Lecture 6: Introduction to Quantitative genetics

Bruce Walsh lecture notes
Uppsala EQG course
version 28 Jan 2012

Quantitative Genetics

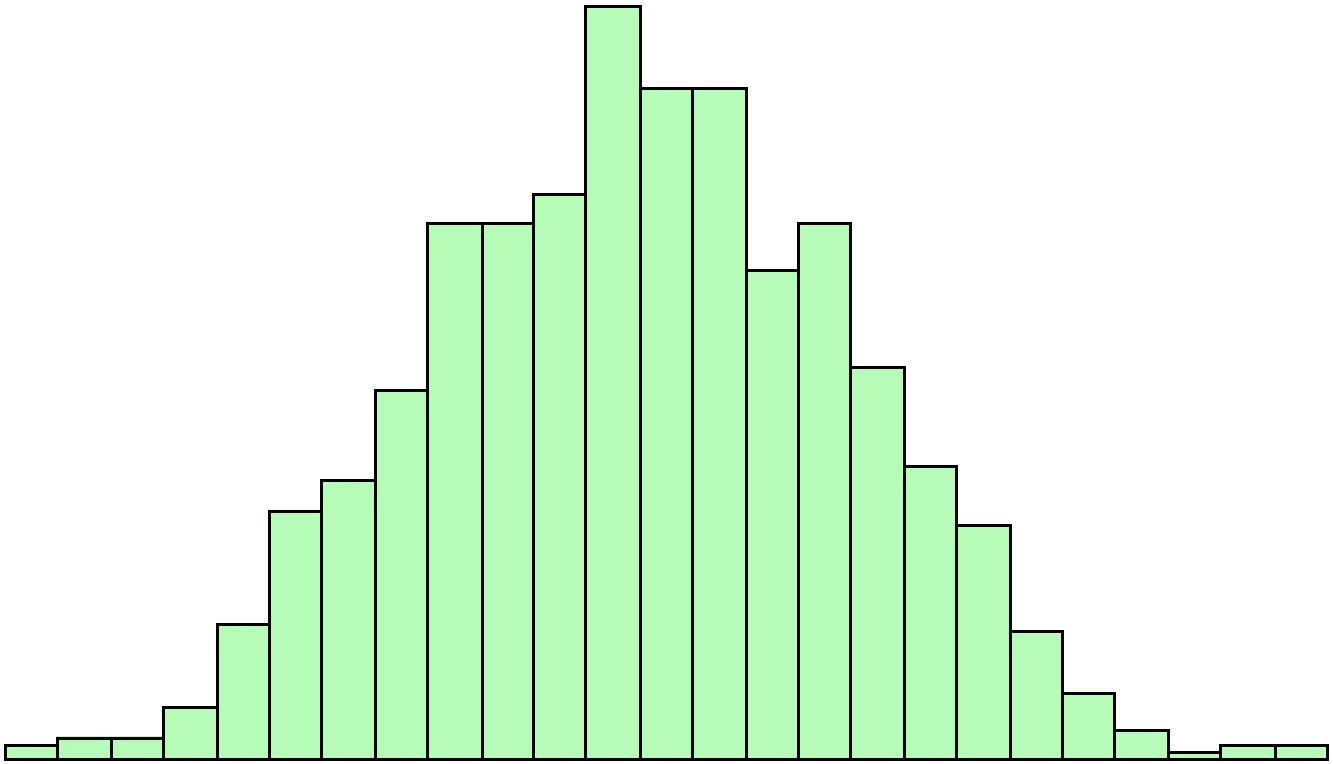
The analysis of traits whose variation is determined by both a number of genes and environmental factors

