5	-.408	.816	.816	-.408	-.166	.25	.50	.25
.8	.0	.1	1.686	2.108	.0	1.349	1.635	.01	.18	.81
.8	.0	.3	.552	1.380	.0	.441	-.396	.09	.42	.49
.8	.0	.5	.0	1.264	.0	.0	-.640	.25	.50	.25
.8	.5	.1	1.057	1.488	.744	1.152	.424	.01	.18	.81
.8	.5	.3	.210	1.110	.555	.040	-1.180	.09	.42	.49
.8	.5	.5	-.298	1.192	.596	-.636	-.521	.25	.50	.25
.8	1.0	.1	.706	1.139	1.139	1.130	.318	.01	.18	.81
.8	1.0	.3	-.017	.894	.894	-.028	-1.278	.09	.42	.49
.8	1.0	.5	-.516	1.032	1.032	-.826	-.426	.25	.50	.25

Table 4.1.2

Genotypic Values and Genotypic Frequencies

Genotype	Genotypic Value	Genotypic Frequency $p_1=p_2=.5$	$p_1=.2$, $p_2=.8$	$p_1=p_2=.8$
$A_1A_1B_1B_1$	$2a$	.0625	.0256	.4096
$A_1A_1B_1B_2$ $A_1A_2B_1B_1$	a	.25	.2176	.4096
$A_1A_1B_2B_2$ $A_1A_2B_1B_2$ $A_2A_2B_1B_1$	0	.375	.5136	.1536
$A_1A_2B_2B_2$ $A_2A_2B_1B_2$	$-a$	.25	.2176	.0256
$A_2A_2B_2B_2$	$-2a$	.0625	.0256	.0016

4.2 Descriptive Statistics of Simulated Data

A total of one hundred-twenty samples were generated for twenty-four sets of given values of parameters, heritability (h^2), degree of dominance (x), number of loci (n), gene frequencies (p_i for i = 1 to n), degree of linkage disequilibrium (D*) and sample size (N). There were five samples in each set and each sample was generated in different gene frequencies based on random deviates between zero and one. The parameter values for h^2, x, n, D* and N of each set were described in Table 3.5. The sample mean $\hat{\mu}_y$, sample variance $\hat{\sigma}_y^2$, gene frequencies p_i for i = 1 to n, minimum and maximum values (Y_{min} and Y_{max}) and the results of the normality test for each sample are given in Table 4.2.1. Since the parameters of the phenotypic mean and the variance were zero and unity respectively for data simulation, sample statistics showed good agreement. Figure 4.2.1 shows examples of phenotypic distributions as observed from generated data (the first samples of data set 1 to 8). As it was observed in theoretical distributions (4.1), when the parameter value of heritability was low (data 4-1) and the number of loci was large (data 5-1 and data 6-1), distributions were close to normal. The data 2-1 was skewed and multimodal as x = 1 was assumed. The coefficients of skewness (λ_1) and kurtosis (λ_2) for the eight sample distributions were:

Data	λ_1	λ_2
1-1	-.083	-.295
2-1	.606	-.221
3-1	.167	-.067
4-1	.010	-.045
5-1	-.053	-.138

Data	λ_1	λ_2
6-1	-.133	-.116
7-1	-.102	-.565
8-1	.004	-.219

Sample variations of distributions are shown in Figure 4.2.2 (five samples of data set 1). These variations were due only to random environmental effects and different pairs of gene frequencies which were chosen at random.

Table 4.2.1

Descriptive Statistics

Data Set	Sample	p_i		$\hat{\mu}_y$	$\hat{\sigma}^2_y$	Y_{min}	Y_{max}	Non-normality
1	1	.74	.46	.002	1.017	-3.356	2.693	
	2	.63	.51	-.012	.986	-3.175	2.911	
	3	.14	.45	-.002	.990	-2.788	3.751	
	4	.45	.38	.013	1.010	-2.774	3.283	**
	5	.26	.12	.004	.979	-2.232	4.448	**
2	1	.11	.04	.024	.989	-2.119	3.820	**
	2	.24	.61	-.012	1.040	-3.270	2.354	**
	3	.27	.12	-.012	1.000	-2.387	3.144	**
	4	.85	.70	-.001	1.004	-6.225	1.895	**
	5	.61	.29	.014	.954	-3.508	2.275	**
3	1	.06	.35	-.029	.972	-3.337	3.606	
	2	.11	.25	-.002	1.021	-3.164	4.069	*
	3	.03	.27	-.020	1.007	-2.627	3.835	*
	4	.43	.19	-.003	.974	-3.379	3.426	
	5	.52	.71	.007	.981	-3.865	3.192	
4	1	.06	.35	-.023	.986	-3.725	3.315	
	2	.99	.99	.007	.990	-4.807	3.325	**
	3	.05	.77	.018	.959	-4.037	3.004	
	4	.36	.37	-.016	.985	-3.959	3.824	
	5	.43	.50	.016	.994	-3.212	3.367	

Table 4.2.1--Continued

Data Set	Sample	p_i	$\hat{\mu}_y$	$\hat{\sigma}^2_y$	Y_{min}	Y_{max}	Non-normality
5	1	.15 .67 .61 .64 .31	-.000	1.028	-3.869	3.102	
	2	.81 .44 .26 .64 .44	-.030	.992	-3.709	3.242	
	3	.55 .54 .26 .72 .94	-.003	.971	-3.658	3.726	
	4	.78 .73 .33 .40 .43	-.015	.999	-3.809	3.293	
	5	.38 .48 .45 .38 .21	-.016	1.004	-3.377	3.907	
6	1	.66 .98 .92 .67 .73	-.003	.978	-3.715	3.181	
		.39 .74 .85 .48 .17					
	2	.59 .83 .67 .53 .17	.020	1.026	-3.694	3.563	
		.21 .69 .19 .92 .80					
	3	.12 .91 .36 .96 .46	-.003	1.041	-3.564	3.519	
		.10 .45 .80 .73 .19					
	4	.66 .98 .94 .75 .06	-.006	1.027	-4.023	3.293	
		.55 .78 .74 .33 .35					
	5	.78 .16 .95 .19 .63	-.020	1.013	-3.116	3.524	
		.02 .42 .33 .15 .90					
7	1	.56 .91	-.040	.984	-2.876	2.154	
	2	.94 .10	.043	.907	-2.809	3.314	**
	3	.39 .62	.023	1.046	-2.951	2.749	
	4	.17 .35	.027	.961	-1.859	2.882	
	5	.94 .02	-.016	.995	-3.146	3.337	**

Table 4.2.1--Continued

Data Set	Sample	p_i		$\hat{\mu}_y$	$\hat{\sigma}^2_y$	Y_{min}	Y_{max}	Non-normality
8	1	.06	.87	.007	.995	-2.372	2.495	
	2	.52	.21	.177	.772	-2.159	2.287	
	3	.42	.03	.074	1.032	-2.022	2.618	
	4	.49	.95	.097	1.144	-2.346	2.204	
	5	.24	.43	-.157	.893	-2.017	2.280	
9	1	.29	.08	-.024	.984	-2.571	4.313	*
	2	.11	.73	-.007	1.000	-3.091	3.521	**
	3	.46	.19	.034	1.001	-2.490	3.585	
	4	.94	.72	.011	1.001	-4.572	2.312	**
	5	.83	.39	-.021	1.029	-3.426	3.027	
10	1	.01	.50	.002	.991	-2.946	3.355	**
	2	.66	.13	-.010	.991	-3.327	2.751	**
	3	.77	.36	.004	1.009	-3.741	2.214	**
	4	.38	.84	-.019	1.007	-4.420	2.357	**
	5	.51	.19	.003	.980	-2.816	2.772	**
11	1	.08	.55	-.014	.992	-3.371	3.568	
	2	.29	.31	-.008	.986	-3.230	3.516	
	3	.98	.11	.013	.974	-3.966	3.955	**
	4	.88	.01	-.004	.953	-4.223	3.052	**
	5	.87	.89	.016	1.016	-4.353	3.172	**

Table 4.2.1--Continued

Data Set	Sample	p_i					$\hat{\mu}_y$	$\hat{\sigma}^2_y$	Y_{min}	Y_{max}	Non-normality
12	1	.08	.55				-.012	.988	-3.559	3.407	
	2	.50	.83				-.001	1.011	-3.663	3.649	
	3	.47	.61				-.011	.987	-3.353	3.266	
	4	.95	.41				-.015	1.000	-3.667	3.068	
	5	.70	.80				.006	.981	-3.503	3.259	
13	1	.13	.61	.31	.67	.82	.008	.967	-3.429	3.242	
	2	.36	.34	.40	.65	.28	.015	.999	-3.357	3.450	
	3	.45	.52	.39	.82	.80	.019	1.000	-3.425	3.267	
	4	.96	.50	.19	.63	.21	-.023	.999	-3.273	4.134	
	5	.19	.34	.23	.26	.18	-.020	.979	-2.815	3.423	**
14	1	.80	.88	.17	.57	.59	.008	.998	-3.541	3.310	
		.87	.96	.42	.40	.37					
	2	.32	.18	.83	.09	.70	.012	1.016	-3.377	3.755	
		.34	.00	.87	.27	.79					
	3	.81	.17	.33	.79	.40	-.010	1.010	-3.257	3.787	
		.09	.65	.09	.47	.61					
	4	.89	.23	.85	.26	.65	-.020	1.016	-3.804	4.061	
		.51	.28	.13	.17	.15					
	5	.30	.03	.11	.70	.08	-.009	.993	-3.314	4.090	
		.08	.86	.50	.72	.47					

Table 4.2.1--Continued

Data Set	Sample	p_i		$\hat{\mu}_y$	$\hat{\sigma}^2_y$	Y_{min}	Y_{max}	Non-normality
15	1	.85	.10	-.027	1.077	-2.974	3.232	
	2	.51	.35	.078	.927	-2.382	2.931	
	3	.82	.88	-.031	1.039	-4.966	1.958	**
	4	.79	.29	-.099	.884	-3.285	2.683	
	5	.66	.85	.080	.967	-3.057	2.475	
16	1	.05	.20	.072	.969	-1.572	3.742	
	2	.36	.31	-.042	.950	-2.029	2.279	
	3	.69	.48	.052	1.100	-3.145	2.147	
	4	.30	.01	.014	.823	-1.599	2.294	
	5	.38	.11	.079	1.151	-2.113	2.517	
17	1	.50	.08	-.010	1.011	-2.520	3.791	
	2	.50	.84	.031	.977	-3.781	2.711	*
	3	.50	.21	.014	.990	-2.850	3.540	
	4	.50	.46	.011	.996	-3.409	3.104	
	5	.50	.79	.009	.970	-3.576	2.791	
18	1	.50	.69	-.002	1.033	-3.980	2.146	**
	2	.50	.76	-.006	.998	-4.071	1.961	**
	3	.50	.74	.000	.987	-3.778	2.209	**
	4	.50	.10	-.004	1.003	-2.710	2.749	**
	5	.50	.78	.000	.982	-4.022	2.136	**

Table 4.2.1--Continued

Data Set	Sample	p_i	$\hat{\mu}_y$	$\hat{\sigma}_y^2$	Y_{min}	Y_{max}	Non-normality
19	1	.50 .55	-.032	1.007	-3.434	3.289	
	2	.50 .80	-.005	.982	-3.253	3.322	
	3	.50 .24	.020	.999	-3.190	4.171	
	4	.50 .15	-.010	1.034	-3.116	3.951	
	5	.50 .95	.007	1.001	-3.971	3.243	
20	1	.50 .55	-.026	.999	-3.588	3.309	
	2	.50 .51	-.000	.974	-3.649	3.524	
	3	.50 .16	.017	.990	-3.108	3.521	
	4	.50 .46	-.005	.998	-3.497	3.365	
	5	.50 .60	.013	.987	-3.556	3.443	
21	1	.50 .61 .31 .67 .82	.009	1.038	-3.463	3.410	
	2	.50 .29 .30 .87 .23	.007	.973	-3.367	3.685	
	3	.50 .03 .30 .63 .78	.001	1.020	-3.302	3.620	
	4	.50 .05 .36 .69 .19	.011	.986	-3.040	3.733	
	5	.50 .73 .41 .79 .46	-.005	.966	-3.123	3.393	
22	1	.50 .92 .73 .74 .48	.001	.988	-3.865	3.545	
		.70 .54 .33 .23 .30					
	2	.50 .72 .91 .97 .79	-.010	.971	-3.272	3.926	
		.20 .58 .58 .57 .52					
	3	.50 .57 .99 .61 .42	.002	.989	-3.564	3.750	
		.53 .91 .03 .59 .19					
	4	.50 .73 .01 .47 .03	-.013	.993	-3.594	3.329	
		.77 .06 .58 .55 .76					
	5	.50 .62 .45 .83 .02	.003	1.038	-3.965	3.725	
		.75 .70 .93 .35 .67					

Table 4.2.1--Continued

Data Set	Sample	p_i	$\hat{\mu}_y$	$\hat{\sigma}^2_y$	Y_{min}	Y_{max}	Non-normality
23	1	.50 .91	.026	.969	-2.376	2.482	
	2	.50 .92	-.049	1.029	-2.729	2.712	
	3	.50 .20	.013	.981	-2.519	3.186	
	4	.50 .08	-.008	1.008	-2.379	2.866	
	5	.50 .11	-.009	1.065	-2.690	3.575	
24	1	.65 .89	-.039	1.123	-3.332	1.949	
	2	.66 .54	.191	1.094	-2.703	2.464	
	3	.18 .24	.020	1.061	-1.550	3.111	*
	4	.05 .19	-.202	.888	-1.527	3.158	**
	5	.42 .33	-.029	1.009	-2.073	2.937	

* significant at p = .05 level
** significant at p = .01 level

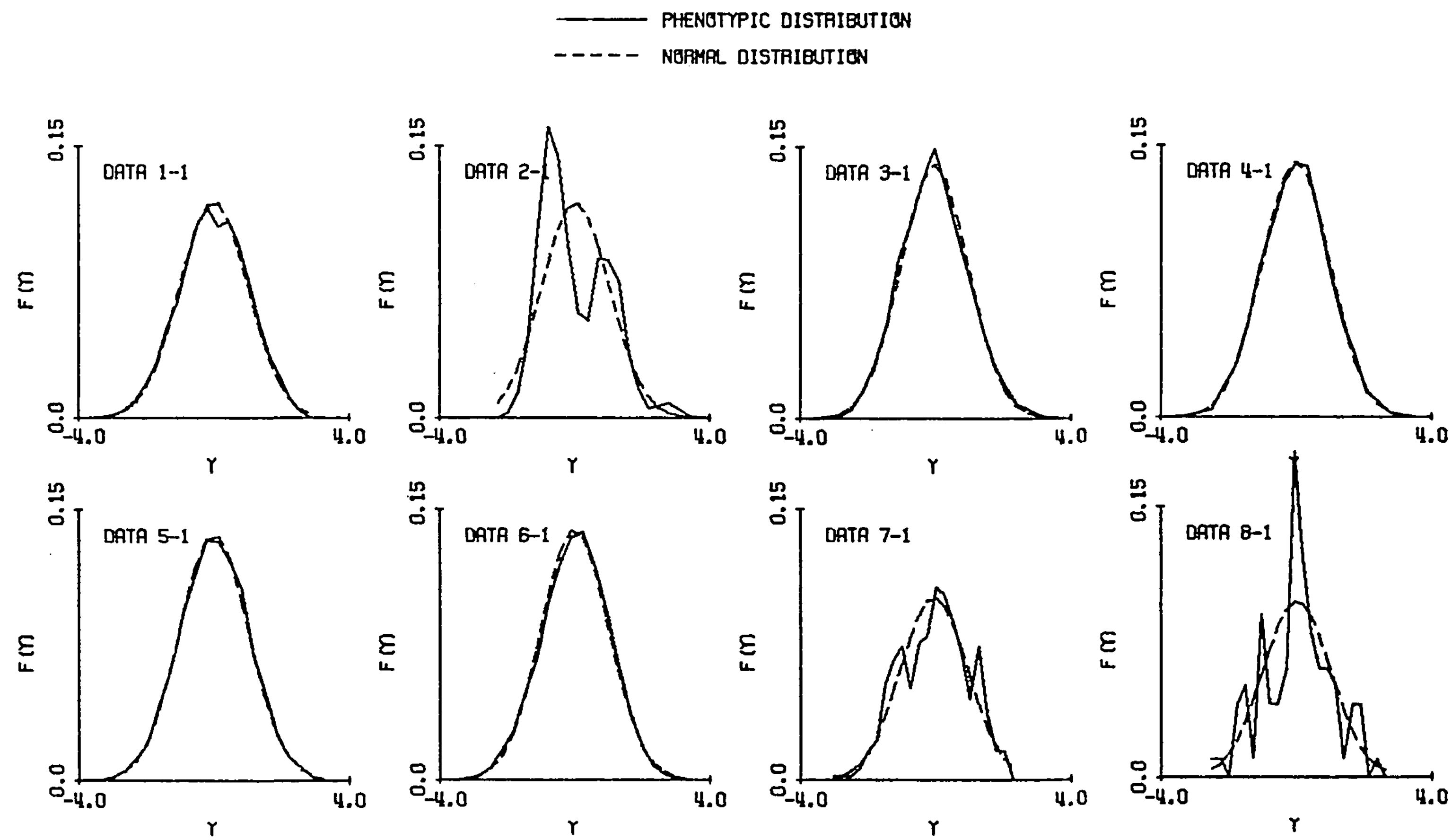


Figure 4.2.1 Observed Phenotypic Distributions

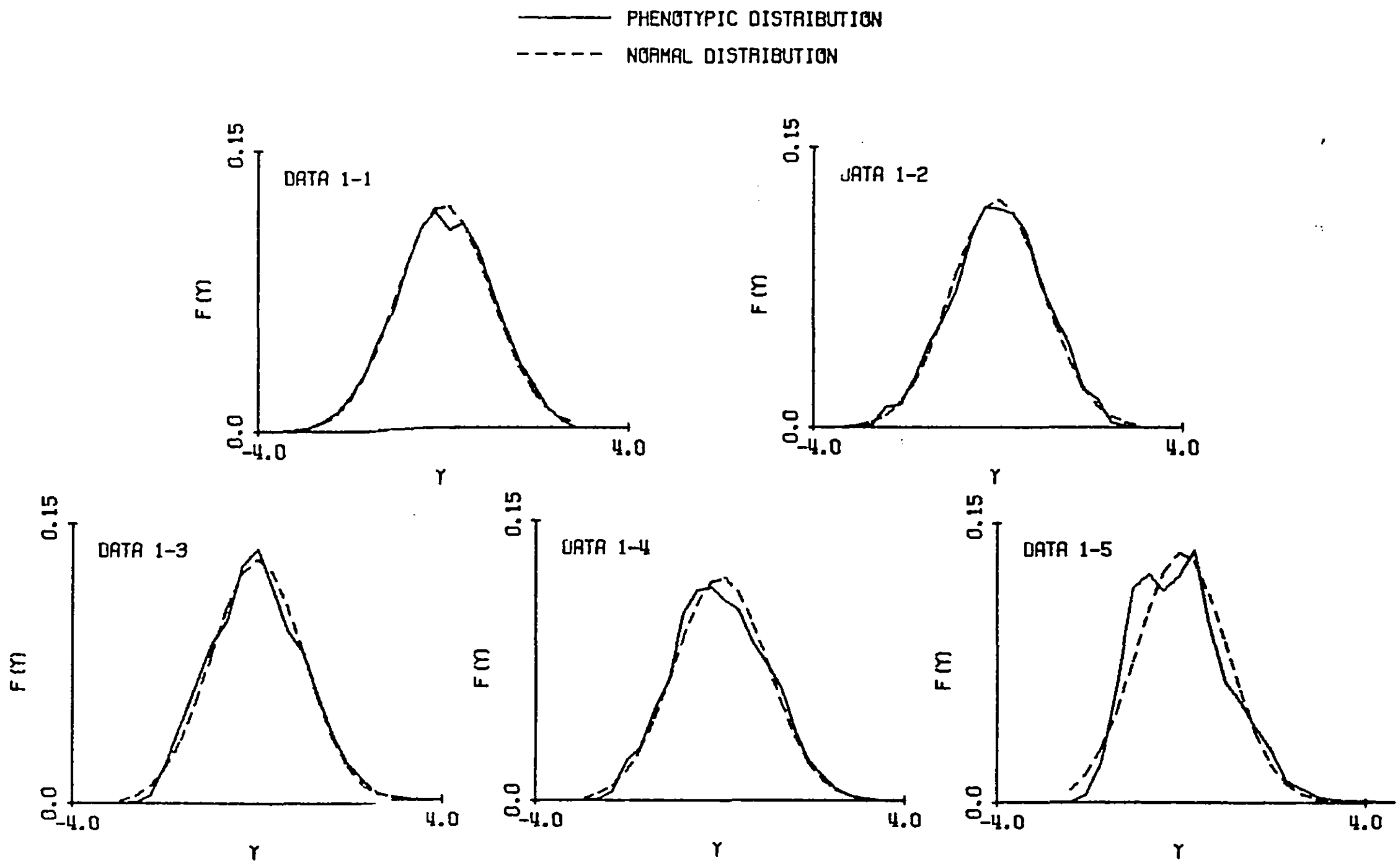


Figure 4.2.2 Variations of Observed Phenotypic Distributions

4.3 Estimation of Genetic Parameters Based on Phenotypic Values Only

Thirty-three of one hundred-twenty samples were found to be significantly different from normal distribution. Departure from normality occurred when the degree of dominance was assumed to be one ($x = 1$) as it was shown in data sets 2, 10 and 18. The results were expected since the degree of dominance was one of the factors which caused non-normality of phenotypic distribution. All but one of the sample distributions were not significantly different from normal when $n = 5$ and 10 (data sets 5, 6, 13, 14, 21, 22). These results were also expected as the exact distribution of phenotypic values approached normal when the number of loci became large by the Central Limit Theorem. When the heritability was assumed to be low ($h^2 = .2$), most of the distributions were normal (data sets 4, 12, and 20), whereas, when the heritability was higher ($h^2 \geq 5$), the deviation of phenotypic distribution from normality was random nature (data sets 1, 3, 9, 11, 17 and 19). Since $x = 0$ was assumed for these data sets, the deviation from normality was due to random combination of gene frequencies. Also, when the sample size was small ($N \leq 500$), the results of the normality test were not consistent. Assuming these samples were real and the parameters were unknown, a search for the best initial values was undertaken. The initial values for heritability ($\hat{h}^2$), number of loci ($\hat{n}$), the degree of dominance ($\hat{x}$), genotypic value ($\hat{a}$), and gene frequencies ($\hat{p}_i$ for $i = 1$ to n), for each sample are shown in Table 4.3.1. The best initial values chosen were generally in good agreement with parameter values for all samples in a data set when heritability was high and the sample size was large. As heritability of the trait became lower and the sample size became smaller, the results were not consistent. However, present results indicated that it is now possible to estimate an integer value of number of loci as well as genotypic values and heritability of a quantitative trait from a random sample of phenotypic values. It was not possible before since the exact distribution of phenotypic values was unknown and methods available were not applicable to such data.

Sample 4 of data set 1 (data 1-4) was selected from the samples which were not normally distributed to illustrate the method for searching for the best combination of initial values and obtaining the best estimates of parameters by the method of maximum likelihood. The histogram of frequency distribution is shown in Figure 4.3. A given set of parameters, h^2, x, n, and ten trial runs of gene frequencies (p_i for $i = 1$ to n), log likelihood function values (log L) were computed and the run with maximum log L among ten was selected. This process was continued for all combinations of heritability values ($h^2 = .5, .6, .7, .8$, and $.9$), degree of dominance ($x = 0$ and 1) and number of loci ($n = 1, 2, 3$, and 4). The log L values are shown in Table 4.3.2. A total of four hundred log L values obtained by different combinations of h^2, x, n, and p_i (for $i = 1$ to n) were compared and log $L = -7109.000$ was found to be maximum. Therefore, it was concluded that the combination of parameters which gave this log L value to be the best set of initial values ($\hat{a} = .864$, $\hat{d} = 0$, $\hat{\mu}_e = .413$, $\hat{\sigma}_e^2 = .303$, $\hat{n} = 2$, $\hat{p}_1 = .376$, and $\hat{p}_2 = .392$). It should be noted that:

$$\hat{h}^2 = \frac{1.010 - .303}{1.010} = .7 \text{ where } \hat{\sigma}_y^2 = 1.010$$

and $\hat{x} = \frac{0}{.864} = 0$. The best initial values were used to obtain the maximum likelihood estimates (ML estimates) of parameters. The parameter values as well as initial values, ML estimates and its standard errors of the estimates given $n = 2$ were:

	Parameter Value	Initial Value	ML Estimates	Standard Error
a	.909	.864	.851	.001
d	.0	.0	.019	.040
μ_e	.304	.413	.358	.009
σ_e^2	.2	.303	.313	-.000
p_1	.452	.376	.380	-.000
p_2	.380	.392	.404	-.000

Since this was a simulated sample data, the relative closeness of estimates to the parameter values indicated the effectiveness of the estimation procedure proposed. Assuming these were real data, the results could be interpreted as:

(1) The sample fits the distribution with ML estimates as its parameters;

(2) The estimate of heritability was $h^2 = .690$;

(3) The number of loci was estimated as $n = 2$; and

(4) The degree of dominance was $x = .022$.

