6. Computation of Mendelian Likelihood

§6.1. General problem and formulation

Given a pedigree of \( n \) members, let \( X_i, i = 1, \ldots, n \), denote the observed phenotypes (e.g., disease status) of the \( n \) members. Let \( G_i, i = 1, \ldots, n \), denote the members’ concerned unobserved genotypes (in the case of multilocus, \( G_i \) is the combined genotypes). Assume:

1. Hardy-Weinberg and linkage equilibrium hold;
2. No recombination interference.

The problem is to compute the likelihood function

\[
L = P(X_1, \ldots, X_n) \\
= \sum_{G_1} \ldots \sum_{G_n} P(X_1, \ldots, X_n | G_1, \ldots, G_n) P(G_1, \ldots, G_n).
\]
The conditional probability $P(X_1, \ldots, X_n | G_1, \ldots, G_n)$ can be factorized as

$$P(X_1, \ldots, X_n | G_1, \ldots, G_n) = \prod_i P(X_i | G_i),$$

since given a person’s genotype the phenotype of the person is independent of other’s genotypes. The probabilities $P(X_i | G_i)$ is called penetrance.

The probability $P(G_1, \ldots, G_n)$ can be factorized as a product of conditional probabilities and founder’s genotype probabilities. For example, consider a nuclear family with child $k$ and parents $i$ and $j$. For fixed genotypes $(G_k, G_i, G_j)$,

$$P(G_k, G_i, G_j) = P(G_k | G_i, G_j)P(G_i, G_j).$$

The probability $P(G_k | G_i, G_j)$ is called transmission probability. The probability $P(G_i, G_j)$ can be further factorized as

$$P(G_i, G_j) = P(G_i)P(G_j),$$
where the probabilities \( P(G_i) \), \( P(G_i) \) are called prior probabilities.

- Remarks on the computation of the probabilities
  
  - Priors

  The computation of priors is based on (i) the Hardy-Weinberg and linkage equilibria and (ii) the independence between founders.

  - Transmission probabilities

  Let \( \text{Tran}(G_k|G_i,G_j) \) denote the transmission probability that a mother \( i \) with genotype \( G_i \) and a father \( j \) with genotype \( G_j \) produce a child \( k \) with genotype \( G_k \). If the phase of \( G_k \) is known as \( U_k/V_k \) (by convention, the first allele is from the mother) then

  \[
  \text{Tran}(G_k|G_i,G_j) = \text{Tran}(U_k|G_i)\text{Tran}(V_k|G_j).
  \]
The factors $\text{Tran}(U_k|G_i)$ and $\text{Tran}(V_k|G_j)$ are **gamete transmission probabilities**. The computation of gamete transmission probabilities are based on Mendelian segregation law and Haldane’s recombination model.

**Mendelian law:** A parent transmit each of his or her two alleles to a child with equal probability $\frac{1}{2}$.

**Haldane’s model:** Recombination occurs independently on disjoint intervals.

- **Penetrance**

Penetrance specifies the likelihood of a phenotype $X$ given an genotype $G$, denoted by $\text{Pen}(X|G)$. If $X$ is a continuous trait, $\text{Pen}(X|G)$ is a density function. If $X$ is discrete, $\text{Pen}(X|G)$ is a probability. It is assumed that phenotypes are independent given the genotypes.
The likelihood function can be expressed in terms of penetrances, transmission probabilities and priors as

\[
L = \sum_{G_1} \cdots \sum_{G_n} P(X_1, \ldots, X_n|G_1, \ldots, G_n)P(G_1, \ldots, G_n)
\]

\[
= \sum_{G_1} \cdots \sum_{G_n} \prod_{i} \text{Pen}(X_i|G_i)\text{Pr}(G_1, \ldots, G_n)
\]

\[
= \sum_{G_1} \cdots \sum_{G_n} \prod_{i} \text{Pen}(X_i|G_i) \prod_{j} \text{Prior}(G_j) \prod_{\{k,l,m\}} \text{Tran}(G_m|G_k, G_l). 
\]

where \(j\) is taken over all founders, and \(\{k, l, m\}\) is taken over all parent-offspring triples.

