Proc. Assoc. Advmt. Anim. Breed. Genet. Vol 13

INFERRING GENOTYPE PROBABILITIES FOR UNTYPED INDIVIDUALS IN COMPLEX PEDIGREES

J. M. Henshall, B. Tier, and R. J. Kerr

Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351

SUMMARY

A method of estimating genotype probabilities in complex pedigrees is described. The method is similar to existing genotype elimination methods, but is based on sampling inheritance rather than genotype. This removes assumptions regarding the distribution of alleles in ungenotyped base animals, reducing bias in the genotype probability estimates.

Keywords: Genetic markers, complex pedigrees

INTRODUCTION

Increasing use of markers to detect quantitative trait loci (QTL) is being made by animal breeders. The process is still expensive and consequently only a proportion of the population is genotyped. The probability that ungenotyped individuals possess particular genotypes at these loci is of interest, and is similar to the problem faced by human geneticists concerned with carriers of genetic disorders.

Currently no algorithm for determining these probabilities unbiasedly is available for large complex pedigrees containing many intersecting loops, which are the norm for animal breeders. Loops cause problems for approximation methods such as peeling based algorithms (eg. Elston and Stewart 1971). A method for computing unbiased genotype probabilities for ungenotyped animals in complex pedigrees would be most useful for both human and animal geneticists.

One approach to the estimation of genotype probabilities is to repeatedly sample the genotype of the entire pedigree. Each sample must be consistent with the observed genotypes and adhere to the mendelian rules of inheritance. If the origin (paternal or maternal) of the alleles is specified in the sample, then it can be referred to as an ordered legal genotype sample.

Sobel and Lange (1996) use the genotype elimination algorithm (see Lange 1997) to generate ordered legal genotypes as a first step in a QTL detection algorithm. The genotype elimination algorithm repeatedly cycles through the pedigree, eliminating any ordered genotypes that are inconsistent with the set of ordered genotypes currently feasible for parents and offspring. The result is a list of ordered genotypes for each animal, which are consistent with the observed marker data. Iteration can be used to narrow this down further to a list in which each animal has only a single ordered genotype (Lange 1997). In each cycle, one ordered genotype is randomly chosen for an animal with more than one possible ordered genotype, and the genotype elimination algorithm applied to the newly constrained pedigree. This is referred to here as genotype elimination through genotype constraint (GEGC).

---

* AGBU is a joint institute of NSW Agriculture and The University of New England

For Sobel and Lange (1997), any legal ordered genotype was satisfactory, as they were only used as starting values in a larger algorithm. However, if legal genotypes are being used to estimate genotype probabilities, not only must the genotypes be legal, they must be random samples from the appropriate distribution. The random parts of the GEGC algorithm are in the choice of the animal to constrain in each cycle, and its genotype. Ideally, a genotype would be selected with probability according to the (constrained) genotype probability for that animal. However this is unknown, and its estimation is likely to be the reason for implementing the algorithm in the first place. Heath (1998) included a peeling step in the GEGC algorithm to estimate constrained genotype probabilities prior to each round of constraint. This should reduce, but not eliminate bias in genotype probabilities estimated from repeated sampling, as peeling produces biased genotype probability estimates in the presence of loops (Fernando *et al.* 1993).

Without markers, but with phenotypic measurements for a quantitative trait, it is possible to draw QTL genotype samples for the entire pedigree (Tier and Henshall, pers comm). If it were feasible to implement this method with marker data, then the samples would provide unbiased genotype probability estimates. Unfortunately, due to the high proportion of illegal marker configurations, the method is probably infeasible for large pedigrees with marker data. However, the idea of sampling inheritance rather than genotype is applicable to genotyped pedigrees and a modification to the GEGC algorithm to incorporate inheritance sampling is described here.

METHOD

Animals inherit a marker allele from each parent. Each of these alleles will have originated from either a grandfather or grandmother. The origin of each allele is its inheritance state. Marker data may limit the inheritance state of an allele to a single configuration. Such an allele is said to be constrained in its inheritance state. Otherwise, the allele's inheritance state is unconstrained by the marker data.

The Genotype Elimination through Inheritance State Constraint (GEIC) Algorithm. The algorithm proceeds as follows:

1. Apply genotype elimination to construct a list of feasible ordered genotypes for each animal from the marker data.
2. From the feasible genotypes found in step 1, construct a list of alleles with unconstrained inheritance states.
3. Repeat the following steps until no unconstrained inheritance states remain:
    a) Randomly chose an allele from the current list with unconstrained inheritance states.
    b) Assign the allele chosen in step 3a a random inheritance state.
    c) Eliminate any infeasible genotypes resulting from Step 3b.
    d) Rebuild the list of alleles with unconstrained inheritance states.

