Theoretical and Applied Genetics 42, 272–273 (1972)

# Computation of Genetic Contributions from Pedigrees

J. W. JAMES

School of Wool and Pastoral Sciences, The University of New South Wales, Kensington, N. S. W. (Australia)

**Summary.** A simple procedure is outlined by which genetic contributions of individuals to later generations can be estimated. The method, which involves simple matrix operations, is well suited to automatic computation.

In pedigreed populations it is possible to estimate, by analyses of the pedigrees, what proportion of the genes in the population have been derived from any specific ancestor. Study of changes in proportions of genes has proved useful in the interpretation of genetic effects of selection, as shown by James and McBride (1958), James (1962 a, b) and Jones (1969 a, b).

The proportion of genes in a population which has been derived from a particular ancestor may be computed by tracing pedigrees of all individuals, and for each individual computing the proportionate genetic contribution of each ancestor as the fraction of pedigree lines tracing back to that ancestor. For each ancestor these proportions can be averaged over all individuals to give the ancestor's proportion of genes in the whole population. This is a tedious method, especially if extended over many generations. Maruyama and Yasuda (1970) have outlined another method of calculating genetic contributions, but it is not convenient for data from selection lines. The method to be described has proved very suitable for such data.

Consider two successive generations, $t$ and $t+1$, in a population with distinct generations. Then the proportion of genes in a member of generation $t+1$ which are direct descendants of genes carried by a member of generation $t$ is $1/2$, if the individuals are parent and offspring, zero otherwise. If selfing is possible the fraction is unity for progeny produced by selfing. These are the probabilities that a random gene from the generation $t+1$ individual is directly descended from a gene carried by the generation $t$ individual. If a more formal title than proportion of genes is preferred, such probabilities could be named *coefficients of descent*. These probabilities are not directly related to co-ancestries or inbreeding coefficients and may be zero for closely related individuals, such as full sibs. Only *direct* descent is involved.

If there are $n(t)$ members of generation $t$ and $n(t+1)$ of generation $t+1$, the proportions of genes can be arrayed in a matrix $\boldsymbol{D}(t, t+1)$ with $n(t)$ rows and $n(t+1)$ columns, where the element $D_{ij}(t, t+1)$ is the probability that a random gene from the $j$th member of generation $t+1$ is descended from a gene carried by the $i$th member of generation $t$. The columns of this *descent matrix* all sum to unity.

In the same way a descent matrix $\boldsymbol{D}(t+1, t+2)$ may be defined connecting generations $t+1$ and $t+2$. Similarly, a matrix $\boldsymbol{D}(t, t+2)$ may be defined connecting generations $t$ and $t+2$, the element $D_{ij}(t, t+2)$ being the probability that a random gene from the $j$th member of generation $t+2$ is descended from a gene carried by the $i$th member of generation $t$. The chance that a random gene in the $j$th member of generation $t+2$ is descended from the $i$th member of generation $t$ through the $k$th member of generation $t+1$ is $D_{ik}(t, t+1)\, D_{kj}(t+1, t+2)$. It then follows that

$$D_{ij}(t, t+2) = \sum_{k=1}^{n(t+1)} D_{ik}(t, t+1)\, D_{kj}(t+1, t+2)\,.$$

Since this is the matrix multiplication rule, the result can be written in matrix notation as

$$\boldsymbol{D}(t, t+2) = \boldsymbol{D}(t, t+1)\, \boldsymbol{D}(t+1, t+2)\,.$$

This result may clearly be extended for as many generations as required. Thus, for a period of $u$ generations,

$$\begin{aligned}\boldsymbol{D}(t, t+u) &= \boldsymbol{D}(t, t+1)\, \boldsymbol{D}(t+1, t+2) \ldots \\ &\quad \ldots \boldsymbol{D}(t+u-1, t+u) = \\ &= \boldsymbol{D}(t, t+u-1)\, \boldsymbol{D}(t+u-1, t+u)\,.\end{aligned}$$

These matrices give the proportions of genes in individuals. To obtain proportions in the population in generation $t+u$ the rows are summed and each row sum is divided by the total of row sums. The total of row sums is clearly $n(t+u)$ since total of row sums = total of column sums = $n(t+u)$ since there are $n(t+u)$ columns each summing to unity. If $\boldsymbol{V}(t+u)$ is defined as a column vector of $n(t+u)$ elements, each of which is $1/n(t+u)$, and $\boldsymbol{G}(t, t+u)$ is defined as a column vector of $n(t)$ elements which are proportions of genes from members of generation $t$ in the population at generation $t+u$, then

