Theor Appl Genet (1994) 88:573–580

R. L. Fernando · C. Stricker · R. C. Elston

The finite polygenic mixed model: an alternative formulation for the mixed model of inheritance

Received: 14 June 1993 / Accepted: 27 September 1993

Abstract This paper presents a mixed model of inheritance with a finite number of polygenic loci. This model leads to a likelihood that can be calculated using efficient algorithms developed for oligogenic models. For comparison, likelihood profiles were obtained for the finite polygenic mixed model, the usual mixed model, with exact and approximate calculations, and for a class D regressive model. The profiles for the finite polygenic mixed model were closest to the profiles for the usual mixed model with exact calculations.

Key words Mixed model inheritance · Likelihood

Introduction

Maximum likelihood is a powerful method for testing hypotheses and estimating parameters. Thus, the development of fast algorithms for calculating likelihoods for pedigree data has been an important area of research in genetics.

When a trait is influenced by a few loci, inheritance of this trait is said to be oligogenic. Under oligogenic inheritance, the phenotypic values of pedigree members are assumed to be conditionally independent, given the genotypes of these pedigree members. Further, the genotype of an individual is conditionally independent of the genotypes of all ancestors and sibs, given the genotypes of the parents. Because of these properties of the phenotypic and genotypic distributions under oligogenic inheritance, it has been possible to develop fast algorithms to calculate the likelihood (Elston and Stewart 1971; Lange and Elston 1975; Cannings et al. 1978; Lange and Boehnke 1983; Goradia et al. 1992).

When a trait is influenced by a large number of loci, each with a small effect, inheritance is said to be polygenic. The genotypic value of polygenic traits is often assumed to have a normal distribution (Fisher 1918; Elston and Stewart 1971; Bulmer 1980). Under this assumption, algorithms used for oligogenic inheritance can be adapted to calculate the likelihood (Elston and Stewart 1971; Elston et al. 1992). Also, by assuming a finite number of polygenic loci, the likelihood can be calculated by direct application of oligogenic algorithms (Thompson and Skolnick 1977). Further, under the assumption of normality, the likelihood can be calculated using Henderson's mixed model equations for very large and complex pedigrees (Henderson 1984; Meyer 1989).

Calculation of the likelihood has also been discussed for traits influenced by both a single major locus with a large effect and a normally distributed polygenic component (Elston and Stewart 1971; Morton and MacLean 1974). These traits are said to have mixed major gene and polygenic inheritance (Morton and MacLean 1974; Elston 1990). From here on, this model of inheritance will be referred to as the mixed model of inheritance, or simply the mixed model, which should not be confused with the mixed linear model.

Under the mixed model, the phenotypic values of pedigree members cannot be assumed to be conditionally independent, given only the major genotypes of these pedigree members, because the phenotypic value is also influenced by the polygenic loci. Thus, for this formulation of the mixed model, fast algorithms to calculate the exact pedigree likelihood do not exist (Elston 1990; Bonney 1992).

The objective of this paper is to present an alternative formulation of the mixed model with a finite number of polygenic loci, leading to a likelihood that can be computed using fast algorithms that have been developed for oligogenic traits. Furthermore, certain aspects of the

Communicated by E. J. Eisen

R. L. Fernando (✉)
Department of Animal Sciences, University of Illinois at Urbana-Champaign, 1207 West Gregory Drive, Urbana, IL 61801 USA

C. Stricker · R. C. Elston
Department of Biometry and Genetics and the Center for Molecular and Human Genetics, Louisiana State University, Medical Center, 1901 Perdido Street, New Orleans 70112-1393, USA

resulting model may be more realistic in some practical situations.

An alternative formulation of the mixed model

In most statistical models, the assumption of a normal distribution is mathematically convenient. In the statistical analysis of quantitative traits, the assumption of normality for the genotypic value has been justified on the basis of the trait being controlled by a large number of polygenic loci, each with a small effect (Fisher 1918; Elston and Stewart 1971; Bulmer 1980). However, the assumption of normality for the polygenic component of the genotypic value in the usual mixed model does not result in a likelihood that is easy to compute. To understand better this difficulty in computing the likelihood for the usual mixed model of inheritance, it is instructive to consider first computing the likelihood for an oligogenic model.