[All animals will now have either a single feasible ordered genotype, or multiple ordered genotypes with equal probability, or the sample may be illegal].

4. Store (or accumulate) the sample if legal.
5. Restore the original list of alleles with unconstrained inheritance states identified in step 2 and repeat steps 3 and 4 until sufficient samples are drawn.
6. Calculate ordered genotype probabilities for all individuals as the mean of the samples.

A special case of multiple ordered genotypes with equal probability occurs when a marker allele is equally likely to take any value, because that particular allele was not measured anywhere in the pedigree.

As the inheritance state of all animals is not sampled in a single step, but in multiple steps conditional on previous steps, some bias is introduced. Some samples are found after fewer cycles than others. A variation of the GEIC algorithm is to weight each sample by a function of the number of cycles used in obtaining the sample. We will refer to this as the GEICW algorithm.

Table 1. Simulated pedigree and observed unordered genotypes (.. denotes genotype unknown).

Animal	1	2	3	4	5	6	7	8
Sire	0	0	0	2	2	4	2	7
Dam	0	0	0	1	3	5	5	6
Genotype	13	..	..	..	12	..	11	11

Evaluation A small pedigree was constructed to compare the algorithms (see Table 1). Three alleles are observed in the pedigree, labelled 1, 2, and 3. The true genotype probabilities for each animal were established using an exhaustive search (ES) of the complete pedigree space (an infeasible process for all but very small pedigrees, unless there are no loops in the pedigree). Three iterative genotype elimination algorithms were tested, genotype elimination through genotype constraint (GEGC), genotype elimination through inheritance state constraint with unweighted samples (GEIC) and GEIC with samples weighted by 2^n where n is the number of cycles used to obtain the sample (GEICW). Animals to be constrained were chosen randomly from a uniform distribution of those available, and the genotype or inheritance state chosen randomly from a uniform distribution of those available. 10,000 samples were drawn for each method, and the results accumulated.

RESULTS AND DISCUSSION

Table 2 contains the probabilities obtained using the four methods for those animals where the genotype probabilities are of interest. None of the sampling methods provided genotype probabilities identical to those obtained using an exhaustive search, but GEIC and GEICW produced estimates closer to those from ES than GEGC did. For example, GEGC severely underestimated the probability of animals 2, 4 and 6 being homozygous. The estimates produced by GEICW were clearly superior to those produced by GEIC. Additional research will establish whether better weight functions exist.

Randomly sampling an inheritance state for a random animal is similar to sampling a genotype for that animal's parents given the genotypes of progeny, which is similar to what occurs in the algorithm of Heath (1998). The properties of the estimates produced by this algorithm and GEIC require further investigation. GEIC should be equally suited to large complex pedigrees and should be quicker and less biased as no peeling step is required.

Table 2. Genotype probabilities (%) obtained using an exhaustive search (ES), genotype elimination through constraint (GEGC) and genotype elimination through inheritance constraint unweighted (GEIC) and weighted (GEICW). The paternal alleles precede maternal alleles, and "*" indicates that the allele was not measured anywhere in the pedigree, and is therefore unknown

	Ordered	Genotype probability (%)			
Animal	genotype	ES	GEGC	GEIC	GEICW
2	*1, 1*	14.3	28.7	13.8	15.2
	11	42.9	5.7	31.1	41.2
	12, 21	14.3	18.4	20.6	14.3
3	*1, 1*	14.3	18.4	20.6	14.3
	2, 2	35.7	31.5	29.4	35.7
4	*1	7.7	19.8	6.6	9.0
	*3	3.6	15.5	3.8	3.2
	11	46.4	18.8	41.7	50.0
	13	32.1	25.2	30.9	26.2
	21	7.1	10.4	10.7	8.6
	23	3.6	10.2	6.4	2.9
5	21	28.6	36.9	41.2	28.5
	12	71.4	63.1	58.8	71.5
6	*1	3.6	15.6	2.5	3.2
	11	53.6	27.2	47.5	52.8
	12	28.6	26.5	27.5	32.3
	21	3.6	9.8	5.1	2.9
	31	10.7	20.9	17.5	8.8

REFERENCES

Elston, R. C. and Stewart, J.(1971) *Human Heredity* **21**:523

Fernando, R. L., Stricker, C. and Elston, R. C. (1993) *Theor. Appl. Genet.* **87**:89

Heath, S.C. (1998) *Human Heredity* **48(1)**:1

Lange, K. (1997) "Mathematical and Statistical Methods for Genetic Analysis" Springer-Verlag New York

Sobel, E. and Lange, K. (1996) *Am. J. Hum. Genet.* **58**:1323