QTL mapping in humans

Karl W Broman
Biostatistics & Medical Informatics
University of Wisconsin – Madison

www.biostat.wisc.edu/~kbroman

Biometrical preliminaries

Consider a single QTL with alleles $a_1, a_2, \ldots, a_n$

at frequencies $p_1, p_2, \ldots, p_n$

Assume random mating and Hardy-Weinberg equilibrium

Let Y = an individual's phenotype; G = QTL genotype

Define $\mu_{ij} \equiv$ E(Y | G = $a_i a_j$) $\quad (\mu_{ij} = \mu_{ji}$ for all $i, j)$

Write $\mu_{ij} = \mu + \alpha_i + \alpha_j + \delta_{ij}$

where $\quad \mu = \sum_{ij} \mu_{ij} p_i p_j$

$\alpha_i = \sum_j (\mu_{ij} - \mu) p_j \quad$ average effect of allele a_i

$\delta_{ij} = \mu_{ij} - (\mu + \alpha_i + \alpha_j) \quad$ "dominance deviation"

Genetic variances

For I, a random variable taking values 1, 2, . . . , n,

with probability $p_1, p_2, \ldots, p_n$

$$E(\alpha_I) = \sum_i \alpha_i p_i = 0$$

$$E(\delta_{Ij}) = \sum_i \delta_{ij} p_i = 0 \qquad \text{for any } j$$

Let J have the same distribution as I, with I & J being independent.

$$\text{var}(\alpha_I + \alpha_J) = 2\,\text{var}(\alpha_I) = 2\sum_i \alpha_i^2 p_i \equiv \sigma_a^2 \qquad \text{additive variance}$$

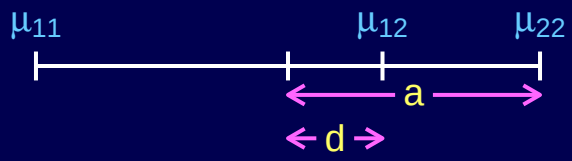
$$\text{var}(\delta_{IJ}) = \sum_{i,j} \delta_{ij}^2 p_i p_j \equiv \sigma_d^2 \qquad \text{dominance variance}$$

Assume $\text{var}(Y \mid G = a_i a_j) \equiv \sigma_e^2$

Then $\text{var}(Y) = \sigma_a^2 + \sigma_d^2 + \sigma_e^2$

Special case

Diallelic QTL, equally frequent alleles



$$\mu = \tfrac{1}{4}\mu_{11} + \tfrac{1}{2}\mu_{12} + \tfrac{1}{4}\mu_{22}$$

$$\begin{aligned}\alpha_1 &= \tfrac{1}{2}(\mu_{11} - \mu) + \tfrac{1}{2}(\mu_{12} - \mu) \\ &= \tfrac{1}{2}\mu_{11} + \tfrac{1}{2}\mu_{12} - \mu \\ &= \cdots = \tfrac{1}{4}(\mu_{11} - \mu_{22})\end{aligned}$$

$$\alpha_2 = -\alpha_1$$

$$\begin{aligned}\delta_{11} &= \mu_{11} - \mu - \alpha_1 - \alpha_2 \\ &= \cdots = \tfrac{1}{2}\{(\mu_{11} + \mu_{22})/2 - \mu_{12}\}\end{aligned}$$

$$\delta_{22} = \delta_{11}$$

$$\delta_{12} = \delta_{21} = -\delta_{11}$$

Kinship coefficient

Consider 2 individuals and a single autosomal locus.

Let I, J = random alleles chosen from each individual.

$$\Phi_{a,b} = \Pr(\text{I \& J are IBD})$$

For noninbred individuals, $\Phi_{a,b} = \frac{1}{4}\pi_1 + \frac{1}{2}\pi_2$

Relationship	π_0	π_1	π_2	Φ
MZ twins	0	0	1	1/2
Parent-offspring	0	1	0	1/4
Full sibs	1/4	1/2	1/4	1/4
Half sibs	1/2	1/2	0	1/8
Uncle-nephew	1/2	1/2	0	1/8
Grandparent-GC	1/2	1/2	0	1/8
Double 1st cousins	9/17	6/16	1/16	1/8
1st cousins	3/4	1/4	0	1/16
1st cousins once rem.	7/8	1/8	0	1/32
2nd cousins	15/16	1/16	0	1/64

Inbreeding coefficient

$$
\begin{aligned}
\mathsf{f} &= \Phi_{\mathsf{mom,dad}} \\
&= \Pr(\text{the individual's two alleles are IBD}) \\
&= 2\Phi_{\mathsf{a,a}} - 1
\end{aligned}
$$

Genetic correlations

Let A, B be non-inbred individuals with phenotypes Y_A, Y_B.

