2.3 IBD of more than 2 genes
2.3.1 Four genes of two people

Ibd pattern B1:p,m B2:p,m	ibd label	ibd group	state description: individuals ibd	genes shared
# # # #	1 1 1 1	1 1 1 1	B1, B2	4 genes ibd
# # # $	1 1 1 2	1 1 1 2	B1	3 genes ibd
# # $ #	1 1 2 1			
# $ # #	1 2 1 1	1 2 1 1	B2	3 genes ibd
# # # $	1 2 2 2			
# # $ $	1 1 2 2	1 1 2 2	B1,B2	none
# # $ @	1 1 2 3	1 1 2 3	B1	none
# $ @ @	1 2 3 3	1 2 3 3	B2	none
# $ # $	1 2 1 2	1 2 1 2	neither	2 genes shared
# $ $ #	1 2 2 1			
# $ # @	1 2 1 3	1 2 1 3	neither	1 gene shared
# $ @ #	1 2 3 1			
# $ $ @	1 2 2 3			
# $ @ $	1 2 3 2			
# $ @ %	1 2 3 4	1 2 3 4	neither	none

2.3.2 Any number of genes

- Order the k individuals, and the two genes within each (e.g. paternal then maternal).
- Label the 2k ordered genes sucessively, giving each the label previously assigned to genes to which it is ibd, and otherwise the next available integer.
- For example, in the state 1 2 1 3 4 4 1 5, the paternal genes of individuals 1,2, and 4 are ibd and the two genes of individual 3 are ibd.
- Reduce to genotypically equivalent classes of states: 12 13 44 15, 12 31 44 15, 12 31 44 51, 12 13 44 51, 12 23 44 25, 12 32 44 25, 12 32 44 52 and12 23 44 52 are all equivalent: Individuals 1,2,4 share one gene, and the two genes of individual 3 are ibd.
- Note that when the two genes of the first individual are interchanged, we must relabel the genes 1 and 2, to obtain a legal state label.
- The case of 4 genes of two individuals is shown in the Table 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.

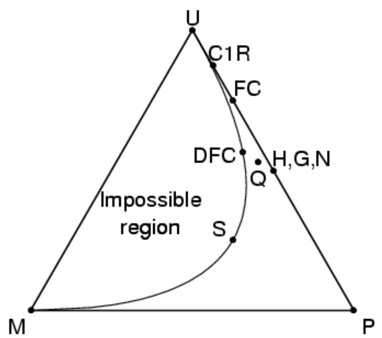
2.3.3 RELATIONSHIPS BETWEEN TWO NON-INBRED RELATIVES

- For two non-inbred relatives, there are 7 states, 3 classes, and 2 probabilities K_i = P(i genes shared ibd), with $K2 + K1 + K0 = 1$.
- The following equations relate ψ and κi, i=0,1,2: .
  $\psi = (1/2) K2 + (1/4) K1 = (1/4) (1 + K2 - K0)$.
  $\psi(B1,B2) = (1/4)(\psi(M1,M2) + \psi(M1,F2) + \psi(F1,M2) + \psi(F1,F2))$.
- Then also .
  $K2(B1,B2) = \psi(M1,M2)\, \psi(F1,F2) + \psi(M1,F2)\psi(F1,M2)$
- $K1(B1,B2) = 4\, \psi(B1,B2) - 2\, K2(B1,B2)$,
- And $K0(B1,B2) = 1 - K1(B1,B2) - K2(B1,B2)$.

Table of K values for relationships

Relationship	K0	K1	K2	ψ
Unrelated (U)	1.0	0.0	0.0	0.0
Parent-offspring (P)	0.0	1.0	0.0	0.25
Monozygous twin (M)	0.0	0.0	1.0	0.5
Half-sib, uncle, niece, grandparent (H)	0.5	0.5	0.0	0.125
Full sib (S)	0.25	0.5	0.25	0.25
First cousin (C)	0.75	0.25	0.0	0.0625
Double first cousin (D)	0.5625	0.375	0.0625	0.125
Quad half first cousin (Q)	0.5312	0.4375	0.0312	0.125

Representation of relationships in a triangle of unit height



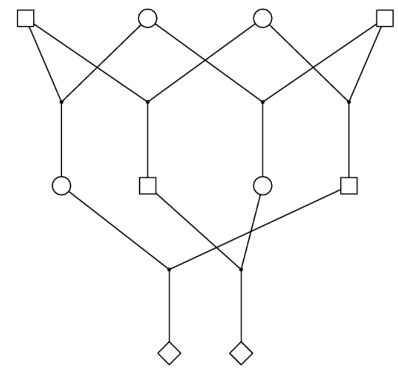
- Any set of three numbers summing to 1 can be represented as a point in a unit-height triangle.

  The relationships of the table are shown in the figure.

  Applying the Arithmetic Geometric means inequality to these same equations shows K1^2 is always at least

  4*K0*K2 for all real relationships. The excluded area is show in the figure. (See book, P.38, for details.)

2.3.4 EXAMPLE OF QUADRUPLE HALF FIRST COUSINS



- The example of QHFC shows it is possible for all four of ψ(M1,M2), ψ(F1,F2), ψ(M1,F2) and ψ (F1,M2) to be non-zero without the children being inbred.
- Apply equations of 2.3.3 for QHFC.
- ψ(M1,M2)= ψ(F1,F2) = ψ(M1,F2) = ψ(F1,M2) =1/8. Then
- K2 = (1/8)*(1/8)+ (1/8)*(1/8) = 1/32 ψ= (1/4)(1/8 +1/8+1/8+1/8) = 1/8, K1= 4 ψ- 2 K2 = 7/16, and K0 = 1 - K2 - K1 = 17/32.
- In the triangle, QHFC are midway between DFC and half-sibs

2.4 DATA ON RELATIVES
2.4.1 FROM INHERITANCE TO DATA

- *Segregation* of genes is fully specified by *meiosis indicators:* . S_i = 0 if gene is parent's maternal gene . = 1 if gene is parent's paternal gene where i = 1,...,m indexes the meioses.
- S_i are independent each with P(S_i = 0) = P(S_i = 1) =(1/2).
- ibd at a locus is a function of the S_i at that locus.
- The general formula for data Y:
  P(Y) = Σ_S P(Y | S) P(S) = Σ_S P(Y | J(S)) P(S) .
  = Σ_J P(Y | J) P(J)
- P(Y | J(S)) is the sum over all possible assignments Ω of allelic types to founder genes (FGL) of the product of allele frequencies q(a(k)) of assigned alleles a(k):
- P(Y | J(S)) = Σ_Ω Π_k q(a(k)) = Σ_Ω Π_a q(a)^n(a)
  where n(a) is the number of FGL that are assigned allelic type a in a particular assignment Ω.

2.4.2 Examples of 1 and 2 individuals

- DATA ON 1 INDIVIDUAL : Suppose we observe someone who is A1A1.
  The possible ibd states are J = (I,N), with P(I) =f, P(N) = 1-f.
  So P(A1A1) = P(A1A1 | I) f + P(A1A1 |N) (1-f) = q f + q^2 (1-f) =q (f + q - qf).

- DATA ON TWO INDIVIDUALS
- The relationship between the two individuals determines the probabilities Δ1, ..., Δ9 of the 9 ibd classes (groups of states). Suppose we observe the individuals to be AA and AC. The left hand side tabulates P(J), for states J and P(AA,AC |J):

Δ1	11 11	0
Δ2	11 12	q(A) q(C)
Δ3	12 11	0
Δ4	11 22	0
Δ5	11 23	q(A) (2 q(A) q(C))
Δ6	12 33	0
Δ7	12 12	0
Δ8	12 13	q(A)q(A) q(C)
Δ9	12 34	q(A)^2 (2 q(A) q(C))

Total probability of observing AA,AC is

P(AA,AC) = Σ_{k=1}^9 Δk P(AA,AC | J =k)

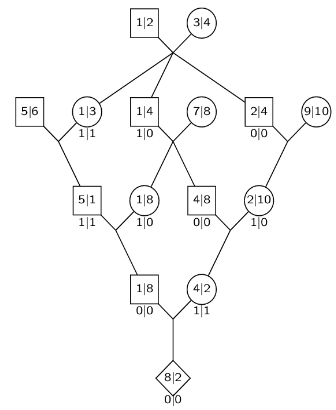
= Δ2 q(A)q(C)+ Δ5 2 q(A)^2 q(C)

+ Δ8 q(A)^2 q(C) + Δ9 2 q(A)^3 q(C)

2.4.3 DATA ON A NON-INBRED PAIR

- Given relationship R and data Y = (G1,G2), the genotypes of the pair of individuals.
- Let J(i), (i=0,1,2) denote the state of ibd in which the two non-inbred individuals share i genes ibd, and Ki = Ki(R) the probabilities of state J(i) implied by R.
- P(G1,G2 ; R) = Σ_i P(Y |J(i)) P(J(i); R)
  = K0(R) P(G1,G2 | J(0)) + K1(R) P(G1,G2 | J(1))
  + K2(R) P(G1,G2 | J(2))
  = K0 P(G1,G2 | Unrel) + K1 P(G1,G2 | Par-offsp)
  + K2 P(G1,G2 | MZ-twins)
  = K0 P(G1) P(G2) + K1 P(G1) P(kid=G2 | par=G1)
  + K2 P(G1) I(G2 = G1).

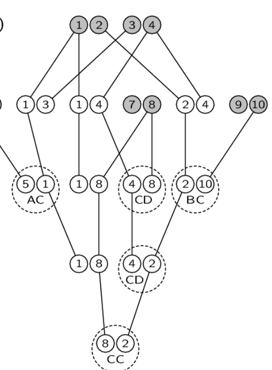
2.4.4 Example of use of general formula, using the JV pedigree.



- Label the founder genomes (FGL).
- Any segregation pattern of genes, S, enables us to track the FGL down the pedigree.
- This in turn specifies which genes of which individuals are ibd.
- In this example, males are on left, and females on the right of each marriage.
- In this example, paternal FGL and S are on left, maternal FGL and S are on the right of each individual.

Adding genotype data Y on five individuals

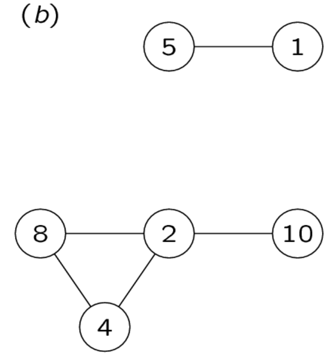
(a)



- Here we have the same pattern of segregation S as before, where now we show paths of descent of genes.
- Five individuals are observed as shown with genotypes AC, CD, BC, CD and CC for a locus with (at least) 4 alleles.
- On the next page, we form the FGL-graph. The nodes are FGL present in observed individuals, and lines connect FGL in the same observed individual.

Computing the probability

(b)



- Next we assign allelic types to the FGL.
- An FGL not present can be anything: total probability 1.
- FGL 1 and 5 can be 1=A,5=C or 1=C,5=A, total probability 2q(A)q (C).
- A bit of thought shows there is 1 assignment possible for the other component: 8=2=C, 4=D, 10=B with probability q(B) q(C)^2 q(D).
- Combining we have total prob-ability 2 q(A) q(B) q(C)^3 q(D).
- There will always be 2,1, or 0 possible assignments.