Phenotype is highly uninformative as to underlying genotype

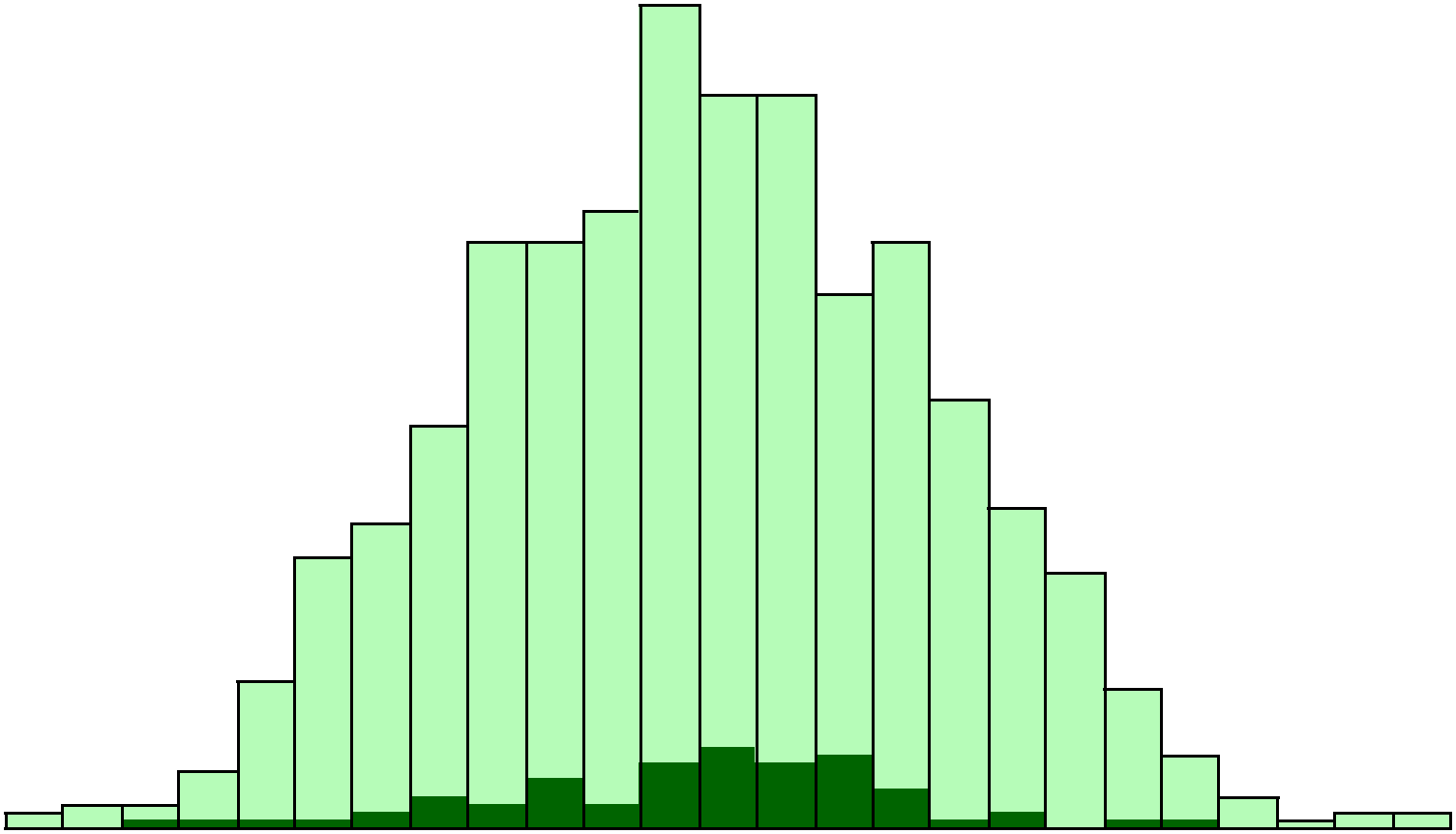
Complex (or Quantitative) trait

- No (apparent) simple Mendelian basis for variation in the trait
- May be a single gene strongly influenced by environmental factors
- May be the result of a number of genes of equal (or differing) effect
- Most likely, a combination of both multiple genes and environmental factors
- Example: Blood pressure, cholesterol levels
  - Known genetic and environmental risk factors
- Molecular traits can also be quantitative traits
  - mRNA level on a microarray analysis
  - Protein spot volume on a 2-D gel

Phenotypic distribution of a trait



Consider a specific locus influencing the trait



For this locus, mean phenotype = 0.15, while overall mean phenotype = 0

Goals of Quantitative Genetics

- Partition total trait variation into genetic (nature) vs. environmental (nurture) components
- Predict resemblance between relatives
  - If a sib has a disease/trait, what are your odds?
- Find the underlying loci contributing to genetic variation
  - QTL -- quantitative trait loci
- Deduce molecular basis for genetic trait variation
- eQTLs -- expression QTLs, loci with a quantitative influence on gene expression
  - e.g., QTLs influencing mRNA abundance on a microarray

Dichotomous (binary) traits

Presence/absence traits (such as a disease) can (and usually do) have a complex genetic basis

Consider a disease susceptibility (DS) locus underlying a disease, with alleles D and d, where allele D significantly increases your disease risk

In particular, Pr(disease | DD) = 0.5, so that the penetrance of genotype DD is 50%

Suppose Pr(disease | Dd) = 0.2, Pr(disease | dd) = 0.05

dd individuals can rarely display the disease, largely because of exposure to adverse environmental conditions

dd individuals can give rise to phenocopies 5% of the time, showing the disease but not as a result of carrying the risk allele

If freq(d) = 0.9, what is Prob (DD | show disease) ?