Consistency of results was examined by repeating the process of finding best initial values. Three samples were selected for this purpose: sample 1 of data set 1 (data 1-1); sample 4 of data set 1 (data 1-4); and sample 1 of data set 4 (data 4-1). The parameter values of heritability for data 1-1 and data 1-4 was .8; however, the phenotypic distribution of data 1-1 was not significantly different and the distribution of data 1-4 was significantly different from normal distribution. The heritability value for data 4-1 was assumed to be .2. Five sets of the best initial values for each of the three samples which were obtained independently are shown in Table 4.3.3. Consistent and good results were obtained for both data 1-1 and data 1-4; however, the results of data 4-1 initial values were not consistent, and deviate significantly from parameter values. Therefore, when the heritability of a trait was high, the probability of obtaining good initial values was much higher than when the heritability of the trait was low. This implied that when initial values of parameters were needed for a trait with unknown heritability, it was important to repeat the process to find out whether or not the results were consistent. If it was consistent, the results could be used to get a better estimate by the maximum likelihood method.

Table 4.3.1

Initial Values

Data Set	Sample	$\hat{n}$	$\hat{a}$	$\hat{h}^2$	$\hat{x}$	$\hat{p}_i$				
1	1	2	.871	.7	0	.51	.67			
	2	3	.783	.8	0	.74	.67	.36		
	3	2	.931	.8	0	.42	.30			
	4	2	.864	.7	0	.37	.39			
	5	2	1.101	.8	0	.25	.15			
2	1	1	1.573	.8	0	.20				
	2	3	.659	.8	1	.29	.70	.57		
	3	1	1.291	.8	0	.40				
	4	3	1.375	.8	1	.82	.72	.90		
	5	3	.738	.8	1	.27	.72	.82		
3	1	3	.748	.7	0	.21	.53	.25		
	2	3	.950	.7	0	.18	.00	.38		
	3	4	.758	.6	0	.13	.20	.03	.30	
	4	4	.813	.8	0	.83	.98	.35	.28	
	5	1	.904	.4	0	.60				
4	1	3	.582	.5	0	.39	.52	.60		
	2	1	1.285	.3	0	.90				
	3	3	.689	.5	0	.84	.81	.33		
	4	1	.554	.1	0	.80				
	5	2	.714	.5	0	.38	.47			
5	1	5	.699	.9	0	.83	.71	.77	.23	.44
	2	3	.804	.8	0	.79	.28	.51		
	3	6	.675	.6	0	.94	.08	.05	.55	.82
						.93				
	4	5	.643	.7	0	.31	.46	.93	.38	.91
	5	4	.606	.6	0	.24	.24	.27	.47	

Table 4.3.1--Continued

Data Set	Sample	$\hat{n}$	$\hat{a}$	$\hat{h}^2$	$\hat{x}$	$\hat{p}_i$				
6	1	10	.493	.9	0	.16	.83	.54	.73	.68
						.78	.78	.63	.51	.94
	2	9	.527	.9	0	.57	.53	.05	.10	.53
						.71	.41	.87	.29	
	3	9	.633	.9	0	.56	.01	.01	.96	.64
						.15	.12	.34	.96	
	4	10	.581	.8	0	.42	.78	.85	.07	.68
						.86	.98	.10	.82	.97
	5	8	.591	.9	0	.43	.66	.13	.53	.02
						.29	.61	.99		
7	1	2	1.040	.8	0	.59	.85			
	2	1	1.587	.5	0	.10				
	3	4	.795	.9	0	.61	.98	.38	.49	
	4	2	1.003	.8	0	.20	.32			
	5	2	2.744	.8	0	.96	.01			
8	1	4	.830	.9	0	.00	.54	.27	.71	
	2	4	.685	.9	0	.60	.22	.89	.38	
	3	2	1.084	.8	0	.40	.12			
	4	1	1.353	.8	0	.50				
	5	2	.983	.8	0	.13	.51			
9	1	3	1.071	.7	0	.08	.14	.11		
	2	2	1.040	.8	0	.71	.20			
	3	2	.875	.7	0	.42	.30			
	4	2	1.239	.8	0	.93	.72			
	5	2	.931	.7	0	.76	.38			
10	1	1	.961	.7	1	.50				
	2	4	.845	.8	1	.88	.10	.65	.00	
	3	4	.790	.8	1	.96	.40	.74	.81	
	4	3	.894	.8	1	.89	.91	.38		
	5	3	.628	.7	1	.20	.46	.98		

Table 4.3.1--Continued

Data Set	Sample	$\hat{n}$	$\hat{a}$	$\hat{h}^2$	$\hat{x}$	$\hat{p}_i$				
11	1	1	.909	.4	0	.40				
	2	1	.969	.4	0	.30				
	3	1	1.471	.4	0	.10				
	4	1	1.455	.4	0	.90				
	5	1	1.503	.4	0	.90				
12	1	1	.769	.3	0	.50				
	2	2	.475	.2	0	.71	.58			
	3	1	.785	.3	0	.60				
	4	2	.710	.5	0	.55	.53			
	5	2	.738	.5	0	.53	.72			
13	1	6	.616	.9	0	.46	.51	.84	.46	.17
						.85				
	2	5	.609	.9	0	.52	.51	.37	.64	.48
	3	4	.682	.8	0	.44	.58	.51	.86	
	4	4	.697	.6	0	.32	.15	.94	.31	
	5	4	.714	.8	0	.10	.37	.28	.37	
14	1	7	.687	.9	0	.02	.78	.47	.46	.86
						.98	.15			
	2	8	.658	.9	0	.36	.52	.07	.45	.02
						.03	.00	.73		
	3	9	.560	.7	0	.78	.13	.76	.23	.02
						.02	.88	.10	.61	
	4	7	.626	.9	0	.53	.76	.31	.98	.09
						.70	.30			
	5	7	.517	.5	0	.11	.72	.10	.15	.97
						.44	.84			
15	1	2	1.284	.8	0	.80	.11			
	2	2	.866	.7	0	.24	.56			
	3	2	1.402	.5	0	.95	.90			
	4	2	.957	.6	0	.19	.84			
	5	1	1.347	.6	0	.80				

Table 4.3.1--Continued

Data Set	Sample	$\hat{n}$	$\hat{a}$	$\hat{h}^2$	$\hat{x}$	$\hat{p}_i$				
16	1	3	.912	.9	0	.28	.28	.13		
	2	1	1.177	.7	0	.40				
	3	1	1.748	.5	0	.90				
	4	1	1.171	.8	0	.40				
	5	1	1.385	.7	0	.30				
17	1	2	1.113	.8	0	.45	.08			
	2	2	.969	.7	0	.58	.85			
	3	2	.913	.8	0	.33	.50			
	4	2	.955	.8	0	.28	.63			
	5	4	.861	.7	0	.49	.90	.00	.86	
18	1	1	1.607	.8	0	.80				
	2	2	.943	.8	1	.45	.86			
	3	3	.882	.8	1	.49	.74	.00		
	4	2	.756	.8	1	.09	.46			
	5	2	.921	.8	1	.54	.72			
19	1	1	.793	.3	0	.60				
	2	3	.804	.8	0	.76	.24	.58		
	3	4	.675	.7	0	.44	.09	.51	.24	
	4	4	.688	.8	0	.45	.23	.26	.52	
	5	3	.816	.7	0	.58	.96	.53		
20	1	2	.385	.1	0	.80	.76			
	2	1	.764	.3	0	.50				
	3	2	.679	.4	0	.62	.26			
	4	1	.645	.2	0	.40				
	5	1	.888	.4	0	.50				
21	1	6	.608	.8	0	.63	.89	.56	.62	.69
						.12				
	2	7	.577	.9	0	.67	.21	.21	.34	.15
						.49	.81			
	3	4	.802	.8	0	.89	.69	.09	.38	
	4	5	.627	.8	0	.13	.62	.52	.48	.18
	5	5	.644	.9	0	.30	.72	.65	.35	.24

Table 4.3.1--Continued

Data Set	Sample	$\hat{n}$	$\hat{a}$	$\hat{h}^2$	$\hat{x}$	$\hat{p}_i$				
22	1	10	.517	.9	0	.61	.49	.46	.03	.99
						.33	.82	.91	.61	.72
	2	8	.613	.9	0	.28	.92	.98	.61	.84
						.55	.66	.03		
	3	8	.523	.9	0	.47	.28	.44	.11	.70
						.16	.65	.47		
	4	7	.685	.9	0	.78	.47	.46	.86	.98
						.15	.02			
	5	7	.690	.9	0	.84	.20	.14	.86	.93
						.78	.31			
23	1	1	1.164	.7	0	.50				
	2	2	.923	.8	0	.62	.45			
	3	1	1.183	.6	0	.30				
	4	3	.788	.9	0	.44	.40	.41		
	5	3	.966	.8	0	.30	.01	.34		
24	1	2	1.158	.9	0	.71	.77			
	2	3	1.427	.7	0	.01	.99	.79		
	3	1	1.523	.7	0	.20				
	4	1	1.858	.7	0	.10				
	5	2	.970	.8	0	.40	.24			

Table 4.3.2

log Likelihood Function Values (log-L)

(Data 1-4)

n	h^2	log L	
		$x = 0$	$x = 1$
1	.5	-7112.695	-7118.738
	.6	-7112.273	-7182.378
	.7	-7138.988	-7400.175
	.8	-7269.179	-8081.636
	.9	-8253.750	-11141.089
2	.5	-7120.253	-7113.558
	.6	-7117.277	-7138.347
	.7	-7109.000*	-7146.460
	.8	-7112.984	-7315.019
	.9	-7366.218	-8953.316
3	.5	-7123.117	-7124.941
	.6	-7121.718	-7151.792
	.7	-7117.847	-7161.394
	.8	-7116.492	-7231.027
	.9	-7145.421	-7450.875
4	.5	-7127.433	-7130.089
	.6	-7127.515	-7131.402
	.7	-7121.148	-7141.417
	.8	-7114.773	-7133.523
	.9	-7122.753	-7259.511

* the largest log L value

Table 4.3.3

Results of Consistency Test

Data	Run Number	$\hat{a}$	$\hat{d}$	$\hat{\mu}_e$	$\hat{\sigma}_e^2$	$\hat{h}^2$	$\hat{x}$	$\hat{n}$	$\hat{p}_i$				ML
1-1	1	.871	0	-.334	.305	.7	0	2	.51	.67			-7130.976
	2	.924	0	-.295	.203	.8	0	2	.50	.67			-7125.644
	3	1.004	0	-.487	.203	.8	0	2	.80	.43			-7131.308
	4	.923	0	-.222	.203	.8	0	2	.64	.47			-7125.933
	5	.920	0	-.109	.203	.8	0	2	.43	.62			-7131.066
	(Parameters:	.955	0	-.410	.2	.8	0	2	.74	.46)			
1-4	1	.882	0	.539	.303	.7	0	2	.32	.38			-7110.367
	2	.920	0	.406	.202	.8	0	2	.41	.37			-7103.796
	3	.905	0	-.616	.202	.8	0	3	.98	.52	.34		-7110.441
	4	.970	0	.434	.202	.8	0	2	.23	.54			-7108.234
	5	.925	0	.316	.202	.8	0	2	.50	.33			-7103.359
	(Parameters:	.909	0	.304	.2	.8	0	2	.45	.38)			
4-1	1	.767	0	-.010	.493	.5	0	2	.29	.69			-7086.898
	2	.889	0	-.028	.394	.6	0	2	.25	.75			-7086.777
	3	.648	0	-.294	.394	.6	0	4	.91	.25	.33	.70	-7086.902
	4	.686	0	-.128	.394	.6	0	3	.38	.80	.38		-7086.921
	5	.909	0	-.512	.295	.7	0	3	.95	.63	.17		-7086.312
	(Parameters:	.586	0	.676	.8	.2	0	2	.06	.35)			

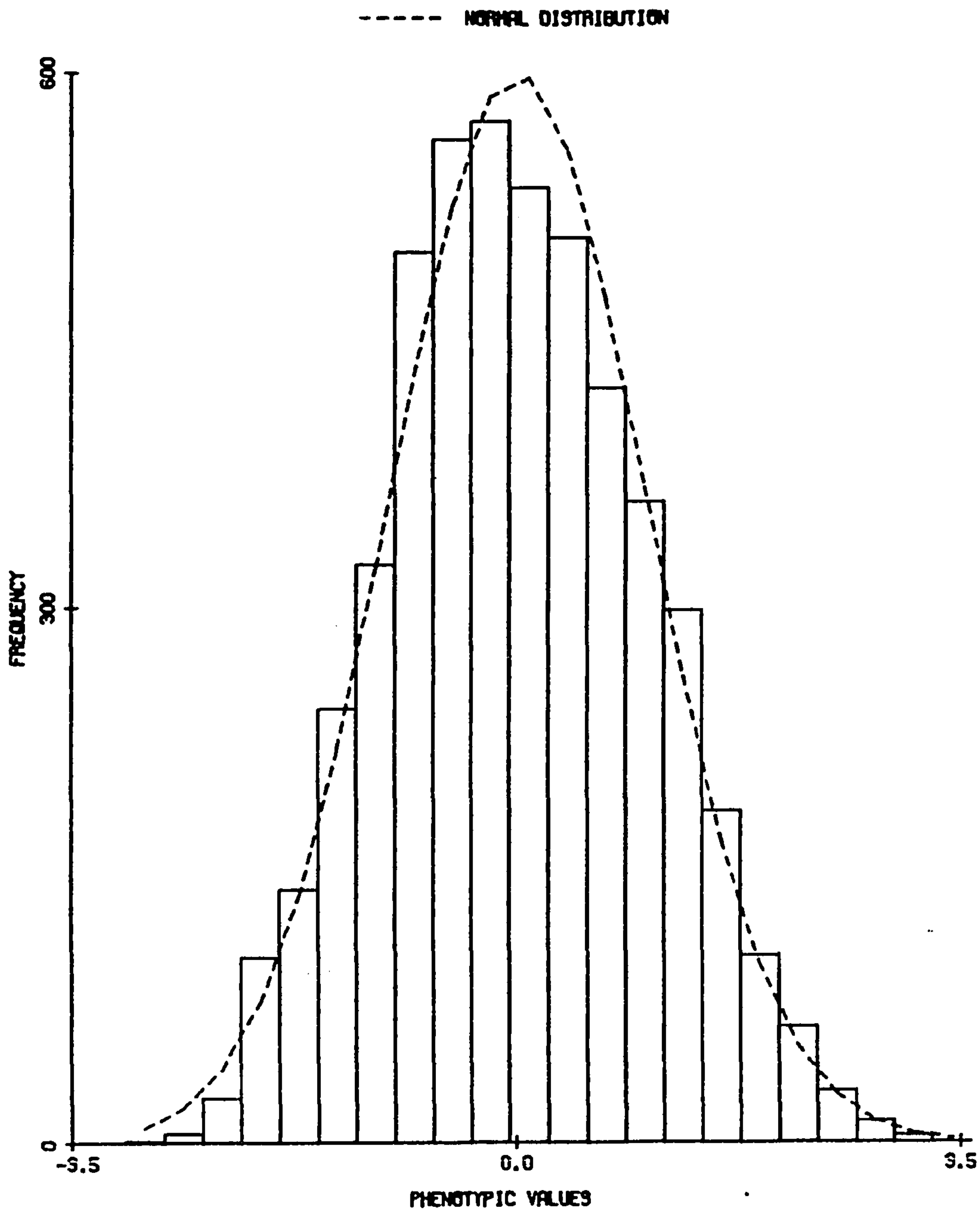


Figure 4.3 Histogram of Data 1-4

Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=2$, $p_1=.37$, $p_2=.39$

4.4 Estimation of Parameters Using Genetic Marker Information

Presented in Table 4.4.1 are: three subclass means ($\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\mu}_3$), variances ($\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, $\hat{\sigma}_3^2$), sample sizes (N_1, N_2, N_3), minimum and maximum values of subclasses, the expected gene frequency of A_1-gene (p_n) at G_n-locus, and observed gene frequency of M-gene ($\hat{p}_m$) at M-locus. The twenty-four data sets were divided into three subclasses: $D^* = 0$ for data sets 1 to 8; $D^* = .5$ for data sets 9 to 16, and $D^* = 1$ for data sets 17 to 24. The nth locus (G_n-locus) was arbitrarily named among n loci to be the locus associated with the marker locus. When $D^* = 0$, there was no association; when $D^* = .5$, there was partial linkage; when $D^* = 1$, there was complete linkage and $p_n = p_m$ were assumed a priori to the generation of data sets. Results of the association test are given in Table 4.4.2. All except one of the data sets 1 to 6 showed no significant difference between three subclasses. These results were expected since the simulation parameter for D^* was zero for these data sets. Although $D^* = 0$, data sets 7 and 8 indicated some significant differences between subclasses, presumably due to the small sample sizes (500 and 100) of the data sets. Expected differences between subclasses were noted in most of data sets 9 to 14 and all of data sets 17 to 23. Association of quantitative trait and marker genotypes was not detected when the sample size was small; for example, when $N = 500$ and 100, as in data sets 15 and 16 ($D^* = .5$), or when $N = 100$, as in data set 24 ($D^* = 1$). One sample from each data set was chosen and its frequency and cumulative frequency distributions were plotted (Figure 4.4.1 to Figure 4.4.24). Distributions of three subclasses were almost identical when $D^* = 0$, as was expected. As the value of D^* departed from zero, the differences between distributions increased due to the degree of linkage disequilibrium. When two or three pairs of subclass distributions were significantly different, the method of estimation for the degree of disequilibrium was employed. A total of sixty samples were selected. When the phenotypic distribution was not normal, the exact distributions were used. The degree of linkage disequilibrium ($\hat{D}$) was calculated by using means of subclasses ($\hat{\mu}_{MM}$, $\hat{\mu}_{Mm}$, and $\hat{\mu}_{mm}$), genotypic values ($\hat{a}$ and $\hat{d}$), mean of environmental effects ($\hat{\mu}_e$) and the number of loci ($\hat{n}$). These estimates were obtained by the method discussed in Section 3.6. Then, the gene frequency of A_1-gene at G_n-locus was selected from $\hat{p}_i$ for $i = 1$ to n and the remaining gene frequencies were considered as estimates of $\hat{p}_i$ for $i = 1$ to $(n - 1)$. When the total phenotypic distribution was normal, approximate distributions were used. $\hat{D}$ was computed by assuming $n = 1$. The mean ($\hat{\mu}$) and the variance ($\hat{\sigma}^2$) of normal distribution were estimated as:

$$\hat{\mu} = \hat{\mu}_y - \hat{\mu}_{gn}$$
$$\hat{\sigma}^2 = \hat{\sigma}^2_y - \hat{\sigma}^2_{gn}$$

where $\hat{\mu}_y$ = mean of phenotypic values, $\hat{\sigma}^2_y$ = variance of phenotypic values, $\hat{\mu}_{gn}$ = mean of genotypic values at G_n-locus, and $\hat{\sigma}^2_{gn}$ = variance of genotypic values at G_n-locus. $\hat{\mu}_{gn}$ and $\hat{\sigma}^2_{gn}$ were computed as follows:

$$\hat{\mu}_{gn} = (\hat{p}_n - \hat{q}_n)\hat{a} + 2\hat{p}_n\hat{q}_n\hat{d}$$
$$\hat{\sigma}^2_{gn} = 2\hat{p}_n\hat{q}_n\ (\hat{a} + (\hat{q}_n - \hat{p}_n)\hat{d})^2 + (2\hat{p}_n\hat{q}_n\hat{d})^2$$

$\hat{a}$ and $\hat{d}$ were the estimates of genotypic values obtained by the method described in Section 3.6; $p_n (q_n = 1 - p_n)$ was selected from $\hat{p}_i$ for

$i = 1$ to n. Therefore, $\hat{\sigma}^2$ included the variance of environmental effects and the variance due to the other $(n - 1)$ loci. Both the exact distributions and approximate distributions were given in Section 3.6.

A log likelihood function value of each sample was computed for all possible $\hat{p}_n$ values. Then a set of estimates with maximum log likelihood function value was selected for each sample. A summary of estimates for D, D^* and p_n is shown in Table 4.4.3. Some estimates for the samples among data sets 9 to 14 did not agree with parameter values, particularly samples 2, 3, and 4 of data set 10. This presumably was due to values of estimates used to calculate $\hat{D}$, $\hat{D}^*$ and $\hat{p}_n$, since initial values of these estimates were different from parameter values (Table 4.3.1). For example, the parameter value of the number of loci for data set 10 was 2, whereas estimates were 4, 4, and 3 for data 10-2, 10-3 and 10-4, respectively. However, the estimates were in good agreement with parameter values for data sets 17 to 23. This indicated that when there was a complete linkage between a quantitative trait locus and a genetic marker, the probability of detecting this linkage and obtaining good estimates was high.

Sample 2 of data set 9 (Data 9-2) was chosen to demonstrate an example of a complete procedure for testing the association and estimation of the degree of linkage disequilibrium between one of the loci in a quantitative trait and a genetic marker locus. Frequency and cumulative frequency distributions of the subclasses were plotted (Figure 4.4.25). Three subclass sample distributions were found to be significantly different by the Kolomogrov-Smirnov two-sample test.

This difference was also obvious in the plots of the three cumulative frequency distributions. Since the sample phenotypic distribution of Data 9-2 (sample 2 of data set 9) was significantly different from the normal distribution (Table 4.4.2), the exact distributions were used for the next procedure. The gene frequency of M-gene at M-locus was estimated as:

$$p_m = \frac{2(1235) + 2543}{2(5000)} = .501$$

where 1235, 2543 and 5000 were the sample sizes of MM-class, Mm-class and total, respectively. Initial values for $\hat{a}$, $\hat{d}$, $\hat{\mu}_e$, $\hat{\sigma}_e^2$, and $\hat{n}$ were obtained by the method derived in Section 3.6. These values were: $\hat{a}$ = 1.040, $\hat{d}$ = 0, $\hat{\mu}_e$ = .163, $\hat{\sigma}_e^2$ = .200 and $\hat{n}$ = 2. The initial value for the gene frequency of A_1-gene at G_n-locus ($\hat{p}_n$) was selected from .20 and .71, where .20 and .71 were the best initial values obtained from the phenotypic distribution. The means of subclasses for MM, Mm, and mm-class were .107, -.001 and -.134, respectively. The values of $\hat{D}$, $\hat{D}^*$ and log likelihood function value (log L) were computed for $\hat{p}_n$ = .20 and $\hat{p}_n$ = .71. The maximum log L was observed when $\hat{p}_n$ = .2 and $\hat{p}_1$ = .6. Therefore, $\hat{p}_n$ = .2 was assumed as the estimate of gene frequency at G_n-locus. $\hat{D}^*$ = .292 and $\hat{D}$ = .029 were obtained. Since the expected value of D^* was .5 and D was .027, the relative closeness of the parameter values and the estimate indicated that the procedure proposed was effective in estimating the degree of linkage disequilibrium and the gene frequency at G_n-locus. If this were real data, the results could be interpreted as:

(1) Data 9-2 fit the distribution with the estimates shown above;

(2) the linkage disequilibrium value between one of the quantitative trait loci and the genetic marker locus was estimated as .029;

(3) the gene frequency at G_n-locus was estimated as .20. Since $\hat{D}$ = .029, $\hat{p}_n$ = .20, and $\hat{p}_m$ = .501, $\hat{D}^*$ = .292.

The efficiency of the method was lower when the degree of linkage disequilibrium (D^*) was .5 than when $D^* = 1$. However, the overall results of simulation studies indicated that some useful information could be obtained when genetic marker data as well as phenotypic values of individuals from a random mating population were available, such as:

(1) detecting association between a quantitative trait and a genetic marker; and

(2) estimating the degree of linkage disequilibrium and gene frequency of a quantitative trait locus associated with the genetic marker when the association was detected.

There was no need to assume that a quantitative trait was controlled only by a single locus nor a complete linkage between a quantitative trait locus and a genetic marker locus. All the parameters of exact distributions or approximate distributions could be estimated by the method of maximum likelihood.

