Segregation, IBD and Linkage from the Perspective of Bayesian Networks

G. Nuel

Institute of Mathematics (INSMI), CNRS,
Laboratory of Applied Mathematics (MAP5),
University Paris Descartes,
France

JFRB'2012, May 11-13,
Kerkennah, Tunisia

Outline

Outline

Definition

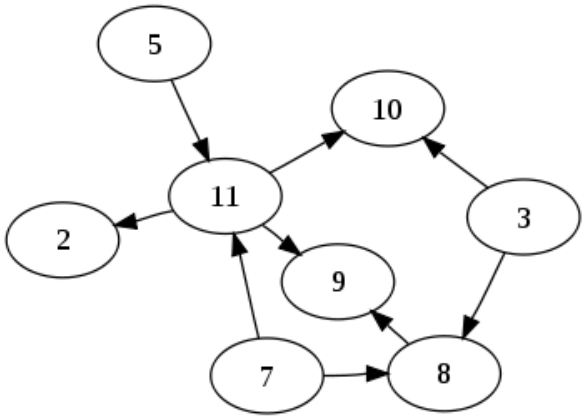
Bayesian Network (BN):

- $X_{\mathcal{U}} = (X_u)_{u \in \mathcal{U}}$ set of p random variables in $\mathbb{R}^{d_u}$;
- $\mathcal{F} \subset \mathcal{U} \times \mathcal{U}$ such that $(\mathcal{U}, \mathcal{F})$ defines a DAG over $\mathcal{U}$;
- $\forall v \in \mathcal{U}$, the parent set $\text{pa}(X_v) = \{X_u \in \mathcal{U}, (u, v) \in \mathcal{F}\}$;

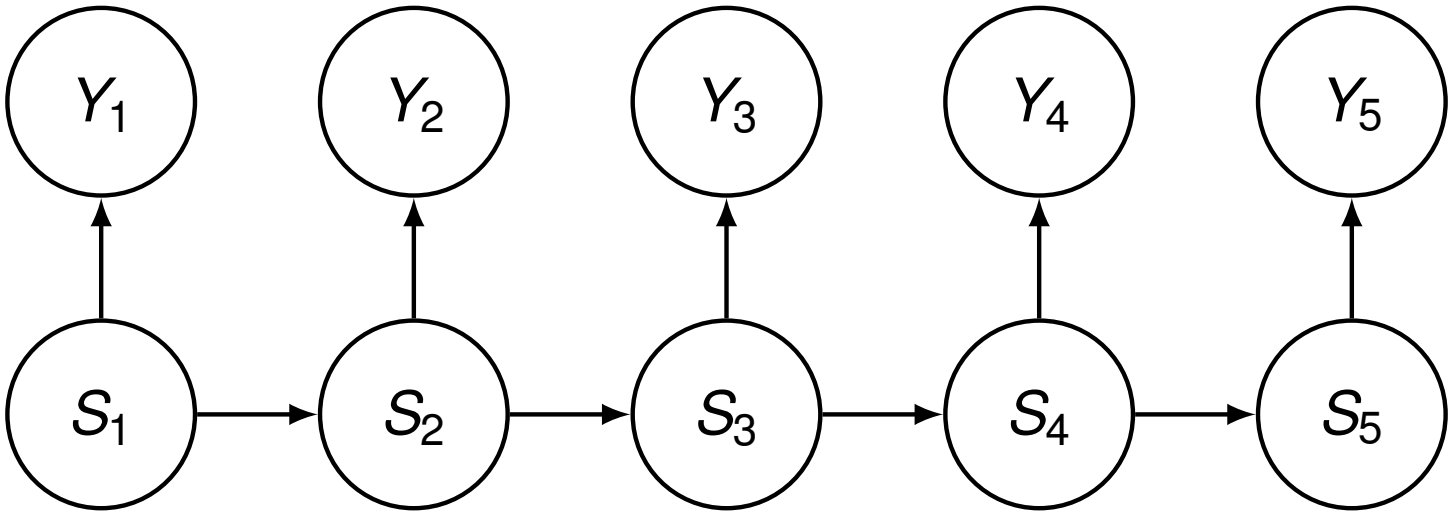
$$\mathbb{P}(X_{\mathcal{U}}) = \prod_{u \in \mathcal{U}} \mathbb{P}(X_u | \text{pa}(X_u))$$

Examples:

- Markov chains
- Hidden Markov Models
- Genotypes in a pedigree



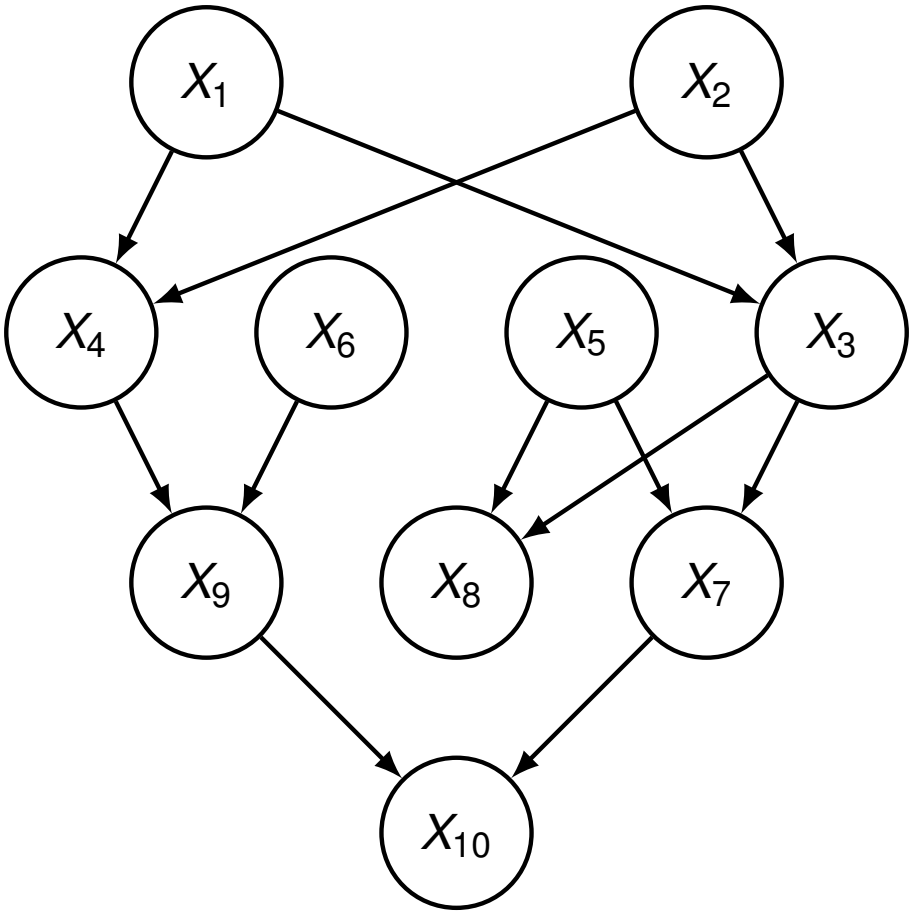
BN for a classical HMM



$Y_{1:n}$ are the observed states and $S_{1:n}$ the hidden states

Example: $Y_i =$ mm of precipitation, $S_i =$ atmospheric pressure

BN for a pedigree (family structure)



with a consanguinity loop between cousins 7 and 9

Notion of Evidence

Evidence:
For any $u \in \mathcal{U}$, $\mathcal{X}_u \subset \mathbb{R}^{d_u}$ be the set of possible outcomes for X_u. For any $\mathcal{V} \subset \mathcal{U}$, we define the *evidence function* by:

$$e(X_\mathcal{V}) = \begin{cases} 1 & \text{if } X_\mathcal{V} \in \mathcal{X}_\mathcal{V} \\ 0 & \text{else} \end{cases}$$

and the *evidence* by the event: $\mathcal{E} = \{e(X_\mathcal{U}) = 1\} = \{X_\mathcal{U} \in \mathcal{X}_\mathcal{U}\}$.