Let $X_i = Y_i - \mu$

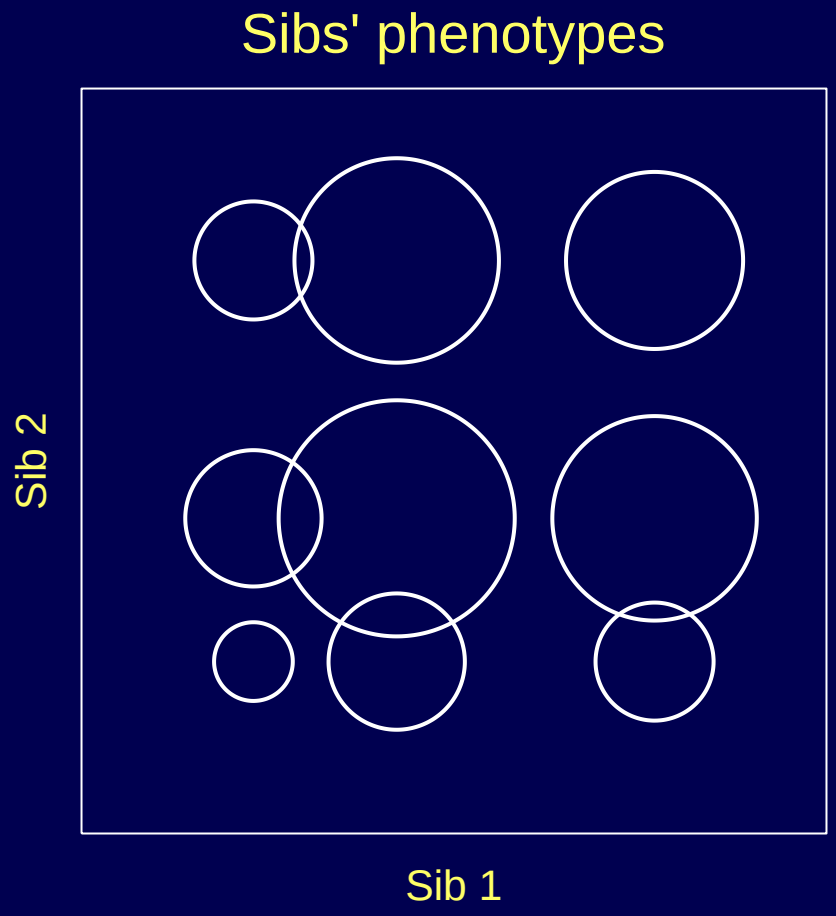
Assuming one QTL:

$$\begin{aligned} \mathrm{cov}(Y_A, Y_B) &= E(X_A X_B) \\ &= \sum_k E(X_A X_B \mid k \text{ alleles IBD at QTL})\, \pi_k \end{aligned}$$