**An example:** The likelihood of the pedigree with ordered ABO genotypes uncertain for some child:

1,2 — parents of 3; 3,4 — parents of 5.

1 \( (G_1 = AO) \), 2 \( (G_2 = AO) \), 3 \( (G_3 = A) \),

4 \( (G_4 = AO) \), 5 \( (G_5 = AA) \).
\[ P(G_1, G_2, G_3, G_4, G_5) \]

\[ = \sum_{G_3 \in \{A/A, A/O, O/A\}} P(G_1)p(G_2)P(G_4)P(G_3|G_1, G_2)P(G_5|G_3, G_4) \]

\[ = P(G_1)p(G_2)P(G_4)P(A/A|G_1, G_2)P(G_5|A/A, G_4) + P(G_1)p(G_2)P(G_4)P(A/O|G_1, G_2)P(G_5|A/O, G_4) + P(G_1)p(G_2)P(G_4)P(O/A|G_1, G_2)P(G_5|O/A, G_4) \]

\[ = P(G_1)p(G_2)P(G_4) \times \{P(A/A|G_1, G_2)P(G_5|A/A, G_4) + P(A/O|G_1, G_2)P(G_5|A/O, G_4) + P(O/A|G_1, G_2)P(G_5|O/A, G_4)\} \]

\[ = (2p_A p_O)^3 \left[ \frac{1}{4} \times \frac{1}{4} + \frac{1}{4} \times \frac{1}{4} + \frac{1}{4} \times \frac{1}{2} \right]. \]

§6.2. Techniques for efficient computation

- Genotype elimination and allele consolidation

The algorithm for genotype elimination
(A) For each pedigree member, list only those of ordered genotypes compatible with his or her phenotype.

(B) For each nuclear family:

(1) Consider each mother-father genotype pair.

(a) Determine which zygotes can arise from the genotype pair.

(b) If each child in the nuclear family has one or more of these zygote genotypes among his or her current list of genotypes, then save the parental genotypes and any child genotype matching one of the created zygote genotypes.

(c) If any child has none of these zygote genotypes among his or her current list of genotypes, then take no action to save any genotypes.
(2) For each person in the nuclear family, exclude any genotypes not saved during step (1) above.

(C) Repeat part (B) until no more genotypes can be excluded.

An example

The pedigree: 1,2 — parents of 3; 3,4 — parents of 5. 1 (unknown), 2 (O), 3 (A), 4 (unknown), 5 (O).

The procedure:

<table>
<thead>
<tr>
<th>Person</th>
<th>After Applying (A)</th>
<th>After Applying (B) to {3, 4, 5}</th>
<th>After Applying (B) to {1, 2, 3}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All 9 genotypes</td>
<td>All 9 genotypes</td>
<td>{A/A, A/O, O/A, A/B, B/A}</td>
</tr>
<tr>
<td>2</td>
<td>{O/O}</td>
<td>{O/O}</td>
<td>{O/O}</td>
</tr>
<tr>
<td>3</td>
<td>{A/A, A/O, O/A}</td>
<td>{A/O, O/A}</td>
<td>{O/A}</td>
</tr>
<tr>
<td>4</td>
<td>All 9 genotypes</td>
<td>{A/O, O/A, B/O, O/B, O/O}</td>
<td>{A/O, O/A, B/O, O/B, O/O}</td>
</tr>
<tr>
<td>5</td>
<td>{O/O}</td>
<td>{O/O}</td>
<td>{O/O}</td>
</tr>
</tbody>
</table>
Allele consolidation

For a multi-allele locus, all those alleles which are not observed within the pedigree can be consolidated into a lumped allele with appropriate summed frequency.

- *Array transformations and iterated sums*

  **Joint sum vs. iterated sums**

  An example: Computing

  \[
  \sum_{G_1 \in S_1} \sum_{G_2 \in S_2} \sum_{G_3 \in S_3} A(G_1, G_2)B(G_2)C(G_2, G_3),
  \]

  where \( S_j \) is the set of possible \( G_j \) with \( m_j \) elements.

