CHAP 2: RELATED INDIVIDUALS: ONE LOCUS
2.1 PEDIGREES AND RELATIONSHIPS

2.1.1 TERMINOLOGY

Founders and non-founders; founders have no parents specified. They are assumed unrelated.

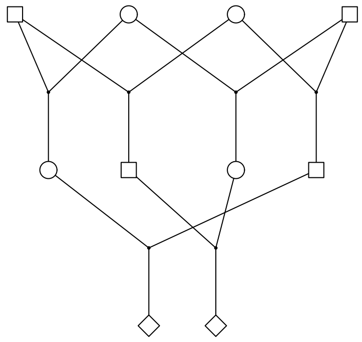
Related : individuals having a common ancestor (implies a biological relationship)

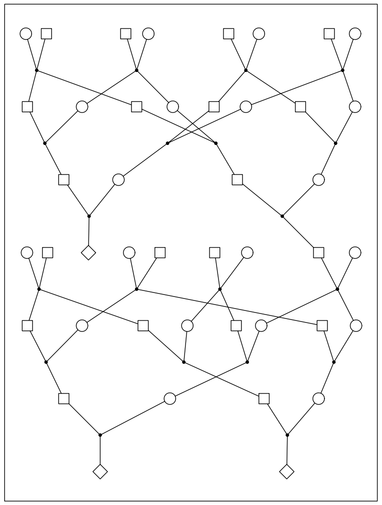
Inbred: individuals whose parents are related (implies the maternal and paternal genes can descend from single ancestral gene).

Unilateral (one-sided) and bilateral (two-sided) relationships:
unilateral: half-sibs, aunt, niece, cousins
bilateral: sibs, double first cousins, etc.

Cousin-type relationships:
Half, full, and double (cousin) relationships
Cousins of different degree: n^{th} cousins k times removed

Complex relationships:
quadruple half first cousins; quadruple second cousins





2.1.2 GENE IDENTITY BY DESCENT (*ibd*)

RELATIVES ARE SIMILAR because they have *ibd* genes, that are copies of the same gene in a common ancestor.
NOTE: *ibd* is defined relative to given pedigree or time point

***ibd* genes are of the same allelic type, non-*ibd* genes are of independent types.**
(See the mother-baby pairs in 1.3.6.)

A pedigree or relationship determines probabilities of *ibd*,
which determine probabilities of joint genotypes
which determine probabilities of joint phenotypes
that is, similarity among relatives.

2.2.1 KINSHIP and INBREEDING

The simplest pedigree-defined probabilities of gene *ibd* are the coefficients of kinship (ψ) and inbreeding (f), which measure *ibd* between two genes.

$$\begin{aligned}\psi(B,C) &= \Pr(\mathit{homologous\ genes\ segregating}\\ &\quad \mathit{from\ B\ and\ C\ are\ ibd})\\ f(B) &= \Pr(\mathit{homologous\ genes\ in\ B\ are\ ibd})\\ &= \psi(M_B, F_B)\end{aligned}$$

where M_B and F_B are the parents of B.

2.2.2 EXAMPLES OF PATH COUNTING

Half sibs: $(1/2) \times (1/2) \times (1/2) \ = \ 1/8$

Two genes from an inbred (f) parent:
$1 \times f + (1/2) \times (1-f) \ = \ (1/2)(1+f)$

Half sibs with inbred (f) parent: $(1+f)/8$

Full sibs: $1/8 + 1/8 = 1/4$

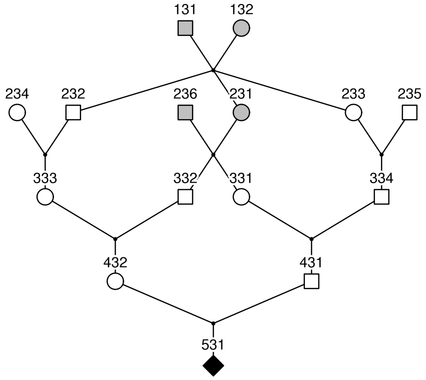
First cousins: $(1/4) \times (1/2) \times (1/2) \ = \ (1/16)$