The probability density of the phenotypic values, expressed as a function of the unknown parameters of the density, is the likelihood of a pedigree. Under oligogenic inheritance, phenotypic values are assumed to be conditionally independent, given the genotypes. Thus, the conditional density of the phenotypic values given the genotypes can be written as

$$\Pr(\mathbf{y}|\mathbf{g}) = \prod_{i=1}^{n} \Pr(y_i|g_i), \tag{1}$$

where $\mathbf{y}$ is the vector of n phenotypic values, $\mathbf{g}$ is the vector of n genotypes for the pedigree members, and $\Pr(y_i|g_i)$ is the penetrance function or the conditional density of the phenotypic value given the genotype. Under Mendelian inheritance, the probability of the genotypes can be written as

$$\Pr(\mathbf{g}) = \prod_{i=1}^{n_1} \Pr(g_i) \prod_{i=n_1+1}^{n} \Pr(g_i|g_m, g_f), \tag{2}$$

where pedigree members 1 through n_1 are founders and the rest are non-founders, $\Pr(g_i)$ is the population frequency of genotype g_i, and $\Pr(g_i|g_m, g_f)$ is the transition probability or the conditional probability that a child will have genotype g_i given the parents have genotypes g_m and g_f (Elston and Stewart 1971; Bonney 1984). Now the likelihood of the pedigree can be written as

$$\Pr(\mathbf{y}) = \sum_{g_1}\sum_{g_2}\cdots\sum_{g_n}\prod_{i=1}^{n}\Pr(y_i|g_i) \times \prod_{i=1}^{n_1}\Pr(g_i)\prod_{i=n_1+1}^{n}\Pr(g_i|g_m,g_f). \tag{3}$$

The summations in Eq. 3 are over the genotypes. For founders, let $f(g_i) = \Pr(y_i|g_i)\Pr(g_i)$, and for non-founders, let $h(g_i, g_m, g_f) = \Pr(y_i|g_i)\Pr(g_i|g_m, g_f)$. Then, the likelihood can be written as

$$\Pr(\mathbf{y}) = \sum_{g_1}\sum_{g_2}\cdots\sum_{g_n}\prod_{i=1}^{n_1} f(g_i)\prod_{i=n_1+1}^{n} h(g_i, g_m, g_f). \tag{4}$$

If the summations are over m genotypes, the number of calculations required to compute the likelihood as indicated by Eq. 4 is proportional to m^n. However, because the function $f(g_i)$ involves only the genotype of a founder and the function $h(g_i, g_m, g_f)$ involves only the genotypes of a non-founder and parents m and f, the order of adding and multiplying in Eq. 4 can be rearranged such that the number of calculations required to compute the likelihood is proportional to n (Elston and Stewart 1971; Lange and Elston 1975; Cannings et al. 1976, 1978; Lalouel 1980; Lange and Boehnke 1983; Goradia et al. 1992).

Now consider the mixed model of inheritance. Suppose the phenotypic value of an individual can be modeled as

$$y_i = \mu_{u_i} + \mu_{v_i} + e_i, \tag{5}$$

where μ_{u_i} is the effect of the major locus, μ_{v_i} is the effect of the polygenic loci, and e_i is a residual. (More generally, μ_{u_i} can represent the sum of a major locus effect and a set of fixed effects for individual i). The effects μ_{v_i} and e_i are usually assumed to be normally distributed with null means. Further, the e_i are assumed to be identically and independently distributed. Under these assumptions, the conditional distribution of the phenotypic values, given the major genotypes, cannot be written as Eq. 1 but is

$$\Pr(\mathbf{y}|\mathbf{u}) \propto |\Sigma|^{-1/2}\prod_{i=1}^{n}\exp\left\{-\frac{\sigma^{ii}}{2}(y_i-\mu_{u_i})^2\right\} \times \prod_{i=1}^{n-1}\prod_{j=i+1}^{n}\exp\{-\sigma^{ij}(y_i-\mu_{u_i})(y_j-\mu_{u_i})\} \tag{6}$$

where $\Sigma = \mathrm{Var}(\mathbf{y}|\mathbf{u})$, $\mathbf{u}$ is a vector of major genotypes, and σ^{ij} is the ijth element of the inverse of Σ. Thus, the likelihood for the mixed model is

$$\Pr(\mathbf{y}) \propto \sum_{u_1}\sum_{u_2}\cdots\sum_{u_n}|\Sigma|^{-1/2}\prod_{i=1}^{n}\exp\left\{-\frac{\sigma^{ii}}{2}(y_i-\mu_{u_i})^2\right\} \times \prod_{i=1}^{n-1}\prod_{j=i+1}^{n}\exp\{-\sigma^{ij}(y_i-\mu_{u_i})(y_j-\mu_{u_i})\} \times \prod_{i=1}^{n_1}\Pr(u_i)\prod_{i=n_1+1}^{n}\Pr(u_i|u_m,u_f), \tag{7}$$

and because Eq. 7 cannot be written as Eq. 4, the order of adding and multiplying in Eq. 7 cannot be rearranged for efficient calculation of the likelihood (Elston 1990; Hasstedt 1991; Bonney 1992).

Therefore, consider a mixed model where the genotypic value is determined by a major locus and by a finite number (k) of unlinked polygenic loci rather than by the infinite number of polygenic loci implied by the assumption of normality for the polygenic component in the usual mixed model. Further, assume that the polygenic component of the genotypic value is additive and that the aggregate genotypic value is the sum of the genotypic values of the major locus and of the polygenic component.