  **A.** Computation by joint sum. Total number of operations equals \( 3m_1m_2m_3 - 1 \).

  **B.** Computation by iterated sums in the order \((3, 2, 1)\):
1. Compute

\[ D(G_2) = \sum_{G_3 \in S_3} C(G_2, G_3), \]

in \( m_2(m_3 - 1) \) additions.

2. Compute

\[ E(G_1) = \sum_{G_2 \in S_2} A(G_1, G_2) B(G_2) D(G_2), \]

in \( 2m_1m_2 \) multiplications and \( m_1(m_2 - 1) \) additions.

3. Compute \( \sum_{G_1 \in S_1} E(G_1) \) in \( m_1 - 1 \) additions.

Total number of operations equals \( m_2(3m_1 + m_3 - 1) - 1 \).

C. Computation by iterated sums in the order (3,1,2): Total number of operations equals \( m_2(m_1 + m_3 + 1) - 1 \).
Greedy algorithm for choosing the order of iterated sums

1. Always sum on that index requiring fewest current arithmetic operations to eliminate.
2. Always multiply arrays pairwise until only two arrays involving the current summation index survive. For example, to compute

\[ E(G_1) = \sum_{G_2 \in S_2} A(G_1, G_2) B(G_2) D(G_2), \]

first compute the array

\[ F(G_2) = B(G_2) D(G_2), \]

then compute

\[ E(G_1) = \sum_{G_2 \in S_2} A(G_1, G_2) F(G_2). \]

3. At each stage of multiplication, always
pick the two arrays that cost the least to multiply.

- **Array factoring**

  - **Factorization of priors**

    Suppose a $m$-locus model is under investigation. $G_i = (G_{i1}, \ldots, G_{im})$. By linkage equilibrium,

    \[
    \text{Prior}(G_i) = \prod_{j=1}^{m} \text{Prior}(G_{ij}).
    \]

  - **Factorization of penetrances**

    If $i$’s phenotype $X_i$ decomposes into separate observations $X_{ij}$ at each locus,

    \[
    \text{Pen}(X_i|G_i) = \prod_{j=1}^{m} \text{Pen}(X_{ij}|G_{ij}).
    \]
Factorization of transmission probabilities

A gamete transmission probability Tran($H_k|G_i$) factors into terms encompassing blocks of loci, with each block delimited by two heterozygous loci in the parent $i$.

Example: $i$ has the only heterozygous at loci $r$ and $s$, $1 < r < s < m$. Then

$$
\begin{align*}
\text{Tran}(H_k|G_i) &= \text{Tran}(H_{k1}, \ldots, H_{kr}|G_{i1}, \ldots, G_{ir}) \\
&\times \text{Tran}(H_{k,r+1}, \ldots, H_{ks}|G_{ir}, \ldots, G_{is}; H_{kr}) \\
&\times \text{Tran}(H_{k,s+1}, \ldots, H_{km}|G_{is}, \ldots, G_{im}; H_{ks}).
\end{align*}
$$
\[ \text{Tran}(H_{k1}, \ldots, H_{kr}|G_{i1}, \ldots, G_{ir}) = \frac{1}{2}. \]

\[
\text{Tran}(H_{k,r+1}, \ldots, H_{ks}|G_{ir}, \ldots, G_{is}; H_{kr}) = \begin{cases} 
\theta_{rs} & \text{for recombination on } [r, s] \\
1 - \theta_{rs} & \text{for non-recombination on } [r, s]. 
\end{cases}
\]

\[ \text{Tran}(H_{k,s+1}, \ldots, H_{km}|G_{is}, \ldots, G_{im}; H_{ks}) = 1. \]

where \( \theta_{rs} \) is the recombination fraction between loci \( r \) and \( s \).

Remark: For array factorization to be useful, the factorization for all possible \( G_i \) must have the same form. This can be realized when person \( i \) is an obligate heterozygote at certain loci.
§6.3. Examples of pedigree analysis

• **Paternity testing**

  Paternity testing confirms or eliminates a putative father as the actual father of a child given the genotypes at a number of loci of the mother, child and the putative father.

