Modelling Epistasis

Fisher (1918) first partitioned the genetic variance
using the least squares principle
variance components:
additive, dominance, and epistatic effects

Cockerham (1954) partitioned two-gene epistatis
into four variance components
– additive×additive
– additive×dominance
– dominance×additive
– dominance×dominance

Mather Jinks (1982) model based on RI (F_∞)
usual model developed in genetics classes
$G_2 = \mu + a$, $G_1 = \mu + d$, $G_0 = \mu - a$
does not extend well to epistasis
estimates of a, d depend on interactions

Gene effects and variances with epistasis under the least square genetic model

G_{jl}^{ik} genotypic value for offspring from
male gamete A_iB_k with female gamete A_jB_l
assume no maternal effects: $G_{jl}^{ik} = G_{ik}^{jl}$

model effects classified in two ways
within columns: loci effects g_j^i, g_l^k
loci interaction $(g_j^i g_l^k)$
within rows: gametic effects g^{ik} and g_{jl}
gametic interaction $(g^{ik} g_{jl})$

$$
\begin{array}{rcl}
G_{jl}^{ik} = \mu & & +g_j^i \ +g_l^k \ +(g_j^i g_l^k) \\
 & & = \ = \ = \\
+g^{ik} & = & +a^i \ +a^k \ +(a^i a^k) \\
+g_{jl} & = & +a_j \ +a_l \ +(a_j a_l) \\
+(g^{ik} g_{jl}) & = & +d_j^i \ +d_l^k \ +(a^i a_l) + (a_j a^k) \\
 & & +(a^i d_l^k) + (d_j^i a^k) \\
 & & +(a_j d_l^k) + (d_j^i a_l) \\
 & & +(d_j^i d_l^k)
\end{array}
$$

connections between gametic effects and loci effects

all epistatic effects represent
interactions among nonallelic genes
interactions among loci

each gametic effect contains average effects
and additive by additive effect

gamete interactions include dominance effects
and remaining epistatic effects ($a \times d$, $d \times a$, $d \times d$)

assume random mating and linkage equilibrium
allow different frequencies in male, female gametes
$P_{jl}^{ik} = p^i p_j p^k p_l, \quad \sum p^i = \sum p_j = \sum p^k = \sum p_l = 1$

recall least squares genetic model from 4.1, 4.4

effects in terms of genetic means G 4-factor factorial design

additive main effects

$$a^i = G_{\cdot\cdot}^{i\cdot} - G_{\cdot\cdot}^{\cdot\cdot}, \quad a^k = G_{\cdot\cdot}^{\cdot k} - G_{\cdot\cdot}^{\cdot\cdot}$$

$$a_j = G_{j\cdot}^{\cdot\cdot} - G_{\cdot\cdot}^{\cdot\cdot}, \quad a_l = G_{\cdot l}^{\cdot\cdot} - G_{\cdot\cdot}^{\cdot\cdot}$$

dominance as two-allele intralocus interactions

$$d_j^i = G_{j\cdot}^{i\cdot} - G_{\cdot\cdot}^{i\cdot} - G_{j\cdot}^{\cdot\cdot} + G_{\cdot\cdot}^{\cdot\cdot}, \quad d_l^k = G_{\cdot l}^{\cdot k} - G_{\cdot\cdot}^{\cdot k} - G_{\cdot l}^{\cdot\cdot} + G_{\cdot\cdot}^{\cdot\cdot}$$

additive × additive two-loci interactions

$$(a^i a^k) = G_{\cdot\cdot}^{ik} - G_{\cdot\cdot}^{i\cdot} - G_{\cdot\cdot}^{\cdot k} + G_{\cdot\cdot}^{\cdot\cdot}, \quad (a^i a_l) = G_{\cdot l}^{i\cdot} - G_{\cdot\cdot}^{i\cdot} - G_{\cdot l}^{\cdot\cdot} + G_{\cdot\cdot}^{\cdot\cdot}, \quad \cdots$$

additive × dominance three-factor interactions

$$(a^i d_l^k) = G_{\cdot l}^{ik} - G_{\cdot\cdot}^{ik} - G_{\cdot l}^{i\cdot} - G_{\cdot l}^{\cdot k} + G_{\cdot\cdot}^{i\cdot} + G_{\cdot\cdot}^{\cdot k} + G_{\cdot l}^{\cdot\cdot} - G_{\cdot\cdot}^{\cdot\cdot}, \quad \cdots$$

