Mean test

The test based on the statistic

$$T_{\text{mean}}(n_2, n_1, n_0) = \frac{2n_2 + n_1}{2n}$$

is called the *mean test*.

Since

$$2 \cdot n \cdot T_{\text{mean}}(n_2, n_1, n_0) = 2n_2 + n_1 = \sum_{j=1}^{n} \text{IBD}_j$$

with IBD_j denoting the number of alleles shared ibd by the children of family j and since IBD_j is binomial distributed Bin$(2, 1/2)$ under the null hypothesis of no linkage, it follows that $2 \cdot n \cdot T_{\text{mean}}(n_2, n_1, n_0)$ is binomial distributed Bin$(2n, 1/2)$ under the null hypothesis of no linkage.

Nonparametric linkage (NPL) score

Often, the null distribution of the statistic T_{mean} is approximated by a normal distribution (c.f. S/22):

Since

$$\begin{aligned} \mathrm{E}_{H_0}(T_{\text{mean}}) &= \frac{1}{2n} \cdot \mathrm{E}_{H_0}(2 \cdot n \cdot T_{\text{mean}}) = 1/2, \\ \mathrm{Var}_{H_0}(T_{\text{mean}}) &= \frac{1}{4n^2} \cdot \mathrm{Var}_{H_0}(2 \cdot n \cdot T_{\text{mean}}) = 1/(8 \cdot n) \end{aligned}$$

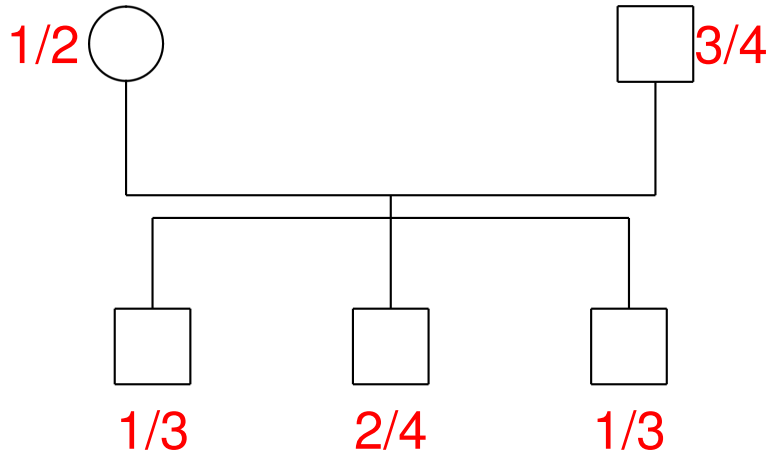
it follows that the null distribution of

$$T_S = \frac{T_{\text{mean}} - \mathrm{E}_{H_0}(T_{\text{mean}})}{\sqrt{\mathrm{Var}_{H_0}(T_{\text{mean}})}} = \sqrt{\frac{2}{n}} \cdot (n_2 - n_0)$$

is approximated by a normal distribution.

Name: Nonparametric linkage (NPL) score

Mean test: More than two siblings



$$S_{12} = \text{IBD(1st child, 2nd child)} = 0$$

$$S_{13} = \text{IBD(1st child, 3rd child)} = 2$$

$$S_{23} = \text{IBD(2nd child, 3rd child)} = 0$$

Mean test: More than two siblings

For sibships with $k > 2$ affected children, the IBD scores $(S_{ij})_{1 \le i < j \le k}$ are not independent. However, it can be shown that the IBD scores $(S_{ij})_{1 \le i < j \le k}$ are uncorrelated under H_0, which is sufficient (c.f. P/32) for

$$\mathrm{Var}_{H_0}\Big(\sum_{1 \le i < j \le k} S_{ij}\Big) = \sum_{1 \le i < j \le k} \mathrm{Var}_{H_0}(S_{ij}).$$

Therefore, the NPL score is easily generalized to arbitrary sibship sizes, i.e., in case of a family with k affected children, IBD scores for all $\binom{k}{2}$ pairs of affected children are calculated and treated as if they would stem from different families.

