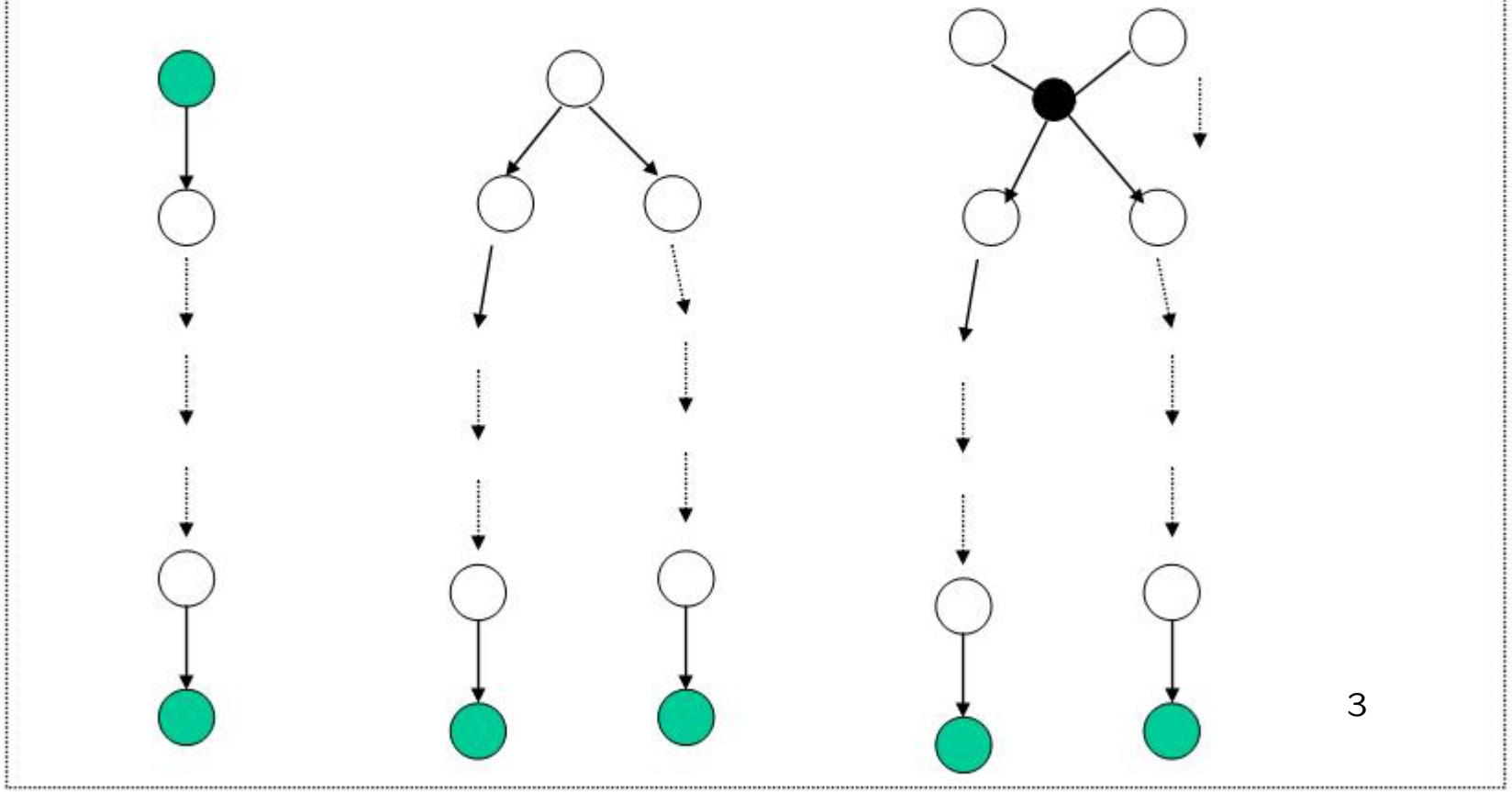
Basic Theory of IBD

Chris Cannings ,University of Sheffield, UK.

Bologna, May 2012.