Table 4.4.1

Statistics of Subclasses

Data	Subclass	$\hat{\mu}_y$	$\hat{\sigma}_y^2$	$\hat{N}$	Y_{min}	Y_{max}	p_n	$\hat{p}_m$
1-1	MM	.037	.989	1164	-2.932	2.612	.746	.490
	Mm	-.010	.995	2579	-3.022	2.626		
	mm	-.005	1.087	1257	-3.356	2.693		
1-2	MM	.001	1.008	1286	-3.154	2.911	.636	.499
	Mm	-.002	.989	2421	-3.175	2.903		
	mm	-.045	.955	1293	-3.070	2.413		
1-3	MM	-.013	.971	1337	-2.588	3.048	.140	.512
	Mm	.017	.996	2452	-2.788	3.393		
	mm	-.032	.998	1211	-2.490	3.751		
1-4	MM	-.024	.968	1214	-2.508	2.781	.452	.498
	Mm	.049	1.017	2559	-2.727	3.283		
	mm	-.022	1.031	1227	-2.774	3.232		
1-5	MM	.015	.959	1165	-2.232	4.035	.264	.485
	Mm	-.014	.975	2526	-2.229	4.448		
	mm	.031	1.004	1309	-1.990	3.842		
2-1	MM	.043	1.010	1198	-2.119	3.674	.177	.492
	Mm	.008	.987	2524	-2.096	3.820		
	mm	.037	.973	1278	-1.990	3.229		
2-2	MM	-.011	1.035	1234	-2.776	2.280	.247	.499
	Mm	.007	1.035	2522	-3.004	2.354		
	mm	-.053	1.052	1244	-3.270	2.288		
2-3	MM	-.030	1.010	1235	-2.241	3.144	.278	.495
	Mm	.003	1.006	2481	-2.387	2.763		
	mm	-.026	.977	1284	-2.233	2.830		
2-4	MM	-.011	1.051	1254	-6.225	1.566	.852	.501
	Mm	-.003	1.012	2502	-5.469	1.895		
	mm	.012	.938	1244	-5.152	1.873		
2-5	MM	.008	.931	1266	-3.508	2.275	.610	.500
	Mm	.018	.965	2470	-3.447	2.164		
	mm	.010	.955	1264	-3.440	2.174		

Table 4.4.1--Continued

Data	Subclass	$\hat{\mu}_y$	$\hat{\sigma}_y^2$	$\hat{N}$	Y_{min}	Y_{max}	p_n	$\hat{p}_m$
3-1	MM	-.034	1.005	1245	-3.337	3.026	.065	.496
	Mm	-.021	.972	2470	-2.840	3.252		
	mm	-.039	.937	1285	-2.613	3.606		
3-2	MM	.015	.965	1251	-2.519	3.442	.117	.499
	Mm	-.017	1.045	2490	-3.164	4.069		
	mm	.008	1.028	1259	-2.801	3.703		
3-3	MM	-.024	1.041	1249	-2.577	3.549	.039	.497
	Mm	.032	1.013	2477	-2.627	3.695		
	mm	.039	.959	1274	-2.431	3.835		
3-4	MM	-.031	.939	1236	-2.918	3.366	.433	.497
	Mm	.012	1.010	2498	-3.379	3.426		
	mm	-.010	.936	1266	-3.062	2.954		
3-5	MM	-.022	1.004	1205	-3.234	2.605	.524	.493
	Mm	.010	.958	2525	-3.003	3.192		
	mm	.028	1.004	1270	-3.865	2.683		
4-1	MM	-.035	.995	1245	-3.725	3.151	.065	.496
	Mm	-.022	1.003	2470	-3.298	3.315		
	mm	-.015	.943	1285	-2.809	3.299		
4-2	MM	.017	.960	1248	-4.799	2.914	.994	.498
	Mm	-.006	1.000	2492	-4.807	3.163		
	mm	.026	1.000	1260	-4.339	3.325		
4-3	MM	-.020	.935	1238	-3.309	2.859	.054	.495
	Mm	.027	.977	2482	-4.037	2.943		
	mm	.039	.942	1280	-2.738	3.004		
4-4	MM	-.022	.978	1242	-3.492	3.455	.362	.497
	Mm	-.010	.996	2493	-3.959	3.824		
	mm	-.023	.971	1265	-3.115	3.416		
4-5	MM	.008	.939	1199	-2.745	3.033	.436	.492
	Mm	.013	.999	2525	-3.011	3.253		
	mm	.029	1.034	1276	-3.212	3.367		

Table 4.4.1--Continued

Data	Subclass	$\hat{\mu}_y$	$\hat{\sigma}_y^2$	$\hat{N}$	Y_{min}	Y_{max}	p_n	$\hat{p}_m$
5-1	MM	.009	1.041	1220	-3.869	2.756	.157	.494
	Mm	.028	1.035	2508	-3.073	3.102		
	mm	-.067	.996	1272	-3.287	2.797		
5-2	MM	-.012	.966	1244	-3.144	3.242	.819	.502
	Mm	-.018	1.003	2541	-3.541	3.055		
	mm	-.072	.995	1215	-3.709	2.924		
5-3	MM	-.033	1.000	1256	-3.658	2.916	.553	.507
	Mm	.003	.973	2562	-3.223	3.726		
	mm	.013	.933	1182	-3.604	2.707		
5-4	MM	.002	1.002	1274	-3.438	3.293	.784	.501
	Mm	-.034	.992	2470	-3.707	3.078		
	mm	.005	1.010	1256	-3.809	3.054		
5-5	MM	-.041	1.057	1242	-3.377	3.617	.383	.498
	Mm	-.008	.991	2500	-2.775	3.741		
	mm	-.009	.975	1258	-2.819	3.907		
6-1	MM	.042	.963	1269	-3.715	2.747	.664	.500
	Mm	-.006	.973	2468	-3.160	3.181		
	mm	-.044	.998	1263	-3.339	2.999		
6-2	MM	-.001	.996	1237	-2.658	3.503	.591	.498
	Mm	.049	1.056	2514	-3.431	3.563		
	mm	-.013	.994	1249	-3.694	3.136		
6-3	MM	.008	.991	1264	-2.728	3.519	.125	.503
	Mm	-.022	1.080	2510	-3.564	3.472		
	mm	.021	1.012	1226	-3.430	2.864		
6-4	MM	.003	1.040	1312	-3.314	2.901	.665	.510
	Mm	-.003	1.017	2485	-4.023	3.293		
	mm	-.021	1.034	1203	-3.679	2.631		
6-5	MM	.005	.994	1243	-2.980	3.524	.783	.495
	Mm	-.017	1.008	2465	-3.116	2.966		
	mm	-.051	1.039	1292	-2.932	3.029		

Table 4.4.1--Continued

Data	Subclass	$\hat{\mu}_y$	$\hat{\sigma}_y^2$	$\hat{N}$	Y_{min}	Y_{max}	p_n	$\hat{p}_m$
7-1	MM	.520	.744	116	-1.666	2.040	.563	.487
	Mm	-.095	.820	255	-2.719	2.148		
	mm	-.437	1.078	129	-2.876	2.154		
7-2	MM	.023	.501	120	-1.885	2.042	.942	.511
	Mm	.068	1.044	271	-2.809	3.314		
	mm	-.000	1.009	109	-2.111	2.165		
7-3	MM	.389	.894	117	-2.633	2.614	.394	.494
	Mm	.043	1.025	260	-2.321	2.749		
	mm	-.366	.955	123	-2.951	2.149		
7-4	MM	.115	1.142	131	-1.837	2.688	.175	.507
	Mm	-.010	.968	245	-1.747	2.882		
	mm	.011	.745	124	-1.859	2.244		
7-5	MM	.180	.616	130	-2.315	2.841	.940	.494
	Mm	-.028	.974	234	-3.112	3.318		
	mm	-.186	1.327	136	-3.146	3.337		
8-1	MM	.042	1.308	22	-2.164	2.002	.067	.495
	Mm	-.001	.944	55	-2.372	2.495		
	mm	-.006	.816	23	-1.884	1.310		
8-2	MM	.208	.758	29	-1.778	1.951	.526	.495
	Mm	-.197	.848	41	-2.159	2.287		
	mm	-.522	.419	30	-1.823	.935		
8-3	MM	.498	.804	29	-1.870	2.618	.423	.520
	Mm	.052	.999	46	-2.002	2.414		
	mm	-.376	.945	25	-1.985	1.818		
8-4	MM	.750	.713	30	-1.083	1.911	.499	.550
	Mm	.050	.872	50	-1.964	2.204		
	mm	-.765	1.081	20	-2.346	2.064		
8-5	MM	.092	.947	19	-1.930	1.614	.249	.450
	Mm	-.047	.985	52	-1.634	2.280		
	mm	-.518	.500	29	-2.017	.887		

Table 4.4.1--Continued

Data	Subclass	$\hat{\mu}_y$	$\hat{\sigma}^2_y$	$\hat{N}$	Y_{min}	Y_{max}	p_n	$\hat{p}_m$
9-1	MM	.345	1.083	1253	-2.396	4.313	.296	.498
	Mm	-.038	.927	2479	-2.571	3.672		
	mm	-.360	.749	1268	-2.065	3.353		
9-2	MM	.107	1.096	1235	-2.880	3.347	.110	.501
	Mm	-.001	1.013	2543	-3.091	3.521		
	mm	-.134	.847	1222	-3.029	2.505		
9-3	MM	.466	.942	1312	-2.454	3.276	.464	.507
	Mm	.023	.874	2452	-2.490	3.585		
	mm	-.403	.925	1236	-2.466	3.286		
9-4	MM	.106	.886	1181	-2.925	2.032	.943	.493
	Mm	.002	1.014	2575	-3.652	2.312		
	mm	-.061	1.070	1244	-4.572	2.064		
9-5	MM	.128	.887	1253	-3.032	3.027	.837	.503
	Mm	-.023	1.020	2526	-3.426	2.970		
	mm	-.171	1.147	1221	-3.269	2.766		
10-1	MM	.037	1.090	1297	-2.828	3.305	.018	.507
	Mm	-.022	.973	2480	-2.946	3.153		
	mm	.015	.918	1223	-2.651	3.355		
10-2	MM	.130	.721	1271	-2.867	2.538	.662	.500
	Mm	.053	.938	2464	-3.314	2.751		
	mm	-.276	1.266	1265	-3.327	2.578		
10-3	MM	.081	.900	1242	-3.416	2.085	.770	.497
	Mm	.023	.972	2488	-3.656	2.214		
	mm	-.107	1.172	1270	-3.741	1.901		
10-4	MM	.284	.749	1210	-3.586	2.075	.385	.493
	Mm	.058	.997	2517	-4.128	2.357		
	mm	-.462	.977	1273	-4.420	1.974		
10-5	MM	.251	.787	1254	-2.690	2.538	.511	.502
	Mm	.076	.867	2514	-2.770	2.772		
	mm	-.398	1.171	1232	-2.816	2.677		

Table 4.4.1--Continued

Data	Subclass	$\hat{\mu}_y$	$\hat{\sigma}^2_y$	$\hat{N}$	Y_{min}	Y_{max}	p_n	$\hat{p}_m$
11-1	MM	.057	1.073	1220	-3.371	3.568	.087	.492
	Mm	-.019	.992	2486	-3.180	3.525		
	mm	-.071	.909	1294	-3.003	2.973		
11-2	MM	.205	1.003	1226	-2.827	3.170	.294	.492
	Mm	.001	.952	2477	-3.230	3.516		
	mm	-.228	.943	1297	-2.851	3.029		
11-3	MM	.000	.943	1249	-3.966	3.427	.987	.501
	Mm	.022	.997	2521	-3.332	3.955		
	mm	.006	.956	1230	-3.739	3.316		
11-4	MM	.149	.743	1248	-2.761	2.991	.886	.500
	Mm	.017	.947	2511	-3.912	3.052		
	mm	-.203	1.111	1241	-4.223	2.928		
11-5	MM	.169	.904	1216	-3.332	2.776	.874	.498
	Mm	.001	1.003	2553	-4.353	3.172		
	mm	-.105	1.116	1231	-4.026	2.721		
12-1	MM	.039	1.043	1220	-3.559	3.407	.087	.492
	Mm	-.021	.972	2486	-3.317	3.237		
	mm	-.041	.961	1294	-3.093	3.314		
12-2	MM	.250	.945	1215	-2.866	3.130	.504	.494
	Mm	.009	.998	2510	-3.663	3.649		
	mm	-.263	.970	1275	-3.345	3.372		
12-3	MM	.202	.936	1205	-2.916	3.266	.473	.489
	Mm	-.008	.974	2484	-2.986	3.183		
	mm	-.212	.977	1311	-3.353	3.197		
12-4	MM	.013	.966	1277	-3.161	2.808	.950	.506
	Mm	.001	.967	2513	-3.306	3.068		
	mm	-.082	1.099	1210	-3.667	3.018		
12-5	MM	.168	.943	1230	-2.753	3.185	.701	.496
	Mm	-.007	.985	2505	-3.315	2.760		
	mm	-.121	.966	1265	-3.503	3.259		

Table 4.4.1--Continued

Data	Subclass	$\hat{\mu}_y$	$\hat{\sigma}^2_y$	$\hat{N}$	Y_{min}	Y_{max}	p_n	$\hat{p}_m$
13-1	MM	.115	1.005	1257	-3.020	3.206	.138	.501
	Mm	-.004	.964	2500	-3.221	3.242		
	mm	-.074	.915	1243	-3.429	2.940		
13-2	MM	.205	1.004	1273	-2.825	3.283	.363	.500
	Mm	.006	1.021	2461	-3.357	3.089		
	mm	-.157	.886	1266	-2.849	3.450		
13-3	MM	.254	1.001	1268	-2.778	3.157	.450	.501
	Mm	.050	.946	2474	-3.425	3.267		
	mm	-.277	.962	1258	-3.161	2.752		
13-4	MM	-.020	1.003	1223	-2.955	3.961	.963	.499
	Mm	-.003	1.017	2545	-3.273	4.134		
	mm	-.067	.956	1232	-3.057	3.112		
13-5	MM	.089	1.028	1265	-2.635	3.250	.197	.497
	Mm	-.019	.963	2444	-2.637	3.423		
	mm	-.130	.937	1291	-2.815	3.276		
14-1	MM	.109	.995	1297	-2.868	2.961	.802	.508
	Mm	.011	.974	2490	-3.271	3.310		
	mm	-.103	1.026	1213	-3.541	2.970		
14-2	MM	.169	1.067	1235	-3.377	3.755	.326	.495
	Mm	.021	.998	2484	-3.154	3.191		
	mm	-.157	.947	1281	-3.085	3.340		
14-3	MM	.124	.949	1221	-3.257	3.031	.813	.493
	Mm	-.027	.982	2497	-3.218	3.787		
	mm	-.106	1.096	1282	-3.195	3.485		
14-4	MM	.018	.945	1237	-3.144	2.755	.897	.497
	Mm	-.006	1.020	2504	-3.191	4.061		
	mm	-.088	1.071	1259	-3.804	3.514		
14-5	MM	.160	.983	1179	-3.021	3.655	.306	.485
	Mm	.011	1.013	2497	-3.314	4.090		
	mm	-.202	.899	1324	-3.037	2.677		

Table 4.4.1--Continued

Data	Subclass	$\hat{\mu}_y$	$\hat{\sigma}^2_y$	$\hat{N}$	Y_{min}	Y_{max}	p_n	$\hat{p}_m$
15-1	MM	.016	.936	101	-1.905	3.165	.856	.464
	Mm	-.018	.999	262	-2.370	3.108		
	mm	-.074	1.326	137	-2.974	3.232		
15-2	MM	.056	.989	123	-1.903	2.675	.518	.510
	Mm	.116	.942	264	-2.153	2.931		
	mm	.012	.816	113	-2.382	2.512		
15-3	MM	-.050	1.237	126	-3.908	1.958	.821	.507
	Mm	.032	.940	255	-4.966	1.775		
	mm	-.147	1.020	119	-2.389	1.669		
15-4	MM	-.002	.818	135	-2.373	2.683	.797	.516
	Mm	-.065	.938	246	-3.285	2.318		
	mm	-.280	.801	119	-2.461	2.332		
15-5	MM	.066	.995	125	-2.838	2.092	.663	.482
	Mm	.059	.858	232	-2.531	2.250		
	mm	.125	1.117	143	-3.057	2.475		
16-1	MM	.218	1.547	19	-1.337	3.742	.054	.470
	Mm	.182	.898	56	-1.572	2.453		
	mm	-.284	.520	25	-1.315	1.813		
16-2	MM	-.050	.728	19	-1.600	1.619	.361	.480
	Mm	.006	1.076	58	-2.029	2.279		
	mm	-.160	.794	23	-1.587	1.381		
16-3	MM	.369	.518	18	-1.190	2.147	.695	.430
	Mm	-.265	1.251	50	-3.145	2.005		
	mm	.372	.873	32	-1.783	2.071		
16-4	MM	-.059	.682	25	-1.514	1.036	.300	.485
	Mm	.025	.953	47	-1.599	2.294		
	mm	.063	.724	28	-1.338	1.733		
16-5	MM	.073	1.134	22	-1.810	1.391	.385	.475
	Mm	.168	1.494	51	-2.113	2.517		
	mm	-.084	.476	27	-1.231	1.186		

Table 4.4.1--Continued

Data	Subclass	$\hat{\mu}_y$	$\hat{\sigma}_y^2$	$\hat{N}$	Y_{min}	Y_{max}	p_n	$\hat{p}_m$
17-1	MM	1.115	.381	1212	-.641	3.791	.500	.494
	Mm	.009	.403	2518	-1.503	2.524		
	mm	-1.125	.362	1270	-2.520	1.448		
17-2	MM	1.031	.467	1278	-1.931	2.711	.500	.503
	Mm	.024	.478	2479	-2.598	1.707		
	mm	-.983	.436	1243	-3.781	.889		
17-3	MM	.953	.539	1319	-.792	3.540	.500	.511
	Mm	-.004	.504	2479	-1.908	2.330		
	mm	-.975	.541	1202	-2.850	1.561		
17-4	MM	.878	.616	1265	-1.195	3.104	.500	.500
	Mm	.027	.593	2479	-2.175	2.381		
	mm	-.894	.595	1256	-3.409	1.112		
17-5	MM	.992	.513	1211	-1.889	2.791	.500	.494
	Mm	.023	.497	2523	-2.744	1.867		
	mm	-.957	.491	1266	-3.576	1.008		
18-1	MM	.449	.452	1263	-2.462	1.966	.500	.502
	Mm	.423	.453	2497	-2.616	2.146		
	mm	-1.318	.491	1240	-3.980	.306		
18-2	MM	.455	.360	1230	-2.263	1.896	.500	.495
	Mm	.456	.384	2490	-2.425	1.961		
	mm	-1.353	.371	1280	-4.071	.214		
18-3	MM	.447	.397	1269	-2.129	1.928	.500	.503
	Mm	.428	.412	2499	-2.604	2.209		
	mm	-1.330	.411	1232	-3.778	.174		
18-4	MM	.384	.551	1255	-1.018	2.749	.500	.497
	Mm	.382	.585	2468	-1.354	2.733		
	mm	-1.132	.547	1277	-2.710	1.227		
18-5	MM	.460	.315	1233	-2.267	1.953	.500	.497
	Mm	.463	.360	2508	-2.235	2.136		
	mm	-1.372	.356	1259	-4.022	.037		

Table 4.4.1--Continued

Data	Subclass	$\hat{\mu}_y$	$\hat{\sigma}_y^2$	$\hat{N}$	Y_{min}	Y_{max}	p_n	$\hat{p}_m$
19-1	MM	.685	.731	1259	-2.134	3.289	.500	.493
	Mm	-.035	.752	2418	-2.791	2.785		
	mm	-.710	.784	1323	-3.434	2.227		
19-2	MM	.755	.712	1224	-2.705	3.322	.500	.497
	Mm	.005	.705	2524	-2.911	2.506		
	mm	-.770	.653	1252	-3.253	1.906		
19-3	MM	.785	.723	1317	-1.947	4.171	.500	.509
	Mm	-.016	.695	2456	-2.469	3.110		
	mm	-.725	.716	1227	-3.190	2.005		
19-4	MM	.831	.661	1285	-1.660	3.951	.500	.501
	Mm	-.019	.694	2443	-3.049	3.551		
	mm	-.842	.656	1272	-3.116	1.762		
19-5	MM	.931	.563	1233	-1.685	3.243	.500	.498
	Mm	.017	.572	2521	-2.313	2.400		
	mm	-.929	.579	1246	-3.971	1.415		
20-1	MM	.415	.871	1259	-2.654	3.309	.500	.493
	Mm	-.030	.910	2418	-3.051	3.120		
	mm	-.439	.927	1323	-3.588	2.862		
20-2	MM	.418	.879	1233	-2.776	3.409	.500	.497
	Mm	-.003	.892	2513	-3.209	3.524		
	mm	-.407	.896	1254	-3.649	2.585		
20-3	MM	.491	.862	1291	-2.657	3.521	.500	.508
	Mm	.010	.873	2503	-3.007	2.850		
	mm	-.474	.887	1206	-3.108	2.657		
20-4	MM	.430	.914	1277	-2.263	3.365	.500	.503
	Mm	.001	.886	2476	-3.311	3.057		
	mm	-.464	.903	1247	-3.497	2.450		
20-5	MM	.480	.902	1220	-2.540	3.443	.500	.500
	Mm	-.001	.877	2565	-2.846	2.836		
	mm	-.422	.896	1215	-3.556	2.612		

Table 4.4.1--Continued

Data	Subclass	$\hat{\mu}_y$	$\hat{\sigma}_y^2$	$\hat{N}$	Y_{min}	Y_{max}	p_n	$\hat{p}_m$
21-1	MM	.662	.788	1281	-2.643	3.410	.500	.503
	Mm	-.007	.840	2475	-3.311	3.043		
	mm	-.628	.840	1244	-3.463	1.824		
21-2	MM	.635	.740	1243	-2.030	3.685	.500	.498
	Mm	.011	.773	2497	-2.669	3.037		
	mm	-.619	.818	1260	-3.367	2.562		
21-3	MM	.687	.811	1248	-2.130	3.620	.500	.499
	Mm	.001	.760	2496	-3.056	3.546		
	mm	-.678	.812	1256	-3.302	3.186		
21-4	MM	.654	.755	1287	-1.755	3.630	.500	.507
	Mm	.006	.775	2496	-3.040	3.733		
	mm	-.659	.776	1217	-2.923	2.397		
21-5	MM	.578	.867	1235	-2.705	3.393	.500	.495
	Mm	.014	.756	2482	-2.896	3.161		
	mm	-.608	.775	1283	-3.123	2.057		
22-1	MM	.454	.877	1317	-2.584	3.545	.500	.507
	Mm	-.000	.883	2439	-2.871	2.733		
	mm	-.475	.868	1244	-3.865	2.324		
22-2	MM	.451	.890	1238	-2.618	3.926	.500	.496
	Mm	-.006	.849	2484	-3.272	2.704		
	mm	-.465	.874	1278	-3.180	2.331		
22-3	MM	.503	.895	1237	-2.476	3.750	.500	.502
	Mm	-.017	.869	2549	-3.196	3.137		
	mm	-.466	.860	1214	-3.564	2.291		
22-4	MM	.467	.848	1223	-2.179	3.329	.500	.497
	Mm	.005	.893	2531	-3.594	2.917		
	mm	-.525	.847	1246	-3.531	2.148		
22-5	MM	.478	.968	1248	-2.790	3.725	.500	.499
	Mm	.005	.915	2497	-3.255	3.169		
	mm	-.473	.898	1255	-3.965	2.737		

Table 4.4.1--Continued

Data	Subclass	$\hat{\mu}_y$	$\hat{\sigma}_y^2$	$\hat{N}$	Y_{min}	Y_{max}	p_n	$\hat{p}_m$
23-1	MM	1.178	.420	127	-1.035	2.482	.500	.508
	Mm	-.028	.327	254	-1.761	1.264		
	mm	-1.084	.264	119	-2.376	-.051		
23-2	MM	1.074	.424	123	-.663	2.712	.500	.475
	Mm	.026	.439	229	-2.729	1.404		
	mm	-1.102	.278	.48	-2.657	.015		
23-3	MM	1.019	.504	125	-.522	3.186	.500	.502
	Mm	.002	.506	252	-1.395	2.340		
	mm	-.983	.415	123	-2.519	1.303		
23-4	MM	1.123	.384	108	-.229	2.866	.500	.485
	Mm	.057	.420	269	-1.383	2.614		
	mm	-1.148	.409	123	-2.379	.507		
23-5	MM	1.057	.444	139	-.237	3.575	.500	.526
	Mm	-.057	.421	248	-1.422	2.429		
	mm	-1.214	.384	113	-2.690	1.115		
24-1	MM	.089	1.571	22	-2.530	1.509	.500	.510
	Mm	-.083	.801	58	-2.718	1.949		
	mm	-.052	1.542	20	-3.332	1.725		
24-2	MM	.127	.917	24	-1.834	1.614	.500	.530
	Mm	.201	1.126	58	-2.703	2.464		
	mm	.244	1.218	18	-1.770	2.219		
24-3	MM	-.199	.585	20	-1.250	1.492	.500	.445
	Mm	.130	1.392	49	-1.550	3.111		
	mm	-.012	.793	31	-1.462	1.936		
24-4	MM	-.035	.853	23	-1.032	2.251	.500	.475
	Mm	-.098	1.087	49	-1.279	3.158		
	mm	-.519	.425	28	-1.527	1.287		
24-5	MM	-.000	.892	31	-1.744	2.762	.500	.560
	Mm	.030	.932	50	-1.986	2.937		
	mm	-.232	1.349	19	-2.073	1.997		

Table 4.4.2

Results of Association Test

Data	D_{12}	D_{13}	D_{23}	Data	D_{12}	D_{13}	D_{23}
1-1				8-1			
1-2			*	8-2		*	
1-3				8-3		**	
1-4				8-4	**	**	*
1-5				8-5		*	
2-1				9-1	**	**	**
2-2				9-2	*	**	**
2-3				9-3	**	**	**
2-4				9-4	*	**	
2-5				9-5	**	**	**
3-1				10-1			
3-2				10-2		**	**
3-3				10-3		**	*
3-4				10-4	**	**	**
3-5				10-5	**	**	**
4-1				11-1		*	
4-2				11-2	**	**	**
4-3				11-3			
4-4				11-4	**	**	**
4-5				11-5	**	**	
5-1				12-1			
5-2				12-2	**	**	**
5-3				12-3	**	**	**
5-4				12-4			
5-5				12-5	**	**	*
6-1				13-1	*	**	
6-2				13-2	**	**	**
6-3				13-3	**	**	**
6-4				13-4			
6-5				13-5	*	**	**
7-1	**	**	**	14-1	*	**	**
7-2				14-2	*	**	**
7-3	**	**	*	14-3	**	**	*
7-4				14-4		*	
7-5				14-5	**	**	**

