UNIVERSITETET
I OSLO

Genetic Pedigree Estimation

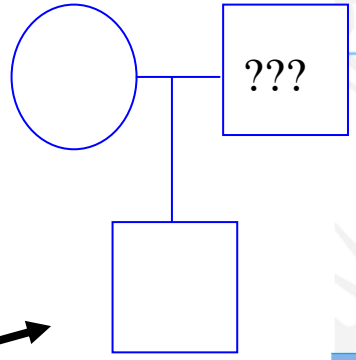
Thore Egeland, *Inst of Forensic Medicine, U of Oslo*
Nuala Sheehan, *Dep of Health Sciences and Genetics, U of Leicester*
Øivind Skare, *Dep of Public Health and Primary Health Care, U of Bergen*

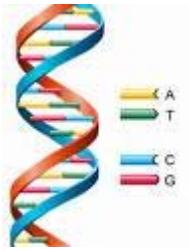
Graphical Models and Genetics Applications Workshop,
Warwick, April 15-17, 2009

UNIVERSITETET I OSLO

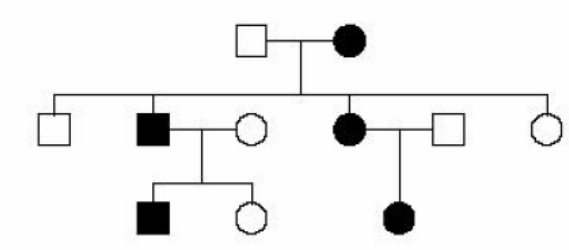
Contents

- *Nuala*: **Pedigrees and the estimation problem**
  - Motivation
  - Bayesian framework
  - Priors on pedigrees
- *Thore*: **How many markers and should they be linked or unlinked?**
  - IBD
  - Linked markers
  - Equivalence classes. (K. Donnelly, 1983)
- *Øivind*: **How far back can we go? Examples and conclusions**
  - R- Freeware http://folk.uio.no/thoree/FEST/
  - General framework for pairwise relationships
  - Graphs and pedigrees
- *Based largely on* Skare Ø, Sheehan N and Egeland T:
  - How distant family relationships can be detected? Submitted, 2009.





Common features: Problem, Data (DNA), Weight of evidence



UNIVERSITETET I OSLO

Still need for new methods I

- Thomas Jefferson og Sally Hemings
  - "In September 1802, political journalist James T. Callender, a disappointed office-seeker who had once been an ally of Jefferson, wrote in a Richmond newspaper that Jefferson had for many years "kept, as his concubine, one of his own slaves." "Her name is Sally," Callender continued, adding that Jefferson had "several children" by her."
  - "In January 2000, the committee reported its finding that the weight of all known evidence - from the DNA study, original documents, written and oral historical accounts, and statistical data - indicated a high probability that Thomas Jefferson was the father of Eston ..."
  - " Since then, a committee commissioned by the Thomas Jefferson Heritage Society, after reviewing essentially the same material, reached different conclusions, …".
  - DNA (Y-chromosome) cannot exclude (Randolph) or his sons

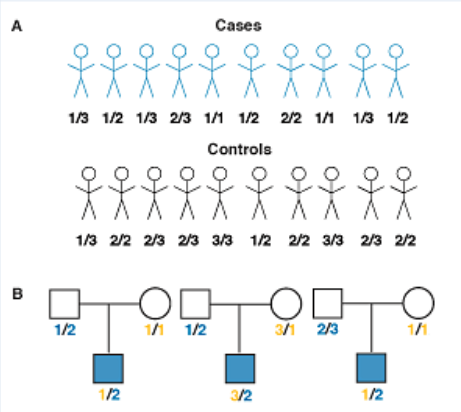
UNIVERSITETET
I OSLO

Haplotypes

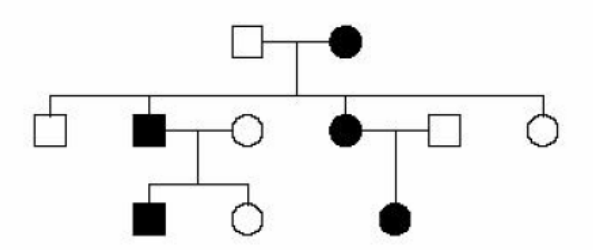
- Y-chromosome
  - Example: Jefferson paternity
- mtDNA
  - Example: Romanovs
- X-chromosome
  - Deficient paternity cases
- Other haplotypic data
  - HLA