Double first cousins: $1/16 \ + \ 1/16 = 1/8$

General formula (Wright, 1922):

$$\psi = \sum_A \sum_{\mathcal{P}(A)} \frac{1}{2}(1+f_A)\left(\frac{1}{2}\right)^{n_1(\mathcal{P}(A))+n_2(\mathcal{P}(A))}$$

EXAMPLE: The JV pedigree (Goddard et al. 1996)



2 ancestors, each with 3 paths, each with $n_1 = n_2 = 3$:
and 2 ancestors, each with 1 path, each with $n_1 = n_2 = 2$.
$2 \times 3 \times (\frac{1}{2})^7 + 2 \times 1 \times (\frac{1}{2})^5 = 7/64$

2.2.3 RECURSIVE METHOD

$$\psi(B,C) \;=\; \frac{1}{2}(\psi(M_B,C)+\psi(F_B,C))$$

provided B is not C nor an ancestor of C

$$\psi(B,B) \;=\; \frac{1}{2}(1+f_B) \quad = \quad \frac{1}{2}(1+\psi(M_B,F_B))$$

Boundary conditions:

$$\psi(A,A) \quad = \quad \frac{1}{2} \;\text{ and }\; \psi(A,C) \;=\; 0$$

if A is a founder, and not an ancestor of C

Expanding up the JV pedigree, among the grandparents, we have 3 first-cousin pairs and a sib pair. The kinship of first cousins is 1/16, and of sibs is 1/4, so overall we have

$$\frac{1}{4}(3\frac{1}{16}+\frac{1}{4}) \;=\; \frac{7}{64}$$

2.2.4
INBREEDING and GENOTYPE FREQUENCIES

MYTHS: Recessive diseases are more frequent in genetic isolates. This is because isolates are ”more inbred”

TRUTH: the more inbred individuals within any population have higher probability of homozygosity.

Consider a recessive allele a with freq q, and an individual with inbreeding coefficient f

$$\begin{aligned}\Pr(aa) &= q^2(1-f)+qf \quad = \quad q^2+fq(1-q)\\ \Pr(Aa) &= 2q(1-q)(1-f)\\ \Pr(AA) &= (1-q)^2+fq(1-q)\end{aligned}$$

See population mixtures and Wahlund variance

In population subdivision, people marry those more similar, hence more homozygosity in offspring.

In inbreeding, people marry relatives, and hence more similar, and hence ...

Inbreeding is a form of population subdivision.

Autozygous ≡ having *ibd* genes

inbred ≡ having non-zero prob of being autozygous

$$\Pr(\text{IBD} \mid \text{affected}) \;=\; \frac{qf}{q^2+fq(1-q)} \;=\; \frac{f}{q+f(1-q)}$$

TRUTH2: the affected people in a population have higher probability of being inbred.

Suppose a proportion α of the population (Pop_1) has inbreeding coefficient f and others (Pop_2) are not inbred:

$$\begin{aligned}
\Pr(\text{affected } aa) &= (1-\alpha)q^2 \; + \; \alpha(q^2 + fq(1-q)) \\
&= q^2 + \alpha f q(1-q) \\
\Pr(Pop_1 \mid \text{affected}) &= \frac{\alpha(q^2 + fq(1-q))}{q^2 + \alpha f q(1-q)} \\
&= \frac{\alpha(q + f - fq)}{(q + \alpha f - \alpha f q)}
\end{aligned}$$

which is always $\geq \alpha$ and $\rightarrow 1$ as $q \rightarrow 0$.

TRUTH3: The affected inbred people in a population have higher probability of being autozygous.

$$\Pr(\text{autozyg} \mid \text{affected}) = \frac{\alpha f q}{q + \alpha f(1-q)}$$

Same form as before with f now becoming αf.

