

*Likelihood Calculation in
Large Families*

Biostatistics 666

Previously

- Analysis of Unrelated Individuals
- Computational Methods Well Suited to Handling Many Markers in Sibling Pairs and Small Families
- **Today:** Computational Methods Well Suited to Larger Pedigrees

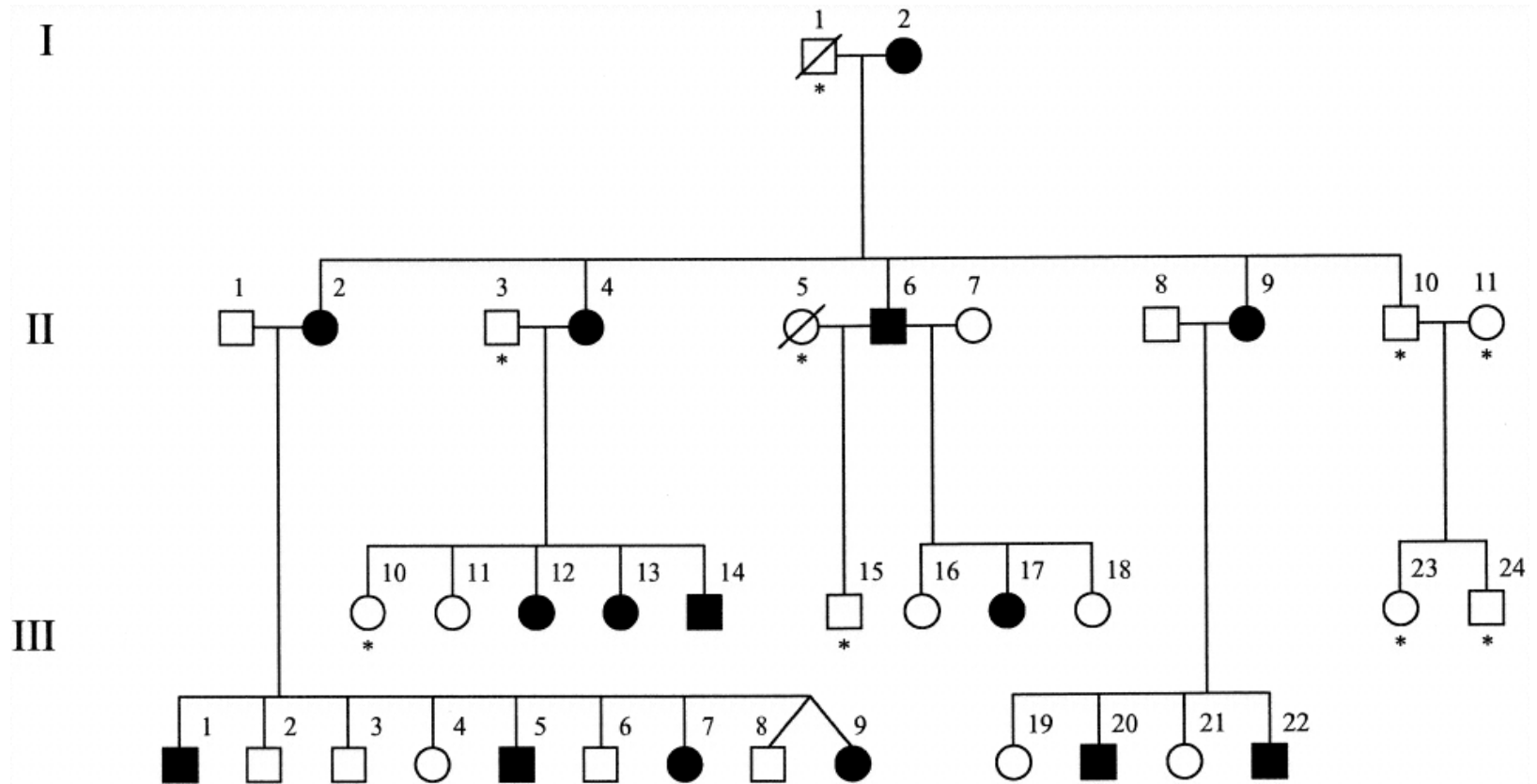
Today ...