Still need for new methods II

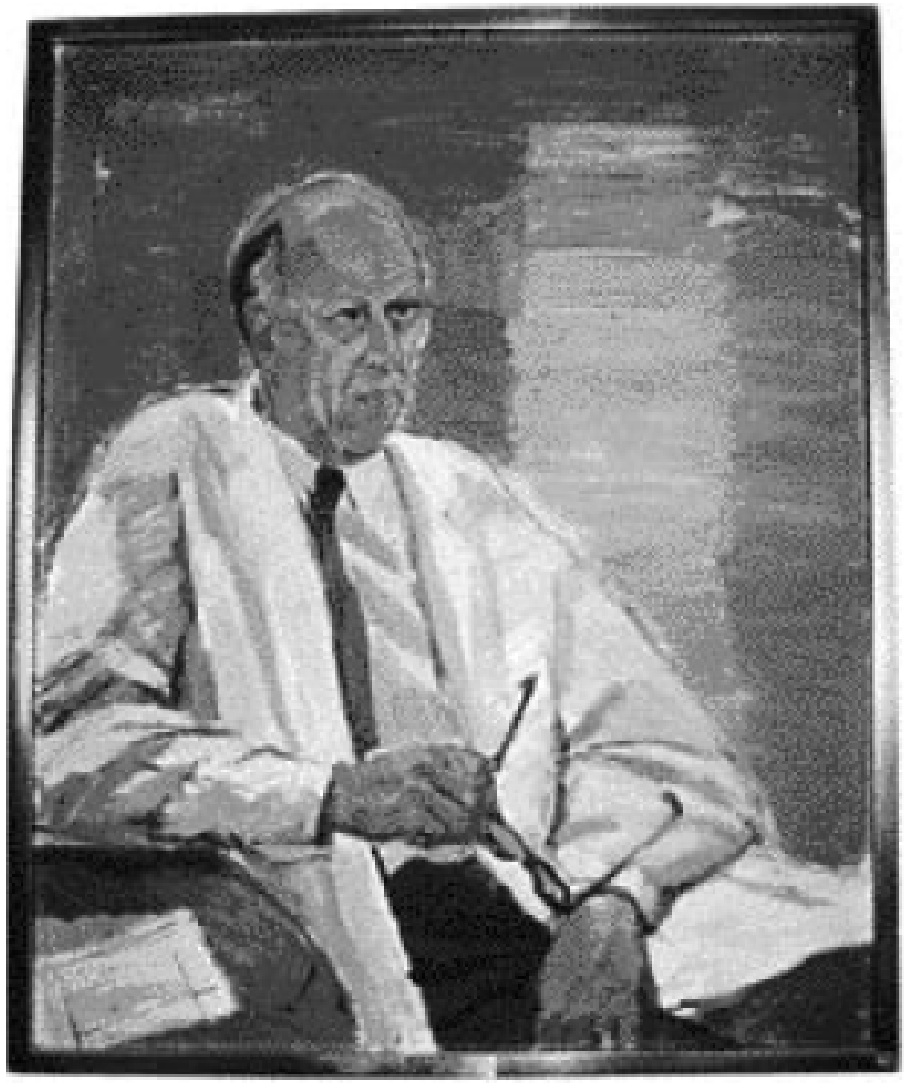
- Association analyses
  - Are individuals unrelated?



- Linkage
  - Are founders unrelated?
  - Correct pedigree?
- …



Methods



Erik Essen-Möller porträtterad av Brita af Klercker.

Forensic Science International, 25 (1984) 1—17
Elsevier Scientific Publishers Ireland Ltd.

ON THE THEORY AND PRACTICE OF ESSEN-MÖLLER'S *W* VALUE AND GÜRTLER'S PATERNITY INDEX (PI)

KONRAD HUMMEL

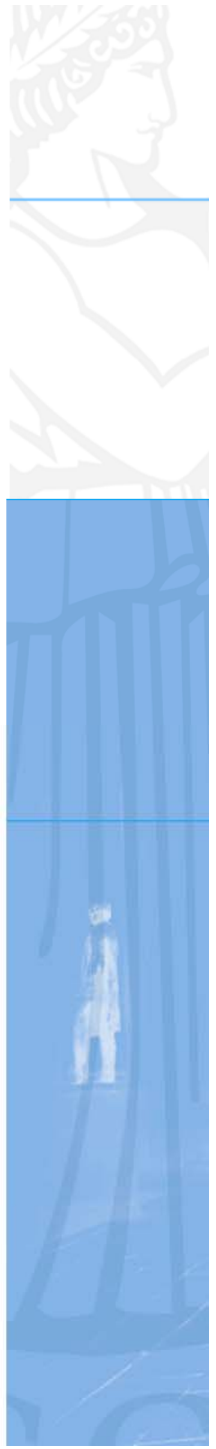
Institute for Blood Group Serology, University of Freiburg (F.R.G.)

