ALGORITHMS FOR MATE SELECTION

R. K. Shepherd[1] and B. P. Kinghorn[2]

[1]Central Queensland University, Rockhampton, QLD 4702
[2]University of New England, Armidale, NSW 2351

SUMMARY

Two types of Mate Selection Index (MSI) are discussed. Algorithms for implementing mate selection are reviewed. Over the last 15 years progress has been made in developing efficient algorithms for mate selection. Evolutionary Algorithms (EA) are most successful when the value of an individual mating depends on what other matings are made. Desireable properties of EA are discussed. More research is required into developing efficient mate selection algorithms, especially if Look Ahead Mate Selection (LAMS) is to be implemented. LAMS schemes involve mate selection among predicted future progeny. LAMS is most useful where non-additive effects, like heterosis or non-additive QTL effects, are of importance.
Keywords: Breeding program, selection, mating, evolutionary algorithm

INTRODUCTION

A challenge facing modern animal breeding is how to simultaneously consider all the important issues when designing a genetic improvement program. Decisions on who to select for breeding and how mates should be allocated, the need to use all relevant information, like EBVs, coancestry, etc and available resources in order to meet the breeding objective. Simultaneously considering the decisions of selection and mate allocation is called mate selection. A tactical approach for simultaneously accommodating all the issues of importance is to:

- develop a Mate Selection Index (MSI) which describes net economic merit in terms of selection and mating decisions, and
- develop and implement a mate selection algorithm which maximises the MSI.

This paper focuses on the issues involved in developing and implementing a mate selection algorithm with particular reference to the advantages and disadvantages of existing algorithms.

EVALUATING A MATE SELECTION INDEX

Selection and mating decisions can be represented by a decision matrix X with elements X_{ij} indicating whether male i is mated ($X_{ij} = 1$), or not mated ($X_{ij} = 0$), to female j. Usually there are constraints placed on X because of available resources and the logistics of mating. For example, without advanced reproductive techniques, females may mate at most once, which constrains the sum of each column of X to be either 0 or 1. Two important scenarios arise when developing and evaluating a mate selection index for a particular application.

S1. *The value of an individual mating is independent of what other matings are made*. This scenario covers most instances of maximising utility in the next generation. For example, this occurs when the value of a mating is its mean progeny merit, defined as the average EBV of the sire and dam, plus maternal effects, breed additive effects and breed heterotic effects as expressed in the progeny. Then the value of a mating set (or the MSI) is the sum of the values of the individual matings.

S2. *The value of an individual mating is NOT independent of what other matings are made.* In this scenario it is not possible to calculate the value of an the individual mating. In fact the MSI is usually only defined for the whole mating set, and cannot be marginalised or parititioned into individual mating contributions.

Scenario S2 is quite common in animal breeding programs. For example, Kinghorn *et al* (1999) defined an MSI containing three components: mean progeny genetic merit (for genetic gain), mean progeny inbreeding coefficient (for short-term inbreeding) and mean coancestry of all parents (for long-term inbreeding). To calculate the mean parental coancestry component, the whole mating set is required. Hence the value of an individual mating is not meaningful as the MSI is only defined for whole mating sets. Another example of an S2 scenario is the Look Ahead Mate Selection (LAMS) scheme (Shepherd and Kinghorn 1998) which looks a number of generations into the future. Crossbred matings in the future depend on which matings are made in the current generation. Thus the value (in the future) of a current mating depends on what other matings are made in the current mating set. Hence the MSI is only defined for a complete mating set.

ALGORITHMS FOR IMPLEMENTING MATE SELECTION

A few procedures (or algorithms) have been advocated for finding the *optimal* mating set ie. the one that maximises the MSI. Each type of algorithm will be presented and discussed in terms of its advantages and disadvantages.

Complete enumeration. There are usually too many possible mating sets for full enumeration to be computationally feasible. For example, if there are M male candidates and F female candidates available for selection and p matings are required, then the total number of possible mating sets is $p!({}^{M}C_p)({}^{F}C_p)$ if males can mate only once, and $M^p({}^{F}C_p)$ if each male can mate up to p times. These formulae assume females can mate only once. If $M=F=15$ and $p=4$ then there are over 44 million possible mating sets if males can only mate once, and over 69 million possible mating sets if males can mate up to 4 ($= p$) times. Hence for any livestock breeding herd or flock it will be a futile exercise to attempt to systematically or randomly search for the *optimal* mating set as the expected number of MSI evaluations to find the *optimal* mating set is N, where N is the total number of possible mating sets.

Linear Programming. If the value of an individual mating is independent of what other matings are made (scenario S1), then a value matrix V can be constructed in which V_{ij} represents the value of the mating between male i and female j. Then the problem is to maximise $\Sigma V_{ij}X_{ij}$ subject to a number of resource and logistical constraints as discussed earlier. If all the constraints are linear, then Linear Programming (LP) techniques can be used to search all possible mating sets to find the *optimal* one. LP techniques evaluate only a small subset of all the possible mating sets and so are much more computationally efficient than complete enumeration. Jansen and Wilton (1984) showed how to formulate mate selection as an LP transportation problem and thus the more efficient LP transportation algorithms can be used instead of more general LP algorithms. The transportation problem will find the optimal mating set in a number of iterations of the order of $(M+1)(F+1)$. For example, if $M=F=15$ then $(M+1)(F+1)=625$.

