Replacing IBS with IBD: The MLS Method

Biostatistics 666
Lecture 14
Scheduling

- Review Session
  - Tuesday, February 22

- Midterm
  - Thursday, February 24

- Reminder:
  - Although it is a good idea to discuss problem sets with colleagues, you should turn in YOUR own work.
Previous Lecture

- Analysis of Affected Relative Pairs
- Test for Increased Sharing at Marker
- Expected Amount of IBS Sharing
Test for Independence

\[ \chi^2_{2df} = \sum_i \frac{\left[ N_{IBS=i} - E(N_{IBS=i}) \right]^2}{E(N_{IBS=i})} \]  
\[ \chi^2_{1df} = \frac{\left[ N_{IBS=0} - E(N_{IBS=0}) \right]^2}{E(N_{IBS=0})} + \frac{\left[ N_{IBS>0} - E(N_{IBS>0}) \right]^2}{E(N_{IBS>0})} \]  

(general test, for sibling pairs)

(grouping often preferable for other relatives)

- Assuming all counts are relatively large
- If counts are small, use binomial or trinomial distribution
Previous Lecture: Expected IBS Sharing

- Calculated probability of IBS for each IBD state

- Probability of IBD state depends on relationship
  - Under the null hypothesis of no linkage

\[ P(\text{IBS} = i \mid R) = \sum_{j=0}^{2} P(\text{IBS} = i \mid \text{IBD} = j)P(\text{IBD} = j \mid R) \]
Results of Bishop and Williamson (1990)

- Effect size, $P(\text{IBD} \mid \text{Affected pair})$
- Number of alleles at marker
- Different relationships
- Recombination fraction
More Alleles Increase Power

Figure 3  Variation in ELOD as a function of $n$, the number of alleles at the marker locus. All alleles are assumed to have frequency $1/n$. This calculation is performed for the grandparent-grandchild relationship with a rare trait allele frequency.
Effect of Recombination Varies According to Relationship

Figure 2  Probability of i.b.d. at a second linked locus conditional on i.b.d. at an index locus, as a function of the recombination fraction $r$ between the loci, for specific genetic relationships. This function is $d_{11}(r)$ in the notation of table 1.
Figure 4  ELOD as a function of $\mu$ for an eight-allele marker system for a range of relationships. The recombination fraction between the trait and marker locus is .1.
With no phenocopies, rare alleles are easier to map

**Figure 1** Probability of i.b.d. at a trait locus for two affected related individuals, as a function of the mode of inheritance of the trait. This figure is computed for the relationships with $\phi = .125$. 
Shortcomings of IBS Method

- All sharing is weighted equally
  - Sharing a rare allele
  - Sharing a common allele
  - Sharing homozygous genotype
  - Sharing heterozygous genotype

- Inefficient.
  - Data contains additional information that is being ignored.
Today

- A likelihood based approach
- Evaluate linkage in fully informative pairs
- An E-M algorithm for practical settings
- MLS method, Risch (1990)
Simple Case

- If IBD could be observed

- Each pair of individuals scored as
  - IBD=0
  - IBD=1
  - IBD=2

- Evaluate likelihood for null and alternative hypothesis
The Model

- Depends on three parameters $z_0$, $z_1$, $z_2$
  - Probability of sharing 0, 1 and 2 alleles IBD

- Under the null, determined by relationship

- Under the alternative, determined by genetic model
Sib Pair Likelihood
(Fully Informative Data)

Under the null hypothesis:

\[ L = \left( \frac{1}{4} \right)^{n_{IBD0}} \left( \frac{1}{2} \right)^{n_{IBD1}} \left( \frac{1}{4} \right)^{n_{IBD2}} \]

Under the alternative hypothesis:

\[ L = \left( z_0 \right)^{n_{IBD0}} \left( z_1 \right)^{n_{IBD1}} \left( z_2 \right)^{n_{IBD2}} \]
Testing for Linkage

- Evaluate likelihood at null hypothesis
- Evaluate likelihood at MLE
- Compare alternatives using likelihood ratio test
Commonly Used Test Statistics

\[
LOD = \log_{10} \frac{L(\hat{z}_0, \hat{z}_1, \hat{z}_2)}{L(z_0 = \frac{1}{4}, z_1 = \frac{1}{2}, z_2 = \frac{1}{4})}
\]

\[
\chi^2 = 2 \ln \frac{L(\hat{z}_0, \hat{z}_1, \hat{z}_2)}{L(z_0 = \frac{1}{4}, z_1 = \frac{1}{2}, z_2 = \frac{1}{4})}
\]

\[
= 2 \ln L(\hat{z}_0, \hat{z}_1, \hat{z}_2) - 2 \ln L(z_0 = \frac{1}{4}, z_1 = \frac{1}{2}, z_2 = \frac{1}{4})
\]
Example

IBD=1

IBD=2
Example

- Assume that 10 sib-pairs are examined
  - 5 share 2 alleles IBD
  - 5 share 1 allele IBD

- Calculate likelihood for null
- Calculate MLEs
- Calculate LOD score
- Evaluate LOD for each pair
In real life...

