Mapping Human Quantitative Trait Locus

(Using Sib-pair Data)

Soham Bonnerjee

Roll No. BS1609

B.Stat 3rd Year

Indian Statistical Institute, Kolkata

March 28, 2020

Content

Quantitative Traits are those *measurable* traits that doesn't behave according to *Mendel's Laws*. Some examples include *blood pressure*, *height*, etc. These are known to be primarily determined by inherited genetic factors, but it's difficult to demonstrate without doubt that a particular genetic component is involved in determination of a qualitative trait.

Introduction
Quantitative Trait Locus Mapping

Quantitative Traits are those *measurable* traits that doesn't behave according to *Mendel's Laws*. Some examples include *blood pressure*, *height*, etc. These are known to be primarily determined by inherited genetic factors, but it's difficult to demonstrate without doubt that a particular genetic component is involved in determination of a qualitative trait.

The primarily used methods of finding evidence of genetic control for a quantitative trait is to determine the association between the *trait locus*, and the *marker locus*, called the *QTL Mapping*. The method that we focus on, was developed by *Haseman* and *Elston* in 1970, and uses *Sibling-Pair Data*.

Definitions and Pre-requisites

We define the following technical terms in easy words, so as all of us can understand the rest of the presentation.

Definitions and Pre-requisites

We define the following technical terms in easy words, so as all of us can understand the rest of the presentation.

Locus

- Any position on the genome/choromosome. Consists of two alleles, from the set of alleles (A, a)

Definitions and Pre-requisites

We define the following technical terms in easy words, so as all of us can understand the rest of the presentation.

Locus

- Any position on the genome/choromosome. Consists of two alleles, from the set of alleles (A, a)

Allele

- Different variations of DNA nucleotide at a locus. In our case they are of two types, A and a.

Definitions and Pre-requisites

We define the following technical terms in easy words, so as all of us can understand the rest of the presentation.

Locus

- Any position on the genome/choromosome. Consists of two alleles, from the set of alleles (A, a)

Allele

- Different variations of DNA nucleotide at a locus. In our case they are of two types, A and a.

Genotype

- Combination of the two alleles at the locus inhabited from two parents.
- If Father has genotype A_1A_2, and mother has genotype B_1B_2, the offpsriing will have genotype of the form A_iB_j, $i, j = 1, 2$, with each genotype having equal probability of occurring at that locus.

Definition and Pre-requisites

Marker Locus

- The locus whose position and genotypes we know.

Definition and Pre-requisites

Marker Locus

- The locus whose position and genotypes we know.

Linkage

- When marker locus and trait locus are in close proximity, the usual Mendelian recombination does NOT occur, instead the alleles (of the trait and marker locus) tend to be inherited together. This is called *Linkage*.

Recombination Fraction

- The probability of recombination between genes (denoted by θ). If their is no linkage, i.e trait and marker locus is independent, then $\theta = 0.5$

Let n sib-pairs be available, and let Y_{1j} and Y_{2j} be the observed quantitative trait values for the first and second sibs, in the j^{th} sib pair. For $i = 1, 2$, $j = 1(1)n$, we assume the following linear model:

$$Y_{ij} = \mu + x_{ij} + \epsilon_{ij}$$

where μ is the overall mean effect, and x_{ij} and ϵ_{ij} are the *genetic* and *environmental* effects on the trait.
Going forward, we make some assumptions to simplify numerical calculations.

Assumptions

- Only one *biallelic locus* determines x_{ij}. Suppose these two alleles are A and a, with allele frequencies p and q, $p + q = 1$, ($p \geq 0.5$) in the offspring generation.
- We assume random mating in the population, i.e P(an offspring has AA genotype)=p^2, P(an offspring has Aa genotype)=$2pq$, P(an offspring has aa genotype)=q^2. This is also known as *Hardy-Weinberg Equilibrium*.

Thus, with this assumptions,we see that x_ij's depend only on the genotype of the offspring, and hence we can write the genotypic values:

$$x_{ij} = \begin{cases} a, & \text{for } AA \text{ individual} \\ b, & \text{for } Aa \text{ individual} \\ c & \text{for } aa \text{ individual} \end{cases}$$

Assuming no dominance between the alleles A and a, this is equivalent to :

$$x_{ij} = \begin{cases} \alpha, & \text{for } AA \text{ individual} \\ 0, & \text{for } Aa \text{ individual} \\ -\alpha & \text{for } aa \text{ individual} \end{cases}$$

where α can be thought of as the *marginal affect* of the Quantitative Trait Locus.