Two individuals are related if they share a common ancestor. The relationship is uni-lineal if one is an ancestor of the other.

Related Individuals

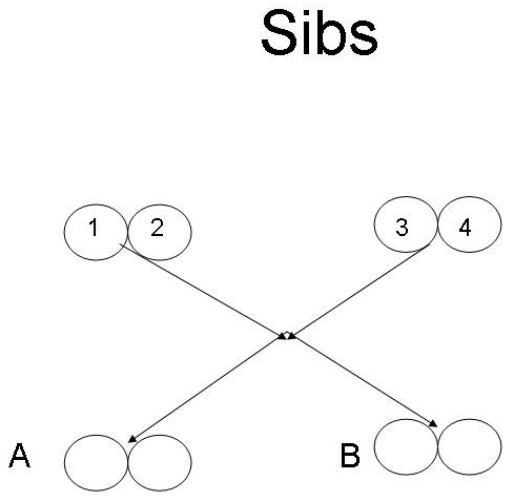


Our interest is in the inheritance of individual genes and how they may co-occur in related individuals. The fundamental concept is that of **IDENTITY BY DESCENT**.

Two genes are said to be **Identical by Descent** if they are copies of some single ancestral gene. Write **IBD**.

Given two (or more) individuals we wish to make calculations regarding the probability of their joint genotypes, etc. and for this we introduce various coefficients which specify the probabilities that genes are identical by descent.

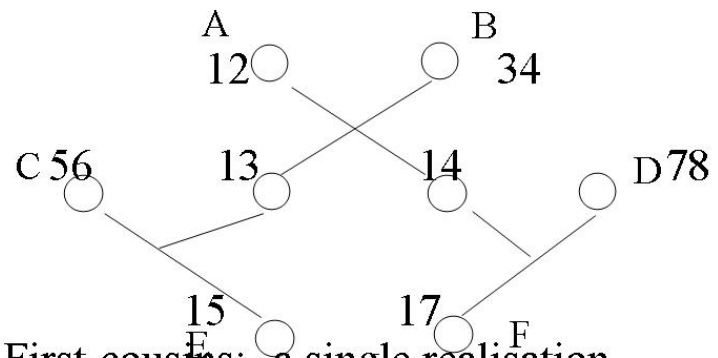
Given two distinct individuals A and B we define the **Coefficient of Kinship** as
$\phi(A, B) = P$[a randomly selected gene from A is **IBD** to a randomly selected gene from B].



Example. Pair of sibs A & B.

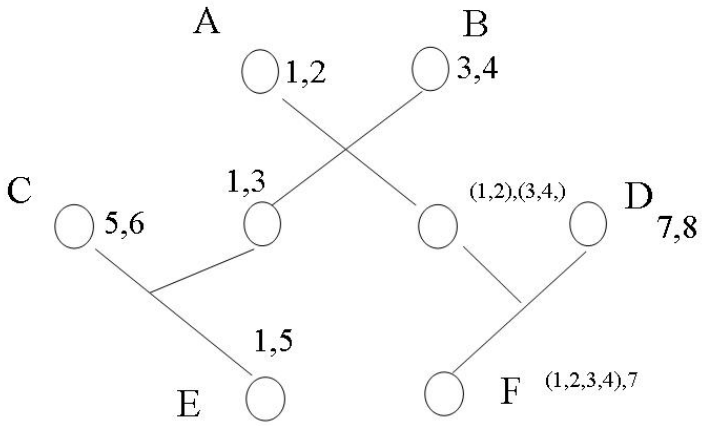
Random gene from A is 1,2,3 or 4 and pick from B, chance of being same is $\phi(A, B) = 1/4$.

IBD



First-cousins:- a single realisation

FIRST COUSINS.



Note that wlog (without loss of generality) we can choose for individual E to have genes (1,5) and individual F ((1,2,3,4),7) we have probability 1/2 for each of 1 and 5, and from F we have probabilities 1/8th for 1,2,3,4 and 1/2 for 7. Thus we find 1/16th as the probability that the same gene is picked, being 1. So $\phi(FC) = 1/16$.

