Modeling IBD for Pairs of Relatives

Biostatistics 666
Lecture 16
Previously ...

- Linkage Analysis of Relative Pairs

- IBS Methods
  - Compare observed and expected sharing

- IBD Methods
  - Account for frequency of shared alleles
  - Provide estimates of IBD sharing at each locus
IBS Linkage Test

\[ \chi^2_{2df} = \sum_{i} \frac{(N_{IBS=i} - E[N_{IBS=i}])^2}{E(N_{IBS=i})} \]

- \( E(N_{IBS=i}) \) depends on \( N \) and allele frequencies
- Bishop and Williamson (1990)
Likelihood for Sibpair Data

\[ L_i \propto \sum_{j=0}^{2} P(\text{IBD} = j \mid ASP)P(\text{Genotypes} \mid \text{IBD} = j) \propto \sum_{j=0}^{2} z_j w_{ij} \]

Risch (1990) defines

\[ w_{ij} \propto P(\text{Genotypes}_i \mid \text{IBD} = j) \]

\[ z_i = P(\text{IBD} = i \mid \text{affected relative pair}) \]
MLS Statistic of Risch (1990)

\[ L(z_0, z_1, z_2) = \prod_i \sum_j z_j w_{ij} \]

\[ LOD = \log_{10} \prod_i \frac{\hat{z}_0 w_{i0} + \hat{z}_1 w_{i1} + \hat{z}_2 w_{i2}}{\frac{1}{4} w_{i0} + \frac{1}{2} w_{i1} + \frac{1}{4} w_{i2}} = \frac{\chi^2}{2 \ln 10} \]

The MLS statistic is the LOD evaluated at the MLEs of \( z_0, z_1, z_2 \)
The \( \hat{z}_0, \hat{z}_1, \hat{z}_2 \) can be estimated using an E-M algorithm
Today ...

- Predicting IBD for affected relative pairs
  - Modeling marginal effect of a single locus
  - Relative risk ratio ($\lambda_R$)

- The Possible Triangle for Sibling Pairs
  - Plausible IBD values for affected siblings
  - Refinement of the model of Risch (1990)
Single Locus Model

1. Allele frequencies
   • For normal and susceptibility alleles

2. Penetrances
   • Probability of disease for each genotype

   ● Useful in exploring behavior of linkage tests
     • A simplification of reality

   ● Ignore effect of other loci and environment
Penetrance

- $f_{ij} = P(\text{Affected} \mid G = ij)$
- Probability someone with genotype $ij$ is affected
- Models the marginal effect of each locus
Using Penetrances

- Allele frequency $p$
- Genotype penetrances $f_{11}, f_{12}, f_{22}$
- Probability of genotype given disease
  - $P(G = ij | D) =$
- Prevalence
  - $K =$
Pairs of Individuals

- A genetic model can predict probability of sampling different affected relative pairs

- We will consider some simple cases:
  - Unrelated individuals
  - Parent-offspring pairs
  - Monozygotic twins

- What do the pairs above have in common?
What we might expect ...

- Related individuals have similar genotypes

- For a genetic disease...

- Probability that two relatives are both affected must be greater or equal to the probability that two randomly sampled unrelated individuals are affected
Relative Risk and Prevalence

- In relation to affected proband, define
  - $K_R$ prevalence in relatives of type R
  - $\lambda_R = K_R/K$ increase in risk for relatives of type R
  - $\lambda_R$ is a measure of the overall effect of a locus
    - Useful for predicting power of linkage studies
Unrelated Individuals

- Probability of affected pair

\[ P(a \text{ and } b \text{ affected}) = P(a \text{ affected})P(b \text{ affected}) \]
\[ = P(\text{affected})^2 \]
\[ = \left[ p^2 f_{11} + 2p(1-p)f_{12} + (1-p)^2 f_{22} \right] \]
\[ = K^2 \]

- For any two related individuals, probability that both are affected should be greater
Monozygotic Twins

- Probability of affected pair

\[ P(MZ \text{ pair affected}) = \sum_G P(G)P(a \text{ affected} | G)P(b \text{ affected} | G) \]

\[ = p^2 f_{11}^2 + 2p(1-p)f_{12}^2 + (1-p)^2 f_{22}^2 \]

\[ = K_{MZ} K \]

\[ = \lambda_{MZ} KK \]

- \( \lambda_{MZ} \) will be greater than for any other relationship
### Probability for Genotype Pairs

