The Elston-Stewart Algorithm

Biostatistics 666

Lecture 24

Scheduling – Important Dates

- Remaining Lectures, April 5, 7, 14
- Polio Symposium, April 12
  - Rackham Auditorium, starts at 9:30
- Review Session, April 19
- Final Exam, April 27

Last Lecture

- The Lander Green Algorithm in practice
- Computational refinements
  - Speeding up transitions
  - Reducing inheritance space
- Non-parametric linkage analysis

Today …

- Elston-Stewart Algorithm
  - Another approach to pedigree likelihoods
  - Can handle very large pedigrees
  - Limited to a few markers
- Prelude to discussion of parametric linkage analysis

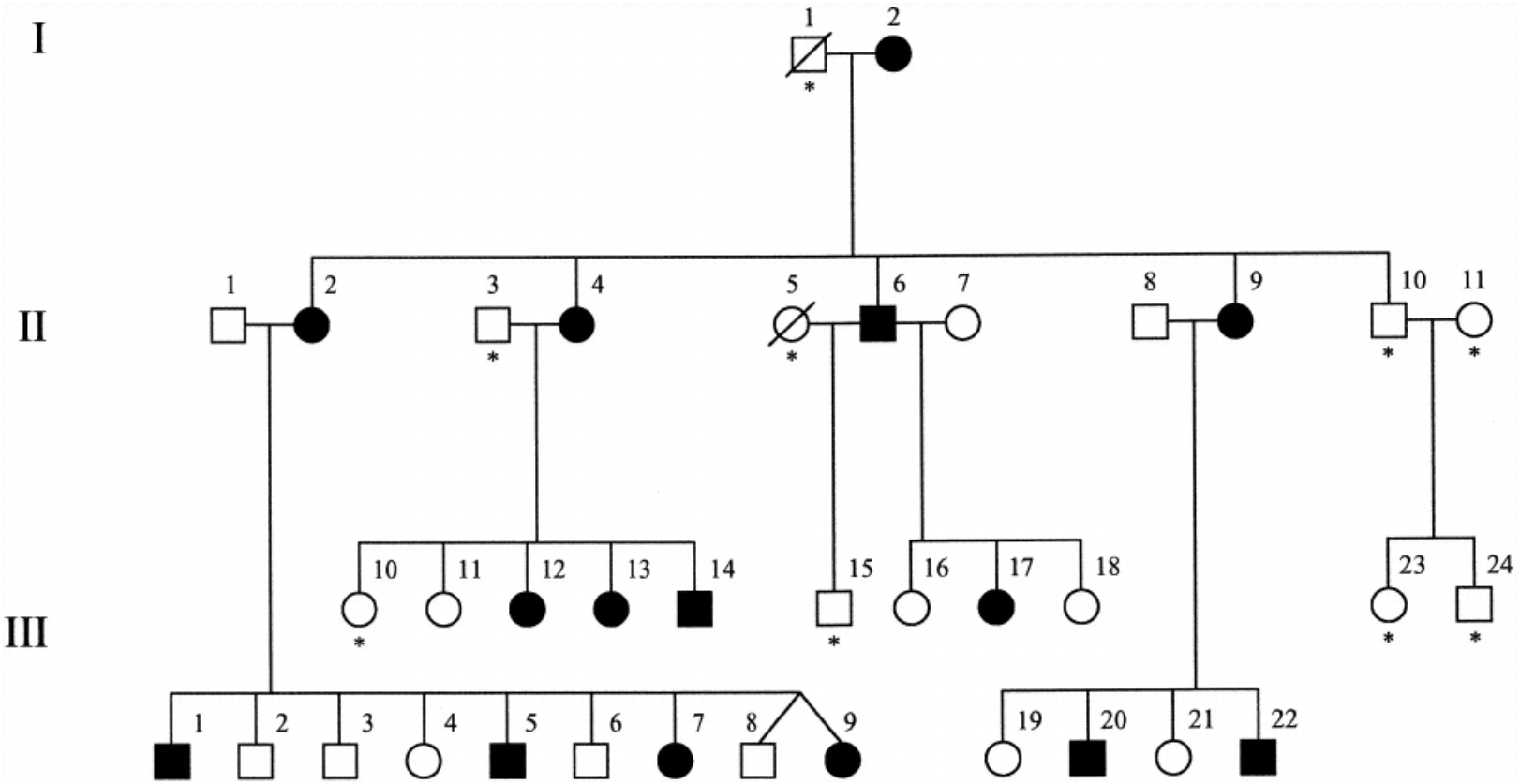
So far ...