(Received November 10, 1982)
(Revision received
(Accepted

Summary

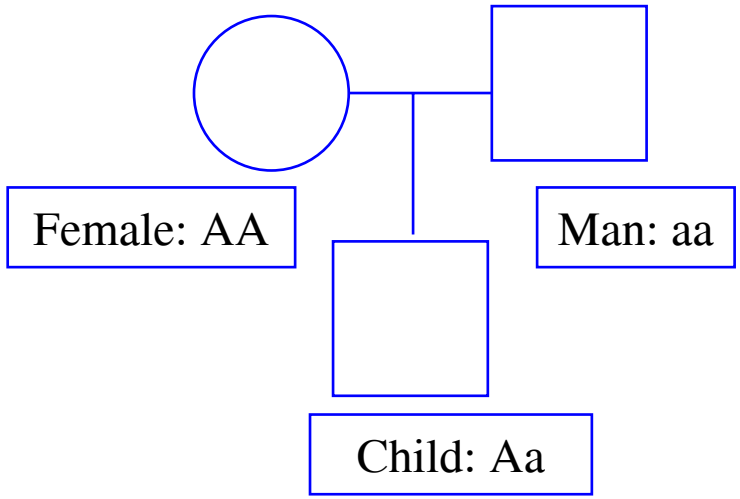
In cases of disputed parentage the biostatistical information is contained in the frequencies X and Y (as defined by Essen-Möller); X denotes the hypothesis "paternity", Y the hypothesis "non-paternity". Essen-Möller proposed a probability of paternity which includes both values: $W = X/(X + Y)$ (where $X + Y$ becomes 1). Gürtler recommends the ratio X/Y as a "Paternity Index" (=PI). Both W and PI are based on a *neutral* prior probability (=0.5 in normal triplet cases) and contain the *same* information, though differing in form. It is this difference which can lead to different results in forensic practice. W% is the common form for expressing probabilities, and each range of W values has an appropriate, easily understood verbal predicate. By contrast, the PI value is more abstract and can be interpreted as providing fixed decision limits, a possibility increased by the lack of distinct subdivisions with verbal predicates. Tables and computer programs are available for calculating W values even in complex cases. If one chooses to use PI values instead of W they must be calculated by the following formula:

$$\mathrm{PI} = \frac{W}{1 - W}.$$



Genetic terminology

Marker 1	
Allele	Frequency
A	0.9
a	0.1



Mendel

Bayesian framework

- Find a set of "possible" pedigrees $P_1,...,P_N \in \Omega$
- Set up prior probabilities $\pi(P_1),...,\pi(P_N)$ based on non-DNA information.
- Compute $\pi(\text{DNA}-\text{data}\,|\,P_i)$ for each pedigree P_i
- Make inferences from the posterior distribution:

$$\pi(P_i\,|\,\text{DNA}-\text{data}) = \frac{\pi(\text{DNA}-\text{data}\,|\,P_i)\pi(P_i)}{\sum_{j=1}^{N}\pi(\text{DNA}-\text{data}\,|\,P_j)\pi(P_j)}$$

Prior

- Sample space $\{P_1, \ldots, P_N\}$
- $$\pi(P_i) = const. \prod_{i=1}^{s} M_i^{b_i(P_i)} \prod_{\substack{j,k=1 \\ j \neq k}}^{n} R_{jk}^{o_{jk}(P_i)}$$

Global features

Local features; parent-child

Prior. Parameters and exponents

Calculated from pedigree

$$\pi(P_i) = const. \prod_{i=1}^{s} M_i^{b_i(P_i)} \prod_{\substack{j,k=1 \\ j \neq k}}^{n} R_{jk}^{o_{jk}(P_i)}$$

User specified parameters

Prior. Global part

- Example: One global feature, inbreeding.
  - Calculated from pedigree

$$b(P_i) = 1 \text{ if inbreeding, } 0 \text{ otherwise}$$

  - User specified prior belief of inbreeding:

$M = 0$ no inbreeding

M<1 downweigths

M=1 flat

M>1 upweights

$$\pi(P_i) = const.\ M^{b(P_i)}$$

Prior. Local part added

$o_{jk}(P_i) = 1$ if j parent of k, 0 otherwise

$R_{jk} = 0$ j not parent of k

$R_{jk} < 1$ downweigths parent-child relationship

$R_{jk} = 1$ flat

$R_{jk} > 1$ upweights parent-child relationship

$$\pi(P_i) = const. \prod_{i=1}^{s} M_i^{b_i(P_i)} \prod_{\substack{j,k=1 \\ j \neq k}}^{n} R_{jk}^{o_{jk}(P_i)}$$

$$= const.\ M^{b(P_i)} R_{jk}^{o_{jk}(P_i)}$$

Wine example from
Sheehan and Egeland.
Ann. Hum. Gen. (2007)

Table 5 Data from Bowers *et al.* (1999) giving the genotypes of Chardonnay, C, and its assumed parents Pinot, P, and Gouais blanc, G, at four loci. The estimated allele frequencies relate to the progeny alleles and so the frequency of 221 at locus VVMD28 is 0.057637 (or 0.06 as reported by Bowers *et al.* (1999).