  ◦ **Paternity index**

  Let $X^j$ denote the phenotypes of trio of the mother, child and putative father at locus $j$. Consider two pedigrees:
  Pedigree 1 — The trio with the putative father as real father;
  Pedigree 2 — The putative father is substituted by a random male with all phenotypes unknown as the actual father, the putative father is an isolated individual unrelated to the child.
The paternity index is defined as
\[
\frac{\prod_j P(X^j|\text{Ped}_1)}{\prod_j P(X^j|\text{Ped}_2)}.
\]

If \(\alpha\) is the prior probability for the putative father to be the real father and \(\beta\) denotes its posterior probability, then
\[
\frac{\beta}{1 - \beta} = \frac{\alpha \prod_j P(X^j|\text{Ped}_1)}{(1 - \alpha) \prod_j P(X^j|\text{Ped}_2)}.
\]

⊙ **Exclusion probability**

The exclusion probability is the probability that a random male would be excluded by at least one of the tests based on the phenotypes of the mother and child.

Let \(S_j\) be the set of non-excluded genotypes for the random male. The exclusion probabil-
ity is given by
\[ 1 - \prod_j \left[ \sum_{G_j \in S_j} P(G_j) \right]. \]

○ An example

Phenotypes for a paternity test:

<table>
<thead>
<tr>
<th>Person</th>
<th>ABO Phenotype</th>
<th>ADA Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother</td>
<td>AB</td>
<td>1/1</td>
</tr>
<tr>
<td>Child</td>
<td>B</td>
<td>1/2</td>
</tr>
<tr>
<td>Putative Father</td>
<td>B</td>
<td>1/2</td>
</tr>
</tbody>
</table>

Allele frequencies at ABO locus:
\[ p_A = 0.28, \ p_B = 0.06, \ p_O = 0.66. \]

Allele frequencies at ADA locus:
\[ p_1 = 0.934, \ p_2 = 0.066. \]
Computation for ABO locus:

Genotype elimination procedure: Pedigree 1

<table>
<thead>
<tr>
<th>Person</th>
<th>Before Elimination</th>
<th>After Elimination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother</td>
<td>{AB}</td>
<td>{AB}</td>
</tr>
<tr>
<td>Father</td>
<td>{BB, BO}</td>
<td>{BB, BO}</td>
</tr>
<tr>
<td>Child</td>
<td>{B/B, B/O, O/B}</td>
<td>{B/B, B/O}</td>
</tr>
</tbody>
</table>

Genotype elimination procedure: Pedigree 2 (without the observed male).

<table>
<thead>
<tr>
<th>Person</th>
<th>Before Elimination</th>
<th>After Elimination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother</td>
<td>{AB}</td>
<td>{AB}</td>
</tr>
<tr>
<td>Father</td>
<td>All 9 genotypes</td>
<td>{AO, AB, OO}</td>
</tr>
<tr>
<td>Child</td>
<td>{B/B, B/O, O/B}</td>
<td>{B/B, B/O}</td>
</tr>
</tbody>
</table>

Computation of paternity index at ABO locus:

\[
P(ABO|\text{Ped}_1) = P(AB) \times \{P(BB)P(B/B|AB, BB) + P(BO)[P(B/O|AB, BO) + P(B/B|AB, BO)]\}
\]

\[
= 2p_Ap_B[p_B^2 \times \frac{1}{2} + 2p_Bp_O \times \frac{1}{2}]
\]

\[
\]
\[
P(ABO|\text{Ped}_2) \\
= (2p_Bp_O + p_B^2)\{0.00139104 \\
+ P(AB) \times [P(AO)P(B/O|AB, AO) \\
+ P(AB)P(B/B|AB, AB) \\
+ P(OO)P(B/O|AB, OO)]\} \\
= 0.0828(0.00139104 + 0.01070496) = 0.001001549.
\]

Note: the sum within \{\} is the likelihood of Pedigree 2 without the observed male. The paternity index at ABO locus is then
\[
\frac{P(ABO|\text{Ped}_1)}{P(ABO|\text{Ped}_2)} = \frac{0.00139104}{0.001001549} = 1.39.
\]

Exclusion probability at ABO locus:
\[
1 - [P(AO) + P(AB) + P(BB) + P(BO) + P(OO)] \\
= P(AA) = p_A^2 = 0.0784.
\]

