Inbreeding and Kinship in Pedigreed Populations

Jessica Rutkoski

October 17, 2016

1 Inbreeding coefficients of individuals

The coefficient of inbreeding (F), which is the probability of any individual having two IBD alleles at a particular locus, can be estimated from the size of the breeding population, and the particular breeding structure (close matings excluded, unequal numbers of males and females, etc.). This F is the average inbreeding coefficient of all individuals of a generation. When populations are small, it is more accurate and more convenient to deduce the rate of inbreeding using pedigrees. When using pedigrees to estimate inbreeding rates, it is not necessarily a reflection of the random genetic drift which is occurring. Processes (such as selection) that may counteract drift must not be occurring.

The coefficient of inbreeding of an individual is the probability that the pair of gametes that gave rise to that individual carried alleles (at a particular locus) that were IBD. Thus, computing the inbreeding coefficients requires tracing the pedigree back to common ancestors of the parents and computing the probabilies of transmitting alleles from the common ancestor at each segregation.

As a simple example, lets look at calculating the inbreeding coefficient of an individual with a simple pedigree (Figure 1)

X is the individual whose inbreeding coefficient we would like to compute. The parents of X are P and Q. P and Q are related through their common parent A. They are not related in any other way, so we only have to consider the transmission of alleles from A, to P and Q, and then to X in order to calculate the probability IBD in X. Figure 1: Simple pedigree representing a mating between half-sibs

Lets think about one bi-allelic locus with two alleles A_1 and A_2 . We assume that A, the common ancestor of P and Q, is not inbred, thus its genotype is A_1A_2 . First lets think about the probability that A_1 is transmitted to X from both parents. Remember that to calculate the probability of more than one event occurring, we multiply the probabilities that each event occurs on its own.

The probability that X receives A_1 from A via P, is the probability that A passes A_1 to P multiplied by the probability that P passes A_1 to X. Because at each generation, parents transmit half of their alleles to their progeny. This probability is 1=2 1=2 = 1=4. The probability that X receives A_1 from A via Q is the same as above, the probability that A passes A_1 to Q multiplied by the probability that Q passes A_1 to X, 1=2 1=2 = 1=4. Next, we need to know the probability that X receives A_1 from both P and Q. To do this we multiple the probabilities that X received A_1 via P and that X received A_1 via Q. 1=4 1=4 = 1=16

Now we know the probability that A_1 is IBD in X. However, there are two alleles, A_1 and A_2 . Thus, X could also be IBD by receiving two copies A_2 . Remember that in order to get the probability that either event 1 or

event 2 occurs, we take the sum of the probabilities of the two events. Thus, the probability of IBD in X is 1/16 + 1/16 = 2/16 = 1/8.

In order to define a general rule for calculating the probability of IBD in X. Lets look at this problem again. From before we figured out that the probability that A_1 is IBD is $(1/2)^4$, because A_1 must move from A to P, P to X and then from A to Q, Q to X. Likewise the probability that A_2 is IBD is $(1/2)^4$. Thus the probability that either A_1 or A_2 are IBD is $2(1/2)^4 = \frac{2}{2\cdot 2\cdot 2\cdot 2} = \frac{1}{2\cdot 2\cdot 2} = (1/2)^3$. Notice that there are 3 individuals in the path connecting the parents of X to their common ancestor (these are P, A,and Q). We can therefore use this general rule to estimate the inbreeding coefficients by counting the number of individuals in the path.

This probability of IBD in X represents the new inbreeding arising from A as a common ancestor of P and Q. However, the common ancestor, A may itself be IBD at a given locus. The probability that A is IBD at a given locus is simply its inbreeding coefficient. If the inbreeding coefficient of A is greater than zero, there is now an additional probability that X is IBD due to previous inbreeding. This is the probability that A is IBD multiplied by the probability of transmission to X $(1/2)^3 \cdot F_A$ where F_A is the inbreeding coefficient of the common ancestor. Now there are two ways inbreeding can occur, one is from 'new' inbreeding where the common ancestor is not IBD, and the other is from 'old' inbreeding where the common ancestor is IBD. To get the total probability that X is IBD we sum the two probabilities. Thus, $F_X = (1/2)^3 + (1/2)^3 F_A = (1/2)^3(1 + F_A)$. Note again that 3 in our case is the number of individuals in the path connecting the parents of X to the common ancestor.

