4. Genetic Identity Coefficients

§4.1. Kinship and inbreeding coefficients

- **Some definitions**

  Identity by state (ibs) : Two alleles are ibs if they are functionally the same.

  Identity by descent (ibd) : Two alleles are ibd if one is a physical copy of the other, or if they are both physical copies of the same ancestral allele.

  Kinship coefficient Φ_{ij}: The kinship coefficient Φ_{ij} between two individuals i and j is the probability that an allele selected randomly from i and an allele selected randomly from the same autosomal locus of j are ibd.

Inbreeding coefficient f_i: The Inbreeding coefficient f_i of an individual i is the probability that his two alleles at any autosomal locus are ibd. If $f_i > 0$, i is said to be **inbred**.

Relation between kinship and inbreeding coefficients:

$$\Phi_{ii} = \frac{1}{2}(1 + f_i), \quad f_i = \Phi_{kl},$$

where k and l are parents of i.

- **Calculation of kinship coefficients**

Simple examples

Parent-offspring: $\Phi_{ij} = 1/4$.

Full sibs: 3 and 4.

1×2: parents; 3, 4: childreen of 1 and 2.

$$\Phi_{34} = \frac{1}{2}\Phi_{31} + \frac{1}{2}\Phi_{32} = \frac{1}{4}.$$

Half sibs: 4 and 5.

1×2: husband and wife;

2×3: husband and wife;

4: child of 1 and 2;

5: child of 2 and 3.

$$\begin{aligned} \Phi_{45} &= \frac{1}{2}\Phi_{42} + \frac{1}{2}\Phi_{43} \\ &= \frac{1}{2} \times \frac{1}{4} + 0 = \frac{1}{8}. \end{aligned}$$

First cousins: 7 and 8.

1×2: husband and wife;

3, 4: children of 1 and 2;

3×5: husband and wife;

4×6 husband and wife;

7: child of 3 and 5;

8: child of 4 and 6.

$$
\begin{aligned}
\Phi_{78} &= \frac{1}{2}\Phi_{74} + \frac{1}{2}\Phi_{76} \\
&= \frac{1}{2}\Phi_{74} + 0 \\
&= \frac{1}{2}(\frac{1}{2}\Phi_{34} + \frac{1}{2}\Phi_{54}) \\
&= \frac{1}{2}(\frac{1}{2}\Phi_{34} + 0) \\
&= \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}.
\end{aligned}
$$

General algorithm for calculating kinship coefficients between members of a pedigree:

(i) Any person should have either both or neither of his or her parents in the pedigree.

(ii) Members in the pedigree are numbered in such a way that every parent precedes his or her children.

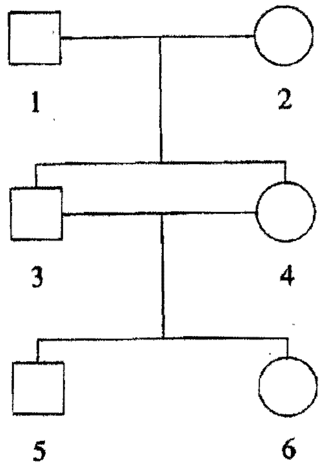
(iii) The kinship coefficients between any two persons in the pedigree are computed in a symmetric matrix from left top downwards recursively.

(iv) The recursive calculation formulas are:

- For Φ_{ii}: if i is a founder, $\Phi_{ii} = 1/2$, otherwise, $\Phi_{ii} = \frac{1}{2} + \frac{1}{2}\Phi_{kl}$, where k and l are parents of i.
- For Φ_{ij}, $(i > j)$: if i is a founder, $\Phi_{ij} = 0$, otherwise, $\Phi_{ij} = \frac{1}{2}\Phi_{jk} + \frac{1}{2}\Phi_{jl}$, where k and l are parents of i.

Basic Rule: Substitution of parental alleles for the allele of the child.

A brother-sister mating example:



$$\Phi = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \frac{3}{8} & \frac{3}{8} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & \frac{3}{8} & \frac{3}{8} \\ \frac{1}{4} & \frac{1}{4} & \frac{3}{8} & \frac{3}{8} & \frac{5}{8} & \frac{3}{8} \\ \frac{1}{4} & \frac{1}{4} & \frac{3}{8} & \frac{3}{8} & \frac{3}{8} & \frac{5}{8} \end{pmatrix}$$

A remark on the substitution rule

The substitution of parental alleles in the calculation of the kinship coefficient between two persons should always operate on the higher numbered person.

