IBS Methods for Affected Pairs Linkage

Biostatistics 666
Lecture 12
Course is Almost Half Done!

- All comments are very welcome
  - Lectures
  - Lecture notes
  - Weekly Homework
  - Content
Genetic Mapping

“Compares the inheritance pattern of a trait with the inheritance pattern of chromosomal regions”

Positional Cloning

“Allows one to find where a gene is, without knowing what it is.”
Where are the genes influencing a particular trait?
Intuition for Linkage Analysis

- Millions of variations could potentially be involved
  - Costly to investigate each individually

- Within families, variation is organized into a limited number of haplotypes
  - Sample modest number of markers to determine whether each stretch of chromosome is shared
Tracing Chromosomes

A pedigree with several affected individuals
Tracing Chromosomes

Segregation pattern for chromosome carrying disease alleles
Tracing Chromosomes

Segregation of a specific marker near the disease locus
Today ...

- Linkage analysis with sibling pairs
  - Identity-by-State (IBS) based method

- Find markers that are near disease locus
  - Near means recombination fraction $\theta < \frac{1}{2}$

- Minimalist approach …
Reference for Today...

- Power of IBS Methods for Linkage Analysis


- Recommended Reading
"The availability of a large number of DNA markers has made possible mapping projects with the certainty that if:

(a) a major gene exists for a trait;
(b) the trait is reasonably homogeneous;
(c) there is sufficient family material available;

then a linked marker can be found."
Data for a Linkage Study

- Pedigree
  - Set of individuals of known relationship

- Observed marker genotypes
  - SNPs, VNTRs, microsatellites

- Phenotype data for individuals
Minimalist Approach

- Pedigree
  - Two individuals of known relationship

- Observed Marker Genotypes
  - A single marker

- Phenotypes
  - Both individuals are affected
Allele Sharing Analysis

- Reject random sharing at a particular region
- Less powerful than classic methods
  - When disease model is known
- More robust than classic methods
  - When disease model is unknown
Consider
Autosomal Recessive Locus ...

- For a collection of sibling pairs...

- What patterns of sharing do you expect at the disease locus?

- What patterns of sharing do you expect as you move away from the disease locus?
Historical References

- Penrose (1953) suggested comparing IBD distributions for affected siblings.
  - Possible for highly informative markers (eg. HLA)

- Thomson (1986) suggested discarding partially informative families.

- Lange (1986) proposed using IBS information instead of IBD.
IBS Based Methods

- Sample of affected relative pairs
- Examine a marker of interest
- Count alleles shared for each pair
  - This includes both …
  - Chromosomes that are identical-by-descent
  - Chromosomes that simply carry identical alleles
Examples of IBS States

IBS = 2

IBS = 1

IBS = 0
Examples of IBS States

IBS = 2

IBS = 1

IBS = 0
Evidence for Linkage

- Increased similarity in affected pairs

- Compared to:
  - Unselected pairs
  - Unaffected pairs
  - Discordant pairs
  - Expectations derived from allele frequencies
Test for Independence

\[ \chi^2_{2df} = \sum_i \left( \frac{N_{IBS=i} - E(N_{IBS=i})}{E(N_{IBS=i})} \right)^2 \]  
(general test, for sibling pairs)

\[ \chi^2_{1df} = \left( \frac{N_{IBS=0} - E(N_{IBS=0})}{E(N_{IBS=0})} \right)^2 + \left( \frac{N_{IBS>0} - E(N_{IBS>0})}{E(N_{IBS>0})} \right)^2 \]  
(grouping often preferable for other relatives)

- Assuming all counts are relatively large
- If counts are small, use binomial or trinomial distribution
Modeling IBS Sharing

For any relative pair, calculate:

- Probability of IBD sharing
  - 0, 1 or 2 alleles

- Conditional probability of IBS sharing
  - 0, 1, 2 alleles

- IBS sharing $\geq$ IBD sharing
  - Assumption?
IBD

- The underlying sharing of chromosomes segregating within a family

- Siblings share 0, 1 or 2 alleles
  - Probabilities $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$

- Unilineal relatives share 0 or 1 alleles
  - Probability of sharing is kinship coefficient $\phi$ * 4
# P(Marker Genotype|IBD State)

<table>
<thead>
<tr>
<th>Relative</th>
<th>IBD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>(a,b)</td>
<td>4p_\text{ap_bpcp_d}</td>
</tr>
<tr>
<td>(a,a)</td>
<td>2p_\text{a}^2p_c</td>
</tr>
<tr>
<td>(a,a)</td>
<td>p_\text{a}p_b</td>
</tr>
<tr>
<td>(a,b)</td>
<td>4p_\text{ap_bpc}</td>
</tr>
<tr>
<td>(a,a)</td>
<td>2p_\text{a}^3p_b</td>
</tr>
<tr>
<td>(a,b)</td>
<td>4p_\text{a}^2p_b^2</td>
</tr>
<tr>
<td>(a,a)</td>
<td>p_\text{a}^4</td>
</tr>
</tbody>
</table>