$$\text{freq(disease)} = 0.1^2*0.5 + 2*0.1*0.9*0.2 + 0.9^2*0.05$$
$$= 0.0815$$

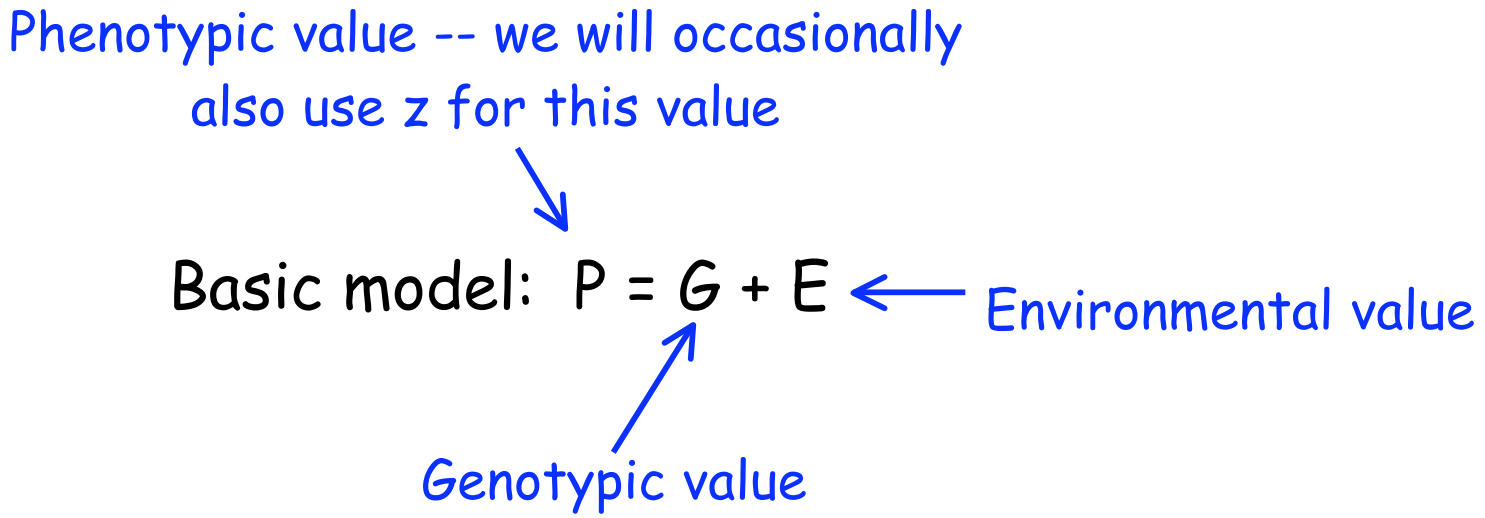
From Bayes' theorem,

$$\Pr(DD \mid \text{disease}) = \Pr(\text{disease} \mid DD)*\Pr(DD)/\text{Prob(disease)}$$
$$= 0.1^2*0.5 / 0.0815 = 0.06\ (6\ \%)$$

Pr(Dd | disease) = 0.442, Pr(dd | disease) = 0.497

Thus about 50% of the diseased individuals are phenocopies

Basic model of Quantitative Genetics



G = average phenotypic value for that genotype
if we are able to replicate it over the universe
of environmental values, G = E[P]

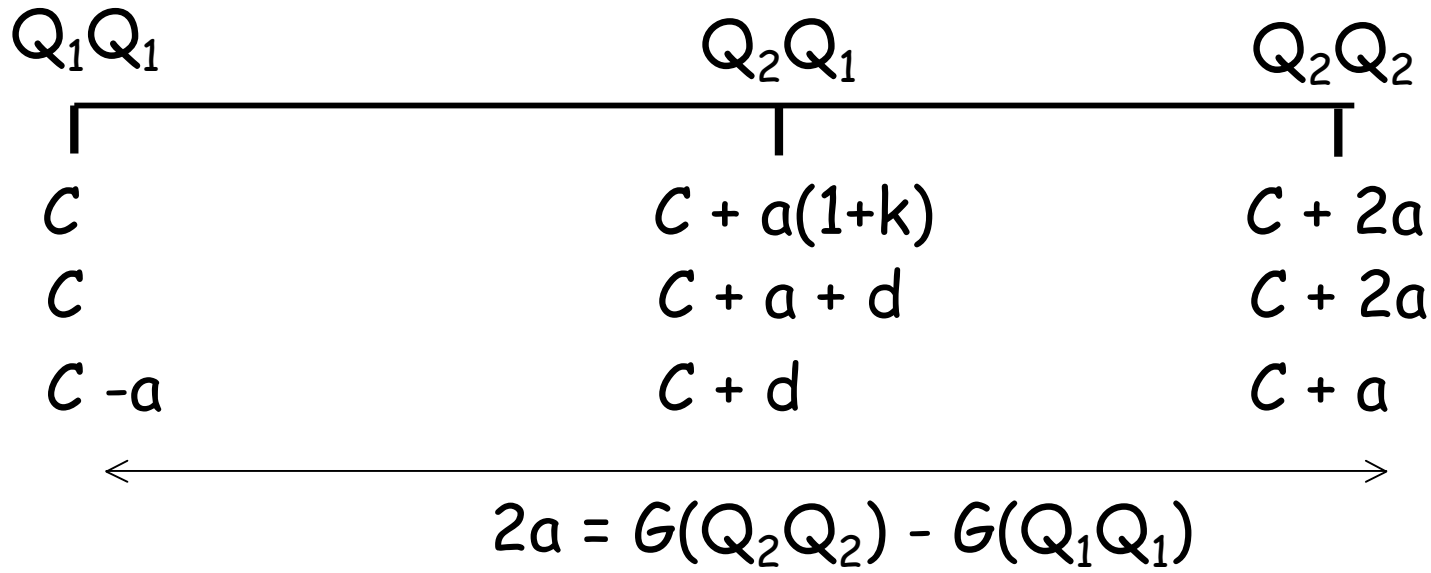
Basic model of Quantitative Genetics

Basic model: $P = G + E$

G = average phenotypic value for that genotype if we are able to replicate it over the universe of environmental values, $G = E[P]$

$G \times E$ interaction --- G values are different across environments. Basic model now becomes $P = G + E + GE$

Contribution of a locus to a trait



d measures dominance, with d = 0 if the heterozygote is exactly intermediate to the two homozygotes

$$d = ak = G(Q_1Q_2) - [G(Q_2Q_2) + G(Q_1Q_1)]/2$$

$k = d/a$ is a scaled measure of the dominance

Example: Apolipoprotein E & Alzheimer's

Genotype	ee	Ee	EE
Average age of onset	68.4	75.5	84.3

$2a = G(EE) - G(ee) = 84.3 - 68.4 \rightarrow a = 7.95$

$ak = d = G(Ee) - [\, G(EE)+G(ee)]/2 = -0.85$

$k = d/a = 0.10$ Only small amount of dominance

Example: Booroola (B) gene

Genotype	bb	Bb	BB
Average Litter size	1.48	2.17	2.66

$2a = G(BB) - G(bb) = 2.66 - 1.46 \rightarrow a = 0.59$

$ak = d = G(Bb) - [G(BB) + G(bb)]/2 = 0.10$

$k = d/a = 0.17$

Fisher's (1918) Decomposition of *G*

One of Fisher's key insights was that the genotypic value consists of a fraction that can be passed from parent to offspring and a fraction that cannot.

