Variance Components for Modeling Quantitative Traits

Biostatistics 666

Guest Lecturer

Michael Boehnke

Today

- Analysis of quantitative traits
- Kinship coefficients
  - measure of genetic similarity between two individuals
- Modeling covariance for pairs of individuals
  - estimating heritability
  - estimating locus-specific heritability
- Extending the model to larger pedigrees

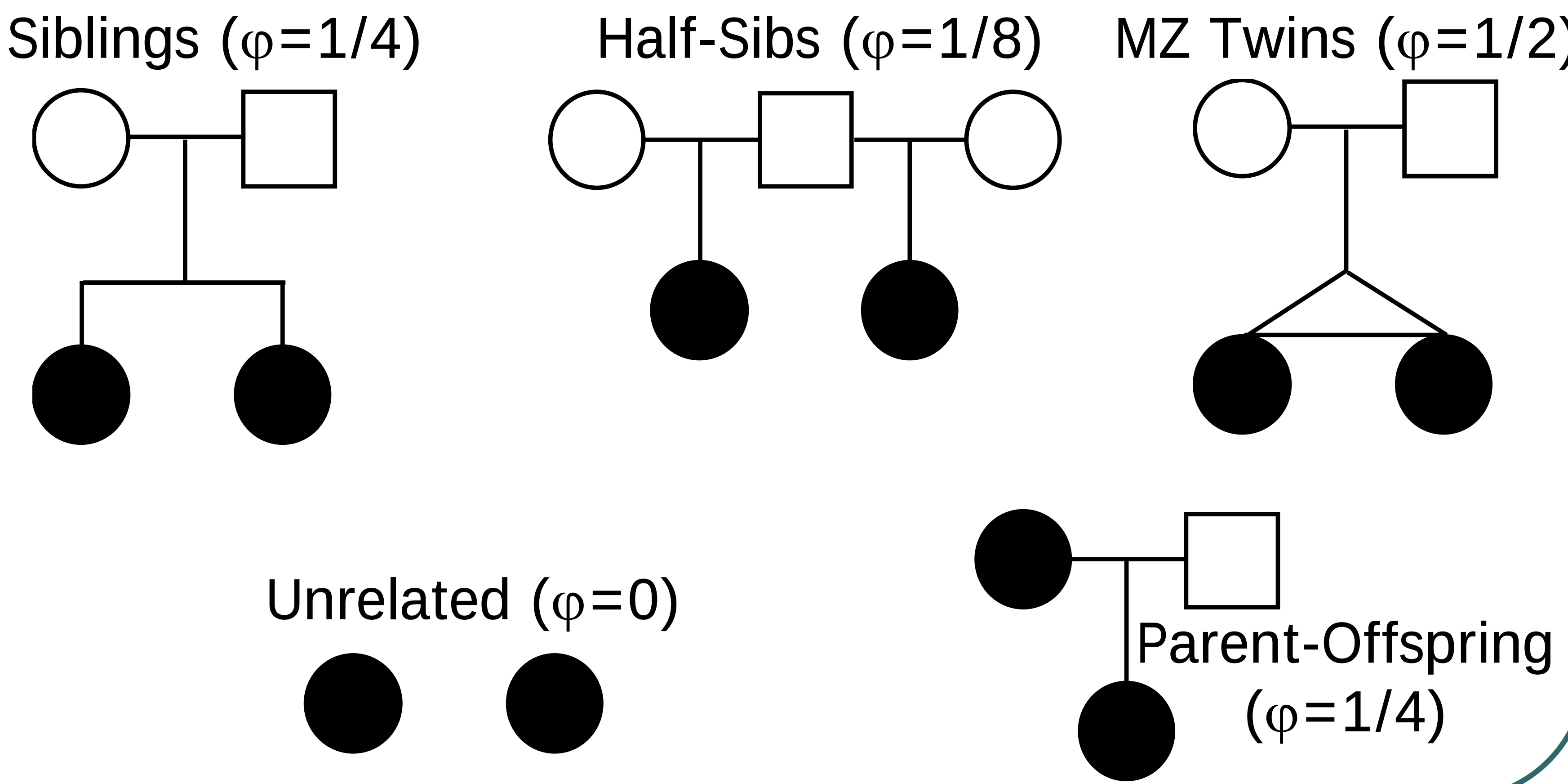
Kinship Coefficients

- Summarize genetic similarity between pairs of individuals.
- In a variance components model, they predict the phenotypic similarity between individuals.