$$\boldsymbol{G}(t, t+u) = \boldsymbol{D}(t, t+u)\, \boldsymbol{V}(t+u)\,.$$

A reduction in computation can be achieved by working with matings rather than individuals as units. This is especially so with equal numbers of males and females, since a matrix based on matings will then have half the numbers of rows and columns used in a matrix based on individuals. In a mating descent matrix, $D_{ij}(t, t+u)$ is the probability that a gene taken at random from mating $j$ in generation $t+u$ is descended from a gene carried by mating $i$ in generation $t$. Thus $D_{ij}(t, t+1)$ is zero if neither member of mating $j$ in generation $t+1$ is a progeny of mating $i$ in generation $t$; is one-half if one is but the other is not; and is unity if both are, i. e. mating $j$ of generation $t+1$ is of full sibs. With this definition all the preceding results apply to mating descent matrices. It should be noted that contributions from generation $t$ matings are identical with contributions from generation $t+1$ families, since genes in a generation $t+1$ family are all contributed by a generation $t$ mating.

Usually one's object is to trace changes in gene proportions from members of generation $t$ over a period of $u$ generations. One then finds $\boldsymbol{D}(t, t+1)$, next $\boldsymbol{D}(t+1, t+2)$ and hence $\boldsymbol{D}(t, t+2)$. Next one finds $\boldsymbol{D}(t+2, t+3)$ whence $\boldsymbol{D}(t, t+3) = \boldsymbol{D}(t, t+2)\,\boldsymbol{D}(t+2, t+3)$ and so on. If one is interested in which ancestors from previous generations have contributed importantly to generation $t$, the process is reversed. One begins with $\boldsymbol{D}(t-1, t)$, then $\boldsymbol{D}(t-2, t-1)$ and so finds $\boldsymbol{D}(t-2, t) = \boldsymbol{D}(t-2, t-1)\,\boldsymbol{D}(t-1, t)$, and so on.

It is normally best to use mating descent matrices as a basic method, since this reduces matrix size and computing time. If a male is mated to several females his total contribution can be found by summing over all matings involving him and dividing by 2, since he contributes only half the genes in each mating. It is of course possible to start and finish at any points in the sequence of generations.

It is often useful to compute gene proportions for some stage of the life cycle other than at mating, such as at sexual maturity before artificial selection is practised. If $\boldsymbol{W}(t+u)$ is a column vector of $n(t+u)$ elements whose $j$th element is the number of progeny of the $j$th mating of generation $t+u$ which are present at the stage in question, then the proportions of genes in the population at that stage are given by the vector $\boldsymbol{D}(t, t+u)\,\boldsymbol{W}(t+u)/w$, where $w$ is the sum of the elements in the $n(t)$ rows of the vector $\boldsymbol{D}(t, t+u)\,\boldsymbol{W}(t+u)$, or equivalently the sum of the elements of $\boldsymbol{W}(t+u)$. Several stages in the life cycle may be treated by using several $\boldsymbol{W}$ vectors.

The method can be adapted to deal with sex-linked genes. If the male is heterogametic it must get its $X$ chromosome from its dam. Hence probabilities in the $\boldsymbol{D}(t, t+1)$ matrix are 0 for sire and son, 1 for dam and son, $^1/_2$ for sire and daughter, $^1/_2$ for dam and daughter, and 0 otherwise. If one were dealing with sex linked genes, sexes would probably be treated separately and analysis with matings as units would not be considered.

It is possible to use this procedure with populations in which generations overlap, though perhaps at the expense of very large matrices. An individual present at both times $t$ and $t+1$ would be included in both rows and columns of the $\boldsymbol{D}(t, t+1)$ matrix, with a probability of unity for its coefficient of descent from itself.

The construction of the $\boldsymbol{D}(t, t+1)$ matrices from records of parents of individuals involved, and the other matrix operations, are easily set up in a computer program.

## Literature

1. James, J. W.: The spread of genes in random mating control populations. Genetic. Res. **3**, 1–10 (1962a). — 2. James, J. W.: The spread of genes in populations under selection. Proc. 12th Wld's Poult. Congr. Sydney, pp. 14–16 (1962b). — 3. James, J. W., McBride, G.: The spread of genes by natural and artificial selection in a closed poultry flock. J. Genet. **56**, 55–62 (1958). — 4. Jones, L. P.: Effects of artificial selection on rates of inbreeding in populations of *Drosophila melanogaster*. I. Effect in early generations. Aust. J. biol. Sci. **22**, 143–155 (1969a). — 5. Jones, L. P.: Effects of artificial selection on rates of inbreeding in populations of *Drosophila melanogaster*. II. Effect of previous selection on rates of inbreeding. Aust. J. biol. Sci. **22**, 157–169 (1969b). — 6. Maruyama, T., Yasuda, N.: Use of graph theory in computation of inbreeding and kinship coefficients. Biometrics **26**, 209–219 (1970).

Received November 29, 1971
Communicated by J. S. F. Barker

J. W. James
School of Wool and Pastoral Sciences
The University of New South Wales
Kensington, N. S. W. 2033 (Australia)