- Alternative formulation of likelihood for related individuals
- Elston-Stewart Algorithm
 - Alternative to Hidden Markov Models
 - Can handle large pedigrees
 - Limited to a few markers

Why Study Large Families?

- Traditionally, large families have been used to map Mendelian traits
- In general, large families are an ideal setting to study the behavior of rare alleles
 - Greatly increased likelihood that multiple individuals with the same rare allele will be observed
- As our ability to examine rare variation increases, there is renewed interest in large pedigrees

Typical Family for Mapping Mendelian Trait...



The Problem

- These families are typically too large for the Hidden Markov Models we have discussed
- Impractical to enumerate all potential IBD configurations / 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

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

Segregation Probabilities

- $P(G_o | G_f, G_m)$
- 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

Penetrances

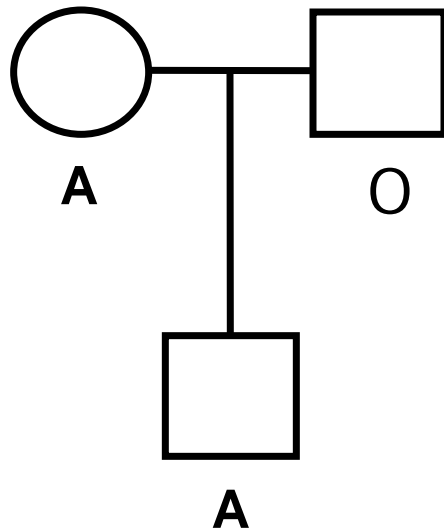
- $P(X_i | G_i)$
- Probability of observed phenotype conditional on genotype
- Generally, assume that phenotypes are independent within families

Overall Pedigree Likelihood

$$L = \sum_{G_1} \dots \sum_{G_n} \prod_f P(G_f) \prod_{\{o,f,m\}} P(G_o | G_f, G_m) \prod_i P(X_i | G_i)$$

- 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 mother being A/A vs A/O

Computationally ...

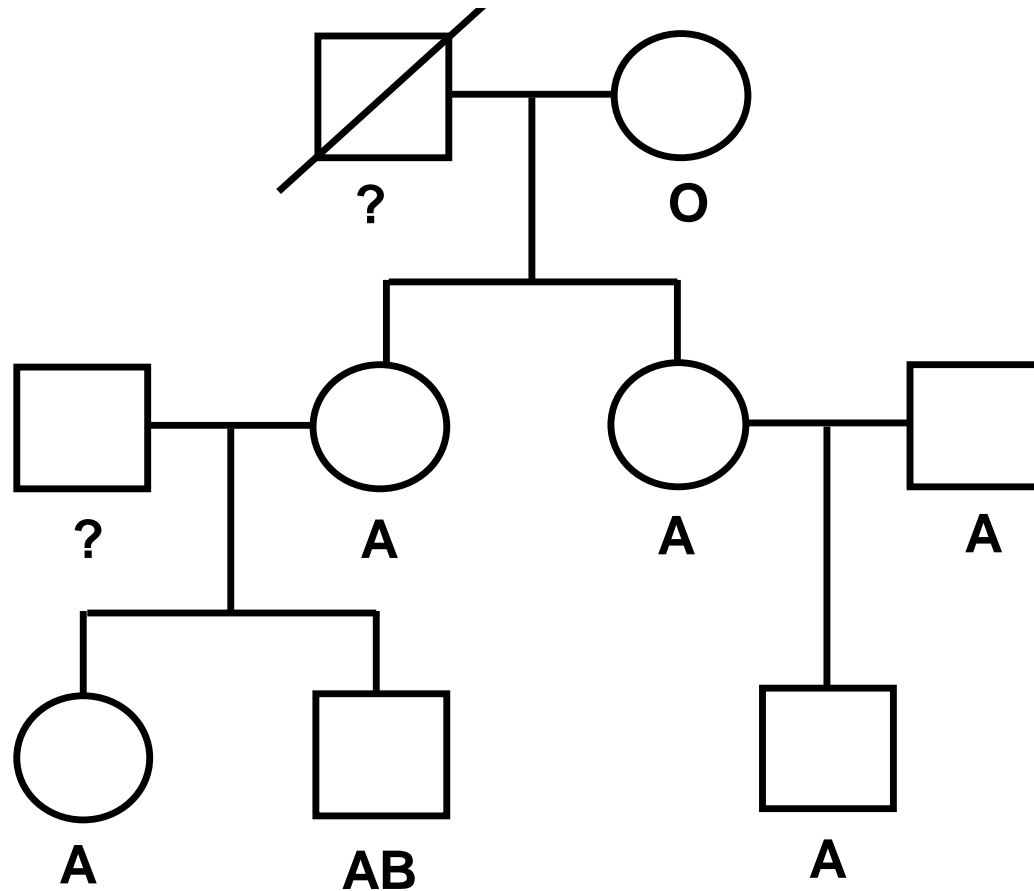
$$L = \sum_{G_1} \dots \sum_{G_n} \prod_i P(X_i | G_i) \prod_{founder} P(G_{founder}) \prod_{\{o,f,m\}} P(G_o | G_f, G_m)$$