Table 4.4.2--Continued

Data	D_{12}	D_{13}	D_{23}	Data	D_{12}	D_{13}	D_{23}
15-1				20-1	**	**	**
15-2				20-2	**	**	**
15-3				20-3	**	**	**
15-4				20-4	**	**	**
15-5				20-5	**	**	**
16-1				21-1	**	**	**
16-2				21-2	**	**	**
16-3			*	21-3	**	**	**
16-4				21-4	**	**	**
16-5				21-5	**	**	**
17-1	**	**	**	22-1	**	**	**
17-2	**	**	**	22-2	**	**	**
17-3	**	**	**	22-3	**	**	**
17-4	**	**	**	22-4	**	**	**
17-5	**	**	**	22-5	**	**	**
18-1		**	**	23-1	**	**	**
18-2		**	**	23-2	**	**	**
18-3		**	**	23-3	**	**	**
18-4		**	**	23-4	**	**	**
18-5		**	**	23-5	**	**	**
19-1	**	**	**	24-1			
19-2	**	**	**	24-2			
19-3	**	**	**	24-3			
19-4	**	**	**	24-4			
19-5	**	**	**	24-5			

D_{12} = difference between *MM*-class and *Mm*-class

D_{13} = difference between *MM*-class and *mm*-class

D_{23} = difference between *Mm*-class and *mm*-class

* = significant at p = .05 level

** = significant at p = .01 level

Table 4.4.3

Estimates for $\hat{D}$, $\hat{D}^*$ and $\hat{p}_n$

Data	$\hat{D}$	$\hat{D}^*$	$\hat{p}_n$	Data	$\hat{D}$	$\hat{D}^*$	$\hat{p}_n$
7-1	.115	.576	.590	18-1	.100	1.000	.800
7-3	.118	.479	.490	18-3	.137	.565	.490
				18-4	.045	1.000	.090
8-4	.138	.615	.500	18-5	.137	.600	.540
9-1	.055	1.000	.110	19-1	.197	1.000	.600
9-2	.029	.292	.200	19-2	.208	1.000	.580
9-3	.124	.600	.420	19-3	.249	1.000	.510
9-4	.016	.122	.720	19-4	.240	1.000	.520
9-5	.040	.333	.760	19-5	.234	1.000	.530
10-2	.000	.000	.100	20-1	.118	1.000	.760
10-3	.000	.000	.810	20-2	.135	.542	.500
10-4	.000	.000	.910	20-3	.177	1.000	.620
10-5	.087	.461	.380	20-4	.173	.873	.400
				20-5	.127	.508	.500
11-2	.055	.367	.300				
11-4	.030	.605	.900	21-1	.221	1.000	.560
11-5	.022	.458	.900	21-2	.245	1.000	.490
				21-3	.190	1.000	.380
12-2	.135	.652	.580	21-4	.243	1.000	.520
12-3	.066	.337	.600	21-5	.176	1.000	.350
12-5	.049	.353	.720				
				22-1	.197	1.000	.610
13-1	.038	.168	.460	22-2	.186	1.000	.550
13-2	.074	.402	.370	22-3	.231	.991	.470
13-3	.097	.444	.440	22-4	.181	.767	.470
13-5	.038	.207	.370	22-5	.155	1.000	.310
14-1	.038	.167	.470	23-1	.243	.988	.500
14-2	.062	.341	.360	23-2	.236	1.000	.450
14-3	.051	.867	.880	23-3	.149	1.000	.300
14-5	.077	1.000	.150	23-4	.226	1.000	.440
				23-5	.161	1.000	.340
17-1	.227	1.000	.450				
17-2	.211	1.000	.580				
17-3	.244	1.000	.500				
17-4	.185	1.000	.630				
17-5	.247	1.000	.490				

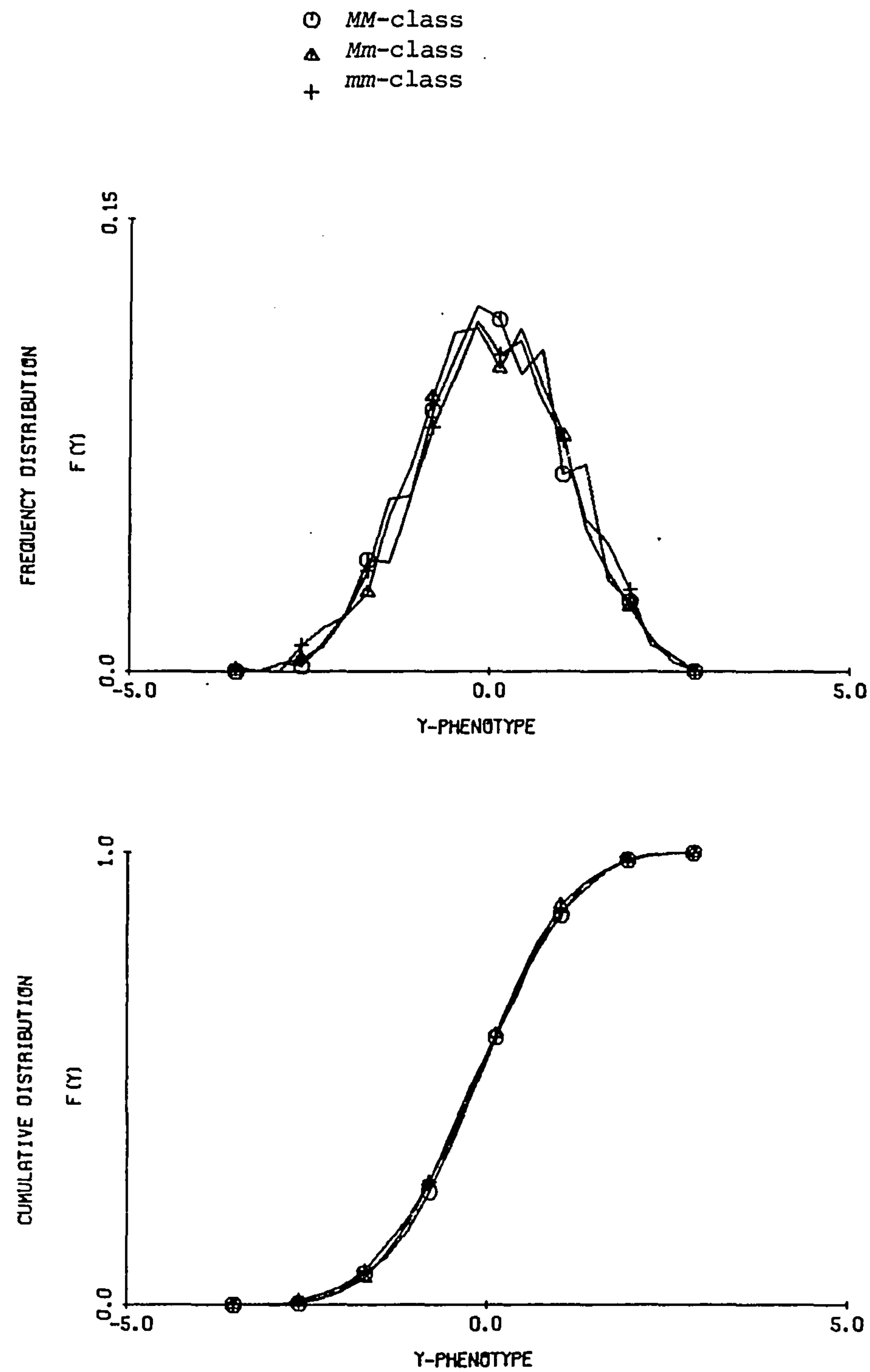


Figure 4.4.1 Subclass Distributions (data 1-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=2$, $N=5000$, $D^*=0$

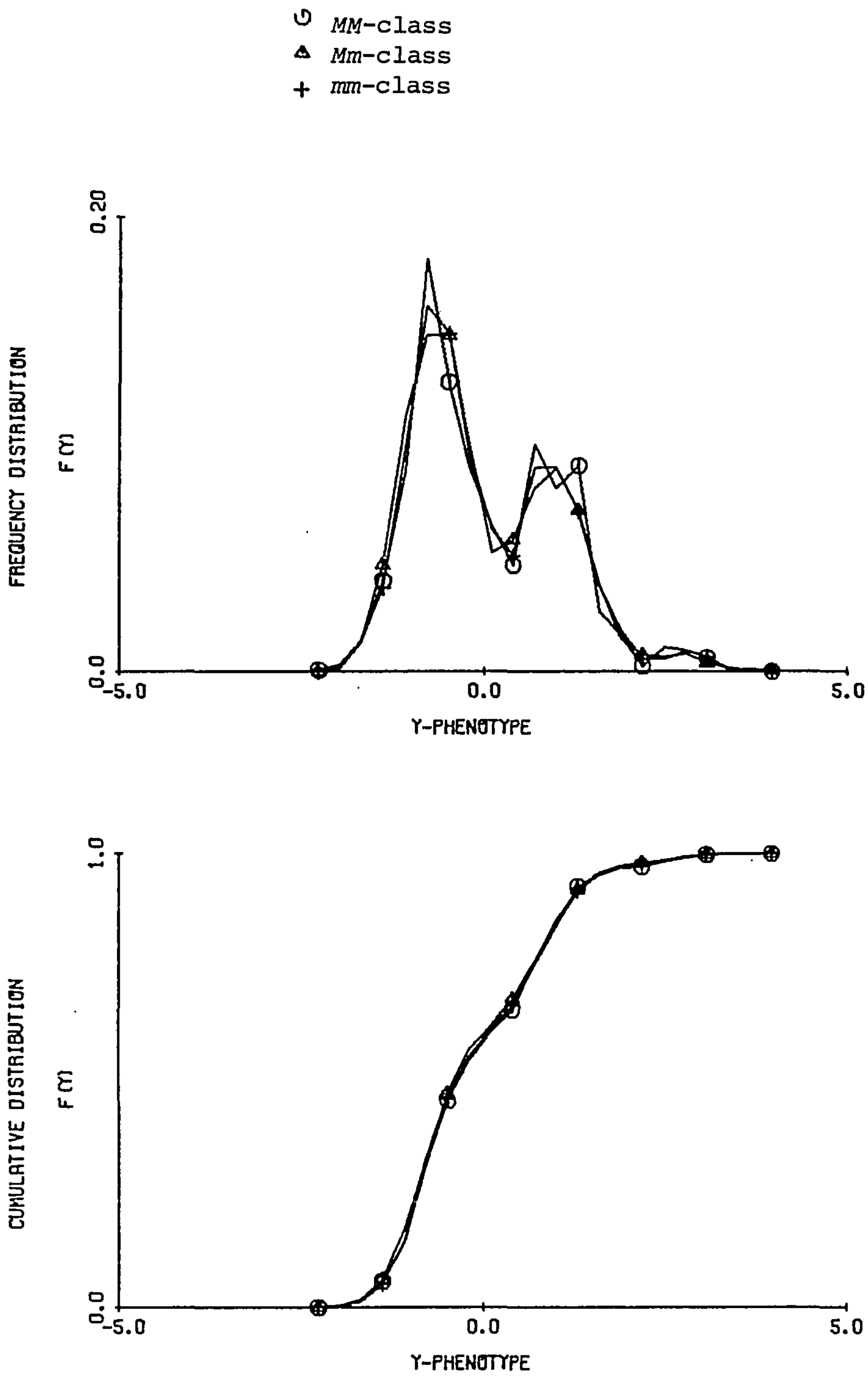


Figure 4.4.2 Subclass Distributions (data 2-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=1$, $n=2$, $N=5000$, $D^*=0$

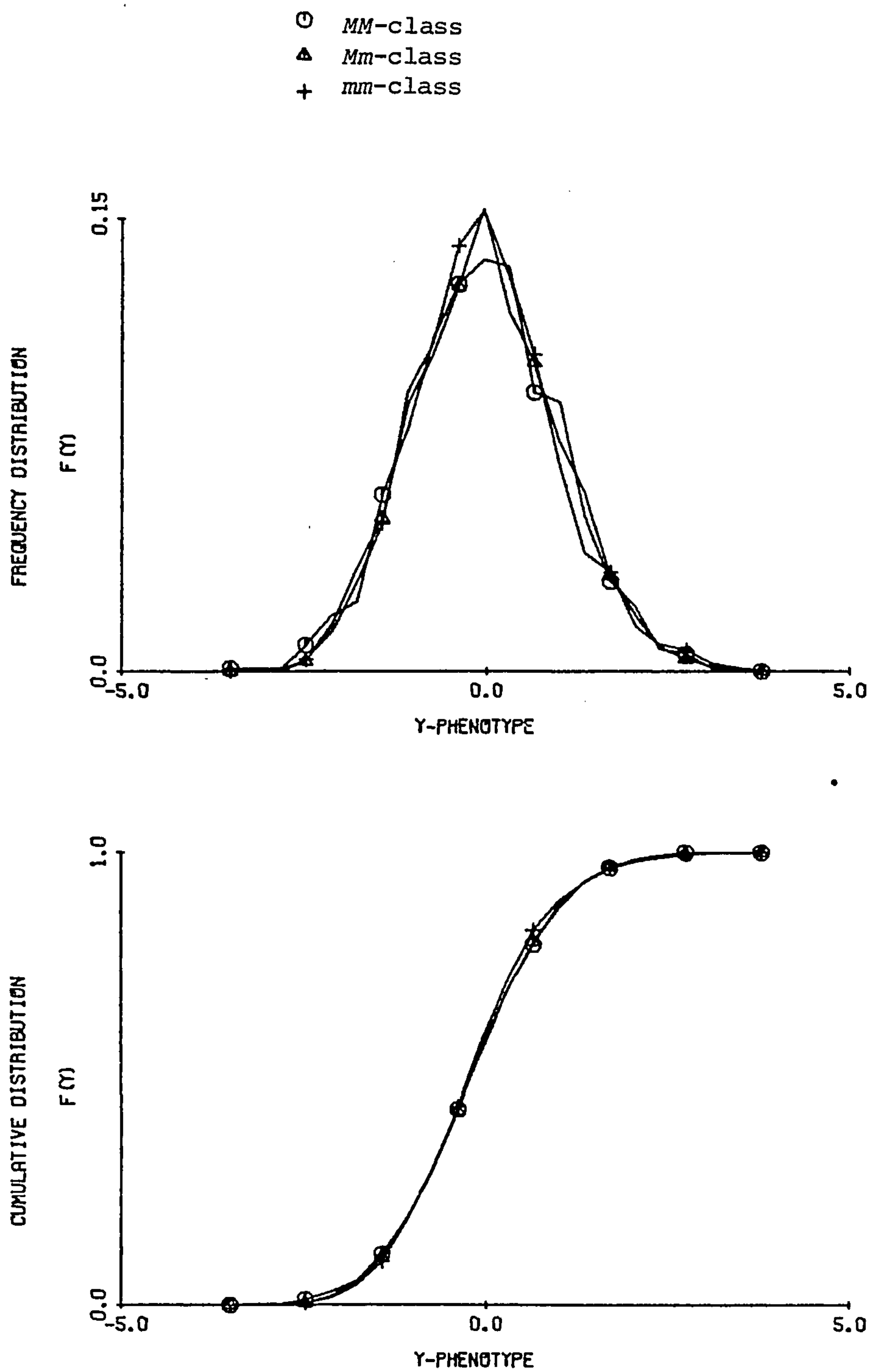


Figure 4.4.3 Subclass Distributions (data 3-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.5$, $x=0$, $n=2$, $N=5000$, $D^*=0$

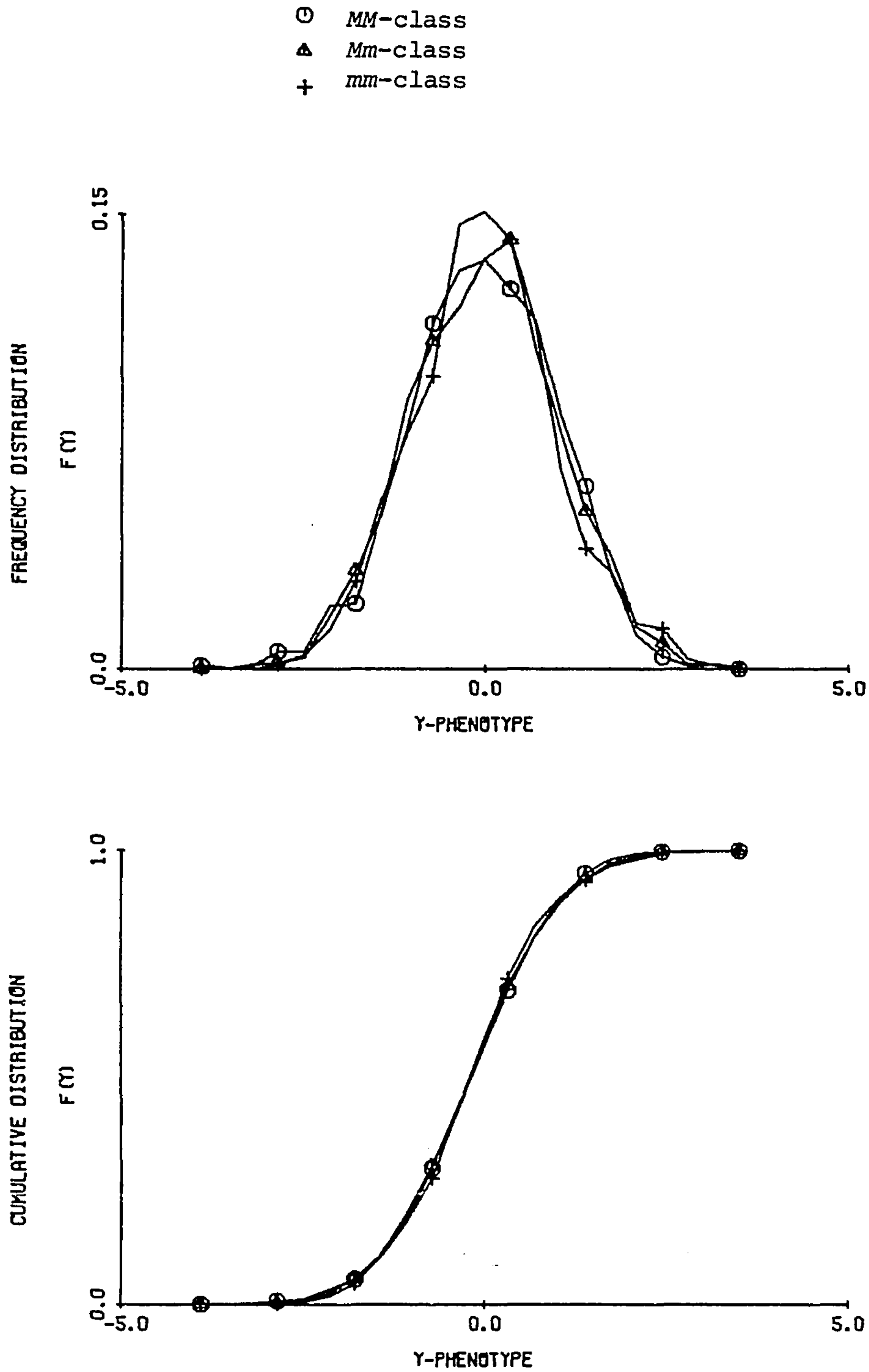


Figure 4.4.4 Subclass Distributions (data 4-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.2$, $x=0$, $n=2$, $N=5000$, $D^*=0$

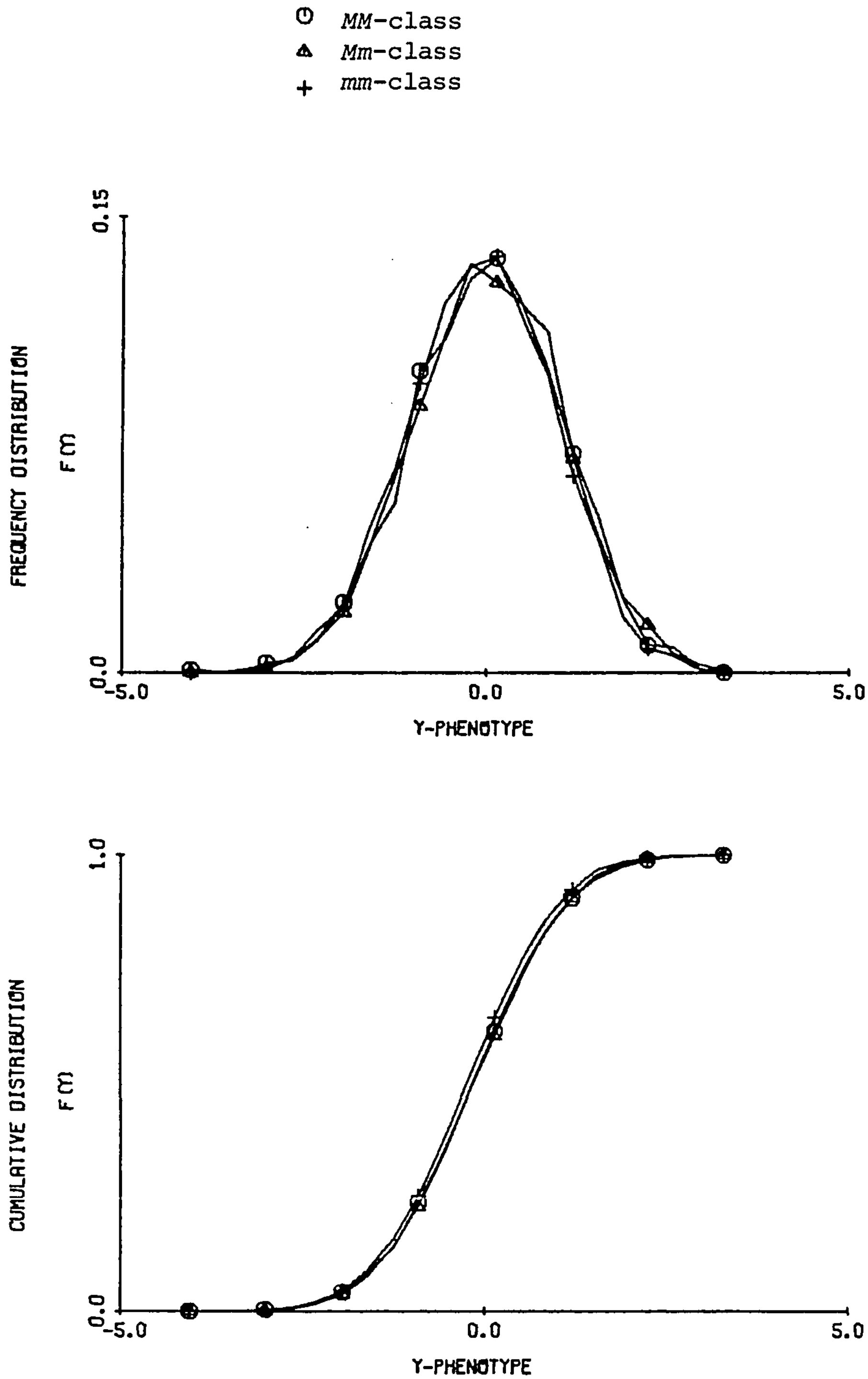


Figure 4.4.5 Subclass Distributions (data 5-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=5$, $N=5000$, $D^*=0$

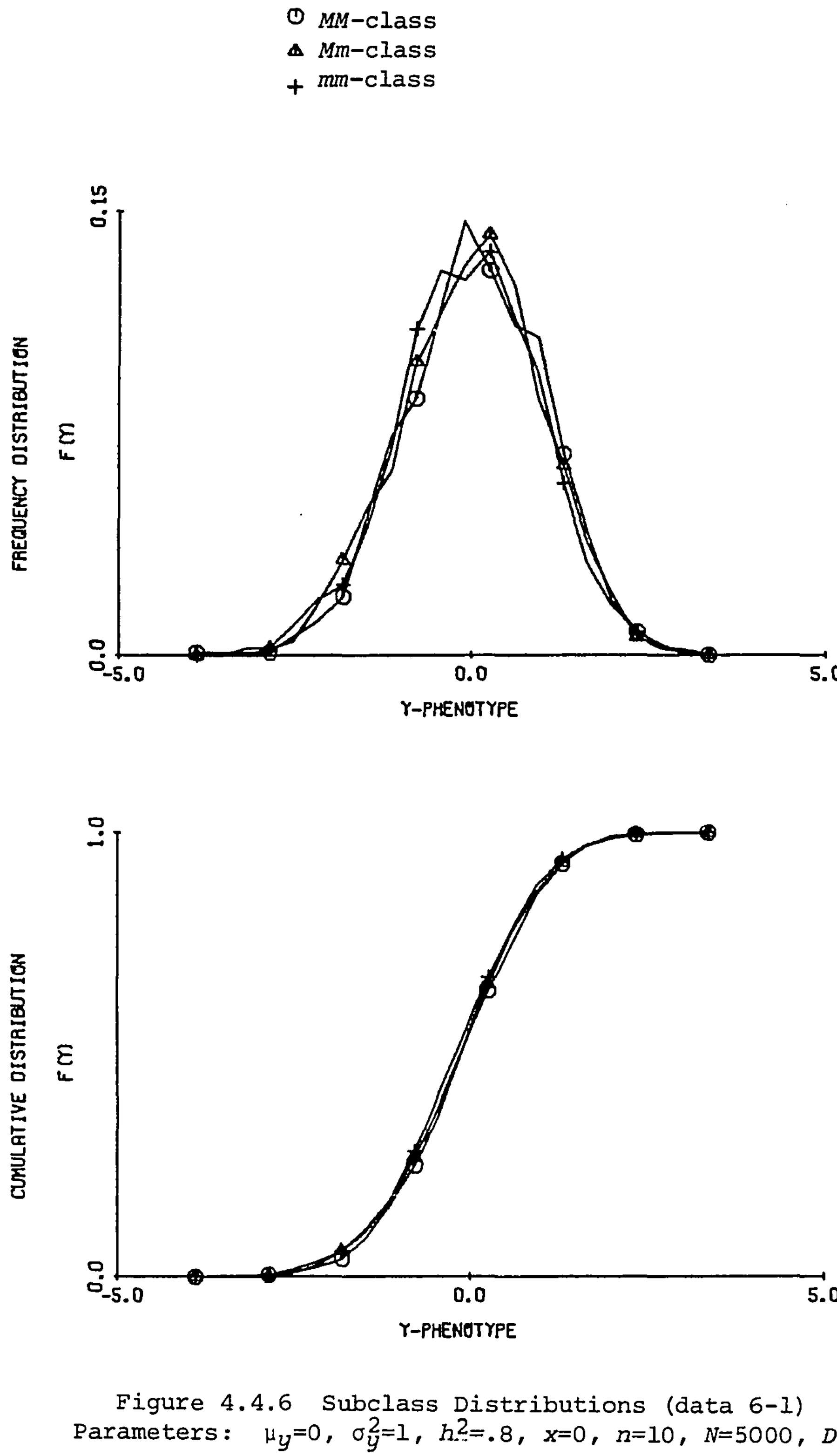


Figure 4.4.6 Subclass Distributions (data 6-1)
Parameters: μ_y=0, σ_y^2=1, h^2=.8, x=0, n=10, N=5000, D^*=0