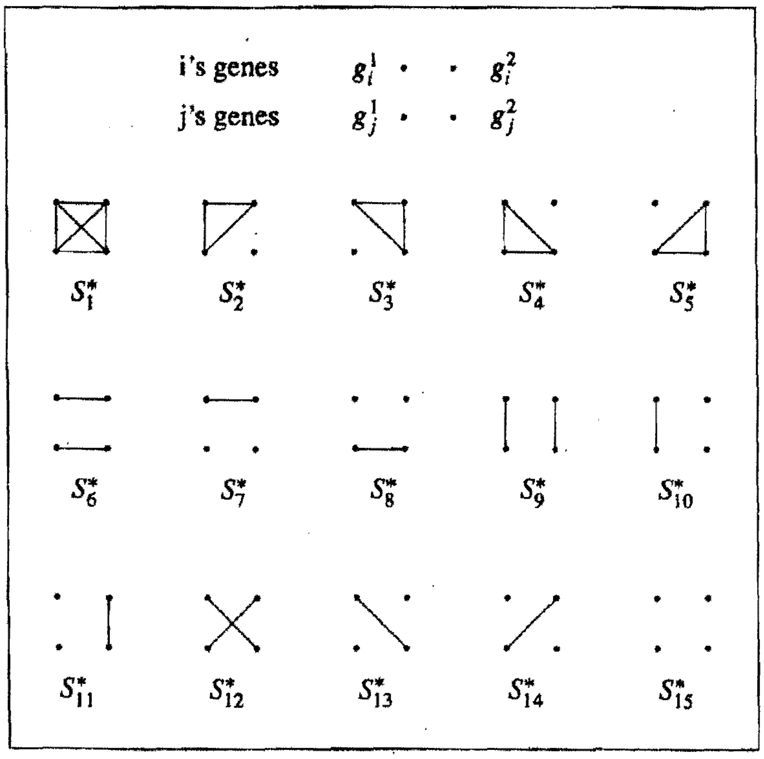
A counterexample:

$$\Phi_{35} = \frac{1}{2}\Phi_{33} + \frac{1}{2}\Phi_{34},$$
$$\Phi_{35} \neq \frac{1}{2}\Phi_{15} + \frac{1}{2}\Phi_{25}.$$

Note: While the parental allele passed to 3 is randomly chosen, once this choice is made, it limits what can be passed to 5. Sampling from 5 depends on what have been sampled for 3.

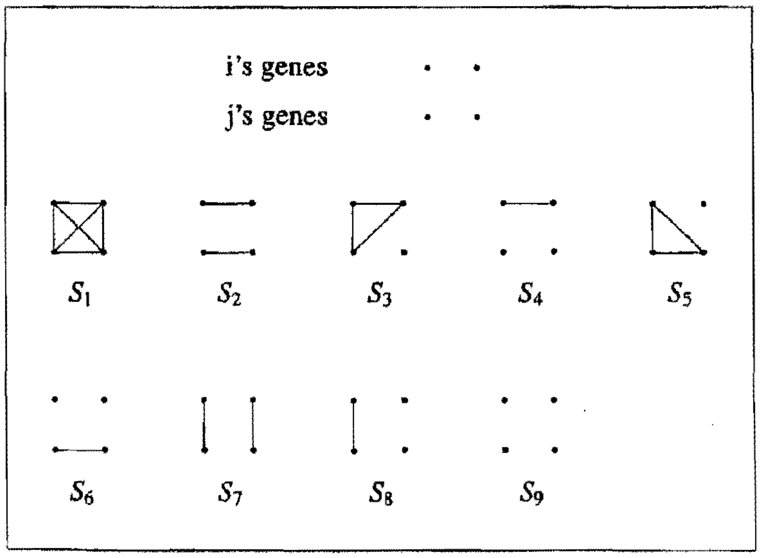
§4.2. Identity states and identity coefficients

• Detailed identity states



• Condensed identity states

$$S_1 = S_1^*,\ S_2 = S_6^*,\ S_3 = S_2^* \bigcup S_3^*$$
$$S_4 = S_7^*,\ S_5 = S_4^* \bigcup S_5^*,$$
$$S_6 = S_8^*,\ S_7 = S_9^* \bigcup S_{12}^*$$
$$S_8 = S_{10}^* \bigcup S_{11}^* \bigcup S_{13}^* \bigcup S_{14}^*$$
$$S_9 = S_9^*.$$



- **Identity coefficients Δ_k**

  Definition:

  $$\Delta_k = P(S_k).$$

  If a person is not inbred, the two alleles of the person cannot be IBD. Therefore

  $\Delta_k = 0$, for $k = 1, 2, 3, 4$, if i is not inbred;

  $\Delta_k = 0$, for $k = 1, 2, 5, 6$, if j is not inbred;

  $\Delta_k = 0$ except $k = 7, 8, 9$, if neither of i and j is inbred.