Name: S_{pairs} statistic (e.g. GENEHUNTER)

Optimal test against a simple alternative

Suppose that the disease model and the recombination fraction θ between disease locus and marker locus would be known. Then, the corresponding distribution $(z_2^\star, z_1^\star, z_0^\star)$ of IBD scores in affected sib pairs can be calculated (c.f. NPL/6). Now, consider the test problem consisting of two simple hypotheses

$$H_0 : (z_2, z_1, z_0) = (1/4, 1/2, 1/4) \text{ vs. } H_1 : (z_2, z_1, z_0) = (z_2^\star, z_1^\star, z_0^\star)$$

Optimal test against a simple alternative

Neyman-Pearson Lemma (c.f. S/17):

$$\frac{L(n_2, n_1, n_0 \mid (z_2^\star, z_1^\star, z_0^\star))}{L(n_2, n_1, n_0 \mid (1/4, 1/2, 1/4))}$$

$$= \frac{(z_2^\star)^{n_2} \cdot (z_1^\star)^{n_1} \cdot (z_0^\star)^{n_0}}{(1/4)^{n_2} \cdot (1/2)^{n_1} \cdot (1/4)^{n_0}}$$

$$= (4z_0^\star)^n \cdot \left(\frac{z_2^\star}{z_0^\star}\right)^{n_2} \cdot \left(\frac{z_1^\star}{2z_0^\star}\right)^{n_1}$$

$$= (4z_0^\star)^n \cdot \exp\left(\underbrace{\ln(z_2^\star/z_0^\star)}_{>0} \cdot n_2 + \underbrace{\ln(z_1^\star/(2z_0^\star))}_{\geq 0} \cdot n_1\right)$$

$\Rightarrow$ Optimal test is based on the statistic

$$T(n_2, n_1, n_0) = n_2 + \underbrace{(\ln(z_1^\star/(2z_0^\star))/\ln(z_2^\star/z_0^\star))}_{=:w} \cdot n_1$$

Properties of the mean test

Mean test for affected sib pairs

- is optimal against alternatives induced by single locus disease models with $f_1^2 = f_2 \cdot f_0$.
- is only slightly less powerful than the most powerful test for most disease models.
- is equivalent to parametric linkage analysis in case that a disease model with $f_1^2 = f_2 \cdot f_0$ is used for the calculation of the maximum lod score.

$\Rightarrow$

Samples consisting of affected sib pairs are quite robust against misspecification of the disease model.

Linkage analysis and imprinting

Genomic imprinting is the phenomenon in which there is differential expression of a gene on whether is was maternally or paternally inherited. Possible causes of imprinting include chemical modifications of the DNA (no base changes!) such as methylation, which can turn off gene expression through preventing RNA polymerase from transcribing a gene.

Paternal imprinting means that an allele inherited from the father is not expressed in offspring.

Maternal imprinting means that an allele inherited from the mother is not expressed in offspring.