- Studying linkage to complex diseases
  - Multiple environmental factors
  - Multiple genetic susceptibility factors
- Study affected sib pairs or small families
- Find regions of excess similarity among affected individuals

What about Mendelian traits?

- Diseases caused by a single genetic defect
- Typically, these are extremely rare
  - Due to natural selection
- We don’t need many families for mapping, but rather one or a few large families, where it is possible to track segregation of mutant alleles

Typical Family for Mapping Mendelian Trait...



The Problem

- These families are typically too large for the Lander-Green algorithm
- Impractical to enumerate all potential inheritance graphs…
- Need an alternative formulation for the pedigree likelihood

Elements of Pedigree Likelihoods

- **Prior Probabilities**
  - For founder genotypes
- **Segregation probabilities**
  - For offspring genotypes, given parents
- **Penetrances**
  - For individual phenotypes, given genotype

Prior Probabilities for Founders

- $P(G_{founder})$
- Assume Hardy-Weinberg equilibrium
  - Based on allele frequencies
- May be multilocus frequencies
  - Typically, assuming linkage equilibrium

Segregation Probabilities

- $P(G_o \mid G_f , G_m)$
- Probability of offspring genotype conditional on parental genotypes
  - Follows from Mendel's laws
- For multiple loci, the probability of offspring haplotypes conditional on parental haplotypes

Segregation Probabilities

- For multiple markers, use “haplo-genotypes”

- $P(G_o \mid G_f , G_m)$
  - $G_o = (H_{o1}, H_{o2})$
  - $G_f = (H_{f1}, H_{f2})$
  - $G_f = (H_{m1}, H_{m2})$

- $P(G_o \mid G_f , G_m) =$

  $P(H_{o1} \mid H_{f1}, H_{f2})P(H_{o2} \mid H_{m1}, H_{m2}) +$
  $P(H_{o2} \mid H_{f1}, H_{f2})P(H_{o1} \mid H_{m1}, H_{m2})$

Penetrances

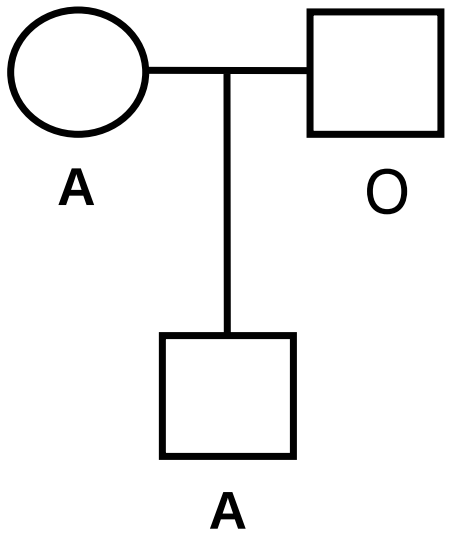
- $P(X_i \mid G_i)$
- Probability of observed phenotype conditional on genotype
- Generally, assume that phenotypes are independent within families

Overall Pedigree Likelihood

$$L = \sum_{G_1} \cdots \sum_{G_n} \prod_{f} P(G_f) \prod_{\{o,f,m\}} P(G_o \mid G_f, G_m) \prod_{i} P(X_i \mid G_i)$$

- Notice the three elements:
  - Probability of founder genotypes
  - Probability of children given parents
  - Probability of phenotypes given genotypes

Simple Example...



- Phenotypes are for the ABO locus
- Calculate:
  - Likelihood for pedigree
  - Likelihood conditional on alternative genotypes for I-1

Computationally …

$$L = \sum_{G_1} \cdots \sum_{G_n} \prod_{i} P(X_i \mid G_i) \prod_{founder} P(G_{founder}) \prod_{\{o,f,m\}} P(G_o \mid G_f, G_m)$$