When we consider a single individual we define the **Inbreeding Coeffient** $F(A) = P$(the two genes of A are **IBD**). Now if the parents of A are P(A) and M(A) then $F(A) = \phi(P(A), M(A))$. So the inbreeding coefficient for the offspring of a pair of First Cousins is 1/16.

Suppose we have an individual A with inbreeding coefficient F(A) and there is some locus with an alleles G and g with frequencies p and q where ($p+q=1$). Then $P(A|GG) = F(A)p+(1-F(A))p^2$.

Suppose an individual with F is GG then
$P(IBD|GG) = \frac{Fp}{Fp+(1-F)p^2}$,
Example. If $p = 0.01$ and $F = 1/16$ then
$P(IBD|GG) = \frac{0.01/16}{0.01/16+0.0015/16} = 100/115 \approx .87$.

The above suggest a possible way of detecting the location of a recessive disease. Select affected individuals whoes parents are first cousins and look for regions of IBD. We return to this shortly.

Now we might also wish to look at patterns of IBD across more than one individual. For this we need the concept of an **Identity State**. We confine ourselves to two individuals.

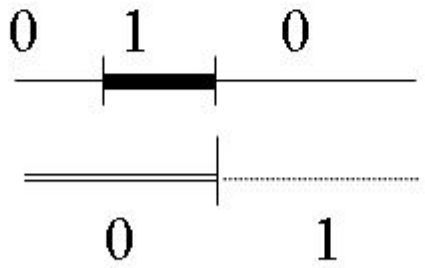
For the two individuals we have four genes and we might have **IBD** pattern $(*, \dagger)$ and $(\dagger, *)$. This we would abbreviate as $(1, 2, 2, 1)$.

IDENTITY STATES

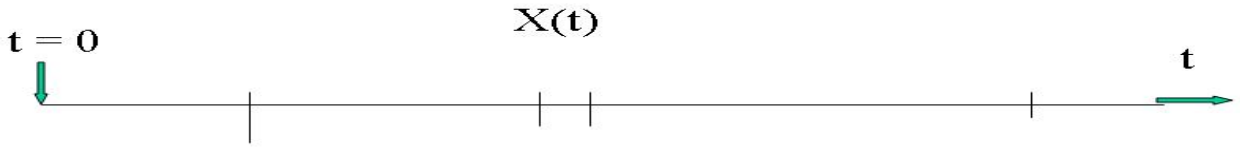
Identity State	Probability
(1,1,1,1)	π_1
(1,1,1,2) = (1,1,2,1)	π_2
(1,2,1,1) = (1,2,2,2)	π_3
(1,1,2,2)	π_4
(1,2,1,2) = (1,2,2,1)	π_5
(1,1,2,3)	π_6
(1,2,3,3)	π_7
(1,2,1,3) = (1,2,2,3) = (1,2,3,1) = (1,2,3,2)	π_8
(1,2,3,4)	π_9

RECOMBINATION PROCESS.

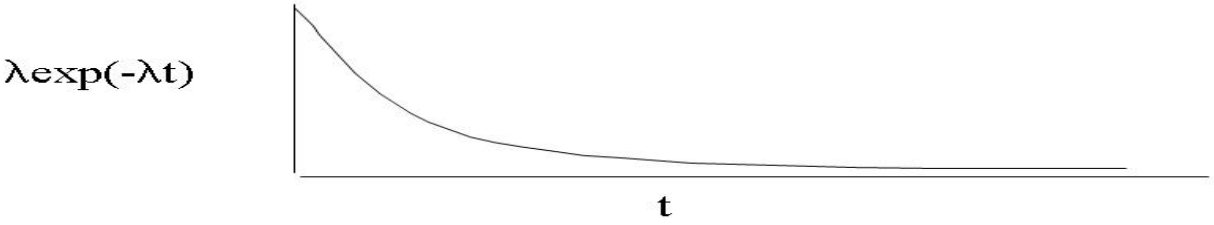
Then with prob=1/4



POISSON PROCESS.