	Genotype			
Locus	P	G	C	Frequency
VVMD28	221	231	221	0.057637
	239	249	231	0.115274
VVS2	137	133	137	0.040346
	151	143	143	0.17147
VVMD31	216	212	214	0.086455
	216	214	216	0.214697
VrZAG79	239	237	243	0.094697
	245	243	245	0.108696

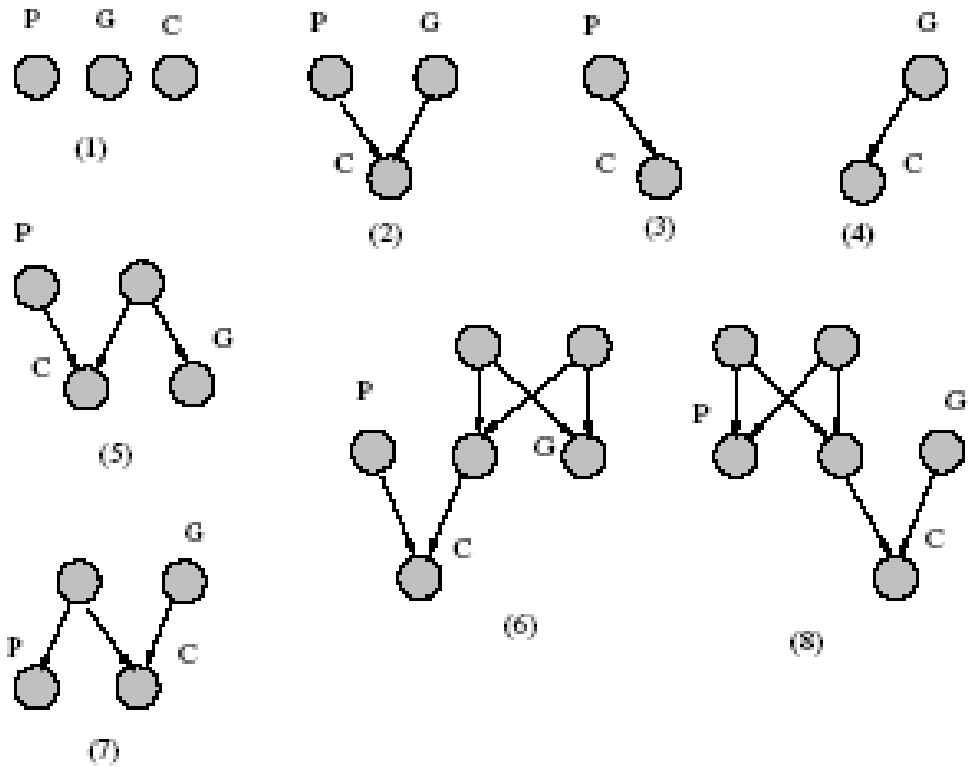


Figure 6 Eight alternative pedigrees for the relationship of Chardonnay with Pinot and Gouais blanc where we make no distinction between male and female plants.

Pedigree	Prior 1	Prior 2	Likelihood	Posterior 1	Posterior 2
1	0.125	0.085470085	1.00994E-17	0.000012	0.000101
2	0.125	0.008547009	5.62162E-13	0.660551	0.564966
3	0.125	0.854700855	1.45349E-15	0.001708	0.146074
4	0.125	0.008547009	2.44133E-16	0.000287	0.000245
5	0.125	0.008547009	8.09692E-14	0.095140	0.081373
6	0.125	0.008547009	8.09692E-14	0.095140	0.081373
7	0.125	0.008547009	6.26209E-14	0.073581	0.062933
8	0.125	0.008547009	6.26209E-14	0.073581	0.062933