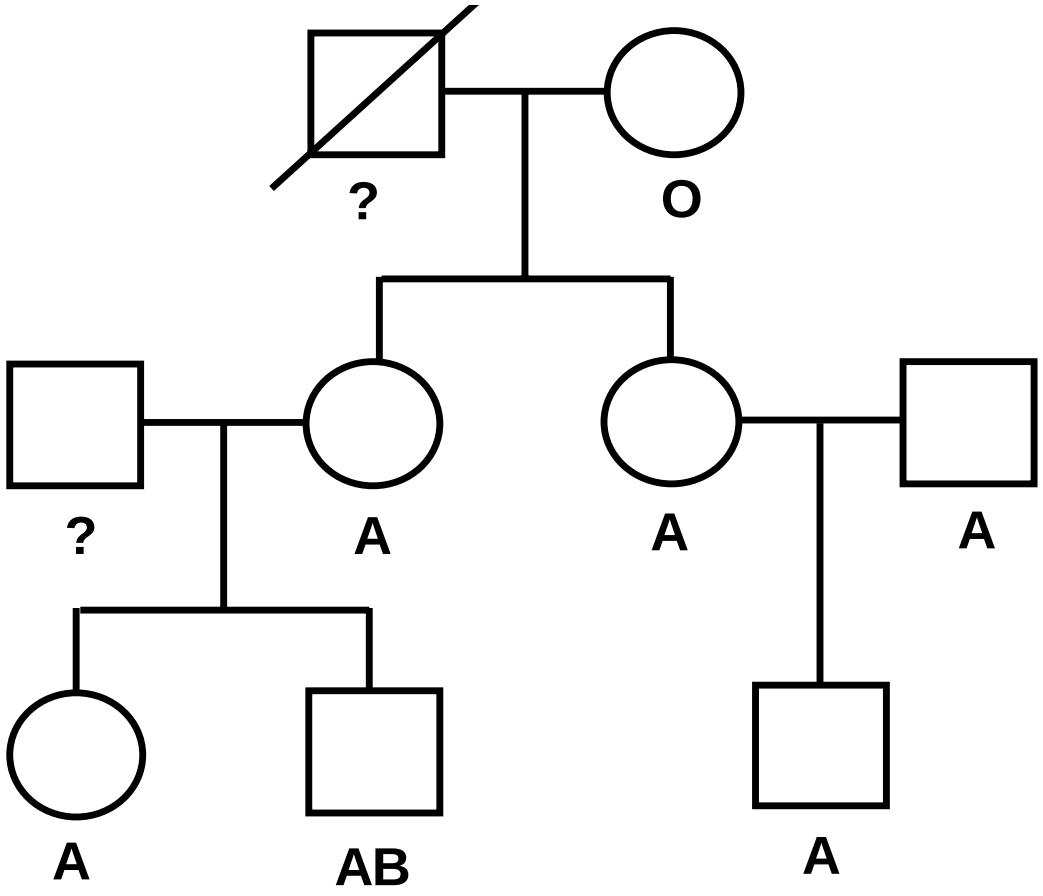
- Computation rises exponentially with #people
- Computation rises exponentially with #markers

- Challenge is summation over all possible genotypes (or haplotypes) for each individual

Typical calculation

- List all possible genotypes
- Create reduced lists
  - Eliminate those where $P(X|G) = 0$
  - Eliminate those where $P(G_o|G_f,G_m) = 0$
- Iterate over all possibilities

Example Pedigree



Iteration over All Genotypes

- 9 individuals
- 3 ABO alleles
  - 6 possible genotypes
- Potential genotype sets
  - $6^9 = 10,077,696$

Condition on Phenotype

Person	Genotypes	#Genotypes
I-1	{AA, AO, BB, BO, AB, OO}	6
I-2	{OO}	1
II-1	{AA, AO, BB, BO, AB, OO}	6
II-2	{AA, AO}	2
II-3	{AA, AO}	2
II-4	{AA, AO}	2
III-1	{AA, AO}	2
III-2	{AB}	1
III-3	{AA, AO}	2

1152 possibilities to consider

Condition on Family Members

Person	Genotypes	#Genotypes
I-1	{AA, AO, AB}	3
I-2	{OO}	1
II-1	{BO, AB}	2
II-2	{AO}	1
II-3	{AO}	1
II-4	{AA, AO}	2
III-1	{AA, AO}	2
III-2	{AB}	1
III-3	{AA, AO}	2

