IBD computations in large pedigrees

February 23, 2006, pm
Yurii Aulchenko

IBD computation in large pedigrees

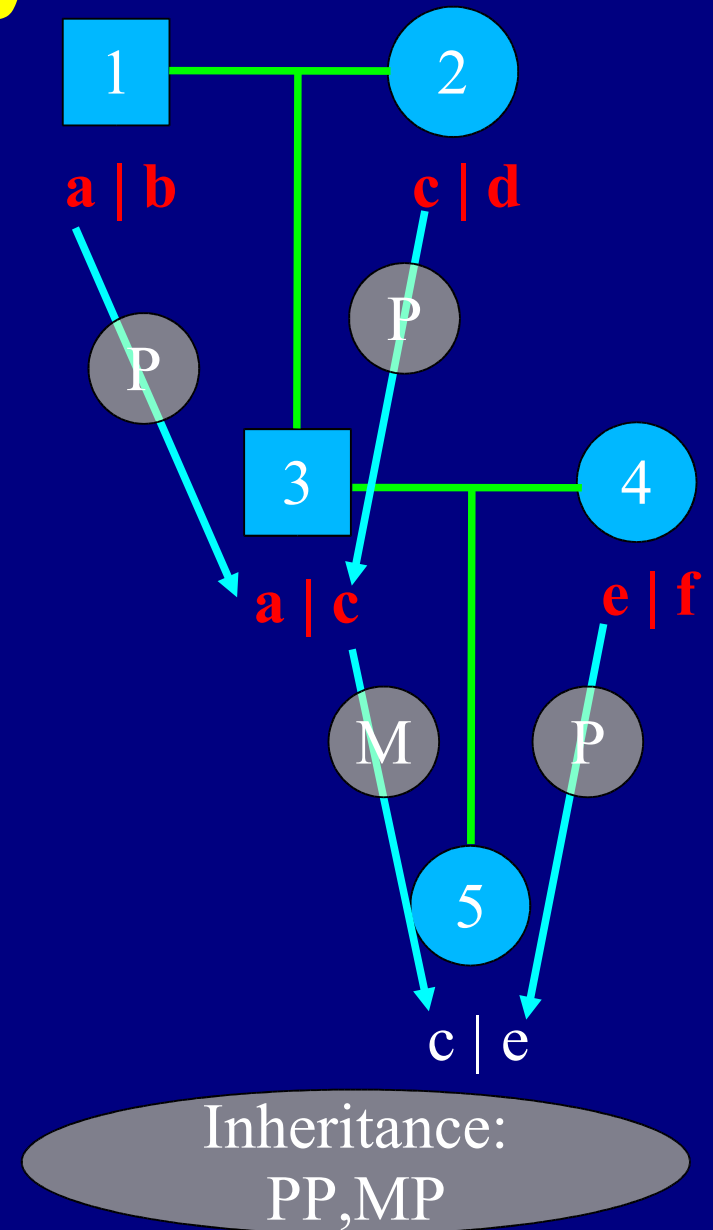Is part of a problem of computation of the likelihood function of a pedigree

Computation of the likelihood

- X – matrix of phenotypes (trait, marker data) of pedigree members

- G – matrix of underlying genotypes of pedigree members

- Likelihood:

$$\Sigma_{\text{all } G} \; P(X|G)P(G)$$

Complexity

- How much is "all G"?
  - Number of genotypes possible for a founder
    - by
  - number of inheritance patterns

Complexity

- How much is "all G"?
  - Number of genotypes possible for a founder:
    - One trait locus, 1 marker with 10 alleles:
      - $2*2*10*10 = 400$ combinations
    - One trait locus, 5 markers with 10 alleles:
      - $4*100^5 = 40,000,000,000$ combinations
  - Number of inheritance patterns
    - 5 non-founders = $2^{2*5} = 1024$
    - 15 non founders = $2^{2*15} = 1,073,741,824$

Computation time

- Sibship of 5:
    - 1 marker   => 2 sec.
    - 2 markers   => 4.5 hours
    - 3 markers   => 5 years

Trait locus + 2 markers
    - 6 sibs   => 18 hours
    - 8 sibs   => 12 days
   - 10 sibs   => ½ year