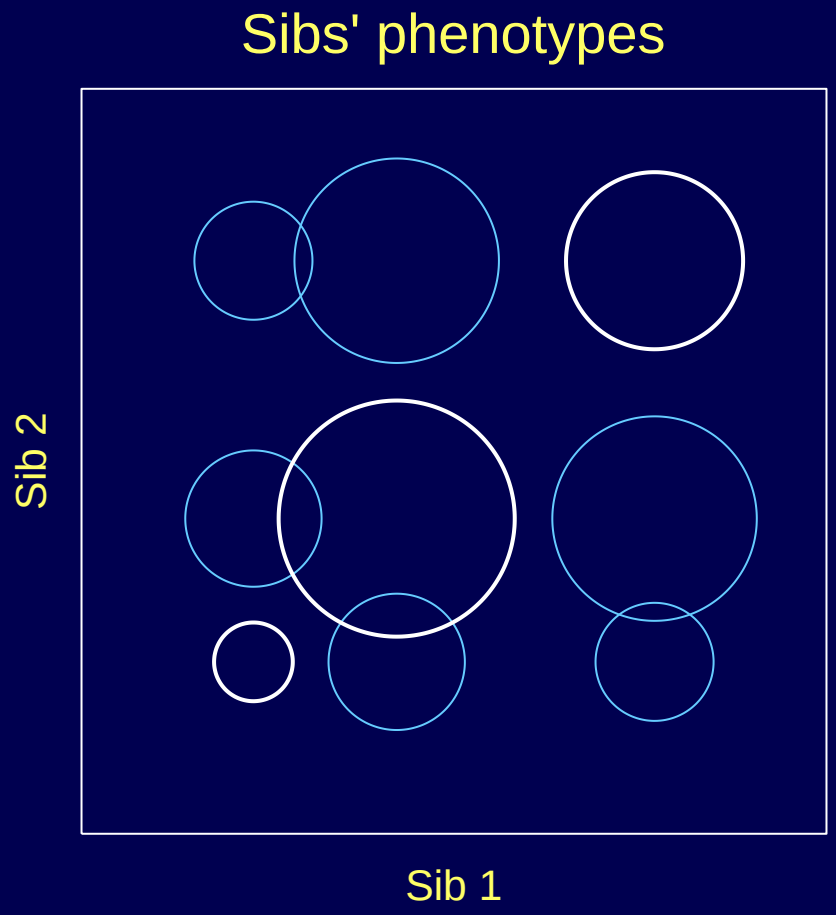
$$\begin{cases} E(X_A X_B \mid 2 \text{ alleles IBD}) = \cdots = \sigma_a^2 + \sigma_d^2 \\ E(X_A X_B \mid 1 \text{ alleles IBD}) = \cdots = \frac{1}{2}\,\sigma_a^2 \\ E(X_A X_B \mid 0 \text{ alleles IBD}) = \cdots = 0 \end{cases}$$

$$\begin{aligned} \text{Thus } \mathrm{cov}(Y_A, Y_B) &= (\sigma_a^2 + \sigma_d^2)\,\pi_2 + \tfrac{1}{2}\,\sigma_a^2\,\pi_1 \\ &= (\tfrac{1}{2}\pi_1 + \pi_2)\,\sigma_a^2 + \pi_2\,\sigma_d^2 \\ &= 2\,\Phi_{A,B}\,\sigma_a^2 + \pi_2\,\sigma_d^2 \end{aligned}$$