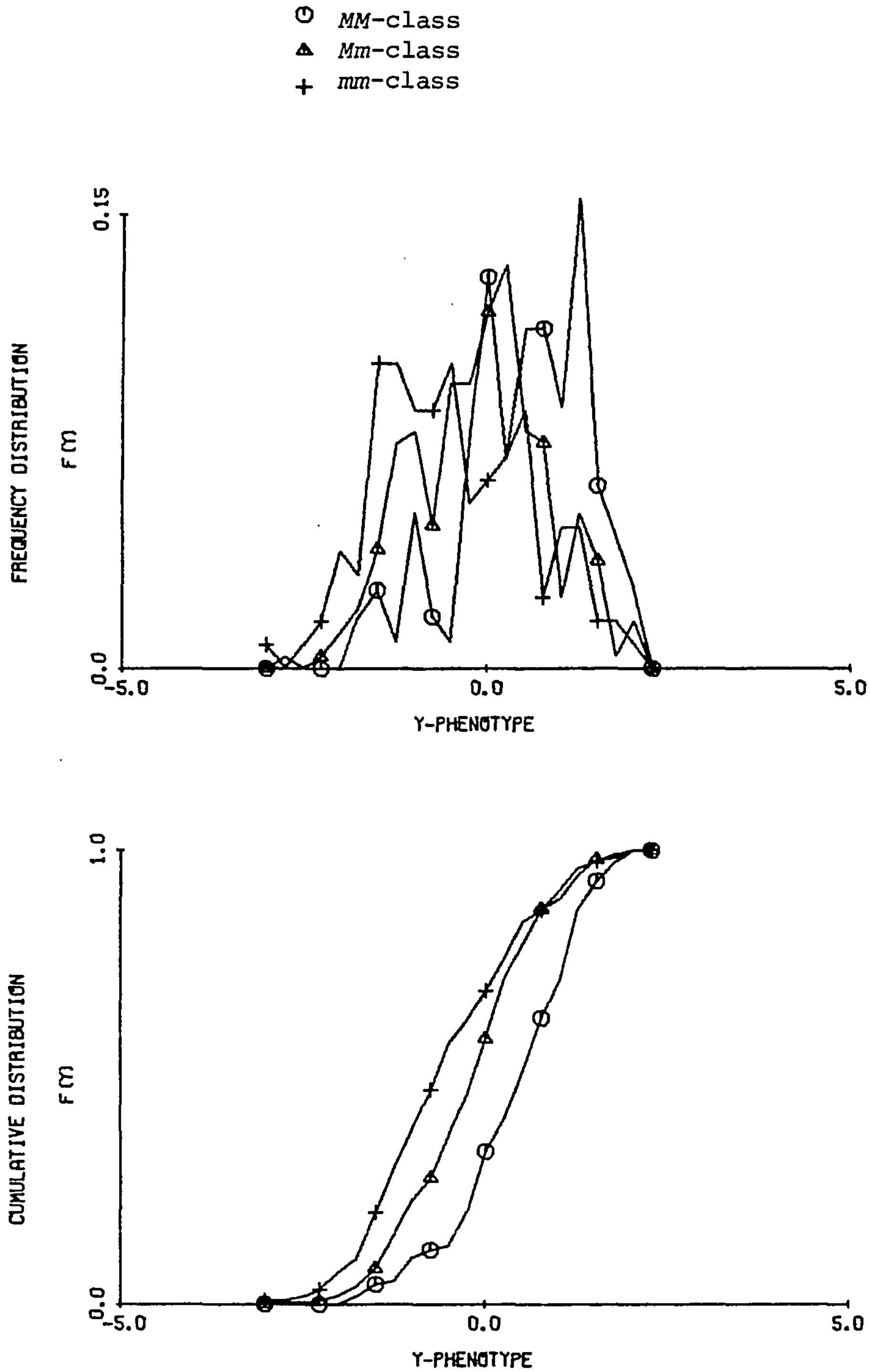


Figure 4.4.7 Subclass Distributions (data 7-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=2$, $N=500$, $D^*=0$

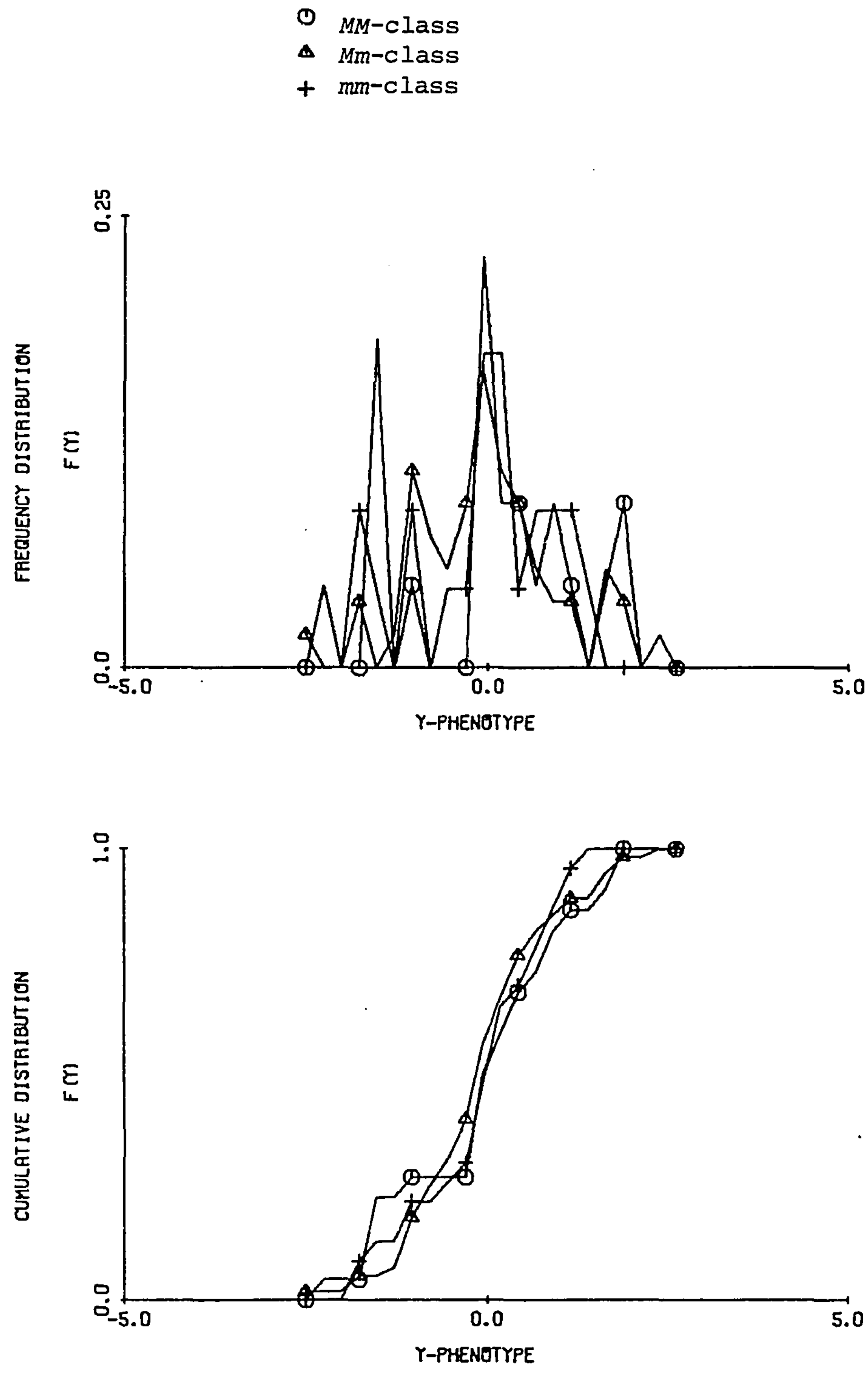


Figure 4.4.8 Subclass Distributions (data 8-1)
Parameters: μ_y=0, σ_y^2=1, h^2=.8, x=0, n=2, N=100, D^*=0

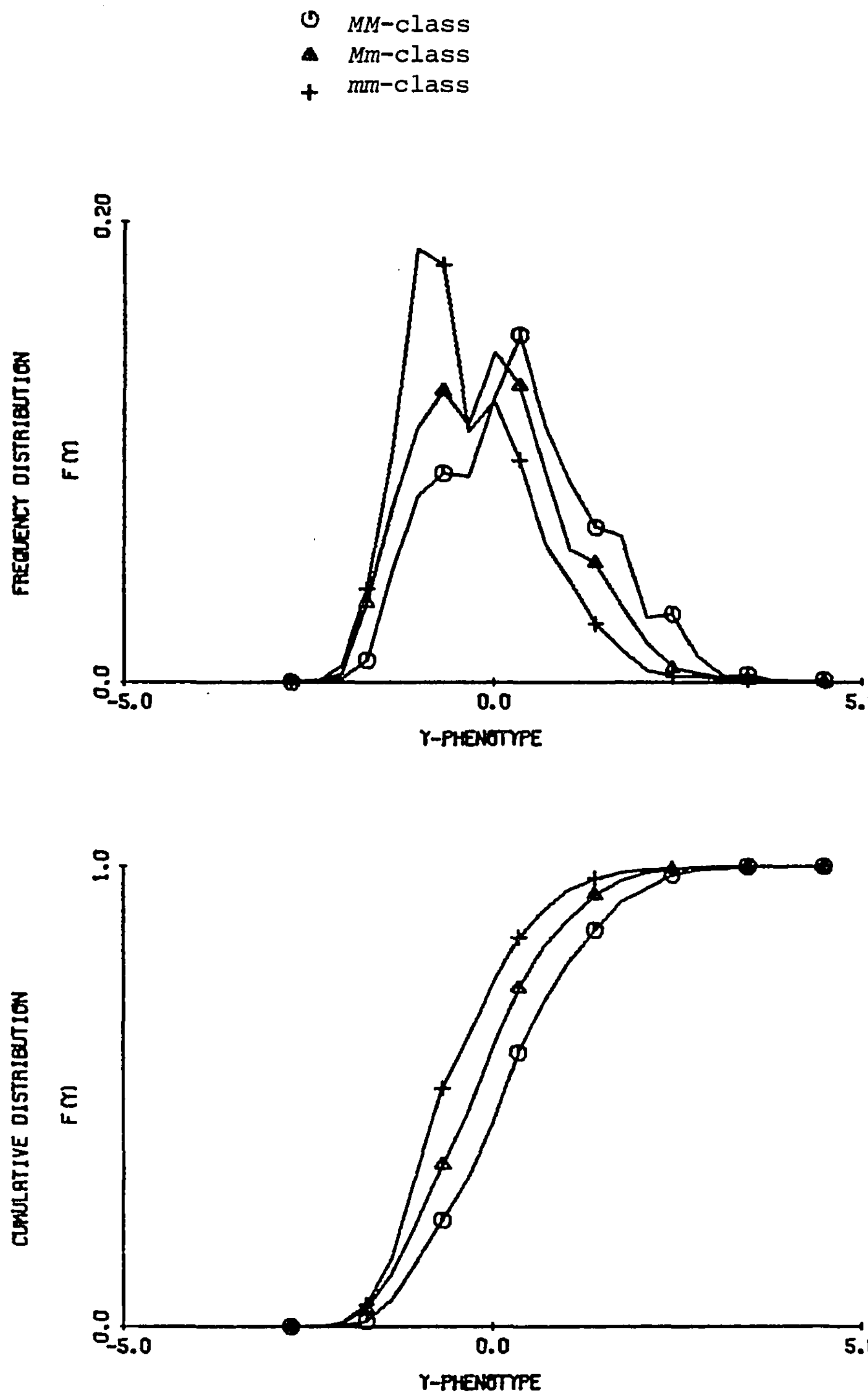
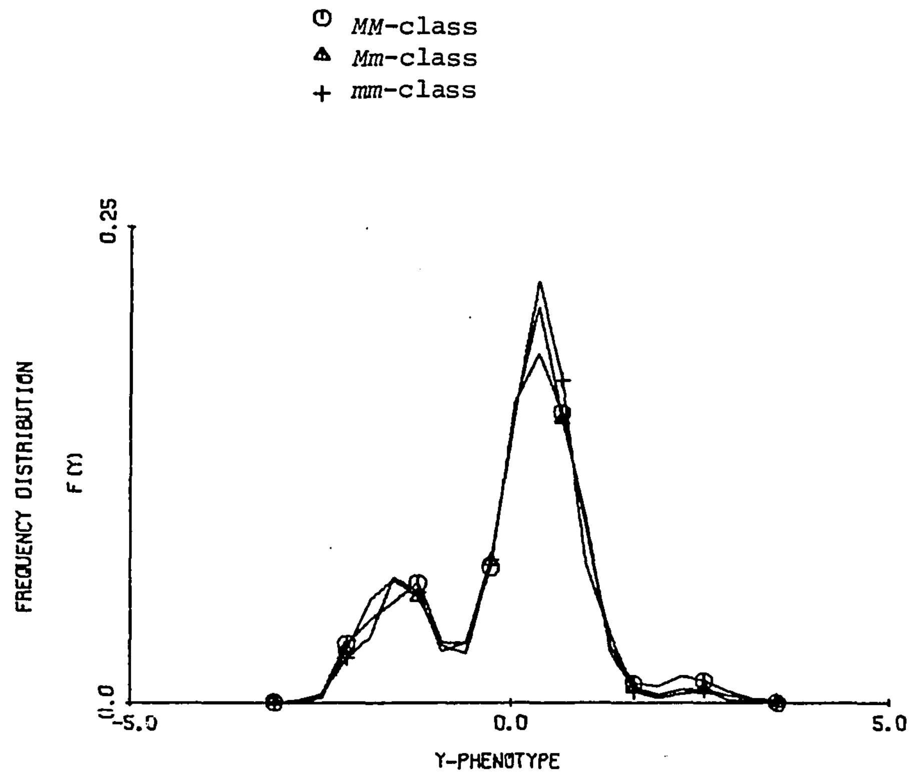


Figure 4.4.9 Subclass Distributions (data 9-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=2$, $N=5000$, $D^*=.5$



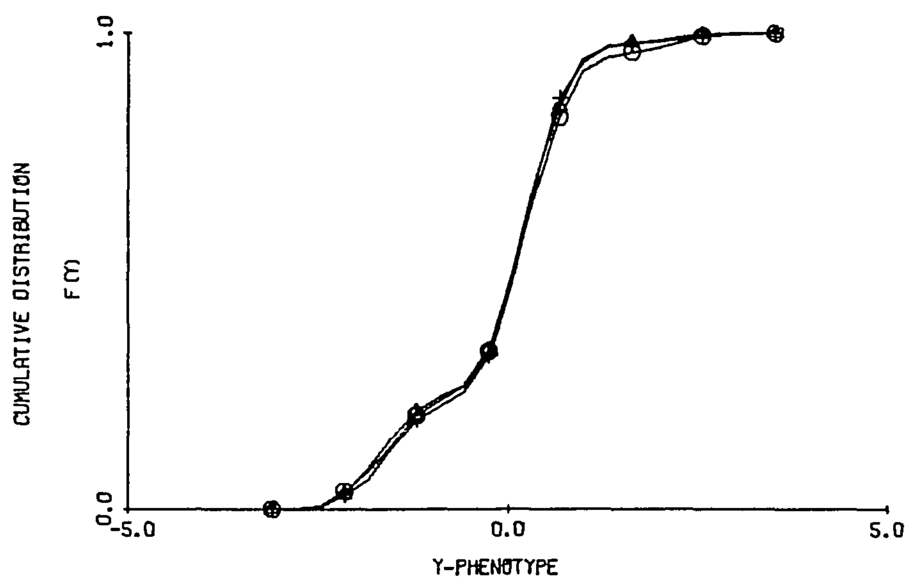


Figure 4.4.10 Subclass Distributions (data 10-1)
Parameters: μ_y=0, σ_y^2=1, h^2=.8, x=1, n=2, N=5000, D^*=.5

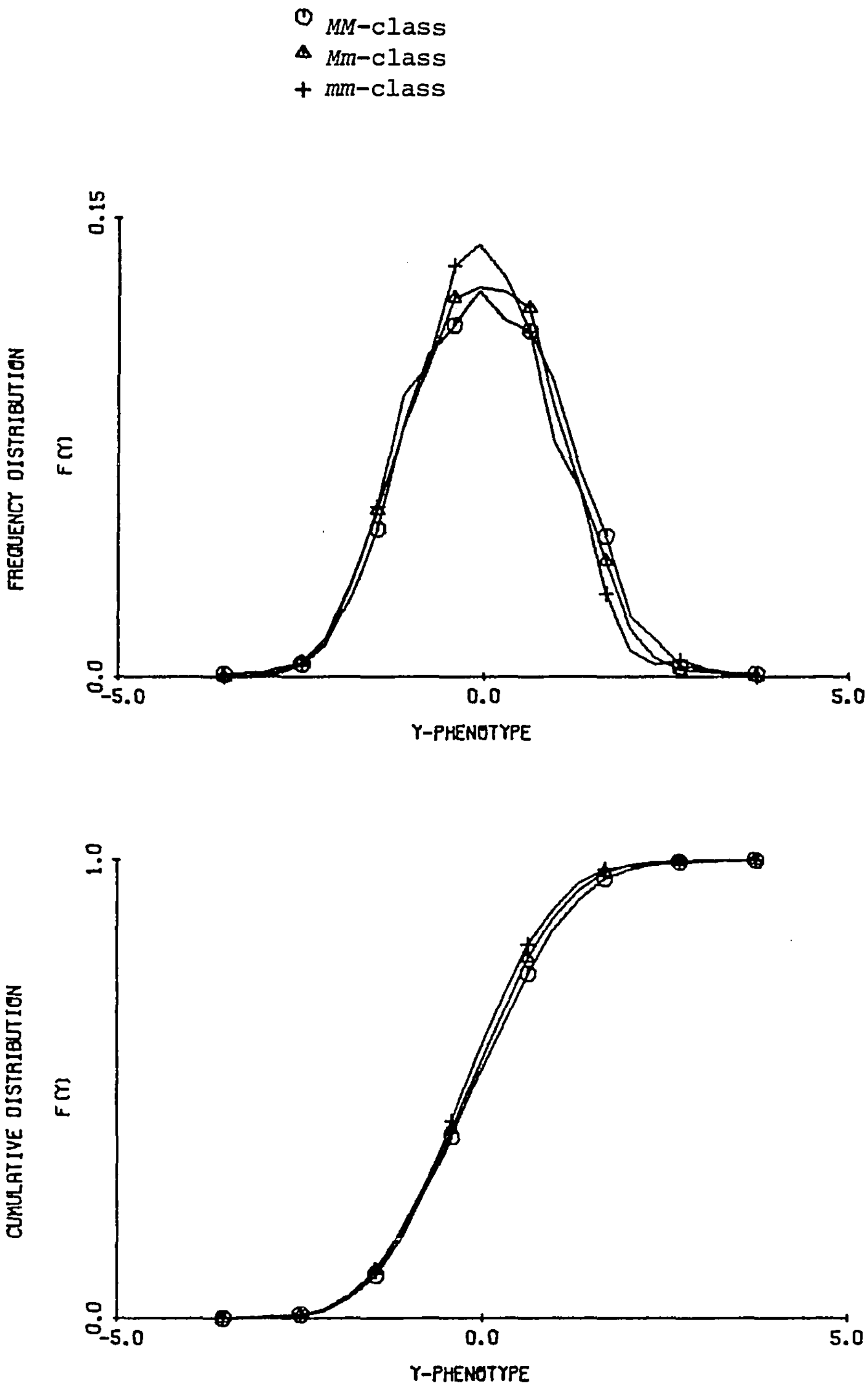


Figure 4.4.11 Subclass Distributions (data 11-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.5$, $x=0$, $n=2$, $N=5000$, $D^*=.5$

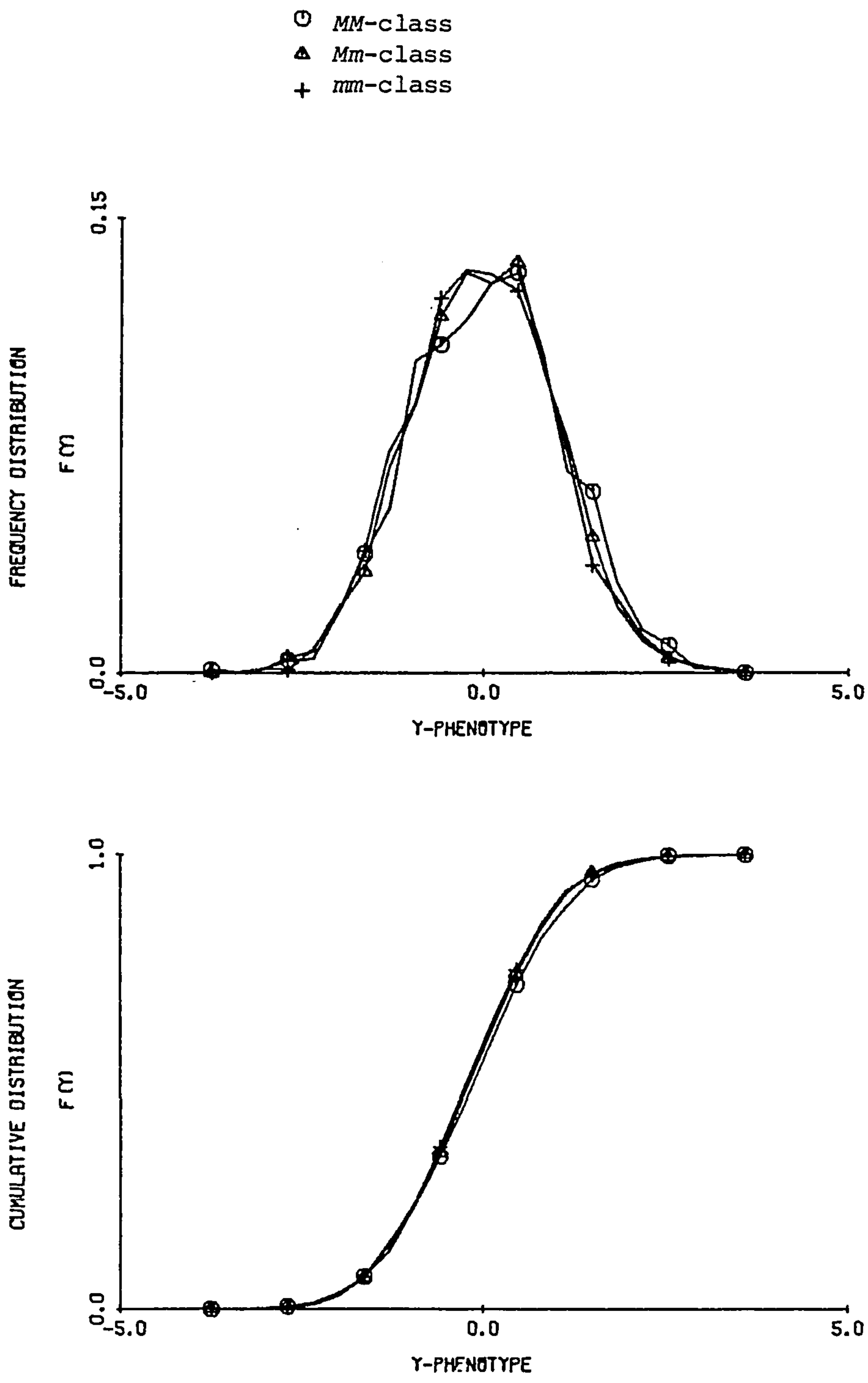


Figure 4.4.12 Subclass Distributions (data 12-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.2$, $x=0$, $n=2$, $N=5000$, $D^*=.5$