Thus, with this assumptions,we see that x_ij's depend only on the genotype of the offspring, and hence we can write the genotypic values:

$$x_{ij} = \begin{cases} a, & \text{for } AA \text{ individual} \\ b, & \text{for } Aa \text{ individual} \\ c & \text{for } aa \text{ individual} \end{cases}$$

Assuming no dominance between the alleles A and a, this is equivalent to :

$$x_{ij} = \begin{cases} \alpha, & \text{for } AA \text{ individual} \\ 0, & \text{for } Aa \text{ individual} \\ -\alpha & \text{for } aa \text{ individual} \end{cases}$$

where α can be thought of as the *marginal affect* of the Quantitative Trait Locus. Denote by σ_x^2 the genetic variance $var(x_{ij})$ at this trait locus. Then:

$$\sigma_x^2 = E(x_{ij}^2) - E(x_{ij})^2 = \alpha^2(p^2 + q^2) - (\alpha(p^2 - q^2))^2 = 2pq\alpha^2$$

Thus, with this assumptions,we see that x_ij's depend only on the genotype of the offspring, and hence we can write the genotypic values:

$$x_{ij} = \begin{cases} a, & \text{for } AA \text{ individual} \\ b, & \text{for } Aa \text{ individual} \\ c & \text{for } aa \text{ individual} \end{cases}$$

Assuming no dominance between the alleles A and a, this is equivalent to :

$$x_{ij} = \begin{cases} \alpha, & \text{for } AA \text{ individual} \\ 0, & \text{for } Aa \text{ individual} \\ -\alpha & \text{for } aa \text{ individual} \end{cases}$$

where α can be thought of as the *marginal affect* of the Quantitative Trait Locus. Denote by σ_x^2 the genetic variance $var(x_{ij})$ at this trait locus. Then:

$$\sigma_x^2 = E(x_{ij}^2) - E(x_{ij})^2 = \alpha^2(p^2 + q^2) - (\alpha(p^2 - q^2))^2 = 2pq\alpha^2$$

We also let $\epsilon_j = \epsilon_{1j} - \epsilon_{2j}$. Let $E(\epsilon_j^2) = \sigma^2$

To estimate the trait values for j^{th} pair, we try to combine the paired data points. Let $Y_j = (Y_{1j} - Y_{2j})^2$ be the squared pair difference (of trait values) for j^{th} sibling pair.
Let π_{jt} be the proportion of trait-locus alleles inherited *identical by descent* in the j^{th} sibling-pair. This only means that the each of the sibling pairs inherit $2\pi_{jt}$ alleles from the same parent. Note that π_{jt} can take values either 0, $\frac{1}{2}$ or 1.

We give the the conditional distribution of the sib pairs given π_{jt} in Table 1.

Sib-pair	Y_j	Probability of genotype given π_{jt}		
		$\pi_{jt} = 0$	$\pi_{jt} = \frac{1}{2}$	$\pi_{jt} = 1$
AA-AA	ϵ_j^2	p^4	p^3	p^2
aa-aa	ϵ_j^2	q^4	q^3	q^2
Aa-Aa	ϵ_j^2	$4p^2q^2$	pq	$2pq$
AA-Aa	$(\alpha + \epsilon_j)^2$	$2p^3q$	p^2q	0
Aa-AA	$(-\alpha + \epsilon_j)^2$	$2p^3q$	p^2q	0
Aa-aa	$(\alpha + \epsilon_j)^2$	$2pq^3$	pq^2	0
aa-Aa	$(-\alpha + \epsilon_j)^2$	$2pq^3$	pq^2	0
AA-aa	$(2\alpha + \epsilon_j)^2$	p^2q^2	0	0
aa-AA	$(-2\alpha + \epsilon_j)^2$	p^2q^2	0	0