2.3.1 *ibd* OF MORE THAN TWO GENES

Label $2k$ genes of k individuals sucessively, giving each the label previously assigned to genes to which it is *ibd*, and otherwise the next available integer.

1 2 1 3 4 4 1 5**: the paternal genes of individuals 1,2,4 are *ibd* and the two genes of individual 3 are *ibd*.**
Reduce to genotypically equivalent classes of states:
1 2 1 3 4 4 1 5 ≡ 1 2 3 1 4 4 1 5 ≡ 1 2 3 1 4 4 5 1 ≡
1 2 1 3 4 4 5 1 ≡ 1 2 2 3 4 4 2 5 ≡ 1 2 3 2 4 4 2 5 ≡
1 2 3 2 4 4 5 2 ≡ 1 2 2 3 4 4 5 2
Note that when the two genes of the first individual are interchanged, we must relabel the genes $1 \leftrightarrow 2$, to obtain a legal state label.

The case of 4 genes of two individuals is shown in the Table: there are 15 states and 9 state classes.

For 12 genes in 6 individuals there are more than 4 million states, but only about 198,000 state classes.

2.3.2 Table for two individuals

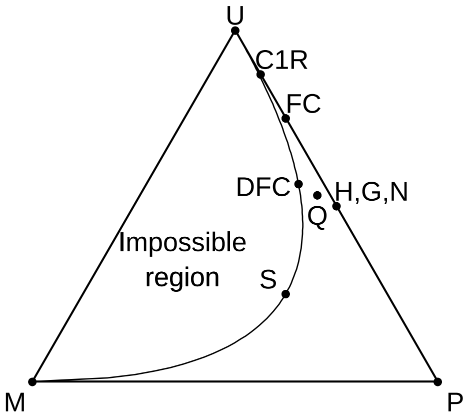
ibd pattern				*ibd* label	*ibd* group	state description	
B_1		B_2				individuals	genes
p	m	p	m			autozygous	shared
•	•	•	•	1 1 1 1	1 1 1 1	B_1, B_2	4 genes *ibd*
•	•	•	○	1 1 1 2	1 1 1 2	B_1	3 genes *ibd*
•	•	○	•	1 1 2 1			
•	○	•	•	1 2 1 1	1 2 1 1	B_2	3 genes *ibd*
•	○	○	○	1 2 2 2			
•	•	○	○	1 1 2 2	1 1 2 2	B_1, B_2	none
•	•	○	†	1 1 2 3	1 1 2 3	B_1	none
•	○	†	†	1 2 3 3	1 2 3 3	B_2	none
•	○	•	○	1 2 1 2	1 2 1 2	none	2 genes
•	○	○	•	1 2 2 1			shared
•	○	•	†	1 2 1 3	1 2 1 3	none	1 gene
•	○	†	•	1 2 3 1			shared
•	○	○	†	1 2 2 3			
•	○	†	○	1 2 3 2			
•	○	†	⋆	1 2 3 4	1 2 3 4	none	none

2.3.3 RELATIONSHIPS BETWEEN TWO NON-INBRED RELATIVES

For two non-inbred relatives, 7 states, 3 classes, 2 probs $\kappa_i = \Pr(i \text{ genes } ibd)$, $\kappa_2 + \kappa_1 + \kappa_0 = 1$.

Pairwise relationship	κ_0	κ_1	κ_2	ψ
Unrelated	1.00	0	0	0
Parent-offspring	0	1.00	0	0.25
Monozygous twin	0	0	1.00	0.50
Full Sib	0.25	0.50	0.25	0.25
Half sib, grandparent, aunt	0.50	0.50	0.00	0.125
First cousin	0.75	0.25	0	0.0625
Double first cousin	0.5625	0.375	0.0625	0.125
Quadruple half first cousin	0.5312	0.4375	0.0312	0.125

Relationships may be represented as points in an equilateral triangle.