Generalised paternity example

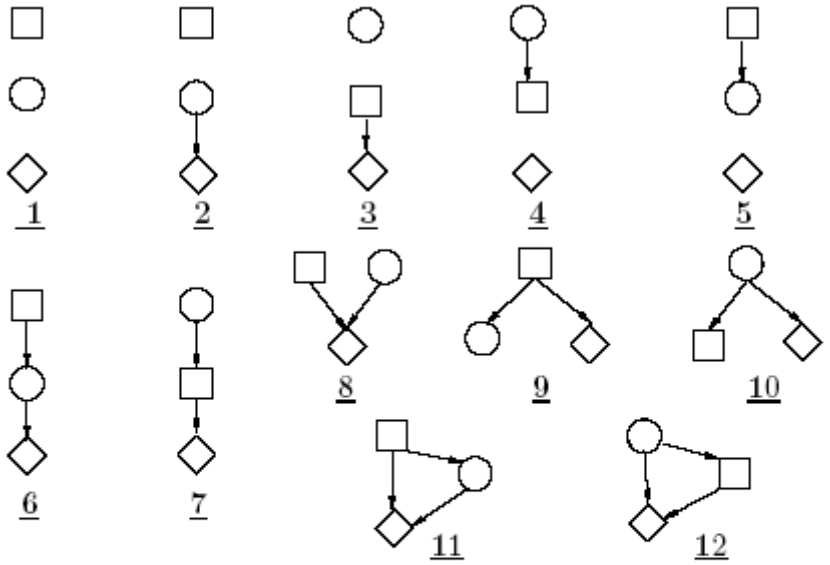


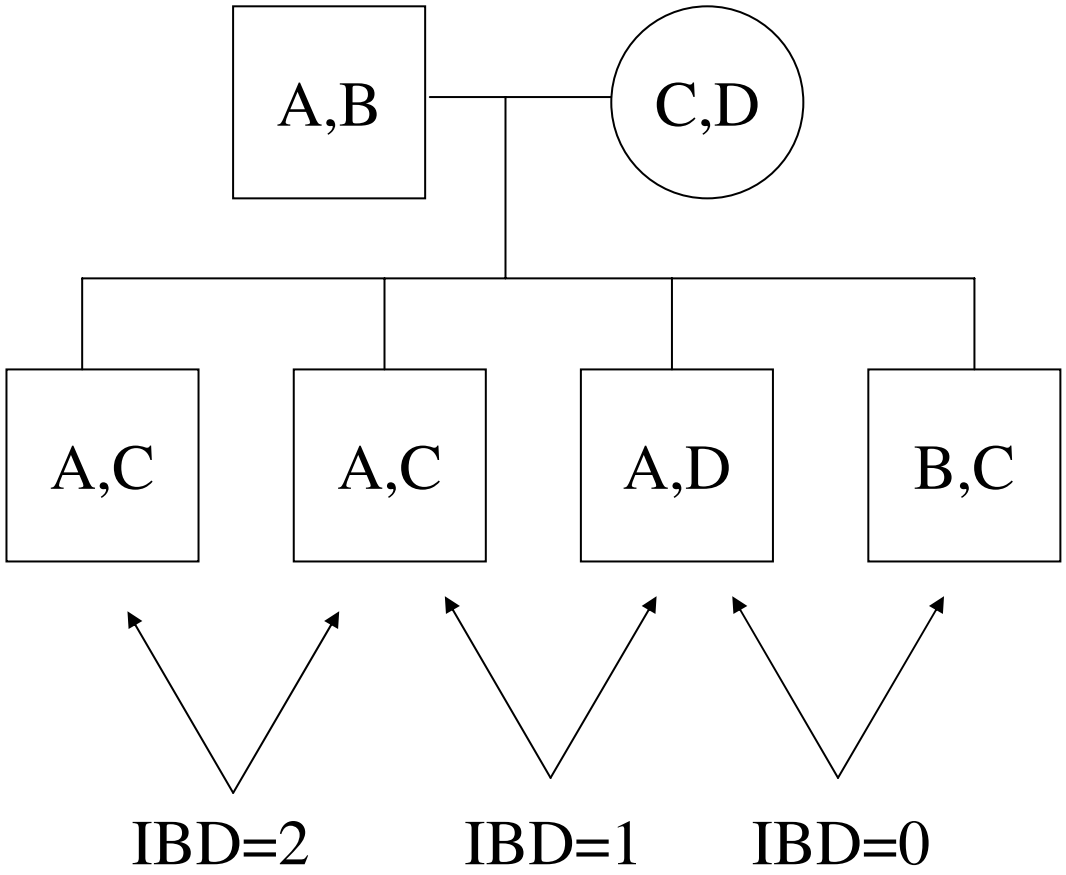
Figure 2 The twelve possible pedigree structures involving an adult male, an adult female and a juvenile. As is consistent with tradition, males are depicted by squares, females by circles and children by diamonds. A parent-offspring relationship is depicted by an arrow directed from the parent to the offspring individual.

How many markers and should they be linked or unlinked?

Likelihoods

UNIVERSITETET I OSLO

Identical By Descent (IBD)



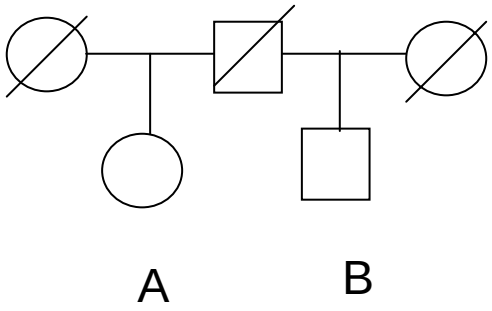
	Prob
IBD=2	0,25
IBD=1	0,50
IBD=0	0,25

Table 1. Probabilities for ordered autosomal genotypes as a function of there being 0, 1 or 2 IBD alleles. This table is used to perform exact calculation for a pairwise family relation by means of Equation 2. Genotypes with no common alleles are omitted.

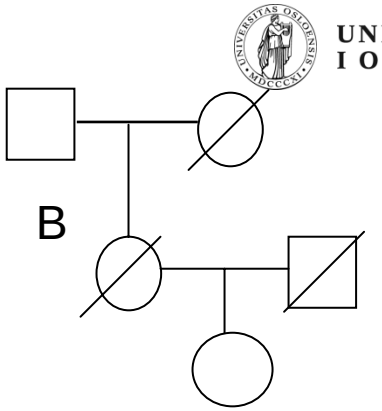
Genotypes for pair	No IBD alleles	One IBD allele	Two IBD alleles
$(a,a),(a,a)$	p_a^4	p_a^3	p_a^2
$(a,a),(a,b)$	$2p_a^3p_b$	$p_a^2p_b$	0
$(a,a),(b,b)$	$p_a^2p_b^2$	0	0
$(a,b),(a,b)$	$4p_a^2p_b^2$	p_ap_b	$2p_ap_b$

$$\begin{aligned} L(data \mid pedigree) &= L(data \mid IBD = 0)P(IBD = 0) \\ &+ L(data \mid IBD = 1)P(IBD = 1) \\ &+ L(data \mid IBD = 2)P(IBD = 2) \end{aligned}$$