With this formulation of the mixed model, the conditional distribution of the phenotypic values, given the genotypes determined by the major and polygenic loci, can be written as Eq. 2, where **g** is now the genotype determined by both the major and polygenic loci. Thus, algorithms applicable to oligogenic traits can be used to calculate the likelihood. A problem with this approach, however, is that the number of genotypes increases exponentially with the number of loci, and so does computing time. For example, suppose there are two alleles at the major locus and at each of k polygenic loci. Then, the number of genotypes that have to be summed over is $3^{(k+1)}$.

To reduce the computations in calculating the likelihood, the following assumptions and definitions are made.

1) There are only two alleles, a with effect α and b with effect β, at each of the k polygenic loci. The frequency of allele a is p at each locus.
2) The polygenic effect, μ_v, is the sum of the effects of the alleles at the polygenic loci. Thus, for an individual possessing v alleles with effect α and $(2k - v)$ alleles with effect β,

$$\mu_v = v\alpha + (2k - v)\beta.$$

The number v will be referred to as the polygenic number.
3) The polygenic number v_i of a pedigree member i is conditionally independent of the polygenic number v_j of any ancestor or sib j, given v_m and v_f, the polygenic numbers of the parents of i. So, the probability of the polygenic numbers can be written as

$$\Pr(v) = \prod_{i=1}^{n_1} \Pr(v_i) \prod_{i=n_1+1}^{n} \Pr(v_i|v_m, v_f),$$

where v is a vector of polygenic numbers. It is shown in Appendix A that this assumption is not strictly consistent with Mendelian inheritance. However, v_i and v_j are conditionally uncorrelated given v_m and v_f, as shown in Appendix B. Recursive calculation of transition probabilities, $\Pr(v_i|v_m, v_f)$, for polygenic numbers is described in Appendix C.
4) The effects α and β are such that μ_v has expected value zero and variance σ_v^2. Thus, setting

$$2k[p\alpha + (1-p)\beta] = 0$$

and

$$2k[p\alpha^2 + (1-p)\beta^2] = \sigma_v^2,$$

after some algebra,

$$\alpha = \sqrt{\frac{\sigma_v^2(1-p)}{2pk}}$$

and

$$\beta = \frac{-p\alpha}{1-p}.$$

Thus, $\alpha = -\beta$ when $p = 0.5$.

From Eq. 5 it can be seen that the phenotypic values are conditionally independent, given μ_u and μ_v, the effects of the genotypes at the major locus and the polygenic loci. Thus, under these assumptions, and because of the one-to-one correspondence between u_i and μ_{u_i}, and between v_i and μ_{v_i}, the likelihood can be written as

$$\Pr(\mathbf{y}) = \sum_{u_1}\sum_{v_1}\sum_{u_2}\sum_{v_2}\cdots\sum_{u_n}\sum_{v_n}\prod_{i=1}^{n}\Pr(y_i|u_i, v_i) \times \prod_{i=1}^{n_1}\Pr(u_i)\Pr(v_i)\prod_{i=n_1+1}^{n}\Pr(u_i|u_m, u_f) \times \Pr(v_i|v_m, v_f) \quad (8)$$

which can be rearranged as

$$\Pr(y) = \sum_{u_1}\sum_{v_1}\sum_{u_2}\sum_{v_2}\cdots\sum_{u_n}\sum_{v_n}\prod_{i=1}^{n_1} f(u_i, v_i) \prod_{i=n_1+1}^{n} h(u_i, v_i, u_m, v_m, u_f, v_f) \quad (9)$$

Further, by assumptions 1 and 2 above, μ_v can take on only one of $2k + 1$ values corresponding to the $2k + 1$ polygenic numbers. (In contrast, there are 3^k possible genotypes for the polygenic loci). Suppose, for example, there are two alleles at the major locus. Then, the number of calculations to compute Eq. 8 is equivalent to that for computing the likelihood for a monogenic model with $3(2k + 1)$ possible genotypes.

Comparison of models

The primary difference between the usual mixed model of inheritance (M_∞) (Elston and Stewart 1971; Morton and MacLean 1974) and that presented here (M_k) is in the assumed number of polygenic loci. The assumption of a normal distribution for the polygenic value in M_∞

implies an infinite number of polygenic loci, while k polygenic loci are assumed in M_k.

As shown in Appendix B, under both M_∞ and M_k Mendelian inheritance implies that the polygenic effect of pedigree member i is conditionally uncorrelated with those of its sibs and ancestors, given the polygenic effects of the parents of i. As polygenic effects are normally distributed under M_∞, the null conditional correlation implies that the polygenic effect of i is conditionally independent of the polygenic effects of its ancestors and sibs, given the polygenic effects of the parents of i. However, as polygenic effects do not have a normal distribution under M_k, the null conditional covariance does not necessarily imply conditional independence. The counterexample in Appendix A demonstrates this lack of conditional independence of polygenic effects between full sibs given the polygenic effects of their parents. However, it is assumed for M_k that the polygenic effect of i is conditionally independent of the polygenic effects of its ancestors and sibs, given the polygenic effects of the parents of i, to enable fast calculation of likelihoods by the Elston-Stewart algorithm (Elston and Stewart 1971) and other algorithms based on it (Cannings et al. 1978; Lange and Boehnke 1983; Fernando et al. 1993).