<table>
<thead>
<tr>
<th>Parent</th>
<th>A1A1</th>
<th>A1A2</th>
<th>A2A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1A1</td>
<td>$p_1^3$</td>
<td>$p_1^2p_2$</td>
<td>0</td>
</tr>
<tr>
<td>A1A2</td>
<td>$p_1^2p_2$</td>
<td>$p_1p_2$</td>
<td>$p_1p_2^2$</td>
</tr>
<tr>
<td>A2A2</td>
<td>0</td>
<td>$p_1p_2^2$</td>
<td>$p_2^3$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$p_1^2$</th>
<th>$2p_1p_2$</th>
<th>$p_2^2$</th>
<th>N pairs</th>
</tr>
</thead>
</table>

**Note:** This table represents the probability distribution for different genotype pairs based on the alleles $A_1$ and $A_2$. The values $p_1$ and $p_2$ represent the frequencies of the alleles $A_1$ and $A_2$, respectively.
# Probability of Genotype Pairs and Being Affected

<table>
<thead>
<tr>
<th>Parent</th>
<th>Child</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A₁A₁</td>
<td>A₁A₂</td>
<td>A₂A₂</td>
</tr>
<tr>
<td>A₁A₁</td>
<td>$p_1^3 f_{11}^2$</td>
<td>$p_1^2 p_2 f_{12} f_{11}$</td>
<td>0</td>
</tr>
<tr>
<td>A₁A₂</td>
<td>$p_1^2 p_2 f_{11} f_{12}$</td>
<td>$p_1 p_2 f_{12}^2$</td>
<td>$p_1 p_2^2 f_{12} f_{22}$</td>
</tr>
<tr>
<td>A₂A₂</td>
<td>0</td>
<td>$p_1 p_2^2 f_{12} f_{22}$</td>
<td>$p_2^3 f_{22}^2$</td>
</tr>
</tbody>
</table>

N pairs
**Parent Offspring Pairs**

- **Probability of Affected Pair**
  
  \[ P = P(\text{parent and child affected}) \]

  \[
  = \sum_{G_p} \sum_{G_o} P(G_p, G_o) f_{G_p} f_{G_o} \\
  = \sum_{i} \sum_{j} \sum_{k} P(i, j, k) f_{ij} f_{ik} \\
  = p^3 f_{11}^2 + (1-p)^3 f_{22}^2 + p(1-p)f_{12}^2 + 2p^2(1-p)f_{11}f_{12} + 2p(1-p)^2 f_{22}f_{12} \\
  = KK_o \\
  = \lambda_o KK
  \]

- \(\lambda\) will be lower for other unilineal relationships
- \(\lambda_o\) will be between 1.0 and \(\lambda_{MZ}\)
Point of Situation

- Probabilities of affected pairs for
  - Unrelated Individuals
  - Monozygotic Twins
  - Parent-Offspring Pairs

- Each of these shares a fixed number of alleles IBD ...
For a single locus model...

\[ \lambda_{IBD=2} = \lambda_{MZ} \]
\[ \lambda_{IBD=1} = \lambda_{O} \]
\[ \lambda_{IBD=0} = 1 \]

- Model ignores contribution of other genes and environment

\[ K_{IBD=2} = K_{MZ} \]
\[ K_{IBD=1} = K_{O} \]
\[ K_{IBD=0} = 1 \]

- Simple model that allows for useful predictions
  - Risk to half-siblings
  - Risk to cousins
  - Risk to siblings
Affected Half-Siblings

- IBD sharing
  - 0 alleles with probability 50%
  - 1 allele with probability 50%

- This gives ...

\[
\lambda_H = \frac{1}{2} \lambda_O + \frac{1}{2} = \frac{1}{2} (\lambda_O + 1) \\
K_H = \frac{1}{2} K_O + \frac{1}{2} K = \frac{1}{2} (K_O + K)
\]
Uni-lineal Relationships

\[
\lambda_R = P(\text{IBD} = 1 \mid R)\lambda_o + P(\text{IBD} = 0 \mid R)
\]

\[
K_R = P(\text{IBD} = 1 \mid R)K_o + P(\text{IBD} = 0 \mid R)K
\]

\[
P(\text{IBD} = 1) \text{ decreases 50\% with increasing degree of relationship}
\]

\[
(\lambda_R - 1) \text{ also decreases 50\% with increasing degree of unilineal relationship}
\]
Affected Sibpairs