Kinship Coefficients – Definition

- Given two individuals
  - One with genes (g_i, g_j)
  - The other with genes (g_k, g_l)
- The kinship coefficient is:
  - $\frac{1}{4}P(g_i \equiv g_k) + \frac{1}{4}P(g_i \equiv g_l) + \frac{1}{4}P(g_j \equiv g_k) + \frac{1}{4}P(g_j \equiv g_l)$
  - where "$\equiv$" represents identity by descent (IBD)
- Probability that two genes sampled at random from each individual are (IBD)

Some kinship coefficients



What about other relatives?

- For any two related individuals i and j …
- … use a recursive algorithm allows calculation of kinship coefficient
- Algorithm requires an order for individuals in the pedigree where ancestors precede descendants
    - That is where for any $i>j$, i is not ancestor of j
    - Such an order always exists!

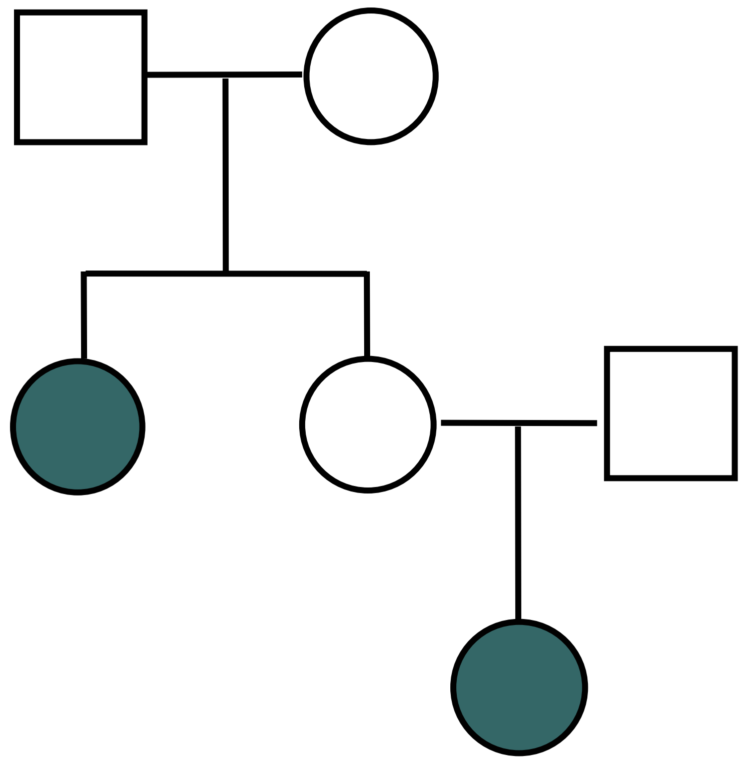
Computing Kinship Coefficients

- The recursive definition is then (for i ≥ j):

$$\varphi_{ij} = \begin{cases} 0 & i \text{ and } j \text{ are founders} \\ \tfrac{1}{2} & i = j, i \text{ is a founder} \\ \tfrac{1}{2}(\varphi_{mother(i)\,j} + \varphi_{father(i)\,j}) & i \neq j \\ \tfrac{1}{2}(1 + \varphi_{mother(i)\,father(i)}) & i = j \end{cases}$$

An example pedigree...

- Can you find …
- Suitable ordering for recursive calculation?
- Calculate kinship coefficient between shaded individuals?



So far …

- Summarize genetic similarity between any two individuals …

- Next, we will proceed to build a simple model for their phenotypes

Simplest Data Structure

- Pairs of related individuals
  - Siblings (or twins!)
  - Parent-Offspring
- Corresponding phenotype measurements
  - **$y = (y_1, y_2)'$**

Elements for a simple model …

- If the trait is normally distributed …

- Model mean and variance for $\mathbf{y}_1$ and $\mathbf{y}_2$
  - Mean and variance could be assumed equal …
  - … or they could depend on some covariates

- But we are also interested in covariance between the two …

Variance-Covariance Matrix

$$\Omega = \begin{bmatrix} V(y_1) & Cov(y_1, y_2) \\ Cov(y_1, y_2) & V(y_2) \end{bmatrix}$$

Model must describe not only variance of each observation but also covariance for pairs of observations

Bivariate density function

- Normal density function

$$L(y) = \frac{1}{\sqrt{2\pi}} \sigma^{-1} e^{-\frac{1}{2}(y-\mu)^2/\sigma^2}$$

- Bivariate normal density function

$$L(\mathbf{y}) = \frac{1}{2\pi} |\Omega|^{-1/2} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\Omega^{-1}(\mathbf{y}-\boldsymbol{\mu})}$$