Prior Probability

\[ \frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{4} 

Note: Assuming alleles unordered within genotypes
\[ \textbf{P(\text{IBS} = i \mid \text{IBD} = j)} \]

\[ P(\text{IBS} = 2 \mid \text{IBD} = 0) = 2 \sum_{i \neq j} p_i^2 p_j^2 + \sum_i p_i^4 \]

\[ P(\text{IBS} = 1 \mid \text{IBD} = 0) = 4 \sum_{i \neq j} p_i^2 p_j (1 - p_i - p_j) + 4 \sum_i p_i^3 (1 - p_i) \]

\[ P(\text{IBS} = 0 \mid \text{IBD} = 0) = \sum_{i \neq j} p_i p_j (1 - p_i - p_j)^2 + \sum_i p_i^2 (1 - p_i)^2 \]
\[
P(\text{IBS} = i \mid \text{IBD} = j)
\]

\[
P(\text{IBS} = 2 \mid \text{IBD} = 2) =
\]

\[
P(\text{IBS} = 1 \mid \text{IBD} = 2) =
\]

\[
P(\text{IBS} = 0 \mid \text{IBD} = 2) =
\]

\[
P(\text{IBS} = 2 \mid \text{IBD} = 1) =
\]

\[
P(\text{IBS} = 1 \mid \text{IBD} = 1) =
\]

\[
P(\text{IBS} = 0 \mid \text{IBD} = 1) =
\]
Example, Assuming Equal Allele Frequencies

<table>
<thead>
<tr>
<th></th>
<th>P(ibs=0)</th>
<th>P(ibs=1)</th>
<th>P(ibs=2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 alleles, IBD=0</td>
<td>0.125</td>
<td>0.500</td>
<td>0.375</td>
</tr>
<tr>
<td>2 alleles, IBD=1</td>
<td>0.000</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>3 alleles, IBD=0</td>
<td>0.222</td>
<td>0.592</td>
<td>0.185</td>
</tr>
<tr>
<td>3 alleles, IBD=1</td>
<td>0.000</td>
<td>0.666</td>
<td>0.333</td>
</tr>
</tbody>
</table>
IBS approaches IBD as number of alleles increases.

If linkage is being tested with chi-square test, how does the number of alleles (and marker informativeness) affect these two tests:
- A test of whether $N_{IBS} \geq 1$ increases?
- A test of whether $N_{IBS} > 1$ increases?
<table>
<thead>
<tr>
<th>No. of Alleles</th>
<th>P(ibs=0)</th>
<th>P(ibs=1)</th>
<th>P(ibs=2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.03</td>
<td>.37</td>
<td>.60</td>
</tr>
<tr>
<td>3</td>
<td>.05</td>
<td>.48</td>
<td>.47</td>
</tr>
<tr>
<td>4</td>
<td>.08</td>
<td>.51</td>
<td>.40</td>
</tr>
<tr>
<td>20</td>
<td>.21</td>
<td>.52</td>
<td>.27</td>
</tr>
<tr>
<td>∞</td>
<td>.25</td>
<td>.50</td>
<td>.25</td>
</tr>
</tbody>
</table>

Sibling IBS as a function of allele count, for marker with equally frequent alleles
Results of Bishop and Williamson (1990)

- Effect size, $P(\text{IBS} \mid \text{Affected pair})$
- 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.
Power vs. \( P(\text{IBS} \mid \text{Affected Pair}) \)

**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 $\phi = .125$. 
In general, phenocopies decrease power.

**Table 2**

Average Informativeness for Mapping a Partially Penetrant Dominant Trait with Phenocopies

<table>
<thead>
<tr>
<th>$p$ and $x$</th>
<th>$\mu^a$</th>
<th>Phenocopy Rate</th>
<th>Relative Information Content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.000</td>
<td>.96</td>
<td>.00</td>
<td>100</td>
</tr>
<tr>
<td>.001</td>
<td>.86</td>
<td>.05</td>
<td>98</td>
</tr>
<tr>
<td>.01</td>
<td>.92</td>
<td>.33</td>
<td>81</td>
</tr>
<tr>
<td>.02</td>
<td>.88</td>
<td>.50</td>
<td>61</td>
</tr>
<tr>
<td>.05</td>
<td>.74</td>
<td>.71</td>
<td>23</td>
</tr>
<tr>
<td>.10</td>
<td>.61</td>
<td>.83</td>
<td>5</td>
</tr>
<tr>
<td>.10:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.000</td>
<td>.75</td>
<td>.00</td>
<td>100</td>
</tr>
<tr>
<td>.001</td>
<td>.75</td>
<td>.00</td>
<td>99</td>
</tr>
<tr>
<td>.01</td>
<td>.74</td>
<td>.04</td>
<td>89</td>
</tr>
<tr>
<td>.02</td>
<td>.73</td>
<td>.08</td>
<td>80</td>
</tr>
<tr>
<td>.05</td>
<td>.69</td>
<td>.18</td>
<td>56</td>
</tr>
<tr>
<td>.10</td>
<td>.64</td>
<td>.30</td>
<td>31</td>
</tr>
</tbody>
</table>

*Note.* The recombination fraction is .1, and the marker system has eight equally frequent alleles.

*a* For a grandparent-grandchild affected pair.
Shortcomings of IBS Method

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

- Inefficient.
Recommended Reading


- Good introduction to linkage analysis in affected relative pairs, discusses
  - Marker choice
  - Recombination fraction
  - Disease model
  - Type of relative pair