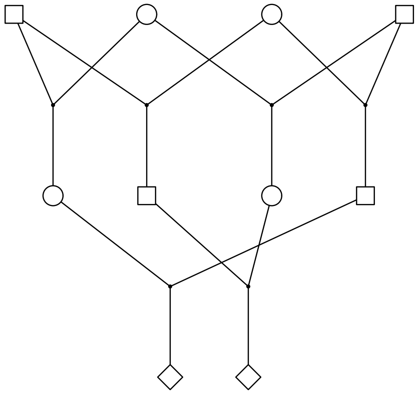


The following equations relate ψ and κ_i, $i = 0, 1, 2$.

$$\begin{aligned}
\psi &= \frac{1}{2}\kappa_2 + \frac{1}{4}\kappa_1 \quad = \quad \frac{1}{4}(1 + \kappa_2 - \kappa_0) \\
\psi(B_1, B_2) &= (1/4)(\psi(M_1, M_2) + \psi(M_1, F_2) + \psi(F_1, M_2) + \psi(F_1, F_2)) \\
\kappa_2(B_1, B_2) &= \psi(M_1, M_2)\psi(F_1, F_2) + \psi(M_1, F_2)\psi(F_1, M_2) \\
\kappa_1(B_1, B_2) &= 4\psi(B_1, B_2) - 2\kappa_2(B_1, B_2) \\
\kappa_0(B_1, B_2) &= 1 - \kappa_1(B_1, B_2) - \kappa_2(B_1, B_2)
\end{aligned}$$

Applying the Arithmetic-Geometric means inequality to these same equations shows $\kappa_1^2 \geq 4\kappa_0\kappa_2$ for all real relationships. (See book, P.38, for details.)

2.3.4 EXAMPLE OF QUAD HALF FIRST COUSINS:



It is possible for all four of $\psi(M_1, M_2)$, $\psi(F_1, F_2)$, $\psi(M_1, F_2)$ and $\psi(F_1, M_2)$ to be non-zero without the children being inbred.

That is, each of the mother and the father of each child is related to both the mother and the father of the other. But, for each child, the mother is not related to the father.

For QHFC,
$\psi(M_1, M_2) = \psi(F_1, F_2) = \psi(M_1, F_2) = \psi(F_1, M_2) = 1/8$
so $\kappa_2 = 1/32$, $\psi = 1/8$, $\kappa_1 = 4\psi - 2\kappa_2 = 7/16$,
$\kappa_0 = 1 - \kappa_2 - \kappa_1 = 17/32$

2.4 DATA ON RELATIVES

2.4.1 SPECIFYING INHERITANCE

Segregation* of genes is fully specified by *meiosis indicators

$$\begin{aligned} S_i &= 0 \quad \text{if gene is parent's maternal gene} \\ &= 1 \quad \text{if gene is parent's paternal gene} \end{aligned}$$

where $i = 1, ..., m$ indexes the meioses.

S_i are i.i.d with

$$\Pr(S_i = 0) = \Pr(S_i = 1) = \frac{1}{2}.$$

ibd **at a locus is a function of the $\{S_i\}$ at that locus.**

2.4.2 The general formula:

$$\begin{aligned}\Pr(\mathbf{Y}) &= \sum_{\mathbf{S}} \Pr(\mathbf{Y} \mid \mathbf{S}) \Pr(\mathbf{S}) \\ &= \sum_{\mathbf{S}} \Pr(\mathbf{Y} \mid \mathbf{J}(\mathbf{S})) \Pr(\mathbf{S}) \\ &= \sum_{\mathbf{J}} \Pr(\mathbf{Y} \mid \mathbf{J}) \Pr(\mathbf{J})\end{aligned}$$

$\Pr(\mathbf{Y} \mid \mathbf{J}(\mathbf{S}))$ **is the sum over all possible assignments** $\mathcal{A}$ **of allelic types to genes of the product of allele frequencies** $q_{a(k)}$ **of assigned alleles** $a(k)$**:**

$$\Pr(\mathbf{Y} \mid \mathbf{J}(\mathbf{S})) = \sum_{\mathcal{A}} \prod_{k} q_{a(k)}.$$

