

Technical Note: Recursive Algorithm for Inbreeding Coefficients Assuming Nonzero Inbreeding of Unknown Parents

I. Aguilar*† and I. Misztal*¹

*Animal and Dairy Science Department, University of Georgia, Athens 30602

†Instituto Nacional de Investigación Agropecuaria, Las Brujas, Uruguay

ABSTRACT

A recursive algorithm to calculate inbreeding coefficients was modified to account for nonzero inbreeding of unknown parents. The modification was done by changing one part of a recursive formula in which the inbreeding of an animal with at least one unknown parent is not zero and replacing it by the mean inbreeding of all animals born the same year. The algorithm is iterative. Testing involved 17 million US Holsteins. Convergence was reached in 6 rounds. The computing time per round was 4 min, twice as fast as the VanRaden algorithm based on the tabular method. The recursive algorithm is very simple; however, it requires that the recursion takes into account the order of animals. After a simple modification, the algorithm provides a very good approximation of inbreeding when the pedigree is unordered.

Key words: inbreeding coefficient, unknown parent, recursive algorithm

VanRaden (1992) presented a method to calculate inbreeding while accounting for missing parents. Animals with missing parents were assumed to have inbreeding coefficients equal to the mean of the inbreeding coefficients for animals with known parents born during the same year.

VanRaden's algorithm recovered most of the inbreeding caused by incomplete pedigrees when the number of missing pedigrees was not too large (Lutaaya et al., 1999). In that study, different proportions of missing information were simulated by randomly deleting 10 to 50% of dam identification in a Holstein population with almost complete pedigrees. VanRaden's method recovered most of the inbreeding when the missing dam information was 10 to 20% (Lutaaya et al., 1999).

The algorithm by VanRaden is based on the tabular method (Emik and Terrill, 1949). For large pedigrees,

this algorithm requires extracting pedigrees for each individual, which makes the algorithm complicated. In addition, for long pedigrees, the size of the table is large and computations are long. The computing cost of the tabular method is approximately $n \times 2^{2p}$, where n is the population size and p is an average number of generations of pedigree per animal.

Different methods are currently used to compute inbreeding coefficients efficiently for large populations (Tier, 1990; Meuwissen and Luo, 1992; Sargolzaei et al., 2005). The last 2 algorithms are optimized versions of the algorithm by Quaas (1976), which has a computational cost of n^2 . The algorithms based on the recursive algorithms (RA) seem to be the simplest (Tier, 1990; Miglior et al., 1992), and their computing cost is approximately $n \times 2^p$. All of these methods as described assign an inbreeding coefficient of zero when at least one parent is missing. Recently, Croquet et al. (2006) applied a modification of the method of Meuwissen and Luo that assigned nonzero inbreeding to animals with unknown parents.

The objectives of this study were to investigate a simple RA to calculate inbreeding coefficients by using rules from the tabular method (Emik and Terrill, 1949) and to expand it to consider animals with missing parent information (VanRaden, 1992).

The RA assumes that animals must be renumbered based on year of birth (YOB) so that parents precede their progeny. For each animal (x) present in the pedigree, the inbreeding coefficient F_x is calculated as $F_x = 0.5R_{sd}$, where R_{sd} is the numerator relationship between the sire (s) and the dam (d) of the animal (x) (Emik and Terrill, 1949). Computation of R_{sd} is recursive and involves tracing the ancestors and computing the relationship between s and d . Three cases are considered:

$$R_{xy} = \begin{cases} 0 & \text{if } x = 0 \text{ or } y = 0 \\ 1 + F_x & \text{if } x = y \\ 0.5(R_{sy} + R_{dy}) & \text{otherwise} \end{cases}$$

where animal x is younger than y , and s and d are the sire and dam of x , respectively.

Received August 3, 2007.

Accepted December 17, 2007.

¹Corresponding author: ignacy@uga.edu

In the last case, the order of animals x and y should be switched if necessary so that subsequent relationships are those of the sire and dam of the younger animal (x) to the older animal (y), which is essential if the 2 animals are in the direct line of descent (Emik and Terrill, 1949).

In the algorithm above, $x = 0$ or $y = 0$ indicates an unknown parent or parents. To extend the algorithm to nonzero inbreeding of unknown parents, let a negative code denote the YOB of their progeny. For example, $x = -1998$ would indicate a parent of an animal born in 1998. Let b_i be an average inbreeding of all animals born in year i . Then, the first rule above ($R_{xy} = 0$) is modified as follows:

$$R_{xy} = 2 \begin{cases} b_{-x} & \text{if } x < 0 \text{ and } y > 0 \\ b_{-y} & \text{if } x > 0 \text{ and } y < 0 \\ \max(b_{-x}, b_{-y}) & \text{if } x < 0 \text{ and } y < 0 \end{cases}$$

In the iterative procedure, b is calculated each round. Let $yob(x)$ be a YOB of known animal x . In the first round b_i is zero. In subsequent rounds,

$$b_i = \bar{F}_x \quad x: yob(x) = i,$$

and the algorithm is repeated until convergence. The average above includes only animals with both parents identified (VanRaden, 1992).