In particular, under sexual reproduction, parents only pass along SINGLE ALLELES to their offspring

Consider the genotypic value G_{ij} resulting from an A_iA_j individual

$$G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

Average contribution to genotypic value for allele i

Mean value, with $\mu_G = \sum G_{ij} \cdot \mathrm{freq}(Q_iQ_j)$

$$G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

Since parents pass along single alleles to their offspring, the α_i (the average effect of allele i) represent these contributions

The average effect for an allele is POPULATION-SPECIFIC, as it depends on the types and frequencies of alleles that it pairs with

The genotypic value predicted from the individual allelic effects is thus $\widehat{G}_{ij} = \mu_G + \alpha_i + \alpha_j$

$$G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

The genotypic value predicted from the individual allelic effects is thus $\widehat{G}_{ij} = \mu_G + \alpha_i + \alpha_j$

Dominance deviations --- the difference (for genotype A_iA_j) between the genotypic value predicted from the two single alleles and the actual genotypic value,

$$G_{ij} - \widehat{G}_{ij} = \delta_{ij}$$

Fisher's decomposition is a Regression

$$G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

Predicted value

Residual error

A notational change clearly shows this is a regression,

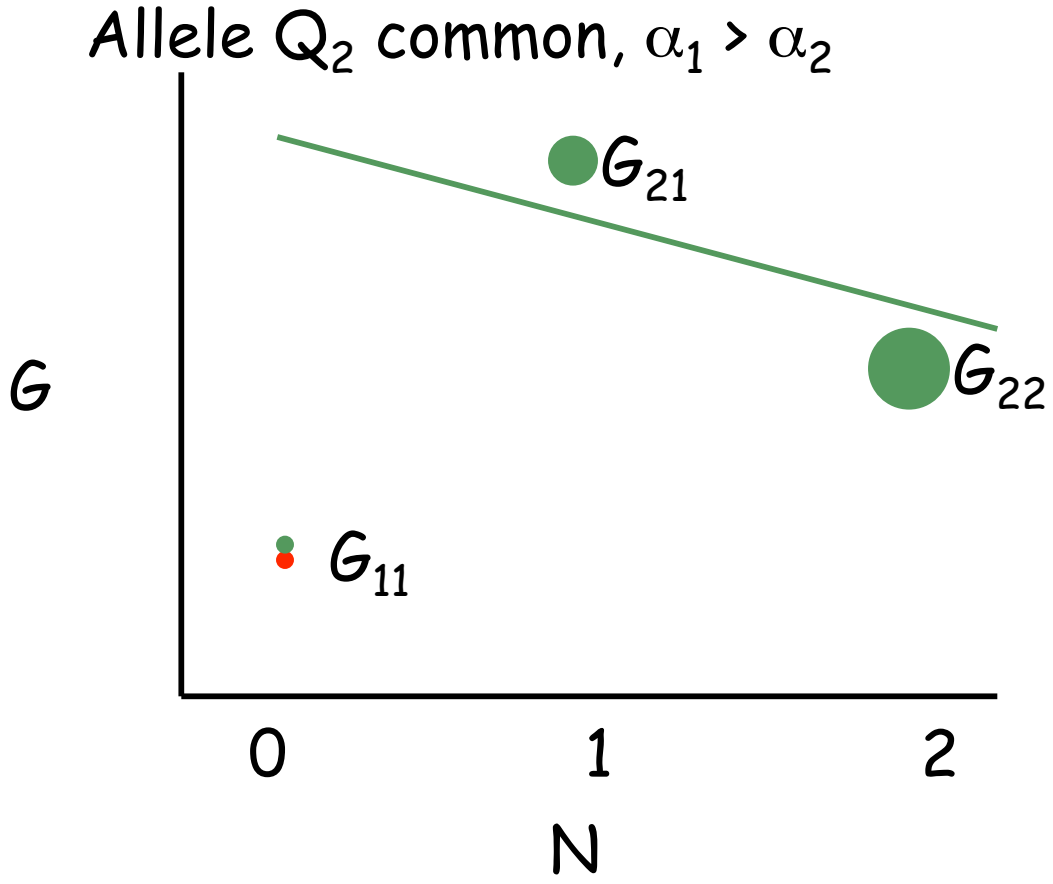
$$G_{ij} = \mu_G + 2\alpha_1 + (\alpha_2 - \alpha_1)N + \delta_{ij}$$