- IBD sharing …
  - 0 alleles with probability 25%
  - 1 alleles with probability 50%
  - 2 alleles with probability 25%
- This gives …

\[
\lambda_S = \frac{1}{4} \lambda_{MZ} + \frac{1}{2} \lambda_O + \frac{1}{4} = \frac{1}{4}(\lambda_{MZ} + 2\lambda_O + 1)
\]

which implies

\[
\lambda_{MZ} = 4\lambda_S - 2\lambda_O - 1
\]
### Examples: Full Penetrance

#### Recessive

<table>
<thead>
<tr>
<th>p</th>
<th>f11</th>
<th>f12</th>
<th>f22</th>
<th>K</th>
<th>MZ</th>
<th>Offspring</th>
<th>Sibling</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.000001</td>
<td>1000000</td>
<td>1000</td>
<td>250500</td>
</tr>
<tr>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.0001</td>
<td>10000</td>
<td>100</td>
<td>2550</td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.01</td>
<td>100</td>
<td>10</td>
<td>30</td>
</tr>
</tbody>
</table>

#### Dominant

<table>
<thead>
<tr>
<th>p</th>
<th>f11</th>
<th>f12</th>
<th>f22</th>
<th>K</th>
<th>MZ</th>
<th>Offspring</th>
<th>Sibling</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.002</td>
<td>500.25</td>
<td>250.50</td>
<td>250.56</td>
</tr>
<tr>
<td>0.01</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.02</td>
<td>50.25</td>
<td>25.50</td>
<td>25.56</td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.19</td>
<td>5.26</td>
<td>3.02</td>
<td>3.08</td>
</tr>
</tbody>
</table>
Examples: Incomplete Penetrance

<table>
<thead>
<tr>
<th>p</th>
<th>f11</th>
<th>f12</th>
<th>f22</th>
<th>K</th>
<th>MZ</th>
<th>Offspring</th>
<th>Sibling</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>1</td>
<td>0.001</td>
<td>2.0</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td>0.01</td>
<td>0.001</td>
<td>0.001</td>
<td>1</td>
<td>0.001</td>
<td>83.5</td>
<td>1.8</td>
<td>22.0</td>
</tr>
<tr>
<td>0.1</td>
<td>0.001</td>
<td>0.001</td>
<td>1</td>
<td>0.01</td>
<td>82.8</td>
<td>8.4</td>
<td>25.2</td>
</tr>
</tbody>
</table>

**Dominant**

<table>
<thead>
<tr>
<th>p</th>
<th>f11</th>
<th>f12</th>
<th>f22</th>
<th>K</th>
<th>MZ</th>
<th>Offspring</th>
<th>Sibling</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.001</td>
<td>1</td>
<td>1</td>
<td>0.003</td>
<td>223</td>
<td>112</td>
<td>112</td>
</tr>
<tr>
<td>0.01</td>
<td>0.001</td>
<td>1</td>
<td>1</td>
<td>0.02</td>
<td>46</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>0.1</td>
<td>0.001</td>
<td>1</td>
<td>1</td>
<td>0.19</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
### Examples: Small Effects

<table>
<thead>
<tr>
<th>p</th>
<th>f11</th>
<th>f12</th>
<th>f22</th>
<th>K</th>
<th>MZ</th>
<th>Offspring</th>
<th>Sibling</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
<td>0.012</td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>0.1</td>
<td>0.01</td>
<td>0.08</td>
<td>0.16</td>
<td>0.024</td>
<td>2.6</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>0.1</td>
<td>0.02</td>
<td>0.16</td>
<td>0.32</td>
<td>0.048</td>
<td>2.6</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>0.2</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
<td>0.014</td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>0.2</td>
<td>0.01</td>
<td>0.08</td>
<td>0.16</td>
<td>0.038</td>
<td>2.1</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>0.2</td>
<td>0.02</td>
<td>0.16</td>
<td>0.32</td>
<td>0.08</td>
<td>2.1</td>
<td>1.6</td>
<td>1.6</td>
</tr>
</tbody>
</table>

#### Smaller Effects

<table>
<thead>
<tr>
<th>Lambdas</th>
<th>MZ</th>
<th>Offspring</th>
<th>Sibling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>2.6</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>2.6</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>2.1</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>2.1</td>
<td>1.6</td>
<td>1.6</td>
</tr>
</tbody>
</table>
Multiple susceptibility loci...