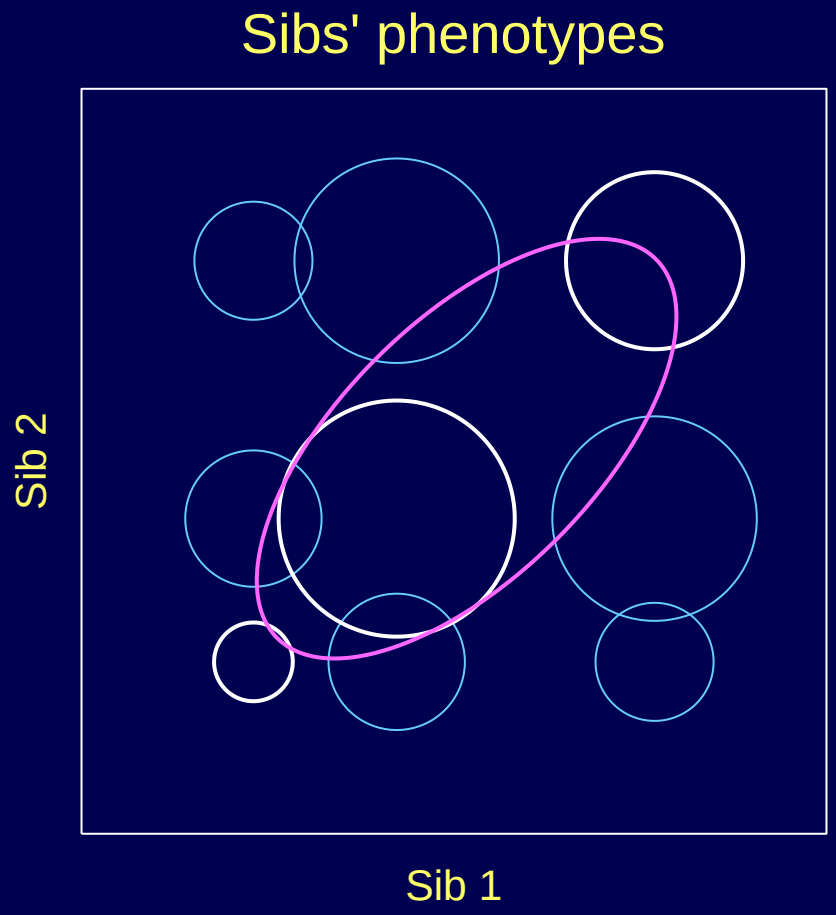
Sibpair, diallelic QTL



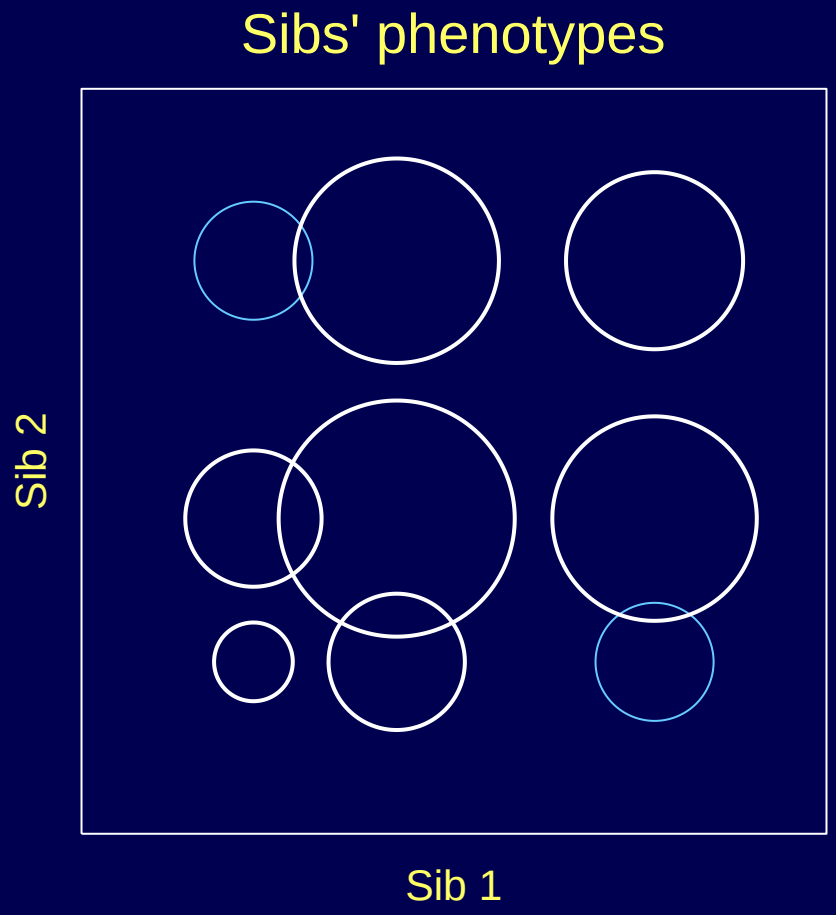
IBD=2



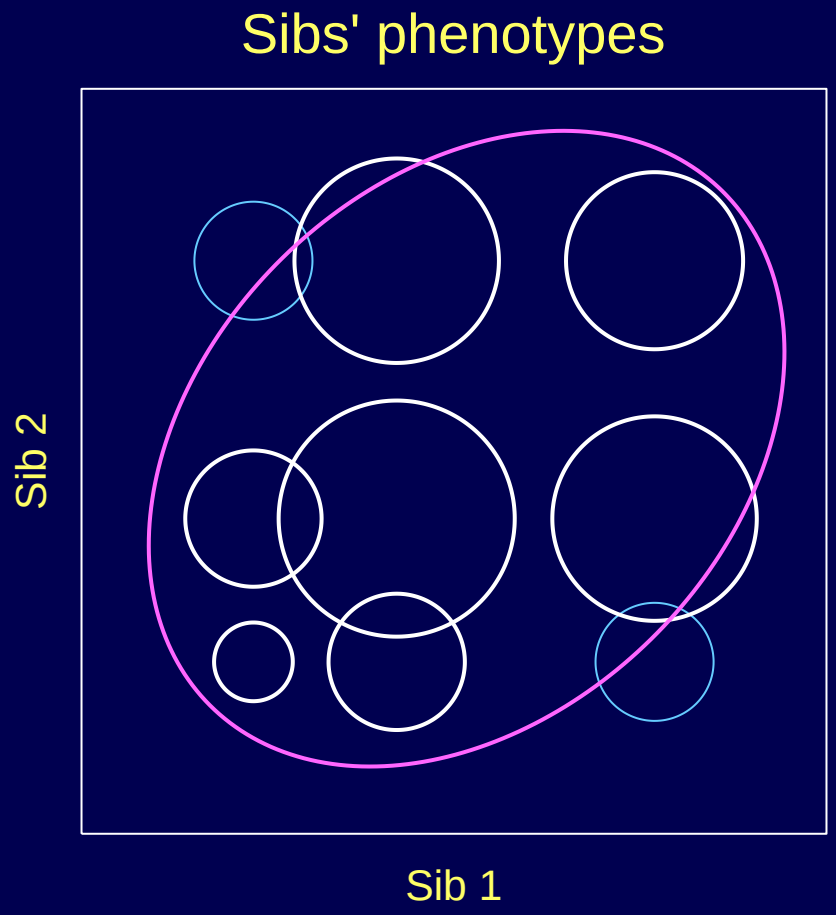
IBD=2



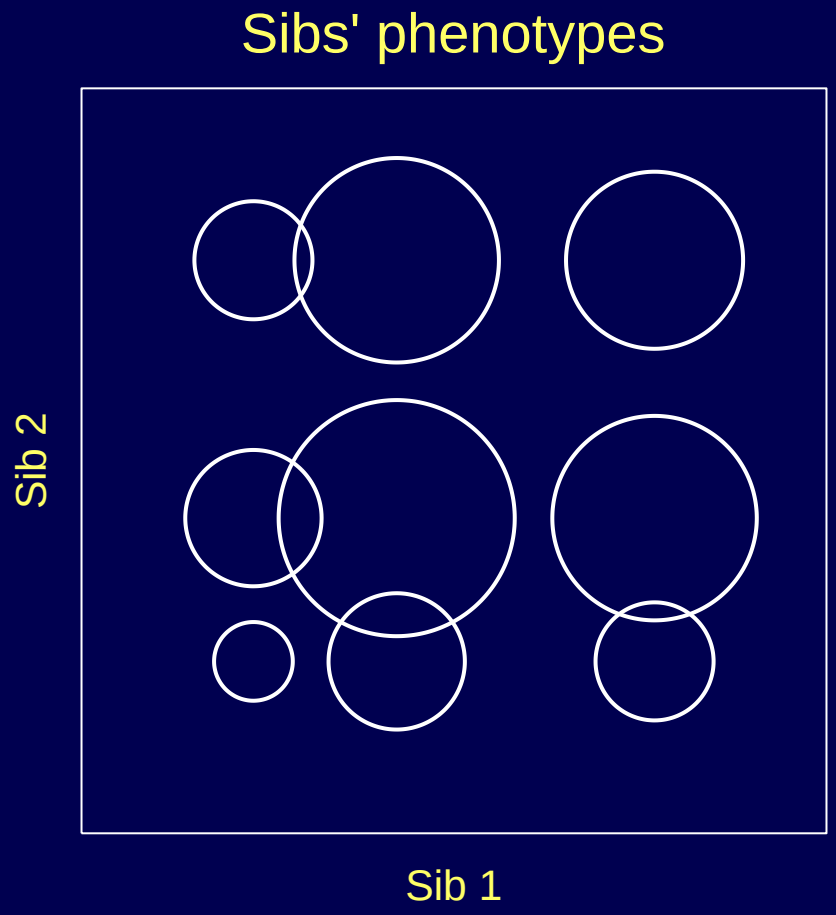
IBD = 1



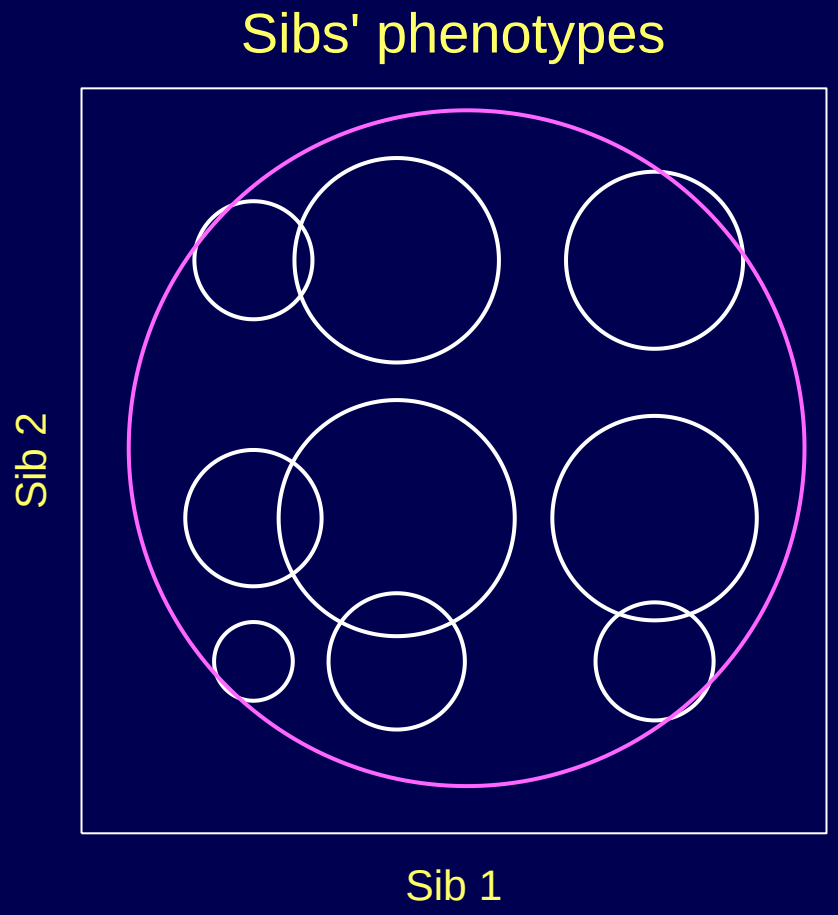
IBD = 1



IBD = 0



IBD = 0



A few more calculations...

Let K = # alleles IBD at marker, V = # alleles IBD at QTL

r = recombination fraction between marker and QTL

$\psi = r^2 + (1 - r)^2$

For sib pairs, Pr(V = v | K = k) =

	v		
k	0	1	2
0	$(1-\psi)^2$	$2\,\psi\,(1-\psi)$	ψ^2
1	$\psi\,(1-\psi)$	$(1-\psi)^2+\psi^2$	$\psi\,(1-\psi)$
2	ψ^2	$2\,\psi\,(1-\psi)$	$(1-\psi)^2$

$$\mathrm{cov}(Y_A, Y_B \mid K = k) = \sum_k \mathrm{cov}(Y_A, Y_B \mid V = v)\ \Pr(V = v \mid K = k)$$

$$= \begin{cases} \psi\,(\sigma_a^2 + \psi\,\sigma_d^2) & \text{if k = 0} \\ \frac{1}{2}\,\sigma_a^2 + \psi(1-\psi)\sigma_d^2 & \text{if k = 1} \\ (1-\psi)\left[\sigma_a^2 + (1-\psi)\sigma_d^2\right] & \text{if k = 2} \end{cases}$$

One more thing...