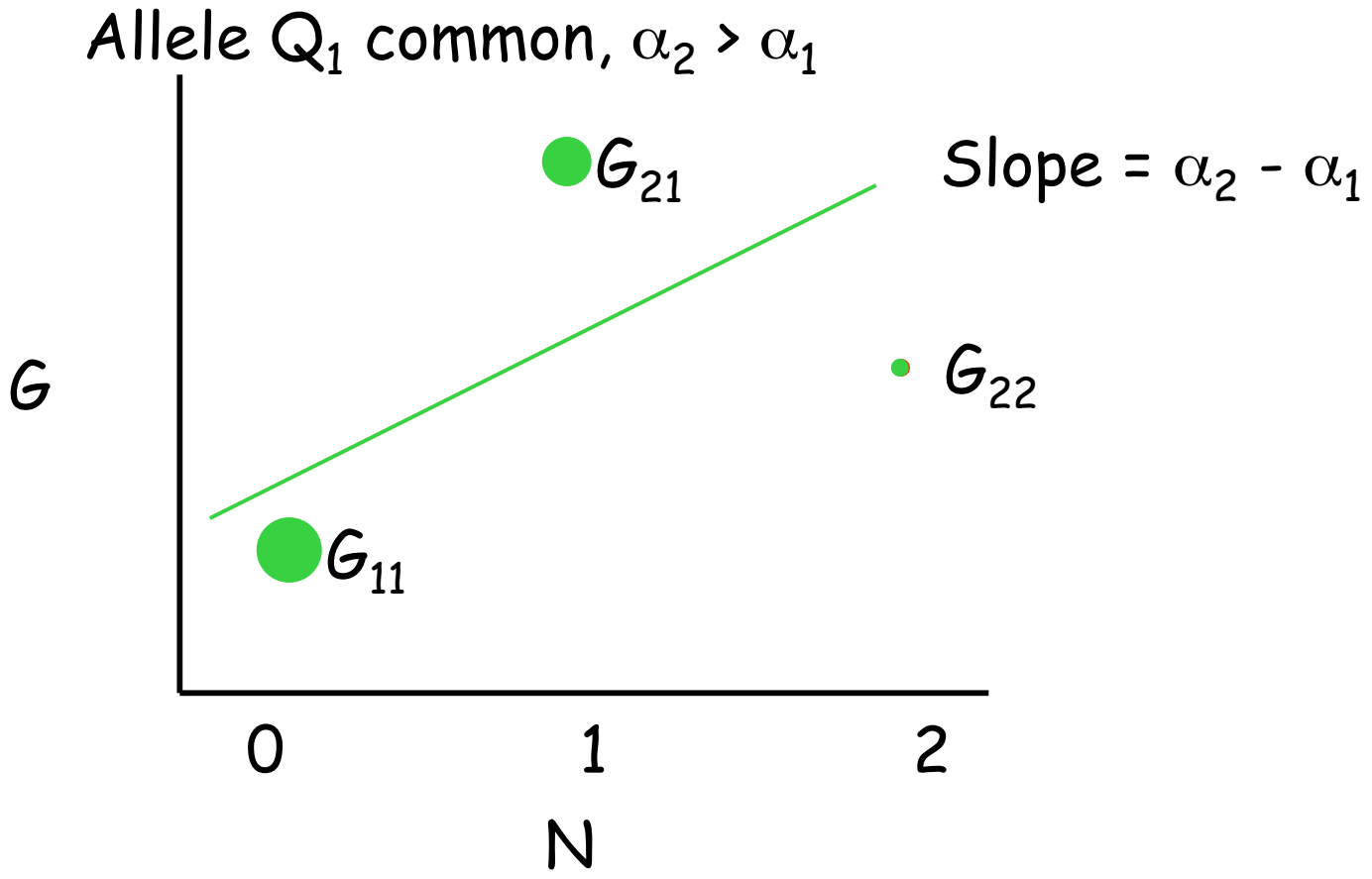
Independent (predictor) variable N = # of Q_2 alleles