- $\lambda$ are upper bound on effect size for one locus
- $\lambda$ decay rapidly for distant relatives
- If genes act multiplicatively, we can multiply marginal $\lambda$ together
Another interpretation...

\[ \lambda_{IBD=2} = \lambda_{MZ} = \frac{P(affected \mid IBD = 2 \text{ with affected relative})}{P(affected)} \]

\[ \lambda_{IBD=1} = \lambda_{O} = \frac{P(affected \mid IBD = 1 \text{ with affected relative})}{P(affected)} \]

\[ \lambda_{IBD=0} = 1 = \frac{P(affected \mid IBD = 0 \text{ with affected relative})}{P(affected)} \]
Bayes' Theorem: Predicting IBD Sharing

\[
P(\text{IBD} = i \mid \text{affected pair}) =
\]

\[
= \frac{P(\text{IBD} = i)P(\text{affected pair} \mid \text{IBD} = i)}{\sum_{j} P(\text{IBD} = j)P(\text{affected pair} \mid \text{IBD} = j)}
\]

\[
= \frac{\lambda_{\text{IBD}=i}}{\sum_{j} P(\text{IBD} = j)\lambda_{\text{IBD}=i}}
\]
Sibpairs
Expected Values for $z_0$, $z_1$, $z_2$

\[ z_0 = 0.25 \frac{1}{\lambda_s} \]
\[ z_1 = 0.50 \frac{\lambda_o}{\lambda_s} \]
\[ z_2 = 0.25 \frac{\lambda_{MZ}}{\lambda_s} \]

$1 \leq \lambda_o \leq \lambda_s \leq \lambda_{MZ}$ for any genetic model
Maximum LOD Score (MLS)

- Powerful test for genetic linkage
- Likelihood model for IBD sharing
  - Accommodates partially informative families
- MLEs for IBD sharing proportions
  - Can be calculated using an E-M algorithm
- Shortcoming:
  - Sharing estimates may be implausible
Possible Triangle

Area covering all possible values for sharing parameters

$z_0 = \frac{1}{4}, z_1 = \frac{1}{2}$
The yellow triangle indicates possible true values for the sharing parameters for any genetic model.

\[ H_0: \quad z_0 = \frac{1}{4}, \quad z_1 = \frac{1}{2} \]
**Intuition**

- **Under the null**
  - True parameter values are \((\frac{1}{4}, \frac{1}{2}, \frac{1}{4})\)
  - Estimates will wobble around this point

- **Under the alternative**
  - True parameter values are within triangle
  - Estimates will wobble around true point
Idea (Holmans, 1993)

- Testing for linkage
  - Do IBD patterns suggest a gene is present?

- Focus on situations where IBD patterns are compatible with a genetic model
  - Restrict maximization to possible triangle
The possible triangle method

1. Estimate $z_0$, $z_1$, $z_2$ without restrictions
2. If estimate of $z_1 > \frac{1}{2}$ then …
   a) Repeat estimation with $z_1 = \frac{1}{2}$
   b) If this gives $z_0 > \frac{1}{4}$ then revert to null (MLS=0)
3. If estimates imply $2z_0 > z_1$ then …
   a) Repeat estimation with $z_1 = 2z_0$
   b) If this gives $z_0 > \frac{1}{4}$ then revert to null (MLS=0)
4. Otherwise, leave estimates unchanged.
Possible Triangle

Holman's Example:

<table>
<thead>
<tr>
<th>IBD</th>
<th>Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
</tr>
</tbody>
</table>

MLS = 4.22 (overall)
MLE = (0.08, 0.60, 0.32)

MLS = 3.35 (triangle)
MLE = (0.10, 0.50, 0.40)
MLS Combined
With Possible Triangle

- Under null, true $z$ is a corner of the triangle
  - Estimates will often lie outside triangle
  - Restriction to the triangle decreases MLS
  - MLS threshold for fixed type I error decreases

- Under alternative, true $z$ is within triangle
  - Estimates will lie outside triangle less often
  - MLS decreases less
  - Overall, power should be increased
Example

- Type I error rate of 0.001

- LOD of 3.0 with unrestricted method
  - Risch (1990)

- LOD of 2.3 with possible triangle constraint
  - Holmans (1993)
  - For some cases, almost doubles power
Recommended Reading

  - Introduces possible triangle constraint
  - Good review of MLS method
Reference

- Risch (1990)

- Recurrence risks for relatives.
- Examines implications of multi-locus models.