- Extends univariate density function

Intuition on Normal Densities

$$L(y) = (2\pi)^{-\frac{1}{2}} \sigma^{-1} e^{-\frac{1}{2}(y-\mu)^2/\sigma^2}$$

Scaling parameter, penalizes settings with large variances

Distance between observation and its expected value

Bivariate Normal Densities

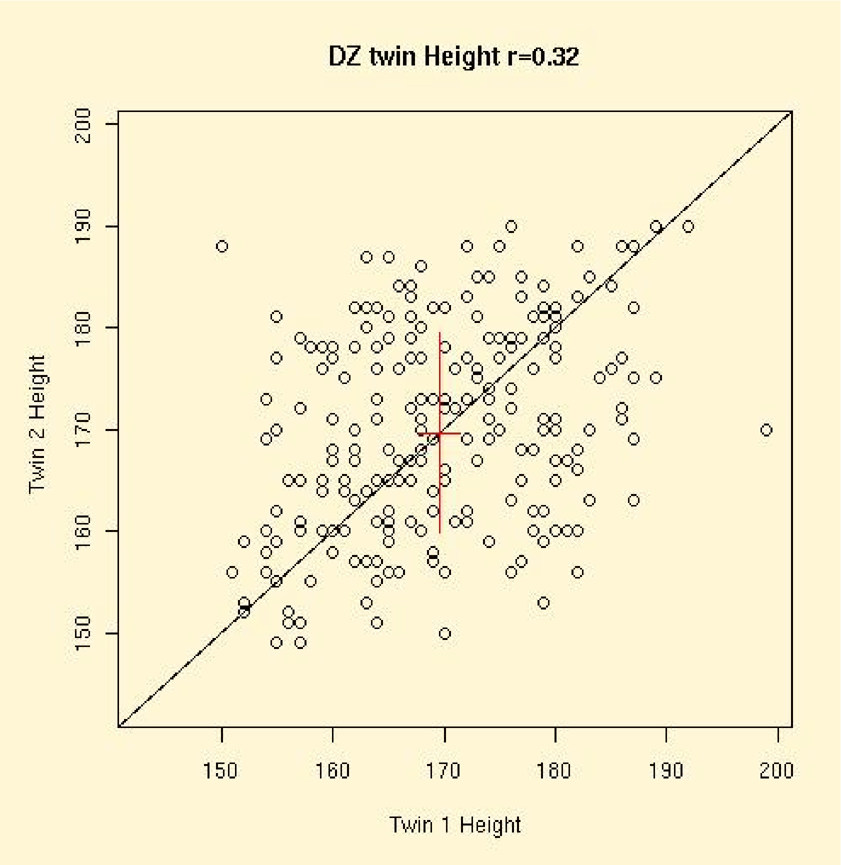
$$L(\mathbf{y}) = (2\pi)^{-1} |\Omega|^{-1/2} e^{-1/2(\mathbf{y}-\boldsymbol{\mu})'\Omega^{-1}(\mathbf{y}-\boldsymbol{\mu})}$$

Scaling parameter, penalizes settings with large variances

Distance between observation and its expected value

Possible Application…

In a sample of twin or sibling pairs, we could use all the data to estimate means, variances and even covariances…



(Data from David Duffy)

Incorporating Kinship Coefficients

- If genes influence trait …
- Covariance will differ for each class of relative pair
- Instead of estimating covariance for each relationship, …
- Impose genetic model that incorporates kinship and relates covariance between different classes of relative pair

A Simple Model for the Variance-Covariance Matrix

$$\Omega = \begin{bmatrix} \sigma_g^2 + \sigma_e^2 & 2\varphi\sigma_g^2 \\ 2\varphi\sigma_g^2 & \sigma_g^2 + \sigma_e^2 \end{bmatrix}$$

Where,

φ is the kinship coefficient for the two individuals

Example...