Exchange algorithm. The exchange algorithm can be used when *the value of an individual mating is NOT independent of what other matings are made* (scenario S2). Kinghorn and Shepherd (1994) described and evaluated the effectiveness of the exchange algorithm. The algorithm starts with a *good* mating set, and then sequentially swaps a currently accepted mate allocation with an allocation not currently accepted (possibly involving animals not currently selected). The new mating set is tested to see if the MSI has increased, whereupon the modified mating set is accepted. The algorithm stops when no further improvements (ie. single mating pair swaps) can be found. There is no guarantee with the exchange algorithm that the final mating set will be *optimal* in terms of the MSI. Kinghorn and Shepherd (1994) showed that as the number of matings increased, the probability of the exchange algorithm finding the *optimal* mating set decreased, even though the efficiency on a merit scale relative to random mating was always very high. However the biggest disadvantage of the exchange algorithm is that it slows down dramatically as the breeding population increases. If no swaps are accepted (eg. we start with the *optimal* mating set), the exchange algorithm has to perform $p(MF - p)$ evaluations of the MSI, and every time a swap occurs, this number of evaluations has to occur before it stops. In general it is difficult to say how long the exchange will take as it depends critically on the initial mating set chosen and as individual matings may move in and out of the current mating set a number of times.

Evolutionary Algorithms (EA). Evolutionary algorithms mimic biological evolution in their quest to find the best solution for a complex optimisation problem. The components of an EA include a genetic representation of a feasible solution, a fitness criterion (eg. the MSI), and genetic operators like recombination and mutation. Then starting with an initial population of solutions, new generations of the population evolve using the concepts of survival (or selection) of the fittest and breeding new solutions using the genetic operators, all in a probabilistic framework. Recently EA have been used for mate selection when *the value of an individual mating is NOT independent of what other matings are made* (scenario S2). Hayes *et al* (1997) described an EA called a Genetic Algorithm (GA) and evaluated its efficiency for small mate allocation problems. Although the efficiency increased dramatically as population size increased, it was much too slow for large industry applications. More recently an EA called Differential Evolution (DE) has been adapted to mate selection and has proved successful in industry applications (Kinghorn and Shepherd 1999). Whether DE will be adequate for all MSI, is currently a debateable question. However lessons from biological evolution, which show that different species evolve to inhabit different niches, would suggest that maybe different EA are needed for MSI which are substantially different.

DESIRABLE PROPERTIES OF AN EA

There are a number of features that could make EA more efficient for mate selection. First, it is important that only feasible mating sets (ie ones that satisfy the constraints) be generated when the genetic operators of recombination and mutation are used. The GA described by Hayes *et al* (1997) allowed the breeding of infeasible *offspring* solutions following recombination. In these situations much computer time can be spent *aborting* infeasible offspring solutions and searching for feasible solutions. Alternatively, the offspring solutions can be *fixed up* to satisfy the constraints. But then they may not closely resemble the parental feasible solutions. Both these outcomes are undesireable. The DE of Kinghorn and Shepherd (1999) always breeds feasible solutions. Second, it is desireable to select parental solutions for breeding *on the basis of offspring fitness* due to the non-additive nature of matings in the MSI. This does not occur with the GA of Hayes *et al* (1997) but does occur in the DE of Kinghorn and Shepherd (1999). Also due to non-additivity it would seem a better strategy to use a number of small breeding populations rather than a single large breeding population of feasible solutions (*cf.* Wright's Shifting Balance Theory). This approach is further enhanced if parallel processing is utilised. Another desireable feature is the monitoring of diversity in a population at each generation in order to *explode* the mutation when diversity is low. This would allow the climbing of MSI hills by the population of solutions and, once at the summit (ie. low diversity), a wider exploration of the MSI landscape would result due to the mutation.

CONCLUSIONS

The utility of mate selection increases with the complexity of the issues and information available for decision making. Mate selection algorithms need to implicitly handle this complexity in the MSI. Over the last 15 years progress has been made in developing efficient algorithms for mate selection, with EA displaying the most promise. Like evolution, EA work best over long time scales and thus require a large number of computer generations. The amount of computer time involved is being continually reduced by the annual increase in processor speed, the use of parallel processing architecture and the development of algorithms which are more adapted to particular MSI. Efficient group mate selection procedures may need to be developed for large populations. Basically more research is required in developing efficient mate selection algorithms. These developments will be necessary if MSI like Look Ahead Mate Selection (LAMS) of Shepherd *et al* (1998) is to be implemented as LAMS require EA driving EA. LAMS is most useful where non-additive effects, like heterosis or non-additive QTL effects, are of importance.

REFERENCES

Jansen G.B. and Wilton J.W. (1985) *J. Dairy Sci.* **68**:1302

Hayes B.J., Shepherd R.K. and Newman S. (1997) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **12**:108

Kinghorn B.P. and Shepherd R.K. (1994) *Proc. 5th World Cong. Genet. App. Livest. Prod.* **18**:255

Kinghorn B.P. and Shepherd R.K. (1999) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **13**: 130

Kinghorn B.P., Shepherd R.K. and Woolliams J.L. (1999) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **13**:412

Shepherd R.K. and Kinghorn B.P. (1998) *Proc. 6th World Cong. Genet. App. Livest. Prod.* **25**:431