- **Relation between identity coefficients and kinship coefficient**

  $$\Phi_{ij} = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \frac{1}{4}\Delta_8.$$

  The relation can be obtained by conditioning on the identity states. For example, given either of S_2, S_4, S_6 and S_9, no alleles of i and

j can be IBD; given S_7, the probablity of two randomly chosen alleles of i and j is $1/2$, etc..

- **Calculation of identity coefficients for simple pedigrees**

  By conditioning on the identity states of parental relatives, the identity coefficients for simple pedigrees can be easily calculated as given in the following table.

Relationship	Δ_7	Δ_8	Δ_9	Φ
Parent-Offspring	0	1	0	1/4
Full siblings	1/4	1/2	1/4	1/4
Half siblings	0	1/2	1/2	1/8
First cousins	0	1/4	3/4	1/16
Double first cousins	1/16	6/16	9/16	1/8
Second cousins	0	1/16	15/16	1/64
Uncle-nephew	0	1/2	1/2	1/8

§4.3. Generalized kinship coefficients

- **Definition**

  Let n alleles $G_1, \ldots, G_n$ be selected at random from n persons (who are not necessarily all different persons). Let the n alleles be partitioned into non-overlapping blocks whose constituent genes are ibd. **A general kinship coefficient** is defined as **the probability that a particular partition occurs**.

  Example: Let G_i, G_j, G_k, G_l be 4 alleles selected at random from 4 persons. There are 15 different partitions of these 4 alleles which correspond to 15 detailed identity states: $S_1^{r*}, \ldots, S_{15}^{r*}$. Thus there are 15 general kinship coefficients: $\Phi(S_k^{r*}), k = 1, \ldots, 15$.

- **Random identity states and their probabilities ψ_k**

  If the 4 alleles are selected from two persons, two from each, the 15 detailed random identity states can be collapsed into 9 condensed random identity states (the order of the two alleles from the same person becomes irrelevant).

  The probability of the condensed random state S_k^r is denoted by ψ_k, i.e.,$\psi_k = P(S_k^r)$.

- **Condensed identity states**

$$
\begin{aligned}
\psi_1 &= \Phi(S_1^{r*}), \qquad \psi_2 = \Phi(S_6^{r*}), \\
\psi_4 &= \Phi(S_7^{r*}), \qquad \psi_9 = \Phi(S_9^{r*}), \\
\psi_3 &= \Phi(S_2^{r*}) + \Phi(S_3^{r*}) = 2\Phi(S_2^{r*}), \\
\psi_5 &= \Phi(S_4^{r*}) + \Phi(S_5^{r*}) = 2\Phi(S_4^{r*}), \\
\psi_7 &= \Phi(S_9^{*}) + \Phi(S_{12}^{r*}) = 2\Phi(S_9^{*}), \\
\psi_8 &= \Phi(S_{10}^{r*}) + \Phi(S_{11}^{r*}) + \Phi(S_{13}^{r*}) + \Phi(S_{14}^{r*}) \\
&= 4\Phi(S_{10}^{r*}).
\end{aligned}
$$

• Relationship between ψ_k and Δ_k

$$
\begin{aligned}
\psi_1 &= \Delta_1 + \frac{1}{4}\Delta_3 + \frac{1}{4}\Delta_5 + \frac{1}{8}\Delta_7 + \frac{1}{16}\Delta_8, \\
\psi_2 &= \Delta_2 + \frac{1}{4}\Delta_3 + \frac{1}{2}\Delta_4 + \frac{1}{4}\Delta_5 + \frac{1}{2}\Delta_6 \\
&\quad + \frac{1}{8}\Delta_7 + \frac{3}{16}\Delta_8 + \frac{1}{4}\Delta_9, \\
\psi_3 &= \frac{1}{2}\Delta_3 + \frac{1}{4}\Delta_7 + \frac{1}{8}\Delta_8, \\
\psi_4 &= \frac{1}{2}\Delta_4 + \frac{1}{8}\Delta_8 + \frac{1}{4}\Delta_9, \\
\psi_5 &= \frac{1}{2}\Delta_5 + \frac{1}{4}\Delta_7 + \frac{1}{8}\Delta_8, \\
\psi_6 &= \frac{1}{2}\Delta_6 + \frac{1}{8}\Delta_8 + \frac{1}{4}\Delta_9, \\
\psi_7 &= \frac{1}{4}\Delta_7, \ \psi_8 = \frac{1}{4}\Delta_8, \ \psi_9 = \frac{1}{4}\Delta_9.
\end{aligned}
$$