Another difference between M_∞ and M_k is in the conditional variance of the polygenic effect in an offspring given the polygenic effects of the parents. Under M_∞, this conditional variance is constant in a random mating population, regardless of the values of the polygenic effects of the parents. However, under M_k this conditional variance is lower with extreme polygenic values for the parents compared to that when the parents have intermediate polygenic values. This occurs because an extreme value of the polygenic effect implies homozygozity at most of the polygenic loci, and in this respect M_k is a more realistic model than M_∞.

Comparison of likelihood profiles and parameter estimates

Regressive models (Bonney 1984; Bonney 1992) are now commonly used for data analysis as an alternative to the mixed model. Under the simplest of the regressive models, class A, the likelihood can be written as Eq. 4, and so can be calculated very efficiently. However, this is often not a good approximation to mixed inheritance (Demenais and Bonney 1989; Konigsberg et al. 1989). Class D regressive models subsume mixed inheritance as a special case when the data consist of only two generations, but are somewhat different for multigenerational data; they assume that the correlation between the components μ_{v_i} of Eq. 5 for r'th degree unilineal relatives equals the parent-offspring correlation raised to the r'th power (instead of the parent-offspring correlation times $(1/2)^{r-1}$). This makes computation of a large pedigree likelihood feasible provided no single sibship is too large. Furthermore, it is possible to approximate the likelihood of a class D regressive model so that it can be written as Eq. 4 (Demenais et al. 1990), which is implemented in the program package S.A.G.E. (1992). Another approach used to analyze data under mixed inheritance is to approximate Eq. 6 such that the likelihood can be written as Eq. 4 (Hasstedt 1982, 1991). Such an approximation has been implemented in the Pedigree Analysis Package (PAP; Hasstedt 1989).

A large sample of data was simulated under M_∞ to compare these approaches. The data consisted of 500 identical families (Fig. 1) of 12 members each. The major locus was simulated to have two additive alleles, c and d, with effects 0 and 10, and a frequency q of 0.5 for each. The polygenic value was simulated to have a variance (σ_v^2) of five, and the residual was simulated to have a variance (σ^2) of seven.

For comparison, likelihood profiles (Figs. 2–4) were obtained assuming:

1) a class D regressive model, with exact calculation of the likelihood, assuming zero spouse correlation and equal sibling and parent-offspring correlations;
2) M_∞ with exact calculation of the likelihood;

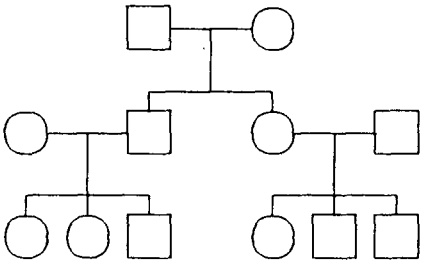
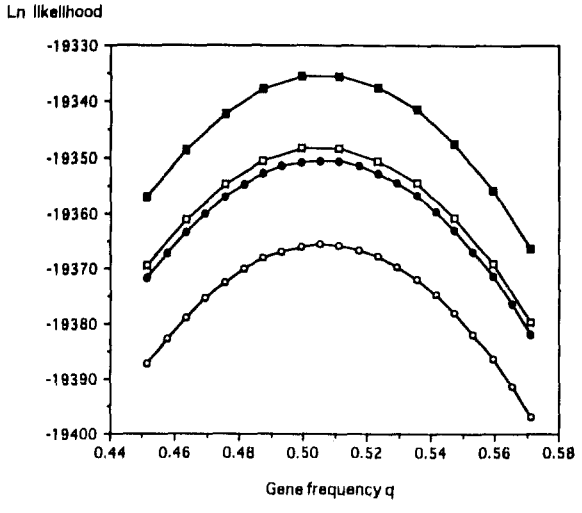


Fig. 1 Pedigree of a family. The simulated data consisted of 500 such independent families

Fig. 2 Likelihood profiles for the gene frequency. ○ Class D regressive model □ M_∞ ■ M_∞ with approximate calculation, ● M_k