Examples:

- the set of observed values in HMM;
- the (partially) observed genotypes in a pedigree

Main Objective: study the conditional distribution

$$\mathbb{P}(X_\mathcal{U}|\mathcal{E}) = \frac{\mathbb{P}(X_\mathcal{U}, \mathcal{E})}{\mathbb{P}(\mathcal{E})}$$

Outline

Junction Tree

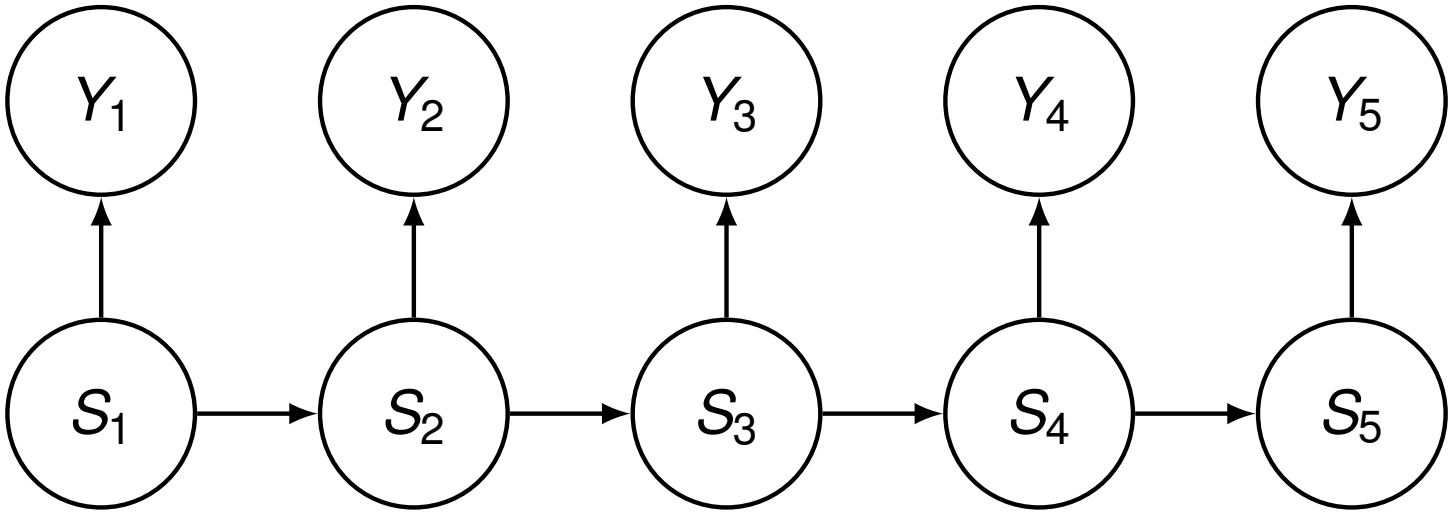
We consider $C_{\mathcal{I}} = (C_i)_{i \in \mathcal{I}}$, $\mathcal{I} = \{1, \ldots, q\}$ a set of q *clusters* such as $C_i \subset X_{\mathcal{U}}$ for all $i \in \mathcal{I}$ and we assume the following three conditions:

JT1) **Tree.** We have a tree structure on $C_{\mathcal{I}}$: for any $i, j \in \mathcal{I}$ it exists a unique connecting path, denoted path(i, j), between C_i and C_j.

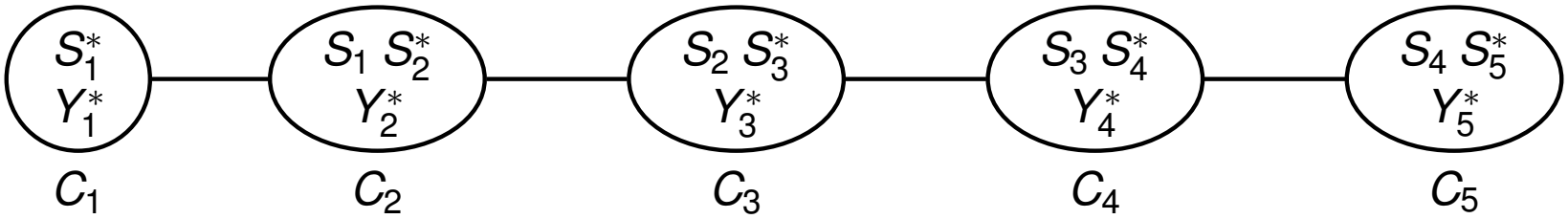
JT2) **Running intersection.** For any $i, j \in \mathcal{I}$, $C_i \cap C_j \subset C_k$ for all $k \in$ path(i, j).

JT3) **Covering.** For any $u \in \mathcal{U}$, it exists at least one $i \in \mathcal{I}$ such as the *family set* $\text{fa}(X_u) = \text{pa}(X_u) \cup \{X_u\} \subset C_i$ (notation: X_u is then marked with a * in C_i)