$$
\begin{aligned}
\mathrm{E}\left[(\mathrm{Y_A}-\mathrm{Y_B})^2 \;\middle|\; \mathrm{V}=\mathrm{v}\right] &= \mathrm{var}\left[\mathrm{Y_A}-\mathrm{Y_B} \mid \mathrm{V}=\mathrm{v}\right] + \left\{\mathrm{E}(\mathrm{Y_A}-\mathrm{Y_B} \mid \mathrm{V}=\mathrm{v})\right\}^2 \\
&= \mathrm{var}(\mathrm{Y_A} \mid \mathrm{V}=\mathrm{v}) + \mathrm{var}(\mathrm{Y_B} \mid \mathrm{V}=\mathrm{v}) - 2\,\mathrm{cov}(\mathrm{Y_A}, \mathrm{Y_B} \mid \mathrm{V}=\mathrm{v}) \\
&= 2(\sigma_a^2+\sigma_d^2+\sigma_e^2) - \mathrm{k}\,\sigma_a^2 + \delta_{\mathrm{k2}}\,(2\,\sigma_d^2)
\end{aligned}
$$

QTL mapping with sib pairs

1. Explicit diallelic model

   (Y_{i1}, Y_{i2}) follows a mixture of 9 bivariate normals, parametrized by $\mu_{11}, \mu_{12}, \mu_{22}, \sigma^2$

2. "Variance components"

   (Y_{i1}, Y_{i2}) | # alleles IBD at QTL = k

   follows a bivariate normal distribution, with mean (μ, μ) and variance matrix

$$V_k = \begin{pmatrix} v_{11} & v_{12}^k \\ v_{12}^k & v_{11} \end{pmatrix}$$

where $v_{11} = \sigma_a^2 + \sigma_d^2 + \sigma_e^2$,

$$v_{12}^k = \begin{cases} 0 & \text{if k =} \quad 0 \\ \frac{1}{2}\,\sigma_a^2 & \quad\quad\quad\;\; 1 \\ \sigma_a^2 + \sigma_d^2 & \quad\quad\quad\;\; 2 \end{cases}$$

QTL mapping with sib pairs

2. “Variance components”

Method 2A: With incomplete data on k, treat as

mixture of 3 bivariate normal dist’ns, with wts Pr(K=k | marker data)

Method 2B: With incomplete data on k, treat as

single bivariate normal, plugging in

$\hat{k}$ = E(k | marker data) for k and

Pr(k = 2 | marker data) for δ_{k2}

$\longrightarrow$ replace V_k with E(V_k | marker data)

QTL mapping with sib pairs

3. Haseman-Elston regression

Consider only $(Y_{i1} - Y_{i2})^2$

$$E\left\{(Y_{i1} - Y_{i2})^2 \mid k \text{ alleles IBD at QTL}\right\} = \beta_0 - (\sigma_a^2)\, k + (2\sigma_d^2)\, \delta_{k2}$$

$\longrightarrow$ Regress $(Y_1 - Y_2)^2$ on k (or on k and δ_{k2})

$\longrightarrow$ Incomplete info on k? Plug in $\hat{k}$

Assumes $(Y_1 - Y_2)^2 \mid k \sim$ normal with constant variance

Larger pedigrees

1. Variance components

   phenotype = mean + add've QTL + "polygenes" + noise

   var(Y) = $\sigma^2_{\mathrm{Q}a} + \sigma^2_{\mathrm{P}a} + \sigma^2_e$

   cov(Y_i, Y_j | k alleles IBD at QTL) = $\frac{1}{2}$ k $\sigma^2_{\mathrm{Q}a} + 2\ \Phi_{ij}\ \sigma^2_{\mathrm{P}a}$

2. Explicit diallelic QTL model (via MCMC)

Larger pedigrees

1. Variance components

phenotype = mean + add've QTL + "polygenes" + noise

var(Y) = $\sigma^2_{\mathrm{Q}a} + \sigma^2_{\mathrm{P}a} + \sigma^2_e$

cov(Y_i, Y_j | k alleles IBD at QTL) = $\frac{1}{2}$ k $\sigma^2_{\mathrm{Q}a} + 2\ \Phi_{ij}\ \sigma^2_{\mathrm{P}a}$

2. Explicit diallelic QTL model (via MCMC)

H-E robust to non-normality, but has low power

V-C powerful when the MVN model is correct, but is not robust

Other issues: selective sampling, treatment of covariates