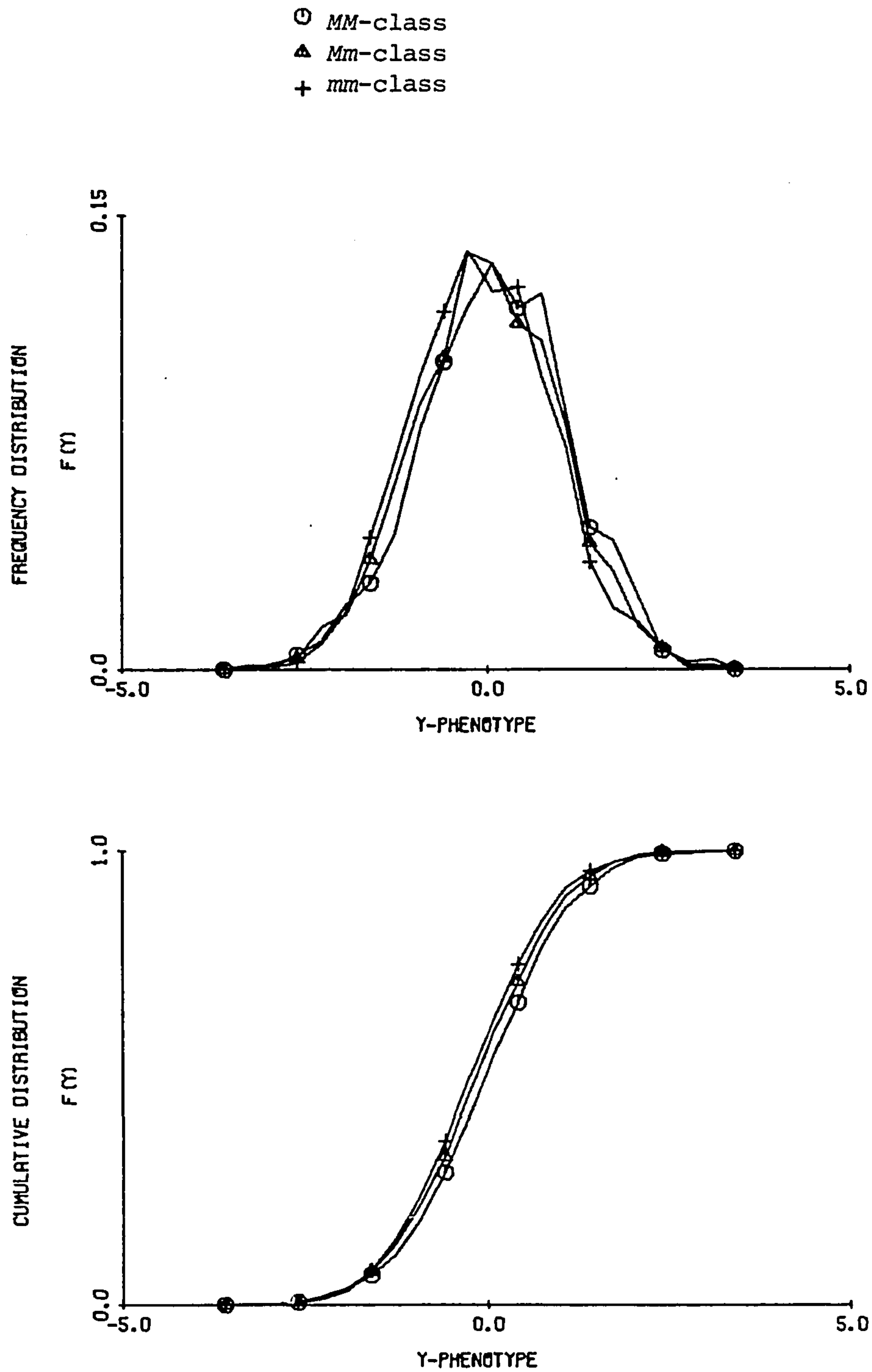


Figure 4.4.13 Subclass Distributions (data 13-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=5$, $N=5000$, $D^*=.5$

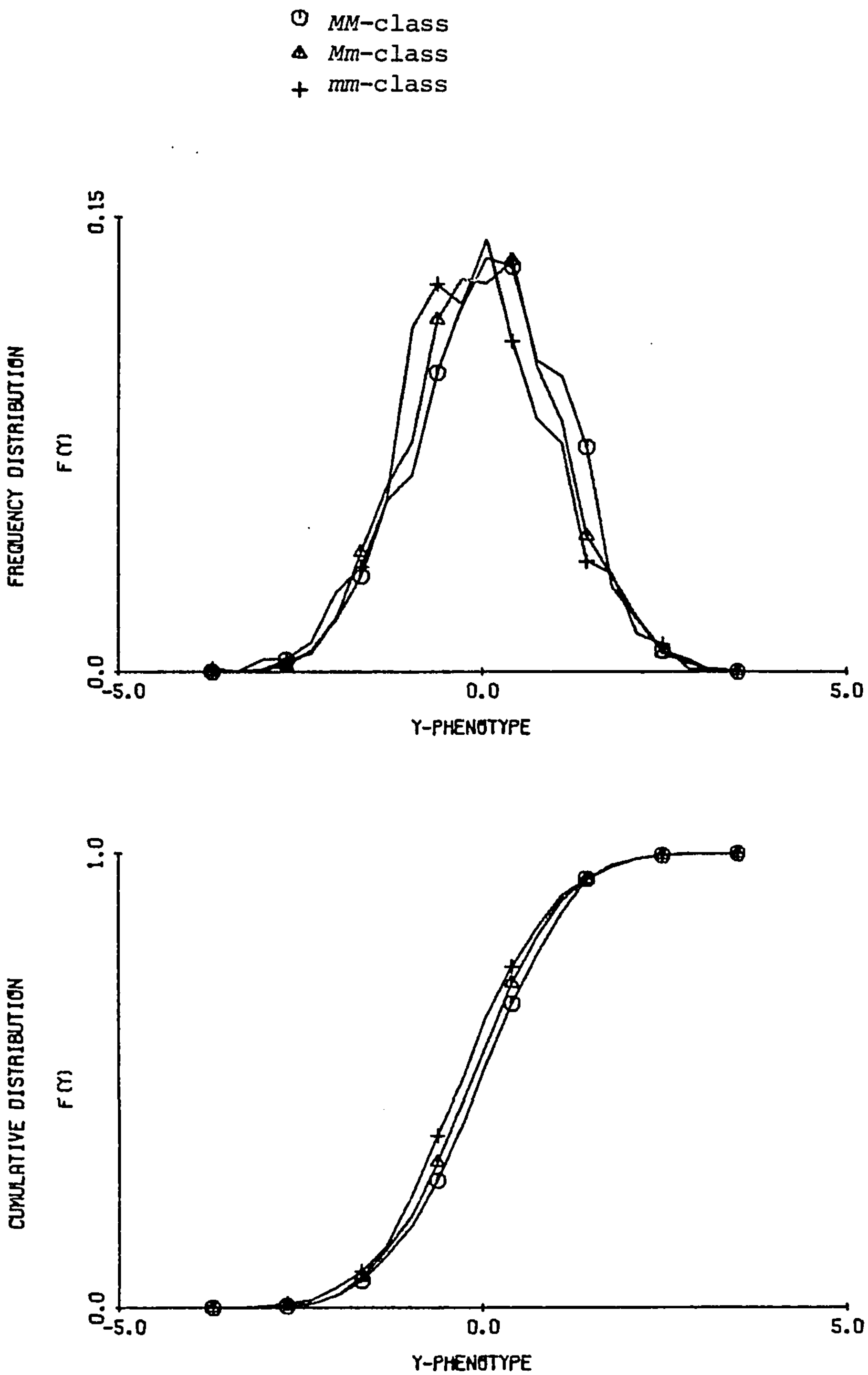


Figure 4.4.14 Subclass Distributions (data 14-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=10$, $N=5000$, $D^*=.5$

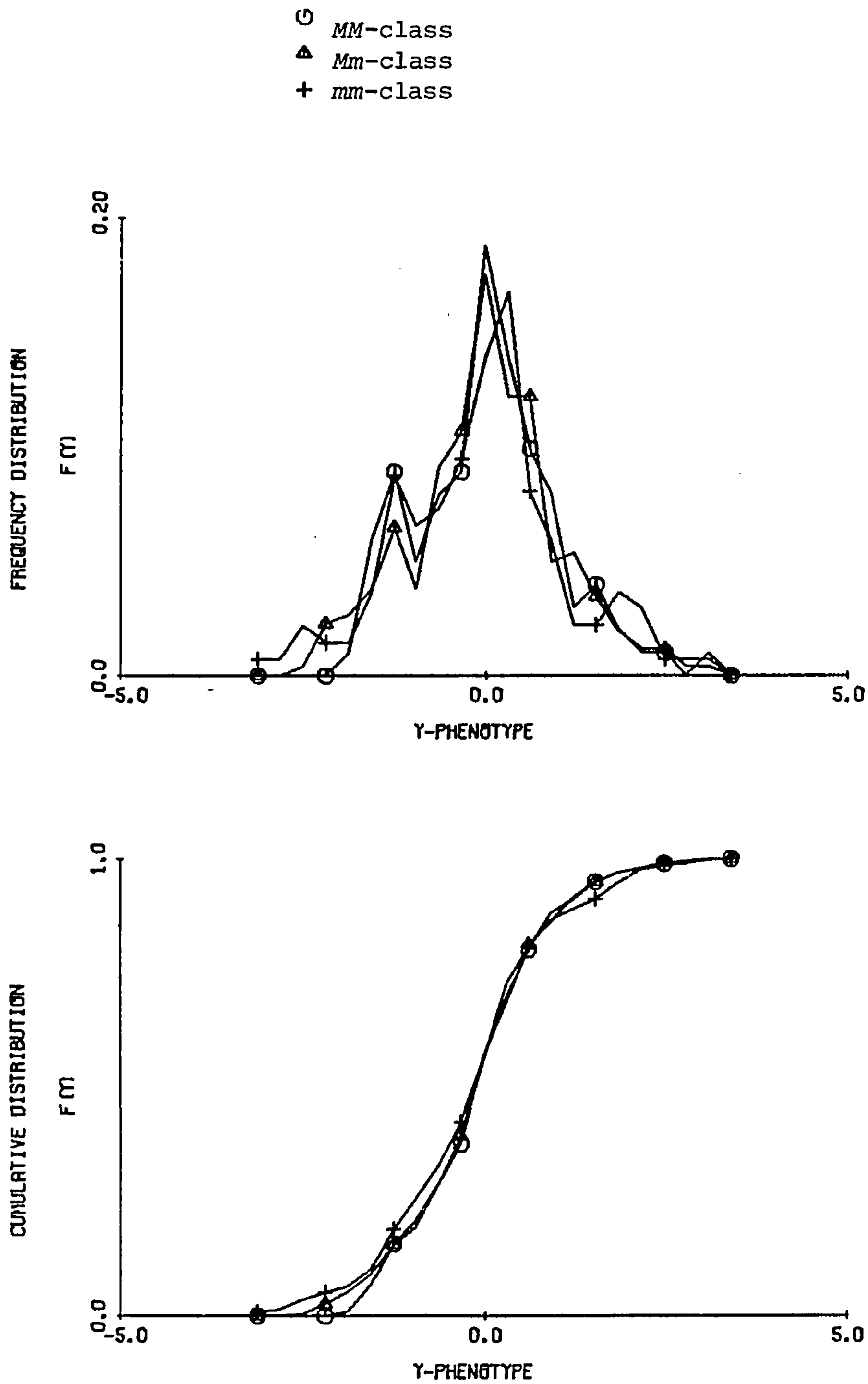


Figure 4.4.15 Subclass Distributions (data 15-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=2$, $N=500$, $D^*=.5$

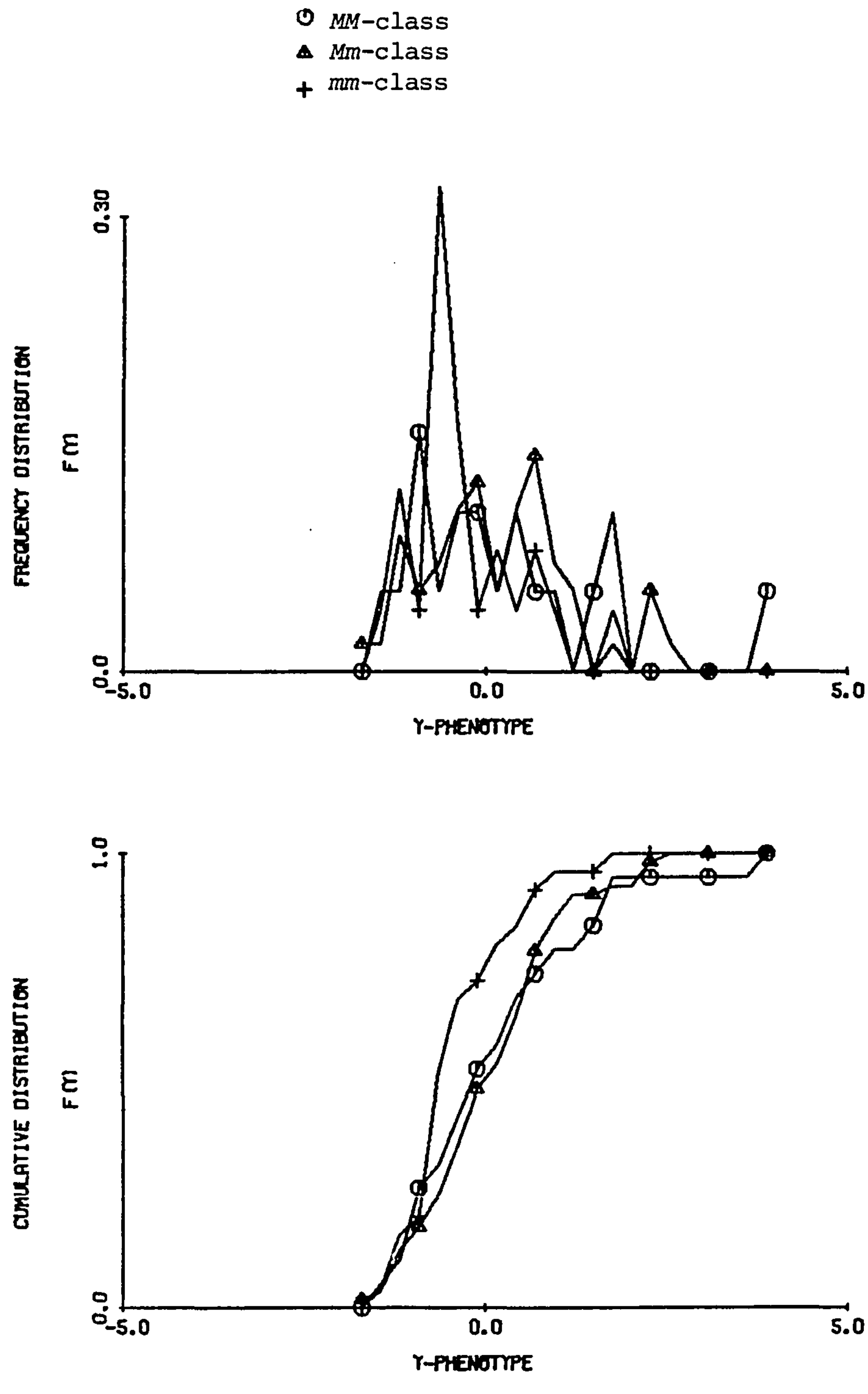


Figure 4.4.16 Subclass Distributions (data 16-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=2$, $N=100$, $D^*=.5$

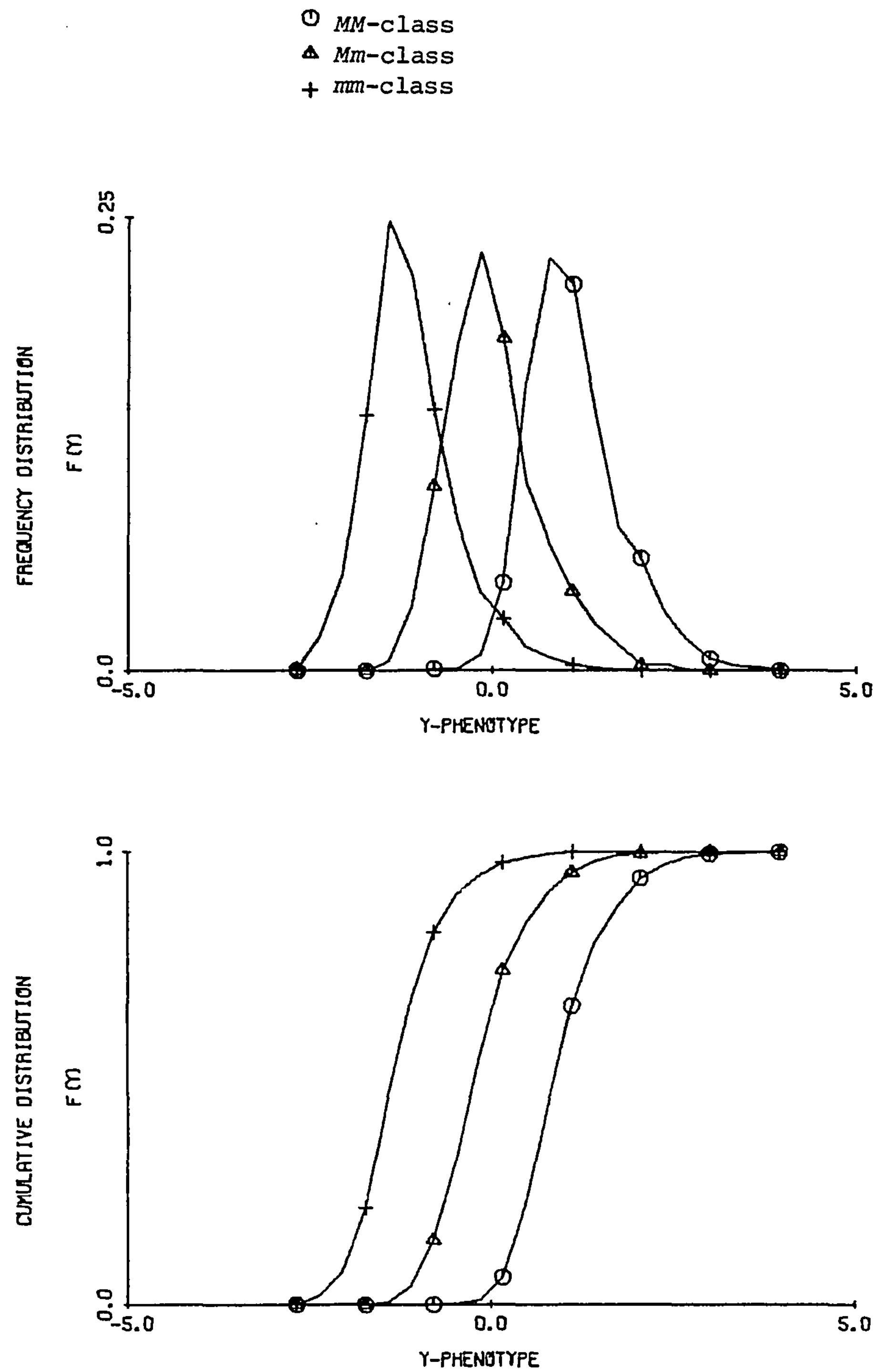


Figure 4.4.17 Subclass Distributions (data 17-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=2$, $N=5000$, $D^*=1$

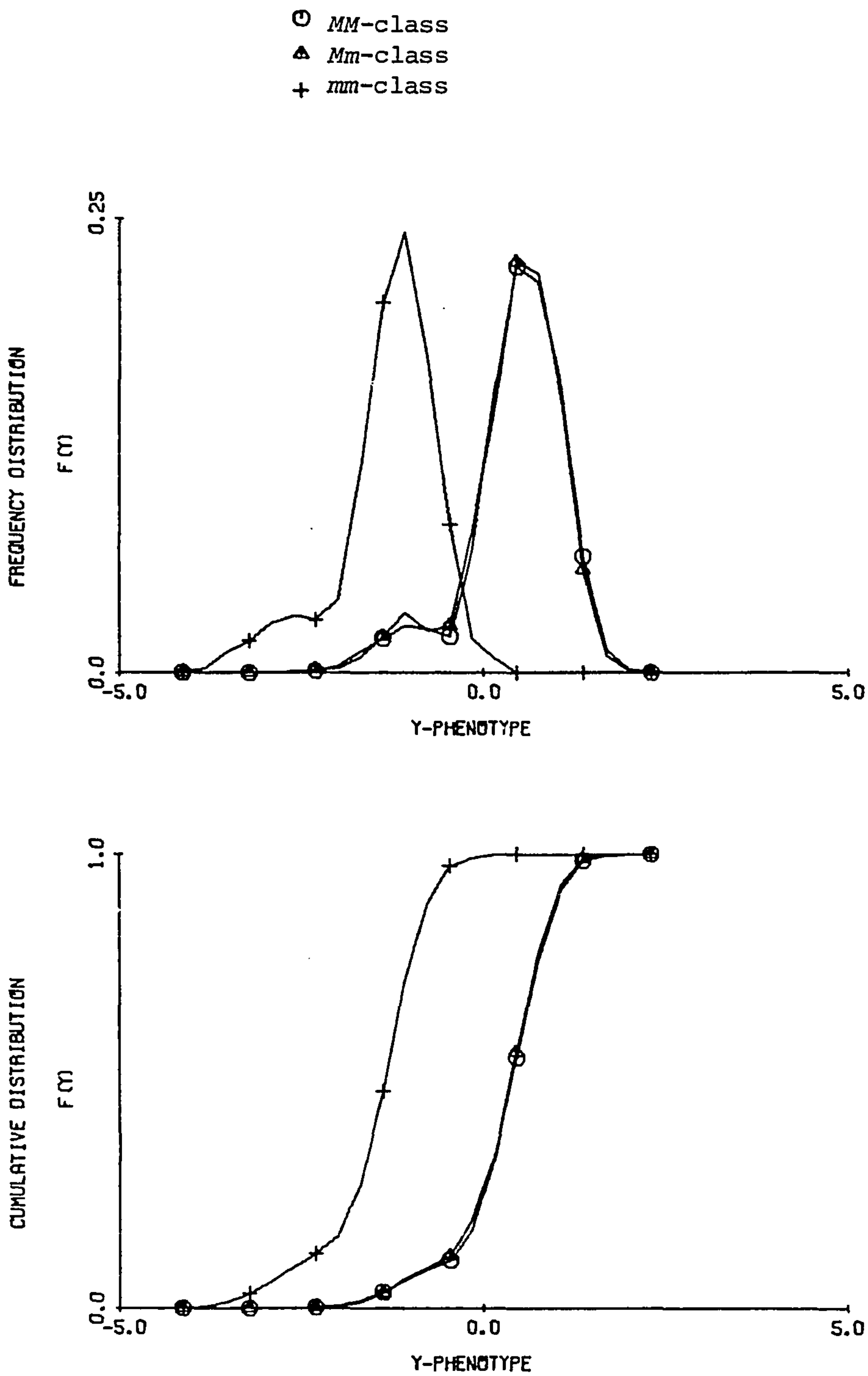


Figure 4.4.18 Subclass Distributions (data 18-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=1$, $n=2$, $N=5000$, $D^*=1$