$$(d_j^i a^k) = G_{j\cdot}^{ik} - G_{\cdot\cdot}^{ik} - G_{j\cdot}^{i\cdot} - G_{j\cdot}^{\cdot k} + G_{\cdot\cdot}^{i\cdot} + G_{\cdot\cdot}^{\cdot k} + G_{j\cdot}^{\cdot\cdot} - G_{\cdot\cdot}^{\cdot\cdot}, \quad \cdots$$

dominance × dominance four-factor interaction

$$(d_j^i d_l^k) = G_{jl}^{ik} - G_{j\cdot}^{ik} - G_{\cdot l}^{ik} - G_{jl}^{i\cdot} - G_{jl}^{\cdot k} + G_{\cdot\cdot}^{ik} + G_{j\cdot}^{i\cdot} + G_{j\cdot}^{\cdot k} + G_{\cdot l}^{i\cdot} + G_{\cdot l}^{\cdot k} + G_{j\cdot}^{\cdot\cdot} + G_{\cdot l}^{\cdot\cdot}$$
$$- G_{\cdot\cdot}^{i\cdot} - G_{\cdot\cdot}^{\cdot k} - G_{j\cdot}^{\cdot\cdot} - G_{\cdot l}^{\cdot\cdot} + G_{\cdot\cdot}^{\cdot\cdot}$$

sum-to-zero side constraints imposed:

$$\sum_i p^i (a^i a^k) = \sum_k p^k (a^i a^k) = \cdots = \sum_j p_j (d_j^i d_l^k) = 0$$

total variance σ^2_G

$$\begin{aligned}
\sigma^2_G &= \sum_{i,j,k,l} p^i p_j p^k p_l (G^{ik}_{jl} - \mu)^2 \\
&= \left[\sum_i p^i (a^i)^2 + \sum_k p^k (a^k)^2 + \sum_j p_j (a_j)^2 + \sum_l p_l (a_l)^2\right] \\
&+ \left[\sum_{i,j} p^i p_j (d^i_j)^2 + \sum_{k,l} p^k p_l (d^k_l)^2\right] + \left[\sum_{i,k} p^i p^k (a^i a^k)^2\right. \\
&+ \left.\sum_{j,l} p_j p_l (a_j a_l)^2 + \sum_{i,l} p^i p_l (a^i a_l)^2 + \sum_{j,k} p_j p^k (a_j a^k)^2\right] \\
&+ \left[\sum_{i,k,l} p^i p^k p_l (a^i d^k_l)^2 + \sum_{i,j,k} p^i p_j p^k (d^i_j a^k)^2\right. \\
&+ \left.\sum_{j,k,l} p_j p^k p_l (a_j d^k_l)^2 + \sum_{i,j,l} p^i p_j p_l (d^i_j a_l)^2\right] \\
&+ \left[\sum_{i,j,k,l} p^i p_j p^k p_l (d^i_j d^k_l)^2\right] \\
&= \sigma^2_a + \sigma^2_d + \sigma^2_{aa} + \sigma^2_{ad} + \sigma^2_{dd}
\end{aligned}$$

Model I: Fisher (1918) Cockerham (1954)
orthogonal partition of genetic variance

consider F_2 population in H-W equilibrium
regress genotypic value G
on quadratic in number of copies of allele A
linear contrast $\sim$ additive effect
quadratic contrast $\sim$ dominance effect

Genotype	AA	Aa	aa
Genotypic value	G_2	G_1	G_0
Frequency	$p^2 = \frac{1}{4}$	$2pq = \frac{1}{2}$	$q^2 = \frac{1}{4}$
linear w_1	1	0	-1
quadratic w_2	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$

Model I: $G_i = \mu + a \times w_{1i} + d \times w_{2i}$
$G_2 = \mu + a - \frac{1}{2}d, \quad G_1 = \mu + \frac{1}{2}d, \quad G_0 = \mu - a - \frac{1}{2}d$
with $\mu = p^2 G_2 + 2pq G_1 + q^2 G_0$
orthogonal genetic model for F_2 population

$$Cov(w_1, w_2) = 0$$

linear and quadratic effects are uncorrelated
call these additive and dominance (why?)