Peeling

- A computational technique which reduces the complexity of computations by sequential computations over conditionally independent events

Elston-Stewart peeling

- **For parts of pedigree, compute probability conditional on all possible genotypes of members who connect this part to the rest**

- **Computation time**

  - **~ No_people * (no_possible_genotypes) * (no_loop-breakers)**

  Grows

    - **linear with no. people**
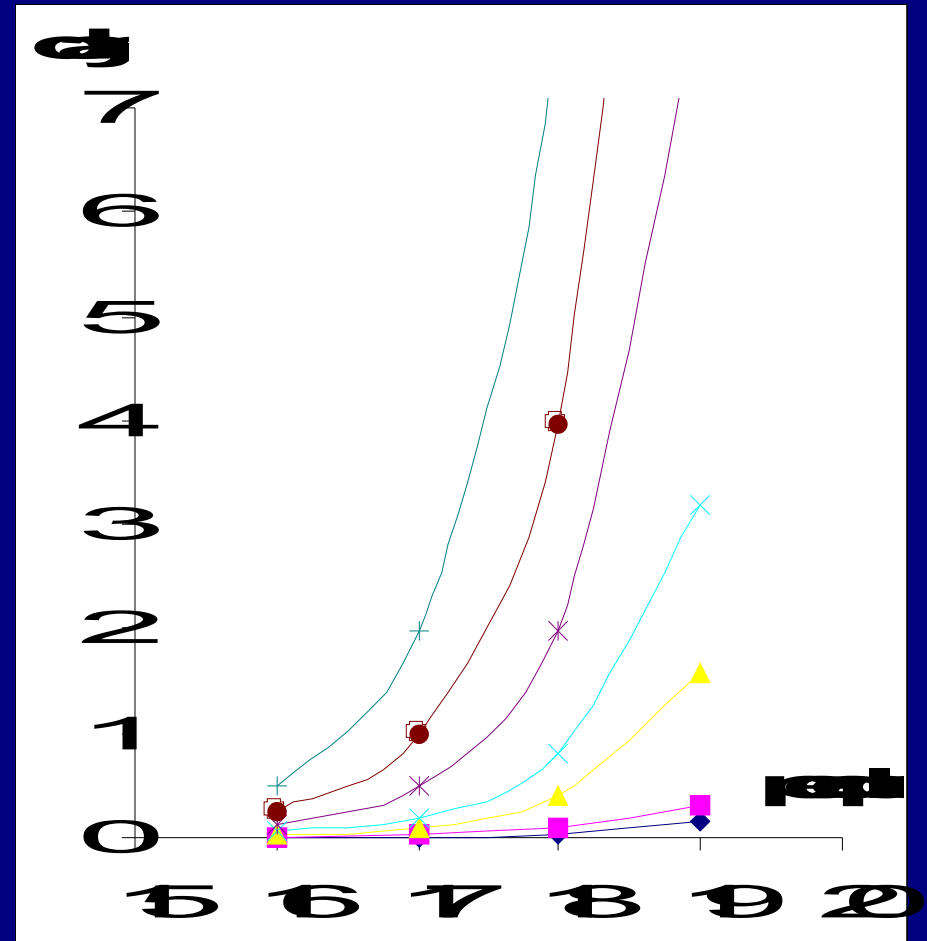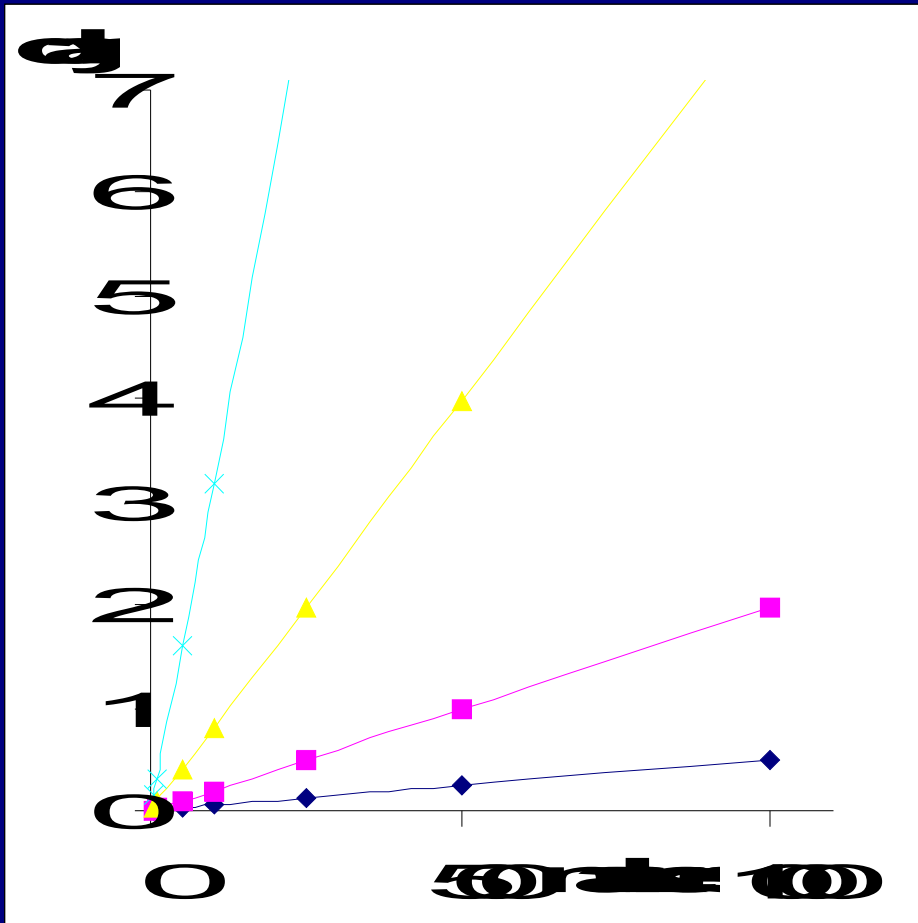   - **exponential with no. markers**