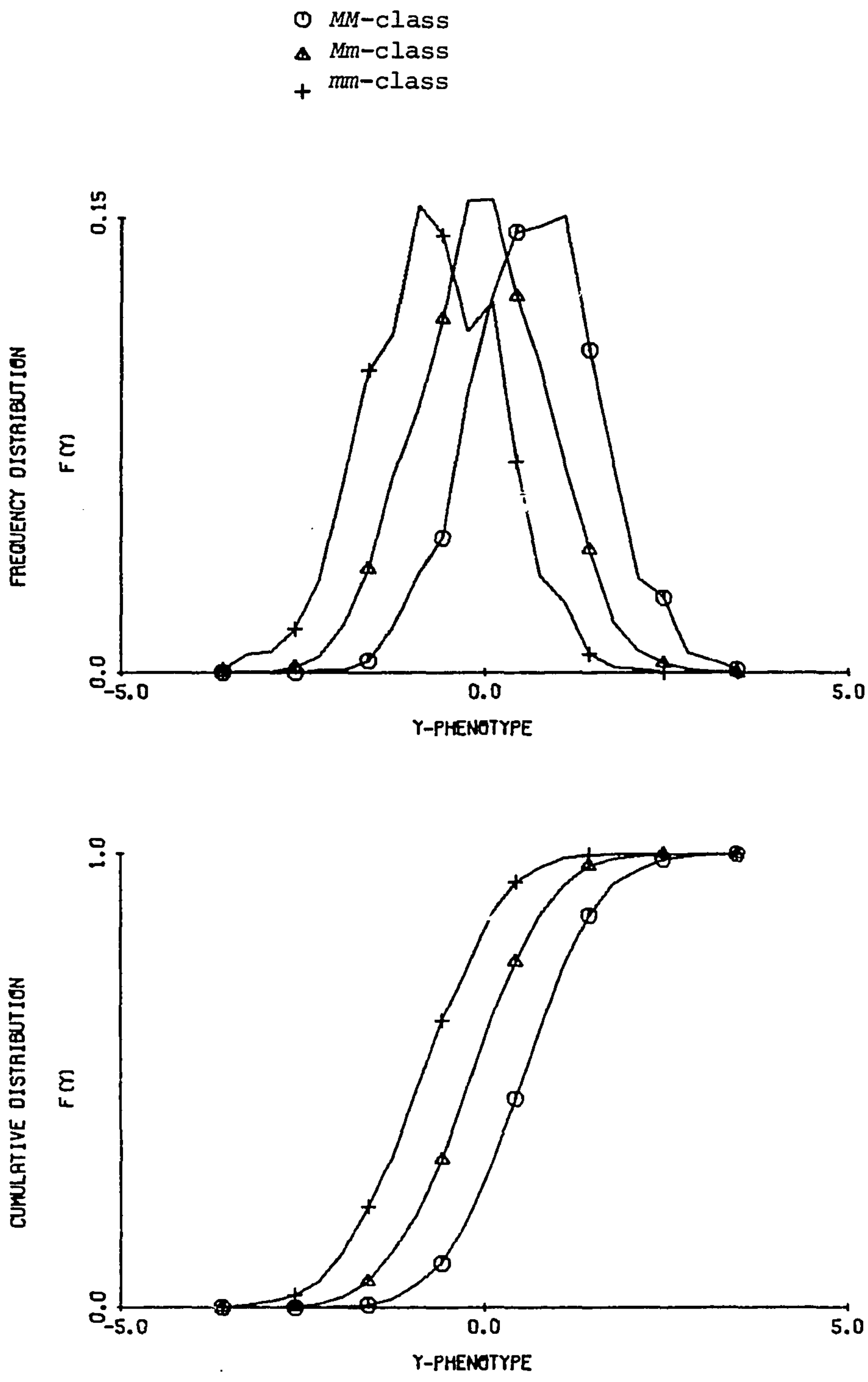


Figure 4.4.19 Subclass Distributions (data 19-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.5$, $x=0$, $n=2$, $N=5000$, $D^*=1$

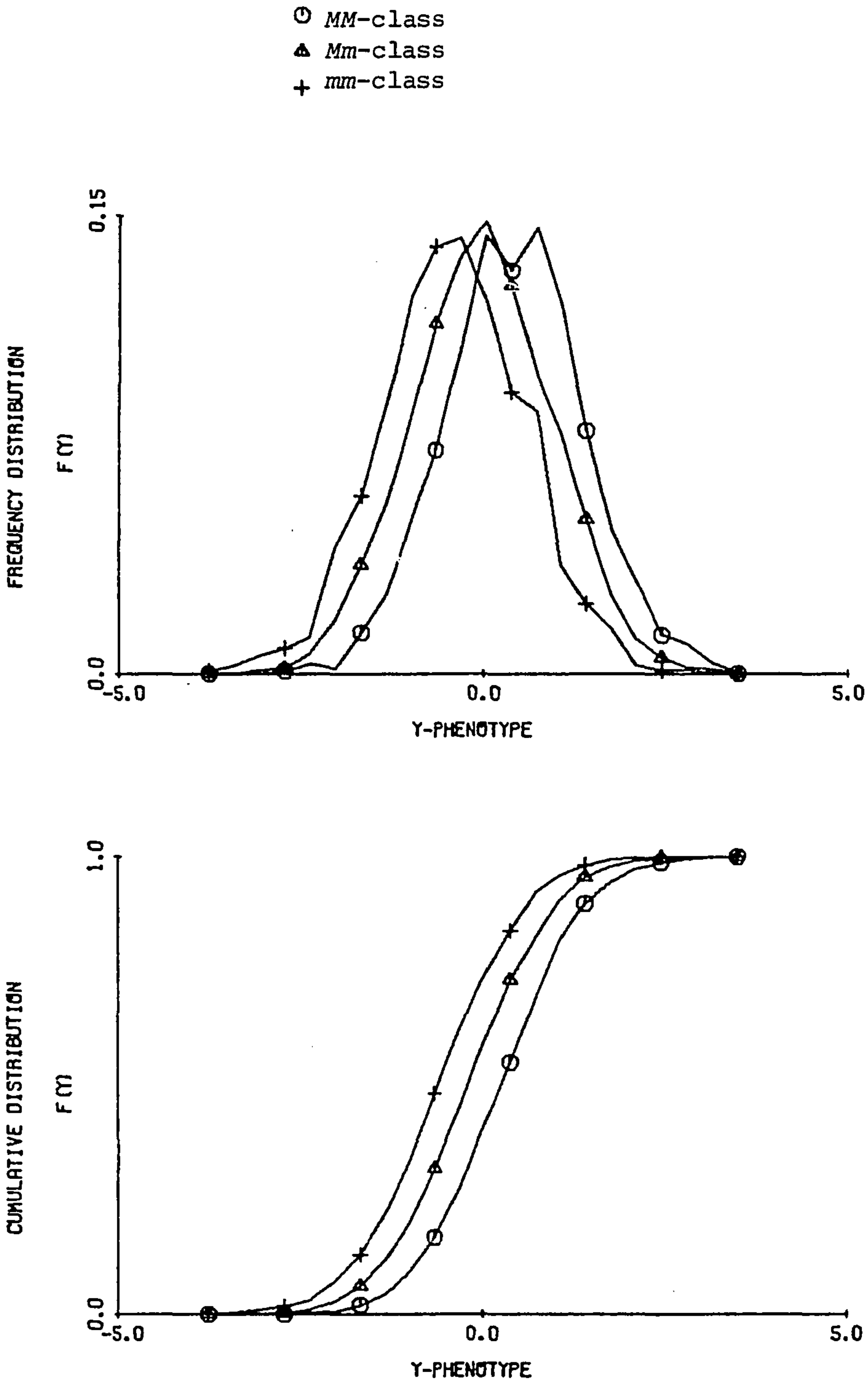


Figure 4.4.20 Subclass Distributions (data 20-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.2$, $x=0$, $n=2$, $N=5000$, $D^*=1$

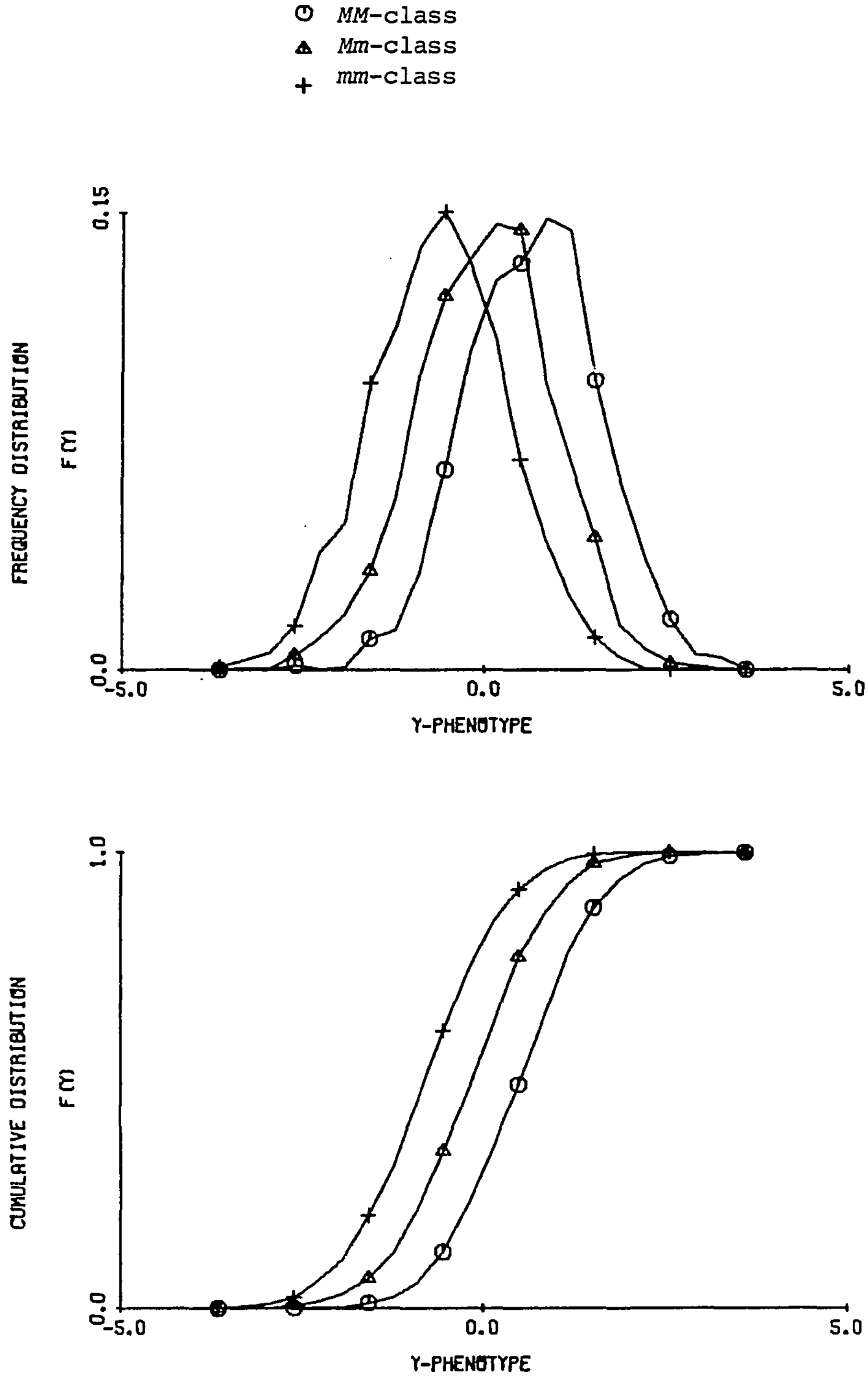


Figure 4.4.21 Subclass Distributions (data 21-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$; $n=5$, $N=5000$, $D^*=1$

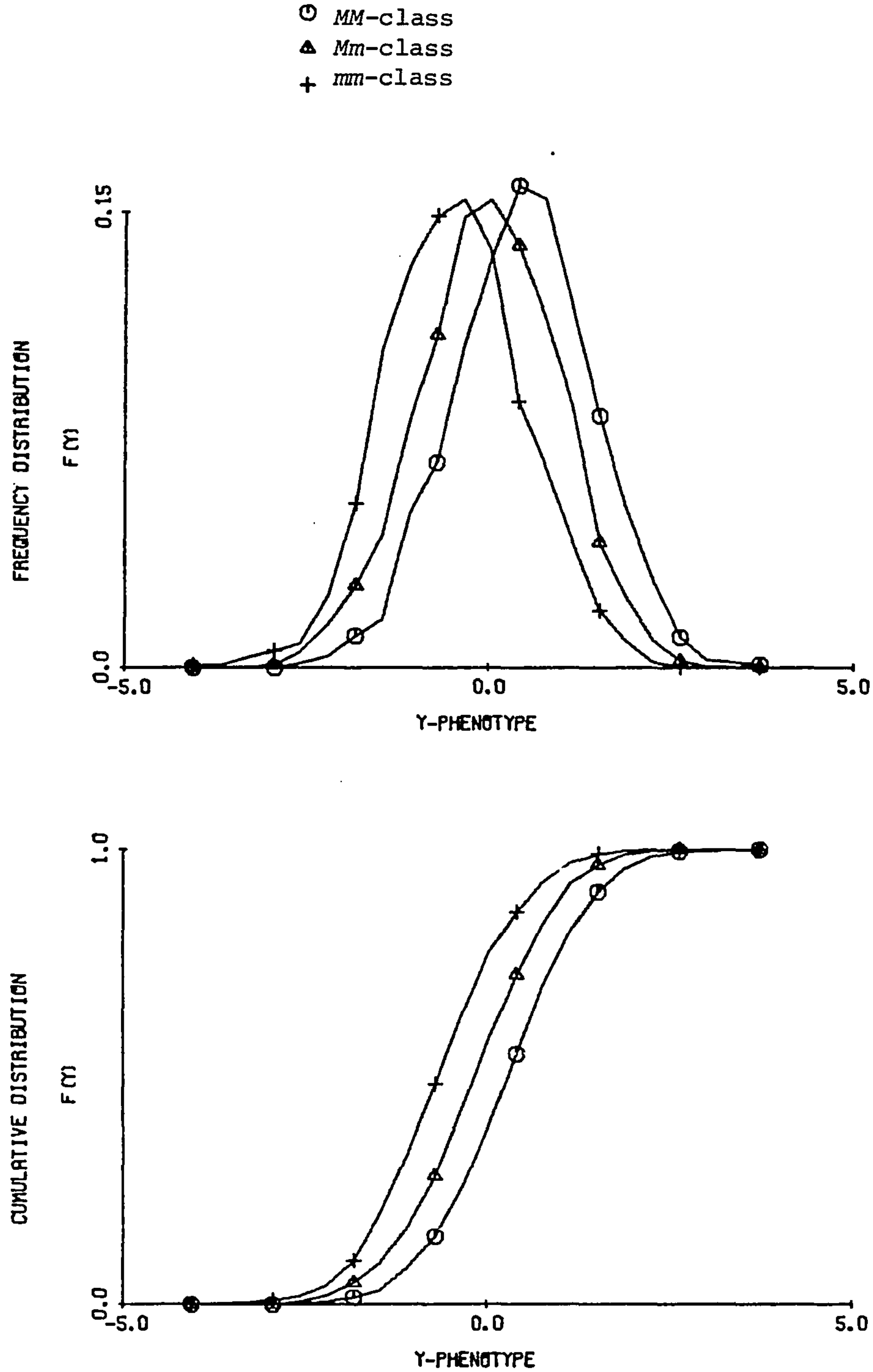


Figure 4.4.22 Subclass Distributions (data 22-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=10$, $N=5000$, $D^*=1$

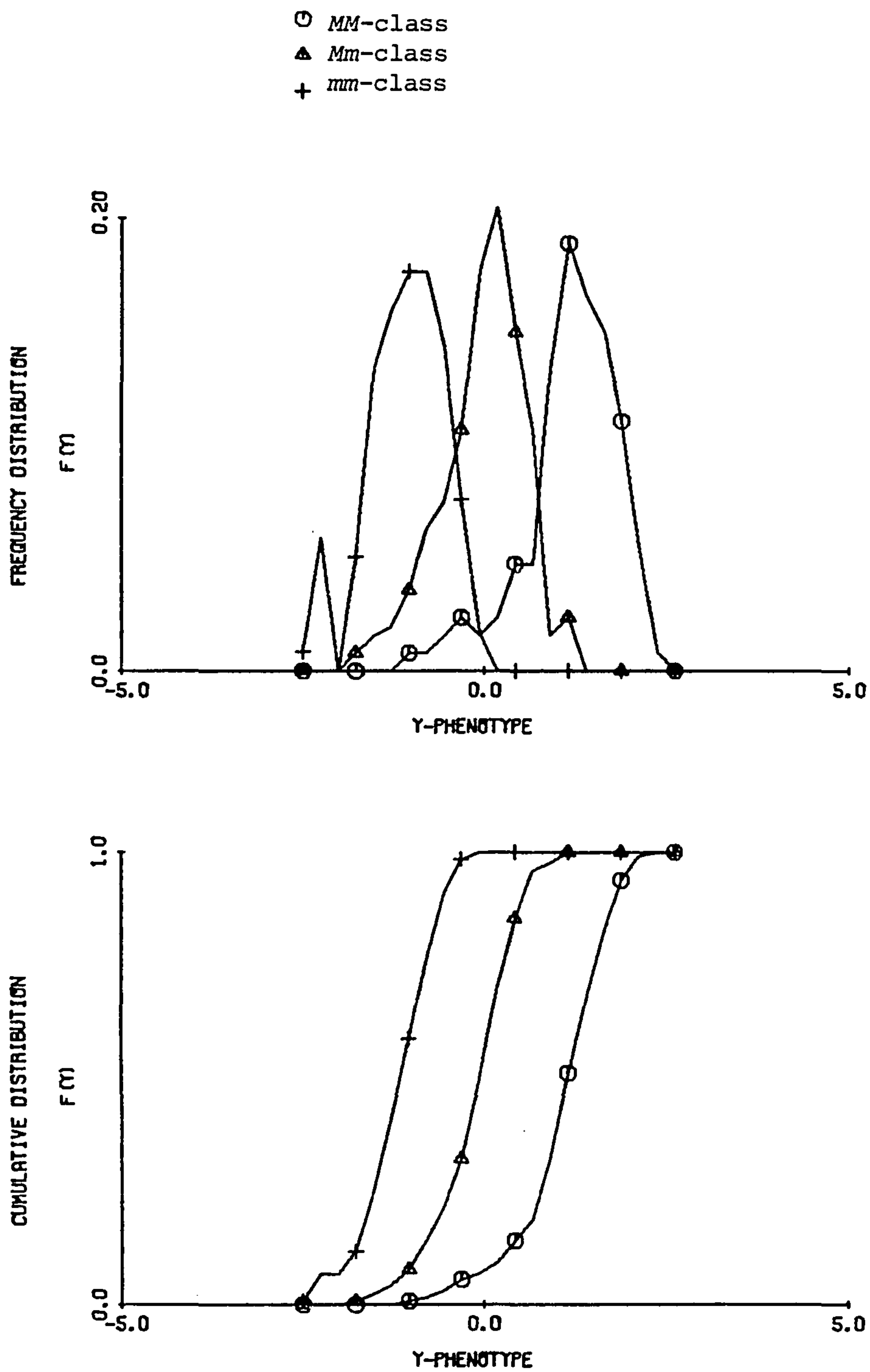


Figure 4.4.23 Subclass Distributions (data 23-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=2$, $N=500$, $D^*=1$

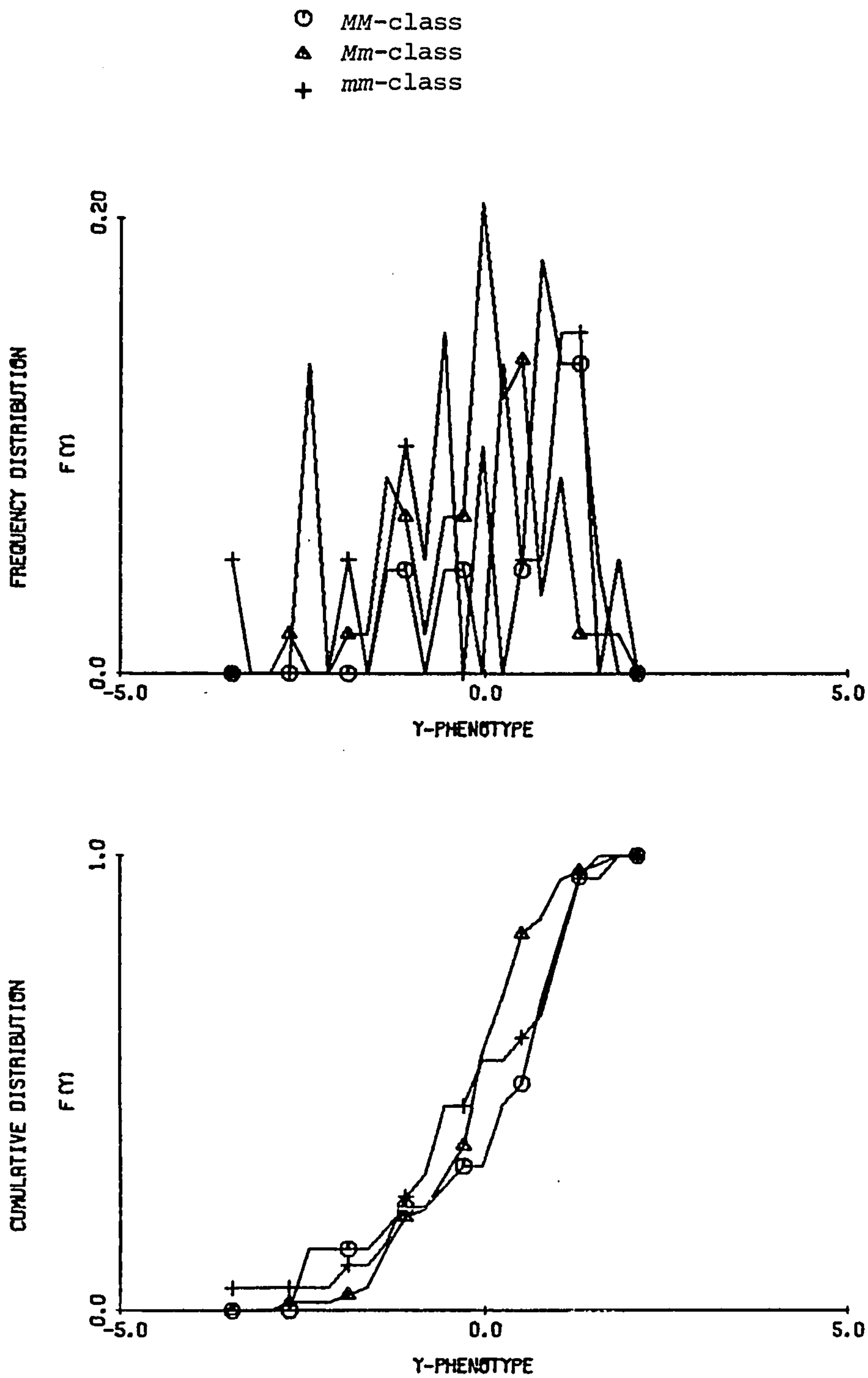


Figure 4.4.24 Subclass Distributions (data 24-1)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=2$, $N=100$, $D^*=1$

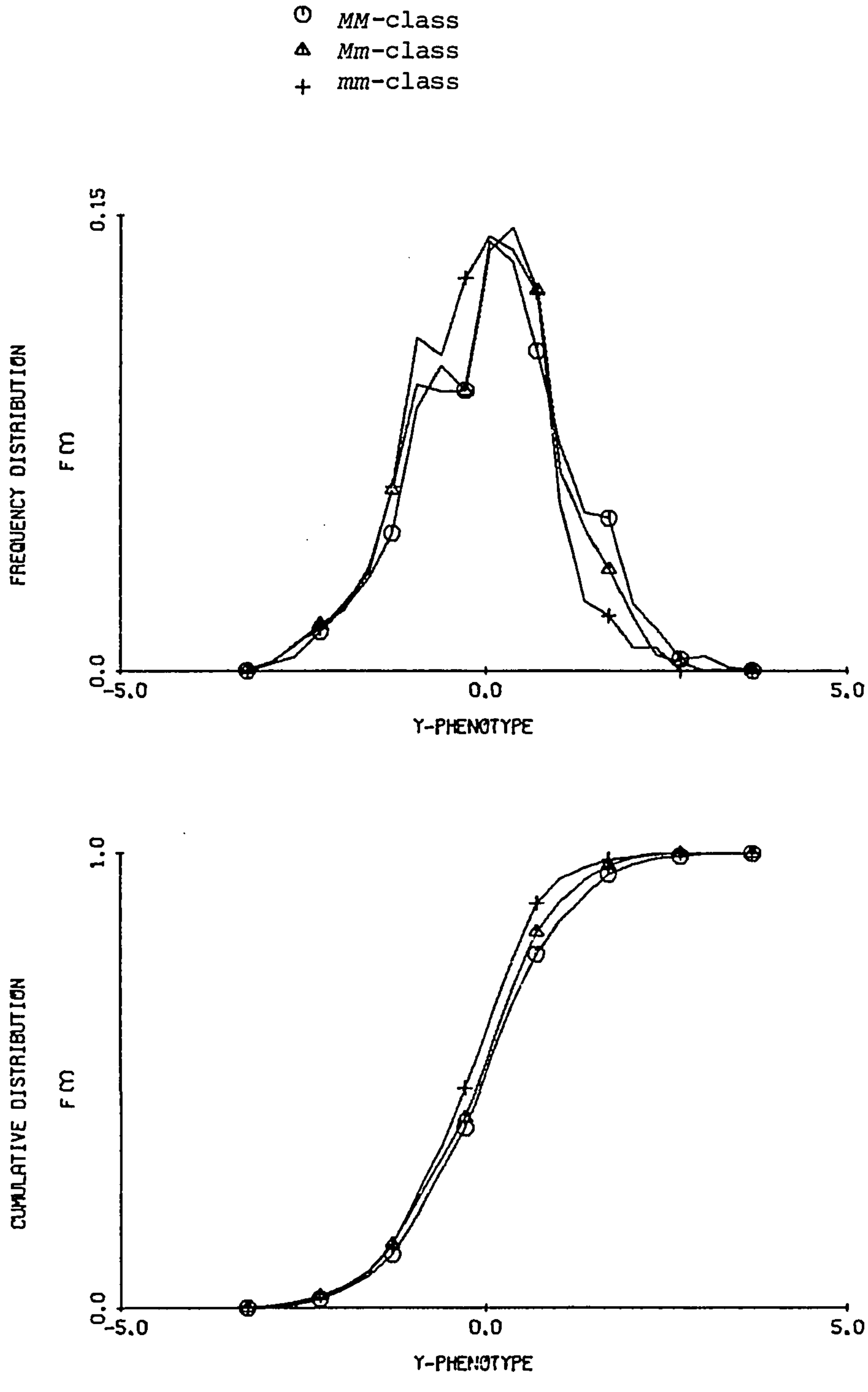


Figure 4.4.25 Subclass Distributions (data 9-2)
Parameters: $\mu_y=0$, $\sigma_y^2=1$, $h^2=.8$, $x=0$, $n=2$, $N=5000$, $D^*=.5$

4.5 General Discussion

A model for quantitative traits employed in the present study had several advantages over other models discussed by others. The advantages were:

(1) gene frequency at each locus was not assumed to be the same;

(2) additive or dominant genetic effects were considered; and

(3) the distribution of phenotypic values was not assumed to be normally distributed.