Table: Distribution of sib-pair genotypes given π_{jt}

The Set-Up
Conditional Expectations of Squared Pair Differences

Using Table 1 and the fact that given sibling genotype, Y_j and π_{jt} are independent, we find the conditional expectaion of Y_j given π_{jt}:

$$\begin{aligned}
E(Y_j|\pi_{jt}=1) &= E(E(Y_j|\pi_{jt}=1, \text{Sibling Genotype})) \\
&= \sum E(Y_j|\pi_{jt}=1, \text{Sibling Genotype}) P(\text{Sibling Genotype}|\pi_{jt}=1) \\
&= \sum E(Y_j|\text{Sibling Genotype}) P(\text{Sibling Genotype}|\pi_{jt}=1) \\
&= E(\epsilon_j^2) \times (p^2+2pq+q^2) \\
&= \sigma^2
\end{aligned}$$

Similarly, using $\sigma_x^2 = 2pq\alpha^2$,

$$\begin{aligned}
E(Y_j|\pi_{jt}=\frac{1}{2}) &= \sigma^2+\sigma_x^2 \\
E(Y_j|\pi_{jt}=0) &= \sigma^2+2\sigma_x^2
\end{aligned}$$

This can be written as

$$E(Y_j|\pi_{jt}) = (\sigma^2+2\sigma_x^2) - 2\sigma_x^2\pi_{jt} \cdots\cdots (1)$$

Regresion Line

If we knew π_{jt}, then from the above equation (1), we could have formed a regression model with Y_j and π_{jt}.

Regresion Line

If we knew π_{jt}, then from the above equation (1), we could have formed a regression model with Y_j and π_{jt}.

But we don't know π_{jt}, since we don't know the location of trait locus.

Regresion Line

If we knew π_{jt}, then from the above equation (1), we could have formed a regression model with Y_j and π_{jt}.

But we don't know π_{jt}, since we don't know the location of trait locus.

At this stage we introduce π_{jm}, the proportion of alleles shared i.b.d at the marker locus in the j^{th} sib-pair.

Regresion Line

If we knew π_{jt}, then from the above equation (1), we could have formed a regression model with Y_j and π_{jt}.

But we don't know π_{jt}, since we don't know the location of trait locus.

At this stage we introduce π_{jm}, the proportion of alleles shared i.b.d at the marker locus in the j^{th} sib-pair.

Do we have the exact values of π_{jm}?

- No. Given the complete information about the parental and sibling genotypes at the marker locus, still we can't exactly point out which allele did the father give, or which allele did the mother give, (except in some special cases), so we can't exactly know the value of π_{jm}.

Regression Line
Estimation of π_j

Let f_{ji} be the probability that the j^{th} sib-pair has i alleles i.b.d at a locus, conditioned on I_m, the complete information available on the sib-pair genotypes and parental genotypes at the marker locus.
i.e $f_{ji} = P(\pi_{jm} = \frac{i}{2} | I_m)$

Regression Line
Estimation of π_j

Let f_{ji} be the probability that the j^{th} sib-pair has i alleles i.b.d at a locus, conditioned on I_m, the complete information available on the sib-pair genotypes and parental genotypes at the marker locus.
i.e $f_{ji} = P(\pi_{jm} = \frac{i}{2}|I_m)$

Estimation of π_j

- $\hat{\pi}_{jm} = E(\pi_{jm}|I_m) = f_{j2} + \frac{1}{2}f_{j1}$.
- $\hat{\pi}_{jm}$ is the *Bayes Estimate* of π_{jm} under squared loss error.

We calculate $\hat{\pi}_j$ for every parental and sibling pair, when sibling and parental information available.

Regression Line

Estimation of π_j

	Sib-Pair	Prob.	f_{j0}	f_{j1}	f_{j2}	$\hat{\pi}_j$
$AA \times AA$	AA-AA	p^4	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
$aa \times aa$	aa-aa	q^4	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
$AA \times aa$	Aa-Aa	$2p^2q^2$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
$AA \times Aa$	AA-AA	p^3q	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
	AA-Aa	$2p^3q$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{4}$
	Aa-Aa	p^3q	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
$aa \times Aa$	aa-aa	pq^3	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
	aa-Aa	$2pq^3$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{4}$
	Aa-Aa	pq^3	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
$Aa \times Aa$	AA-AA	$\frac{p^2q^2}{4}$	0	0	1	1
	aa-aa	$\frac{p^2q^2}{4}$	0	0	1	1
	AA-aa	$\frac{p^2q^2}{2}$	1	0	0	0
	AA-Aa	p^2q^2	0	1	0	$\frac{1}{2}$
	aa-Aa	p^2q^2	0	1	0	$\frac{1}{2}$
	Aa-Aa	p^2q^2	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$

Table: Estimation of π_j

Regression Line

We compute $\hat{\pi}_{jm}$ from Table 2, and use it for regression.

