### 2.3.1 ibd OF FOUR GENES IN TWO INDIVIDUALS:

<table>
<thead>
<tr>
<th>ibd pattern</th>
<th>ibd label</th>
<th>ibd group</th>
<th>state description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_1$ $B_2$</td>
<td>$p m$</td>
<td>$p m$</td>
<td>individuals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>autozygous</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>genes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>shared</td>
</tr>
<tr>
<td>• • • •</td>
<td>1 1 1 1</td>
<td>1 1 1 1</td>
<td>$B_1, B_2$</td>
</tr>
<tr>
<td>• • • o</td>
<td>1 1 1 2</td>
<td>1 1 1 2</td>
<td>$B_1$</td>
</tr>
<tr>
<td>• • o •</td>
<td>1 1 2 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• o • •</td>
<td>1 2 1 1</td>
<td>1 2 1 1</td>
<td>$B_2$</td>
</tr>
<tr>
<td>• o o o</td>
<td>1 2 2 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• • • •</td>
<td>1 1 2 2</td>
<td>1 1 2 2</td>
<td>$B_1, B_2$</td>
</tr>
<tr>
<td>• • o †</td>
<td>1 1 2 3</td>
<td>1 1 2 3</td>
<td>$B_1$</td>
</tr>
<tr>
<td>• o † †</td>
<td>1 2 3 3</td>
<td>1 2 3 3</td>
<td>$B_2$</td>
</tr>
<tr>
<td>• o • o</td>
<td>1 2 1 2</td>
<td>1 2 1 2</td>
<td>none</td>
</tr>
<tr>
<td>• o o •</td>
<td>1 2 2 1</td>
<td></td>
<td>shared</td>
</tr>
<tr>
<td>• o • †</td>
<td>1 2 1 3</td>
<td>1 2 1 3</td>
<td>none</td>
</tr>
<tr>
<td>• o † •</td>
<td>1 2 3 1</td>
<td></td>
<td>shared</td>
</tr>
<tr>
<td>• o † o</td>
<td>1 2 2 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• o † *</td>
<td>1 2 3 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• o † *</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>none</td>
</tr>
</tbody>
</table>

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.

---

### 2.3.2 ibd OF ANY NUMBER OF GENES:

Label $2k$ genes of $k$ individuals successively, 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 $\equiv$ 1 2 3 1 4 4 1 5 $\equiv$ 1 2 3 1 4 4 5 1 $\equiv$

1 2 1 3 4 4 5 1 $\equiv$ 1 2 3 2 4 4 2 5 $\equiv$ 1 2 3 2 4 4 2 5 $\equiv$

1 2 3 2 4 4 5 2 $\equiv$ 1 2 3 2 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 of 2.3.1: 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 (Thompson, 1974).
- For the computers of 1974, even 198,000 was large, but possible.
2.3.3 *Ibd* OF TWO NON-INBRED RELATIVES:

For two non-inbred relatives, 7 states, 3 classes, 2 probs

\[
\kappa_i = \Pr(\text{i genes } \text{ibd}), \quad \kappa_2 + \kappa_1 + \kappa_0 = 1
\]

\[
\psi = \frac{1}{2}\kappa_2 + \frac{1}{4}\kappa_1 + 0\kappa_0 = \frac{1}{4}(2\kappa_2 + \kappa_1)
\]

If \( \kappa_2 = 0 \), \( \kappa_1 = 4\psi \).

<table>
<thead>
<tr>
<th>Pairwise relationship</th>
<th>( \kappa_0 )</th>
<th>( \kappa_1 )</th>
<th>( \kappa_2 )</th>
<th>( \psi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrelated (U)</td>
<td>1.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Parent-offspring (P)</td>
<td>0</td>
<td>1.00</td>
<td>0</td>
<td>0.25</td>
</tr>
<tr>
<td>Monozygous twin (M)</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>Half sib (H), grandad (G), aunt (N)</td>
<td>0.50</td>
<td>0.50</td>
<td>0.00</td>
<td>0.125</td>
</tr>
<tr>
<td>Full Sib (S)</td>
<td>0.25</td>
<td>0.50</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>First cousin (FC)</td>
<td>0.75</td>
<td>0.25</td>
<td>0</td>
<td>0.0625</td>
</tr>
<tr>
<td>Double first cousin (DFC)</td>
<td>0.5625</td>
<td>0.375</td>
<td>0.0625</td>
<td>0.125</td>
</tr>
<tr>
<td>QHFC (Q)</td>
<td>0.5312</td>
<td>0.4375</td>
<td>0.0312</td>
<td>0.125</td>
</tr>
</tbody>
</table>

2.3.4 THE RELATIONSHIP TRIANGLE:

- Three numbers that sum to 1 \((\kappa_2 + \kappa_1 + \kappa_0 = 1)\) can be represented as a point in an equilateral triangle of unit height.