- Markers are only partially informative
- IBD sharing is equivocal
  - Some uncertainty removed by examining relatives
- Need an alternative likelihood
  - Should allow for partially informative data
Desirable Properties

- Also depends on parameters $z_0, z_1, z_2$
  - Probability of sharing 0, 1 and 2 alleles IBD

- Can incorporate partial information on IBD sharing

- For fully informative data, equivalent to previous likelihood
For A Single Family

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

Risch (1990) defines

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

We only need proportionate \( w_{ij} \).
Likelihood and LOD Score

\[ L(z_0, z_1, z_2) = \prod_{i} \sum_{j} z_j w_{ij} \]

\[ LOD = \log_{10} \prod_{i} \frac{z_0 w_{i0} + z_1 w_{i1} + z_2 w_{i2}}{\frac{1}{4} w_{i0} + \frac{1}{2} w_{i1} + \frac{1}{4} w_{i2}} \]

The MLS statistic is the LOD evaluated at the MLEs of \( z_0, z_1, z_2 \).
Example: Scoring of $w_{ij}$

In this case, only one of the weights is non-zero for each family.
More interesting examples: \( w_{ij} \)

In these cases, multiple weights are non-zero (but equal) for each family.
More interesting examples: $w_{ij}$

In this case, relative weights depend on allele frequency.
How to maximize likelihood?

- If all families are informative
  - Use sample proportions of IBD=0, 1, 2

- If some families are uninformative
  - Use an E-M algorithm
  - At each stage generate complete dataset with fractional counts
  - Iterate until estimates of LOD and z parameters are stable
Assigning Partial Counts in E-M

\[ P(\text{IBD} = j \mid \text{Genotypes}) = \]
\[ = \frac{P(\text{IBD} = j \mid \text{ASP})P(\text{Genotypes} \mid \text{IBD} = j)}{L_i} \]
\[ = \frac{P(\text{IBD} = j \mid \text{ASP})P(\text{Genotypes} \mid \text{IBD} = j)}{\sum_{k=0}^{2} P(\text{IBD} = k \mid \text{ASP})P(\text{Genotypes} \mid \text{IBD} = k)} \]
\[ = \frac{z_j \omega_{ij}}{\sum_{k=0}^{2} z_k \omega_{ik}} \]
Example

Assume a bi-allelic marker where the two alleles have identical frequencies.
Example of E-M Steps

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Equivocal Families</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>z0</td>
<td>z1</td>
<td>z2</td>
</tr>
<tr>
<td>0.250</td>
<td>0.500</td>
<td>0.250</td>
</tr>
<tr>
<td>0.056</td>
<td>0.222</td>
<td>0.722</td>
</tr>
<tr>
<td>0.008</td>
<td>0.066</td>
<td>0.926</td>
</tr>
<tr>
<td>0.001</td>
<td>0.017</td>
<td>0.982</td>
</tr>
<tr>
<td>0.000</td>
<td>0.004</td>
<td>0.996</td>
</tr>
<tr>
<td>0.000</td>
<td>0.001</td>
<td>0.999</td>
</tr>
<tr>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Properties of Pair Analyses Explored by Risch

- Effect of marker informativeness
- Effect of adding relative genotypes
- Size of genetic effect
- Degree of relationship
PIC: Measure of Marker Informativeness

- Probability that alleles of parent can be distinguished in offspring
  - Markers that could track dominant alleles

- Probability that parent is heterozygous and informative in relation to spouse
PIC – Definition

- In general:
  \[ PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n} \sum_{j=i+1}^{n} 2(p_i p_j)^2 \]

- For a equally frequent alleles
  \[ PIC = \frac{a-1}{a} - \frac{a-1}{a^3} \]

- PIC <= Heterozygosity
Some PICs and Heterozygositities

<table>
<thead>
<tr>
<th>Alleles</th>
<th>PIC</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.38</td>
<td>0.50</td>
</tr>
<tr>
<td>3</td>
<td>0.59</td>
<td>0.67</td>
</tr>
<tr>
<td>4</td>
<td>0.70</td>
<td>0.75</td>
</tr>
<tr>
<td>5</td>
<td>0.77</td>
<td>0.80</td>
</tr>
<tr>
<td>8</td>
<td>0.86</td>
<td>0.88</td>
</tr>
<tr>
<td>10</td>
<td>0.89</td>
<td>0.90</td>
</tr>
<tr>
<td>20</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Marker Informativeness

Proportion of LOD Retained

Proportion of Expected MLS

Sibs
2nd Degree
3rd Degree

Marker Informativeness
Marker Informativeness
Gene of Modest Effect ($\lambda_o=3$)
Marker Informativeness
Gene of Modest Effect ($\lambda_o=10$)

Expected LOD Score

Expected MLS

Marker Informativeness

- Sibs
- 2nd Degree
- 3rd Degree
Genotypes of Other Family Members

- Expected LOD score decreases
  - by < 33% if only sib-pairs are typed
  - by < 60% for second degree relatives
  - by < 70% for third degree relatives

- Genotyping effort decreases by
  - by 50% if only sib-pairs are typed
  - by 60% if only second degree relatives typed
  - by 75% if only third degree relatives typed
Quick Comment on Literature

- Greenwood and Schork (2004) suggested that uninformative families could bias MLS.

- However, their results use a poor estimate for MLEs.

- If an E-M algorithm is used, there is no problem.
Today ...

- Describe a likelihood model based on IBD sharing for pairs of individuals
- Model accommodates partially informative families
- Maximum LOD score can be calculated using an E-M algorithm
Recommended Reading

- Risch (1990)
  - Linkage Strategies for Genetically Complex Traits. III. The Effect of Marker Polymorphism on Analysis of Affected Relative Pairs

- Introduces MLS method for linkage analysis
  - Still, one of the best methods for analysis pair data

- Evaluates different sampling strategies
  - Results were later corrected by Risch (1992)
Recommended Reading

- Risch (1992)
  - Corrections to Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs.
  - *Am J Hum Genet* 51:673-675

- Evaluates utility of parental genotype data