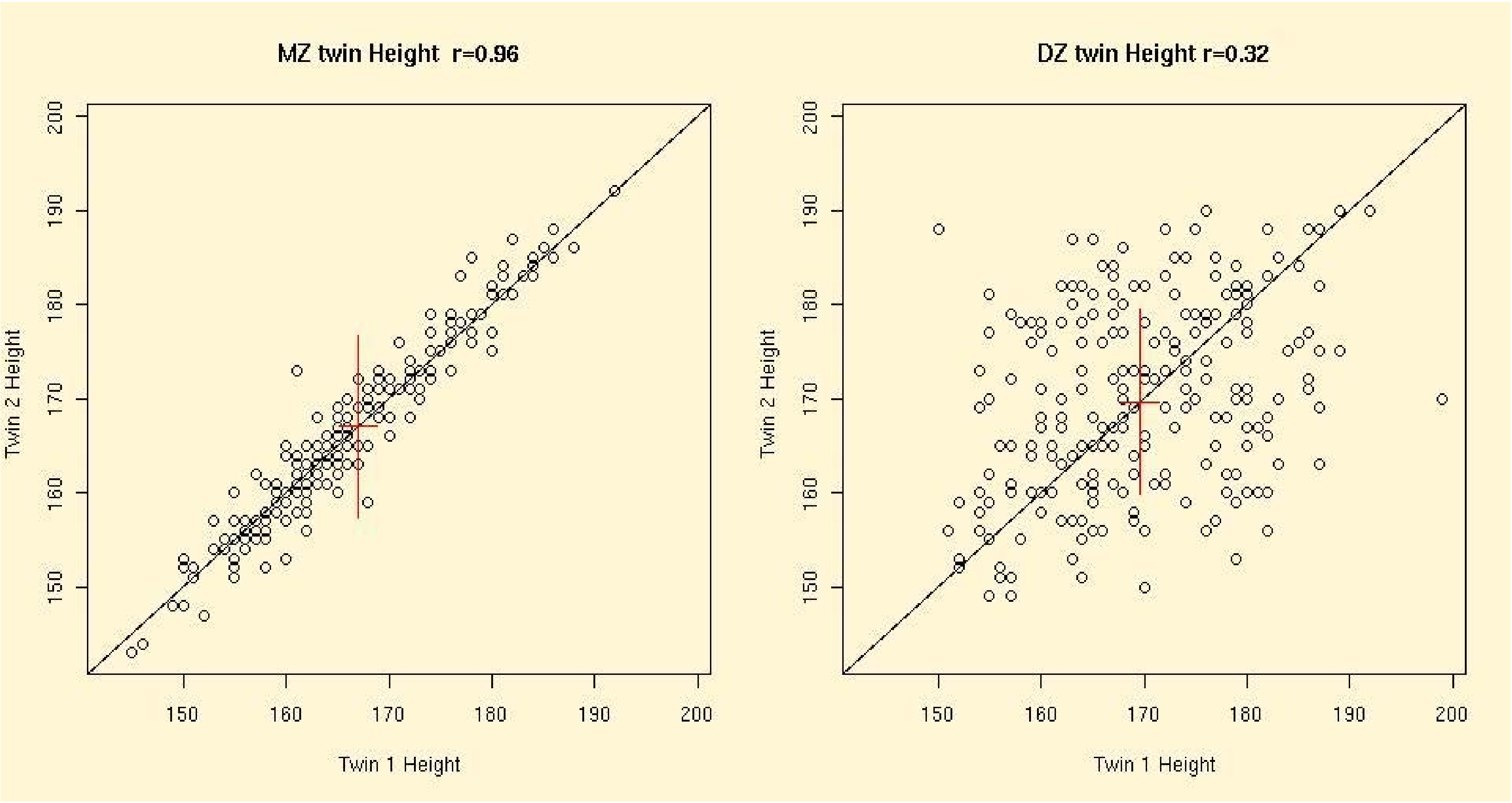
	N	r
MZ males	292	.80
MZ females	380	.80
DZ males	179	.47
DZ females	184	.55
DZ male-female	284	.41

(Reading ability scores from Eaves et al., 1997)

Interpretation…

- Fitting a maximum likelihood model…
  - Eaves et. al estimated
    - $\sigma_g^2 = .81$
    - $\sigma_e^2 = .19$
  - Found no evidence for sex differences
  - Saturated model did not improve fit

Height in DZ and MZ twins



(How would you interpret these data from David Duffy?)

So far …

- Model allows us to estimate the genetic contribution to the variation in any trait
- Incorporates different relative pairs …
- But it doesn't always fit…
  - Fortunately, the model can be easily refined

Another Example…

	N	r
MZ males	271	.56
MZ females	353	.52
DZ males	167	.33
DZ females	165	.45
DZ male-female	260	.41

(Psychomotor retardation scores from Eaves et al., 1997)

Refined Matrix

$$\Omega = \begin{bmatrix} \sigma_g^2 + \sigma_c^2 + \sigma_e^2 & 2\varphi\sigma_g^2 + \sigma_c^2 \\ 2\varphi\sigma_g^2 + \sigma_c^2 & \sigma_g^2 + \sigma_c^2 + \sigma_e^2 \end{bmatrix}$$