Model I regression coefficients

Model I: $G_i = \mu + \beta_{Gw_1} \times w_{1i} + \beta_{Gw_2} \times w_{2i}$

$$\beta_{Gw_1} = \frac{p^2 G_2 - q^2 G_0}{p^2 + q^2} = \frac{1}{2}(G_2 - G_0) = a$$

$$\begin{aligned}
\beta_{Gw_2} &= \frac{-\frac{1}{2}p^2 G_2 + \frac{1}{2} 2pq G_1 - \frac{1}{2} q^2 G_0}{\frac{1}{4}p^2 + \frac{1}{4} 2pq + \frac{1}{4} q^2} \\
&= (G_1 - \frac{1}{2} G_2 - \frac{1}{2} G_0) = d
\end{aligned}$$

Model I variance components

$$\sigma^2_G = \sigma^2_{w_1} + \sigma^2_{w_2} - 2Cov(w_1, w_2) = \sigma^2_1 + \sigma^2_2 - 0$$

$$\sigma^2_1 = 2pq[p(G_2 - G_1) + q(G_1 - G_0)]^2 = \frac{1}{8}[G_2 - G_0]^2$$

$$\sigma^2_2 = p^2 q^2 [G_2 - 2G_1 + G_0]^2 = \frac{1}{16}[G_2 - 2G_1 + G_0]^2$$

Model I due to Fisher (1918)
differs from usual model in quantitative genetics

Model II: Mather Jinks (1982)

Genotype	AA	Aa	aa
Genotypic value	G_2	G_1	G_0
Frequency	$p^2 = \frac{1}{4}$	$2pq = \frac{1}{2}$	$q^2 = \frac{1}{4}$
linear v_1	1	0	-1
quadratic v_2	0	1	0

Model II: $G_i = \mu + a \times v_{1i} + d \times v_{2i}$
$G_2 = \mu + a, \quad G_1 = \mu + d, \quad G_0 = \mu - a$

$Cov(v_1, v_2) = 0$ but $E(v_2) = 1/2 \neq 0$

Model I and II give identical results for single marker or single locus analysis: same estimates of a and d.

They give same analysis for multiple loci if no epistatic terms are in the model.

With epistasis in the model, Model I extends in natural way but Model II has different estimates of a and d.

Model I: Fisher (1918) Cockerham (1954) orthogonal partition of genetic variance

drop assumption of H-W equilibrium
Model I: $G_i = \mu + \beta_{Gw_1} \times w_{1i} + \beta_{Gw_2} \times w_{2i}$
w_1, w_2: linear, quadratic orthogonal polynomials
based on genotype frequencies f_k

$$\sum_{i=0}^{2} f_i w_{1i} = \sum_{i=0}^{2} f_i w_{2i} = \sum_{i=0}^{2} f_i w_{1i} w_{2i} = 0$$

Genotype	AA	Aa	aa
Genotypic value	G_2	G_1	G_0
Frequency	f_2	f_1	f_0
w_1	$2f_0 + f_1$	$f_0 - f_2$	$-(2f_2 + f_1)$
w_2	$-1/(8f_2)$	$1/(4f_1)$	$-1/(8f_0)$

linear steps

$$w_{12} - w_{11} = (2f_0 + f_1) - (f_0 - f_2) = f_\cdot = 1$$

$$w_{11} - w_{10} = (f_0 - f_2) + (2f_2 + f_1) = f_\cdot = 1$$

partition into variance components

$$\sigma_G^2 = \sigma_1^2 + \sigma_2^2$$

with

$$\sigma_t^2 = [Cov(G, w_t)]^2 / \sigma_{w_t}^2 = \beta_{Gw_t}^2 \sigma_{w_t}^2$$

and

$$G = \mu + \beta_{Gw_1} w_1 + \beta_{Gw_2} w_2$$

$$\sigma_k^2 = \left(\sum_{i=0}^{2} f_i G_i w_{ki} \right)^2 / \left(\sum_{i=0}^{2} f_i w_{ki}^2 \right), \quad k = 1, 2$$

F_2 based epistatic model (Cockerham model)

9 genotypes in an F_2 population
need 8 genetic parameters for complete description
assume Hardy-Weinberg and linkage equilibrium
Cockerham (1954)'s orthogonal partition of genetic variance
leads to definition of genotypic value G_{ij}

$$G_{ij} = \beta_0 + \sum_{t=1}^{8} \beta_{Gw_t} w_{tij}$$

eight orthogonal scales or contrasts w_t
Four are marginal scales (Model I)
four are interaction scales made from these marginal scales: linear $\sim$ additive, quadratic $\sim$ dominance
interaction scales:

$w_5 = w_1 \times w_3$	linear $\times$ linear	additive $\times$ additive
$w_6 = w_1 \times w_4$	linear $\times$ quadratic	additive $\times$ dominance
$w_7 = w_2 \times w_3$	quadratic $\times$ linear	dominance $\times$ additive
$w_8 = w_2 \times w_4$	quadratic $\times$ quadratic	dominance $\times$ dominance

epistasis model

$$G_{ij} = \beta_0 + \beta_{Gw_1} w_{1ij} + \beta_{Gw_2} w_{2ij} + \beta_{Gw_3} w_{3ij} + \beta_{Gw_4} w_{4ij}$$
$$+ \beta_{Gw_5} w_{1ij} w_{3ij} + \beta_{Gw_6} w_{1ij} w_{4ij} + \beta_{Gw_7} w_{2ij} w_{3ij} + \beta_{Gw_8} w_{2ij} w_{4ij}$$

with

$$w_1 = \begin{cases} 1 & \text{for } AA \\ 0 & \text{for } Aa \\ -1 & \text{for } aa \end{cases} \qquad w_2 = \begin{cases} -\frac{1}{2} & \text{for } AA \\ \frac{1}{2} & \text{for } Aa \\ -\frac{1}{2} & \text{for } aa \end{cases}$$

$$w_3 = \begin{cases} 1 & \text{for } BB \\ 0 & \text{for } Bb \\ -1 & \text{for } bb \end{cases} \qquad w_4 = \begin{cases} -\frac{1}{2} & \text{for } BB \\ \frac{1}{2} & \text{for } Bb \\ -\frac{1}{2} & \text{for } bb \end{cases}$$

$$w_5 = w_1 w_3, \quad w_6 = w_1 w_4, \quad w_7 = w_2 w_3, \quad w_8 = w_2 w_4$$

matrix notation

$$\mathbf{G} = \mathbf{W}\beta$$

$$\begin{bmatrix} G_{22} \\ G_{21} \\ G_{20} \\ G_{12} \\ G_{11} \\ G_{10} \\ G_{02} \\ G_{01} \\ G_{00} \end{bmatrix} = \begin{bmatrix} 1 & 1 & -\frac{1}{2} & 1 & -\frac{1}{2} & 1 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{4} \\ 1 & 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & -\frac{1}{4} \\ 1 & 1 & -\frac{1}{2} & -1 & -\frac{1}{2} & -1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & \frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & \frac{1}{2} & -\frac{1}{4} \\ 1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{4} \\ 1 & 0 & \frac{1}{2} & -1 & -\frac{1}{2} & 0 & 0 & -\frac{1}{2} & -\frac{1}{4} \\ 1 & -1 & -\frac{1}{2} & 1 & -\frac{1}{2} & -1 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{4} \\ 1 & -1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & -\frac{1}{2} & 0 & -\frac{1}{4} \\ 1 & -1 & -\frac{1}{2} & -1 & -\frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \end{bmatrix}$$

Genotypic values, frequencies, and eight orthogonal scales for F_2 populations

Scale	AABB	AABb	AAbb	AaBB	AaBb	Aabb	aaBB	aaBb	aabb
G	G_{22}	G_{21}	G_{20}	G_{12}	G_{11}	G_{10}	G_{02}	G_{01}	G_{00}
f	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$
W_1	1	1	1	0	0	0	−1	−1	−1
W_2	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
W_3	1	0	−1	1	0	−1	1	0	−1
W_4	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
W_5	1	0	−1	0	0	0	−1	0	1
W_6	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	0	0	0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$
W_7	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{1}{2}$	0	$\frac{1}{2}$
W_8	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$