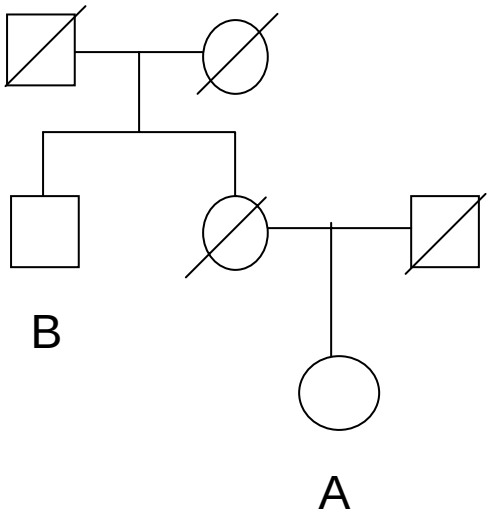
$$Example: L((a,a),(a,a) \mid sibs) = p_a^4\frac{1}{4} + p_a^3\frac{1}{2} + p_a^2\frac{1}{4}$$



half-sibs



grand parent-grand child



uncle-niece

Problem (Thompson, 1986):

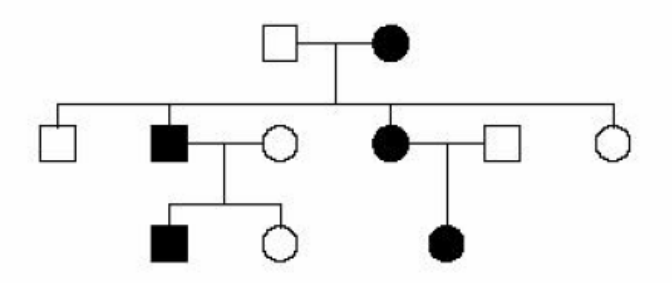
- A and B share no, one or two alleles with probabilities 0.5, 0.5, 0.
- Identical likelihoods.

Linked markers needed

r: recombination fraction
0: completely linked
0.5: unlinked

Classical linkage analysis:

L1: disease mutation
L2: genetic marker
Objective: determine L1-location
We, however, only need null-likelihood