$$
\begin{aligned}
\Delta_1 &= \psi_1 - \frac{1}{2}\psi_3 - \frac{1}{2}\psi_5 + \frac{1}{2}\psi_7 + \frac{1}{4}\psi_8, \\
\Delta_2 &= \psi_2 - \frac{1}{2}\psi_3 - \psi_4 - \frac{1}{2}\psi_5 - \psi_6 \\
&\quad + \frac{1}{2}\psi_7 + \frac{3}{4}\psi_8 + \psi_9, \\
\Delta_3 &= 2\psi_3 - 2\psi_7 - \psi_8, \\
\Delta_4 &= 2\psi_4 - \psi_8 - 2\psi_9, \\
\Delta_5 &= 2\psi_5 - 2\psi_7 - \psi_8, \\
\Delta_6 &= 2\psi_6 - \psi_8 - 2\psi_9, \\
\Delta_7 &= 4\psi_7, \ \Delta_8 = 4\psi_8, \ \Delta_9 = 4\psi_9.
\end{aligned}
$$

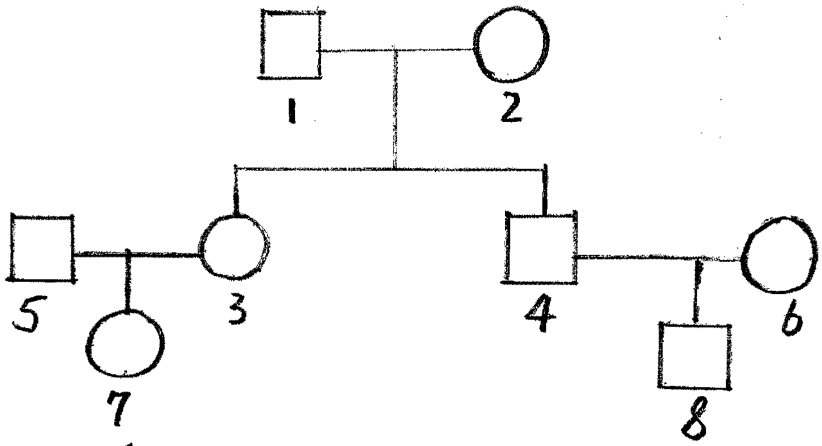
The expression of ψ_k's in terms of Δ_k's can be obtained by conditioning on the (non-random) condensed identity states of the two person's alleles.

For example, given either of S_2, S_4, S_6 and S_9, the randomly sampled four alleles cannot be all IBD; given S_1, the randomly sampled four alleles, two form each person, are IBD with probability 1; etc..

The expression of Δ_k's in terms of ψ_k's can be easily solved by backwards substitution.

§4.4. Calculation of generalized kinship coefficients

- **An example:** General kinship coefficient for first cousins.



$$
\begin{aligned}
&\frac{1}{4}\psi_8 \\
= \;& \Phi(\{G_7^1, G_8^1\}, \{G_7^2\}, \{G_8^2\}) \\
= \;& \frac{1}{4}[\Phi(\{G_3^1, G_8^1\}, \{G_5^1\}, \{G_8^2\}) + \Phi(\{G_5^1, G_8^1\}, \{G_3^1\}, \{G_8^2\})] \\
= \;& \frac{1}{16}[\Phi(\{G_3^1, G_4^1\}, \{G_5^1\}, \{G_6^1\}) + \Phi(\{G_5^1, G_6^1\}, \{G_3^1\}, \{G_4^1\})] \\
= \;& \frac{1}{16}\frac{1}{2}[\Phi(\{G_1^1, G_4^1\}, \{G_5^1\}, \{G_6^1\}) + \Phi(\{G_2^1, G_4^1\}, \{G_5^1\}, \{G_6^1\})] \\
= \;& \frac{1}{16}\frac{1}{2}[\Phi(\{G_1^1, G_1^2\}, \{G_5^1\}, \{G_6^1\}) + \Phi(\{G_2^1, G_2^2\}, \{G_5^1\}, \{G_6^1\})] \\
= \;& \frac{1}{16}\frac{1}{2}\frac{1}{2}.
\end{aligned}
$$

- **General rules**
  - **Recurrence rules**

Recurrence rule 1:
Assume only one allele G_i is sampled from i whose parents are j and k. Then

$$\begin{aligned} & \Phi(\{G_i, \ldots, \}\{\} \ldots \{\}) \\ = \; & \frac{1}{2}[\Phi(\{G_j, \ldots, \}\{\} \ldots \{\}) + \Phi(\{G_k, \ldots, \}\{\} \ldots \{\})]. \end{aligned}$$