$$G_{ij} = \mu_G + 2\alpha_1 + (\alpha_2 - \alpha_1)N + \delta_{ij}$$

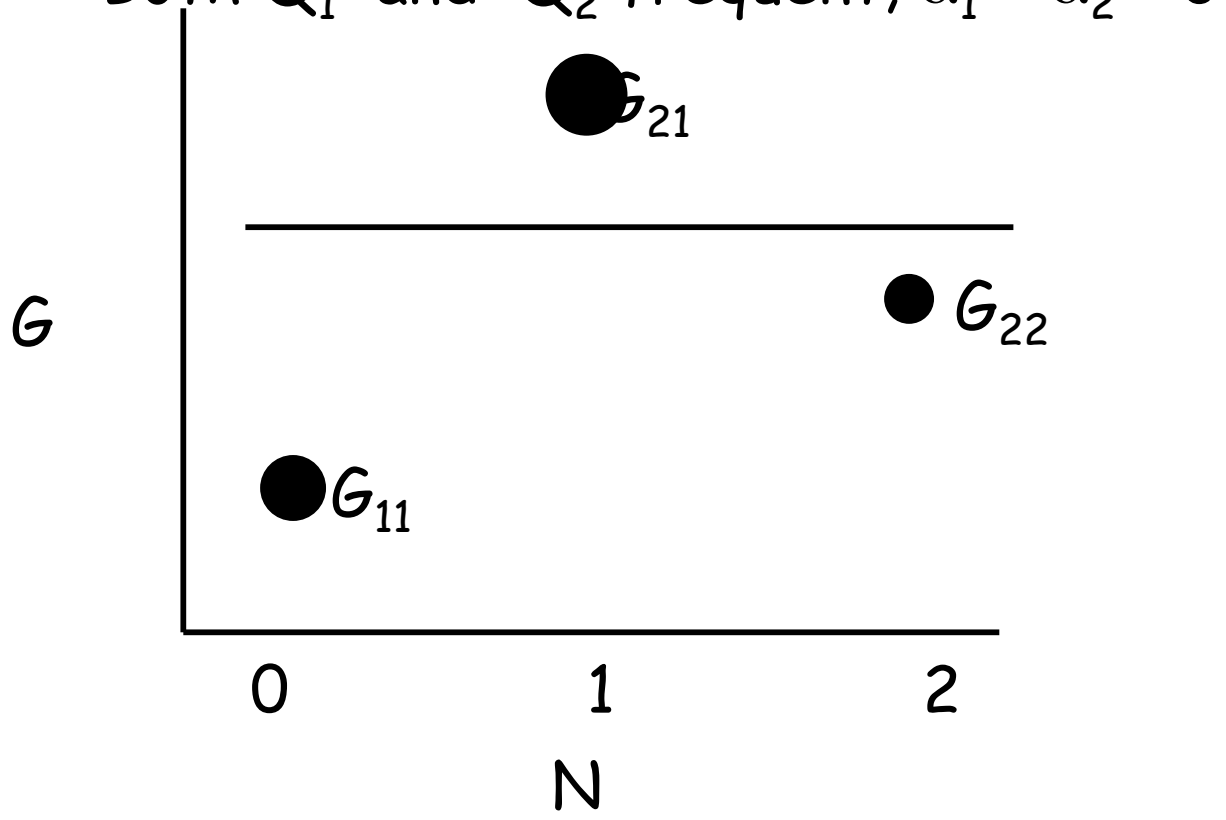
Intercept Regression slope

$$2\alpha_1 + (\alpha_2 - \alpha_1)N = \begin{cases} 2\alpha_1 & \text{for} N = 0,\ \text{e.g},\ Q_1Q_1 \\ \alpha_1 + \alpha_2 & \text{for} N = 1,\ \text{e.g},\ Q_1Q_2 \\ 2\alpha_2 & \text{for} N = 2,\ \text{e.g},\ Q_2Q_2 \end{cases}$$


Allele Q_2 common, $\alpha_1 > \alpha_2$
G_{21}
G_{22}
G
G_{11}
0
1
2
N


Allele Q_1 common, $\alpha_2 > \alpha_1$
G_{21}
Slope = $\alpha_2 - \alpha_1$
G_{22}
G
G_{11}
0
1
2
N

Both Q_1 and Q_2 frequent, $\alpha_1 = \alpha_2 = 0$



Consider a diallelic locus, where p_1 = freq(Q_1)

Genotype	Q_1Q_1	Q_2Q_1	Q_2Q_2
Genotypic value	0	a(1+k)	2a

Mean $\mu_G = 2p_2\, a(1 + p_1 k)$

Allelic effects

$$\alpha_2 = p_1 a\,[\,1 + k\,(p_1 - p_2)\,]$$

$$\alpha_1 = -\,p_2 a\,[\,1 + k\,(p_1 - p_2)\,]$$

Dominance deviations $\delta_{ij} = G_{ij} - \mu_G - \alpha_i - \alpha_j$

Average effects and Additive Genetic Values

The α values are the average effects of an allele

A key concept is the Additive Genetic Value (A) of an individual

$$A\left(G_{ij}\right) = \alpha_i + \alpha_j$$

$$A = \sum_{k=1}^{n} \left(\alpha_i^{(k)} + \alpha_k^{(k)}\right)$$

A is called the Breeding value or the Additive genetic value

$$A = \sum_{k=1}^{n} \left(\alpha_i^{(k)} + \alpha_k^{(k)} \right)$$

Why all the fuss over A?

Suppose father has A = 10 and mother has A = -2 for (say) blood pressure

Expected blood pressure in their offspring is (10-2)/2 = 4 units above the population mean. Offspring A = average of parental A's

KEY: parents only pass single alleles to their offspring. Hence, they only pass along the A part of their genotypic value G

Genetic Variances

$$G_{ij} = \mu_g + (\alpha_i + \alpha_j) + \delta_{ij}$$

$$\sigma^2(G) = \sum_{k=1}^{n} \sigma^2(\alpha_i^{(k)} + \alpha_j^{(k)}) + \sum_{k=1}^{n} \sigma^2(\delta_{ij}^{(k)})$$

$$\sigma^2(G) = \sigma^2(\mu_g + (\alpha_i + \alpha_j) + \delta_{ij}) = \sigma^2(\alpha_i + \alpha_j) + \sigma^2(\delta_{ij})$$

As Cov(α,δ) = 0

Genetic Variances

$$\sigma^2(G) = \sum_{k=1}^{n} \sigma^2(\alpha_i^{(k)} + \alpha_j^{(k)}) + \sum_{k=1}^{n} \sigma^2(\delta_{ij}^{(k)})$$

Additive Genetic Variance
(or simply Additive Variance)

Dominance Genetic Variance
(or simply dominance variance)

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2$$

Key concepts (so far)