Regression Line

We compute $\hat{\pi}_{jm}$ from Table 2, and use it for regression.

So we need to compute $E(Y_j|\hat{\pi}_{jm})$, where $\hat{\pi}_{jm} = f_{j1} + \frac{1}{2}f_{j2}$, where f_{ji}'s are the probabilities that j^{th} sibling-pair share i alleles i.b.d at the marker locus.

Regression Line

Assuming Linkage Equilibrium between trait and marker locus, since given π_{jt}, Y_j and $\hat{\pi}_{jm}$ are independent , hence, conditioning by π_{jt}:

$$E(Y_j|\hat{\pi}_{jm}) = \sum_{\pi_{jt}} E(Y_j|\pi_{jt})P(\pi_{jt}|\hat{\pi}_{jm})$$

Regression Line

Assuming Linkage Equilibrium between trait and marker locus, since given π_{jt}, Y_j and $\hat{\pi}_{jm}$ are independent , hence, conditioning by π_{jt}:

$$E(Y_j|\hat{\pi}_{jm}) = \sum_{\pi_{jt}} E(Y_j|\pi_{jt})P(\pi_{jt}|\hat{\pi}_{jm})$$

Now for fixed π_{jm}, π_{jt} and $\hat{\pi}_{jm}$ are independent, hence conditioning on π_{jm}:

$$E(Y_j|\hat{\pi}_{jm}) = \sum_{\pi_{jt}} E(Y_j|\pi_{jt})P(\pi_{jt}|\hat{\pi}_{jm}) = \sum_{\pi_{jt}} E(Y_j|\pi_{jt}) \sum_{\pi_{jm}} P(\pi_{jt}|\pi_{jm})P(\pi_{jm}|\hat{\pi}_{jm}) \cdots (2)$$

Regression Line

Assuming Linkage Equilibrium between trait and marker locus, since given π_{jt}, Y_j and $\hat{\pi}_{jm}$ are independent , hence, conditioning by π_{jt}:

$$E(Y_j|\hat{\pi}_{jm}) = \sum_{\pi_{jt}} E(Y_j|\pi_{jt})P(\pi_{jt}|\hat{\pi}_{jm})$$

Now for fixed π_{jm}, π_{jt} and $\hat{\pi}_{jm}$ are independent, hence conditioning on π_{jm}:

$$E(Y_j|\hat{\pi}_{jm}) = \sum_{\pi_{jt}} E(Y_j|\pi_{jt})P(\pi_{jt}|\hat{\pi}_{jm}) = \sum_{\pi_{jt}} E(Y_j|\pi_{jt}) \sum_{\pi_{jm}} P(\pi_{jt}|\pi_{jm})P(\pi_{jm}|\hat{\pi}_{jm}) \cdots (2)$$

Hence we need to find two joint distributions: 1) $(\hat{\pi}_{jm}, \pi_{jm})$, 2) (π_{jt}, π_{jm}).
Let the recombination fraction between the trait and the marker locus be θ.

Regression Line

$\hat{\pi}_{jm}$	π_{jm}			Total
	0	$\frac{1}{2}$	1	
0	$\frac{p^2q^2}{2}$	0	0	$\frac{p^2q^2}{2}$
$\frac{1}{4}$	$p^3q + pq^3$	$p^3q + pq^3$	0	$2(p^3q + pq^3)$
$\frac{1}{2}$	$\frac{1}{4}(p^4 + 4p^2q^2 + q^4)$	$\frac{1}{2}(p^4 + 6p^2q^2 + q^4)$	$\frac{1}{4}(p^4 + 4p^2q^2 + q^4)$	$(p^4 + 5p^2q^2 + q^4)$
$\frac{3}{4}$	0	$p^3q + pq^3$	$p^3q + pq^3$	$2(p^3q + pq^3)$
1	0	0	$\frac{p^2q^2}{2}$	$\frac{p^2q^2}{2}$
Total	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Table: Joint Distribution of $\hat{\pi}_{jm}$ and π_{jm} calculated from Table 2