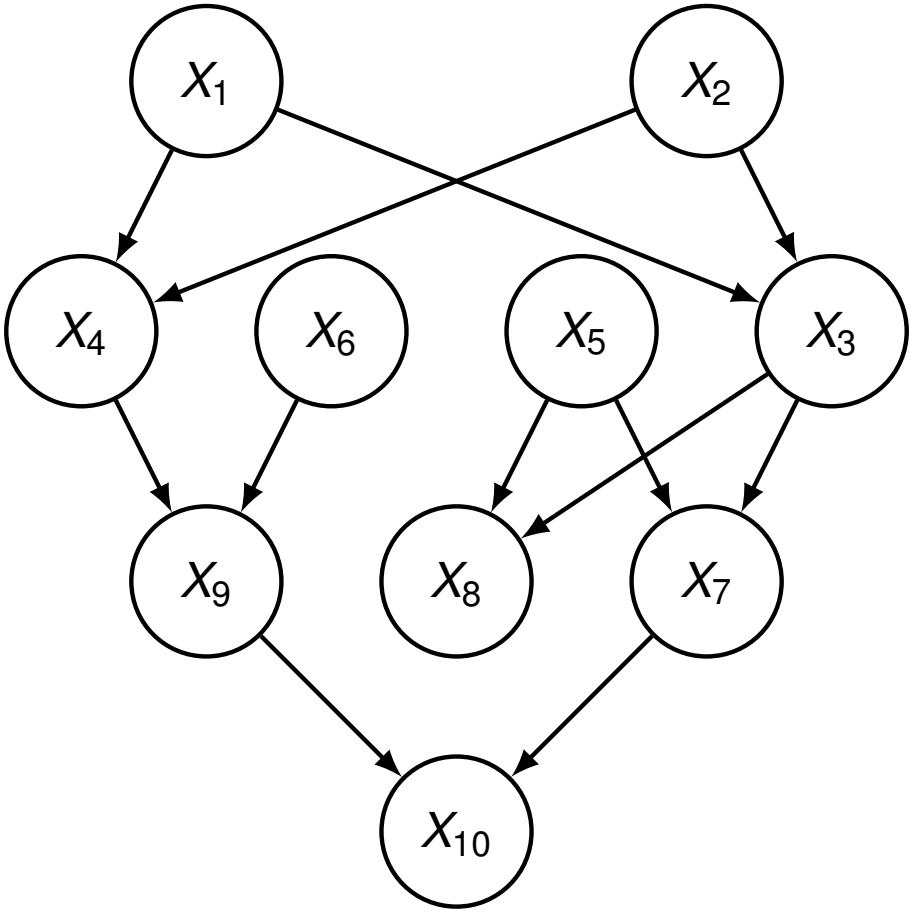
JT for HMM



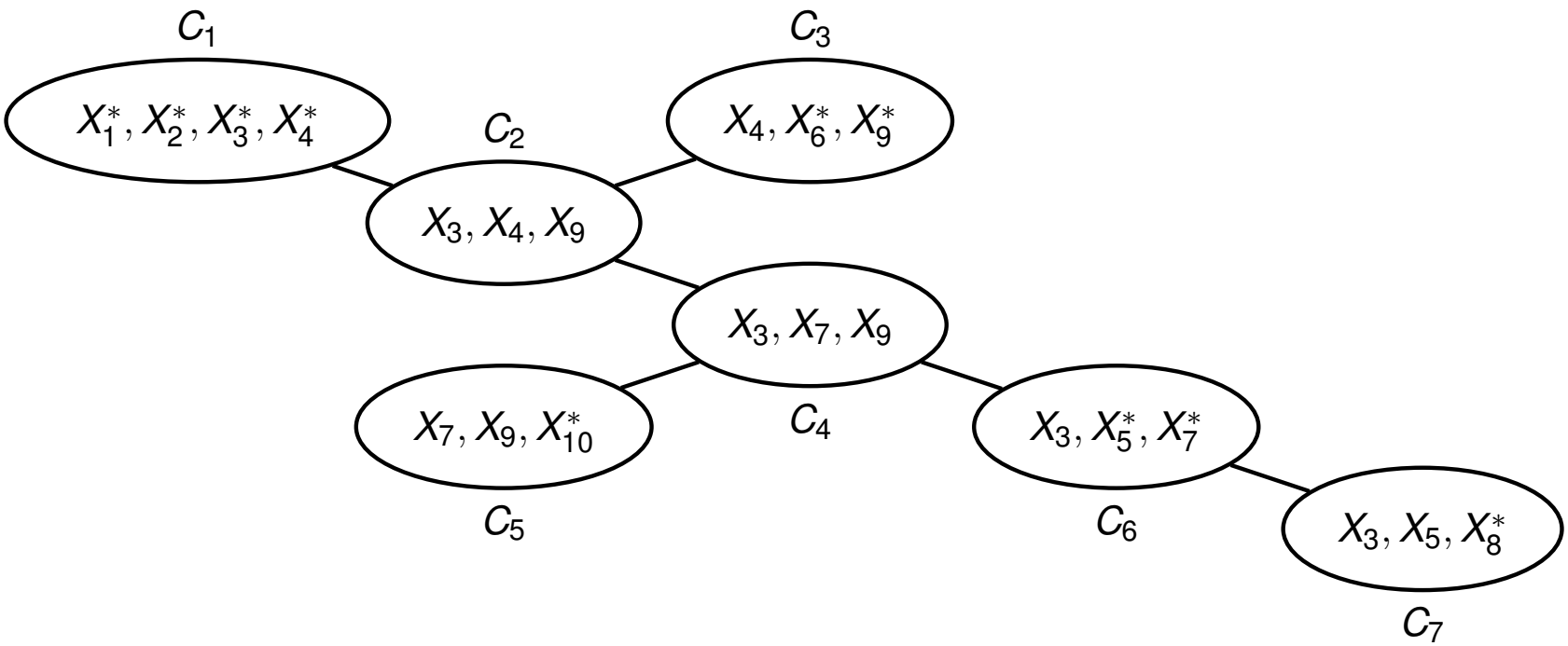
JT for HMM



JT for pedigree



JT for pedigree



Messages

Recusive Definition: For any edge $j - k$ of the JT

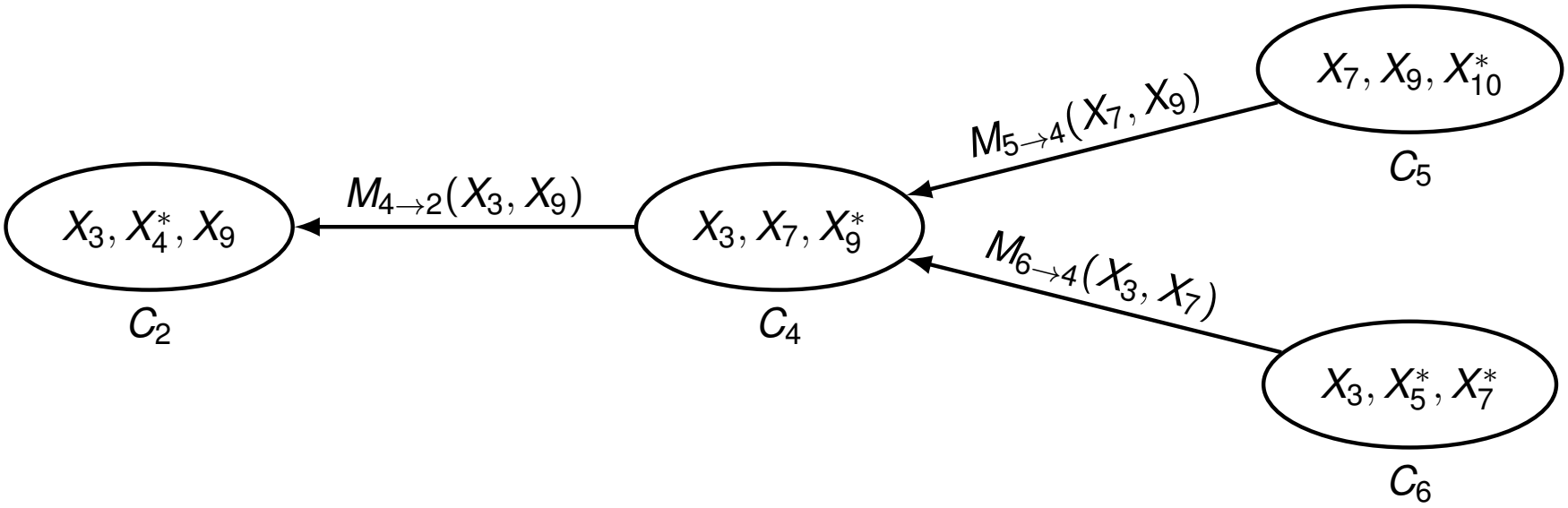
$$M_{j\to k}\left(S_{j,k}\right) = \sum_{C_j \setminus S_{j,k}} \Phi_j\left(C_j\right) \prod_{i\in \mathrm{n}(j), i\neq k} M_{i\to j}\left(S_{i,j}\right)$$

with $S_{j,k} = C_j \cap C_k$ and $\Phi_j\left(C_j\right) = \prod_{X_u \in C_j^*} \mathbb{P}(X_u|\mathrm{pa}(X_u))$.

Algorithms: for any root $r \in \mathcal{I}$ of the JT

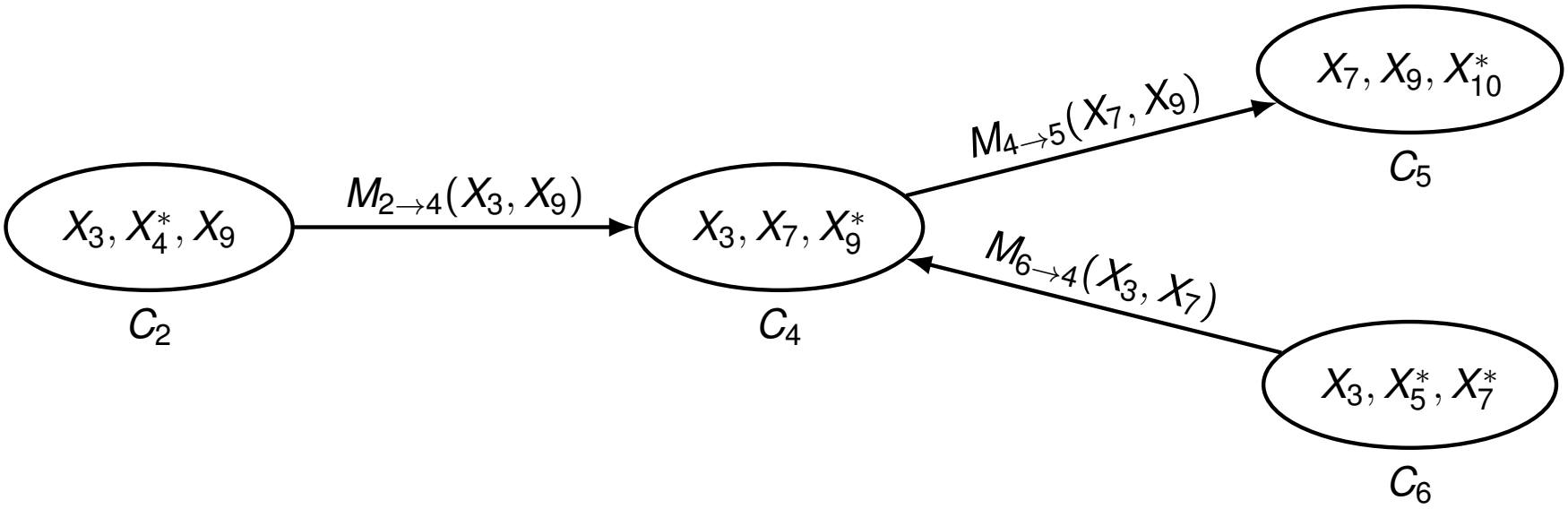
- inward: compute recursively all msg. from leaves to root;
- outward: compute recursively all msg. from root to leaves.

Example of Recursion Steps



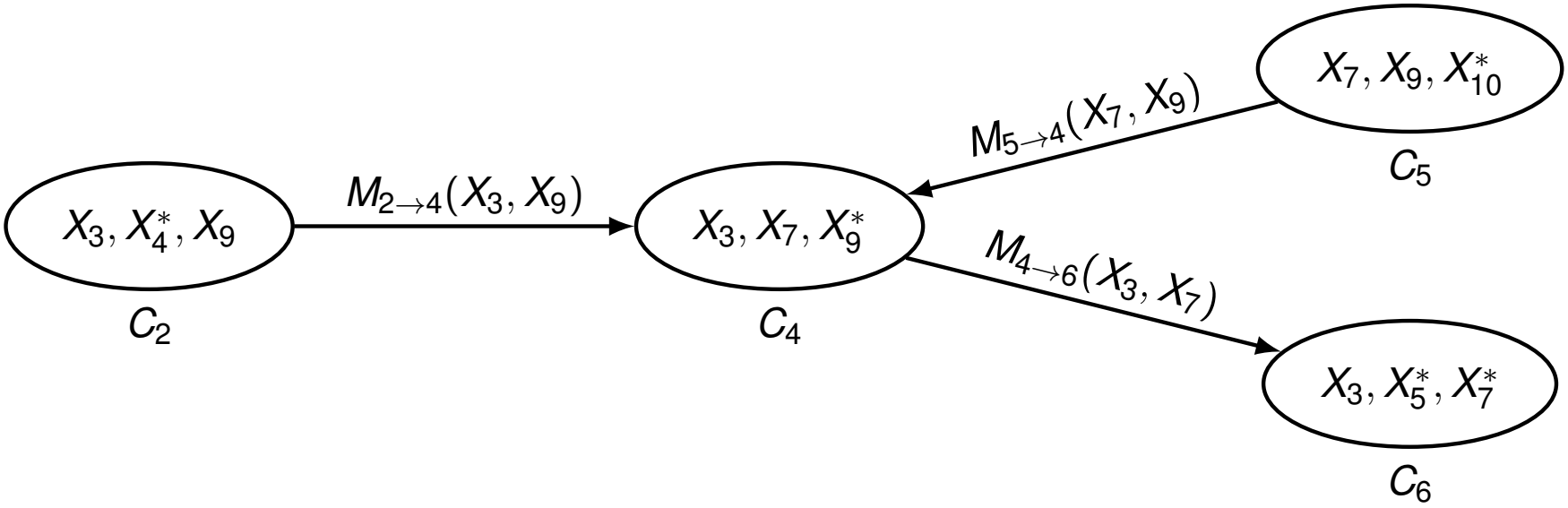
$$M_{4\to2}(X_3, X_9) = \sum_{X_7} \mathbb{P}(X_9|X_3, X_7) M_{5\to4}(X_7, X_9) M_{6\to4}(X_3, X_7)$$