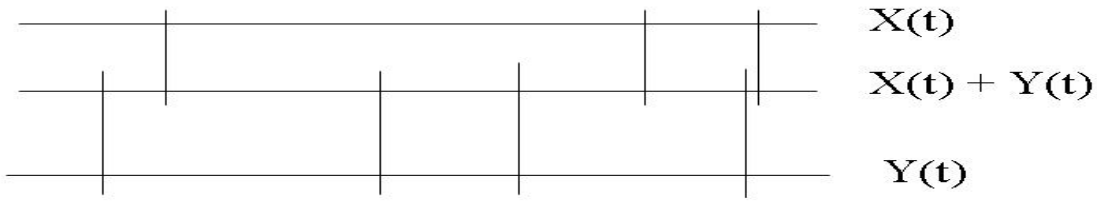


Process with no memory. At any point in the process
Probability (No event in [0,t)) = exp(-λt)
irrespective of history.

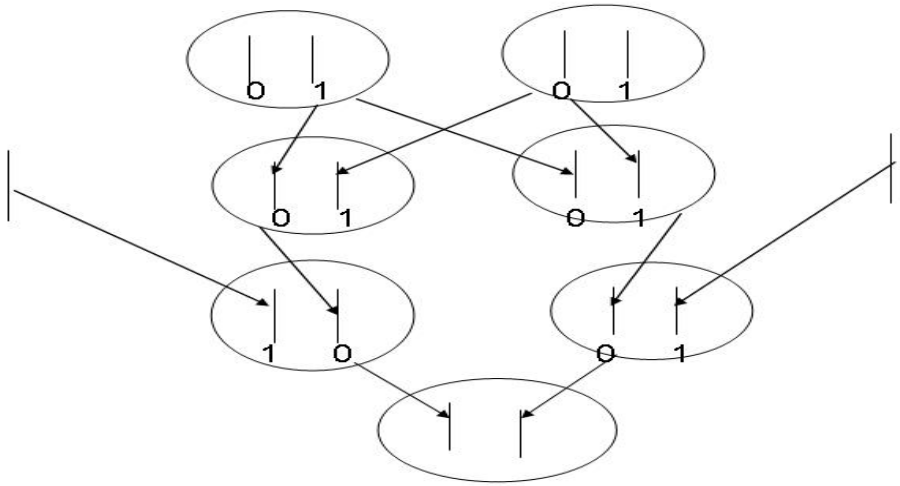


POISSON PROCESS

- 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 $\lambda + \rho$.
-

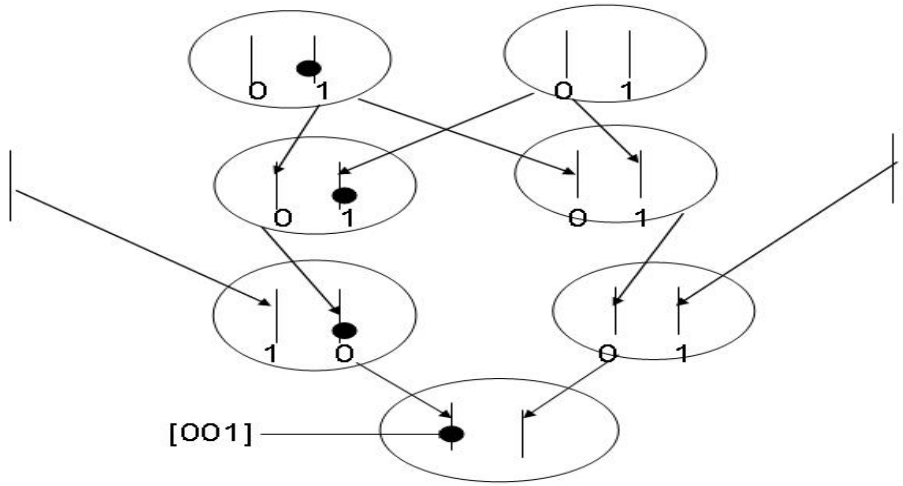


IBD ALONG CHROMOSOME FOR FC



Now we specify the parental chromosomal origin at every point along the chromosomes of the offspring of the FC by tracking back and recording the index.