The convergence criterion was defined as in Lutaaya et al. (1999) as the absolute change in mean of inbreeding for all animals between 2 consecutive rounds less than 10^{-6} . The algorithm was implemented in Fortran 90 and the source code is shown in the Appendix.

The algorithm was tested by using the pedigree of US Holsteins born from 1969 to 2002 ($n = 17,094,297$) obtained from the Animal Improvement Programs Laboratory, Agricultural Research Service, USDA (Beltsville, MD). Missing YOB for parents were calculated as the minimum YOB from the offspring minus 3 yr. Animal identifications were checked to avoid impossible offspring-parent relationships based on the YOB, and they were renumbered from oldest to youngest. Solutions of the RA with unknown parent information were compared for correctness to the algorithm by VanRaden (1992), as used in Lutaaya et al. (1999). After convergence, both methods gave the same estimates of inbreeding. Computing time was 550 s for one round of the VanRaden algorithm and 287 s for one round of RA. Despite differences among methods, the computing was fast for both methods given the size of the pedigree.

The algorithms based on the Quaas algorithm or the tabular method explicitly require ordering animals from the oldest to the youngest. During initial trials,

Table 1. Statistics of calculated inbreeding coefficients for US Holsteins with the recursive algorithm, the modified algorithm¹ and maximum differences between them

Item	Recursive algorithm	Modified ¹	Difference
Minimum	0	0	0
Maximum	0.469	0.469	0.250
Average	0.011	0.010	0.001
SE	0.021	0.020	0.001

¹Animal with missing parents treated as y (older) in the recursion formula.

pedigrees used by RA were not sorted; however, the algorithm was modified to treat an animal without parents known as y (Table 1). For several data sets, including commercial data sets, inbreeding coefficients produced by such an algorithm were almost the same as by the original algorithm with animals sorted, although the differences increased with larger pedigrees with longer generations (Table 2). Some differences arose from pedigree errors, which were detected and eliminated when animals were sorted, and others were due to incorrect use of the recursion formula (“ x younger than y ”). The RA algorithm provides many correct results without explicit sorting because of implicit sorting during the recursion process. Although the modified algorithm may be useful in some situations, renumbering has an additional benefit of detecting errors in the pedigree.

Figure 1 shows the mean inbreeding coefficients per birth year at rounds 1, 2, and 6. Convergence was reached at 6 iterations. Mean inbreeding per year had a pattern similar to the official inbreeding for all the population of US Holsteins as calculated by the Animal Improvement Programs Laboratory (2007).

The method of VanRaden assumes that inbreeding of unknown parents varies by year only. It is possible to generalize this concept to those used in BLUP and to have separate classes of inbreeding of unknown parents by sex, registry status, country of origin, and some dam-sire origin combinations. In such a case, several modifications are needed. First, the vector of averages

Table 2. Statistics of calculated inbreeding coefficients for US Holsteins with the recursive algorithm using pedigrees with different ordering and difference between them

Item	Pedigree ordering		Difference
	Parent first	Random	
Minimum	0	0	0
Maximum	0.469	0.438	0.469
Average	0.011	0.002	0.010
SE	0.020	0.010	0.017

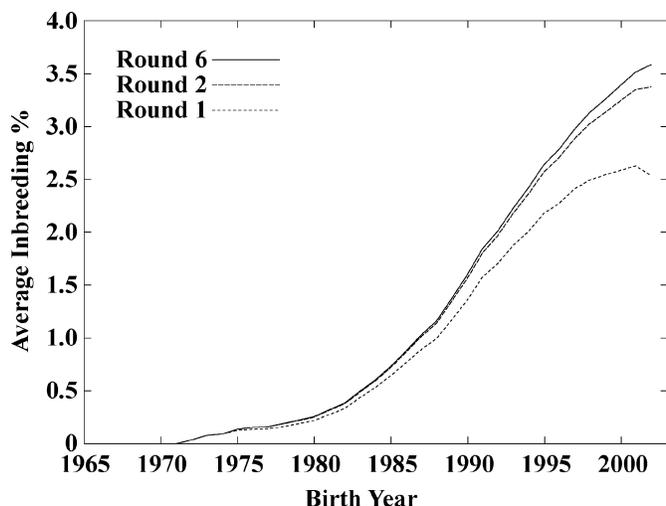


Figure 1. Mean inbreeding (%) per year of birth at different rounds.

needs to be separated for each class. Second, each known animal and each unknown parent needs to be assigned to a class. The estimates of inbreeding may be inaccurate if counts are small for some classes or if some classes are confounded.