Computation time for ES-algorithm
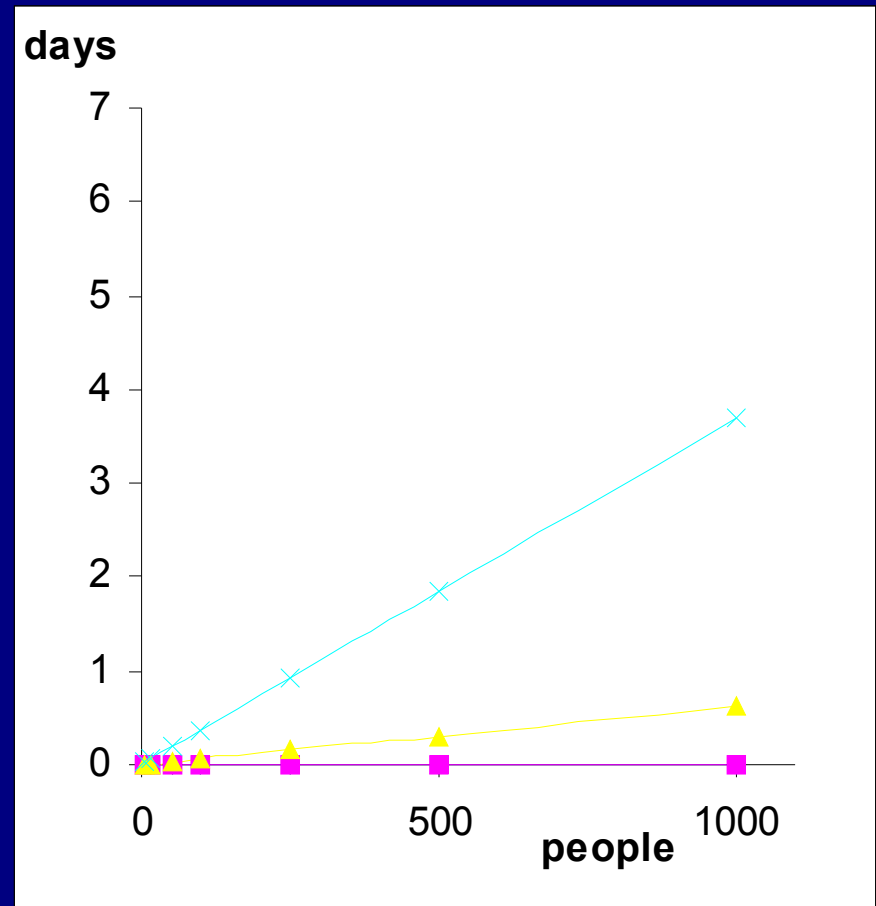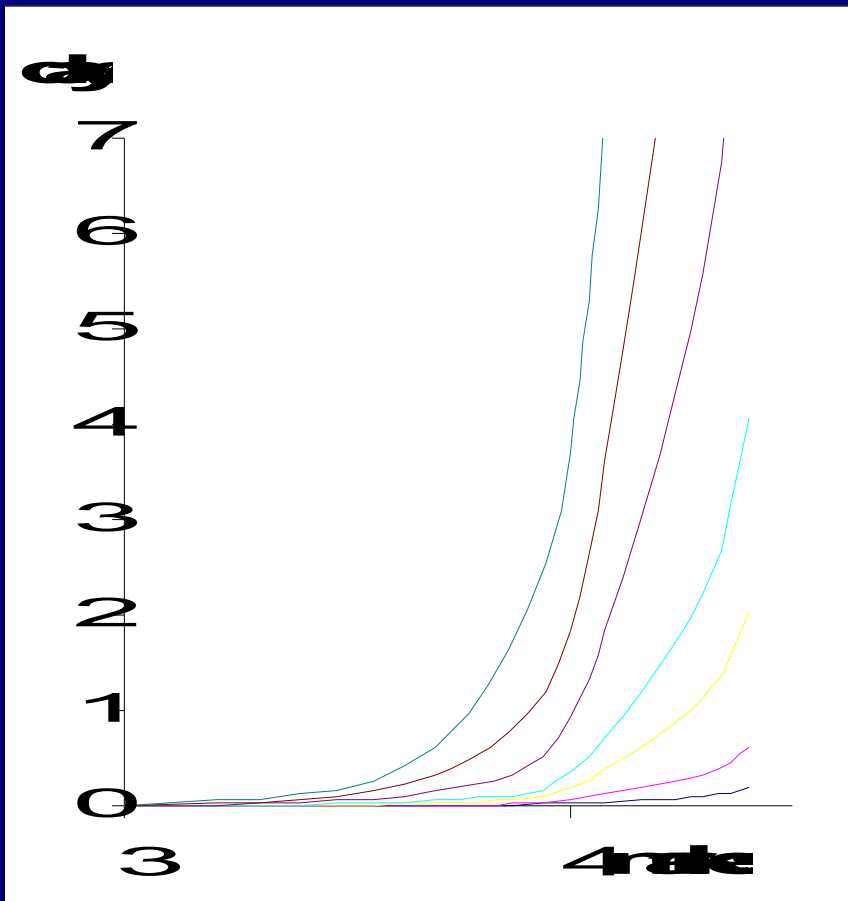
Lander-Green peeling

- **For particular marker (phenotype), compute probability for all pedigree members conditional on flanking genotypes**

- **Computation time**

  ~ No_loci * (no_possible_inheritance_patterns) Grows

  - **Exponential with no. people**
  - **Linear with no. markers**

Computation time for LG-algorithm

Limitations of exact methods

- E-S algorithm:
  - 3-4 markers and 2-3 loops is absolute maximum

- LG algorithm:
  - BIT-SIZE ~ 20-28
    - BIT-SIZE = (2*no_non-founders - no_founders)

Markov Chain Monte-Carlo

- A technique to compute approximate probabilities

- Can run well for few hundreds of people and few dozens of loops (takes weeks to finish)

- Very efficient on pedigrees ~50-70 people

- Results depend on the choice of random numbers

Programs for IBD computation

- Exact IBD using Elston-Stewart algorithm

  MERLIN, ALLEGRO, GENEHUNTER

- Exact single-marker IBD for zero-loop pedigrees

  SOLAR

- MCMC approxmation

  SIMWALK2: great results, but very slow

  LOKI: can be faster, but user has to control convergence