Allele-sharing methods

Idea: affected relatives will be more likely to share DNA in the region of a gene that contributes to the disease.

"Identical by descent":
2 alleles are IBD if they are copies of a single ancestral allele

\[
\begin{array}{ccc}
\text{IBD = 1} & \text{IBD = 2} & \text{IBD = 0 or 2} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\
\frac{1}{3} & \frac{1}{4} & \frac{1}{3} & \frac{1}{4} & \frac{1}{3} & \frac{1}{4} \\
\end{array}
\]

A priori, siblings share 0, 1, 2 alleles w/ prob. \(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\), respectively.

Example: single & allelic gene; disease allele freq = 0.1
penetrances \(P_0 = 0.01\), \(P_1 = 0.1\), \(P_2 = 0.5\)
at pos’n r.f. 5% from gene

| Siblings’ phenotypes | \(Pr(\text{IBD} = v | \text{aff. status})\) | Ave. no. alleles IBD |
|---------------------|-----------------------------------------|---------------------|
| Both affected       | 0.063 0.495 0.492                     | 1.38                |
| Neither affected    | 0.248 0.500 0.252                     | 1.00                |
| One aff, one not    | 0.368 0.503 0.128                     | 0.76                |
Simpler example to show calc’n

- diazepam disease gene; allele freq = p
  - penetrances f_0 = f_1 = 0, f_2 = 1
  - affected sib pair

Let mt = parents’ “mating type”

\[
P(\text{IBD} = V | \text{both sibs affected}) = \sum_{\text{mt}} P(\text{IBD} = V | \text{mt}, \text{both sibs aff})
\]

\[
= \sum_{\text{mt}} P(\text{mt} | \text{both sibs aff}) P(\text{IBD} = V | \text{mt}, \text{both sibs aff})
\]

| Mating type | \(P(\text{mt})\) | \(P(\text{both sibs aff} | \text{mt})\) | \(P(\text{mt} | \text{both sibs aff})\) |
|-------------|-----------------|-------------------|-----------------|
| dd/dd       | \(p^4\)         | 1                 | \(p^4 / S\)     |
| dd/dd       | \(4p^3(1-p)\)   | \(1/4\)           | \(p^3(1-p) / S\) |
| dd/dd       | \(2p^2(1-p)^2\) | 0                 | 0               |
| dd/dd       | \(4p^2(1-p)^2\) | \(1/16\)          | \(1/16 + p^2(1-p)^2 / S\) |
| dd/dd       | \(4p(1-p)^3\)   | 0                 | 0               |
| ++/++       | \((1-p)^4\)     | 0                 | 0               |

\[
S = p^4 + p^3(1-p) + \frac{1}{4} p^2(1-p)^2
\]

<table>
<thead>
<tr>
<th>Mating type</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>dd/dd</td>
<td>(1/4)</td>
<td>(1/2)</td>
<td>(1/4)</td>
</tr>
<tr>
<td>dd/dd</td>
<td>0</td>
<td>(1/2)</td>
<td>(1/2)</td>
</tr>
<tr>
<td>dd/dd</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(p)</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.002</td>
<td>0.091</td>
<td>0.907</td>
</tr>
<tr>
<td>0.10</td>
<td>0.009</td>
<td>0.165</td>
<td>0.826</td>
</tr>
<tr>
<td>0.25</td>
<td>0.04</td>
<td>0.32</td>
<td>0.64</td>
</tr>
</tbody>
</table>
Holman's triangle

The TBD dist'n \((\beta_0, \Pi_1, \Pi_2)\), with \(\Pi_2 > 0\), \(\Sigma \Pi_i = 1\), may be identified with a point in an equilateral triangle.

For a single gene disorder, the TBD distribution must satisfy \(\Pi_1 < \frac{1}{2}\), \(\Pi_1 > \Pi_0\).
Affected sib pairs: the complete data case

- $n$ affected sib pairs
- IBD sharing at a particular position known exactly
- $n_i = \text{no. sib pairs sharing } i \text{ alleles IBD}$

Compare $(n_0/n, n_1/n, n_2/n)$ to $(1/4, 1/2, 1/4)$

Example: 100 aff. sib pairs, $(n_0, n_1, n_2) = (15, 38, 47)$

1. Mean Test

Let $S = n_1/n + 2n_2/n$

Under $H_0: \pi = (1/4, 1/2, 1/4)$

$E(S|H_0) = 1$ \hspace{1cm} $\text{Var}(S|H_0) = \frac{1}{2n}$

Let $Z = \frac{S - E(S|H_0)}{\text{SD}(S|H_0)} = \frac{S}{\sqrt{2n}} (S - 1)$

Example: $S = 1.32$, $Z = 4.53$, $\text{LOD} = 4.45$

$P\text{-value} = 6 \times 10^{-6}$

2. Proportion Test

$S' = n_2/n$ \hspace{1cm} $E(S'|H_0) = 1/4$ \hspace{1cm} $\text{Var}(S'|H_0) = \frac{3}{16n}$