Definition of genetic parameters

Parameters	Definition
$\mu = \beta_0$	mean
$a_1 = \beta_1$	additive effect of locus A
$d_1 = \beta_2$	dominance effect of locus A
$a_2 = \beta_3$	additive effect of locus B
$d_2 = \beta_4$	dominance effect of locus B
$i_{aa} = \beta_5$	additive × additive effect of locus A and B
$i_{ad} = \beta_6$	additive × dominance effect of locus A and B
$i_{da} = \beta_7$	dominance × additive effect of loci A and B
$i_{dd} = \beta_8$	dominance × dominance effect of loci A and B

$$\beta = \mathbf{W}^{-1}\mathbf{G}$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \end{bmatrix} = \begin{bmatrix} \frac{1}{16} & \frac{1}{8} & \frac{1}{16} & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & 0 & 0 & 0 & -\frac{1}{8} & -\frac{1}{4} & -\frac{1}{8} \\ -\frac{1}{16} & -\frac{1}{8} & -\frac{1}{16} & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & -\frac{1}{16} & -\frac{1}{8} & -\frac{1}{16} \\ \frac{1}{8} & 0 & -\frac{1}{8} & \frac{1}{4} & 0 & -\frac{1}{4} & \frac{1}{8} & 0 & -\frac{1}{8} \\ -\frac{1}{16} & \frac{1}{8} & -\frac{1}{16} & -\frac{1}{8} & \frac{1}{4} & -\frac{1}{8} & -\frac{1}{16} & \frac{1}{8} & -\frac{1}{16} \\ \frac{1}{4} & 0 & -\frac{1}{4} & 0 & 0 & 0 & -\frac{1}{4} & 0 & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \\ -\frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} & 0 & -\frac{1}{2} & -\frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} G_{22} \\ G_{21} \\ G_{20} \\ G_{12} \\ G_{11} \\ G_{10} \\ G_{02} \\ G_{01} \\ G_{00} \end{bmatrix}$$

$$\begin{aligned}
\beta &= \mathbf{W}^{-1}\mathbf{G} \\
\beta_0 &= \frac{G_{22}}{16} + \frac{G_{21}}{8} + \frac{G_{20}}{16} + \frac{G_{12}}{8} + \frac{G_{11}}{4} + \frac{G_{10}}{8} + \frac{G_{02}}{16} + \frac{G_{01}}{8} + \frac{G_{00}}{16} = G_{..} \\
\beta_1 &= \frac{G_{22}}{8} + \frac{G_{21}}{4} + \frac{G_{20}}{8} - \frac{G_{02}}{8} - \frac{G_{01}}{4} - \frac{G_{00}}{8} = \frac{G_{2.} - G_{0.}}{2} \\
\beta_2 &= \frac{G_{11}}{4} + \frac{G_{12}}{8} + \frac{G_{10}}{8} - \frac{G_{22}}{16} - \frac{G_{21}}{8} - \frac{G_{20}}{16} - \frac{G_{02}}{16} - \frac{G_{01}}{8} - \frac{G_{00}}{16} \\
&= \frac{2G_{1.} - G_{2.} - G_{0.}}{2} \\
\beta_3 &= \frac{G_{22}}{8} + \frac{G_{12}}{4} + \frac{G_{02}}{8} - \frac{G_{20}}{8} - \frac{G_{10}}{4} - \frac{G_{00}}{8} = \frac{G_{.2} - G_{.0}}{2} \\
\beta_4 &= \frac{G_{11}}{4} + \frac{G_{21}}{8} + \frac{G_{01}}{8} - \frac{G_{22}}{16} - \frac{G_{12}}{8} - \frac{G_{02}}{16} - \frac{G_{20}}{16} - \frac{G_{10}}{8} - \frac{G_{00}}{16} \\
&= \frac{2G_{.1} - G_{.2} - G_{.0}}{2} \\
\beta_5 &= \frac{(G_{22} - G_{02}) - (G_{20} - G_{00})}{4} = \frac{(G_{22} - G_{20}) - (G_{02} - G_{00})}{4} \\
\beta_6 &= \frac{(2G_{21} - G_{22} - G_{20}) - (2G_{01} - G_{02} - G_{00})}{4} \\
\beta_7 &= \frac{(2G_{12} - G_{22} - G_{02}) - (2G_{10} - G_{20} - G_{00})}{4} \\
\beta_8 &= \frac{2(2G_{11} - G_{21} - G_{01}) - (2G_{12} - G_{22} - G_{02}) - (2G_{10} - G_{20} - G_{00})}{4} \\
&= \frac{2(2G_{11} - G_{12} - G_{10}) - (2G_{21} - G_{22} - G_{20}) - (2G_{01} - G_{02} - G_{00})}{4}
\end{aligned}$$

