Linkage Analysis for Large Pedigrees under Mixed Inheritance

C. Stricker[1], R. L. Fernando[2] and R. C. Elston[3]

[1] Institute of Animal Sciences, ETH-Zentrum TAN 1, 8092 Zuerich, Switzerland; [2] Department of Animal Sciences, University of Illinois, 1207 West Gregory Drive, Urbana, IL 61801, U.S.A.; [3] Department of Biometry and Genetics, LSU Medical Center 1901 Perdido Street, New Orleans 70112, U.S.A.

Introduction

Under the usual mixed model of inheritance (Elston and Stewart, 1971; Morton and McLean, 1974), the exact likelihood cannot be calculated for large pedigrees, and likelihood analysis for large pedigrees is usually based on approximations of the likelihood. Fernando et al. (1994) proposed an alternative mixed model, with a finite number of polygenic loci (FPMM), for which the likelihood can be efficiently calculated. Linkage analysis by maximum likelihood under an FPMM was compared with that under the usual mixed model (UMM), where the likelihood under the UMM was approximated as described by Hasstedt (1982).

Material and Methods

Five three-generational pedigrees of the same structure with 4021 individuals each were simulated. The structure of each pedigree was generated by the following matings: 1 grandfather was mated to 10 unrelated grandmothers to produce 1 son each. These 10 halfsibs were mated each to 100 unrelated female individuals to produce 3 daughters each. The genotypic value for the simulated trait was additively determined by a major locus and 40 polygenic loci, with two alleles of equal frequencies at the major locus and at each of the polygenic loci. The difference between homozygotes was 4 at the major locus and 1 at each of the polygenic loci. This gives a variance of 2 for the major locus and of 5 for the polygenic component. The phenotype was simulated by adding a normally distributed residual with mean 0 and variance 7 to the genotypic value. A marker locus with two alleles of equal frequencies was simulated to be unlinked to the polygenic loci and linked to the major locus with a recombination fraction of 0.1. Whereas all individuals were genotyped with respect to the marker locus, all male individuals were assigned a missing phenotype. Linkage analysis under

Table 1: Maximum likelihood estimates under the FPMM

Pedigree	Difference between Homozygotes	Allele Frequency		Recombination Fraction	Variance		
		Major Locus	Marker Locus		Major Locus	Polygenic	Residual
A	3.81	0.54	0.48	0.02	1.8	4.39	7.82
B	3.17	0.57	0.49	0.03	1.2	6.22	6.69
C	3.78	0.44	0.51	0.13	1.8	5.19	6.90
D	4.01	0.54	0.51	0.08	2.0	4.48	7.10
E	4.51	0.56	0.50	0.15	2.5	4.46	6.74

the FPMM, assuming 5 polygenic loci for the polygenic component, was done using the computer package SALP (Stricker et al. 1994) and under the UMM using the computer package PAP (Hasstedt, 1982). The same number of parameters was estimated under both models from each pedigree. It should be noted that SALP uses the Downhill Simplex method (Nelder and Mead, 1965) to maximize the likelihood, whereas PAP maximizes the likelihood function using the variable metric method (GEMINI; Lalouel, 1979).

Results and Discussion

The maximum likelihood estimates for each of the five pedigrees are listed in table 1 for the FPMM and in table 2 for the UMM. The calculation, and thus also the maximization, of the exact likelihood (under the model that was used to simulate the pedigree data) was computationally not feasible. Thus it was not possible to infer the realized parameter values under the model used to simulate the pedigrees. However, because we simulated five large pedigrees, the differences between the realized parameter values and the parameter values used to generate the pedigrees could be expected to be quite small. When the approximate likelihood for the UMM was maximized with the program package PAP, the estimate for the recombination fraction was at its lower boundary zero in pedigrees B, D and E (indicated by '*' in table 2). To examine if other maxima exist for these pedigrees within the parameter space of the recombination fraction, the maximization process was restarted at the previous maximum but with the initial value 0.1 for the recombination fraction.

Table 2: Maximum likelihood estimates under the UMM

Pedigree	Difference between Homozygotes	Allele Frequency		Recombination Fraction	Variance		
		Major Locus	Marker Locus		Major Locus	Polygenic	Residual
A	2.74	0.38	0.51	0.07	0.9	5.02	7.66
B	1.76	0.43	0.49	0.0*	0.4	6.92	6.76
C	4.11	0.44	0.51	0.23	2.1	5.54	6.32
D	0.77	0.52	0.51	0.0*	0.1	6.34	7.07
E	2.05	0.58	0.50	0.0*	0.5	6.41	6.57

* Likelihood maximized at lower boundary

Because the likelihoods for pedigrees B, D and E were again maximized at the same boundary, the lower boundary for the recombination fraction was increased from 0.0 to 0.0001 and the maximization process was restarted again. Because all these restarted maximization processes converged to the boundary 0.0 or 0.0001 for the recombination fraction, respectively, for all three pedigrees, it was concluded that there was no other maximum within the parameter space for pedigrees B, D and E.

The maximum likelihood estimates under the FPMM show only relatively small differencies among the five simulated pedigrees and are within close range of the parameters that were used to simulate the pedigree data. The maximum likelihood estimates under the UMM approximated by Hasstedt (1982) showed more variability among the 5 pedigrees, especially for the parameters recombination fraction, major gene frequency and major genotypic means. As a consequence of the parameter estimates for the major genotypic means and frequencies under the UMM approximated by Hasstedt (1982), the amount of variation the model fitted to the major locus within each pedigree is also different and deviates considerably from the values obtained by the FPMM and those used to simulate the pedigree data.

Conclusions

Fernando et al. (1994) showed, in a segregation analysis of pedigree data simulated under the UMM, that the FPMM fits the data better than do regressive models (Bonney, 1992) or the mixed model approximation of Hasstedt (1982). For the simulated pedigree data in the present study, exact calculation of the maximum likelihood estimates was not feasible; but due to the large number of simulated individuals in each pedigree, the realized parameter values in the pedigree samples could be expected to be within close range of the parameters that were used in the simulation. Thus, the results of the present study extend to linkage analysis the findings of Fernando et al. (1994) with respect to the mixed model approximation of Hasstedt (1982).

References

Bonney, G.E. (1992) Hum. Hered. 42:28-41.

Elston, R.C. and Stewart, J. (1971) Hum. Hered. 21:523-542.

Fernando, R.L., Stricker, C. and Elston, R.C. (1994) Theor. Appl. Genet. (in press).

Lalouel, J-M. (1979) Tech. Report 14, Pop. Genet. Lab. Univ. Hawaii, Honolulu, U.S.A.

Hasstedt, S.J. (1982) Comput. Biomed. Res. 15:295-307.

Morton, N.E. and McLean, J.C. (1974) Amer. J. Hum. Genet. 26:489-503.

Nelder, J. A. and Mead, R. (1965) Computer Journal 7:308-313.

Stricker, C., Fernando, R.L. and R.C. Elston (1994) Proc. 5^{th} World Congr. Genet. Livest. Prod. (submitted).