Example of Recursion Steps



$$M_{4\to5}(X_7, X_9) = \sum_{X_3} \mathbb{P}(X_9|X_3, X_7) M_{2\to4}(X_3, X_9) M_{6\to4}(X_3, X_7)$$

Example of Recursion Steps



$$M_{4\to6}(X_3, X_7) = \sum_{X_9} \mathbb{P}(X_9|X_3, X_7) M_{2\to4}(X_3, X_9) M_{5\to4}(X_7, X_9)$$

Outline

Quantities of Interest

Using Exact BP (and variants), we can:

- compute $\mathbb{P}(\mathcal{E})$

  $\Rightarrow$ **likelihood** computations;
- $\forall u \in \mathcal{U}$, compute $\mathbb{P}(X_u, \text{pa}(X_u)|\mathcal{E})$

  $\Rightarrow$ **marginal distributions**, EM algorithms;
- sample $X_{\mathcal{U}}$ from $\mathbb{P}(X_{\mathcal{U}}|\mathcal{E})$

  $\Rightarrow$ empirical studies, **SEM algorithms**;
- compute $X^*_{\mathcal{U}} = \arg\max \mathbb{P}(X_{\mathcal{U}}|\mathcal{E})$

  $\Rightarrow$ **maximum a posteriori** configuration;

Complexity

Tree-width:

- the complexity of BP depend on the JT topology;
- finding the “best” JT topology is a NP-hard problem;
- heuristic algorithms are available.

Examples:

- HMM: linear complexity;
- Pedigree with no loops: linear complexity;
- Pedigree with loops: exponential with number of loops.

If complexity is too high:

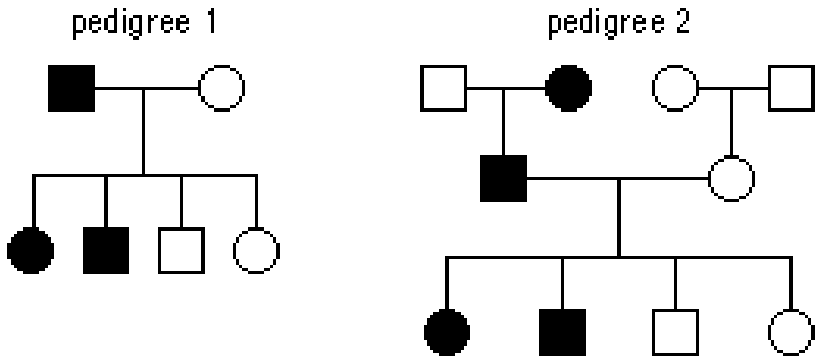
- Approximate Belief Propagation (ex: Loopy BP);
- MCMC methods (ex: Blocking Gibbs-sampling)

Outline

What is segregation analysis ?

Data:

- ascertained families
- only disease status



Objective:

- suggest a disease model that explains the observation
- estimate its parameters

Monogenic disease:

- q: disease allele frequency, $\mathbb{P}(\mathtt{D}) = q$, $\mathbb{P}(\mathtt{d}) = 1 - q$
- f_0, f_1, f_2: the penetrances, $\mathbb{P}(\mathrm{Y}i = \text{affected} | \mathrm{G}i = \mathtt{DD}) = f_2$

Variables

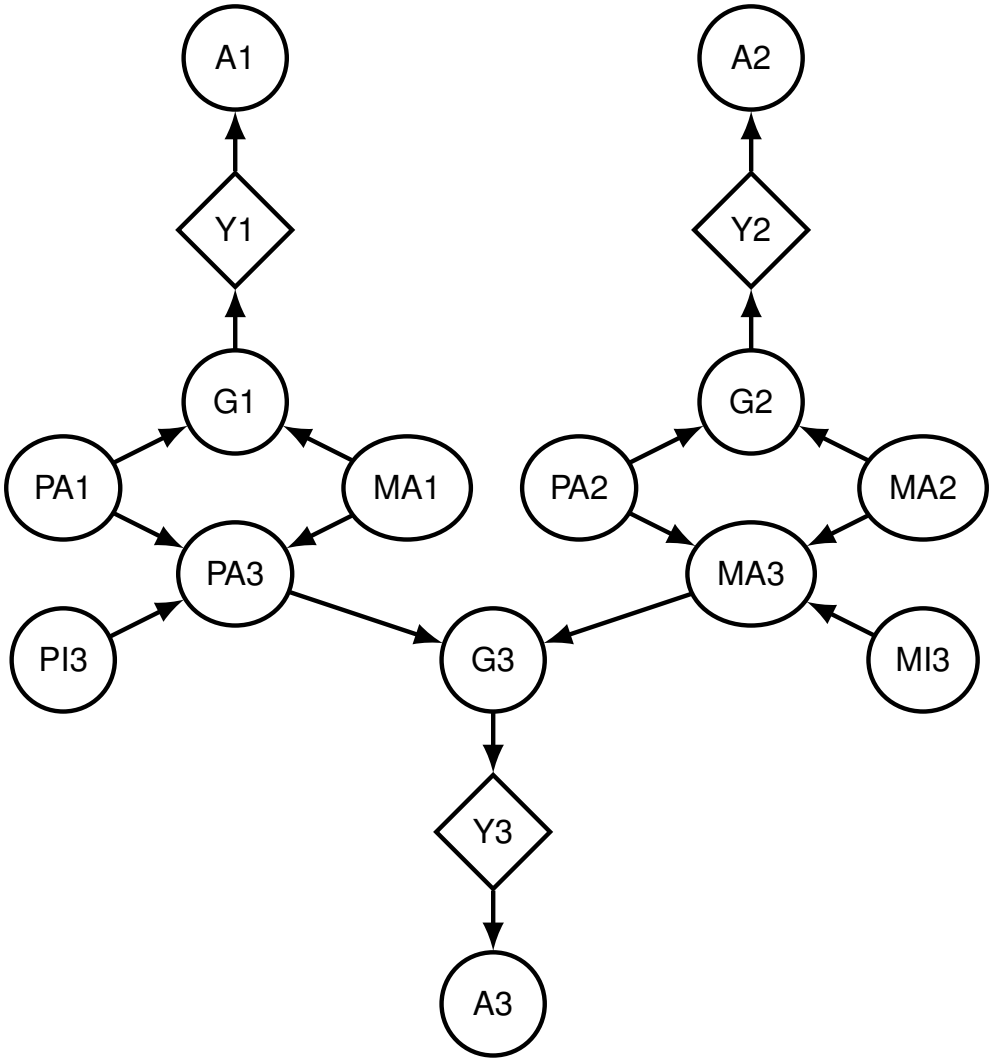
Disease genotypes:

- paternal and maternal alleles of individual *i* (non observed): $\mathrm{PA}i, \mathrm{MA}i \in \{\mathtt{d}, \mathtt{D}\}$;
- genotype of individual *i* (non observed): $\mathrm{G}i \in \{\mathtt{dd}, \mathtt{dD}, \mathtt{DD}\}$;
- paternal and maternal inheritance of non-founder *i* (non observed): $\mathrm{PI}i, \mathrm{MI}i \in \{\mathrm{paternal}, \mathrm{maternal}\}$

Disease status:

- disease status of individual *i* (partially observed): $\mathrm{Y}i \in \{\mathrm{unaffected}, \mathrm{affected}\}$;
- ascertainment status of individual *i* (non observed): $\mathrm{A}i \in \{0, 1\}$;

BN for Segregation Analysis



Conditional Distributions (deterministic)

- Gi:

$$\begin{aligned}
\mathbb{P}(\mathrm{G}i = \mathtt{dd} | \mathrm{PA}i = \mathtt{d}, \mathrm{MA}i = \mathtt{d}) &= 1 \\
\mathbb{P}(\mathrm{G}i = \mathtt{dD} | \mathrm{PA}i = \mathtt{d}, \mathrm{MA}i = \mathtt{D}) &= 1 \\
\mathbb{P}(\mathrm{G}i = \mathtt{DD} | \mathrm{PA}i = \mathtt{D}, \mathrm{MA}i = \mathtt{D}) &= 1 \\
\mathbb{P}(\mathrm{G}i = \mathtt{dD} | \mathrm{PA}i = \mathtt{D}, \mathrm{MA}i = \mathtt{d}) &= 1
\end{aligned}$$

- PAk, MAk for non-founder k with father i and mother j:

$$\begin{aligned}
\mathbb{P}(\mathrm{PA}k = \mathrm{PA}i | \mathrm{PA}i, \mathrm{MA}i, \mathrm{PI}k = \mathrm{paternal}) &= 1 \\
\mathbb{P}(\mathrm{PA}k = \mathrm{MA}i | \mathrm{PA}i, \mathrm{MA}i, \mathrm{PI}k = \mathrm{maternal}) &= 1 \\
\mathbb{P}(\mathrm{MA}k = \mathrm{PA}j | \mathrm{PA}j, \mathrm{MA}j, \mathrm{MI}k = \mathrm{paternal}) &= 1 \\
\mathbb{P}(\mathrm{MA}k = \mathrm{MA}j | \mathrm{PA}j, \mathrm{MA}j, \mathrm{MI}k = \mathrm{maternal}) &= 1
\end{aligned}$$