We demonstrated that the RA to compute inbreeding coefficients can be modified to account for nonzero inbreeding of unknown parents. The algorithm is simple, is relatively inexpensive, and can also be used for other purposes, including computing of relationships of groups of animals (e.g., sires) by using the full pedigree or the creation of nonadditive relationships (e.g., the dominance relationships). For example, most exploitable dominance relationships are due to sire \times maternal grandsire classes with large numbers of members. The dominance relationship matrix for such sire combina-

tions can be constructed by computing relationships among such sires and grandsires, and possibly of their ancestors. Because only relationships of interest are computed, the cost of computations can be reasonable.

ACKNOWLEDGMENTS

We are grateful to Juan Pablo Sanchez, Andres Legarra, and 2 anonymous reviewers for helpful comments and suggestions on the manuscript. We acknowledge the Animal Improvement Programs Laboratory, Agricultural Research Service, USDA (Beltsville, MD), for providing the pedigrees. This study was partly supported by the Holstein Association.

REFERENCES

- Animal Improvement Programs Laboratory. 2007. Inbreeding trend. <http://aipl.arsusda.gov/eval/summary/inbrd.cfm> Accessed July 1, 2007.
- Croquet, C., P. Mayeres, A. Gillon, S. Vanderick, and N. Gengler. 2006. Inbreeding depression for global and partial economic indexes, production, type, and functional traits. *J. Dairy Sci.* 89:2257–2267.
- Emik, L. O., and C. R. Terrill. 1949. Systematic procedures for calculating inbreeding coefficients. *J. Hered.* 40:51–55.
- Lutaaya, B. E., I. Misztal, J. K. Bertrand, and J. W. Mabry. 1999. Inbreeding in populations with incomplete pedigrees. *J. Anim. Breed. Genet.* 116:475–480.
- Meuwissen, T. H. E., and Z. Luo. 1992. Computing inbreeding coefficients in large populations. *Genet. Sel. Evol.* 24:305–313.
- Miglior, F., B. Szkotnicki, and E. B. Burnside. 1992. Analysis of levels of inbreeding and inbreeding depression in jersey cattle. *J. Dairy Sci.* 75:1112–1118.
- Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* 32:949–953.
- Sargolzaei, M., H. Iwaisaki, and J. J. Colleau. 2005. A fast algorithm for computing inbreeding coefficients in large populations. *J. Anim. Breed. Genet.* 122:325–331.
- Tier, B. 1990. Computing inbreeding coefficients quickly. *Genet. Sel. Evol.* 22:419–430.
- VanRaden, P. M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *J. Dairy Sci.* 75:3136–3144.

APPENDIX

The Algorithm in Fortran 90

```
! Assumes animals renumbered from 1 to n, oldest to youngest
integer:: n, & ! number of animals
          k    ! number of years
integer:: ped(3,n), & ! pedigree matrix sire, dam, year of birth
          navg(k)    ! number of animals per year with known parents
real:: f(n), & ! inbreeding
          avginb(k), avginbn(k) ! average inbreeding per year
...
avginbn = 0
do
    f = -1; avginb = avginbn; avginbn = 0; navg = 0
```

```

do i = 1, n
  if (f(i) == -1) f(i) = inbrec (i)
  if (ped(1,i) > 0 .and. ped(2,i) > 0) then
    avginbn(ped(3,i)) = avginbn(ped(3,i)) + f(i)
    navg(ped(3,i)) = navg(ped(3,i)) + 1
  endif
enddo
minb=sum(f)/n
where (navg/=0)
  avginbn=avginbn/navg
end where
if (abs(minbold-minb) < 1.e-6) exit
minbold=minb
enddo
real function inbrec (an)
! Returns inbreeding coefficient for animal = an
! f(an) = 0.5 * cffa(s,d)
! negative s or d means UPG code
  integer:: an, s, d
  s = ped(1,an); d = ped(2,an)
  if (s <= 0 .or. d <= 0) then
    inbrec = avginb(abs(min(s,d)))
  else
    inbrec = 0.5 * (cffa(s,d))
  endif
end function inbrec
recursive real function cffa (a1, a2)
! Returns relationship between a1 and a2
  integer:: a1, a2
  if (a1 <= 0 .or. a2 <= 0) then
    cffa = 2 * avginb(abs(min(a1,a2)))
  elseif (a1==a2) then
    cffa = 1 + f(a1)
  else
    if (a1 < a2) then
      cffa = 0.5 * (cffa(a1,ped(1,a2)) + cffa(a1,ped(2,a2)))
    else
      cffa = 0.5 * (cffa(a2,ped(1,a1)) + cffa(a2,ped(2,a1)))
    endif
  endif
end function cffa

```