- 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

*Can We Avoid Iteration
over all Genotypes*

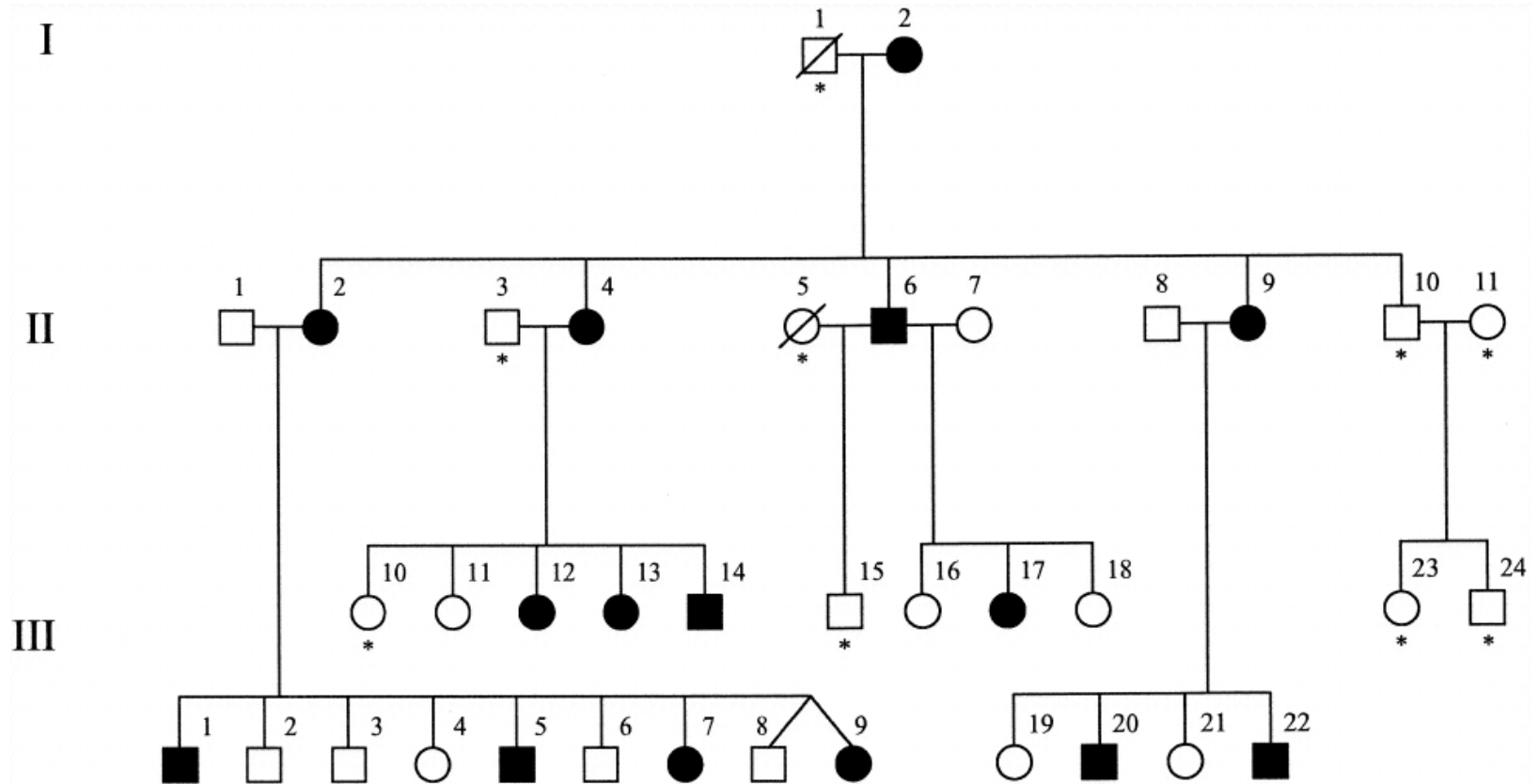
Quickly becomes
extremely costly!