Conditional Distributions (random)

- PAi, MAi for founder i:

$$\mathbb{P}(\text{PA}i = \texttt{D}) = \mathbb{P}(\text{MA}i = \texttt{D}) = q \quad \mathbb{P}(\text{PA}i = \texttt{d}) = \mathbb{P}(\text{MA}i = \texttt{d}) = 1-q$$

- PIk, MIk for non-founder k:

$$\mathbb{P}(\text{PI}k = \text{maternal}) = \mathbb{P}(\text{PI}k = \text{paternal}) = 1/2$$

$$\mathbb{P}(\text{MI}k = \text{maternal}) = \mathbb{P}(\text{MI}k = \text{paternal}) = 1/2$$

- Yi:

$$\mathbb{P}(\text{Y}i = \text{affected}|\text{G}i = \texttt{dd}) = f_0 \quad \mathbb{P}(\text{Y}i = \text{affected}|\text{G}i = \texttt{dD}) = f_1$$

$$\mathbb{P}(\text{Y}i = \text{affected}|\text{G}i = \texttt{DD}) = f_2$$

- Ai:

$$\mathbb{P}(\text{A}i = 1|\text{Y}_i = \text{affected}) = \pi \quad \mathbb{P}(\text{A}i = 1|\text{Y}_i = \text{unaffected}) = 0$$

Likelihood ignoring the ascertainment

Likelihood: $\theta = (f_0, f_1, f_2, q)$

$$L(\theta) = \mathbb{P}(Y_{\text{obs}}; \theta) = \mathbb{P}(\mathcal{E}; \theta)$$

with $\mathcal{E} = \{Y_{\text{obs}}\}$.

Maximum Likelihood Estimator:

$$\widehat{\theta} = \arg\max_{\theta} L(\theta)$$

Problem:

- we observed only ascertained family
- ascertained family have more affected members

$\Rightarrow$ correct the bias by conditioning on ascertainment

Likelihood conditionally on the ascertainment

Likelihood:
$L(\theta) = \mathbb{P}(Y_{\text{obs}}|\text{A} = 1; \theta)$ with $\theta = (f_0, f_1, f_2, q, \pi)$ and where the ascertainment event is defined by $\{\text{A} = 1\} = \{\exists i, \text{A}i = 1\}$

$$L(\theta) = \frac{\mathbb{P}(Y_{\text{obs}}; \theta)\mathbb{P}(\text{A} = 1|Y_{\text{obs}}; \theta)}{\mathbb{P}(\text{A} = 1; \theta)} = \frac{\mathbb{P}(Y_{\text{obs}}; \theta)[1 - \mathbb{P}(\text{A} = 0|Y_{\text{obs}}; \theta)]}{1 - \mathbb{P}(\text{A} = 0; \theta)}.$$

Need to compute:

- $\mathbb{P}(Y_{\text{obs}}; \theta)$ which is $\mathbb{P}(\mathcal{E})$ with $\mathcal{E} = \{Y_{\text{obs}}$ is observed$\}$;
- $\mathbb{P}(\text{A} = 0|Y_{\text{obs}}; \theta)$ which is $\mathbb{P}(\mathcal{E})/\mathbb{P}(Y_{\text{obs}}; \theta)$ with $\mathcal{E} = \{Y_{\text{obs}}$ is observed and A$i = 0$ for all $i\}$;
- $\mathbb{P}(\text{A} = 0; \theta)$ which is $\mathbb{P}(\mathcal{E})$ with $\mathcal{E} = \{\text{A}i = 0$ for all $i\}$;

Maximizing the conditional likelihood

EM algorithm:

$$Q(\theta'|\theta) = \sum_{X_{\mathcal{U}}} \mathbb{P}(X_{\mathcal{U}}|\, Y_{\text{obs}}, \mathsf{A} = 1; \theta) \log \mathbb{P}(X_{\mathcal{U}},\, Y_{\text{obs}}|\mathsf{A} = 1; \theta')$$

Marginal distributions: for all $u \in \mathcal{U}$ we have

$$\mathbb{P}(\text{fa}(X_u)|\, Y_{\text{obs}}, \mathsf{A} = 1; \theta) = \frac{1 - \mathbb{P}(X_{\text{fa}(u)}, \mathsf{A} = 0|\, Y_{\text{obs}}; \theta)}{1 - \mathbb{P}(\mathsf{A} = 0|\, Y_{\text{obs}}; \theta)}$$

which is obtained through $\mathbb{P}(\mathcal{E})$ and $\mathbb{P}(\text{fa}(X_u), \mathcal{E})$ with $\mathcal{E} = \{\, Y_{\text{obs}}$ is observed and $\mathsf{A}i = 0$ for all $i\}$

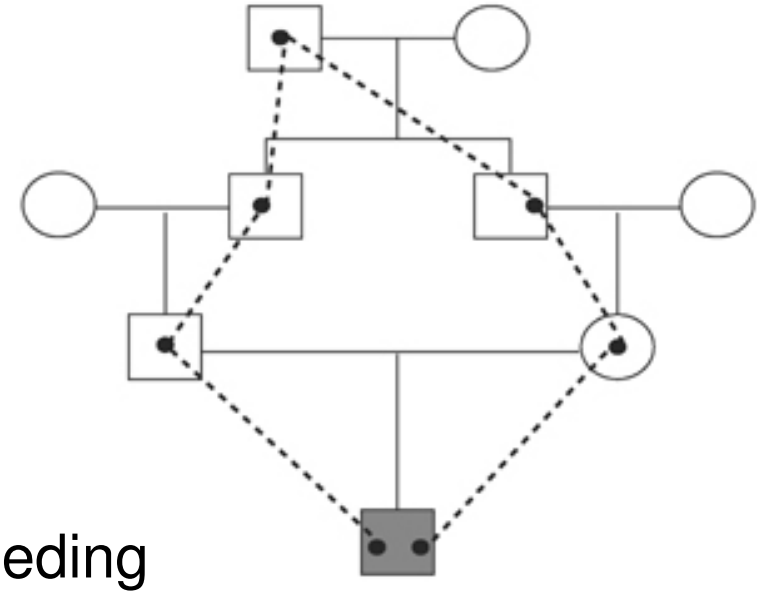
Outline

What is the Identity-By-Descent ?

IBD: two alleles are IBD if they are copies of an ancestral allele.

IBS: two alleles are IBS if they have the same value.

Rq: IBD $\Rightarrow$ IBS, but IBS $\not\Rightarrow$ IBD



Interest of IBD:

- tells about of population inbreeding
- tells about the history of an allele
- because of recombinations, IBD occurs by block

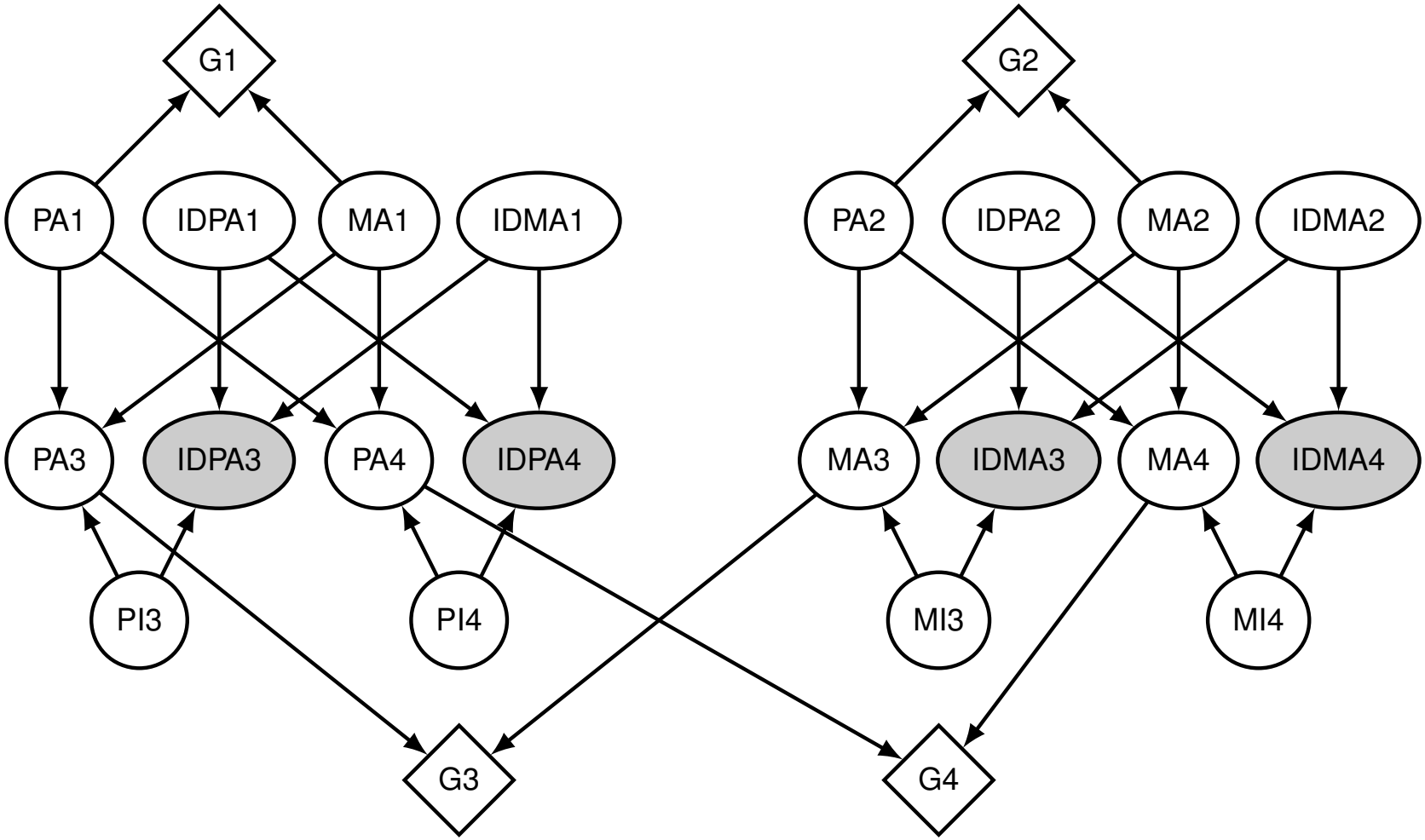
$\Rightarrow$ might be useful for linkage analysis