Similarly it can be computed that the paternity index at ADA locus is 7.58 and the exclusion probability at ADA locus is 0.872.
Paternity index over both loci:
\[ 1.39 \times 7.58 = 10.5362. \]

Exclusion probability over both loci:
\[ 1 - (1 - 0.0784)(1 - 0.872) = 0.882. \]

• Gamete competition model

Testing of Mendelian segregation ratio using GCM

Nuclear families are divided into different mating types. For each given pair of parental genotypes, there are from one to four possible offspring genotypes. The numbers of offsprings observed in the various genotype categories are used to test whether Mendelian segregation ratio holds.

Traditional \( \chi^2 \) test can be applied if there is no ambiguity in all the genotypes.
In Gamete competition model, each allele $A_i$ is assigned a segregation parameter $\tau_i$. One of the $\tau_i$’s is fixed as 1, the remaining $\tau_i$’s takes value between 0 and $\infty$. The segregation probability is determined as

$$P(A_i|A_i/A_j) = \frac{\tau_i}{\tau_i + \tau_j}.$$ 

Under the Mendelian segregation law, all $\tau_i$’s equal 1. The Mendelian law is tested by the likelihood ratio test.

Likelihood function of a given mating type:

Mating type: $A_1A_2|A_1A_2$.

Offspring type: $A_1A_1$, $A_1A_2$ and $A_2A_2$.

Observation: $n_{11}, n_{12}, n_{22}$.

$$L = \left[ \frac{\tau_1}{\tau_1 + \tau_2} \right]^{2n_{11}} \left[ \frac{\tau_2}{\tau_1 + \tau_2} \right]^{2n_{22}} \left[ \frac{2\tau_1\tau_2}{(\tau_1 + \tau_2)^2} \right]^{n_{12}}$$

$$= \left[ \frac{1}{4} \right]^{n_{11}} \left[ \frac{1}{4} \right]^{n_{22}} \left[ \frac{1}{2} \right]^{n_{12}}, \text{ under HWE.}$$
TDT with gamete competition model

The gamete competition model can be used for TDT in the following way: the gamete competition transmission probabilities are applied to the affecteds.

The following extension of the transmission probabilities allows for the consideration of quantitative traits or covariates:

\[
P(A_i|A_i/A_j, X_k) = \frac{e^{\omega_i X_k}}{e^{\omega_i X_k} + e^{\omega_j X_k}}
\]

\[= \frac{1}{1 + e^{(\omega_j - \omega_i)X_k}},\]

where \(X_k\) is the quantitative trait or covariate of the \(k\)th person.
Risk prediction

Risk prediction is done by calculating conditional probabilities. The probability of risk for a person in a pedigree with uncertain disease status is the conditional probability that the person has disease genotypes given the observed phenotypes of the members of the pedigree.

Let the pedigree consists of \( n \) persons. Let \( S_i \) be the set of possible ordered haplotypes for person \( i \). Suppose the risk calculation concerns person \( n \). Let \( S_n \) be decomposed as \( S_n = S_{n1} \cup S_{n2} \) where \( S_{n1} \) consists of non-disease haplotypes and \( S_{n2} \) consists of disease haplotypes. The risk probability is given by

\[
\frac{\sum_{G_1 \in S_1} \cdots \sum_{G_{n-1} \in S_{n-1}} \sum_{G_n \in S_{n2}} P(G_1, \ldots, G_{n-1}, G_n)}{\sum_{G_1 \in S_1} \cdots \sum_{G_{n-1} \in S_{n-1}} \sum_{G_n \in S_n} P(G_1, \ldots, G_{n-1}, G_n)}.
\]
Risk prediction for a pedigree segregating myotonic dystrophy

The pedigree data:

Sets of possible genotypes given the phenotypes:

\[ S_1 = \{ D^+D^-, D^+D^-, D^+D^-, S^+S^+, S^+S^-, S^-S^+, D^-D^+, D^-D^+, D^-D^+ \} \]

\[ S_2 = \{ D^-D^-, D^-D^-, D^-D^-, S^+S^+ , S^+S^- , S^-S^+ \} \]

\[ S_3 = \{ D^-D^+, D^-D^+, D^-D^+, S^+S^+ , S^+S^- , S^-S^+ \} \quad S_4 = S_3 \]

\[ S_5 = \{ D^-D^- \} \quad S_6 = \{ D^+D^- \} \]

\[ S_7 = \{ D^-D^- \} = S_{71} \cup S_{72} \]

24
The risk probability calculation:

\[ P_1 = \sum P(G_1)P(G_2)P(G_5)\text{Tran}(G_3|G_1, G_2) \]
\[ \text{Tran}(G_4|G_1, G_2)\text{Tran}(G_6|G_4, G_5)\text{Tran}(G_71|G_4, G_5), \]

\[ P_2 = \sum P(G_1)P(G_2)P(G_5)\text{Tran}(G_3|G_1, G_2) \]
\[ \text{Tran}(G_4|G_1, G_2)\text{Tran}(G_6|G_4, G_5)\text{Tran}(G_72|G_4, G_5), \]

\[ P_R = \frac{P_2}{P_1 + P_2}. \]

Let

\[ A(G_1, G_2) = \sum_{G_3\in S_3} \text{Tran}(G_3|G_1, G_2), \]

\[ B(G_1, G_2) = \sum_{G_4\in S_4} \text{Tran}(G_4|G_1, G_2)\text{Tran}(G_6|G_4, G_5) \]
\[ \times \text{Tran}(G_71|G_4, G_5), \]

\[ C(G_1) = \sum_{G_2\in S_2} P(G_2)A(G_1, G_2)B(G_1, G_2). \]

Then

\[ P_1 = P(G_5) \sum_{G_1\in S_1} P(G_1)C(G_1). \]

Similarly, \( P_2 \) can be computed.
Some details: Take $G_4 = \frac{D^-|D^+}{S^+|S^-}$ and $G_5 = \frac{D^-|D^-}{S^-|S^-}$.

Then

$$\text{Tran}(G_{71}|G_4, G_5)$$

$$= \text{Tran}(\frac{D^-}{S^-} | \frac{D^-|D^+}{S^+|S^-}) \text{Tran}(\frac{D^-}{S^-} | \frac{D^-|D^-}{S^-|S^-})$$

$$= \frac{\theta}{2}.$$ 

$$\text{Tran}(G_{72}|G_4, G_5)$$

$$= \frac{1 - \theta}{2}.$$ 

$$P(G_5) = p_D^2 p_S^2.$$ 

- Lod scores and location scores

Lod scores

A lod score $Z(\theta)$ is the log of the ratio of two likelihoods of the recombination fraction between a disease locus and a marker based
on the observed phenotypes of the disease and marker loci.

\[ Z(\theta) = \log_{10} \frac{L(\theta)}{L(1/2)}. \]

where \( L(\theta) \) is the likelihood function and \( \theta \) is the recombination fraction.

Lod scores are used by the geneticists in the following way: Let \( \hat{\theta} \) be the MLE of \( \theta \). If \( Z(\hat{\theta}) \geq 3 \), it is claimed that there is linkage between the disease locus and the marker. If \( Z(\hat{\theta}) \leq -2 \), it is claimed that there is no linkage between the disease locus and the marker.

**Remark:** Lod scores use only one marker for the linkage analysis. The major use of the lod scores is to identify the chromosome region of the disease locus.
Location scores

A Location score is the log of the ratio of the joint likelihoods of the phenotypes of the disease locus and a number of marker loci as a function of the position of the disease locus.

In the computation of location scores, the recombination fraction between any two loci is converted from the genetic distance between them. The genetic distance is measured in units of Morgan (or centiMorgan (cM), $1cM = 0.01$ Morgan). The distance between two loci is 1 Morgan if the expected number of crossovers between the two loci is 1. The genetic distance is related to recombination fraction by the Haldane’s map function:

$$r = -\frac{1}{2}\ln(1 - 2\theta), \quad \theta = \frac{1}{2}(1 - e^{-2r}),$$
where \( r \) is in units of Morgan.