Solution for Nuclear Families

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

- Conditional on parental genotypes, offspring are independent
- Avoids nested sums and results in computation whose cost increases linearly with the number of offspring

What about large pedigrees?



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 person k having 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:

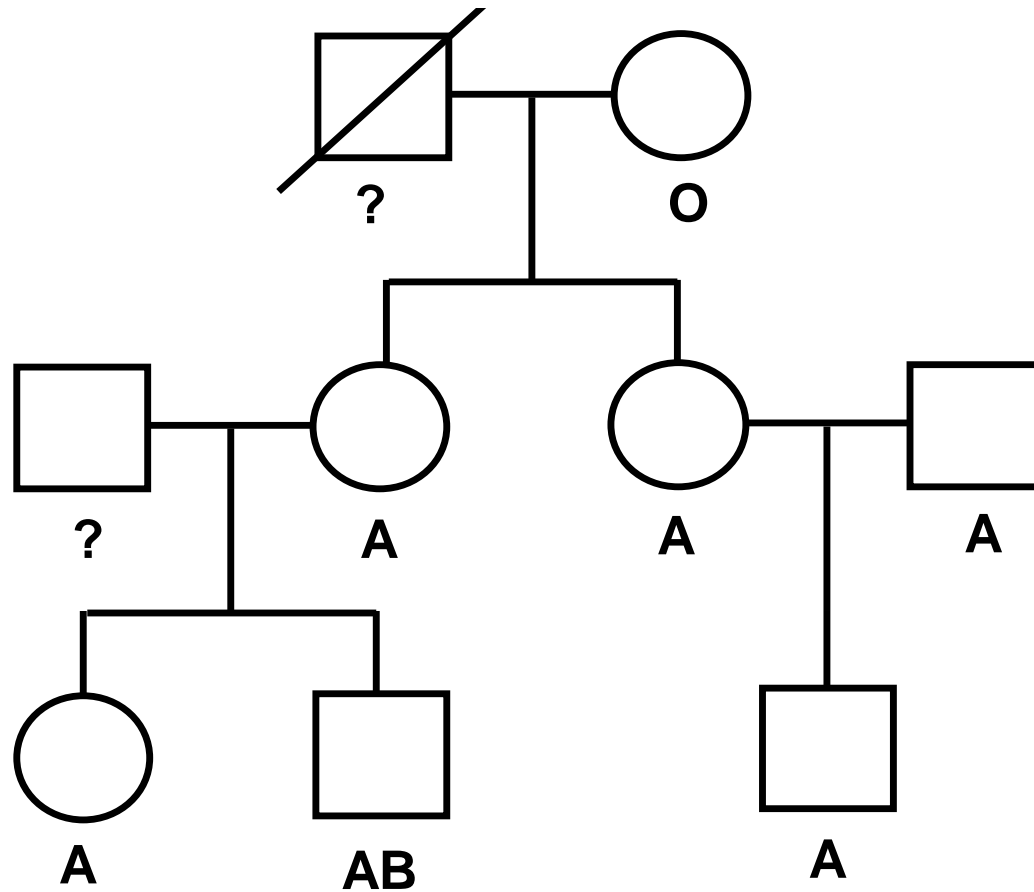
$$H_{kf}(G_{kf}) = 1$$

Final Likelihood

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

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

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

Dealing With Many Markers in Complex Pedigrees

- Hidden Markov Models based on IBD configurations can handle many markers in small pedigrees
- Pedigree “Peeling” algorithms that sum over individual genotypes can deal with a few markers in large pedigrees
- Pedigrees beyond the reach of these strategies must be handled using approximate Monte Carlo methods

Today

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