- \( \kappa_i \) is the perpendicular distance from the side \( \kappa_i = 0 \), opposite the vertex \( \kappa_i = 1, i = 0, 1, 2 \)

By direct computation of expressions using the equations on next page we can show

\[ \kappa_1^2 \geq 4\kappa_0\kappa_2 \]

for all real relationships.
2.3.5 COMPUTATION OF \( \text{ibd} \) PROBABILITIES:

- The following equations relate \( \psi \) and \( \kappa_i, i = 0, 1, 2 \).
  
  \[
  \psi = (1/2)\kappa_2 + (1/4)\kappa_1 = (1/4)(1 + \kappa_2 - \kappa_0)
  \]
  
  \[
  \psi = (1/4)(\psi_{mm} + \psi_{mf} + \psi_{fm} + \psi_{ff})
  \]
  
  \[
  \kappa_2 = \psi_{mm}\psi_{ff} + \psi_{mf}\psi_{fm} \quad \text{— new equation}
  \]
  
  \[
  \kappa_1 = 4\psi - 2\kappa_2, \quad \kappa_0 = 1 - \kappa_1 - \kappa_2
  \]

- Example: double first cousins:
  
  \( \psi_{mm} = \psi_{ff} = 1/4 \) and \( \psi_{mf} = \psi_{fm} = 0 \) or vv.
  
  \( \kappa_2 = 1/16, \psi = 1/8, \text{ so } \kappa_1 = 3/8, \kappa_0 = 9/16. \)

- The inequality:
  
  \[
  4\kappa_2 \leq 4\psi_{mm}\psi_{ff} + 4\psi_{mf}\psi_{fm} \leq (\psi_{mm} + \psi_{ff})^2 + (\psi_{mf} + \psi_{fm})^2 \\
  \leq (\psi_{mm} + \psi_{ff} + \psi_{mf} + \psi_{fm})^2 = (4\psi)^2 = (2\kappa_2 + \kappa_1)^2 \\
  4\kappa_2 \leq 4\kappa_2(\kappa_2 + \kappa_1) + \kappa_1^2 \quad \text{or } 4\kappa_2(1 - \kappa_2 - \kappa_1) \leq \kappa_1^2
  \]

  So \( 4\kappa_2\kappa_0 \leq \kappa_1^2 \)

  with equality if and only if \( \psi_{mm} = \psi_{ff} \) and \( \psi_{mf} = \psi_{fm} = 0 \) or vv.

2.3.6 DOUBLE COUSINS & QUAD HALF COUSINS:

- Each shares 1/4 of her maternal and of her paternal genome \( \text{ibd} \) with the other individual (on average).
- For QHFC, each of the mom and dad of each individual is related to both the mom and the dad of the other individual, but mom is not related to dad.
- For DFC, probability of sharing maternal and paternal genome \( \text{ibd} \) with the other individual is \( (1/4) \times (1/4) = 1/16. \)
  For QHFC this is 1/32.
2.3.7 EXAMPLE OF QUAD HALF FIRST COUSINS:

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.

Then all four of $\psi(M_1, M_2)$, $\psi(F_1, F_2)$, $\psi(M_1, F_2)$ and $\psi(F_1, M_2)$ are non-zero without the children being inbred.

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 $\kappa_2 = 1/32$, $\psi = 1/8$, $\kappa_1 = 4\psi - 2\kappa_2 = 7/16$, $\kappa_0 = 1 - \kappa_2 - \kappa_1 = 17/32$

2.4.1 DATA ON NON-INBRED RELATIVES:

• IDEA: given relationship $\mathcal{R}$, $\Pr(Y|\mathcal{R}) = \sum_J \Pr(Y|J) \Pr(J|\mathcal{R})$ where $J$ are all possible relevant patterns of $ibd$.

• EXAMPLE: one individual; 2 genes; 2 states—$ibd$ or not;

$J = (I, N)$, $\Pr(I) = f$, $\Pr(N) = 1 - f$,

$\Pr(AA) = \Pr(AA|I)f + \Pr(AA|N)(1 - f) = qf + q^2(1 - f) = q^2 + fq(1 - q)$

• EXAMPLE: two non-inbred individuals; 3 states — 2, 1, or 0 $ibd$

$Y = (G_1, G_2) = $ data on $B_1, B_2$. $\mathcal{R} = $ relationship: $\Pr(Y|\mathcal{R})$

$= \kappa_0(\mathcal{R})\Pr(Y|J_0) + \kappa_1(\mathcal{R})\Pr(Y|J_1) + \kappa_2(\mathcal{R})\Pr(Y|J_2)$

$= \kappa_0(\mathcal{R})\Pr(Y|Unrel) + \kappa_1(\mathcal{R})\Pr(Y|Par -- offsp)$

$\quad + \kappa_2(\mathcal{R})\Pr(Y|MZ -- twins)$

$= \kappa_0(\mathcal{R})\Pr(G_1)\Pr(G_2) + \kappa_1(\mathcal{R})\Pr(G_1)\Pr(kid = G_2|par = G_1) + \kappa_2(\mathcal{R})\Pr(G_1)I(G_2 \equiv G_1)$
2.4.2 PARENT-OFFSPRING PROBABILITIES:

- Offspring should share allele with parent; provided there are no typing errors.