Variables

IBD with three alleles:

- paternal and maternal alleles of individual i (non observed): $\mathrm{PA}i, \mathrm{MA}i \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$;
- paternal and maternal allele identifier of individual i (non observed): $\mathrm{IDPA}i, \mathrm{IDMA}i \in \{1, 2, \ldots, 2k\}$ where k is the number of founders;
- genotype of individual i (partially observed): $\mathrm{G}i \in \{\mathtt{aa}, \mathtt{ab}, \mathtt{ac}, \mathtt{bb}, \mathtt{bc}, \mathtt{cc}\}$;
- paternal and maternal inheritance of non-founder i (non observed): $\mathrm{PI}i, \mathrm{MI}i \in \{\mathrm{paternal}, \mathrm{maternal}\}$;

BN for IBD


G1
G2
PA1
IDPA1
MA1
IDMA1
PA2
IDPA2
MA2
IDMA2
PA3
IDPA3
PA4
IDPA4
MA3
IDMA3
MA4
IDMA4
PI3
PI4
MI3
MI4
G3
G4

Conditional distributions (deterministic)

- IDPAi, IDMAi for founder i: unique integer identifier for each allele. In our nuclear family example we get: IDPA1 $= 1$, IDMA1 $= 2$, IDPA2 $= 3$, IDMA2 $= 3$;
- Gi:

$$\begin{aligned}\mathbb{P}(\mathrm{G}i = \mathtt{aa} | \mathrm{PA}i = \mathtt{a}, \mathrm{MA}i = \mathtt{a}) &= 1 \\ \mathbb{P}(\mathrm{G}i = \mathtt{ab} | \mathrm{PA}i = \mathtt{a}, \mathrm{MA}i = \mathtt{b}) &= 1\end{aligned}$$

- PAk, MAk for non-founder k with father i and mother j:

$$\begin{aligned}\mathbb{P}(\mathrm{PA}k = \mathrm{PA}i | \mathrm{PA}i, \mathrm{MA}i, \mathrm{PI}k = \mathrm{paternal}) &= 1 \\ \mathbb{P}(\mathrm{PA}k = \mathrm{MA}i | \mathrm{PA}i, \mathrm{MA}i, \mathrm{PI}k = \mathrm{maternal}) &= 1\end{aligned}$$

- IDPAk, IDMAk for non-founder k with father i and mother j:

$$\mathbb{P}(\mathrm{IDPA}k = \mathrm{IDPA}i | \mathrm{IDPA}i, \mathrm{IDMA}i, \mathrm{PI}k = \mathrm{paternal}) = 1$$

Conditional distributions (random)

- PAi, MAi for founder i (random):

$$\mathbb{P}(\text{PA}i = \mathtt{a}) = \mathbb{P}(\text{MA}i = \mathtt{a}) = q_{\mathtt{a}} \quad \mathbb{P}(\text{PA}i = \mathtt{b}) = \mathbb{P}(\text{MA}i = \mathtt{b}) = q_{\mathtt{b}}$$

$$\mathbb{P}(\text{PA}i = \mathtt{c}) = \mathbb{P}(\text{MA}i = \mathtt{c}) = q_{\mathtt{c}} \quad \text{with } q_{\mathtt{a}} + q_{\mathtt{b}} + q_{\mathtt{c}} = 1$$

- PIk, MIk for non-founder k (random):

$$\mathbb{P}(\text{PI}k = \text{maternal}) = \mathbb{P}(\text{PI}k = \text{paternal}) = 1/2$$

$$\mathbb{P}(\text{MI}k = \text{maternal}) = \mathbb{P}(\text{MI}k = \text{paternal}) = 1/2$$

Non-Parametric Linkage (Whittemore & Halpern'94)

Evidence:

- $X = \{\text{G}i, i \text{ affected}\}$;
- $Y = \{\text{G}j, j \in \mathcal{J}\}$ with $\mathcal{J} \subset \mathcal{I}$ (ex: the unaffected founders);
- $\mathcal{E} = \{X, Y\}$ or $\mathcal{E} = \{Y\}$

Testing statistic: with $\Phi = \{\text{IDMA}i, \text{IDPA}i, i \text{ affected}\}$

$$T(X, Y) = \sum_{\Phi} \mathbb{P}(\Phi|X, Y)S(\Phi)$$

Distribution under null hypothesis $H0$:

- $T(X, Y)$ with $X \sim \mathbb{P}(X|Y)$;
- $\mathbb{E}_{H0}[T(X, Y)] = \sum_{\Phi} \mathbb{P}(\Phi|Y)S(\Phi)$

The statistic $S(\Phi)$

$S_{\mathbf{pairs}}$:

$$S_{\text{pairs}}(\Phi) = \sum_{i<j \text{ aff.}} \text{F}ij(\Phi)$$

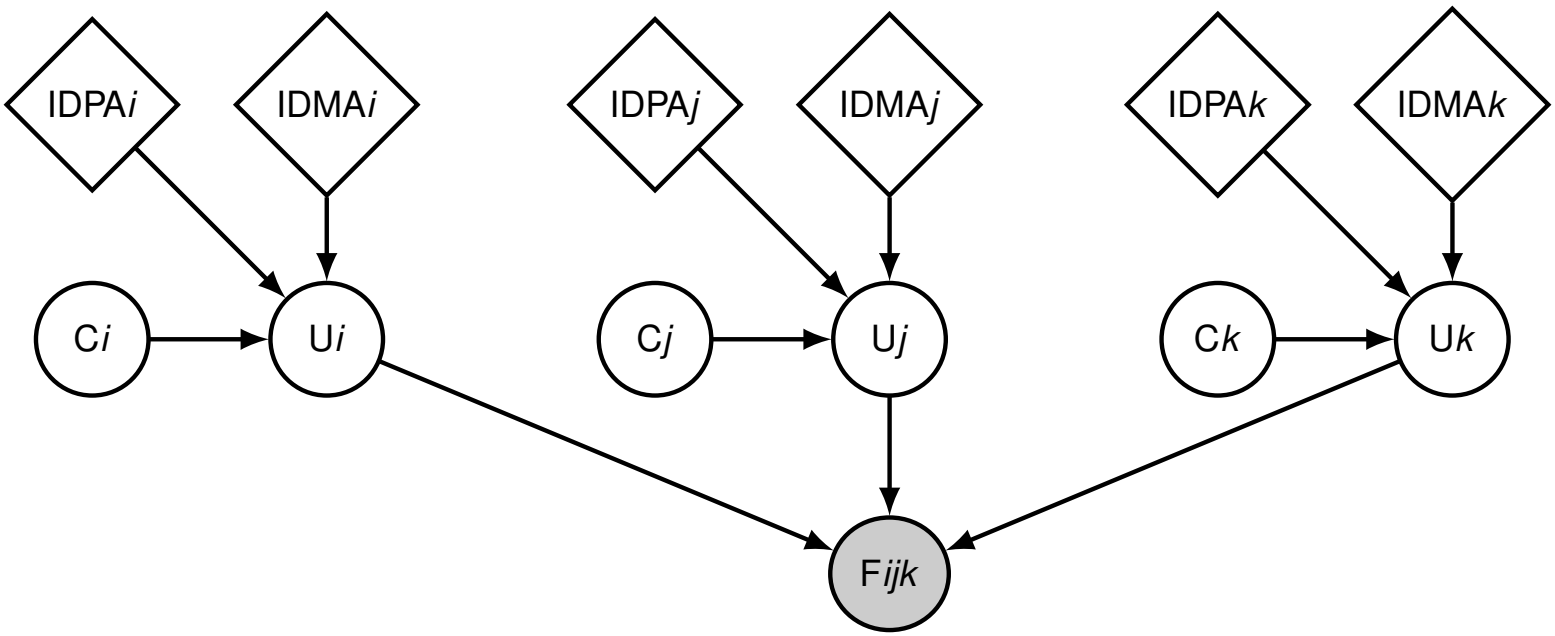
$$\text{F}ij(\Phi) = \frac{1}{4}\left(\mathbf{1}_{\text{IDMA}i=\text{IDMA}j} + \mathbf{1}_{\text{IDMA}i=\text{IDPA}j} + \mathbf{1}_{\text{IDPA}i=\text{IDMA}j} + \mathbf{1}_{\text{IDPA}i=\text{IDPA}j}\right)$$

$S_{\mathbf{all}}$:

$$S_{\text{all}}(\Phi) = \frac{1}{2^n}\sum_{\text{U}_{1:n}} H(\text{U}_{1:n})$$

where Ui is either IDMAi or IDPAi, and $H(\text{U}_{1:n})$ is the number of non-trivial permutations that leave $\text{U}_{1:n}$ unchanged.