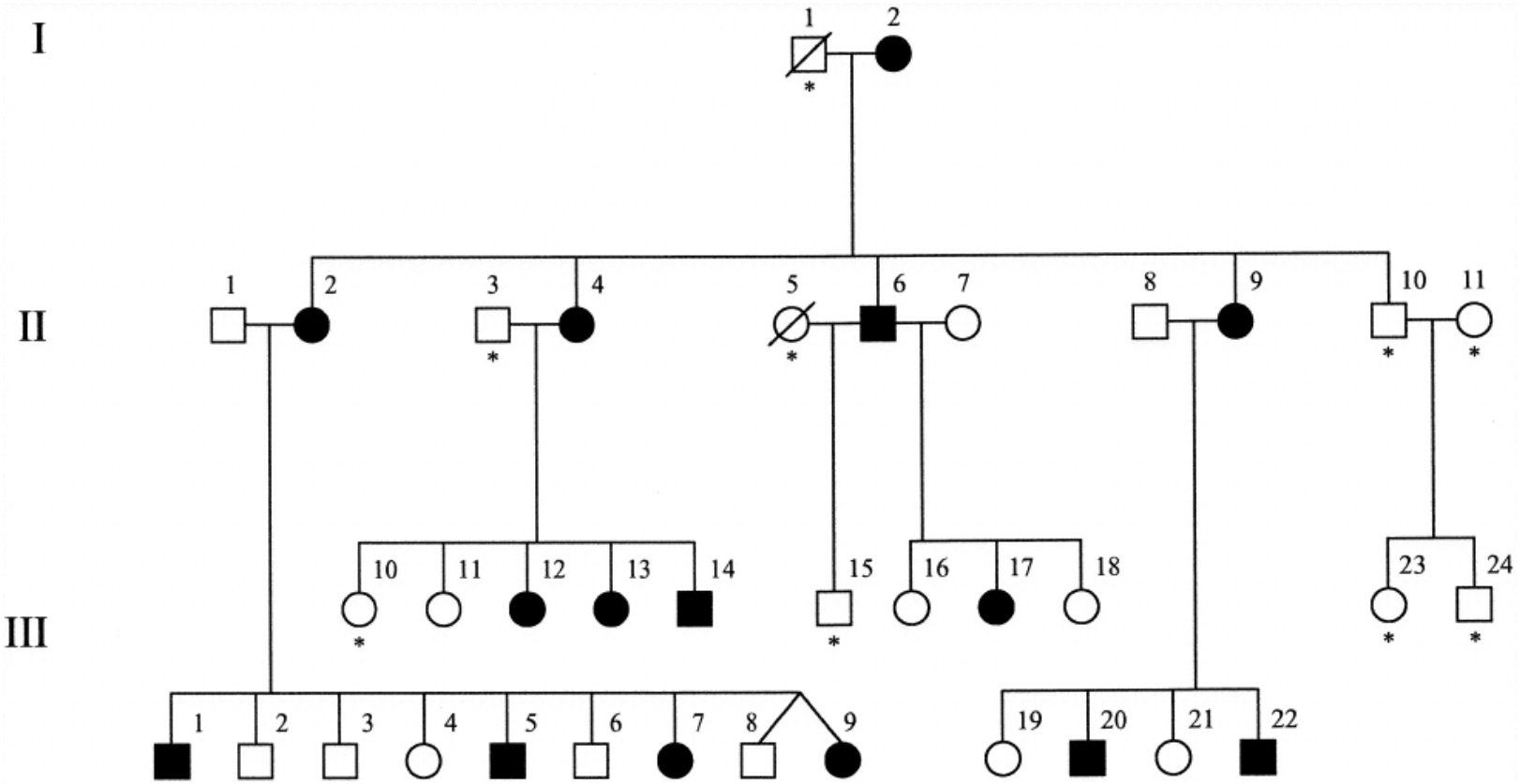
48 possibilities

Simplification for Nuclear Families

$$L = \sum_{G_m} P(X_m \mid G_m) P(G_m) \sum_{G_f} P(X_f \mid G_f) P(G_f) \prod_o \sum_{G_o} P(X_o \mid G_o) P(G_o \mid G_m, G_f)$$

- Conditional on parental genotypes, offspring are independent
- Thus avoid nested sums, and produce likelihood whose cost increases linearly with the number of offspring

What about our large pedigree?



Elston and Stewart's (1971) insight...

- Focus on "special pedigrees" where
  - Every person is either:
    - Related to someone in the previous generation
    - Marrying into the pedigree
  - No consanguineous marriages

- Process nuclear families, by fixing the genotype for one parent …

Successive Conditional Probabilities

- Starting at the bottom of the pedigree…
- Calculate conditional probabilities by fixing genotypes for one parent
- Specifically, calculate $H_k(G_k)$
  - Probability of descendants and spouse for person k
  - Conditional on a particular genotype G_k

Formulae ...