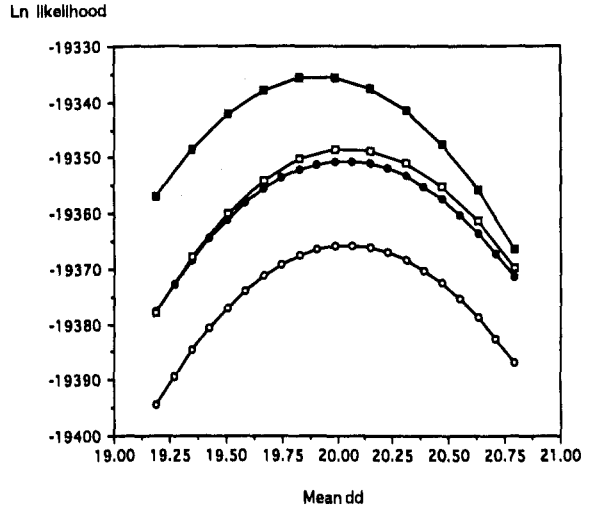


Fig. 3 Likelihood profiles for the mean of major genotype *dd* ○ Class D regressive model, □ M_∞, ■ M_∞ with approximate calculation, ● M_k

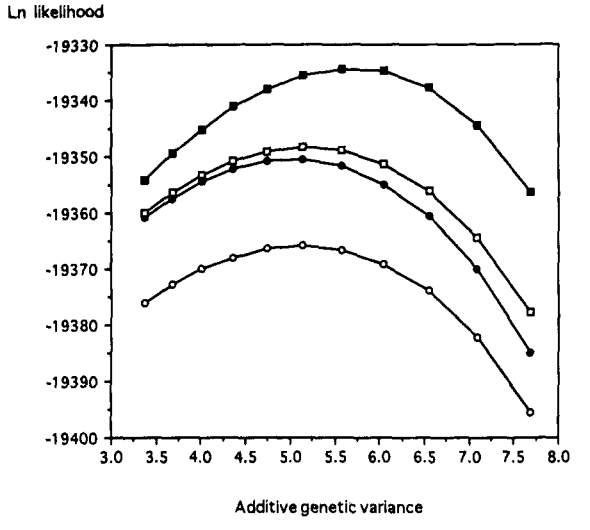


Fig. 4 Likelihood profiles for the additive genetic variance, $(\sigma_v^2 + \sigma^2)$ held constant. ○ Class D regressive model, □ M_∞, ■ M_∞ with approximate calculation, ● M_k

3) M_∞ with approximate calculation of the likelihood using PAP (Hassted 1989); and
4) M_k with exact calculation of the likelihood for $k = 5$ and $p = 0.5$ (gene frequency for the polygenic loci).

Profiles were obtained for the gene frequency at the major locus, the mean of the major genotype *dd*, and for the additive genetic variance. Maximization of the exact likelihood for M_∞ was computationally not feasible, therefore the profiles for a specific parameter were computed instead by fixing all other parameters at their maximum likelihood estimates under M_k for $k = 5$ and varying the parameter under consideration; in the case of the additive genetic variance, the total variance $(\sigma_v^2 + \sigma^2)$ was held fixed.

In each of the Figs. 2–4, the profile for M_k was closest to that for exact M_∞. The approximate likelihood values for M_∞ from PAP were the highest, and the exact likelihood values for the class D regressive model were the lowest. The exact likelihood values for M_k were very close to, but lower than, the exact values for M_∞.

Thus, among the exact likelihood values, that for M_∞ was highest. This is to be expected as the data were generated under M_∞. The likelihood values for M_k were closer to those for M_∞ than were those for the class D regressive model. This indicates that M_k more closely resembles M_∞ than does the class D regressive model. The likelihood values from PAP were approximations for M_∞, but those for M_k were closer than the PAP values to the exact likelihood values for M_∞. This indicates that the exact likelihood for M_k better approximates the likelihood for M_∞ than does the PAP approximation.

Table 1 shows the maximum likelihood estimates obtained by the different approaches. The variance parameters σ_v^2 and σ^2 were obtained for the class D regressive model by equating the residual variance to $\sigma_v^2 + \sigma^2$ and the parent-offspring correlation to $\sigma_v^2/2(\sigma_v^2 + \sigma^2)$. PAP provides estimates of $(\sigma_v^2 + \sigma^2)$ and the polygenic heritability $h_p^2 = \sigma_v^2(\sigma_v^2 + \sigma^2)$. By fitting a quadratic to the four highest ln likelihoods in each profile to estimate the second derivative, the standard errors in Table 2 were estimated. Both the parameter estimates and the standard errors were very similar for all of the models compared. It must be remembered that except under M_∞, the true model, the standard error estimates may not be consistent. Thus, the small differences in the standard errors should not be interpreted as indicating which method of analysis leads to the most efficient estimates.

Discussion

The finite polygenic mixed model (M_k) described here has the advantage that its likelihood can be calculated using efficient algorithms developed for oligogenic models. Similar algorithms (van Arendonk et al. 1989; Fernando et al. 1993) can also be used to calculate the posterior probabilities of genotypes at the major locus and at the polygenic loci. One can thus calculate the posterior mean of the aggregate genotypic value, which is widely used in animal breeding to make selection decisions. Also, compared to the traditional mixed model (M_∞), M_k has some genetically desirable properties such as heterogeneous within-family variances, the accommodation (by letting $p \neq 0.5$) of a non-symmetric distribution for the polygenic effect, and the possibility of estimating k, the number of polygenic loci involved.