Modelling of imprinting: parametric linkage analysis

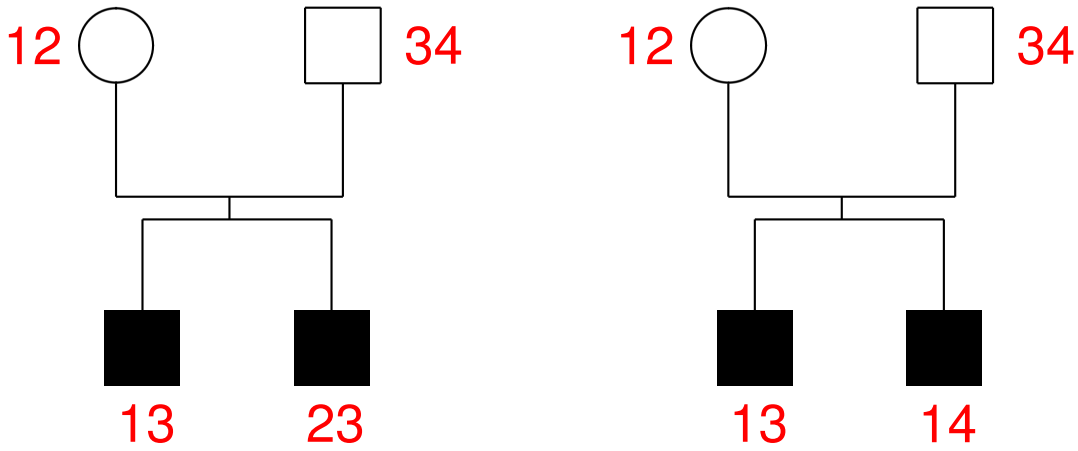
Replace $f_1 = P(\text{“affected”} \mid Dd)$ with

$$f_F = P(\text{“affected”} \mid Dd, D \text{ inherited from the father})$$

$$f_M = P(\text{“affected”} \mid Dd, D \text{ inherited from the mother})$$

Software: GENEHUNTER-IMPRINTING

Modelling of imprinting: nonparametric linkage analysis



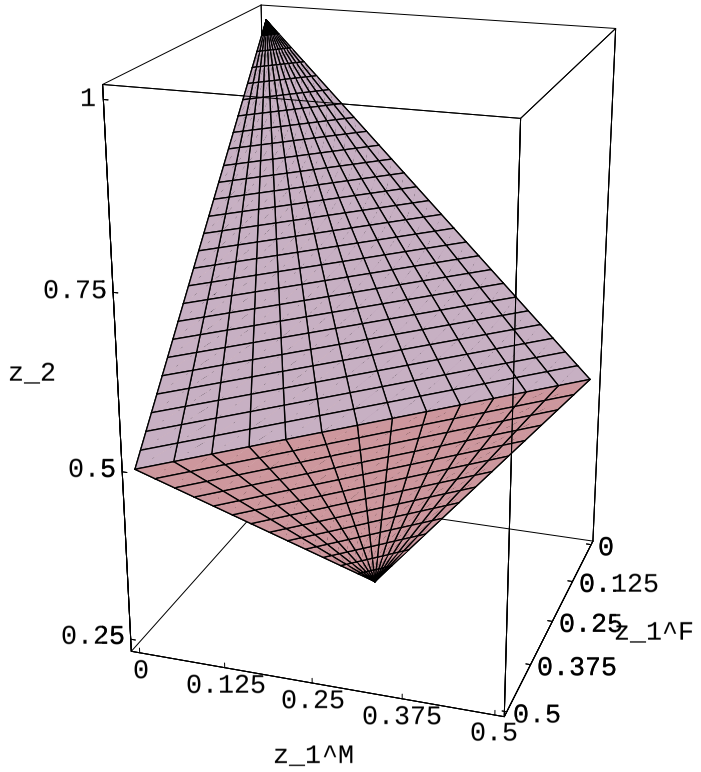
$$z_k := P(IBD = k),\ k = 0, 1^F, 1^M, 2$$

Testing for linkage:

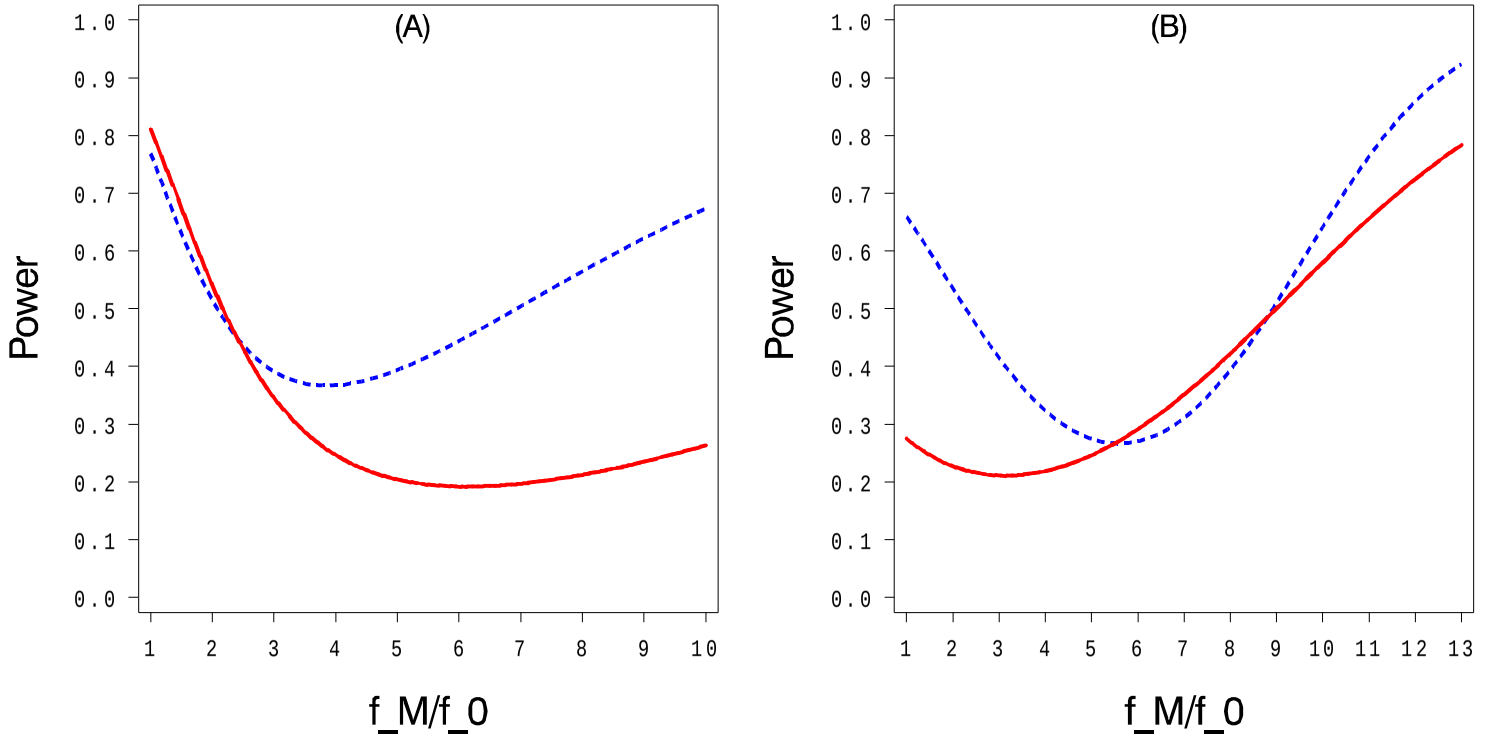
$$H_0 : (z_2, z_1^F, z_1^M, z_0) = (1/4, 1/4, 1/4, 1/4)$$

Genetic constraints for (z_2, z_1^F, z_1^M, z_0)

$$C := \{(z_2, z_1^F, z_1^M, z_0) : z_1^F + z_1^M \leq 1/2, z_1^F, z_1^M \in [z_0, z_2], z_2 + z_1^F + z_1^M + z_0 = 1\}$$



Power comparison



Power of MLS with (– –) and without (—) allowing for imprinting for a completely informative marker and $\alpha = .0001$. (A) Recessive model ($f_2/f_0 = 10$, $f_F/f_0 = 1$, $p_D = .27217$, $N = 120$); (B) Additive model ($f_2/f_0 = 13$, $f_F/f_0 = 7$, $p_D = .055555$, $N = 140$).