- So for each parent, calculate:

$$H_{parent}(G_{parent}) = \sum_{G_{spouse}} P(X_{spouse} | G_{spouse}) P(G_{spouse}) \prod \sum P(X_o | G_o) P(G_o | G_{parent}, G_{spouse}) H_o(G_o)$$

- By convention, for individuals with no descendants:

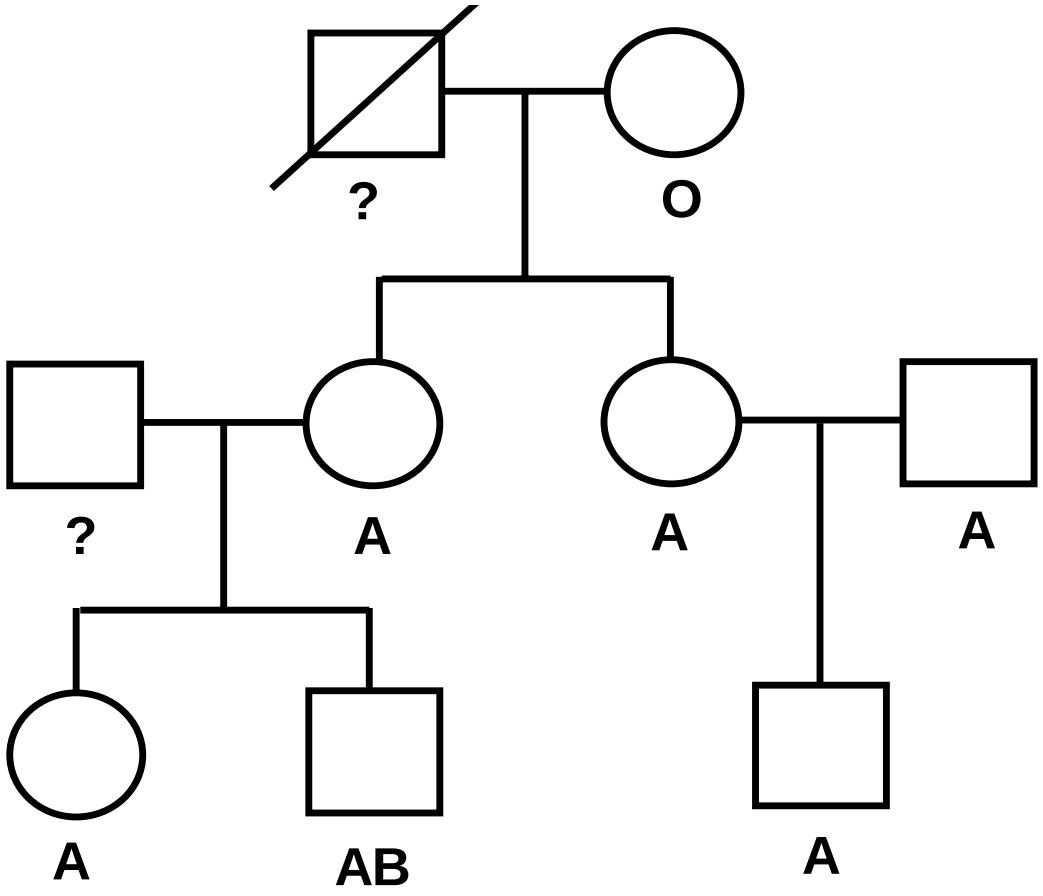
$$H_{leaf}(G_{leaf}) = 1$$

Final Likelihood

- After processing all nuclear family units …
- Simple sum gives the overall pedigree likelihood:

$$L = \sum_{G_{founder}} P(X_{founder} \mid G_{founder}) H(G_{founder}) P(G_{founder})$$

Example Pedigree



Condition on Family Members

Person	Genotypes	#Genotypes
I-1	{AA, AO, AB}	3
I-2	{OO}	1
II-1	{BO, AB}	2
II-2	{AO}	1
II-3	{AO}	1
II-4	{AA, AO}	2
III-1	{AA, AO}	2
III-2	{AB}	1
III-3	{AA, AO}	2

48 possibilities

Steps

- Conditional Probabilities at II-3
  - Using phenotypes at II-4 and III-3
- Conditional Probabilities at II-2
  - Using phenotypes for II-1, III-1 and III-2
- Conditional Probabilities at I-1 (or I-2)
  - Using phenotypes for II-2 and II-3 and conditional probabilities for their descendants

Elston Stewart Applicability

- Potentially large pedigrees
  - But structure of the pedigree must be simple
  - Only a little inbreeding can be accommodated
- Limited to a small number of markers
  - Complexity exponential on number of markers

Today

- Elston Stewart Algorithm
  - Alternative approach for pedigree analysis
  - Can handle relative large pedigrees
- Implemented in the LINKAGE and FASTLINK computer packages