Regression Line

π_{jt}	π_{jm}			Total
	0	$\frac{1}{2}$	1	
0	$\frac{\psi^2}{4}$	$\frac{(\psi)(1-\psi)}{2}$	$\frac{(1-\psi)^2}{4}$	$\frac{1}{4}$
$\frac{1}{2}$	$\frac{(\psi)(1-\psi)}{2}$	$\frac{(1-2\psi+2\psi^2)}{2}$	$\frac{(\psi)(1-\psi)}{2}$	$\frac{1}{2}$
1	$\frac{(1-\psi)^2}{4}$	$\frac{(\psi)(1-\psi)}{2}$	$\frac{\psi^2}{4}$	$\frac{1}{4}$
Total	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Table: Joint Distribution of π_{jt} and π_{jm}. $\psi = \theta^2 + (1-\theta)^2$

Regression Line

From equation (1) and (2), and table 3 and 4, we obtain:

$$E(Y_j|\hat{\pi}_{jm}=0) = \sigma^2[\frac{\frac{(1-\psi)^2}{4}}{\frac{1}{4}}] + (\sigma^2+\sigma_x^2)^2[\frac{\frac{(\psi)(1-\psi)}{2}}{\frac{1}{4}}] + (\sigma^2+2\sigma_x^2)^2[\frac{\frac{\psi^2}{4}}{\frac{1}{4}}]$$
$$= \sigma^2 + 2\psi\sigma_x^2$$

Regression Line

From equation (1) and (2), and table 3 and 4, we obtain:

$$E(Y_j|\hat{\pi}_{jm}=0) = \sigma^2[\frac{\frac{(1-\psi)^2}{4}}{\frac{1}{4}}] + (\sigma^2+\sigma_x^2)^2[\frac{\frac{(\psi)(1-\psi)}{2}}{\frac{1}{4}}] + (\sigma^2+2\sigma_x^2)^2[\frac{\frac{\psi^2}{4}}{\frac{1}{4}}]$$
$$= \sigma^2 + 2\psi\sigma_x^2$$

Similarly:

$$E(Y_j|\hat{\pi}_{jm}=\frac{1}{4}) = \sigma^2 + (\frac{1}{2}+\psi)\sigma_x^2$$
$$E(Y_j|\hat{\pi}_{jm}=\frac{1}{2}) = \sigma^2 + \sigma_x^2$$
$$E(Y_j|\hat{\pi}_{jm}=\frac{3}{4}) = \sigma^2 + (\frac{3}{2}-\psi)\sigma_x^2$$
$$E(Y_j|\hat{\pi}_{jm}=1) = \sigma^2 + 2(1-\psi)\sigma_x^2$$

Combining:

$$E(Y_j|\hat{\pi}_{jm}) = (\sigma^2 + 2\psi\sigma_x^2) + 2(1-2\psi)\sigma_x^2\hat{\pi}_{jm}$$
$$= [\sigma^2 + 2(1-2\theta+2\theta^2)\sigma_x^2] + [-2(1-2\theta)^2\sigma_x^2]\hat{\pi}_{jm}$$

Testing for Linkage

The linear model is : $Y_j = \beta_0 + \beta_1 \hat{\pi}_{jm} + e_j$ where e_j's are iid $N(0, \tau^2)$.
$E(Y_j|\hat{\pi}_{jm}) = \beta_0 + \beta_1 \hat{\pi}_{jm}$, so have $\beta_1 = -2(1-2\theta)^2\sigma_x^2$.

Testing for Linkage

The linear model is : $Y_j = \beta_0 + \beta_1 \hat{\pi}_{jm} + e_j$ where e_j's are iid $N(0, \tau^2)$.
$E(Y_j|\hat{\pi}_{jm}) = \beta_0 + \beta_1 \hat{\pi}_{jm}$, so have $\beta_1 = -2(1-2\theta)^2\sigma_x^2$.