orthogonality and variance components

regression coefficient β_{Gw_t} associated with scale w_t
is corresponding genetic effect by orthogonality
each effect contributes to its variance component

$$\sigma^2_G = \frac{1}{2}a_1^2+\frac{1}{4}d_1^2+\frac{1}{2}a_2^2+\frac{1}{4}d_2^2+\frac{1}{4}i_{aa}^2+\frac{1}{8}i_{ad}^2+\frac{1}{8}i_{da}^2+\frac{1}{16}i_{dd}^2$$

no covariance between different effects

comparison with Mather and Jinks model

extend Model II to include epistasis

$$G_{ij} = \beta_0 + \sum_{t=1}^{8} \beta_{Gw_t} v_{tij}$$

$v_1, \cdots, v_4$ defined by Model II, $v_5, \cdots, v_8$ below
Mather Jinks (1982 *Biometrical Genetics*) model:

$$G_{ij} = \beta_0 + \beta_{Gv_1}v_{1ij} + \beta_{Gv_2}v_{2ij} + \beta_{Gv_3}v_{3ij} + \beta_{Gv_4}v_{4ij}$$
$$+\beta_{Gv_5}v_{1ij}v_{3ij} + \beta_{Gv_6}v_{1ij}v_{4ij} + \beta_{Gv_7}v_{2ij}v_{3ij} + \beta_{Gv_8}v_{2ij}v_{4ij}$$

$$v_1 = \begin{cases} 1 & \text{for } AA \\ 0 & \text{for } Aa \\ -1 & \text{for } aa \end{cases} \qquad v_2 = \begin{cases} 0 & \text{for } AA \\ 1 & \text{for } Aa \\ 0 & \text{for } aa \end{cases}$$

$$v_3 = \begin{cases} 1 & \text{for } BB \\ 0 & \text{for } Bb \\ -1 & \text{for } bb \end{cases} \qquad v_4 = \begin{cases} 0 & \text{for } BB \\ 1 & \text{for } Bb \\ 0 & \text{for } bb \end{cases}$$

$$v_5 = v_1v_3, \;\; v_6 = v_1v_4, \;\; v_7 = v_2v_3, \;\; v_8 = v_2v_4$$

$$\mathbf{G} = \mathbf{V}\beta$$

$$\begin{bmatrix} G_{22} \\ G_{21} \\ G_{20} \\ G_{12} \\ G_{11} \\ G_{10} \\ G_{02} \\ G_{01} \\ G_{00} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ a_1 \\ d_1 \\ a_2 \\ d_2 \\ i_{aa} \\ i_{ad} \\ i_{da} \\ i_{dd} \end{bmatrix}$$

$$\beta = \mathbf{V}^{-1}\mathbf{G}$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & -\frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 & -\frac{1}{4} \\ -\frac{1}{4} & 0 & -\frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{4} & 0 & -\frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 & -\frac{1}{4} & 0 & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} & 0 & 0 & 0 & -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \\ \frac{1}{4} & 0 & -\frac{1}{4} & 0 & 0 & 0 & -\frac{1}{4} & 0 & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \\ -\frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} & 0 & -\frac{1}{2} & -\frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} G_{22} \\ G_{21} \\ G_{20} \\ G_{12} \\ G_{11} \\ G_{10} \\ G_{02} \\ G_{01} \\ G_{00} \end{bmatrix}$$

Mather Jinks model derivation

start with four homozygote genotypes (G_{22}, G_{20}, G_{02}, G_{00})
population of recombinant inbred lines, RI=F_∞)
define μ, a_1, a_2 and i_{aa} for RI lines

$$\mu = \frac{1}{4}G_{22} + \frac{1}{4}G_{20} + \frac{1}{4}G_{02} + \frac{1}{4}G_{00}$$

consider heterozygote genotypes
add terms d_1, d_2, i_{aa}, i_{ad}, i_{da} and i_{dd}

usually called F_∞ metric model

(Va der Veen 1959 *Genetica* 30: 201-232)
not based on population with nine genotypes observed
not based on orthogonal partitions of genetic effect and variance

F_∞ metric model applied to F_2 population

v_2 and v_8 are not independent
neither are v_4 and v_8

Thus some genetic effects defined by the model do not only contribute to their own genetic variance component.

We prefer Cockerham model, also called F_2 metric model, for analyzing QTL epistasis in an F_2 population.