- Probabilities $P_{r}(\text{child}|\text{parent})$: any number of alleles

<table>
<thead>
<tr>
<th>parent genotype</th>
<th>$A_iA_i$</th>
<th>$A_iA_j$</th>
<th>$A_iA_k$</th>
<th>$A_jA_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_iA_i$</td>
<td>$p_i^2$</td>
<td>$\frac{1}{2}p_i$</td>
<td>$\frac{1}{2}(p_i + p_j)$</td>
<td>$\frac{1}{2}p_k$</td>
</tr>
<tr>
<td>$A_iA_j$</td>
<td>$\frac{1}{2}p_i$</td>
<td>$\frac{1}{2}(p_i + p_j)$</td>
<td>$\frac{1}{2}p_k$</td>
<td>$\frac{1}{2}p_k$</td>
</tr>
</tbody>
</table>

- For markers with just 2 alleles:

<table>
<thead>
<tr>
<th>parent geno.</th>
<th>$P_{r}(\text{parent, child})$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>$q^3$</td>
</tr>
<tr>
<td>AB</td>
<td>$q^2(1 - q)$</td>
</tr>
<tr>
<td>BB</td>
<td>$q(1 - q)^2$</td>
</tr>
</tbody>
</table>

Data count

<table>
<thead>
<tr>
<th>child genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
</tr>
<tr>
<td>$n_{00}$</td>
</tr>
<tr>
<td>$n_{10}$</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

2.4.3 ESTIMATING $q$ FROM DATA ON RELATIVES:

For simplicity we consider just mother-baby pairs and assume HWE.

$$\ell = \sum_{(i,j)} n_{ij} \log P_{r}(G_i, G_j)$$

$$= n_{00} \log(q^3) + n_{01} \log(q^2(1 - q)) + n_{10} \log(q^2(1 - q))$$
$$+ n_{11} \log(q(1 - q)) + n_{12} \log(q(1 - q)^2)$$
$$+ n_{21} \log(q(1 - q)^2) + n_{22} \log((1 - q)^3)$$

$$= (3n_{00} + 2(n_{01} + n_{10}) + n_{11} + n_{12} + n_{21}) \log q +$$
$$+ (3n_{22} + 2(n_{21} + n_{12}) + n_{11} + n_{10} + n_{01}) \log(1 - q)$$

$$= m_A \log q + m_B \log(1 - q)$$

The MLE of $q$ is $m_A/(m_A + m_B)$, where $(m_A + m_B) = 3n - n_{11}$ and $m_A = (3n_{00} + 2(n_{01} + n_{10}) + n_{11} + n_{12} + n_{21})$. 
2.4.4 ALTERNATIVES TO THE MLE:

The MLE is “best”, but there are simpler estimators that are not so bad.

(a) Use only founders (the moms):
estimate \( q \) by \( \frac{2n_{AA} + n_{AB}}{2n} \) where \( n_{AA} \) is number of \( AA \) moms, and \( n_{AB} \) is number of \( AB \) moms. \( (n_{AA} = n_{00} + n_{01}) \).

(b) Use everyone, disregarding relationship:
estimate \( q \) by \( \frac{2m_{AA} + m_{AB}}{4n} \), where \( m_{AA} \) is total number of \( AA \) individuals, and \( m_{AB} \) is total number of \( AB \) individuals. \( (m_{AA} = 2n_{00} + n_{01} + n_{10}) \).

These are both unbiased estimators, but asymptotically the MLE has smaller variance.

2.4.5 EFFECTS OF RELATEDNESS IN ESTIMATING \( q \):

- Results due to Tim Thornton.
- COGA data set; \( \sim 1214 \) individuals, in 105 pedigrees, \( \sim 992 \) observed (1984 gene copies).
- For the naive estimators we count alleles.
- For independent alleles the variance is \( \sim q(1 - q)/m \). All these curves are very close to \( q(1 - q)/m \) for some \( m \), and we can think of \( m \) as the “effective sample size”.
  (Larger variance \( \equiv \) smaller \( m \))

- Naive estimator; For \( X \): eff-m =375 (125 female, 125 male)
  For autosomal; eff-m = 500 (250 people) (Factor of 1/.75 = 1.33)
- For BLUE (\( \sim \) MLE): For \( X \): eff-m = 515 (approx) For autosomes:
  eff-m = 680 (Naive: 500/1984 \( \approx \) 0.25. BLUE 680/1984 \( \approx \) 0.34)
2.5.1 SPECIFYING INHERITANCE:

- Segregation of genes is fully specified by meiosis indicators
  
  \[ S_i = 0 \] if gene is parent’s maternal gene
  \[ S_i = 1 \] if gene is parent’s paternal gene

  where \( i = 1, \ldots, m \) indexes the meioses.