Where,

φ is the kinship coefficient for the two individuals

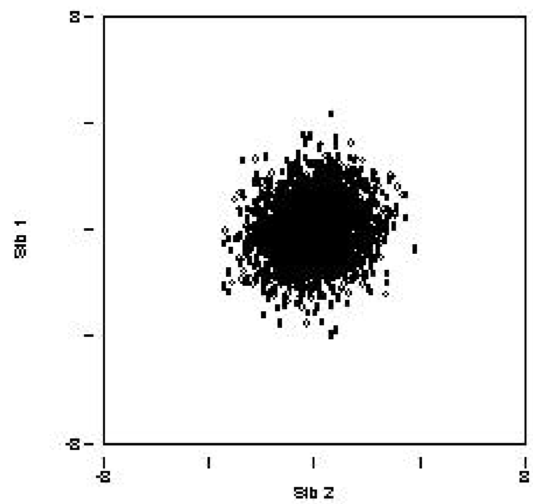
Interpretation…

- Fitting a maximum likelihood model…
  - Eaves et. al estimated (for males)
    - $\sigma_g^2 = .29$
    - $\sigma_c^2 = .24$
    - $\sigma_e^2 = .46$
  - Additive genetic effects could not explain similarities. Any idea why?

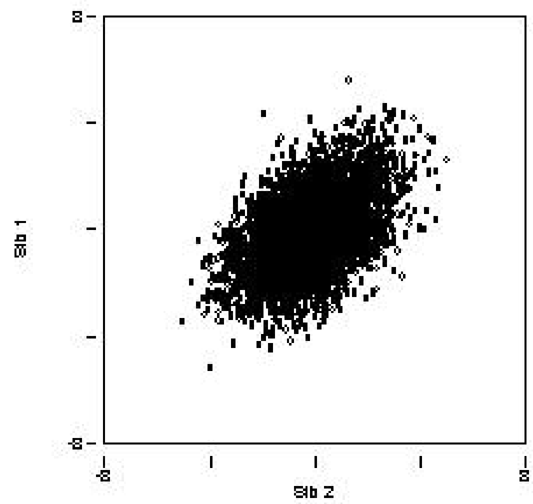
Incorporating IBD Coefficients

- Covariance might differ according to sharing at a particular locus
  - If locus contains genes that influence the trait
- Again, impose a genetic model and estimate model parameters