Note: Estimates of i_{aa}, i_{ad}, i_{da}, and i_{dd} are the same under both models, but those of a_1, d_1, a_2 and d_2 are different.

Examples of analysis

As an example, we showed the epistatic analysis of two QTL in a series of experiments of introgressing maize allele into teosinte genetic background (Doebley, Stec, and Gustus 1995 *Genetics* 141:333-346). Two QTL, one located on chromosome arm 1L and one on 3L, were introgressed by repeated backcross to teosinte and then intercrossed to produce an F_2 population to have these two QTL segregating simultaneously in teosinte background. We showed the mean values and sample sizes of the nine genotypes and the analyses of gene effects on three traits.

Mean value and sample size of two unlinked QTL genotypes for trait 1 (John Doebley)

	bb	Bb	BB
aa	17.98 (10)	54.57 (21)	47.80 (11)
Ab	40.94 (24)	47.55 (42)	83.62 (20)
AA	61.11 (3)	66.50 (22)	101.65 (8)

Estimate and test of gene effects: Trait 1

Effect	Estimate	S.E.	t	P
a_1	15.11	4.47	3.41	0.0008
d_1	-3.92	5.84	-0.67	0.5035
a_2	19.46	4.42	4.40	0.0001
d_2	-5.66	5.84	-0.97	0.3336
i_{aa}	2.68	7.07	0.38	0.7054
i_{ad}	-18.28	8.87	-2.06	0.0411
i_{da}	3.75	8.85	0.42	0.6725
i_{dd}	-18.13	11.68	-1.55	0.1227

Mean value and sample size of two unlinked QTL genotypes for trait 4 (John Doebley)

	bb	Be	BB
aa	4.46 (12)	3.05 (22)	2.09 (11)
Aa	3.23 (26)	2.02 (46)	1.12 (21)
AA	1.43 (7)	1.08 (25)	1.00 (9)

Estimate and test of gene effects: Trait 4

Effect	Estimate	S.E.	t	P
a_1	-1.01	0.09	-10.67	0.0001
d_1	-0.06	0.13	-0.43	0.6712
a_2	-0.88	0.09	-9.29	0.0001
d_2	-0.17	0.13	-1.29	0.1994
i_{aa}	0.48	0.14	3.45	0.0007
i_{ad}	0.05	0.19	0.25	0.8017
i_{da}	-0.36	0.19	-1.89	0.0606
i_{dd}	0.03	0.26	0.11	0.9126

Mean value and sample size of two unlinked QTL genotypes for trait 9 (John Doebley)

	bb	Bb	BB
aa	7.32 (12)	0.41 (21)	1.12 (10)
Aa	1.34 (26)	0.22 (46)	0.28 (21)
AA	0.79 (7)	0.00 (25)	0.00 (9)

Estimate and test of gene effects: Trait 9

Effect	Estimate	S.E.	t	P
a_1	-1.06	0.29	-3.68	0.0003
d_1	-0.74	0.39	-1.88	0.0621
a_2	-1.14	0.28	-3.98	0.0001
d_2	-1.35	0.39	-3.42	0.0008
i_{aa}	1.35	0.43	3.16	0.0019
i_{ad}	1.71	0.57	2.98	0.0033
i_{da}	1.22	0.57	2.13	0.0346
i_{dd}	1.52	0.79	1.93	0.0550

Issues on detecting QTL epistasis

- How to detect and analyze QTL epistasis is a very important issue for QTL mapping analysis. If we know where QTL are located, we can build appropriate genetic and statistical models to test and estimate the effects and interactions of QTL at the locations. But the problem is we do not know where QTL are, and the analyses of QTL effects and interactions are complicated by the search for QTL.

- A common practice to detect QTL epistasis in many published QTL mapping analyses is to perform pairwise marker interaction analysis and to compare the percentage of significant marker pairs on interaction at a given significance level with that expected at the null hypothesis of no interaction.

This analysis, of course, does not analyze interaction of individual QTL. It generally has low power to detect QTL epistasis because of segregation between markers and QTL and possible aggregation and *cancellation* effects from multiple QTL. Also some interactions at some regions can be swamped by the huge number of possible pair-wise tests of markers.

- An appropriate method to analyze QTL epistasis is to integrate QTL epistatic effects in the analysis of QTL and to search and map multiple QTL including epistasis simultaneously. Multiple interval mapping discussed below is such a method.