$Z' = \ldots = \frac{S'/n - 1}{\sqrt{\frac{3}{16}}} (4S' - 1)$

Example: $S' = 0.47$, $Z' = 5.08$, $\text{LOD}' = 5.61$

$P\text{-value} = 4 \times 10^{-7}$
3. $x^2$ test

$$X^2 = \sum_i \frac{(n_i - n \pi_{oi})^2}{n \pi_{oi}}$$

$\pi_{oi}$: null IBD model

Example: $X^2 = 26.2$  \quad $LOD = 5.70$

$p$-value = $2 \times 10^{-6}$

4. Likelihood ratio test

$$l(\pi) = \log P_0 + n_1 \log \pi_1 + n_2 \log \pi_2$$

$$LRT = 2 \left\{ \sum_i n_i \log(n_i/n) - \frac{\pi}{2} n \log \pi_{oi} \right\}$$

Example: $LRT = 23.2$  \quad $LOD = 5.03$

$p$-value = $9 \times 10^{-6}$

Note: we might restrict alternatives to the "possible triangle"

**INCOMPLETE DATA**

- We seldom know the alleles shared IBD exactly
- Calculate, for sib pair $i$
  $$p_{ij} = \text{Pr(sib pair } i \text{ has IBD } = j \mid \text{marker data})$$
- Means test: use $\sum_i p_{ij}$ in place of $n_i$

Problem: The denominator, $\sqrt{n/2}$, is correct for perfect IBD information, but is too large in the case of incomplete data $\Rightarrow$ conservative

- Alternatives: computer simulation, like methodode 1... (4)
General pedigrees (e.g., larger sibships)

\[
\begin{array}{c}
\Box \quad \Box \\
1/2 \quad 3/4 \\
\Box \quad \Box \\
3/6 \quad (v_1, v_2) \\
\Box \quad \Box \\
(v_3, v_4) \quad 7/8 \\
\Box \quad \Box \\
(v_5, v_6) \\
\end{array}
\]

Mom's allele: \( v \) = \( (v_1, v_2), \ldots, (v_{12}, v_{12}) \)

dad's allele: \( v \) = \( (v_1, v_2), \ldots, (v_{12}, v_{12}) \)

\[ = \text{inheritance vector} \]

- indicates which parental alleles were transmitted

**Score Function** \( S(v) \) measures allele sharing

e.g. \( S_{\text{pairs}}(v) = \sum_{\text{pairs of \text{aff. rel.}}} \# \text{alleles shared IBD} \)

\[
S_{\text{all}}(v) = \frac{1}{2^a} \sum_h \left\{ \frac{2^f}{\prod_{i=1}^{h-1} b_i(h)!} \right\}
\]

where \( a = \# \text{aff. relatives} \)

\( h = \) pick one allele from each affected relative

\( b_i(h) = \# \text{times the } i\text{th founder allele appears in } h \)

**Example:** aff. sib trio

\[
\begin{array}{c}
\Box \quad \Box \\
1/2 \quad 3/4 \\
\Box \quad \Box \quad \Box \\
1/3 \quad 1/3 \quad 2/3 \\
\end{array}
\]

\( S_{\text{pairs}} = 2 + 1 + 1 = 4 \)

\( S_{\text{all}} = \frac{1}{2^a} \left\{ 2 \cdot 1 \cdot 1 + 2 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 + 2 \cdot 1 \cdot 1 + 1 + 2 + 2 + 6 \right\} = \frac{1}{2^a} \left\{ 1 + 6 + 1 + 1 + 2 + 2 + 2 + 6 \right\} = \frac{1}{2^a} \left\{ 16 \right\} = 8 \)
Normalized Score \( z(y) = \frac{S(y) - \mu}{\sigma} \)

where \( \mu = \mathbb{E}[S(y)|H_0] \), \( \sigma = \text{SD}[S(y)|H_0] \)

\( H_0: \) no linkage \( \Rightarrow \) \( v's \) equally likely

E.g. 3 affected sibs

Spairs : \( \mu = 3 \) \( \sigma = \sqrt{3/2} \div 1.22 \)

Sall : \( \mu = \frac{15}{8} \div 1.875 \) \( \sigma = \frac{3}{8} \sqrt{2} \div 0.459 \)

For prev. example

Spairs = 4 \( \rightarrow \) \( z = 0.817 \)

Sall = 9/4 \( \rightarrow \) \( z = 0.817 \)

(For 3 alt. sibs, Spairs \( \neq \) Sall.

More generally they differ, and

Sall better for add'Ve (dominant model,

Spairs better for recessive model)

Combining families

calc. normalized score for ea. family: \( z_i = \frac{S_i - \bar{S_i}}{\sigma_i} \)

combine families w/ \( \gamma_i \)'s \( \gamma_i \geq 0 \)

\( z = \frac{\sum \gamma_i z_i}{\sqrt{\sum \gamma_i^2}} \)

choices of \( \gamma \):

\( \equiv 1 \) for all families

\( \equiv \) no. sib pairs

\( \equiv \sigma_i \) (i.e., combine \( \gamma's \), then normalize)

Incomplete data

In place of \( S \), use \( \bar{S} = \sum y S(y) p(y | \text{marker data}) \)

\( \rightarrow \) really need \( \sigma \)'s to take account of uncertainty in \( v's \).