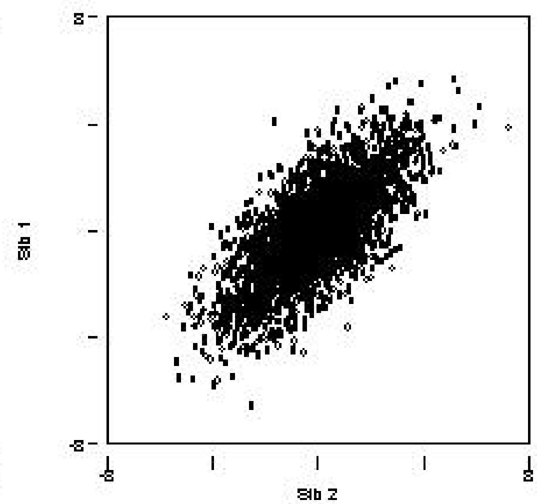
Linkage



IBD 0

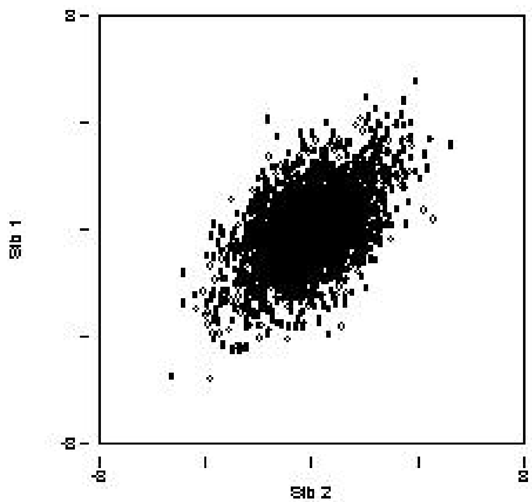


IBD 1

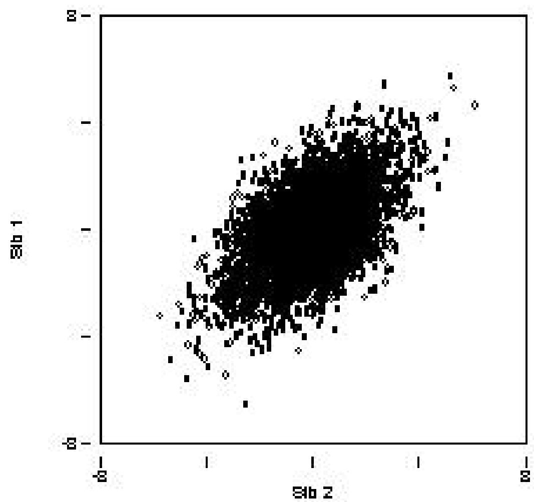


IBD 2

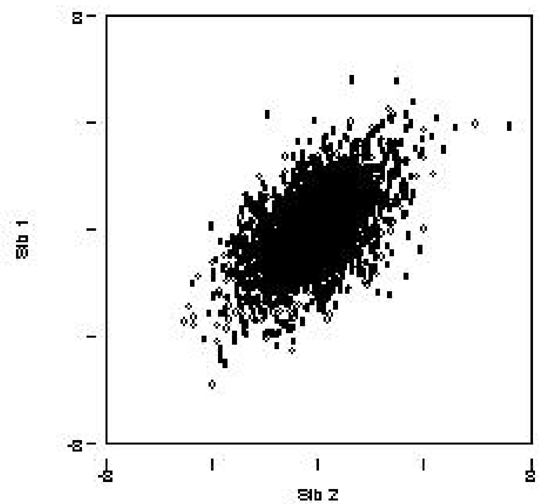
No Linkage



IBD 0



IBD 1



IBD 2

Relationship to IBD probabilities

- For non-inbred pair of relatives, marker or locus-specific kinship coefficients can be derived from IBD probabilities:

$$\varphi_{marker} = \tfrac{1}{4} P(IBD_{marker} = 1) + \tfrac{1}{2} P(IBD_{marker} = 2)$$

Variance-Covariance Matrix

$$\Omega = \begin{bmatrix} \sigma_a^2 + \sigma_g^2 + \sigma_e^2 & 2\varphi_{marker}\sigma_a^2 + 2\varphi\sigma_g^2 \\ 2\varphi_{marker}\sigma_a^2 + 2\varphi\sigma_g^2 & \sigma_a^2 + \sigma_g^2 + \sigma_e^2 \end{bmatrix}$$

Where,

φ is the kinship coefficient for the two individuals

φ_{marker} depends on the number of alleles shared IBD

Likelihood function, Incorporating Uncertain IBD

$$L = \prod_i \sum_{j=0,1,2} Z_{ij} (2\pi)^{-1} | \Omega_{IBD=j} |^{-1/2} \, e^{-1/2 (\mathbf{y}-\boldsymbol{\mu})' \boldsymbol{\Omega}_{IBD=j}^{-1} (\mathbf{y}-\boldsymbol{\mu})}$$

$$\approx \prod_i (2\pi)^{-1} | \Omega^* |^{-1/2} \, e^{-1/2 (\mathbf{y}-\boldsymbol{\mu})' \boldsymbol{\Omega}^{*-1} (\mathbf{y}-\boldsymbol{\mu})}$$