EXAMPLE: DATA ON 1 INDIVIDUAL

Suppose we observe someone who is A_1A_1

$$\begin{aligned}\mathbf{J} &= (I, N), \qquad \Pr(I) = f,\ \Pr(N) = 1 - f, \\ \Pr(A_1A_1) &= \Pr(A_1A_1|I)f \ + \ \Pr(A_1A_1|N)(1 - f) \\ &= qf + q^2(1 - f) \quad = \quad q(f + q - qf)\end{aligned}$$

EXAMPLE: DATA ON TWO INDIVIDUALS

We know the relationship between two individuals, so can (we suppose) compute the probabilities $\Delta_1, ..., \Delta_9$ of the 9 IBD classes (groups of states). Suppose we observe the individuals to be AA and AC.

$P(\mathbf{J})$	$\mathbf{J}$	$P(AA, AC \mid \mathbf{J})$
Δ_1	**1 1 1 1**	**0**
Δ_2	**1 1 1 2**	$q_A q_C$
Δ_3	**1 2 1 1**	**0**
Δ_4	**1 1 2 2**	**0**
Δ_5	**1 1 2 3**	$q_A(2q_Aq_C)$
Δ_6	**1 2 3 3**	**0**
Δ_7	**1 2 1 2**	**0**
Δ_8	**1 2 1 3**	$q_A q_A q_C$
Δ_9	**1 2 3 4**	$q_A^2.(2q_Aq_C)$

Total probability of observing (AA, AC) is

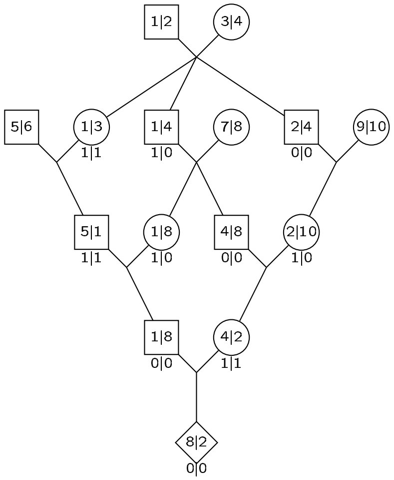
$$\begin{aligned} P(AA, AC) &= \sum_{k=1}^{9} \Delta_k P(AA, AC \mid \mathbf{J} = i) \\ &= \Delta_2 q_A q_C \;+\; \Delta_5 2q_A^2 q_C \;+\; \Delta_8 q_A^2 q_C + \Delta_9 2 q_A^3 q_C \end{aligned}$$

2.4.3 DATA ON A NON-INBRED PAIR

$$\begin{aligned} \Pr(G_1, G_2; R) &= \sum_{\mathbf{J}} \Pr(\mathbf{Y} \mid \mathbf{J}) \Pr(\mathbf{J}; R) \\ &= \kappa_0(R)\Pr(G_1, G_2|J_0) \;+ \\ &\quad \kappa_1(R)\Pr(G_1, G_2|J_1) \;+\; \kappa_2(R)\Pr(G_1, G_2|J_2) \\ &= \kappa_0(R)\Pr(G_1, G_2|\mathrm{Unrel}) \\ &\qquad + \;\kappa_1(R)\Pr(G_1, G_2|\mathrm{Par-offsp}) \\ &\qquad + \;\kappa_2(R)\Pr(G_1, G_2|\mathrm{MZ-twins}) \\ &= \kappa_0(R)\Pr(G_1)\Pr(G_2) \;+ \\ &\qquad \kappa_1(R)\Pr(G_1)\Pr(kid = G_2|par = G_1) \;+ \\ &\qquad \kappa_2(R)\Pr(G_1)I(G_2 \equiv G_1) \end{aligned}$$

2.4.4 Example showing the general formula

Consider the following segregation pattern of genes:



Consider the possible allelic types of these genes given the genotypes of 5 individuals shown