Likelihoods for two linked markers

$$L(data \mid ped.\ \mathrm{i}) = a k_{11}^{i}(r) + b$$

a and b depend only on allele frequencies

Linked markers

$k_{11}^{i}(r)$

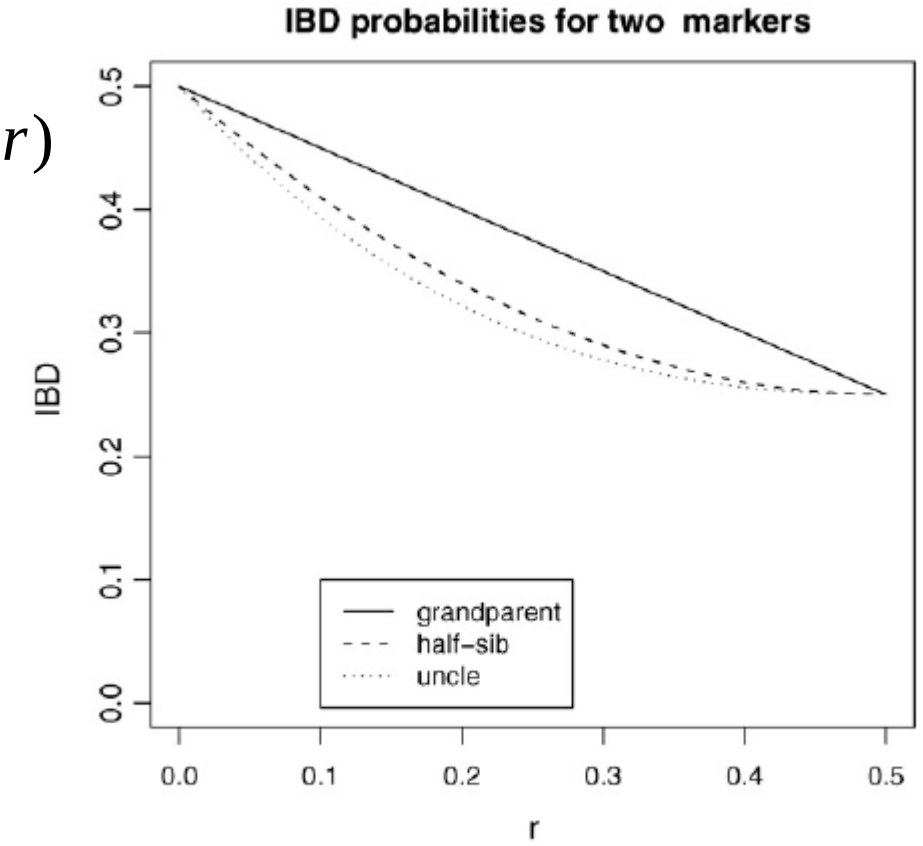


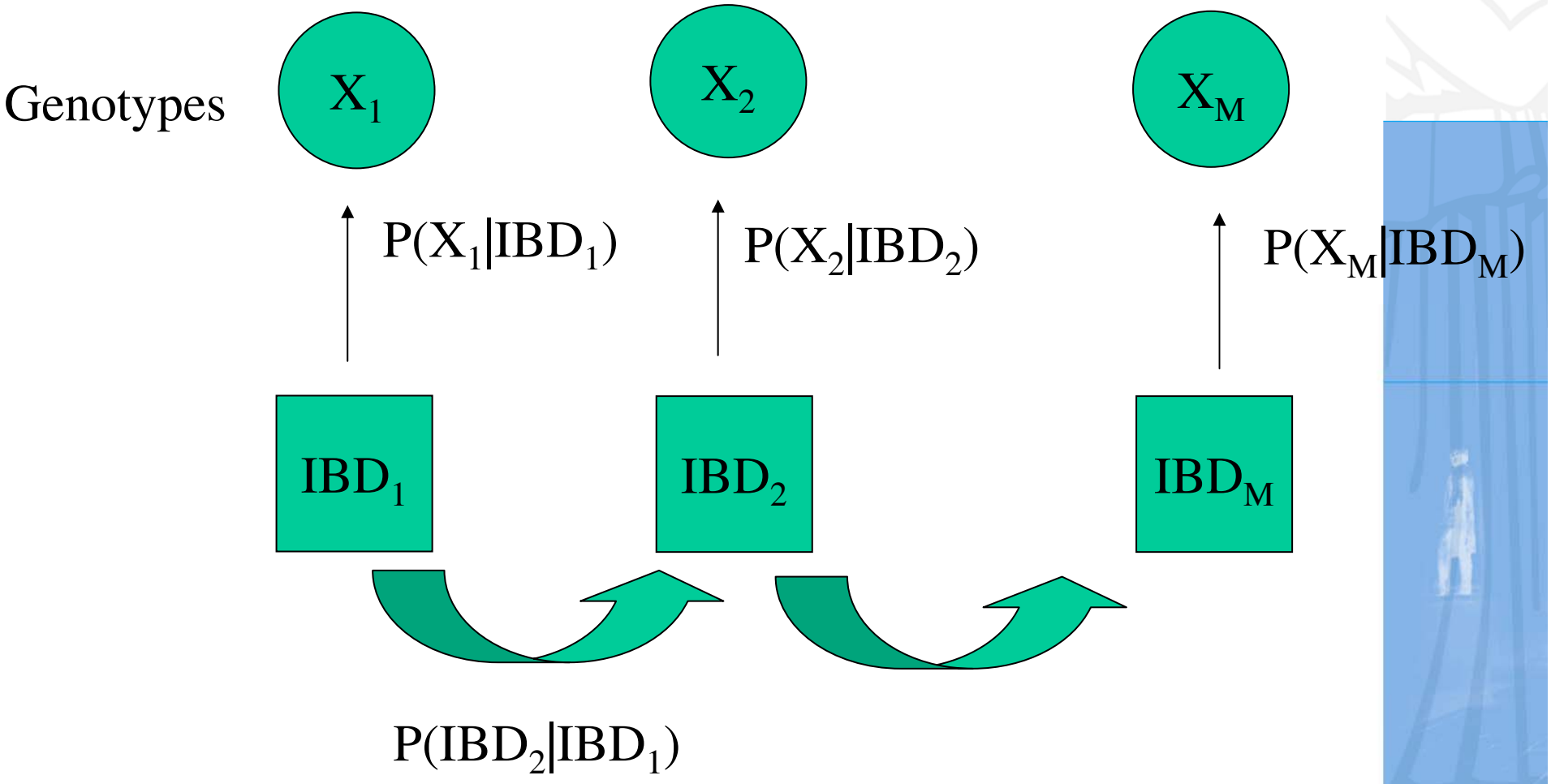
Fig. 2. The probability that two individuals are IBD at each of two loci is shown for the pedigrees of Fig. 1.