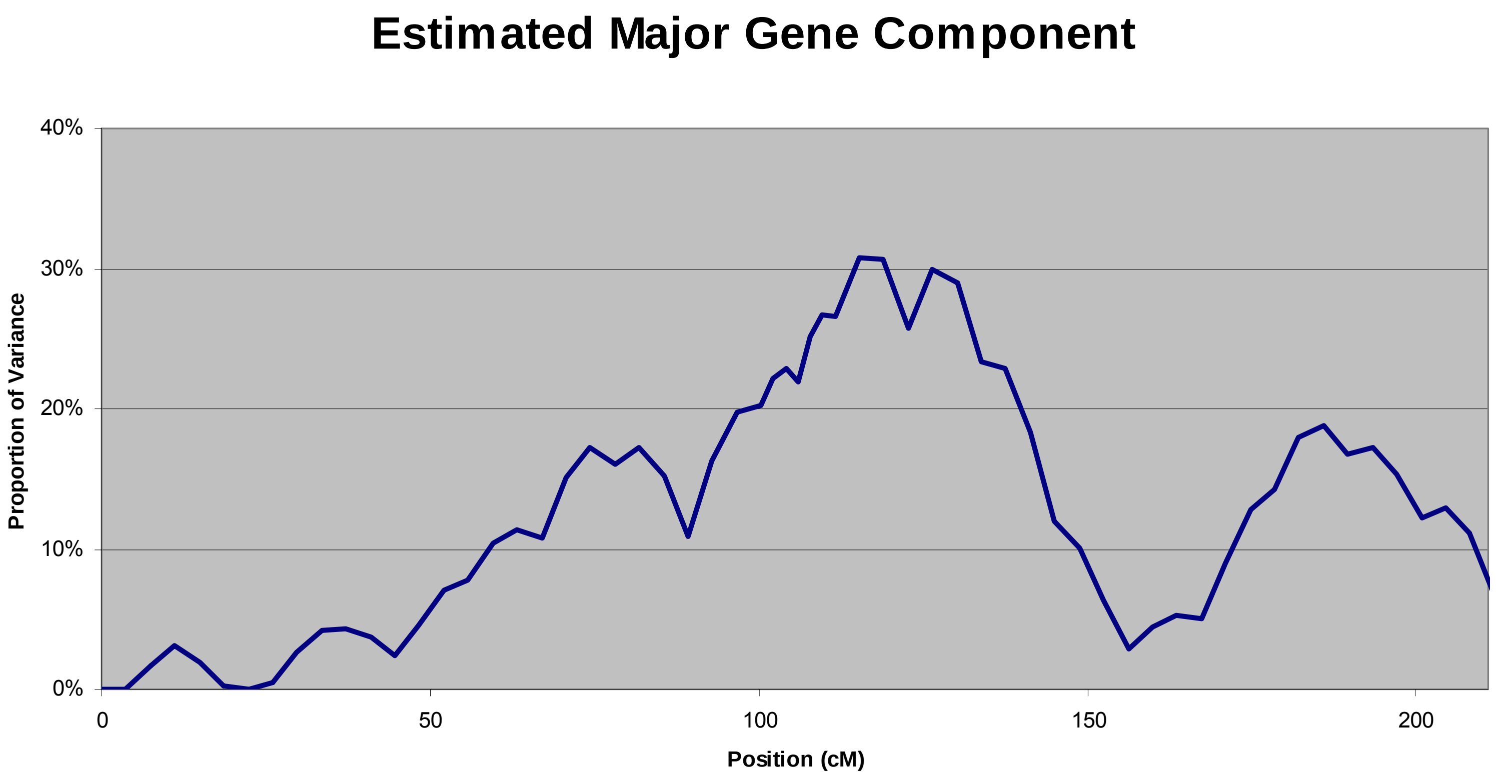
$$Z_{ij} = P(IBD_i = j \mid \text{marker data}) \quad \text{IBD sharing probabilities}$$

$$\Omega^* = \sum_{j=0,1,2} Z_{ij} \Omega_{IBD=j} \quad \text{"Expected"}\ \Omega$$

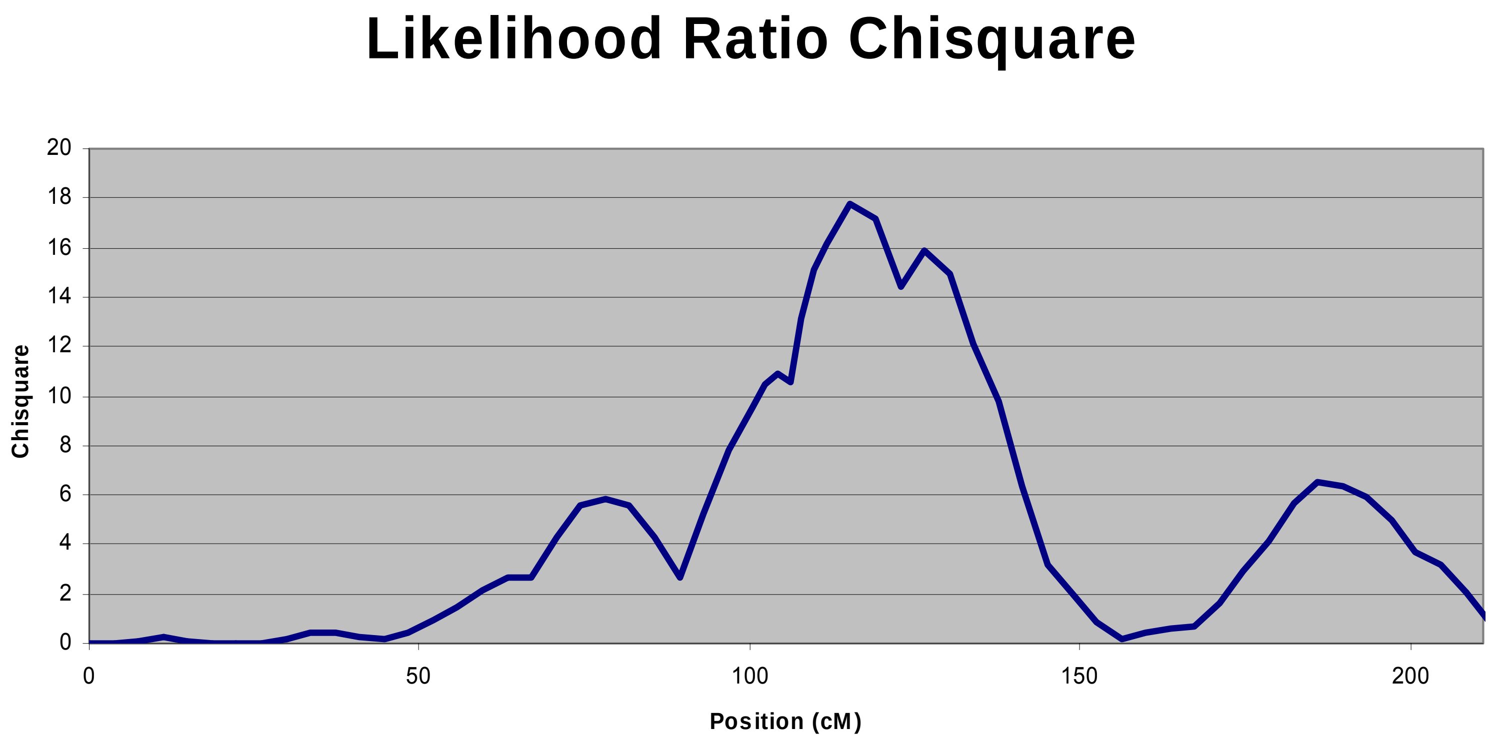
How it works ...