Hence noting that for β_1 is an increasing function of θ for $\theta \leq 0.5$, and $\beta_1 = 0 \implies \theta = 0.5$, we have that a test for linkage at the trait locus (i.e $H_0 : \theta = 0.5$ vs $H_1 : \theta < 0.5$), is equivalent to the test for $H_0 : \beta_1 = 0$ vs $\beta_1 < 0$ which is the usual t-Test.

Testing for Linkage

The linear model is : $Y_j = \beta_0 + \beta_1 \hat{\pi}_{jm} + e_j$ where e_j's are iid $N(0, \tau^2)$.
$E(Y_j|\hat{\pi}_{jm}) = \beta_0 + \beta_1 \hat{\pi}_{jm}$, so have $\beta_1 = -2(1-2\theta)^2\sigma_x^2$.

Hence noting that for β_1 is an increasing function of θ for $\theta \leq 0.5$, and $\beta_1 = 0 \implies \theta = 0.5$, we have that a test for linkage at the trait locus (i.e $H_0 : \theta = 0.5$ vs $H_1 : \theta < 0.5$), is equivalent to the test for $H_0 : \beta_1 = 0$ vs $\beta_1 < 0$ which is the usual t-Test.
Our Test statistic is:

$$T = \frac{\hat{\beta}_1}{\hat{se}(\hat{\beta}_1)} \sim t_{n-2} \text{ under } H_0$$

where

$$\hat{se}(\hat{\beta}_1) = \sqrt{\frac{S_{11} R_0^2}{n-2}}$$

where, $S = (X'X)^{-1}$,X being the design matrix $[\mathbf{1} : \pi_{\mathbf{jm}}]$. R_0^2 is the residual sum of squares$= \sum_{j=1}^{n} (Y_j - \beta_0 - \beta_1 \hat{\pi}_{jm})^2$.

The critical region for the test at level-γ is $\{T < t_{n-2,1-\gamma}\}$.
If n is large, using CLT, we approximate the critical region by $\{T < z_{1-\gamma}\}$, z_{1-p} being the p^{th} quantile of the standard normal distribution.

The critical region for the test at level-γ is $\{T < t_{n-2,1-\gamma}\}$.
If n is large, using CLT, we approximate the critical region by $\{T < z_{1-\gamma}\}$, z_{1-p} being the p^{th} quantile of the standard normal distribution.
Thus the power function is

$$P(\beta_1) = \Phi\left(z_{1-\gamma} - \frac{\beta_1}{\hat{se}(\hat{\beta}_1)}\right)$$

where Φ is the standard normal c.d.f.

The test will have a power at β at β_1 if:

$$\Phi\left(z_{1-\gamma} - \frac{\beta_1}{\hat{se}(\hat{\beta}_1)}\right) = \beta$$

$$\implies z_{1-\gamma} - \frac{\beta_1}{\sqrt{\frac{S_{11}R_0^2}{n-2}}} = z_{1-\beta}$$

$$\implies n = \frac{(z_{1-\gamma} - z_{1-\beta})^2 S_{11} R_0^2}{\beta_1^2} + 2$$

Testing for Linkage
Sample Size required to detect Linkage

n is a decreasing function of β_1^2, and noting that

$$\beta_1^2 = 4(1-2\theta)^4\sigma_x^4 = 4(1-2\theta)^4 4p^2q^2\alpha^4$$

is a decreasing function of θ for $\theta \leq 0.5$, an increasing function of α, and a decreasing function of p for $p \geq 0.5$, we have:

n is a decreasing function of β_1^2, and noting that

$$\beta_1^2 = 4(1-2\theta)^4\sigma_x^4 = 4(1-2\theta)^4 4p^2q^2\alpha^4$$

is a decreasing function of θ for $\theta \leq 0.5$, an increasing function of α, and a decreasing function of p for $p \geq 0.5$, we have:

Observations

- n is increasing function of θ (Note that for linkage $\theta < 0.5$). This is evident as, if the strength of linkage between trait and marker locus is higher, a smaller sample is required to do detect linkage.
- n is increasing function of p, ($p > 0.5$), i.e if a locus is controlled by several loci with comparable affects, then the sample size required to map the QTL with the highest level of heterozygosity is the smallest.
- n is decreasing function of α , i.e if among several other QTL's, the marginal effect of one QTL increases, then smaller sample sizes are required to map that locus.
- n is independent of σ^2.


Thank you!