A GENERALISATION OF AN APPROXIMATE METHOD TO CALCULATE INBREEDING COEFFICIENTS

B. Cilliers

Department of Genetics, University of Natal

OPSOMMING: 'N VERALGEMENING VAN 'N BENADERDE METODE OM KOËFFISIËNTE VAN INTELING TE BEREKEN

Die benaderde metode vir die berekening van koëffisiënte van inteling vanaf stambome soos dit deur Wright & McPhee (1925) uiteengesit is, word veralgemeen sodat verskillende metodes van bemonstering van lyne van afstamming gehanteer kan word.

SUMMARY:

The approximate method for calculating coefficients of inbreeding from pedigrees presented by Wright & McPhee (1925), is generalized to accommodate different methods of sampling random lines of descent.

Sewall Wright (1922) used his theory of path coefficients to derive the following well known formula for calculating the inbreeding coefficient (F_X) of an individual, X:—

$$F_X = \Sigma \left(\tfrac{1}{2}\right)^{n+n'+1}(1 + F_A).$$

Where, A is a common ancestor of the parents of X; n is the number of generations from the sire of X to A and n' the number of generations from the dam of X to A.

Subsequently, Wright & McPhee (1925) developed a method for calculating an approximate inbreeding coefficient from a pedigree which is completed for only a limited number of generations, the residual pedigree being represented by randomly selected lines of descent. Complete pedigrees are compiled for i generations. Thus 2^i individuals are recorded in the ith generation. Only one parent, chosen at random, of each of these 2^i individuals is recorded in the $(i + 1)$th generation. Thereafter only one parent of a recorded individual is randomly selected and recorded in every generation. This procedure is continued with as far back into the ancestry as the research worker wishes to go. The estimated inbreeding coefficient is then obtained by adding the contributions of all ties, i.e. all paths joining sire and dam through a common ancestor.

Three kinds of ties are distinguished:

a. Those involving common ancestors within the completed part of the pedigree, where each tie has the value

$$\left(\tfrac{1}{2}\right)^{n + n' + 1}(1 + F_A)$$

b. Those involving common ancestors in the randomly selected lines of descent only, where each tie has the value

$$\left(\tfrac{1}{2}\right)^{2k+1}(1 + F_A) \qquad (k = i-1)$$

c. Those involving a common ancestor that appears in the completed part of the pedigree of one parent and in the random part of the other. The appropriate formulae for each tie in this case are:

$$\left(\tfrac{1}{2}\right)^{n+k+1}(1+F_A) \text{ where the common ancestor is in the sire's side of the completed pedigree and,}$$

$$\left(\tfrac{1}{2}\right)^{n'+k+1}(1+F_A) \text{ where it is on the dam's side.}$$

This approximate method was successfully applied by many research workers all over the world and in particular by Lush and his school in the genetic analyses of many breeds of farm animals.

Wright and McPhee's approximate method has two limitations.

(i) The number of random lines must be equal to the number of individuals in the ith generation, i.e. the last generation of the completed portion of the pedigree.

(ii) Only one individual per generation in each of the random lines is recorded.

The purpose of this paper is to present a method whereby these limitations can be overcome. Consider a pedigree which is constructed as follows:

The pedigree is compiled completely for i generations, thus, it will contain 2^i individuals in the ith generation. From here on r random lines of descent are constructed on the sire's side of the pedigree. These lines originate from r randomly selected individuals of the ith generation. In a similar way s lines are developed on the dam's side of the pedigree. Assume further that a total of u, ($u \geq r$), individuals are recorded in each generation of the random lines in the

sire's half of the pedigree, and v, $(v \geq s)$, individuals in the dam's half of the pedigree. (In a symmetrical situation it will be possible to express u and v as functions of r and s respectively, and of another variable which defines the number of individuals per random line per generation.)

The inbreeding coefficient of the pedigree is then obtained by adding the contributions of the following three sources:—

a   Where n and $n' \leq (i-1)$ the contribution of each tie is

$$(\tfrac{1}{2})^{n+n'+1}(1 + F_A).$$

b.   Where n and $n' \geq i$. The total number of individuals which is sampled in each generation beyond the completed portion of the pedigree is $(u + v)$. The nth generation of the sire contains 2^n individuals of which only u are sampled, and only v individuals of the $2^{n'}$ individuals in the n'th generation of the dam are sampled. Thus, the number of combinations which is investigated for possible ties is uv out of a possible $2^{n+n'}$ combinations. The contribution of a tie in a completed pedigree is

$$(\tfrac{1}{2})^{n+n'+1}(1 + F_A),$$

which has to be multiplied by $\dfrac{2^{n+n'}}{uv}$ in order to obtain the contribution of a tie in the randomly selected part of the pedigree. The contribution of such a tie is then,

$$(2uv)^{-1}(1 + F_A) \quad \text{..............................} \quad (1)$$

c.   Where $n \leq (i-1) < n'$. In this instance the ratio of actual to possible number of comparisons is

$$\frac{v2^n}{2^{n+n'}}$$

The contribution of each tie is thus,

$$\frac{1}{v}(\tfrac{1}{2})^{n+1}(1 + F_A) \quad \text{..........................} \quad (2)$$

Obviously, where $n' \leq (i-1) < n$ the contribution of each tie is

$$\frac{1}{u}(\tfrac{1}{2})^{n'+1}(1 + F_A).$$

In the special case which was handled by Wright and McPhee $k = i-1$ and $u = v = r = s = 2^k$. Substitution of these values into formulae (1) and (2) yields,

$$(\tfrac{1}{2})^{2k+1}(1 + F_A) \text{ and}$$
$$(\tfrac{1}{2})^{n+k+1}(1 + F_A)$$

respectively, which are identical to their formulae.

The u and v individuals referred to above can be obtained by any procedure which the researcher wishes to use. Consequently, different generations of the random lines of descent may contain different numbers of individuals which necessitates modification of formulae (1) and (2).

Let generation j, where $j > i$, contain u_j individuals in the sire's half of the pedigree where $u_j \geq r$, and v_j individuals in the dam's half where $v_j \geq s$.

a.   If n & $n' \geq i$ and the common ancestor appears in the jth generation of the sire's half of the pedigree and in the lth generation of the dam's half, the actual number of combinations is $u_j v_l$ out of a possible $2^{n+n'}$ combinations. Thus the contribution of each tie is

$$(2u_j v_l)^{-1}(1 + F_A) \quad \text{.............................} \quad (3)$$

b.   If $n \leq (i-1) < n'$ the ratio of possible to actual combinations which can produce ties when the common ancestor appears in the lth generation on the dam's side is $\dfrac{2^{n+n'}}{v_l 2^n}$. The contribution of each tie is then,

$$\frac{1}{v_l}(\tfrac{1}{2})^{n+1}(1 + F_A) \quad \text{.............................} \quad (4)$$

In the special case dealt with by Wright & McPhee $u_j = v_j = r = s = 2^k$.

On substitution of these relations into equations (3) and (4) the latter reduce to those given by Wright & McPhee (1925).

References

WRIGHT, S. 1922. Coefficients of inbreeding and relationship. *Amer. Nat.* 56, 330.

WRIGHT, S. & McPHEE, H.C. 1925. An approximate method of calculating coefficients of inbreeding and relationship from livestock pedigrees. *J. Agric. Res.* 31, 377.