In more complex pedigrees, parents may be related to each other through more than one common ancestor, or from the same common ancestor, but through different paths. Earlier we defined the probability that X is IBD through one path: $F_X = (1/2)^3 + (1/2)^3 F_A = (1/2)^3 (1 + F_A)$. In the case where there are multiple paths, we simply sum the probabilities of X being IBD across all the possible paths. This gives us the general formula:

$$F_X = \sum (1/2)^n (1 + F_A)$$
(1)

where n is the number of individuals in any path of relationship counting the parents of X, and all individuals in the path connecting the parents to the common ancestor. The summation is over all paths.

Previously, we discussed the concept of the 'base population', in the case where we have pedigrees, the 'base population' is simply the individuals at





the head of the pedigree, whose ancestry further back is not known. We generally assume that the inbreeding coefficient of these individuals is zero.

2 Example: Using pedigree to estimate an inbreeding coefficient

To illustrate how we can use equation 1 to estimate the inbreeding coefficients, lets go through an example of a more complex pedigree (figure 2).

The first step is to identify the common ancestors. These are A and E. The next step is to identify all the paths of relationship, for example (figure 3).



Figure 3: More complex pedigree, identification of common ancestors and paths of relationship

Table 1: Calculating the inbreeding coefficient of X

		<u> </u>	
Paths	n	F of common ancestor	Contribution of F_X
KHDBACFJL	9	0	$(1/2)^9 = 0.002$
$\mathrm{KH}\mathbf{E}\mathrm{IL}$	5	0	$(1/2)^5 = 0.031$
			Total = 0.033

In this example A and E are the common ancestors and there are two paths of relationship.

3 Kinship

When pedigrees are longer and more complicated, it is not practicable to trace all the paths of relationship. Another way to calculate the inbreeding coefficient is to keep track of it at each generation and to shift to thinking about coefficient of kinship between the and individuals' two parents. This is called the tabular method, and it does not differ in princple from the previous method of tracing the paths. The coefficient kinship between two individuals is defined as the probability that two gametes taken at random (one from each individual carry alleles that are IBD. The coefficient of kinship (f) between two individuals is equal to the inbreeding coefficient of their progeny if they were to be mated:

$$F_X = f_{KL} \tag{2}$$

where X is the progeny and K and L are the two parents. Coefficients of kinship are used for planning matings that give rise to the least inbreeding, and they are also used in mixed model analysis for the estimation of breeding value (this will be discussed later on.)

There are general rules that can be used calculate coefficients of kinship generation by generation. Considering the general pedigree in figure 4.

The most basic rule is that the coefficient of kinship between P and Q is simply the mean of the four co-ancestries AC, AD, BC, and BD.

$$f_{PQ} = \frac{1}{4}f_{AC} + \frac{1}{4}f_{AD} + \frac{1}{4}f_{BC} + \frac{1}{4}f_{BD}$$
(3)

This is because the two gametes sampled, one from P and one from Q, will carry alleles from A and C in $1/2 \cdot 1/2 = 1/4$ of the cases, likewise for AD, AD, BC, and BD.

Figure 4: General pedigree



Self Another basic rule states that the coefficient of kinship of an individual (A) with itself f_{AA} is the inbreeding coefficient of progeny that would be produced by self-mating. This is equal to $1/2 + 1/2F_A$ because the probability that the two gametes sampled are either both A_1 or both A_2 is 1/2, and the probability that one is A_1 and the other is A_2 is 1/2, but the probability that A_1 is IBD with A_2 is the coefficient of inbreeding of A thus,

$$f_{AA} = \frac{1}{2}(1 + F_A) \tag{4}$$

Offspring and parent A third basic rule states the the coefficient of kinship between parent and offspring f_{PA} is the mean coefficient of kinship between A and both the parents of P, (A and B), which is

$$f_{PA} = \frac{1}{2}(f_{AB} + f_{AA})$$
(5)

There are several other rules for other specific cases. We will not take the time to go over these because in practice coefficients of kinship and inbreeding are not calculated manually, but rather using software packages.