One possible criticism of M_k could be that it assumes the polygenic effect of an individual to be conditionally

Table 1 Maximum likelihood estimates and ln likelihoods

Parameter	Generating value	M_k	Regular model class D	PAP
q	0.5	0.5114	0.4946	0.5064
μ_{dd}	20	19.9884	20.0230	20.0712
μ_{cc}	0	0.0000	−0.0840	−0.0157
σ_v^2	5	5.1430	5.2480	5.8674
σ^2	7	6.6702	6.6528	6.0861
$\sigma_v^2 + \sigma^2$	12	11.8132	11.9008	11.9535
Ln likelihood	−19 349.07[a]	−19 350.65	−19 365.26	−19 333.65

[a] Ln likelihood for generating parameters under M_∞

Table 2 Standard errors of maximum likelihood estimates

Parameters	M_k	Regular model class D	PAP	M_∞
q	0.0083	0.0083	0.0083	0.0083
μ_{dd}	0.1158	0.1138	0.1112	0.1123
σ_v^2	0.3588	0.3010	0.3027	0.3112

independent of its ancestors and sibs, given the polygenic effects of its parents, which is not consistent with Mendelian inheritance. However, it should be remembered that this assumption, together with the assumptions of homogeneous within-family variances and a symmetric distribution for the polygenic effects, is also implied under M_∞. However, since the number of loci is always finite, none of these assumptions of M_∞ is consistent with Mendelian inheritance. Thus, in the analysis of data using M_k, one makes fewer assumptions that are inconsistent with Mendelian inheritance than when using M_∞.

The approximation to the class *D* regressive model implemented in S.A.G.E. (1992) allows fast computation of the pedigree likelihood under a model that is close to mixed inheritance. However, in the context of sex-limited traits, regressive models require further approximation to maintain their advantage of computational efficiency.

As presented here, M_k does not accommodate non-genetic covariances among relatives. Non-genetic covariances, however, can be accommodated by assuming a regressive model in addition to a finite polygenic component. The likelihood under such a model can also be calculated using algorithms developed for oligogenic models. With this approach, one can test if the residual covariances, after accounting for the major locus, are due to polygenic loci or to non-genetic causes (or both).

An advantage of using the PAP approximation over the use of M_k or regressive models for analyzing data is that PAP can accommodate a completely general structure for the non-genetic residual covariances. The main disadvantage of PAP is that it requires the inversion of a covariance matrix of order n where n is at least the number of individuals in the pedigree, which can be quite large for animal pedigrees. Also, the accuracy of the PAP approximation cannot be determined for such large pedigrees.

Acknowledgements We thank Alexa Sorant for allowing us to use an unreleased version of the computer program REGC (S.A.G.E. 1992) that calculates the exact class D regressive model likelihood. This work was supported in part by the Illinois Agricultural Experiment Station, Hatch Project 35-0345, U.S. Public Health Service research grant GM 28356 from the National Institute of Medical Sciences, resource grant RR 03655 from the Division of Research Resources, and training grant HL 07567 form the National Heart, Lung, and Blood Institute. C. Stricker was supported by the Schweizerischer Nationalfonds, Switzerland.

Appendix

A. Counterexample

This is a counter example to show that under Mendelian inheritance the polygenic numbers of full sibs are not conditionally independent given the polygenic numbers of their parents.

Consider an M_k model with $k = 2$ and $0 < p < 1$. Let i and j be two full sibs with parents m and f. From probability theory, the equality

$$\Pr(v_i = 0 | v_j = 0, v_m = 2, v_f = 0) = \Pr(v_i = 0 | v_j = 1, v_m = 2, v_f = 0) \tag{A.1}$$

must be true if polygenic numbers of the sibs, v_i and v_j, are conditionally independent given polygenic numbers, v_m and v_f, of the parents. We show below that Eq. A.1 is not true.

For parent f, $v_f = 0$ implies that f has genotype *bb* at both loci. However, for parent m, $v_m = 2$ implies that m has either genotype *ab* at both loci or *aa* at one locus and *bb* at the other. Also, if f has genotype *bb* at both loci and m is heterozygous at both loci $\Pr(v_i = 0) = 0.25$. On the other hand, if f is as before and m has genotype *aa* at one locus and *bb* at the other, $\Pr(v_i = 0) = 0$.

But, if $v_j = 0$, m cannot have genotype *aa* at either of the loci and must be heterozygous at both loci. Thus, it is easy to see that the first probability in Eq. A.1 must be equal to 0.25. However, if $v_j = 1$, there is a non-zero probability that m has genotype *aa* at one locus and *bb* at the other. Thus, the second probability in Eq. A.1 is a weighted sum of 0.25 and 0, and is less than 0.25. Thus Eq. A.1 is not true.