IBD ALONG CHROMOSOME FOR FC



Thus the "state" of the chromosome at a point will be described by a triplet (000),(001),(010),(011),(100),(101),(110) or (111), and similarly for the other chromosome of interest. Thus there are 64 states specifying the descent of the genes at a locus.

Now we need to look at the "process" along the chromosome pair, that is the way in which changes are made between the states. Changes occur as a result of any recombination event during any of the 6 reproductions. For example if we had state ((000)(000)) at some point i.e. the genes were both copies of the same ancestral gene. Then there are six possible states to which the process would move as a result of a recombination ((100)(000)) ((010)(000)) ((001)(000)) ((000)(100)) ((000)(010)) ((000)(001)).

These occur as a result of the 6 possible recombinations. It is a property of the Poisson process that these occur with equal probability (i.e.1/6). Further the rate at which a recombination occurs is 6λ where λ is the rate per segregation.

We have 64 states with each state having 6 neighbours. The process spends equal time in each state (on average). However we can reduce the state space if we are interested only in IBD v not-IBD to size 9, if we wish to keep track of the external individuals separately from the first-cousins.

((0**)(0**))	((0**)(1**)) ((1**)(0**))	((1**)(1**))

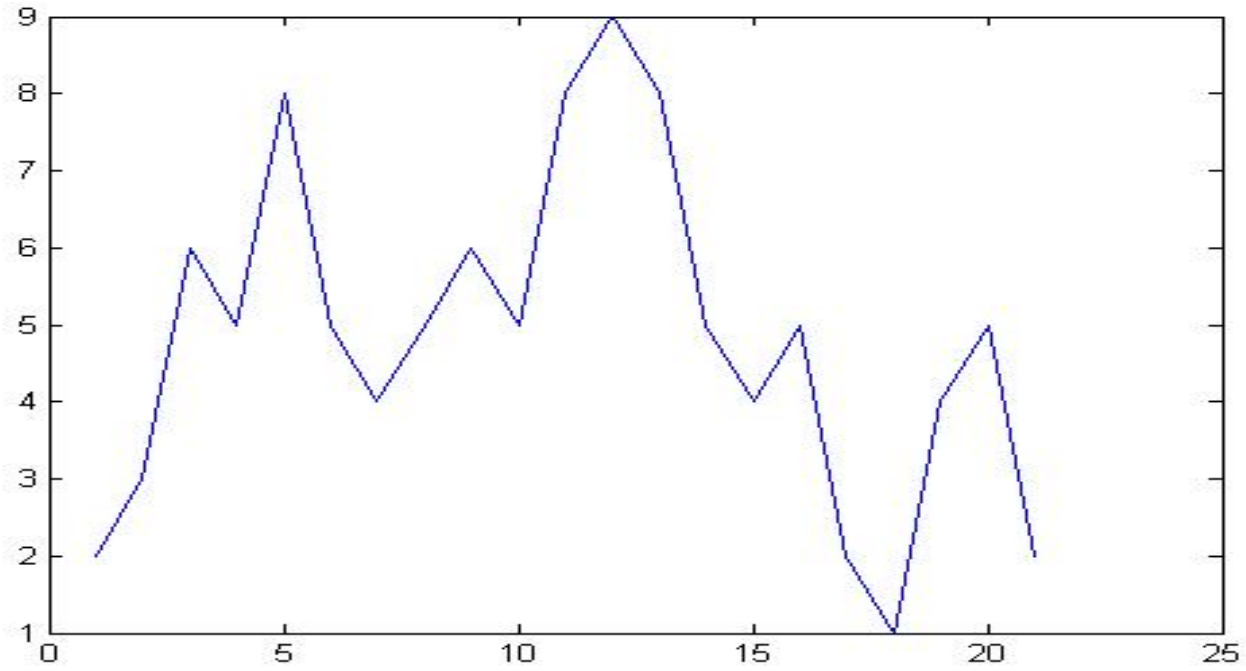
$$X$$

	((*00)(*01)) ((*00)(*10))	
((*00)(*00))	((*01)(*00))	((*00)(*11))
((*01)(*01))	((*01)(*11))	((*01)(*10))
((*10)(*10))	((*10)(*00))	((*10)(*01))
((*11)(*11))	((*10)(*11))	((*11)(*00))
	((*11)(*10))	
	((*11)(*01))	