Although some assumptions made in Section 3.1 may not hold for certain data, it was necessary in order to investigate a phenotypic distribution of quantitative traits at a basic level. In future studies, factors such as genetic-environmental interactions, epistatic effects among loci concerned with a quantitative trait and multiple alleles at each locus, should be added into the phenotypic distribution discussed in this study.

The results of simulation studies have shown that it was possible to establish a genetic basis for the character from a random sample of individual data. Since the phenotypic distribution was a function of genotypic values, number of loci, gene frequencies, and mean and variance of environmental effects, these parameters were possible to estimate. In the first phase of this study, the method of maximum likelihood was employed by using an arbitrary set of initial values for obtaining the maximum likelihood estimates. However, it was found that in obtaining estimates, the results of maximum likelihood estimation depended upon the initial values. An attempt was made to solve for unknown parameters by using higher moments but this failed to give a good set of initial values. Another procedure which is called the Powell-Fletsher search procedure has been suggested by Professor H. D'angelo (Tan, 1976, Personal Communication). This method is independent of initial values and hence, will be applied to this problem in the future. The Powell-Fletsher method is an alternative approach to obtain maximum likelihood value. It was an advantage of simulation study that the results of estimation were examined by comparing with the parameter values. The results were good and consistent when the heritability of a trait was high. The disagreements of some estimates with parameter values were presumably due to: random errors associated with the generation of data set; smaller sample sizes; and larger effects of environment on the trait.

The methods of detecting association between a quantitative trait and a genetic marker and the estimation of the degree of linkage disequilibrium employed also had advantages over the methods proposed by others. Apart from the advantages for the model of quantitative traits mentioned above, there were other advantages such as:

(1) family data were unnecessary; and

(2) assumption of complete linkage between a quantitative locus and a genetic marker locus was not needed.

Such factors as epistatic effects between a quantitative locus and a genetic marker can be considered by introducing these parameters into the subclass distributions.

Although genetic information can be obtained from a random sample by using the exact distributions of phenotypic values, the present procedure should be used in conjunction with other analytic methods in quantitative genetics when additional information from family data is available. The model would be more realistic when other parameters such as genetic-environment interactions and epistatic effects are taken into consideration.

Appendix I

Computational Formula for $W_{uv}^{(n)}$

$$W_{uv}^{(n)} = \Sigma\{\prod_{i=1}^{n} (p_i^2)^{\alpha_i} (2p_iq_i)^{\beta_i} (q_i)^{\gamma_i}$$

where Σ is overall non-negative values of α_i, β_i, and γ_i satisfying $\alpha_i + \beta_i + \gamma_i = 1$ for $i = 1$ to n, $\sum_{i=1}^{n}\alpha_i = u$ and $\sum_{i=1}^{n}\beta_i = v$.

$$W_{un}^{(n)} = q_n^2 W_{uv}^{(n-1)} + p_n^2 W_{(u-1)v}^{(n-1)} + 2p_nq_nW_{u(v-1)}^{(n-1)}$$

where $W_{(-1)v}^{(n)} = W_{u(-1)}^{(n)} = 0$

$$W_{uv}^{(n)} = 0 \text{ when } u + v > n$$

$$W_{00}^{(1)} = q_1^2$$

$$W_{01}^{(1)} = 2p_1q_1$$

$$W_{10}^{(1)} = p_1^2$$

Appendix II

Mean and Variance of Exact Distribution

$$f(y) = \sum_{u=0}^{n} \sum_{v=0}^{n-u} W_{uv}^{(n)} \frac{1}{\sqrt{2\pi\sigma_e^2}} exp\{- \frac{1}{2\sigma_e^2} (y - \mu_e - au - dv + a(n - u - v)^2\}$$

Let $z = y - A$

Where $A = \mu_e + au + dv - a(n - u - v)$

$$\mu_y = E(y) = \int_{-\infty}^{\infty} yf(y)\ dy$$

$$= \sum_{u=0}^{n} \sum_{v=0}^{n-u} W_{uv}^{(n)} \int_{-\infty}^{\infty} (z + A) \frac{1}{\sqrt{2\pi\sigma_e^2}} exp(\frac{-1}{2\sigma_e^2} z^2) d_z$$

$$= \sum_{u=0}^{n} \sum_{v=0}^{n-u} W_{uv}^{(n)} (0 + A)$$

$$= \sum_{u=0}^{n} \sum_{v=0}^{n-u} W_{uv}^{(n)} (\mu_e + (2u - n + v)a + vd)$$

$$= \mu_e + a \sum_{u=0}^{n} \sum_{v=0}^{n-u} W_{uv}^{(n)} (2u - n + v) + d \sum_{u=0}^{n} \sum_{v=0}^{n-u} W_{uv}^{(n)} v$$

$$[\sum_{u=0}^{n} \sum_{v=0}^{n-u} W_{uv}^{(n)} = 1]$$

$$= \mu_e + a\ E(2u - n + v) + d\ E(v)$$

$$= \mu_e + a \sum_{i=1}^{n} (2p_i^2 - 1 + 2p_i q_i) + d \sum_{i=1}^{n} 2p_i q_i$$

$$\begin{bmatrix} E(2u) = 2\ E(u) = 2 \sum_{i=1}^{n} p_i^2 \\ E(-n) = -n = \sum_{i=1}^{n} -1 \\ E(v) = \sum_{i=1}^{n} 2p_i q_i \end{bmatrix}$$

$$= \mu_e + a\sum_{i=1}^{n}(p_i - q_i) + 2d\sum_{i=1}^{n}p_iq_i$$

$$= \mu_e + \mu_a + \mu_d$$

$$= \mu_e + \mu_g$$

$$\{E(y)\}^2 = \{\mu_e + a\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}(2u - n + v) + d\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}v\}^2$$

$$= \mu_e^2 + a^2\{\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}(2u - n + v)\}^2 + d^2\{\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}v\}^2$$

$$+ 2\mu_e a\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}(2u - n + v) + 2\mu_e d\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}v$$

$$+ 2ad\{\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}(2u - n + v)\}\{\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}v\}$$

$$E(y^2) = \int_{-\infty}^{\infty} y^2 f(y)\, dy$$

$$= \sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}\int_{-\infty}^{\infty}(z^2 + 2Az + A^2)\frac{1}{\sqrt{2\pi\sigma_e^2}}exp(-\frac{1}{2\sigma_e^2}z^2)dz$$

$$= \sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}(\sigma_e^2 + 0 + A^2)$$

$$= \sigma_e^2 + \sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}(\mu_e + (2u - n + v)a + vd)^2$$

$$= \sigma_e^2 + \mu_e^2 + a^2\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}(2u - n + v)^2 + d^2\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}v^2$$

$$+ 2\mu_e a\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}(2u - n + v) + 2\mu_e d\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}v$$

$$+ 2ad\sum_{u=0}^{n}\sum_{v=0}^{n-u}W_{uv}^{(n)}(2u - n + v)v$$

$$\sigma_y^2 = E(y^2) - \{E(y)\}^2$$

$$= \sigma_e^2 + a^2\Big[\sum_{u=0}^{n}\sum_{v=0}^{n-u} W_{uv}^{(n)}\,(2u - n + v)^2 - \Big\{\sum_{u=0}^{n}\sum_{v=0}^{n-u} W_{uv}^{(n)}\,(2u - n + v)\Big\}^2\Big]$$

$$+ d^2\Big[\sum_{u=0}^{n}\sum_{v=0}^{n-u} W_{uv}^{(n)}\,v^2 - \Big\{\sum_{u=0}^{n}\sum_{v=0}^{n-u} W_{uv}^{(n)}\,v\Big\}^2\Big]$$

$$+ 2ad\Big[\sum_{u=0}^{n}\sum_{v=0}^{n-u} W_{uv}^{(n)}\,(2u - n + v)v$$

$$- \Big\{\sum_{u=0}^{n}\sum_{v=0}^{n-u} W_{uv}^{(n)}\,(2u - n + v)\Big\}\Big\{\sum_{u=0}^{n}\sum_{v=0}^{n-u} W_{uv}^{(n)}\,v\Big\}\Big]$$

$$= \sigma_e^2 + Var\,\{(2u - n + v)a + dv\}$$

$$= \sigma_e^2 + Var\,(2ua) + Var\{(a + d)v\} + 2Cov\,\{2ua,\ (a + d)v\}$$

$$= \sigma_e^2 + 4a^2 Var(u) + (a + d)^2\,Var(v) + 4a(a + d)Cov\,(u,\ v)$$

$$= \sigma_e^2 + 4a^2\sum_{i=1}^{n} p_i^2 q_i^2 + (a + d)^2\sum_{i=1}^{n} 2p_i q_i(1 - 2p_i q_i) - 8a(a + d)\sum_{i=1}^{n} p_i^3 q_i$$

$$\left[\begin{array}{l} Var(u) = \sum_{i=1}^{n} p_i^2 q_i^2 \\ Var(v) = \sum_{i=1}^{n} 2p_i q_i(1 - 2p_i q_i) \\ Cov(u,v) = -\sum_{i=1}^{n} p_i^2\ 2p_i q_i = -2\sum_{i=1}^{n} p_i^3 q_i \end{array}\right]$$

$$= \sigma_e^2 + a^2\sum_{i=1}^{n}(4p_i^2 - 4p_i^4 + 2p_i q_i - 4p_i^2 q_i^2 - 8p_i^3 q_i)$$

$$+ 4ad\sum_{i=1}^{n}(p_i q_i - 2p_i^2 q_i^2 - 2p_i^3 q_i) + d^2\sum_{i=1}^{n}(1 - 2p_i q_i)\ (2p_i q_i)$$

$$= \sigma_e^2 + a^2\sum_{i=1}^{n} 2p_i q_i + 4ad\sum_{i=1}^{n} p_i q_i\ (q_i - p_i) + d^2\sum_{i=1}^{n} 2p_i q_i(p_i - q_i)^2$$

$$+ d^2\sum_{i=1}^{n}(2p_i q_i)^2$$

$$= \sigma_e^2 + \sum_{i=1}^{n} 2p_i q_i \{a + d(q_i - p_i)\}^2 + 4d^2 \sum_{i=1}^{n} p_i^2 q_i^2$$

$$= \sigma_e^2 + \sigma_a^2 + \sigma_d^2$$

$$= \sigma_e^2 + \sigma_g^2$$

$$\therefore \mu_y = \mu_g + \mu_e$$

$$= a\sum_{i=1}^{n} (p_i - q_i) + 2d\sum_{i=1}^{n} p_i q_i + \mu_e$$

$$\sigma_y^2 = \sigma_g^2 + \sigma_e^2$$

$$= \sum_{i=1}^{n} 2p_i q_i \{a + d(q_i - p_i)\}^2 + 4d^2 \sum_{i=1}^{n} p_i^2 q_i^2 + \sigma_e^2$$

Appendix III

Estimation of a and d given σ_g^2, x and p_i for $i = 1$ to n

$$\sigma_g^2 = \sum_{i=1}^{n} 2p_iq_i\{a + d(q_i - p_i)\}^2 + d^2 \sum_{i=1}^{n} (2p_iq_i)^2$$

$$= \sum_{i=1}^{n} 2p_iq_i\{a + ax(q_i - p_i)\}^2 + a^2x^2 \, \Sigma(2p_iq_i)^2$$

$$[x = \frac{d}{a} \longrightarrow d = ax]$$

$$= a^2 \left[\sum_{i=1}^{n} 2p_iq_i\{1 + x(q_i - p_i)\}^2 + x^2 \sum_{i=1}^{n} (2p_iq_i)^2\right]$$

$$\therefore \; a = \frac{\sigma_g}{\sqrt{\sum_{i=1}^{n} 2p_iq_i\{1 + x(q_i - p_i)\}^2 + x^2 \sum_{i=1}^{n} (2p_iq_i)^2}}$$

$$d = ax$$

Appendix IV

Initial Estimate of Degree of Dominance (D)

Let the probability density function of phenotypic values be:

$$f(y) = p_m^2 \; E_1(y) + 2p_m q_m \, E_2(y) + q_m^2 \, E_3(y)$$

where

$$E_1(y) = aB_1 + dB_2 + 2a$$

$$E_2(y) = aB_1 + dB_2 + (a + d)$$

$$E_3(y) = aB_1 + dB_2$$

$$B_1 = 2\sum_{i=1}^{n-1} p_i^2 + \sum_{i=1}^{n-1} 2p_i q_i - n$$

$$B_2 = \sum_{i=1}^{n-1} 2p_i q_i$$

and genotypic values (a and d), phenotypic means (μ_{MM}, and μ_{mm}) and gene frequencies (p_i for $i = 1$ to n and p_m) are known. Then

$$p_m^2 \, \mu_{MM} = p_m^2 \, \mu_e + f_{11}^2 (aB_1 + dB_2 + 2a)$$

$$+ 2f_{11} f_{21} \{aB_1 + dB_2 + (a + d)\} + f_{21}^2 \, (aB_1 + dB_2)$$

$$= p_m^2 \, \mu_e + 2a \, (e_1 + e_2 D) + 2d \, (e_3 + e_4 D - D^2)$$

where

$$e_1 = \tfrac{1}{2} B_1 p_m^2 + p_n^2 p_m^2 + p_n q_n p_m^2$$

$$e_2 = p_n p_m + q_n p_m = p_m$$

$$e_3 = \tfrac{1}{2}B_2 p_m^2 + p_n q_n p_m^2$$

$$e_4 = -p_n p_m + q_n p_m$$

$$q_m^2\ \mu_{mm} = q_m^2(\mu_e + aB_1 + dB_2) + f_{12}^2(2a) + 2f_{12}f_{22}(a + d)$$

$$= q_m^2\ \mu_e + 2a(e_5 + e_6 D) + 2d(e_7 + e_8 D - D^2)$$

where

$$e_5 = \tfrac{1}{2}B_1 q_m^2 + p_n^2 q_m^2 + p_n q_n q_m^2$$

$$e_6 = -p_n q_m - q_n q_m = -q_m$$

$$e_7 = \tfrac{1}{2}B_2 q_m^2 + p_n q_n q_m^2$$

$$e_8 = p_n q_m - q_n q_m$$

$$f_{11} = p_n p_m + D$$

$$f_{12} = p_n q_m - D$$

$$f_{21} = q_n p_m - D$$

$$f_{22} = q_n q_m + D$$

$$p_m^2 q_m^2(\mu_{MM} - \mu_{mm})$$

$$= q_m^2\{2a(e_1 + e_2 D) + 2d(e_3 + e_4 D - D^2)\}$$

$$- p_m^2\{2a(e_5 + e_6 D) + 2d(e_7 + e_8 D - D^2)\}$$

$$= \{2a(q_m^2 e_1 - p_m^2 e_5) + 2d(q_m^2 e_3 - p_m^2 e_7)\}$$

$$+ \{2a(q_m^2 e_2 - p_m^2 e_6) + 2d(q_m^2 e_4 - p_m^2 e_8)\}D$$

$$+ \{-2q_m^2 d + 2p_m^2 d\} D^2$$

$$c_1 + c_2 D + c_3 D^2 = 0$$

where

$$c_1 = \{2a(q_m^2 e_1 - p_m^2 e_5) + 2d(q_m^2 e_3 - p_m^2 e_7)\} - p_m^2 q_m^2 (\mu_{MM} - \mu_{mm})$$

$$c_2 = \{2a(q_m^2 e_2 - p_m^2 e_6) + 2d(q_m^2 e_4 - p_m^2 e_8)\}$$

$$c_3 = 2d(p_m - q_m)$$

$$\therefore D = \frac{-c_2 \pm \sqrt{c_2^2 - 4c_1 c_3}}{2c_1}$$

Appendix V

Coefficients of Skewness and Kurtosis

Let λ_1 and λ_2 be the coefficients of skewness and kurtosis, respectively. Then:

$$\lambda_1 = nk_3/\sigma_y^3$$

$$\lambda_2 = nk_4/\sigma_y^4$$

where

$$\sigma_y^2 = n(x_2 - x_1^2) + \sigma_e^2$$

$$k_3 = x_3 - 3x_1x_2 + 2x_1^3$$

$$k_4 = x_4 - 4x_1x_3 - 3x_2^2 + 12x_2x_1^2 - 6x_1^4$$

$$x_i = \frac{1}{n} \sum_{i=1}^{n} \{p_i^2a^i + 2p_i(1 - p_i)d^i + (1 - p_i)^2 \ (-a)^i\}$$

where n is the number of loci, p_i is the gene frequency of A_{i1}, a is the genotypic value of $A_{i1}A_{i1}$, d is the genotypic value of $A_{i1}A_{i2}$, and σ_e^2 is the variance of environmental effects.

LITERATURE CITED

Armitage, P. Statistical Methods in Medical Research. New York: Wiley & Sons, 1973.

Bateson, W. and E. R. Saunders. "Experimental Studies in the Physiology of Heredity." Report to the Evolution Committee of the Royal Society, 1 (1902), 1-160.

Breese, E. L. and K. Mather. "The Organization of Polygenic Activity Within a Chromosome in Drosophila: I. Hair Characters." Heredity, 2 (1957), 373-395.

Breese, E. L. and K. Mather. "The Organization of Polygenic Activity Within a Chromosome in Drosophila: II. Viability." Heredity, 14 (1960), 375-400.

Brues, A. M. "Linkage of Body Build with Sex, Eye Color, and Freckling." American Journal of Human Genetics, 2 (1950), 215-238.

Castle, W. E. "On a Method of Estimating the Number of Genetic Factors Concerned in Cases of Blending Inheritance." Science, 54 (1921) 93-96.

Downham, D. Y. and F. D. K. Roberts. "Multiplicative Congruential Pseudo-random Number Generators." Computer Journal, 10 (1967), 74-77.

East, E. M. and H. K. Hayes. "Inheritance in Maize." Bulletin of the Connecticut Agricultural Experiment Station, New Haven, 167, (1911), 142.

East, E. M. "Studies on Size Inheritance in Nicotiana." Genetics, 1 (1916), 164-176.

Elston, R. C. and M. A. Campbell. "Schizophrenia: Evidence for the Major Gene Hypothesis." Behavior Genetics, 1 (1970), 3-10.

Elston, R. C. and J. Stewart. "A General Model for the Genetic Analysis of Pedigree Data." Human Heredity, 21 (1971), 523-542.

Falconer, D. S. Introduction to Quantitative Genetics. New York: Ronald, 1960.

Fisher, R. A. "The Correlation Between Relatives on the Supposition of Mendelian Inheritance." Transactions of the Royal Society of Edinburgh, 52 (1918), 399-433.

Galton, F. Natural Inheritance. London: MacMillan, 1889.

Haseman, J. K. and R. C. Elston. "The Investigation of Linkage Between a Quantitative Trait and a Marker Locus." Behavior Genetics, 2 (1972), 3-19.

Hill, A. P. "Quantitative Linkage: A Statistical Procedure for its Detection and Estimation." Annals of Human Genetics, 38 (1975), 439-449.

Jayakar, S. D. "On the Detection and Estimation of Linkage Between a Locus Influencing a Quantitative Character and a Marker Locus." Biometrics, 26 (1970), 451-464.

Johannsen, W. Elemente der exakten Erblichkeitslehre. Gustar Fischer: Jena, 1909.

Kaplan, E. B. and R. C. Elston. A Subroutine Package for Maximum Likelihood Estimation (MAXLIK). North Carolina: University of North Carolina, 1972.

Koshal, R. S. "Application of the Method of Maximum Likelihood to the Improvement of Curves Fitted by the Method of Moments." Journal of the Royal Statistical Society A, 96 (1933), 303-313.

Lange, E., M. A. Spence and M. B. Frank. "Application of the Lod Method to the Detection of Linkage Between a Quantitative Trait and a Qualitative Marker: A Simulation Experiment." American Journal of Human Genetics, 28 (1976), 1967-173.

Lewis, P. A. W., A. S. Goodman and J. M. Miller. "A Pseudo-random Number Generator for the System/360." IBM System Journal, 8 (1969), 136-147.

Lowry, D. and F. T. Shultz. "Testing Association of Metric Traits and Marker Genes." Annals of Human Genetics, 23 (1959), 83-90.

Lush, J. L. Animal Breeding Plans. Ames: Iowa State College Press, 1945.

MacLaren, M. D. and G. Marsaglia. "Uniform Random Number Generators." Journal of the Association for Computing Machinery, 12 (1965), 83-89.

Mather, K. Biometric Genetics. New York: Dover, 1949.

Mather, K. and J. L. Jinks. Biometrical Genetics. New York: Cornell University, 1971.

Morton, N. E. "Sequential Test for the Detection of Linkage." American Journal of Human Genetics, 7 (1955), 277-318.

Morton, N. E. "The Detection of Major Genes Under Additive Continuous Variation." American Journal of Human Genetics, 19 (1967), 23-34.

Morton, N. E. and C. J. MacLean. "Analysis of Family Resemblance. III. Complex Segregation of Quantitative Traits." American Journal of Human Genetics, 26 (1974), 489-503.

Naylor, T. H., J. L. Balintfy, D. S. Burdick and K. Chu. Computer Simulation Techniques. New York: Wiley & Sons, 1966.

Neimann-Sorenson, A. and A. Robertson. "The Association Between Blood Groups and Several Production Characteristics in Three Danish Cattle Breeds." Acta Agriculture Scandinavica, 11 (1961), 163-196.

Nilsson-Ehle, H. "Kreuzungsuntersuchungen an Hafer and Weizen." Lunds. Univ. Aarskr. NF, 5 (1909), 1-122.

O'Donald, P. "The Distribution of Genotypes Produced by Segregating at a Number of Loci." Heredity, 26 (1971), 23-241.

Pearson, K. "Mathematical Contribution to the Theory of Evolution. On the Law of Ancestral Heredity." Proceedings of the Royal Society, 62 (1898), 386-412.

Pearson, K. "A Mendelian's View of the Law of Ancestral Inheritance." Biometrika, 3 (1904), 109-112.

Penrose, L. S. "The Detection of Autosomal Linkage in Data which Consist of Pairs of Brothers and Sisters of Unspecified Parentage." Annals of Eugenics, 6 (1935), 133-138.

Penrose, L. S. "Genetic Linkage in Graded Human Characters." Annals of Eugenics, 8 (1938), 233-237.

Penrose, L. S. "A Further Note on the Sib-pair Linkage Method." Annals of Eugenics, 13 (1946), 25-29.

Robertson, A. "Linkage Between Marker Loci and those Affecting a Quantitative Trait." Behavior Genetics, 3 (1973), 389-391.

Robson, D. S. "Applications of the K_4 Statistics to Genetic Variance Component Analysis." Biometrics, 12 (1956), 433-444.

Sax, K. "The Association of Size Differences with Seed-Coat Pattern and Pigmentation in *Phaseolus Vulgaris*." Genetics, 8 (1923), 552-560.

Shenton, L. R. "Maximum Likelihood and the Efficiency of the Method of Moments." Biometrika, 37 (1970), 111-116.

Siegel, S. Non Parametric Statistics: For the Behavioral Sciences. New York: McGraw-Hill, 1956.

Smith, C. A. B. "A Non-parametric Test for Linkage with a Quantitative Character." Annals of Human Genetics, 38 (1975), 451-460.

Snedecor, G. W. Statistical Methods. Ames: Iowa State University, 1956.

Sokal, R. R. and F. J. Rohlf. Biometry. The Principles and Practice of Statistics in Biological Research. San Francisco: Freeman, 1969.

Spence, M. A., R. C. Elston, K. K. Namboodiri and W. S. Pollitzer. "Evidence for a Possible Major Gene Effect in Absolute Finger Ridge Count." Human Heredity, 23 (1973), 414-421.

"Student." "A Calculation of the Minimum Number of Genes in Winter's Selection Experiment." Annals of Human Genetics, 6 (1934), 77-82.

Tan, W. Y. "Some Distribution Theories in Quantitative Genetics." Paper read before the Conference of Biomathematics and Statistics, Washington State University, May 1975.

Tan, W. Y. and W. C. Chang. "Convolution Approach to the Genetics Analysis of Quantitative Characters of Self-fertilized Population." Biometrics, 28 (1972a), 1073-1090.

Tan, W. Y. and W. C. Chang. "Some Comparisons of the Method of Moments and the Method of Maximum Likelihood in Estimating Parameters of a Mixture of Two Normal Densities." The Journal of the American Statistical Association, 67 (1972b), 702-708.

Thoday, J. M. "Location of Polygenes." Nature, 191 (1961), 368-370.

Thoday, J. M. "New Insights into Continuous Variation." Proceedings of the Third International Congress of Human Genetics (1967), 339-350.

Whitehouse, H. L. K. Towards an Understanding of the Mechanism of Heredity. New York: St. Martin's, 1965.

Wright, S. "Systems of Mating." Genetics, 6 (1921), 111-178.

Wright, S. "The Genetics of Quantitative Variability" in Quantitative Inheritance, pp. 5-41. Agricultural Research Council, London: Her Majesty's Stationery Office, 1952.

Wright, S. Evolution and the Genetics of Population. Vol. I. Genetic and Biometric Foundation, 1968.