Computing S_{all} for three affected



Remarks

- Computing $T(X, Y)$ could be numerically tedious;
- Empirical computation of $T(X, Y)$ possible:

$$\widehat{T}(X, Y) = \frac{1}{N} \sum_{j=1}^{N} S(\Phi^j) \text{ with } (\Phi^1, \ldots, \Phi^n) \text{ from } \mathbb{P}(\Phi | X, Y)$$

- Choice of Y for $H0$ unclear;
- Distribution of $T(X, Y)$ under $H0$ only empirically

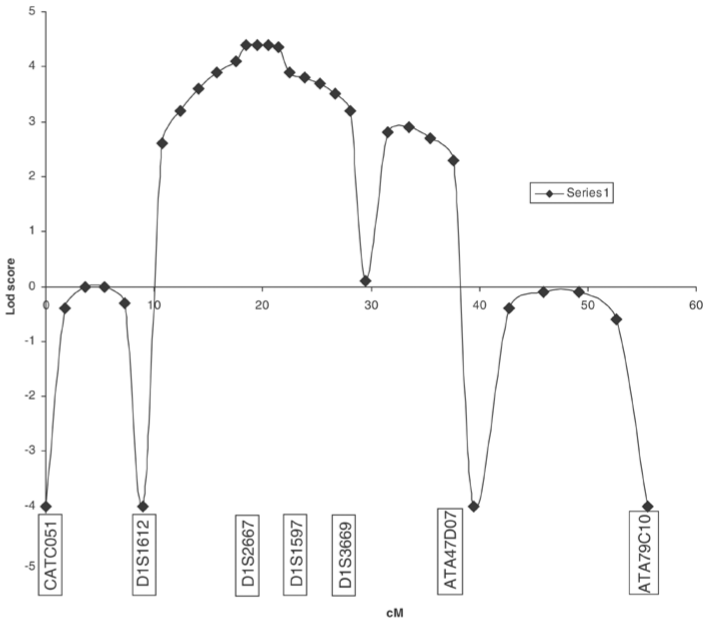
$\Rightarrow$ Approach often limited to simple designs (ex: affected sibs)

Outline

What is the Linkage Analysis ?

LD: two loci have no LD if the recombination rate between them is $\theta = 0.5$. Ex: two loci on two different chromosomes.

Linkage Analysis: localize the disease locus by measuring the LD with observed loci.



Data:

- pedigrees with disease status
- observed genotypes at one or several loci
- a disease model from previous segregation analysis

Two-point linkage

What's new:

- Two loci: disease locus DGi and observed locus Gi;
- Segregation model for disease status Yi: $\mathbb{P}(\text{Y}i|\text{DG}i)$;
- Inheritance variable at each locus: PDIk, MDIk, PIk, MIk

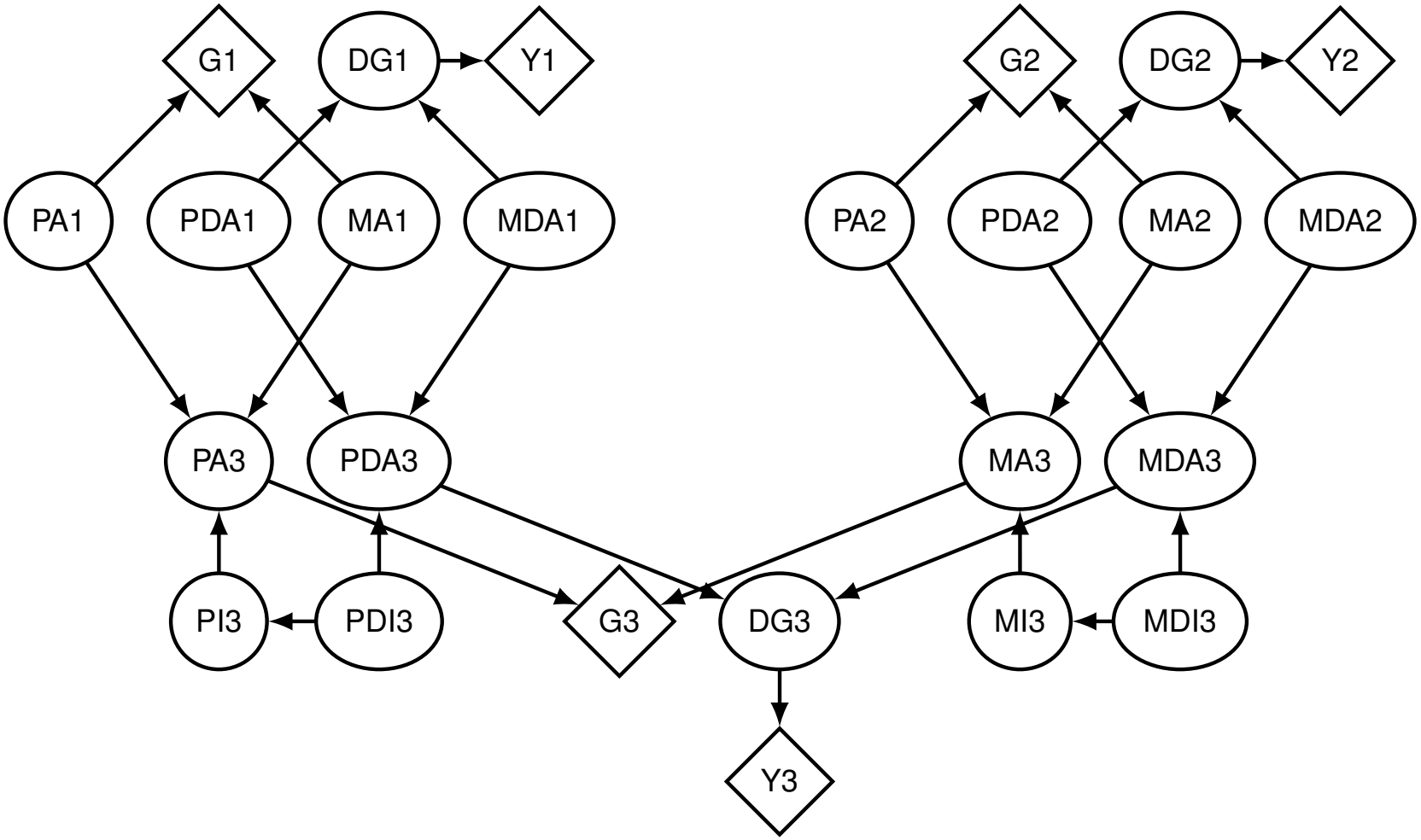
$$\mathbb{P}(\text{PDI}k = \text{maternal}) = \mathbb{P}(\text{PDI}k = \text{paternal}) = 0.5$$

$$\mathbb{P}(\text{MDI}k = \text{maternal}) = \mathbb{P}(\text{MDI}k = \text{paternal}) = 0.5$$

$$\mathbb{P}(\text{PI}k|\text{PDI}k) = \mathbb{P}(\text{MI}k|\text{MDI}k) = \begin{pmatrix} 1-\theta & \theta \\ \theta & 1-\theta \end{pmatrix}$$

with recombination rate θ.