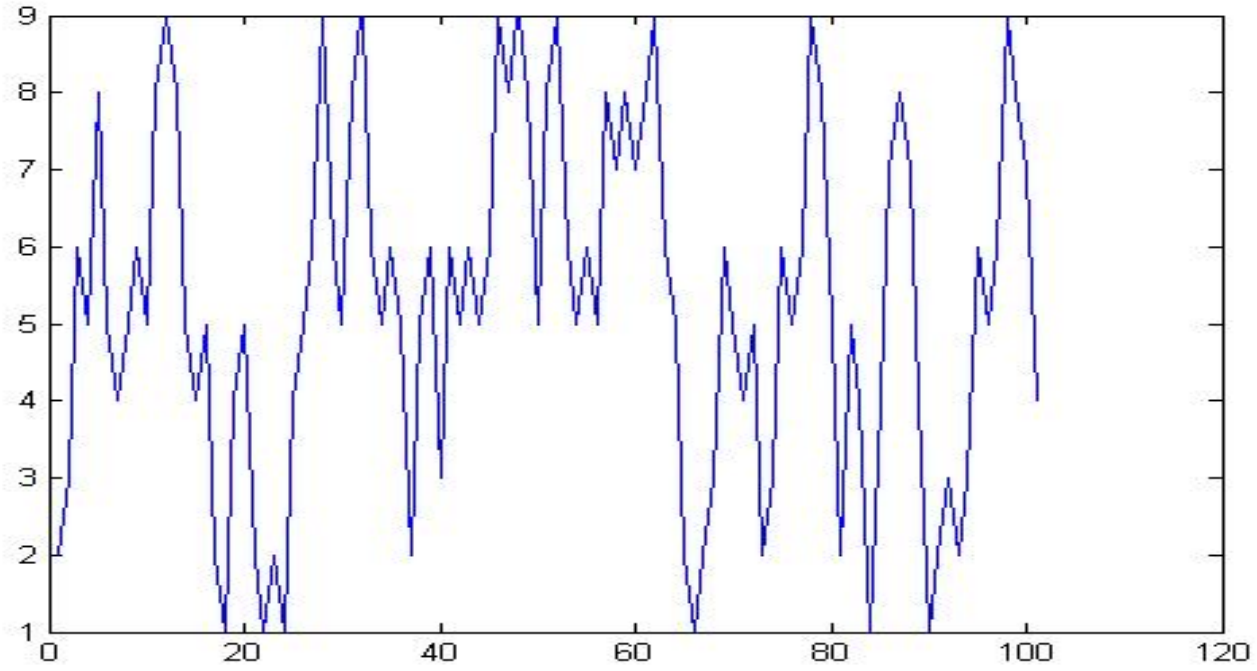
Transition Matrix * 6

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

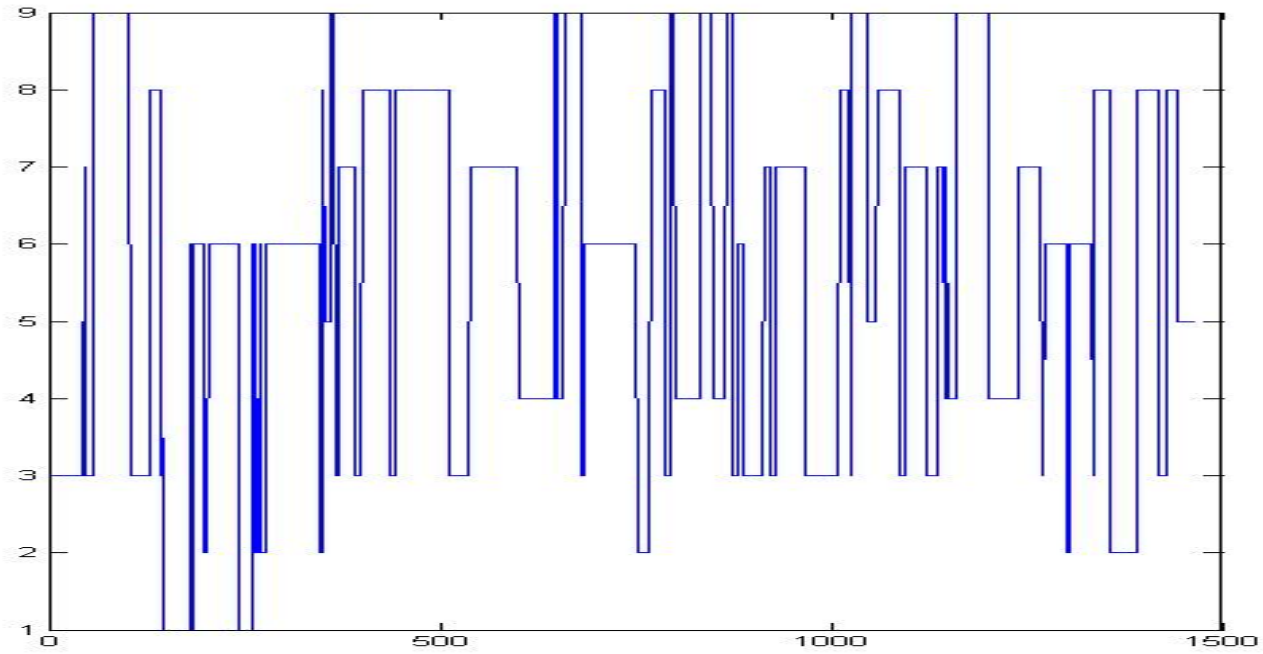