- α_i = average effect of allele i
  - Property of a single allele in a particular population (depends on genetic background)
- *A* = Additive Genetic Value (*A*)
  - *A* = sum (over all loci) of average effects
  - Fraction of *G* that parents pass along to their offspring
  - Property of an Individual in a particular population
- Var(*A*) = additive genetic variance
  - Variance in additive genetic values
  - Property of a population
- Can estimate *A* or Var(*A*) without knowing any of the underlying genetical detail (forthcoming)

$$\sigma_A^2 = 2E[\alpha^2] = 2\sum_{i=1}^{m} \alpha_i^2 p_i$$

Q_1Q_1	Q_1Q_2	Q_2Q_2
0	a(1+k)	2a

Since $E[\alpha] = 0$,

$Var(\alpha) = E[(\alpha - \mu_a)^2] = E[\alpha^2]$

One locus, 2 alleles: $\sigma_A^2 = 2p_1 p_2 a^2[1 + k(p_1 - p_2)]^2$

Dominance alters additive variance

When dominance present, Additive variance is an asymmetric function of allele frequencies

Dominance variance

Q_1Q_1	Q_1Q_2	Q_2Q_2
0	a(1+k)	2a

$$\sigma_D^2 = \quad E[\delta^2] = \sum_{i=1}^{m}\sum_{j=1}^{m} \delta_{ij}^2 \, p_i \, p_j$$