The location score:

\[
Z(d) = \log_{10} \frac{L_{\text{loc}}(d)}{L_{\text{loc}}(\infty)},
\]

where \( L_{\text{loc}}(d) \) is the likelihood at the putative position \( d \), and \( d \) is the genetic distance relative to a fixed position. Note that \( L_{\text{loc}}(\infty) = L(1/2) \).

**Remark:** Suppose there are \( m \) markers involved in the location scores. The markers are ordered with known genetic distance between any two of them. For a fixed position \( d \), the disease locus is considered at position \( d \). The computation of a location score reduces to the computation of the joint likelihood of the phenotypes of the \( m + 1 \) loci with known recombination fractions.
Example: an Episodic Ataxia pedigree with reconstructed haplotypes

The pedigree data:
The pedigree data reduced for the computation of lod score based on marker D12S372:
Computation of lod scores:

Contribution to \( L(\theta) \) of each generation when 2001 has genotype 2\(d\)\(3D\):

- 2nd: \( \left[ \frac{1}{2}(1 - \theta) \right]^5 \)
- 3rd: \( \left[ \frac{1}{4}(1 - \theta) \right]^{3+2+1}\left[ \frac{1}{2}(1 - \theta) \right]^2 \)
- 4th: \( \left[ \frac{1}{4} \right]^3 \frac{1}{4}(1 - \theta) \frac{1}{4} \theta \left[ \frac{1}{4}(1 - \theta) \right]^2 \)

The contributions are the same when 2001 has genotype 2\(D\)\(3d\) except the 2nd generation which has contribution \( \left( \frac{1}{2} \theta \right)^5 \).

The likelihood:

\[
L(\theta) = \frac{1}{2}\left( \left[ \frac{1}{2}(1 - \theta) \right]^5 + \frac{1}{2}\theta^5 \right) \left( \frac{1}{4} \right)^{14} (1 - \theta)^{11} \theta.
\]

\[
L(1/2) = c(1/2)^{18}.
\]

The lod score:

\[
Z(\theta) = \log_{10}(2^{18}\left[ (1 - \theta)^5 + \theta^5 \right] (1 - \theta)^{11} \theta).
\]
An illustration for the computation of location scores based on markers: S91, S100, CACNL1A1, S372, pY2/1 and S93:

**Distances between adjacent markers**

\[ S_{91} \xrightarrow{1cM} S_{100} \xrightarrow{1cM} CACNL1A1 \xrightarrow{3cM} S_{372} \xrightarrow{3cM} pY2/1 \xrightarrow{4cM} S_{93} \]

**Recombination fractions between adjacent markers**

\[ S_{91} \xrightarrow{.0099} S_{100} \xrightarrow{.0099} CACNL1A1 \xrightarrow{.0291} S_{372} \xrightarrow{.0291} pY2/1 \xrightarrow{.0384} S_{93} \]

Let the distance \( d \) be measured relative to the position of S91. Take \( d = 7 \text{ cM} \). Then the putative disease locus is between S371 and pY2/1 and is 2 cM from S371.
Recombination fractions between disease locus and flanking markers

$S372 \rightarrow D$ \hspace{0.5cm} Disease locus \hspace{0.5cm} pY2/1

The location score is the product of 20 transmission probabilities. An example of the computation of these transmission probabilities is as follows: given that the genotype of 2001 at disease locus is $D/d$, then

$$\text{Tran}(G_{1000}|G_{2002}, G_{2001})$$

$$= \text{Tran}([3, 6, 10, 3, D, 2, 4]| [2, 5, 7, 2, D, 1, 4]/[3, 6, 10, 3, d, 2, 4]) \times \text{Tran}([1, 4, 10, 2, d, 1, 7]| [1, 4, 10, 2, d, 1, 7]/[1, 1, 9, 2, d, 5, 4])$$

$$= \frac{1}{2}(0.9901)^2 (0.9719) \frac{1}{2}[1 + e^{-0.02(2)}] \frac{1}{2}[1 + e^{-0.02(1)}]$$

$$\times \frac{1}{2}(0.9901)[1 - \frac{1}{2}(1 + e^{-0.02(6)})](0.9616).$$