If current state is 6 then move to States 3, 5 or 0 with probs. 1/6, 4/6 and 1/6/

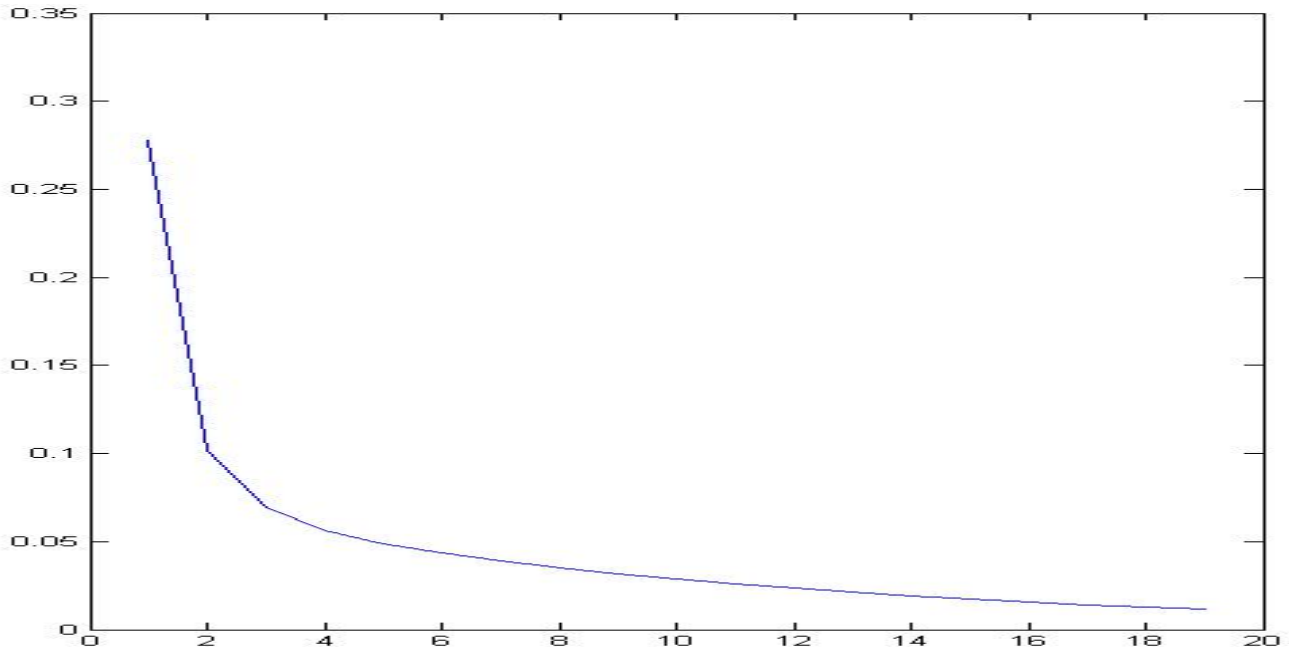


A Realisation of the IBD process switching.