B. Conditional correlation with ancestors or sibs

It is shown here that Mendelian inheritance implies that the polygenic number v_i of a pedigree member i is conditionally uncorrelated with the polygenic number v_j of any ancestor or sib j, given the polygenic numbers v_m and v_f of the parents of i. Because of the linear relationship between polygenic number and polygenic effect, the conditional correlation of polygenic effects is identical to that of polygenic numbers. To show that the above conditional correlation is null, it is sufficient to show that the conditional covariance, $\mathrm{Cov}(v_i, v_j | v_m, v_f)$, is null. This conditional covariance can be written as

$$\mathrm{Cov}(v_i, v_j | v_m, v_f) = \mathrm{E}[\mathrm{Cov}(v_i, v_j | v_m, v_f, g_m^p, g_f^p)] + \mathrm{Cov}[\mathrm{E}(v_i | v_m, v_f, g_m^p, g_f^p), E(v_j | v_m, v_f, g_m^p, g_f^p)] \tag{B.1}$$

where g_m^p and g_f^p are genotypes of the parents that result in (or map to) polygenic numbers v_m and v_f. The outer expectation and covariance are taken over all such genotypes, and the inner covariance and expectations are over restricted ranges of v_i and v_j. From principles of Mendelian inheritance, it is easy to see that the conditional expectation $E(v_i|v_m, v_f, g_m^p, g_f^p)$ is a constant for all genotypes g_m^p and g_f^p that result in (or map to) v_m and v_f. Thus, the covariance of the conditional expectations in Eq. B.1 is null, and Eq. B.1 can be written as

$$\mathrm{Cov}(v_i, v_j|v_m, v_f) = \mathrm{E}[\mathrm{Cov}(v_i, v_j|v_m, v_f, g_m^p, g_f^p)]. \tag{B.2}$$

But the conditional covariance in Eq. B.2 is null for all genotypes of the parents, g_m^p and g_f^p, because given these genotypes of the parents, the genotypes g_i^p and g_j^p are independent, and the polygenic numbers v_i and v_j are functions of the genotypes g_i^p and g_j^p. It follows that the expected value of the conditional covariance in Eq. B.2 is null.

C. Recursive calculation of transition probabilities for polygenic number

Let x denote the polygenic number for loci $1, \ldots, t$, i.e., the number of a alleles at the first t polygenic loci. The polygenic number for the single locus $t+1$, denoted w, can take on one of three values 0, 1, or 2, corresponding to genotypes bb, ab, and aa. Thus, because there is a one-to-one correspondence between the polygenic number for one locus and the genotype at that locus, the transition probabilities for w are identical to the genotype transition probabilities for a locus with two alleles.

Suppose the transition probabilities for x are known. It is shown here how to compute transition probabilities for $z = x + w$, the polygenic number for loci $1, \ldots, t+1$, from the transition probabilities for x and w.

Reasoning entirely from probability theory, the transition probabilities for z can be written as

$$\begin{aligned}\Pr(z_o|z_m, z_f) = \sum_{w_m} \sum_{w_f} \sum_{w_o} & \\ & \Pr(z_o|w_o, w_m, w_f, x_m, x_f, z_m, z_f) \\ & \Pr(w_o|w_m, w_f, x_m, x_f, z_m, z_f) \\ & \Pr(w_m, w_f, x_m, x_f|z_m, z_f)\end{aligned} \tag{C.1}$$

where o, m, and f denote offspring, mother, and father, respectively, and $x_m = z_m - w_m$ and $x_f = z_f - w_f$. The first probability in Eq. C.1 can be written as

$$\Pr(z_o|w_o, w_m, w_f, x_m, x_f, z_m, z_f) = \Pr(x_o = z_o - w_o|x_m, x_f) \tag{C.2}$$

because $z_o = x_o + w_o$, x_o is independent of w_o, w_m, and w_f, and given x_m and x_f, x_o is also independent of z_m and z_f. The second probability in Eq. C.1 can be written as

$$\Pr(w_o|w_m, w_f, x_m, x_f, z_m, z_f) = \Pr(w_o|w_m, w_f) \tag{C.3}$$

because w_0 is independent of x_m and x_f, and given w_m and w_f, w_o is also independent of z_m and z_f. Assuming that the parents m and f are unrelated, the third probability in Eq. C.1 can be written as

$$\Pr(w_m, w_f, x_m, x_f|z_m, z_f) = \Pr(x_m, w_m|z_m)\Pr(x_f, w_f|z_f) \tag{C.4}$$