Two-point linkage



Two-point linkage

Parameters:

- recombination rate: θ
- allele frequencies at disease locus: q
- allele frequencies at observed locus: p
- penetrance: f

Likelihood:

- $L(\theta, f, p, q) = \mathbb{P}(\mathrm{Y}|\mathrm{G}; \theta, p, q, f)$
- $\widehat{\theta} = \arg\max_\theta L(\theta, p, q, f)$ (p estimated, q, f known)

LOD score:

$$\mathrm{lod}(\widehat{\theta}) = \log_{10}\left(\frac{L(\widehat{\theta}; p, q, f)}{L(0.5; p, q, f)}\right) \sim \frac{\chi^2(\mathrm{df}{=}1)}{2\log 10} = \frac{\chi^2(\mathrm{df}{=}1)}{4.6}$$

Joint linkage and segregation

By adding ascertainment variables:

$$\mathbb{P}(Y|A=1,G) = \frac{\mathbb{P}(Y,A=1|G)}{\mathbb{P}(A=1|G)} = \frac{1-\mathbb{P}(Y,A=0|G)}{1-\mathbb{P}(A=0|G)}$$

$$LOD = \log_{10}\left(\frac{\mathbb{P}(Y|A=1,G;\widehat{\theta},\widehat{p},\widehat{q},\widehat{f},\widehat{\pi})}{\mathbb{P}(Y|A=1,G;0.5,\widehat{p},\widehat{q},\widehat{f},\widehat{\pi})}\right)$$

Through the MOD score: using $\mathbb{P}(G|Y,A=1) = \mathbb{P}(G|Y)$

$$MOD = \log_{10}\left(\frac{\mathbb{P}(G|Y;\widehat{\theta},\widehat{p},\widehat{q},\widehat{f})}{\mathbb{P}(G|Y;0.5,\widehat{p},\widehat{q},\widehat{f})}\right)$$

Multi-point linkage

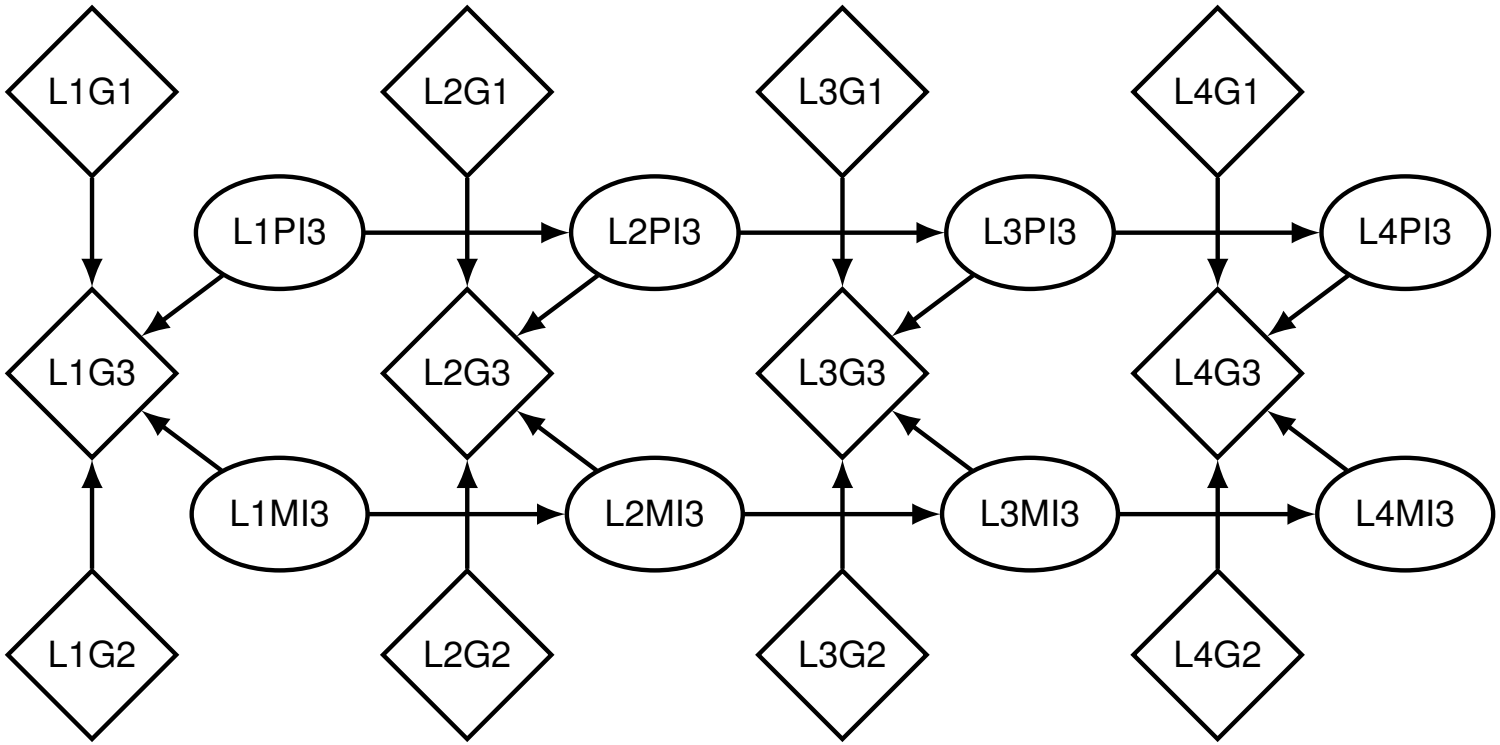
What's new:

- p loci: L1, L2, ..., Lp;
- θ_i: recombination rate between loci i and $i+1$;
- for any non founder k: (LiPIk)$_i$ and (LiPIk)$_i$ heterogeneous Markov chains with transition:

$$\tau_i = \begin{pmatrix} 1-\theta_i & \theta_i \\ \theta_i & 1-\theta_i \end{pmatrix}$$

- a disease locus D (and the observed phenotype) could be added between any two consecutive loci

Multi-point linkage



Multi-point linkage

Elston-Stewart (1971):

- peels one nuclear family after another;
- must store the current loci configuration;
- large pedigree (without loops) and few locus;

Lander-Green (1986):

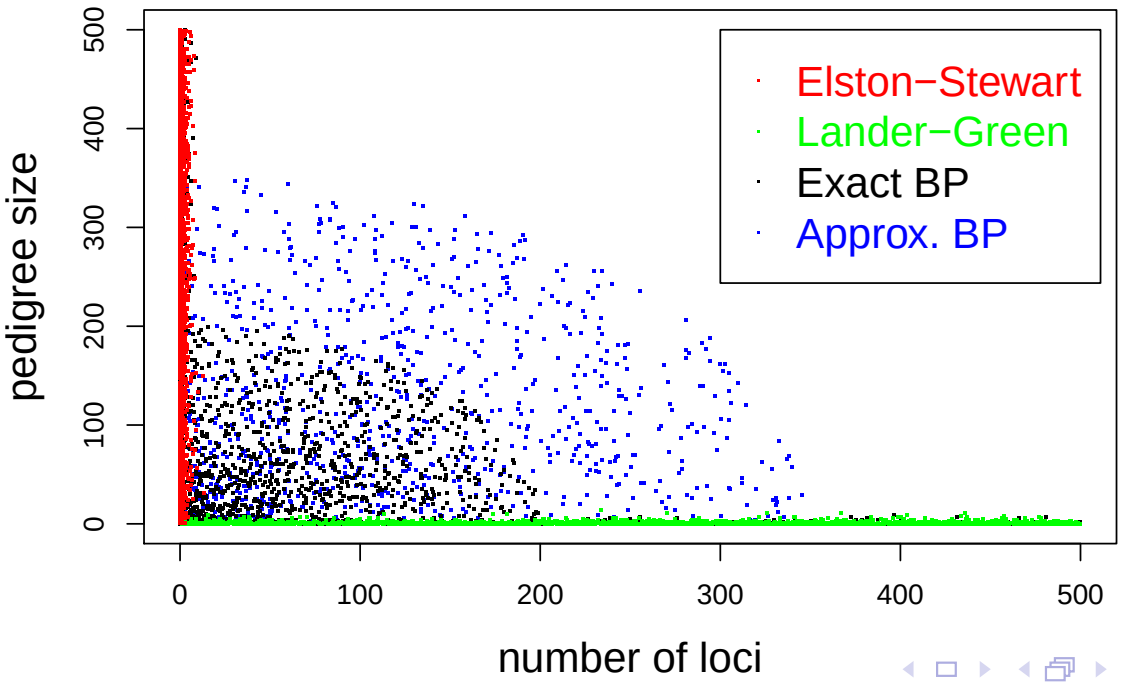
- peels one locus after another
- must store the current inheritance vector;
- small pedigree and many loci

Bayesian networks:

- search for a more efficient elimination order (JT)
- ex: Superlink (2005), RC_Link (2007)
- typical limit: 100 individuals, 50 loci

Why BNs ?

- BNs are natural models in genetics;
- suggests extensions (ex: ascertainement in segregation)
- suggests computation techniques (ex: efficient sampling);
- for complex BNs, powerful approx. BP algorithms available (ex: loopy belief, blocking Gibbs-sampling)



Bibliography I

D. Allen and A. Darwiche.

Rc_link: Genetic linkage analysis using bayesian networks.

International journal of approximate reasoning, 48(2):499–525, 2008.

R.C. Elston and J. Stewart.

A general model for the genetic analysis of pedigree data.

Human Heredity, 21(6):523–542, 1971.

D. Koller and N. Friedman.

Probabilistic graphical models.

MIT press, 2009.

L. Kruglyak, M.J. Daly, M.P. Reeve-Daly, and E.S. Lander.

Parametric and nonparametric linkage analysis: a unified multipoint approach.

American Journal of Human Genetics, 58(6):1347, 1996.

Bibliography II

E.S. Lander and P. Green.

Construction of multilocus genetic linkage maps in humans.

Proceedings of the National Academy of Sciences, 84(8):2363, 1987.

S.L. Lauritzen and N.A. Sheehan.

Graphical models for genetic analyses.

Statistical Science, pages 489–514, 2003.

N.D. Ma'ayan Fishelson and D. Geiger.

Maximum likelihood haplotyping for general pedigrees.

Human Heredity, 59:41–60, 2005.

A.S. Whittemore and J. Halpern.

A class of tests for linkage using affected pedigree members.

Biometrics, pages 118–127, 1994.