- To find linkage to a particular trait…
- Collect sibling pair sample
- Calculate IBD for multiple points along genome
- Model covariance as a function of IBD sharing at each point

Example…


Estimated Major Gene Component
Proportion of Variance
40%
30%
20%
10%
0%
0
50
100
150
200
Position (cM)

Example…



Likelihood Ratio Chisquare
Chisquare
20
18
16
14
12
10
8
6
4
2
0
0
50
100
150
200
Position (cM)

So far ...

- Models for similarity between relative pairs
- Kinship coefficient used to estimate overall genetic effect
- Locus-specific coefficients used to detect genetic linkage

Extensions ...

- The model extends gracefully to other settings:

- For larger pedigrees, we extend the covariance matrix

- To model genetic association, we model specific means for each individual

Larger Pedigrees…

$$\Omega_{jk} = \begin{cases} \sigma_a^2 + \sigma_g^2 + \sigma_e^2 & \text{if } j = k \\ 2\varphi_{marker}\sigma_a^2 + 2\varphi\sigma_g^2 & \text{if } j \neq k \end{cases}$$

Where,

φ is the kinship coefficient for the two individuals

φ_{marker} depends on the number of alleles shared IBD

j and k index different individuals in the family

Multivariate density function

- Normal density function

$$L(y) = (2\pi)^{-1/2} \sigma^{-1} e^{-1/2 (y-\mu)^2/\sigma^2}$$

- Multivariate normal density function

$$L(\mathbf{y}) = 2\pi^{-n/2} \,|\,\Omega\,|^{-1/2}\, e^{-1/2 (\mathbf{y}-\boldsymbol{\mu})'\Omega^{-1}(\mathbf{y}-\boldsymbol{\mu})}$$

- Extends univariate density function

Means Model

Expected Phenotype
for Individual i
(e.g. expected weight)

Estimated effects for covariates
(e.g. expected weight increases
1kg/year with age)

Measured Covariates
for Individual i
(e.g. age, sex, …)

$$E(y_i) = \mu + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}$$

In addition to modeling variances and covariances, we can model the means

Simple Association Model

- Each copy of allele changes trait by a fixed amount
  - Include covariate counting copies for allele of interest
- Evidence for association when $a \neq 0$

$$E(y_i) = \mu + a * [\text{number of copies of mutant allele}]$$

$$\mathrm{E}(\mathrm{y_i}) = \mu + \beta_X X_i$$

X is the number of copies for allele of interest.
β_x is the estimated effect of each copy (the additive genetic value).

Today

- Analysis of quantitative traits
- Kinship coefficients
  - Measure of genetic similarity between two individuals
- Modeling covariance for pairs of individuals
  - estimating heritability
  - estimating locus-specific heritability
- Extending the model to larger pedigrees

Useful References

- Amos (1994)

  Am J Hum Genet **54:**535-543

- Hopper and Matthews (1982)

  Ann Hum Genet **46**:373–383

- Lange and Boehnke (1983)

  Am J Med Genet **14**:513-24