Subtituting Eqs. C.2, C.3, and C.4 in C.1, the transition probabilities for z can be written as

$$\begin{aligned}\Pr(z_o|z_m, z_f) = \sum_{w_m} \sum_{w_f} \sum_{w_o} & \\ & \Pr(x_o = z_o - w_o|x_m, x_f)\Pr(w_o|w_m, w_f) \\ & \Pr(x_m, w_m|z_m)\Pr(x_f, w_f|z_f)\end{aligned} \tag{C.5}$$

The first two probabilities in Eq. C.5 are transition probabilities for x and w, respectively. To calculate the third and fourth probabilities in Eq. C.5 note that x and w are independent binomial variables and that $z = x + w$. From this it follows that

$$\Pr(x, w|z) = \frac{\mathrm{Bin}_{2t,p}(x)\mathrm{Bin}_{2,p}(w)}{\sum_{i+j=z}\mathrm{Bin}_{2t,p}(i)\mathrm{Bin}_{2,p}(j)} \tag{C.6}$$

for $x + w = z$, and $\Pr(x, w|z) = 0$ otherwise, where $\mathrm{Bin}_{q,p}()$ stands for the binomial probability function with parameters q and p. The summation in the denominator of Eq. C.6 is over all values of i and j such that $i + j = z$ for integers i between 0 and $2t$ and integers j between 0 and 2.

References

Bonney GE (1984) On the statistical determination of major gene mechanisms in continuous human traits: regressive models. Am J Med Genet 18:731–749

Bonney GE (1992) Compound regressive models for family data. Hum Hered 42:28–41

Bulmer MG (1980) *The mathematical theory of quantitative genetics.* Clarendon Press, Oxford

Cannings C, Thompson EA, Skolnick MH (1976) The recursive derivation of likelihoods on complex pedigrees. Adv Appl Prob 8: 622–625

Cannings C, Thompson EA, Skolnick MH (1978) Probability functions on complex pedigrees. Adv Appl Prob 10:26–61

Demenais F, Bonney G (1989) The equivalence of the mixed regressive models for genetic analysis. I. Continous traits. Genet Epidemiol 6:597–617

Demenais F, Murigande C, Bonney G (1990) Search for faster methods for fitting the regressive models to quatitative traits. Genet Epidemiol 7:319–334

Elston RC (1990) Models for discrimination between alternative models of inheritance. In: Gianola D, Hammond K (eds). Advances in statistical methods for genetic improvement of livestock. Springer, Berlin, Heidelberg New York, pp 41–55

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

Elston RC, George VT, Severtson F (1992) The Elston-Stewart algorithm for continuous genotypes and environmental factors. Hum Hered 42:16–27

Fernando RL, Stricker C, Elston RC (1993) An efficient algorithm to compute the posterior genotypic distribution for every member of a pedigree without loops. Theor Appl Genet 87:89–93

Fisher RA (1918) The correlation between relatives on the supposition of Mendelain inheritance. Trans Roy Soc Edinburgh 52: 399–433

Goradia TM, Lange K, Miller PL, Nadkarni PM (1992) Fast computation of genetic likelihoods on human pedigree data. Hum Hered 42:42–62

Hasstedt SJ (1982) A mixed model approximation for large pedigrees. Comput Biomed Res 15:295–307

Hassted SJ (1989) PAP: pedigree analysis package, rev 3. Department of Human Genetics, University of Utah, Salt Lake City, Utah

Hasstedt SJ (1991) A variance components/major locus likelihood approximation on quantitative data. Genet Epidemiol 8: 113–125

Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Guelph, Ontario

Konigsberg L, Kammerer C, MacCluer J (1989) Segregation analysis of quatitative traits in nuclear families: comparison of three program packages. Genet Epidemiol 6:713–726

Lalouel JM (1980) Probability calculations in pedigrees under complex modes of inheritance. Hum Hered 30:320–323

Lange K, Boehnke M (1983) Extensions to pedigree analysis. V. Optimal calculation of Mendelian likelihoods. Hum Hered 33: 291–301

Lange K, Elston RC (1975) Extension to pedigree analysis. I. Likelihood calculations for simple and complex pedigrees. Hum Hered 25:95–105

Meyer K (1989) Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. Genet Sel Evol 21:317–340

Morton NE, MacLean CJ (1974) Analysis of family resemblance. III. Complex segregation analysis of quantitative traits. Am J Hum Genet 26:489–503

S.A.G.E. (1992) Statistical analysis for genetic epidemiology, release 2.1. Computer program package available from the Department of Biometry and Genetics, LSU Medical Center, New Orleans, La.

Thompson EA, Skolnick MH (1977) Likelihoods on complex pedigrees for quantitative traits. In: Pollack E, Kempthorne O, Bailey TB (eds). Proc Int Conf Quant Genet. Iowa State University Press, Ames, Iowa, pp 815–818

van Arendonk JAM, Smith C, Kennedy BW (1989) Method to estimate genotype probabilities at individual loci farm livestock. Theor Appl Genet 78:735–740