Recurrence rule 2:
Assume that the alleles $G_i^1, \ldots, G_1^s$ are sampled sampled from i for $s > 1$. If these alleles occur in one block, then

$$\begin{aligned} & \Phi(\{G_i^1, \ldots, G_i^s, \ldots\}\{\} \ldots \{\}) \\ = \; & [1 - 2(\frac{1}{2})^s]\Phi(\{G_j, G_k, \ldots\}\{\} \ldots \{\}) \\ & +(\frac{1}{2})^s\Phi(\{G_j, \ldots, \}\{\} \ldots \{\}) + (\frac{1}{2})^s\Phi(\{G_k, \ldots, \}\{\} \ldots \{\})]. \end{aligned}$$

Recurrence rule 3:

Assume that the alleles $G_i^1, \ldots, G_1^s, G_i^{s+1}, \ldots, G_i^{s+t}$ are sampled sampled from i. If the first s alleles occur in one block and the remaining t alleles occur in another block, then

$$\begin{aligned} & \Phi(\{G_i^1, \ldots, G_i^s, \ldots\}\{G_i^{s+1}, \ldots, G_i^{s+t}, \ldots\}\{\} \ldots \{\}) \\ = \; & (\frac{1}{2})^{s+t}[\Phi(\{G_j, \ldots,\}\{G_k, \ldots,\}\{\} \ldots \{\}) \\ & + \Phi(\{G_k, \ldots,\}\{G_j, \ldots,\}\{\} \ldots \{\})]. \end{aligned}$$

– **Boundary rules**

Boundary rule 1:
If any person is involved in three or more blocks then $\Phi = 0$.

Boundary rule 2:
If two founders appear in the same block then $\Phi = 0$.

Boundary rule 3:
If only founders contribute sampled alleles and neither condition in boundary rule 1 nor conditions in boundary rule 2 pertains, then $\Phi = (\frac{1}{2})^{m_1 - m_2}$, where m_1 is the total number of sampled founder alleles and m_2 is the total number of founders sampled.

– The brother-sister mating example

$$
\begin{aligned}
& \frac{1}{4}\psi_8 \\
= & \Phi(\{G_5^1, G_6^1\}, \{G_5^2\}, \{G_6^2\}) \\
= & \frac{1}{4}[\Phi(\{G_3^1, G_6^1\}, \{G_4^1\}, \{G_6^2\}) + \Phi(\{G_4^1, G_6^1\}, \{G_3^1\}, \{G_6^2\})] \\
= & \frac{1}{2}[\Phi(\{G_3^1, G_6^1\}, \{G_4^1\}, \{G_6^2\}) \\
= & \frac{1}{8}[\Phi(\{G_3^1, G_3^2\}, \{G_4^1\}, \{G_4^2\}) + \Phi(\{G_3^1, G_4^2\}, \{G_4^1\}, \{G_3^2\})] \\
= & \frac{1}{8}[A+B].
\end{aligned}
$$

$$
\begin{aligned}
A = & \frac{1}{2}\Phi(\{G_1^1, G_2^1\}, \{G_4^1\}, \{G_4^2\}) \\
& +\frac{1}{4}[\Phi(\{G_1^1\}, \{G_4^1\}, \{G_4^2\}) + \Phi(\{G_2^1\}, \{G_4^1\}, \{G_4^2\})] \\
= & 0 + \frac{1}{2}\Phi(\{G_1^1\}, \{G_4^1\}, \{G_4^2\}) \\
= & \frac{1}{8}[\Phi(\{G_1^1\}, \{G_1^2\}, \{G_2^1\}) + \Phi(\{G_1^1\}, \{G_2^1\}, \{G_1^2\})] \\
= & \frac{1}{4}\Phi(\{G_1^1\}, \{G_1^2\}, \{G_2^1\}) = \frac{1}{8}.
\end{aligned}
$$

$$
\begin{aligned}
B = & \frac{1}{4}[\Phi(\{G_1^1, G_4^2\}, \{G_4^1\}, \{G_2^1\}) + \Phi(\{G_2^1, G_4^2\}, \{G_4^1\}, \{G_1^1\})] \\
= & \frac{1}{2}\Phi(\{G_1^1, G_4^2\}, \{G_4^1\}, \{G_2^1\}) \\
= & \frac{1}{8}[\Phi(\{G_1^1, G_1^2\}, \{G_2^2\}, \{G_2^1\}) + \Phi(\{G_1^1, G_2^2\}, \{G_1^2\}, \{G_2^1\})] \\
= & \frac{1}{8} \times \frac{1}{4}.
\end{aligned}
$$