- Mendel’s First Law is \( S_i \) are independent with
  
  \[ \Pr(S_i = 0) = \Pr(S_i = 1) = \frac{1}{2}. \]

- \( ibd \) state \( J \) at a locus is a function of the \( \{ S_i \} \) at that locus.

- If \( \{ S_i \} \) are known, then we know which founder genomes (FGL) descend to each individual.

2.5.2 Example showing descent of FGL:

- Consider the following segregation pattern of genes:

- Label the founder genomes.
- Use the \( \{ S_i \} \) to trace descent of FGL.
- Same FGL implies \( ibd \).
- Example: The final individual and his maternal grandfather share FGL 8 – not by direct descent, but because both receive DNA from the founder who carries FGL 8.
2.5.3 The general formula for data probabilities:

\[ \Pr(Y) = \sum_{S} \Pr(Y \mid S) \Pr(S) \]

\[ = \sum_{S} \Pr(Y \mid J(S)) \Pr(S) \]

\[ = \sum_{J} \Pr(Y \mid J) \Pr(J) \]

\( \Pr(Y \mid J(S)) \) is the sum over all possible assignments \( A \) of allelic types to \( ibd \) gene-groups \( k \) of the product of allele frequencies \( q_{a(k)} \) of assigned alleles \( a(k) \):

\[ \Pr(Y \mid J(S)) = \sum_{A} \prod_{k} q_{a(k)}. \]

- **EXAMPLE**: Mom-baby pairs: \( ibd \) state 1 2 1 3: (or equiv.)
  - Data \( AA, AB \); 1 is \( A \), 2 is \( A \), 3 is \( B \): prob \( q^2(1 - q) \)
  - Data \( AB, BB \); 1 is \( B \), 2 is \( A \), 3 is \( B \): prob \( q(1 - q)^2 \)
  - Data \( AB, AB \); 1 is \( A \), 2 is \( B \), 3 is \( B \): prob \( q(1 - q)^2 \)
    - OR 1 is \( B \), 2 is \( A \), 3 is \( A \): prob \( q^2(1 - q) \); sum \( q(1 - q) \).

---

2.5.4 EXAMPLE: DATA ON TWO INDIVIDUALS:

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

<table>
<thead>
<tr>
<th>( P(J) )</th>
<th>( J )</th>
<th>( P(AA, AC \mid J) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta_1 )</td>
<td>1 1 1 1</td>
<td>0</td>
</tr>
<tr>
<td>( \Delta_2 )</td>
<td>1 1 1 2</td>
<td>( q_A q_C )</td>
</tr>
<tr>
<td>( \Delta_3 )</td>
<td>1 2 1 1</td>
<td>0</td>
</tr>
<tr>
<td>( \Delta_4 )</td>
<td>1 1 2 2</td>
<td>0</td>
</tr>
<tr>
<td>( \Delta_5 )</td>
<td>1 1 2 3</td>
<td>( q_A (2q_A q_C) )</td>
</tr>
<tr>
<td>( \Delta_6 )</td>
<td>1 2 3 3</td>
<td>0</td>
</tr>
<tr>
<td>( \Delta_7 )</td>
<td>1 2 1 2</td>
<td>0</td>
</tr>
<tr>
<td>( \Delta_8 )</td>
<td>1 2 1 3</td>
<td>( q_A q_A q_C )</td>
</tr>
<tr>
<td>( \Delta_9 )</td>
<td>1 2 3 4</td>
<td>( q_A^3(2q_A q_C) )</td>
</tr>
</tbody>
</table>

Total probability of observing \( (AA, AC) \) is

\[ P(AA, AC) = \Delta_2 q_A q_C + \Delta_5 2q_A^2 q_C + \Delta_8 q_A^2 q_C + \Delta_9 2q_A^3 q_C \]
2.5.5 Back to JV pedigree example:

- Given the particular descent pattern S, consider the possible allelic types of these genes given the genotypes of 5 individuals shown:

- FGL graph or ibd graph or descent graph (Lange):

- There are always 2, 1, or 0 possible assignments of allelic types to FGL nodes that are consistent with observed (no-error) genotypes.

Blank slide: