FINITE POPULATIONS.

- In reality we often deal with relatively small populations, or wish to focus on related individuals, e.g. for genetic counselling or to understand the population variability and how we can maintain this in a lab. or zoo population.
- We need to know how to specify relatedness between individuals or pops.

FINITE POPULATIONS.

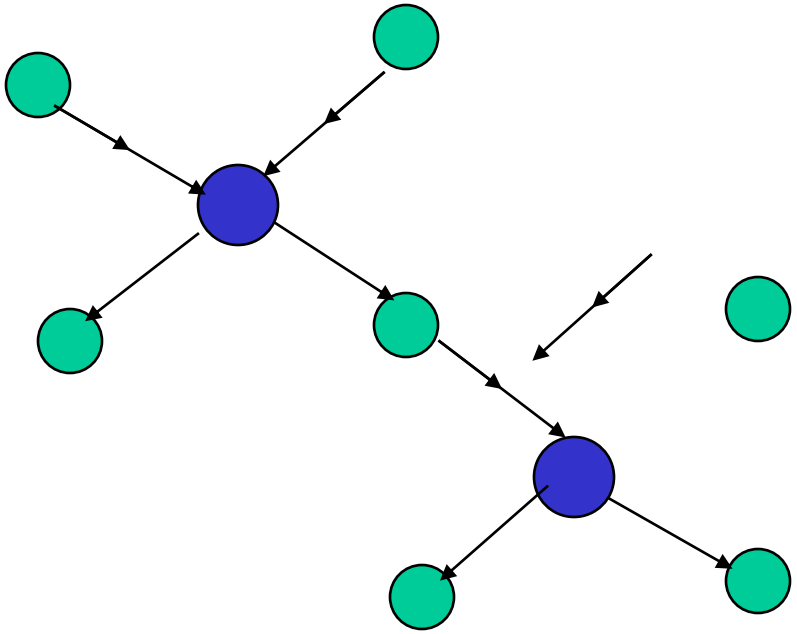
- Relationship.
- A = M(B) = “A is mother of B” , and A = F(B) = “A is father of B”.
  Then we can define other relationships, e.g. C = S(D) = “C is sib of D”, which is symmetric i.e. D = S(C), and is defined by (M(C) = M(D) **and** F(C) = F(D)).

RELATIONSHIPS.

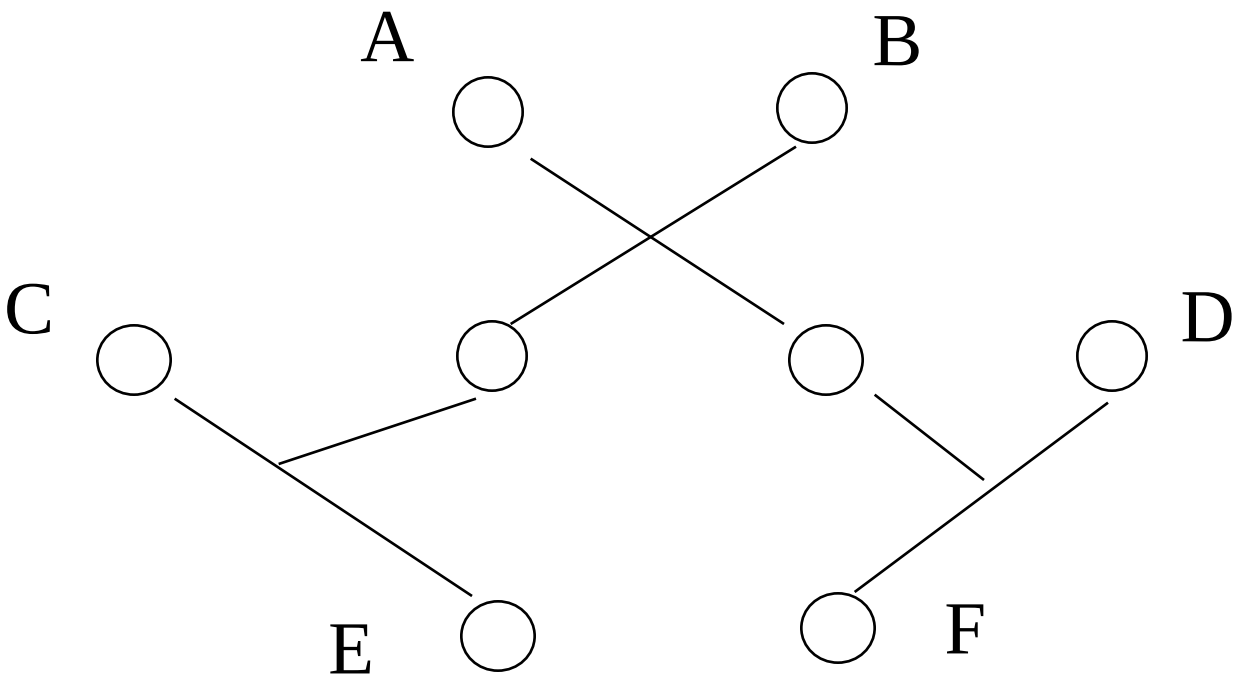
- First Cousins (symmetric relationship)
- A and B are First Cousins A = FC(B) if (P1(A) = S(P1(B)) **and** P2(A) = U(P2(B))) where U( ) = “unrelated to”, and P1 and P2 denote the two parents of an individual.
- A and B are Double First Cousins A = DFC(B) if (P1(A) = S(P1(B)) **and** P2(A) = S(P2(B))) .

MARRIAGE NODE GRAPH.

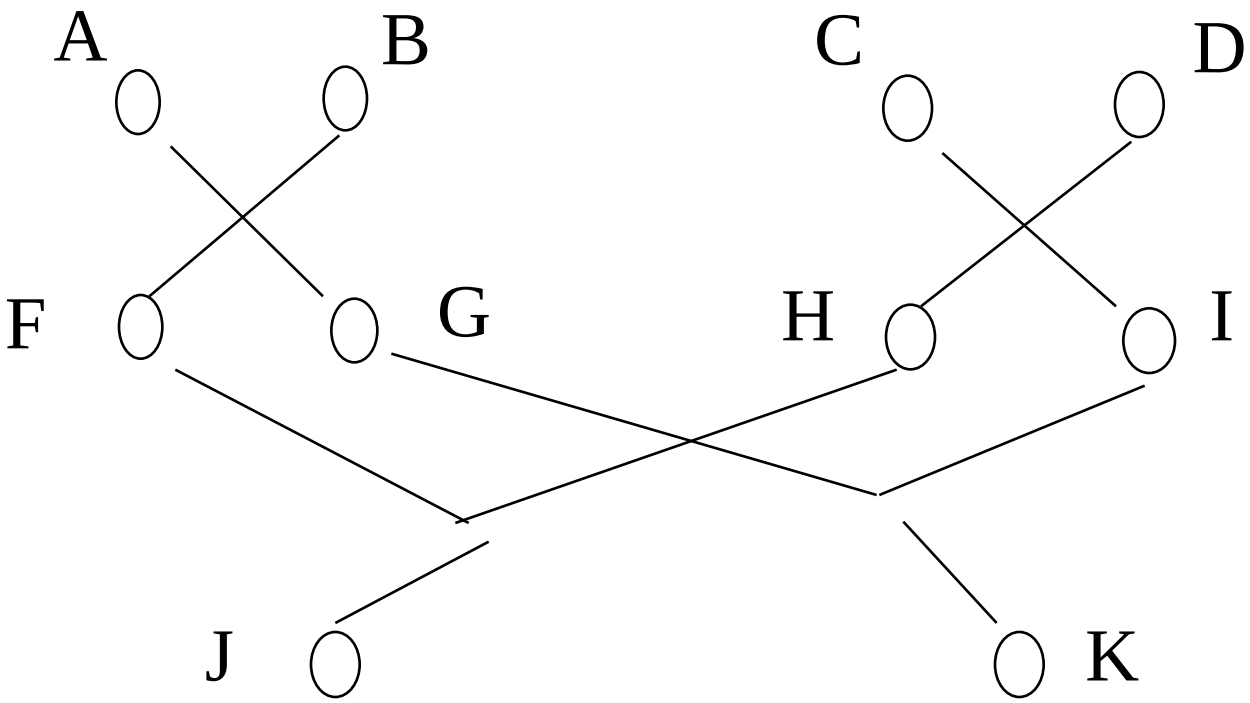
- individual marriage
  reproductive arc marriage arc



FIRST COUSINS. E & F



DFC's, J & K.



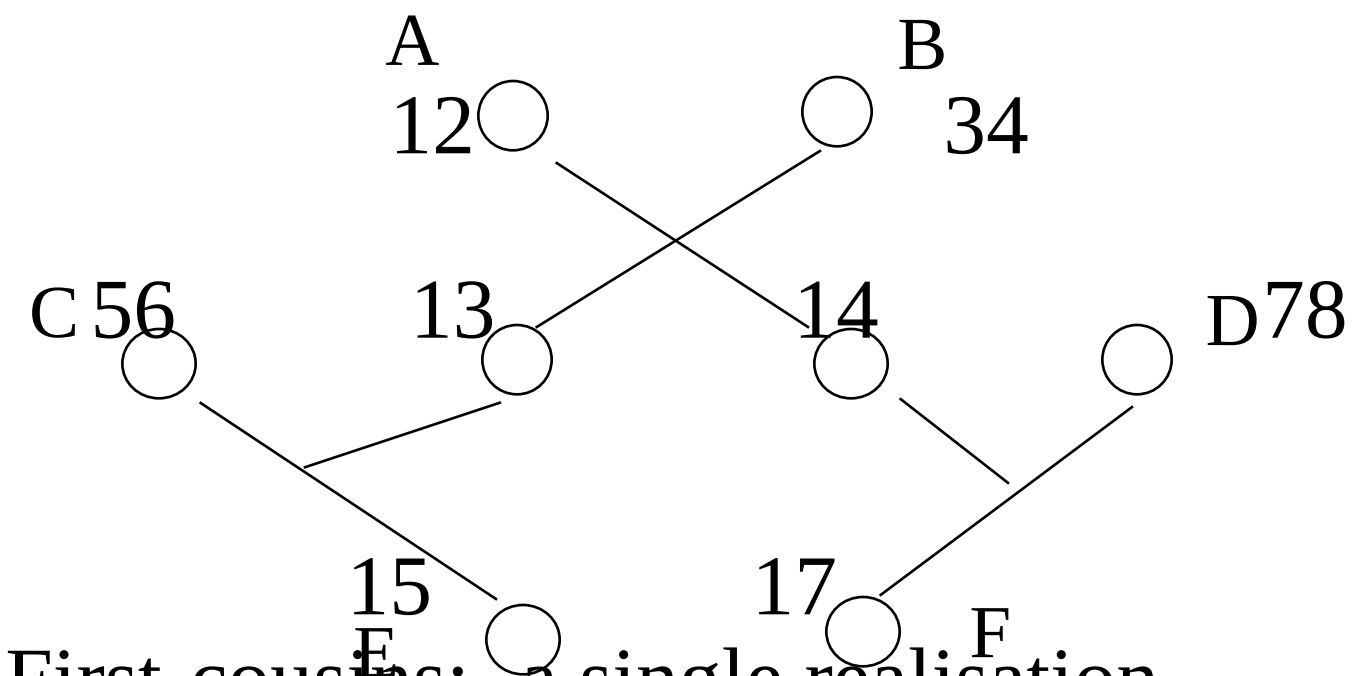
RELATIONSHIPS

- Here and elsewhere we make the assumption that all the relationships are fully specified (unless other stated). Thus in the first cousin definition it is assumed that the parents of the siblings are unrelated, as are the individuals who marry in.
- One can build up an algebra of relationships but here we shall use only in a limited way.

RELATIONSHIPS.

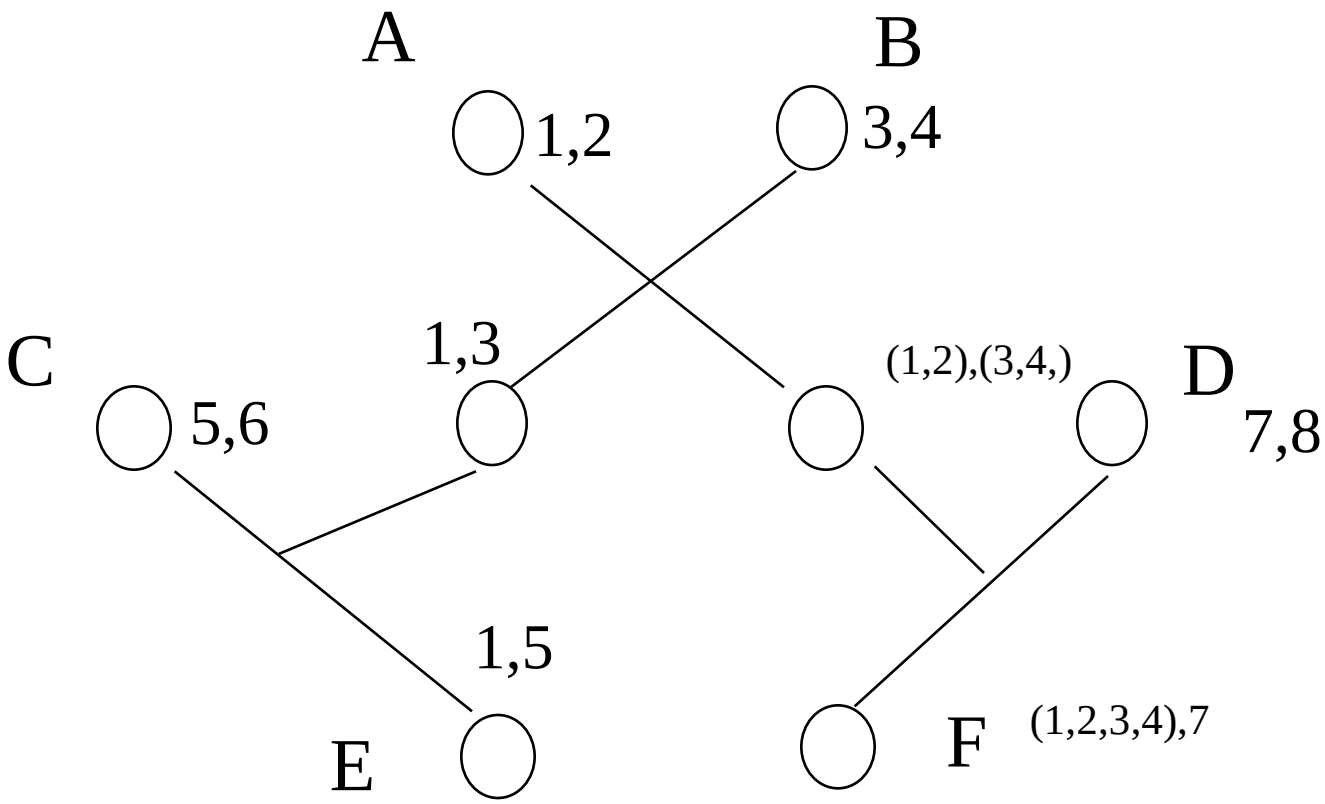
- Consider FC's. Because they share one set of grandparents they can have received copies of the same piece of DNA. We focus on a single locus and introduce the notion of Identity by Descent , IBD.
- Two genes are said to IBD if they are copies of some single ancestral gene.

IBD



First-cousins:- a single realisation

FIRST COUSINS.



RELATIONSHIPS.

- Define the Coefficient of Kinship of A & B Ø(A,B) = Probability that a random gene (at a specific locus) from A is IBD a random gene from B.
- Then Ø(FC) = 1/16 since there is a probability of 1/4 that we pick the two genes which come from the common set of grandparents and 1/4 that we pick the same one from those 4.

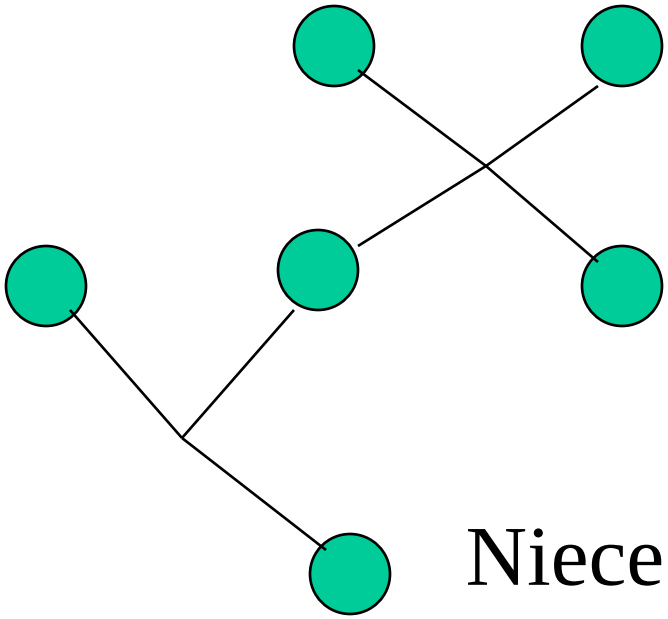
RELATIONSHIPS.

- Ø(S) = 1/4.
- Ø(FC) = Ø(S) / 4 + 3* Ø(U) / 4 where U indicates unrelated individuals and so Ø(U) = 0; i.e. Ø(FC) = 1/16.
- A and B are second cousins if (P1(A) = FC(P1(B))).

  Ø(SC) = Ø(FC) / 4 = 1/64.

RELATIONSHIP

- Uncle-niece



Uncle

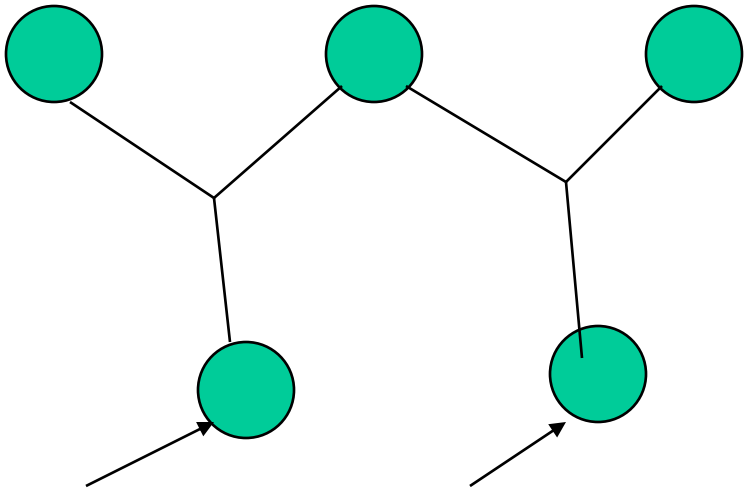
Niece

RELATIONSHIPS.

- Uncle-Niece, (P(A) = S(B)) .
- Ø(UN) = Ø(S) / 2 = 1/8.

RELATIONSHIP

Half-Sibs



RELATIONSHIP

- Half-Sibs, (P1(A) = P1(B)) .
- Ø(HS) = 1/8.

RELATIONSHIPS.

- If individuals can share genes IBD then a single individual can have its two genes IBD.
- Define μ(A) = Inbreeding Coefficient of A = Probability the two genes of A are IBD.
- μ(A) = Ø(P1(A), P2(A)).

INBREEDING.

- Wherever relatives marry the offspring may have their genes IBD, referred to as **being inbred.**
- For individual with $\mu(A) = \mu$ we have
  $Prob(aa) = \mu * p(a) + (1 - \mu) * p(a) * p(a)$
  so $= p(a) * p(a) + \mu * p(a) * q(a)$
  where p(a) = frequency of allele a.

INBREEDING.

- Example. Suppose p(a) = 0.01 for some recessive condition, then for the offspring of unrelated parents, frequency of disease is 0.0001 i.e. 1 in 10,000. For an offspring of first cousins the frequency is
  $p(a)*\mu + (1-\mu)*p(a)*p(a)$
  $= 0.01 / 16 + 15*0.0001/16 \sim 1/1600$ i.e. a six-fold increase.

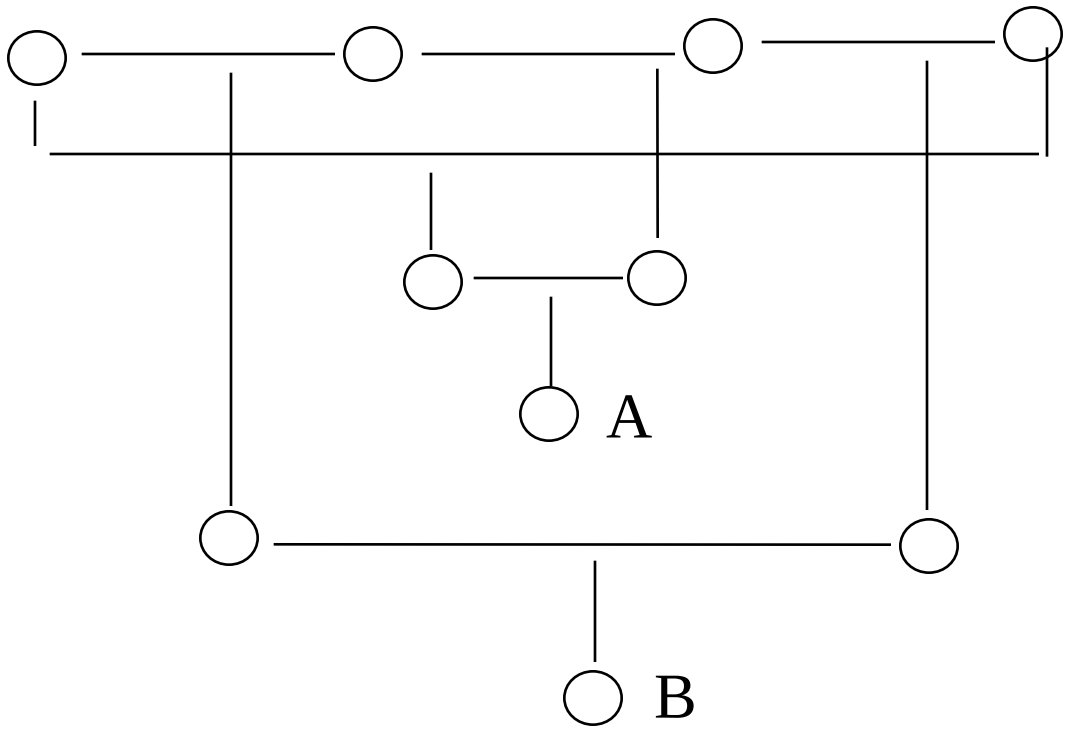
INBREEDING.

- Individuals with μ, frequencies of aa, ab and bb are p*p+μ*p*q, 2*p*q*(1-μ) and q*q+μ*p*q.
- Thus offspring of FC where μ=1/16 and p=0.4 we have frequencies **0.16** + 0.015, **0.48 -** 0.030 and **0.36**+0.015, the **bold** values being Hardy-Weinberg.

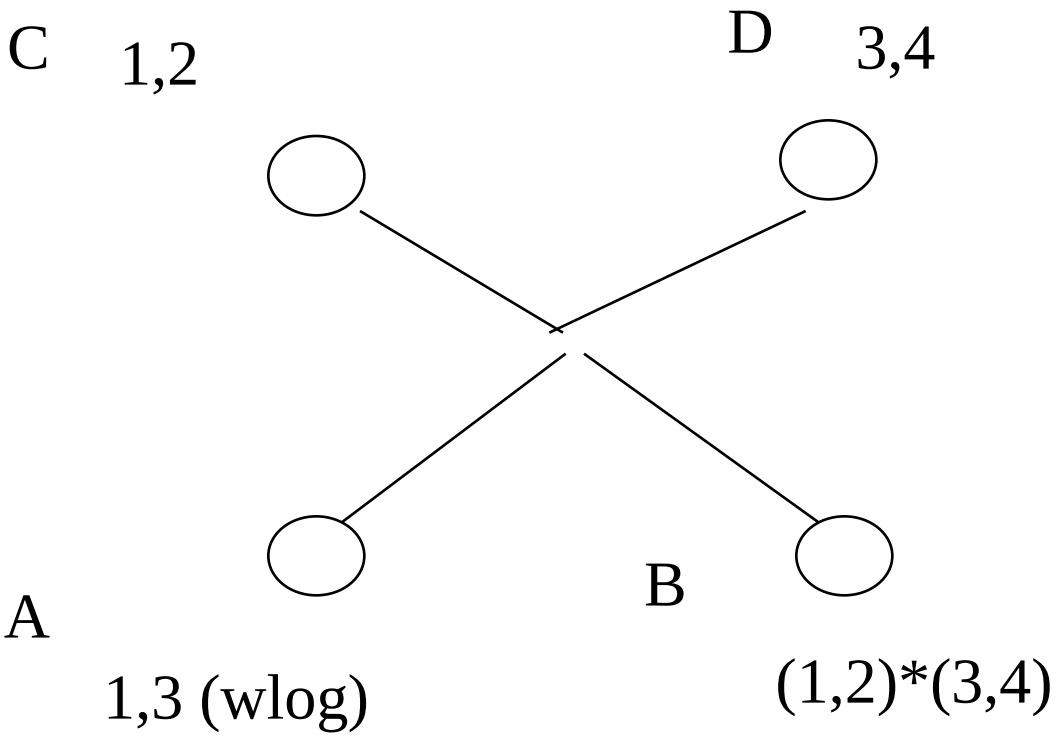
INBREEDING.

- Example. Find the coefficient of Kinship of Sibs, Half-sibs, Uncle-Niece, Double-First Cousins and Quadruple-Half-First Cousins.
  A and B are QHFC if
  (P1(P1(A))=P1(P1(B)) **and**
  P1(P2(A))=P2(P2(B)) **and**
  P2(P1(A))=P1(P2(B)) **and**
  P2(P2(A))=P2(P1(B)))

QHFC, A & B.



SIBS (A & B)



$$\Phi(A,B) = (\Phi(C,C) + 2*\Phi(C,D) + \Phi(D,D)) / 4$$
$$= (2 + \Phi(C) + \Phi(D) + 4 * \Phi(C,D)) / 8$$

SIBS.

- Note that for sibs whose parents are unrelated the probabilities of them having 0, 1 or 2 genes IBD are 1/4, 1/2 and 1/4.

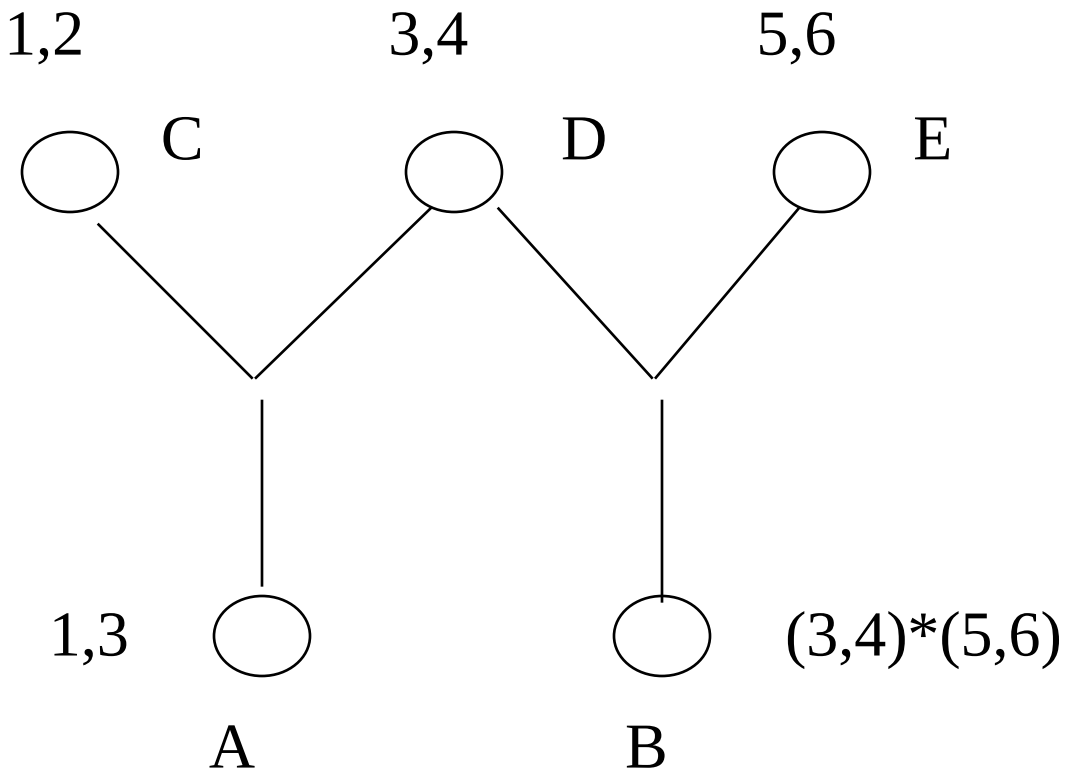
Population Coefficients

- The Coefficient of Kinship of a population is the probability that a gene from individual A is IBD to a gene from individual B where A and B are random individuals from thr population = mean coefficient of kinship over pairs of individuals.
- Coefficient of Inbreeding of a population = mean coefficient of inbreeding over individuals.

SIBS in POPULATION.

- Suppose that we have a population with coefficient of kinship in generation t of $\Phi(t)$ and coefficient of inbreeding $\mu(t)$.
- Then for sibs at time (t+1) we would have $\Phi(S,t+1) = (1 + 2*\Phi(t) + \mu(t)) / 4$ and since $\mu(t) = \Phi(t-1)$ this gives $\Phi(S,t+1) = (1 + 2*\Phi(t) + \Phi(t-1)) / 4$.

HALF-SIBS, A & B.



$$\Phi(A,B) = (\Phi(C,D) + \Phi(C,E) + \Phi(D,E) + \Phi(D,D))/4$$

HALF-SIBS.

$$\Phi(A,B) = (\Phi(C,D) + \Phi(C,E) + \Phi(D,E) + \Phi(D,D))/4$$

$$= (2*\Phi(C,D) + 2*\Phi(C,E) + 2*\Phi(D,E) + \Phi(D) + 1) / 8$$

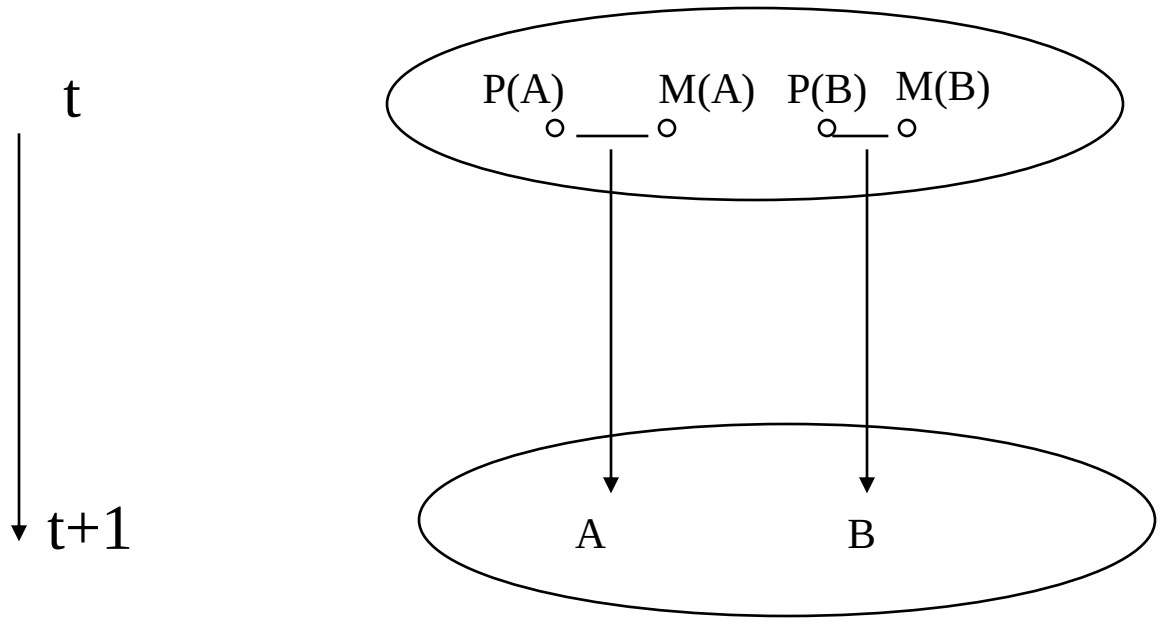
Thus in a population with $\Phi(t)$

$$\Phi(HS, t+1) = (1 + 6*\Phi(t) + \Phi(t-1)) / 8$$

FINITE POPULATION

- We have generalise the coefficient of kinship and the inbreeding coefficient of individuals to populations.
- Population coefficient of kinship $\Phi(t)$ = probability that if we select a random pair of individuals from the population that a random gene from the first individual is IBD to a random gene from the second.

FINITE POPULATION



FINITE POPULATIONS

- $\Phi(A,B)=(\Phi(P(A),P(B)) +\Phi(P(A),M(B)) + \Phi(M(A),P(B)) + \Phi(M(A),M(B)))/4$ for any A and B.
- Thus for a random pair from the population at time t+1

  $\Phi(t+1) = p(sibs)*\Phi(S,t+1)$
  $+ \quad p(half\text{-}sibs)*\Phi(HS,t+1)$
  $+ (1 - p(sibs) - p(half\text{-}sibs)) * \Phi(\text{“U”},t+1)$

FINITE POPULATIONS

- $\Phi(t+1) = p(S)*(1 + 2*\Phi(t) + \Phi(t-1)) / 4 + p(HS)*(1 + 6*\Phi(t) + \Phi(t-1)) / 8 + (1 - p(S) - p(HS))*\Phi(t)$

  $= (p(S) + p(HS)/2) /4$

  $+ \quad \Phi(t)*(1 - (p(S) + p(HS)/2)/2)$

  $+ \quad \Phi(t-1)* (p(S) + p(HS)/2) /4$

  $= p/4 + \Phi(t)*(1 - p/2) + \Phi(t-1)*p/4$

  where $p = (p(S) + p(HS)/2)$

FINITE POPULATIONS

- $\Phi(t+1) = p/4 + \Phi(t)*(1 - p/2) + \Phi(t-1)*p/4$ which is linear.
- Write $(1 - H(t)) = \Phi(t)$ then
  $(1 - H(t+1)) = p/4 + (1 - H(t))*(1 - p/2) +$
  $(1 - H(t-1))*p/4$
  Hence
  $H(t+1) = H(t)*(1 - p/2) + H(t-1)*p/4$
  which is linear and homogeneous.

FINITE POPULATIONS

- $$\mathbf{H}_{t+1} = (H_{t+1},H_t)` = \begin{pmatrix} (1-p/2) & p/4 \\ 1 & 0 \end{pmatrix} (H_t,H_{t-1})`$$

$$= \mathbf{A}\,\mathbf{H}_t\,,$$

Now the eigenvalues are the roots of

$(\lambda - (1-p/2))\,\lambda - p/4 = \lambda^2 - (1-p/2)\,\lambda - p/4 = 0$

FINITE POPULATIONS

- Quadratic Equation:-
  If $ax^2 + bx + c = 0$ then
  $x = (-b \pm \sqrt{} (b^2 - 4ac)) /2a.$
- In this case we have
  $\lambda = ((1-p/2) \pm \sqrt{}((1- p/2)^2 + p)) / 2$
  $= ((1-p/2) \pm \sqrt{}(1- p^2/4))/2$
  $\approx ((1-p/2) \pm \sqrt{}1)/2$ if p is small
  $= (1-p/4)$ and $-p/4$.

FINITE POPULATIONS

- Now H(t) = probability of non-identity and $H(t+1) = H(t)*(1 - p/2) + H(t-1)*p/4$.
- Eigenvalues λ from
  $\lambda^2 - \lambda*(1-p/2) - p/4 = 0$ which for p small give $(1 - p/4)$ and $-p/4$ approximately, since
  $(\lambda - (1 - p/4))(\lambda + p/4)$
  $= \lambda^2 - \lambda*(1 - p/2) - p/4 + o(p^2)$.

FINITE POPULATIONS

- The dominant eigenvalue is $\lambda = 1 - p/4$ where $p = p(S) + p(HS)/2$
  Expected {No of parents in common}= $2*p(S)+1*p(HS)=4p$.

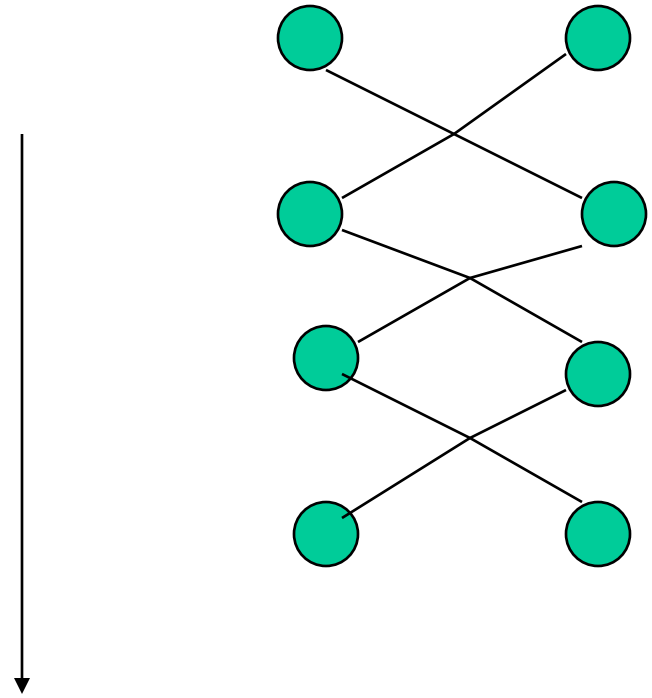
WRIGHT-FISHER MODEL.

- "Wright-Fisher" model = random mating then if population of n individuals (diploids but no sexes) $p(S) = 2 / n(n-1)$ and $p(HS) = 4(n-2) / n(n-1)$ and thus $p = 2 / n$ so $\lambda = 1 - 1/2n$, 2n= number of genes.

REPEATED SIB-MATING

- p(S) = 1 and p(HS) = 0 then p = 1so H(t+1) = H(t)*(1 - p/2) + H(t-1)*p/4 = H(t)/2 + H(t-1)/4 so eigenvalues from $\lambda^2 - \lambda/2 - 1/4 = 0$ or $4\lambda^2 - 2\lambda - 1 = 0$. Thus $\lambda = (1 \pm \text{sqrt}(5)) / 4$ so the dominant eigenvalue is approximately 0.81.

REPEATED SIB-MATING



p(S)

- Suppose we have n males and n females in a constant size population and monogamy. The number of individuals in a family is suppose random and has mean 2 (since population size is constant) and variance=σ^2. We need to derive an expression for p(S) in terms of σ^2.

p(S)

- Consider picking a pair of individuals and checking whether they are sibs.Suppose that t_r = prob(family has r offspring), then when we pick the first individual the probability that that individual is in a family of size r is $r*t_r / \sum r*t_r$ and in that case the chance that the second individual is in the same family is (r-1)/(2n-1) so adding over all r we have

p(S)

- $P(s) = \sum t_r * r(r-1) / (2n-1)\sum t_r * r$
  $= Exp(r(r-1)) / (2n-1)\ Exp(r)$
  $= (\sigma^2 + 2) / 2(2n-1).$

p(S)

- For a monogamous population with equal numbers of males and females n and family size with mean 2 and variance σ^2 we have $p(S) = (\sigma^2 + 2) / 2(2n-1)$
- "Wright-Fisher" monogamous model then family size is B(2n, 1/n) so $\sigma^2 = 2\ n\ (1/n)\ (1 - 1/n) = 2(n-1)/n$ so $p(S) = 1/n$.

p(S)

- Suppose only one male allowed to mate, there being n males and n females. Then if we pick to individuals at random they we have

  $p(S) = 1/n$ and $p(HS) = (n-1)/n$

  so $p = p(S) + p(HS)/2 = (n+1) / 2n$

  and $\lambda = 1-(n+1)/8n$

  n=1 then $\lambda = 3/4$ and n=inf then $\lambda = 7/8$.

OTHER MODELS.

- Males and Females distinct $H_m(t)$ and $H_f(t)$
- Subdivided populations e.g. systems with $H_W(t)$ and $H_B(t)$ which then allow one to treat “avoidance of inbreeding”, and “prohibition of sib-mating etc”.
- General Markov chain then can find eigenvalues for higher moments than variance.

IDENTITY STATES.

- We have used the concept of IBD in a simple way to measure relationship for individuals and populations. This is sufficient for some situations, i.e. for a single individual or for pairs of individuals who have only one or zero genes in common. For generality we need a more sophisticated method.

IDENTITY STATES

- An Identity State specifies completely the IBD status of all the genes.
- We shall limit ourselves to Identity States for at most two individuals.

IDENTITY STATES

- Identity State — Probability

Identity State	Probability
(1,1,1,1)	π_1
(1,1,1,2) = (1,1,2,1)	π_2
(1,2,1,1) = (1,2,2,2)	π_3
(1,1,2,2)	π_4
(1,2,1,2) = (1,2,2,1)	π_5
(1,1,2,3)	π_6
(1,2,3,3)	π_7
(1,2,1,3) = (1,2,2,3) = (1,2,3,1) = (1,2,3,2)	π_8
(1,2,3,4)	π_9

NON-INBRED IND'S

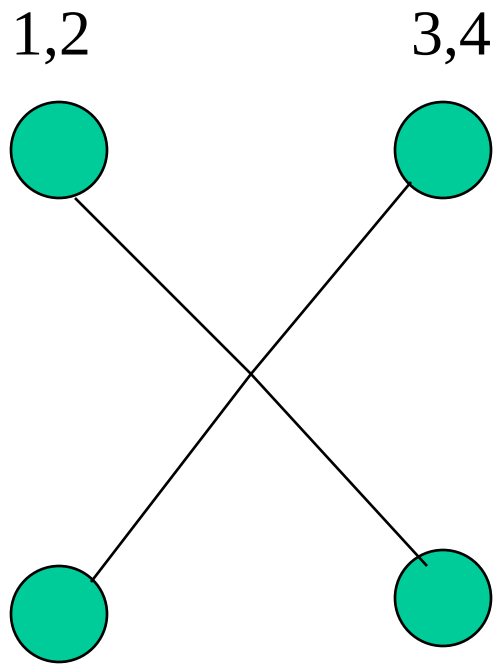
- If the parents of an individual are unrelated then that individual is said to non-inbred, and necessarily has $\mu=0$. If we consider a pair of non-inbred individuals then their identity states are restricted to (1,2,3,4), (1,2,1,4) or (1,2,1,2). In this case we can collapse the set of probabilities to $\mathbf{k}=(k_0,k_1,k_2)$ where subscript=number of IBD genes.

NON-INBRED IND'S

- Sibs **k**(S) = (1/4, 1/2, 1/4),
- First Cousins **k**(FC) = (3/4, 1/4, 0),
- Double-first **k(**DFC) = (9/16, 6/16, 1/16) -cousins.

□ $\Phi(A,B) = k_1(A,B)/4 + k_2(A,B)/2$

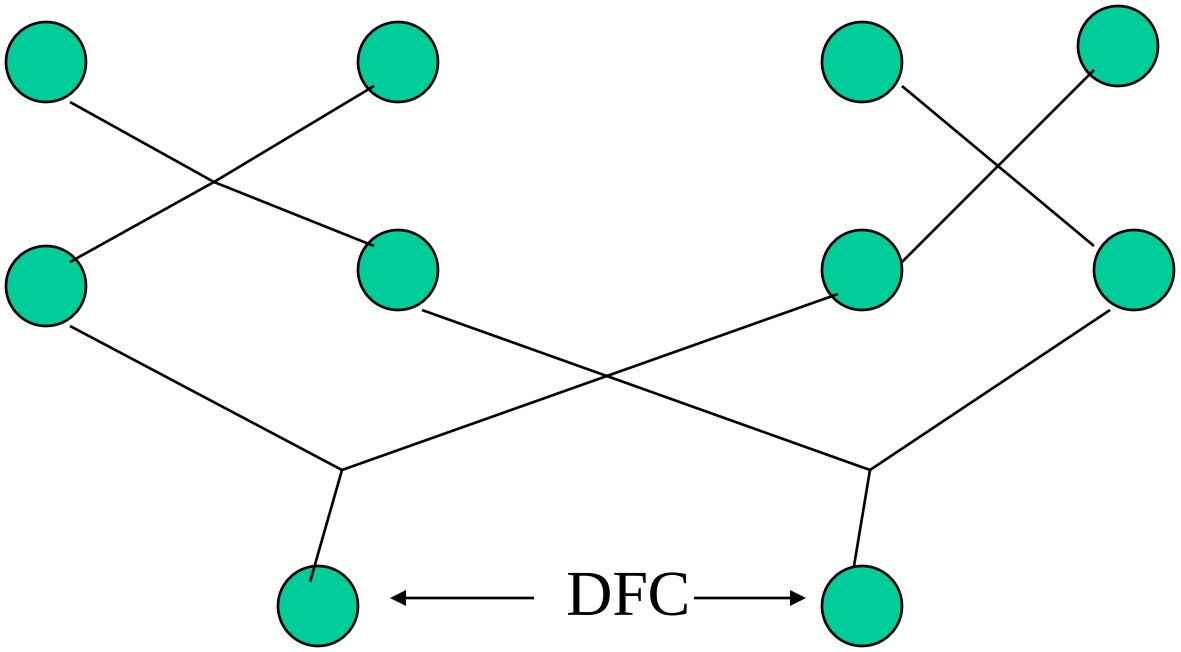
SIBS



1,3 (1,3) or (1,4) or (2,3) or (2,4)

so **k** = (1 ,2, 1) /4

DOUBLE FIRST COUSINS



NON-INBRED IND'S

- $P(J=AA \mid I=AA)$

$$= \frac{P(J=AA \cap I=AA)}{P(I=AA)}$$

$$= \frac{\Sigma\, P(J=AA \cap I=AA \mid k_i)\, k_i}{P(I=AA)}$$

$$= \frac{k_0 * p^4 + k_1 * p^3 + k_2 * p^2}{p^2}$$

NON-INBRED IND'S

- P(J=AA|I=AA) =(k_0*p^4+ k_1*p^3+ k_2*$p^{2)}$/p^2
  Sibs **k**(S) = (1/4, 1/2, 1/4),
  P(J=AA|I=AA) = $(1+p)^2/4$
- First Cousins **k**(FC) = (3/4, 1/4, 0),
  P(J=AA|I=AA)= p(3p + 1)/4
- Double-first **k(**DFC) = (9/16, 6/16, 1/16)
  -cousins. P(J=AA|I=AA)=$((3p+1)/4)^2$.

PAIR OF SIBS OF NON-INBRED PARENTS.

- Suppose parents with given **k** then examine a pair of offspring. We need to calculate the π vector ,
  $\boldsymbol{\pi} = (\pi_1,\pi_2,\pi_3,\pi_4,\pi_5,\pi_6,\pi_7,\pi_8,\pi_9)$ for each of the three cases.

PARENTS STATE (1,2,3,4)

- Offspring (1,3) wlog and ((1,2),(3,4)) so $\pi = (0, 0, 0, 0, 1, 0, 0, 2, 1)/4$.

PARENTS STATE (1,2,1,3)

- Offspring (1,2)*(1,3) and (1,2)*(1,3)
  i.e. Identity States
  (1,1), (1,3), (2,1), (2,3)

(1,1)	1	2	2	6
(1,3)	3	5	8	8
(2,1)	3	8	5	8
(2,3)	7	8	8	5

so

π = (1, 2, 2, 0, 3, 1 , 1, 6, 0) /16.

PARENTS STATE (1,2,1,2)

- Offspring (1,2)*(1,2)
  i.e. Identity States

	(1,1)	(1,2)	(2,1)	(2,2)	
(1,1)	1	2	2	4	
(1,2)	3	5	5	3	
(2,1)	3	5	5	3	
(2,2)	4	2	2	1	so

$\pi = (1, 2, 2, 1, 2, 0, 0, 0, 0)/8$

$\pi_0 \quad \pi_1 \quad \pi_2$

- Thus we have

  $\pi_0 = (0, 0, 0, 0, 1, 0, 0, 2, 1) /4$

  $\pi_1 = (1, 2, 2, 0, 3, 1 , 1, 6, 0) /16$

  $\pi_2 = (1, 2, 2, 1, 2, 0, 0, 0, 0)/8$

- $\pi_0 = (0, 0, 0, 0, 4, 0, 0, 8, 4) /16$

  $\pi_1 = (1, 2, 2, 0, 3, 1 , 1, 6, 0) /16$

  $\pi_2 = (2, 4, 4, 2, 4, 0, 0, 0, 0)/16$

OFFPRING OF FC's

- $\mathbf{k}$(FC) = (3, 1, 0) /4 so
  π(Offspring of FC) = $(3\pi_0 + \pi_1)/4$.
  π_0 = (0, 0, 0, 0, 4, 0, 0, 8, 4) /16
  π_1 = (1, 2, 2, 0, 3, 1 , 1, 6, 0) /16
  π_2 = (2, 4, 4, 2, 4, 0, 0, 0, 0)/16 so
  π(Offspring of FC) =
  (1, 2, 2, 0, 15, 1, 1, 30, 12) / 64.

OFFPRING OF FC's

- All the information is in this π vector.
- Example:- Using
  π =(1,2,2,0,15,1,1,30,12)/64 we have
  Prob(AA,AA) = π_1*p +
  ($\pi_2 + \pi_3 + \pi_4 + \pi_5$)*$p^2$ +
  ($\pi_6 + \pi_7 + \pi_8$)*p^3 + π_9*p^4 =
  p*(1 + 19p + 32p^2 + 12p^3) / 64.

EXERCISE

- For general π calculate the probability that the individuals are AB and AB, and apply for offspring of first cousins.

P(ABAB)

Clearly we cannot have an identity state which requires the two alleles of an individual to be IBD so we only look at states 5 (1212), 8 (1213) and 9 (1234) so obtain

$P(ABAB) = 2pq\,\pi_5 + (pq^2 + p^2q)\,\pi_8 + 4p^2q^2\pi_9$

and for first cousins this gives

$(30pq + 30(pq^2 + p^2q) + 48p^2q^2)/64$

GENOTYPE PAIRS.

- Suppose for a pair of individuals that we know the value of π then all the information about that pair is captured in that value. From it we can calculate the probability of any event concerning the genotypes of those individuals (given the population parameters e.g. allele frequencies).

COUNSELLING

- Suppose for a pair of individuals, I and J, we know the π-vector. Suppose one of them has a known genotype then we may wish to calculate the probability of the other's genotype.
- e.g. P(J=AA | I=AA)

COUNSELLING

- $P(J=AA \mid I=AA)$

$$= \frac{P(J=AA \cap I=AA)}{P(I=AA)}$$

$$= \frac{\Sigma\, P(J=AA \cap I=AA \mid \pi_i)\, \pi_i}{\Sigma\, P(I=AA \mid \pi_i)\, \pi_i}$$

Numerator $= p*\pi_1 + p^2*(\pi_2 + \pi_3 + \pi_4 + \pi_5) + p^3*(\pi_6 + \pi_7 + \pi_8) + p^4*\pi_9.$

COUNSELLING

- Numerator $= p*\pi_1 + p^2*(\pi_2 + \pi_3 + \pi_4 + \pi_5) + p^3*(\pi_6 + \pi_7 + \pi_8) + p^4*\pi_9$.
- Denominator $= p*(\pi_1 + \pi_2 + \pi_4 + \pi_6) + p^2*(\pi_3 + \pi_5 + \pi_7 + \pi_8 + \pi_9)$
- Ratio = P(J = AA | I = AA)

COUNSELLING

- Example. Sibs with unrelated parents so
  $\pi = (0,0,0,0,1,0,0,2,1)/4$ so
  Numerator = $(p^2 + 2*p^3 + p^4)/4$
  $= p^2*(1+p)^2 /4$
  &
  Denominator = p^2 so for sibs I and J
  $P(J = AA \mid I = AA) = (1+p)^2 /4$ which we have already derived from the **k**.

GENOTYPE PAIRS

- Prob of Genotypes given freq. A is p

State	1	2	3	4	5	6	7	8	9
AAAA	p	p^2	p^2	p^2	p^2	p^3	p^3	p^3	p^4
AAAB	0	pq	0	0	0	$2p^2q$	0	p^2q	$2p^3q$
AABB	0	0	0	pq	0	pq^2	p^2q	0	p^2q^2
ABAA	0	0	pq	0	0	0	$2p^2q$	p^2q	$2p^3q$
ABAB	0	0	0	0	$2pq$	0	0	pq	$4p^2q^2$
ABBB	0	0	pq	0	0	0	$2pq^2$	pq^2	$2pq^3$

GENOTYPE PAIRS

- State 1 2 3 4 5 6 7 8 9
  BBAA 0 0 0 pq 0 p^2q pq^2 0 p^2q^2
  BBAB 0 pq 0 0 0 $2pq^2$ 0 pq^2 $2pq^3$
  BBBB q q^2 q^2 q^2 q^2 q^3 q^3 q^3 q^4
- Checks Columns sum to 1.

GENOTYPE PAIRS

- Prob of Genotypes pairs or reverse.

State	1	2	3	4	5	6	7	8	9
AAAA	p	p^2	p^2	p^2	p^2	p^3	p^3	p^3	p^4
AAAB	0	pq	pq	0	0	$2p^2q$	$2p^2q$	$2p^2q$	$4p^3q$
AABB	0	0	0	$2pq$	0	pq	pq	0	$2p^2q^2$
ABAB	0	0	0	0	$2pq$	0	0	pq	$4p^2q^2$
ABBB	0	pq	pq	0	0	$2pq^2$	$2pq^2$	$2pq^2$	$4pq^3$
BBBB	q	q^2	q^2	q^2	q^2	q^3	q^3	q^3	q^4

FIRST COUSINS

π(Offspring of FC) =
(1, 2, 2, 0, 15, 1, 1, 30, 12) / 64.

GENOTYPE PAIRS

- So for offspring of First Cousins we have
- AAAA $p(1 + 19p + 32p^2 + 12p^3)/64$
  AAAB $4pq(1 + 16p + 12p^2)/64$
  AABB $2pq(1 + 12pq)/64$
  ABAB $2pq(30 + + 24pq)/64$
  ABBB $4pq(1 + 16q + 12q^2)/64$
  BBBB $q(1 + 19q + 32q^2 + 12\,q^3)/64$
  Earlier we obtained
  $P(ABAB)=(30pq+30(pq^2+p^2q) + 48p^2q^2)/64$

GENOTYPE PAIRS

- Offspring of first Cousins again, suppose we require probability AA. Then from

  AAAA $p(1 + 19p + 32p^2 + 12p^3)/64$

  AAAB $2pq(1 + 16p + 12p^2)/64$

  AABB $pq(1 + 12pq)/64$

  Prob(AA) = $p(1 +19p +32p^2 + 12p^3 + 3q +$
  $32pq + 24p^2q +12pq^2)/64$

  N.B. we have to take ½ of AAAB and AABB!

GENOTYPE PAIRS

- Prob(AA) = $p(1 + 19p + 32p^2 + 12p^3 + 3q + 32pq + 24p^2q + 12pq^2)/64$

  $= p(1 + 19p + 32p^2 + 12p^3 + 3 - 3p + 32p - 32p^2 + 24p^2 - 24p^3 + 12p - 24p^2 + 12p^3)/64$

  $= p(4 + 60p)/64$

  $= p(4(p+q)+60p)/64$

  $= p^2 + pq/16$ as before

GENOTYPE PAIRS

- Offspring of FC's. Using formulae given above. If p(A) =0.4 then we have frequencies

  AAAA 0.09055 AAAB 0.13980
  AABB 0.02910 ABAB 0.26820
  ABBB 0.22380 BBBB 0.24855,

  as per the assignment for computing.

THE GENERAL CASE

- We have derived appropriate vectors of probabilities for sibs whose parents have probability vector **k.**
- We now specify the complete probabilites for sibs whose parent have vector $\boldsymbol{\pi}$. We write $\boldsymbol{\pi}(S) = A\,\boldsymbol{\pi}(\text{Parents})$

THE GENERAL CASE

- Parents (1,1,1,1) then offspring (1,1,1,1) i.e. offspring π = (1,0,0,0,0,0,0,0,0).
- Parents (1,1,1,2) or (1,2,1,1) then offspring (1,1,1,1), (1,1,1,2), (1,2,1,1) or (1,2,1,2) with equal probabilitites so offspring π=(1,1,1,0,1,0,0,0,0)/4.
- Parents (1,1,2,2) then offspring π=(0,0,0,0,1,0,0,0,0)

THE GENERAL CASE

- Parents (1,2,1,2) then each offspring (11,12,22) with probs 1:2:1 so offspring π=(2,4,4,2,4,0,0,0,0)/16 as before.
- Parents (1,1,2,3) so offspring (1,2,1,2), (1,2,1,3) or (1,3,1,3) so offspring π=(0,0,0,0,1,0,0,1,0)/2.

THE GENERAL CASE

- Parents (1,2,1,3) then offspring π=(1,2,2,0,3,1,1,6,0) /16 as before &
- Parents (1,2,3,4) then offspring π=(0,0,0,0,1,0,0,2,1) /4.

THE GENERAL CASE

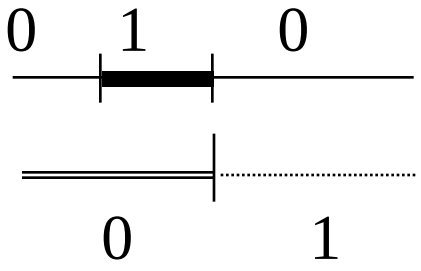
	1	2	3	4	5	6	7	8	9
1	1	1/4	1/4	0	1/8	0	0	1/16	0
2	0	1/4	1/4	0	1/4	0	0	1/8	0
3	0	1/4	1/4	0	1/4	0	0	1/8	0
4	0	0	0	0	1/8	0	0	0	0
5	0	1/4	1/4	1	1/4	1/2	1/2	3/16	1/4
6	0	0	0	0	0	0	0	1/16	0
7	0	0	0	0	0	0	0	1/16	0
8	0	0	0	0	0	1/2	1/2	3/8	1/2
9	0	0	0	0	0	0	0	0	1/4

REPEATED SIB-MATING

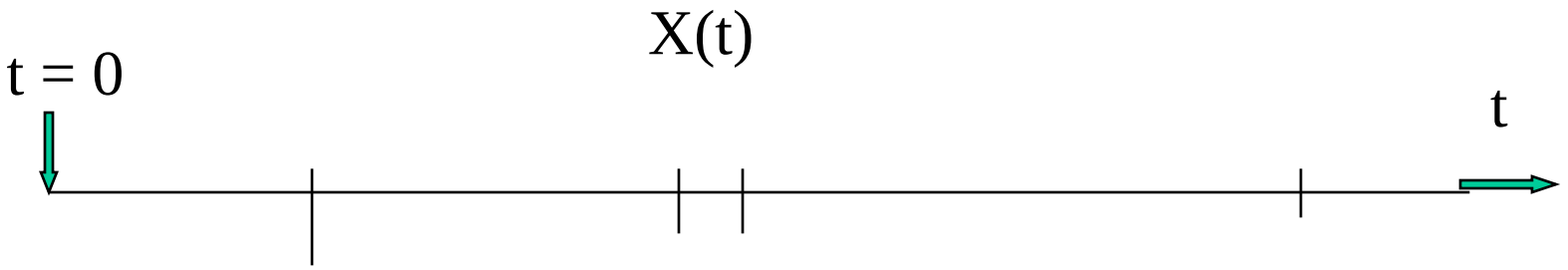
- Since $\pi(S) = A\ \pi(Parents)$ if we have a repeated sib-mating system then the parents are also sibs but in the previous generation, so $\pi_{\tau+1} = A\ \pi_{\tau}$

RECOMBINATION PROCESS.

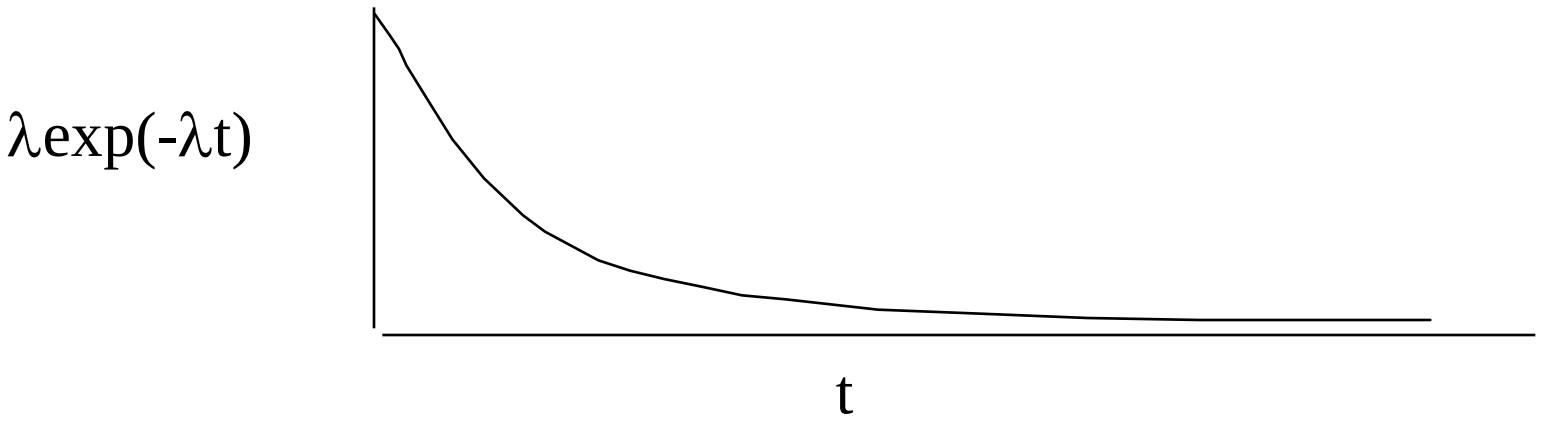
Then with prob=1/4



POISSON PROCESS.



Process with no memory. At any point in the process
Probability (No event in [0,t)) = exp(-λt)
irrespective of history.

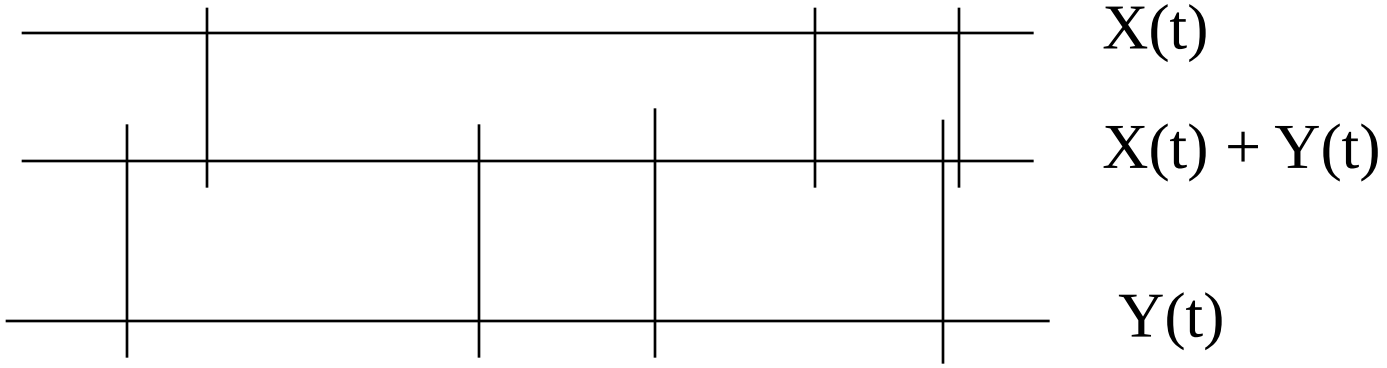


POISSON PROCESS.

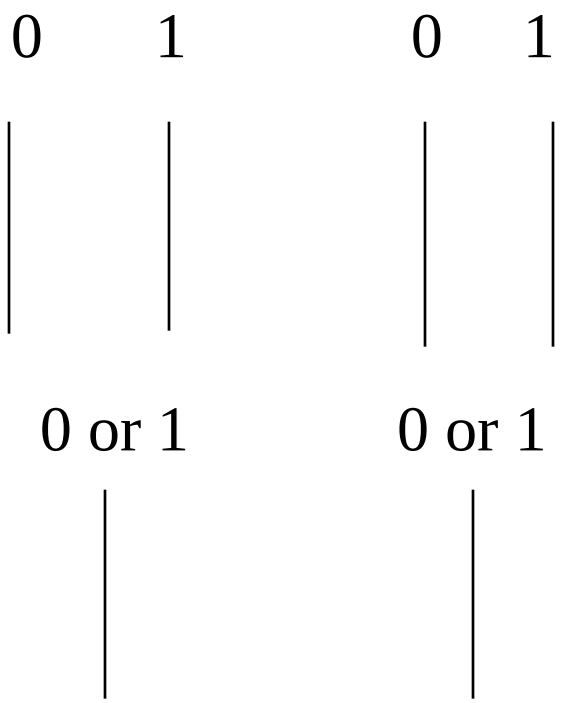
- λ is the rate of events, mean wait is $1/\lambda$.
- Number of events in [u, u+t) = X(t) then Prob(X(t) = r) = $(\lambda t)^r \exp(-\lambda t) / r!$ r = 0,1,…… the Poisson probability.
- Number of events in [v, v+s) = Y(s) then X(t) and Y(s) are independent if [u,u+t) and [v,v+s) non-overlapping.

POISSON PROCESS

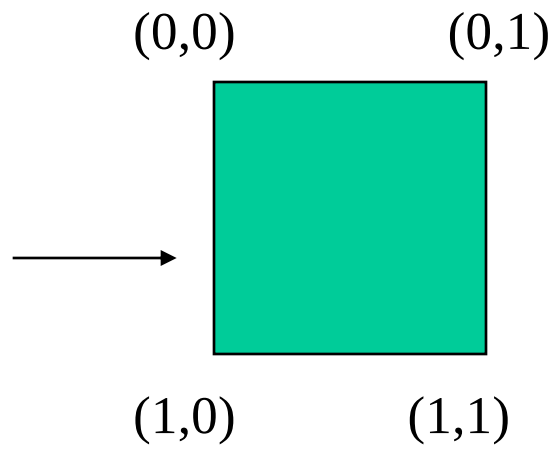
- Adding Poisson Processes.
  If X(t) is PP rate λ and Y(t) is PP rate ρ then X(t) + Y(t) is PP rate $\lambda + \rho$.
-



AN OFFSPRING.

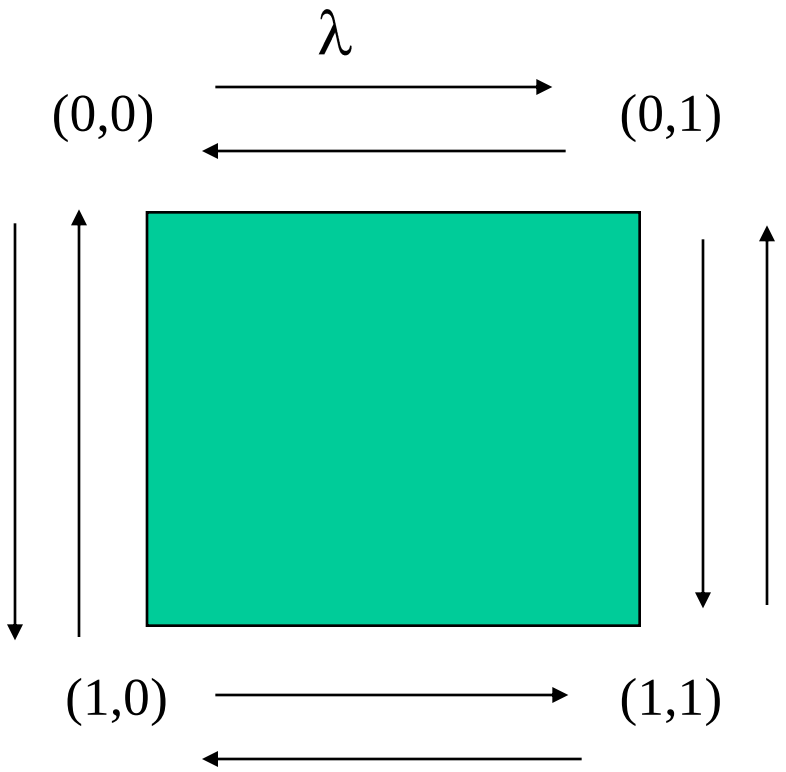


Jointly: start at random corner
move to neighbours at rate λ



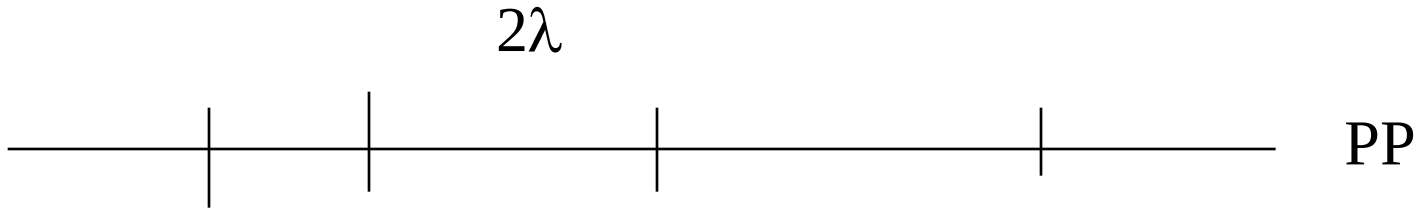
RANDOM WALK ON SQUARE

Rate of movement 2λ. Rate λ on every arrow

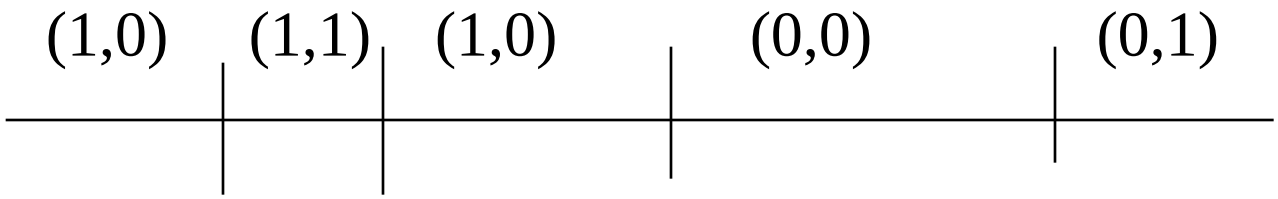


RANDOM WALK ON SQUARE

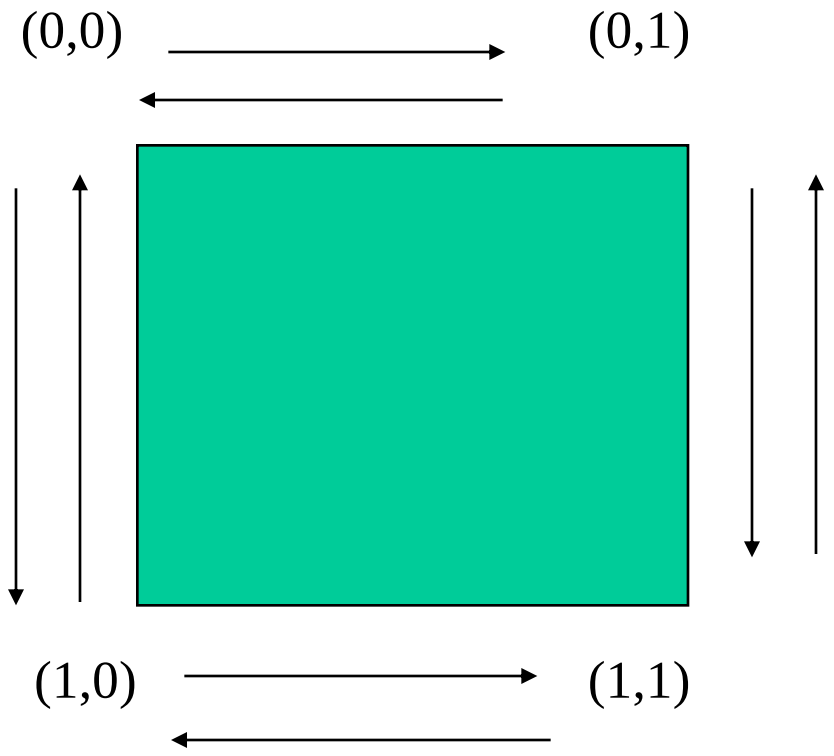
- Realisation of Random Walk



so possibly



IBD ALONG CHROMOSOME FOR SIBS.



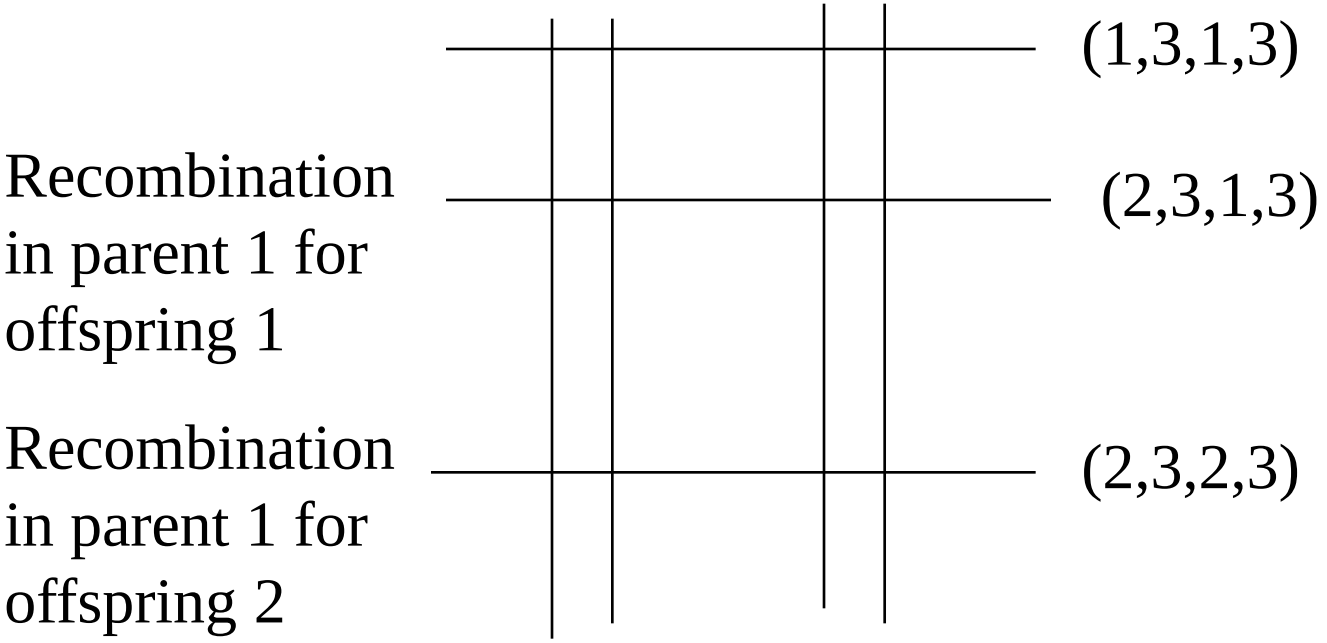
Applies for each
sib produced, and
independent processes..

IBD ALONG CHROMOSOME FOR SIBS.

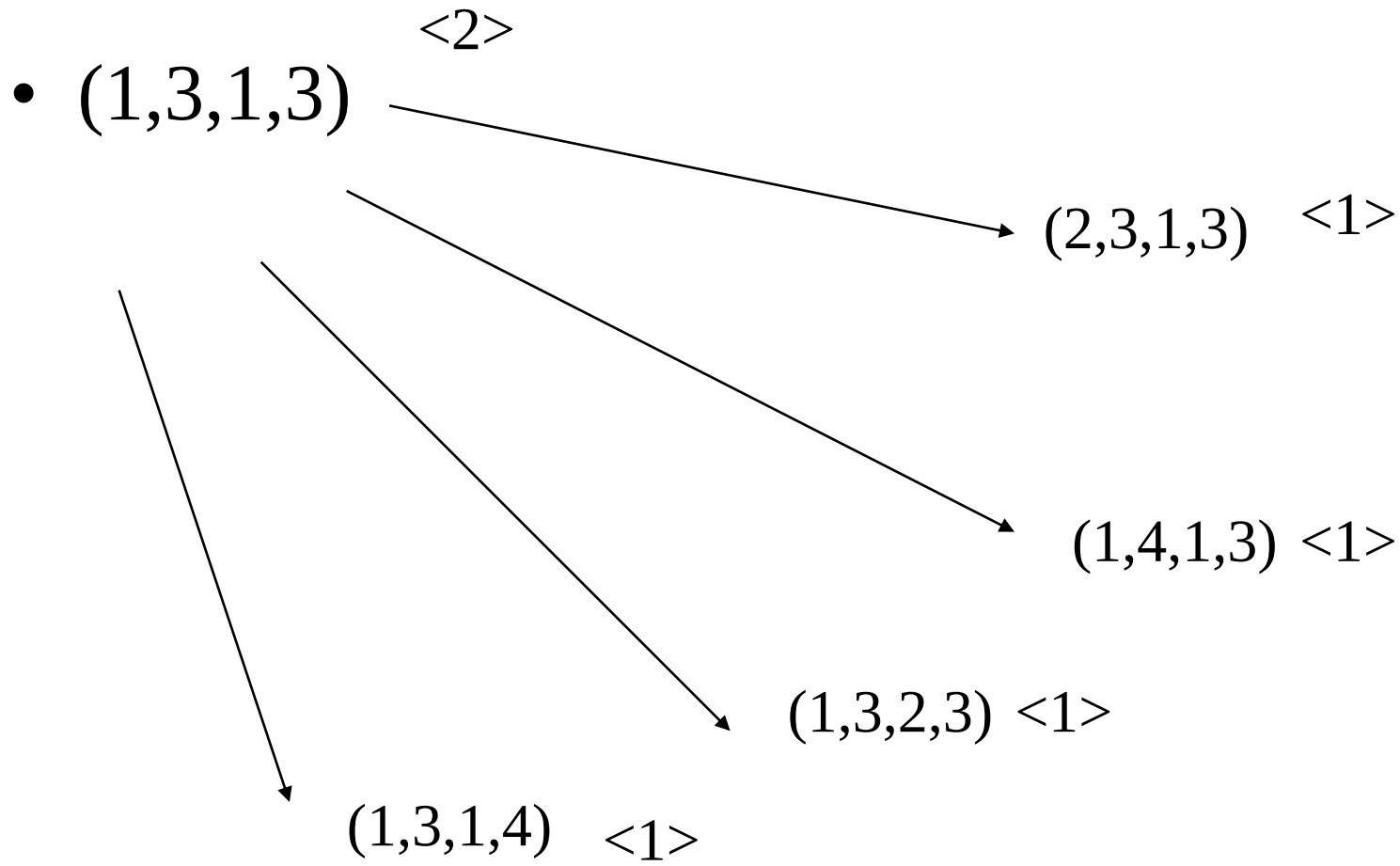
- For two sibs; each is moving around the vertices of the unit square at rate 2λ. We are interested in the state along the four chromosomes, i.e. (0,0,0,0), (0,0,0,1), …. which change with a rates λ for a change of each value. We can collapse the 16 states to four if we are only interested in the IBD according to whether there is a matching.

IBD ALONG CHROMOSOME FOR SIBS.

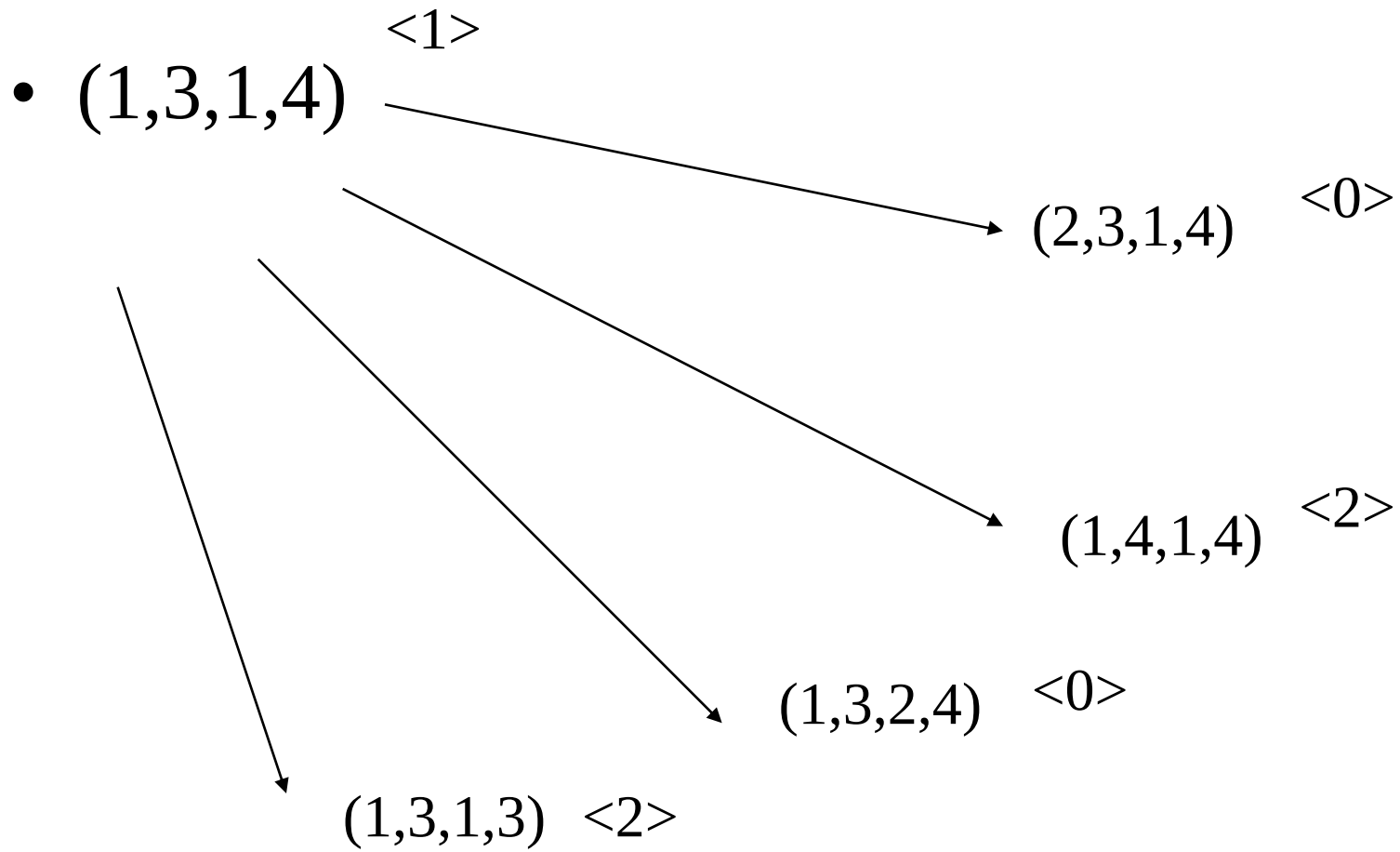
- Parents (1,2) * (3,4)



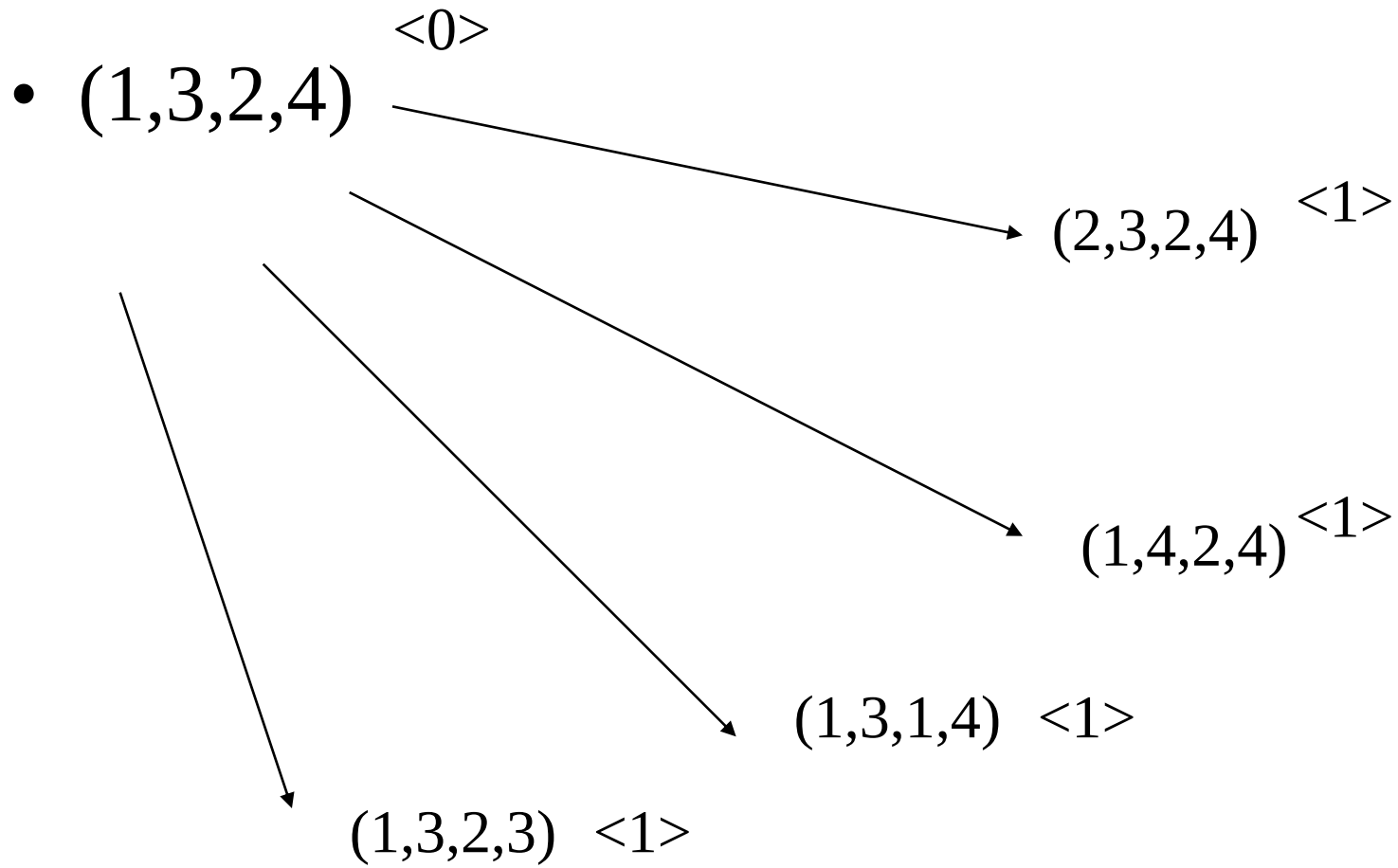
IBD ALONG CHROMOSOME FOR SIBS.



IBD ALONG CHROMOSOME FOR SIBS.

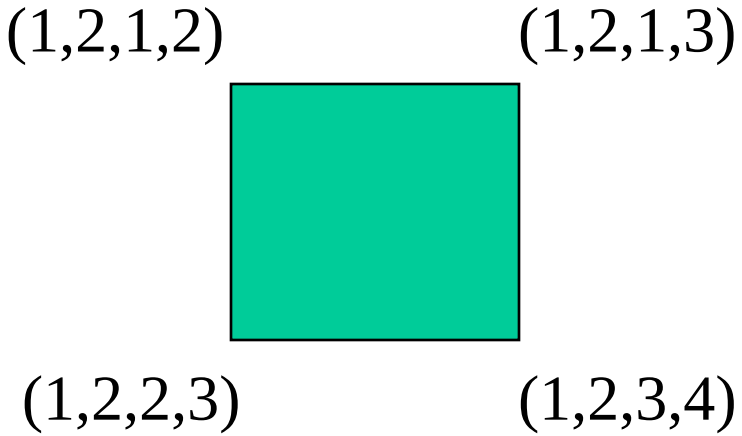


IBD ALONG CHROMOSOME FOR SIBS.



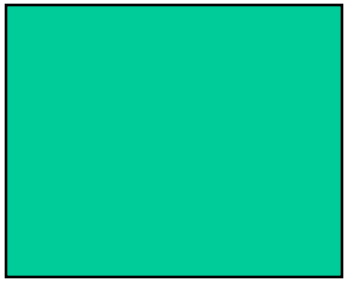
IBD ALONG CHROMOSOME FOR SIBS

- Thus (0,0,0,0) $\longrightarrow$ (1,2,1,2) in IBD, and also (0,1,0,1), (1,0,1,0) and (1,1,1,1).
- The result is that we can collapse the process onto the square again with rate 4λ.



IBD ALONG CHROMOSOME FOR SIBS

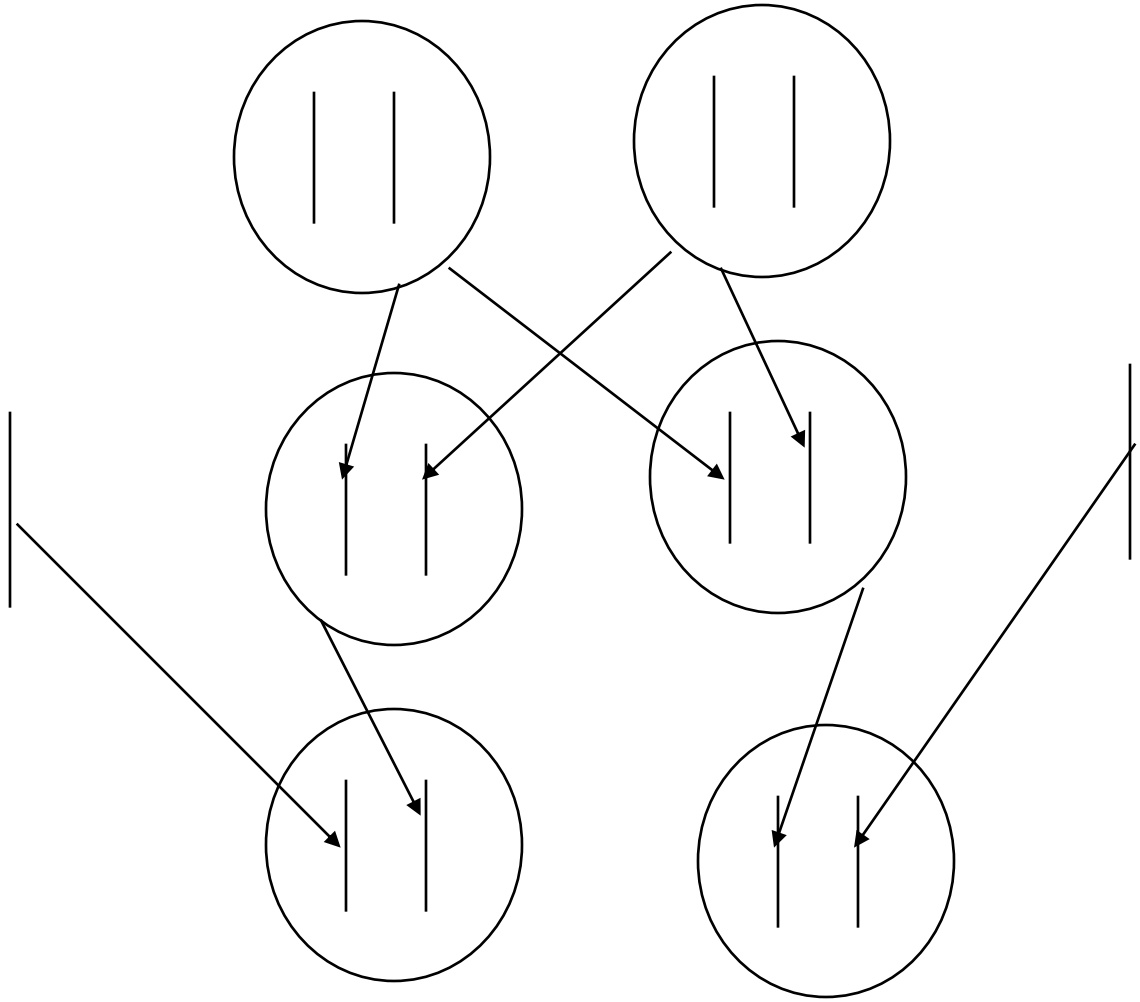
- Spends 1/4 time at each vertex (on average),



IBD ALONG CHROMOSOME FOR SIBS

- So along a chromosome of length t we will have a mean number of regions = $4\lambda t+1$. The average length of a region with 0, 1 or 2 genes IBD will be $t/(4\lambda t+1)$ i.e. approx. $1/4\lambda$

IBD ALONG CHROMOSOME FOR FC



IBD ALONG CHROMOSOME FOR FC

- We are interested in the two chromosomes in the FC's which are copies of material from there common grand-parents. For each individual we need to keep track of the process of recombination from parent to the individual chromosome on (0),(1), and the process in whichever parent supplied that chromosome, also on (0),(1).

IBD ALONG CHROMOSOME FOR FC

- Thus we are interested in a process on the **product** of these two, i.e. on (0,0), (1,1) that is on a square again.
- For the individuals jointly we can collapse onto the same cube with rates 4λ essentially fixing one and letting the other change relatively.

IBD ALONG CHROMOSOME FOR FC

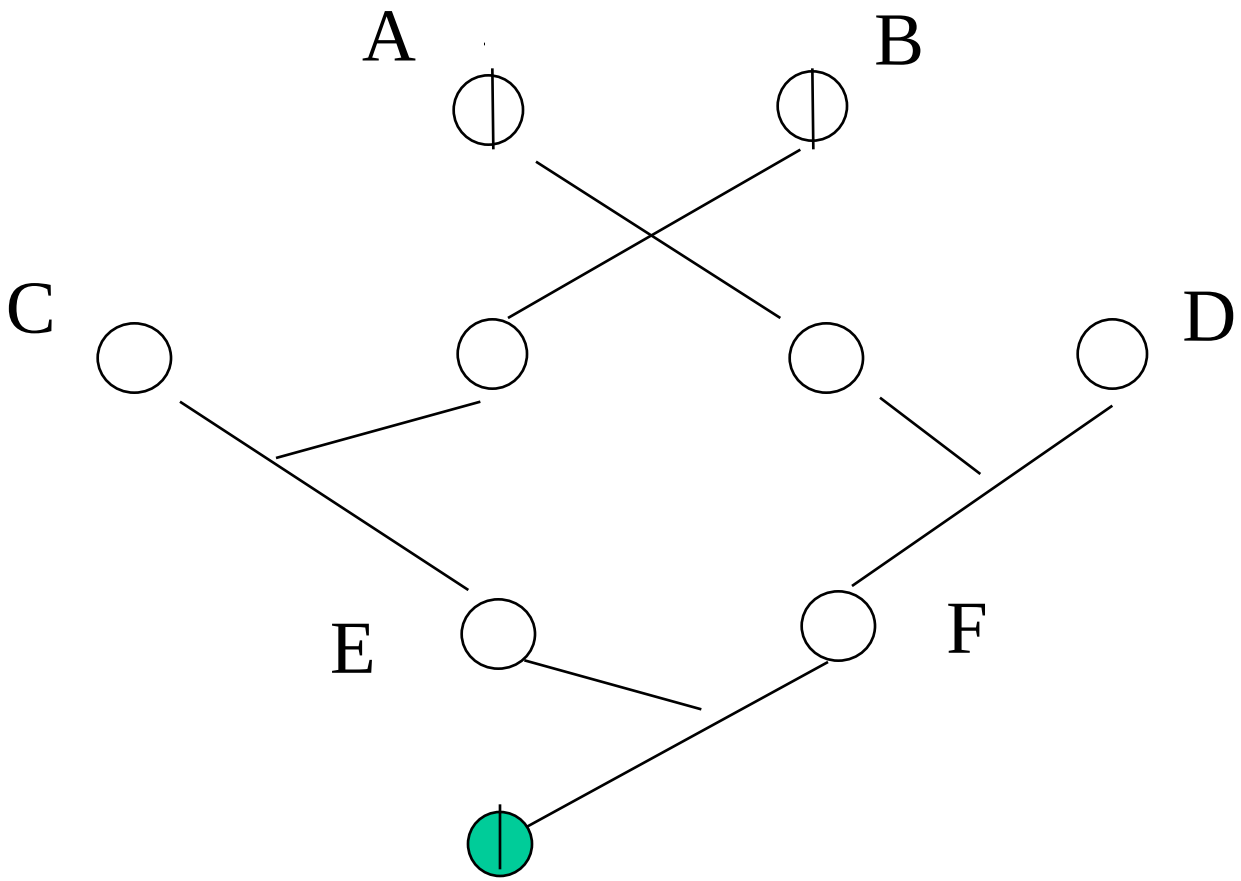
- If we fix the first at (0,0) then the other wanders around as a random walk on a 2-cube and there is IBD at a point if the state of the second is (0,0) also.
- We can see that there is IBD only if the process is in (0,0), which occurs 1/4 of the time, on average.

AUTOZYGOSITY

- For a recessive disease with population frequency of an allele p then the children of first cousins have an elevated frequency. If we were to examine such children who had the disease then many of those children will have acquired the two genes by descent from a common gene in their common ancestors.

AUTOZYGOSITY

- Offspring of FC's



AUTOZYGOSITY

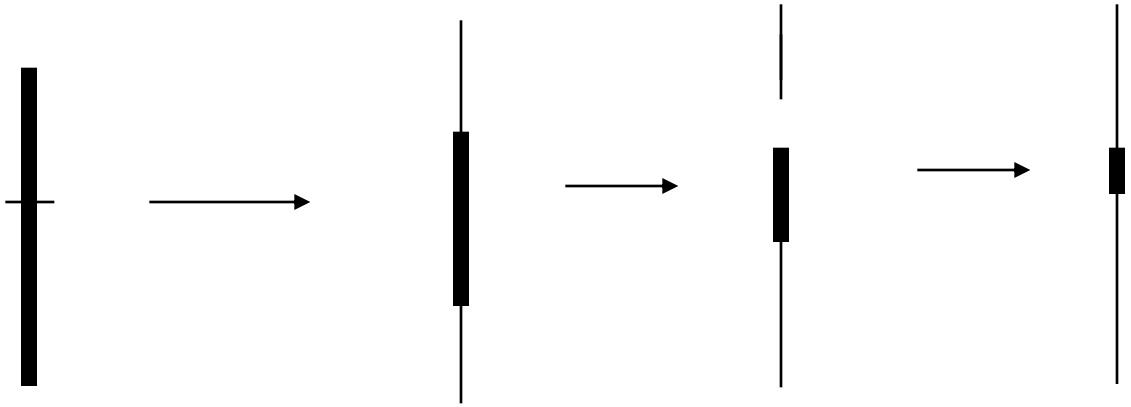
- We know that for a child of first cousins there is probability = 1/16 that the two genes are IBD, and 15/16 that they are not. For a recessive disorder with allele frequency p the probability that that an individual with the disease has got the genes IBD is $\frac{p/16}{p/16 + 15p2/16}$

AUTOZYGOSITY

- For p=0.01 this gives $\frac{0.01}{0.01 + 15*0.0001}$ which is almost 1.
- Thus first cousins will be virtually assured of having received genes IBD.

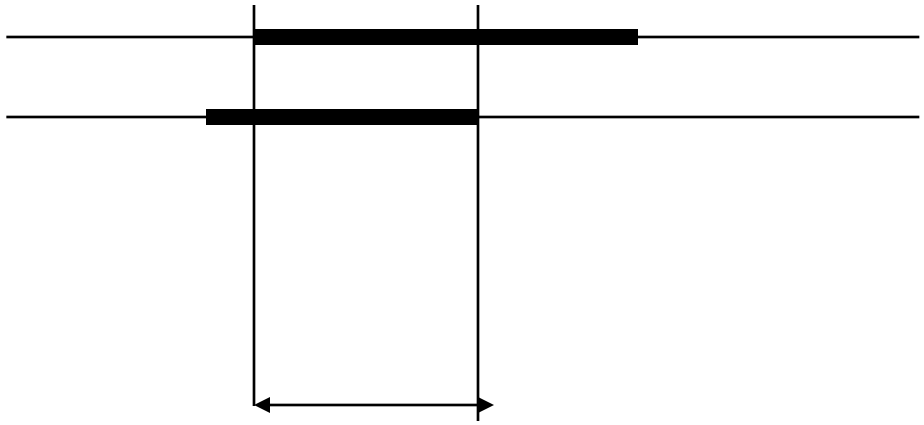
AUTOZYGOSITY

- Thus the chromosome which contained the recessive allele will have been passed down the two routes to the affected child.



AUTOZYGOSITY

- At the locus in question there will be a region of autozygosity



Region of Autozygosity

AUTOZYGOSITY

- On each side of the disease locus there will be a region of autozygosity which has a length X with density $f(x) = 6\lambda \exp(-6\lambda)$, there having been 6 possible crossovers. Thus the mean length will be $1/6\lambda$, where l is one/Morgan so the length will have mean 1/6 of a Morgan.

AUTOZYGOSITY

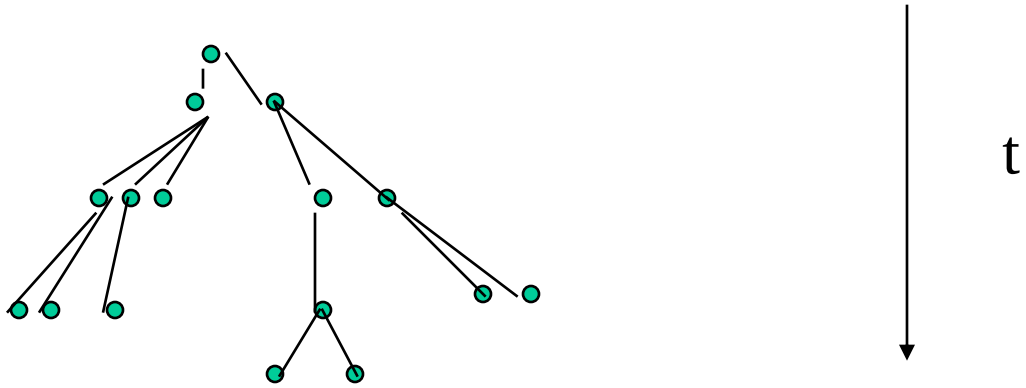
- If we look at a sample of such individuals we should be able to pick out those points on the genome where there is a region of homozygosity matching the special region of autozygosity. There will also be regions of homozygosity elsewhere.

AUTOZYGOSITY

- For a marker at distance r from the disease locus the probability that that marker is IBD on the chromosome with the disease allele is just the probability that there has been no recombination at any of the three segregations = $(1 - r)^3$

POPULATIONS.

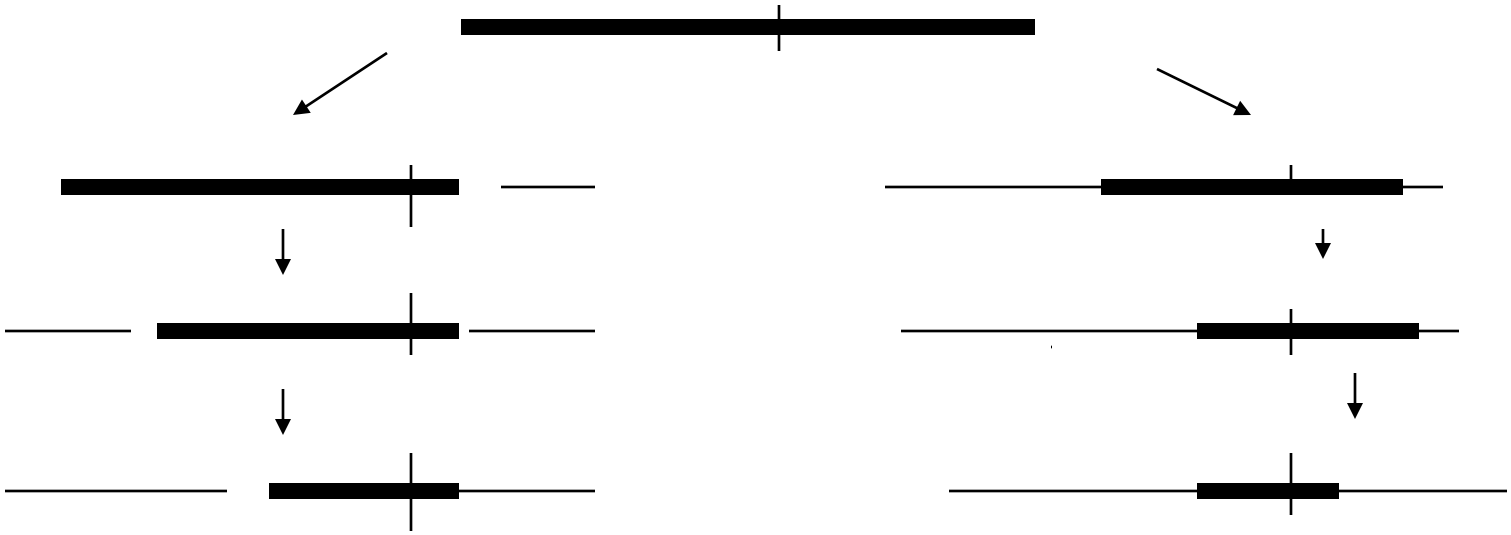
- Suppose that at some unique point in a populations history a recessive disease mutation occurred.



POPULATIONS.

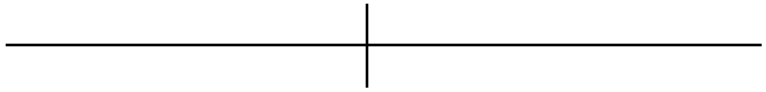
- If that mutation survives, (the chance of which is enhanced if the population underwent a rapid increase shortly after the mutation), then the chromosome on which the mutation occurred will also be passed on. Recombination will gradually break up the region around the mutation, but we may still be able to detect the region.

POPULATIONS.



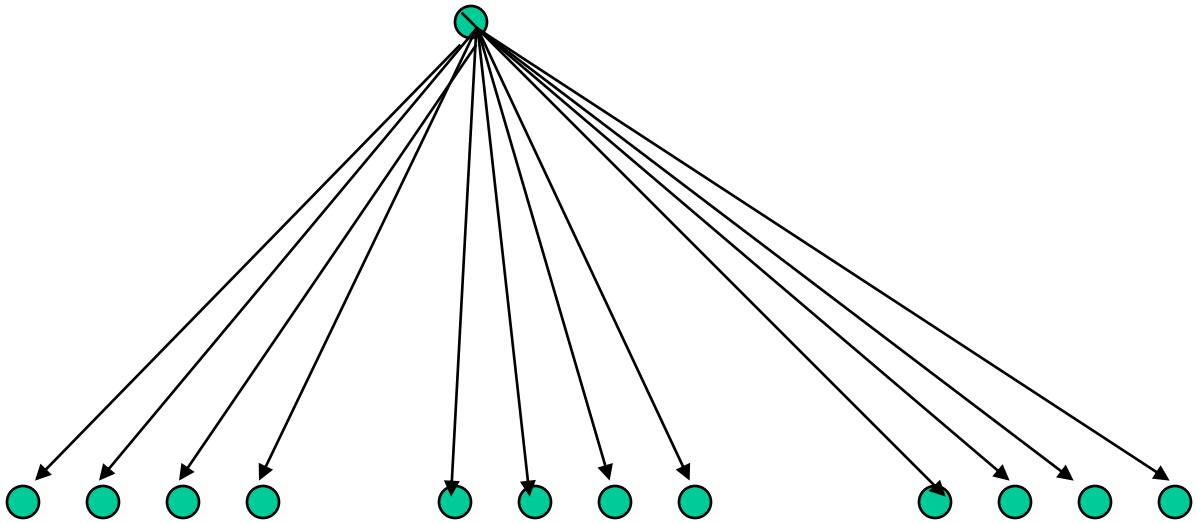
POPULATIONS.

- If we have
  initially
  and after n generations we have m chromosome carrying the disease allele then we would like to work out the region shared by these m around the disease allele.



POPULATIONS.

- The easiest case is



POPULATIONS.

- There have been n*m segregations so the length of common chromosome on each side of the disease allele, X, will be have density $\rho\exp(-\rho x)$, where $\rho=nm\lambda$. This region will be small, average $1/\pi$, but for sufficient large ρ , it will possibly be detectable vis-à-vis the chance that such a region would match by chance across nm.