A Realisation of the **IBD** process switching. (note the clustering of **IBD** regions


9
8
7
6
5
4
3
2
1
0
500
1000
1500



The distribution of inter-IBD gaps.

Alternately we can just keep track of the number of 0's in the state, so there are just 7 states.

Then the matrix is somewhat simpler.

$$\begin{matrix} 0 & 6 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 6 & 0 \end{matrix}$$

Offspring of Second Cousins

In this case we have four segregations for each of the two chromosomes i.e. states $((* * **)(* * **))$ and for **IBD** we need $((00ab)(00ab))$. now the 256 states can be reduced to 15.

In general for nth-cousins there $2^{2(n+2)}$ states reducible to $3(2n+1)$. **OR** $2n+5$ if ignore external/internal difference.

Now note that autosomal chromosomes (in humans) are between about 1 and 3 Morgans (differ between males and females). A Morgan is the length of chromosome in which the expected number of crossovers is 1. Thus for the example explored above we expect a rate of recombination (transition between states) of 6 per Morgan, so between 6 and 18 per human chromosome.

Autozygosity

Suppose now that we have a fully penetrant recessive condition, for example with allele frequency 1%. If we can collect affected individuals who are offspring of FC's then we saw that approximately 87% of these will have an **IBD** region around the location of the gene. In fact since we can observe the Poisson process extending out in both directions from the locus the average length of such an **IBD** region will be double the usual length. More importantly there will be a far greater number of **IBD** regions at this point across the sample than is usually the case.

IBD versus **IBS**

In practice, of course, one observes not the **IBD** state but the genotype. If that is heterozygote then we can infer non-**IBD** but if homozygote it may be that there are distinct lines of descent but common alleles. Note that the best possible SNP has 50% heterozgotes.

IDENTITY STATES

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 $\mathbf{k}(S) = (1/4, 1/2, 1/4)$,
- First Cousins $\mathbf{k}(FC) = (3/4, 1/4, 0)$,
- Double-first -cousins. $\mathbf{k}(DFC) = (9/16, 6/16, 1/16)$

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

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 $\mathbf{k}(S) = (1/4, 1/2, 1/4)$,
  $P(J=AA|I=AA) = (1+p)^2/4$
- First Cousins $\mathbf{k}(FC) = (3/4, 1/4, 0)$,
  $P(J=AA|I=AA) = p(3p + 1)/4$
- Double-first $\mathbf{k}(DFC) = (9/16, 6/16, 1/16)$
  -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

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

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

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

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

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

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

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

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

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.

OFFPRING OF FC's

- $\mathbf{k}(\mathrm{FC}) = (3, 1, 0)/4$ so
  $\pi(\text{Offspring of FC}) = (3\pi_0 + \pi_1)/4.$
  $\pi_0 = (0, 0, 0, 0, 4, 0, 0, 8, 4)/16$
  $\pi_1 = (1, 2, 2, 0, 3, 1, 1, 6, 0)/16$
  $\pi_2 = (2, 4, 4, 2, 4, 0, 0, 0, 0)/16$ so
  $\pi(\text{Offspring of FC}) =$
  $(1, 2, 2, 0, 15, 1, 1, 30, 12)/64.$

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
  $\text{Prob}(AA,AA) = \pi_1 * p +$
  $(\pi_2 + \pi_3 + \pi_4 + \pi_5) * p^2 +$
  $(\pi_6 + \pi_7 + \pi_8) * p^3 + \pi_9 * p^4 =$
  $p*(1 + 19p + 32p^2 + 12p^3) / 64.$

Now the issue arises of how the **IBD** changes along the chromosome for some pair of individuals. Leave this for another day.