Extension: independent pairs

$$L(data \mid ped.\,\mathrm{i}) = \prod_{j=1}^{22} \mathrm{L}(\mathrm{data}_{\mathrm{chr\,j}} \mid ped.\,\mathrm{i})$$

UNIVERSITETET I OSLO

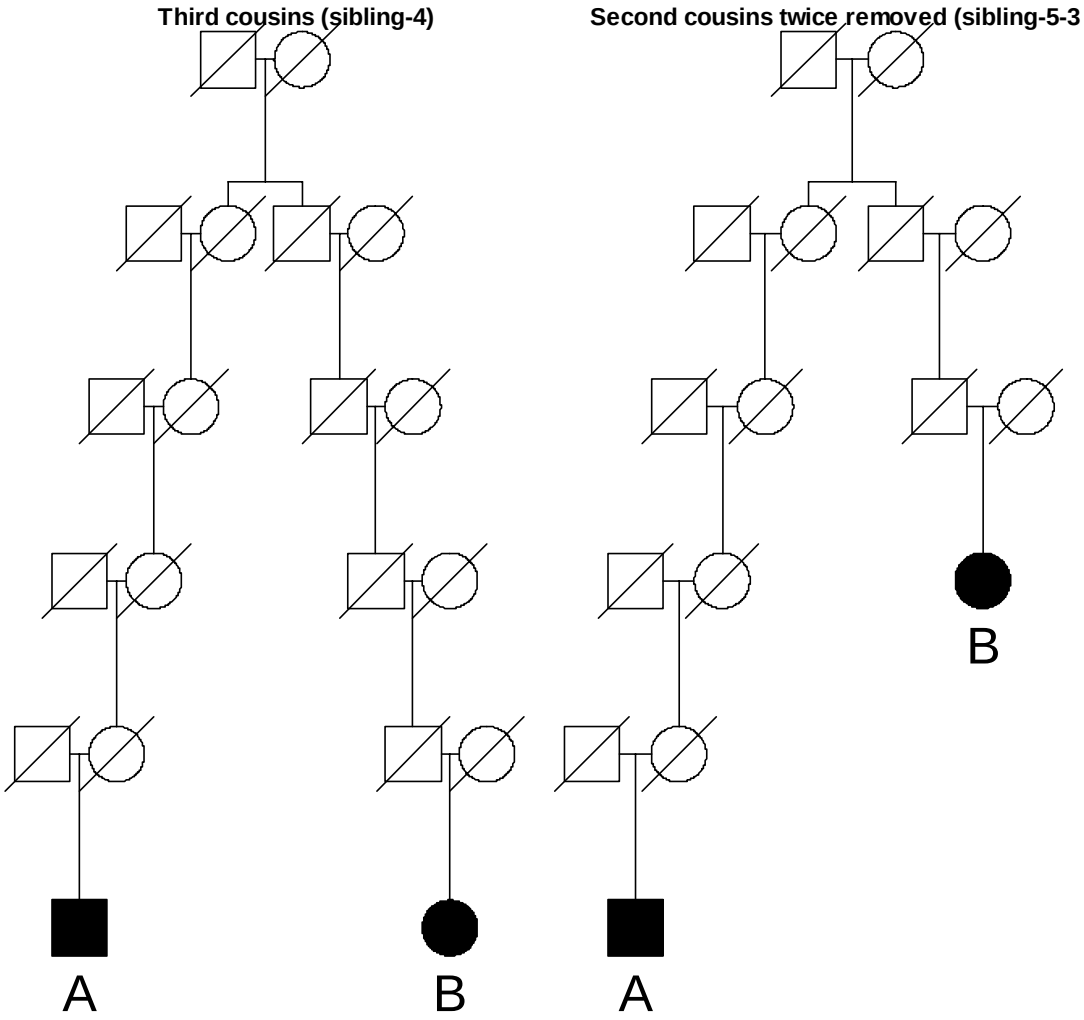
Lander-Green:Hidden markov model for IBD process along chromosome



UNIVERSITETET
I OSLO

Software for linked markers

- Software:
  - Merlin, Allegro, Genehunter,…
    - Lander-Green
  - FEST
  - Morgan. Complex pedigrees
    - MCMC
- Recall:
  - Only null-likelihood needed for relationship estimation



Likelihoods coincide also for linked autosomal markers

KP Donnelly (1983)

UNIVERSITETET I OSLO

Some references

- Abecasis, G *et al*. (2002) *Merlin-rapid analysis of dense genetic maps using sparse gene flow trees*, Nat. Genet. 30 97–101.
- Donnelly, KP (1983). *The probability that related individuals share some section of the genome identical by descent,* Theor. Populat. Biol. 23 34–63.
- Egeland, T *et al.* (2000*) Beyond traditional paternity and identification cases. Selecting the most probable pedigree*, Forensic Science International 110, 47–59.
- Epstein, MP *et al.* (2002) *Improved inference of relationship for pairs of individuals,* Am. J. Hum. Genet. 67 (2002) 1219–1231.
- Essen-Möller, E (1938) Mitteilungen der Anthropologische Gesellschaft (Wien) 68, 9–53.
- Sheehan, N and Egeland T (2007) *Structured incorporation of prior information in relationship identification problems,* Ann. Hum. Genet. 71 501–518
- *Skare, Ø et al.* (2009). *Submitted*
- Thompson, EA (1986) *Pedigree Analysis in Human Genetics. The Johns Hopkins University Press, Baltimore.*