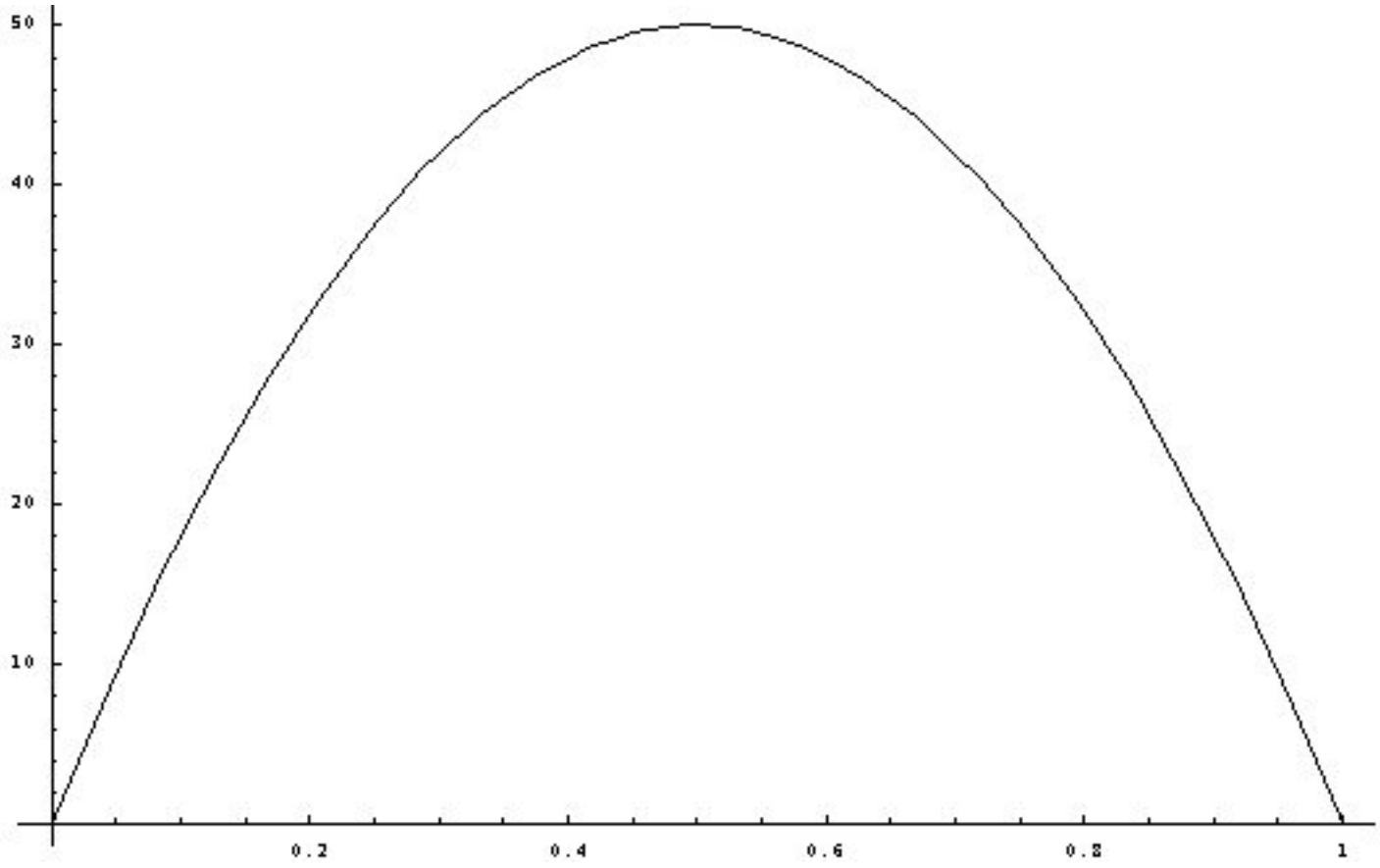
Equals zero if k = 0

One locus, 2 alleles: $\sigma_D^2 = (2p_1 \, p_2 \, ak)^2$

This is a symmetric function of allele frequencies

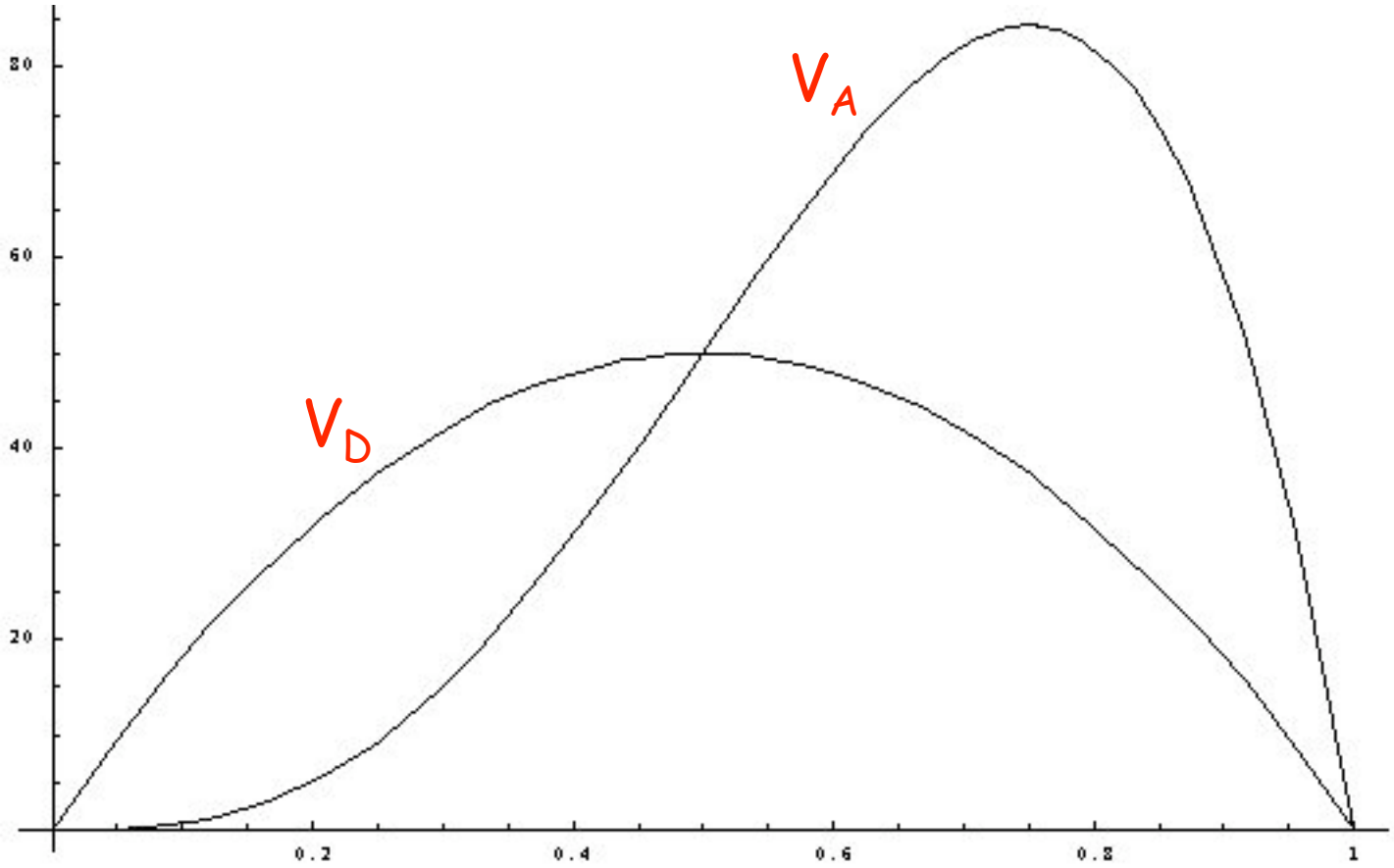
Can also be expressed in terms of d = ak

Additive variance, V_A, with no dominance (k = 0)



Allele frequency, p

Complete dominance (k = 1)



Allele frequency, p

Epistasis

$$
\begin{aligned}
G_{ijkl} = \mu_G &+ (\alpha_i + \alpha_j + \alpha_k + \alpha_l) + (\delta_{ij} + \delta_{kj}) \\
&+ (\alpha\alpha_{ik} + \alpha\alpha_{il} + \alpha\alpha_{jk} + \alpha\alpha_{jl}) \\
&+ (\alpha\delta_{ikl} + \alpha\delta_{jkl} + \alpha\delta_{kij} + \alpha\delta_{lij}) \\
&+ (\delta\delta_{ijkl}) \\
= \mu_G &+ A + D + AA + AD + DD
\end{aligned}
$$

These components are defined to be uncorrelated, (or orthogonal), so that

$$\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_{AA} + \sigma^2_{AD} + \sigma^2_{DD}$$

$$
\begin{aligned}
G_{ijkl} &= \mu_G + (\alpha_i + \alpha_j + \alpha_k + \alpha_l) + (\delta_{ij} + \delta_{kj}) \\
&\quad + (\alpha\alpha_{ik} + \alpha\alpha_{il} + \alpha\alpha_{jk} + \alpha\alpha_{jl}) \\
&\quad + (\alpha\delta_{ikl} + \alpha\delta_{jkl} + \alpha\delta_{kij} + \alpha\delta_{lij}) \\
&\quad + (\delta\delta_{ijkl}) \\
&= \mu_G + A + D + AA + AD + DD
\end{aligned}
$$

Additive x Additive interactions -- $\alpha\alpha$, AA
interactions between a single allele
at one locus with a single allele at another

Additive x Dominance interactions -- $\alpha\delta$, AD
interactions between an allele at one
locus with the genotype at another, e.g.
allele A_i and genotype B_{kj}

Dominance x dominance interaction --- $\delta\delta$, DD
the interaction between the dominance
deviation at one locus with the dominance
deviation at another.