Interval Mapping

- introduced by Lander and Botstein (1989)
- systematic scan of genome for evidence of QTL
- extension of single marker analysis
  to any genomic location flanked by two markers
- genome-wide search for QTL
  employ complete marker linkage maps
  view QTL genotypes as missing data
  use mixture model for maximum likelihood
- concept maintained in extensions of IM analysis
  composite interval mapping (CIM)
  multiple interval mapping (MIM)
  Bayesian interval mapping (BIM)

IM Outline

- QTL model
  trait model
  QTL recombination
- maximum likelihood
  likelihood ratio statistic
  regression approximation
- threshold determination
  theoretical idea
  data-driven permutation test
- sample variance
  theory: Fisher information
  bootstrap estimate
- advantages of IM

general references and extensions to other designs

- Lynch and Walsh (1998)
- Doerge, Zeng and Weir (1997)
- Luo and Williams 1993
- Jansen 1996
- Jiang and Zeng 1997
- Kao and Zeng 1997

QTL Model

- single marker analysis
  test association of quantitative trait and marker
  find marker linked to one or more QTL
  cannot directly map the QTL
  marker effect is confounded with recombination rate
- flanking pair of markers
  Lander and Botstein (1989)
  disentangle r and a from the test statistic
  implemented it using maximum likelihood procedures
- software:
  MapMaker/QTL
  QTL Cartographer
  PLABQTL (regression)

Backcross Design
trait linear model

test for QTL (Q) located on an interval flanked by markers i and $i+1$
assuming loci order M_iQM_{i+1}

$$y_j = \mu + b^* x_j^* + e_j \quad j = 1, 2, \cdots, n$$

where

$$\begin{aligned} b^* &= \text{the effect of the putative QTL} \\ x_j^* &= \begin{cases} 1 & \text{if the QTL genotype is } QQ \\ 0 & \text{if the QTL genotype is } Qq \end{cases} \\ e_j &\sim N(0, \sigma^2) \end{aligned}$$

probability distribution of y_j

$$f(y_j|\mu, b^*) = \begin{cases} \phi\left(\frac{y_j-\mu-b^*}{\sigma}\right) & \text{if } x_j^* = 1 \\ \phi\left(\frac{y_j-\mu}{\sigma}\right) & \text{if } x_j^* = 0 \end{cases}$$

standard normal density

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp[-z^2/2]$$

Backcross Design
unknown QTL genotype

probability distribution of x_j^*
depends on flanking markers M_i, M_{i+1}
depends on quantitative trait position $\theta = \frac{r_{M_iQ}}{r_{M_iM_{i+1}}}$
different for each individual j:

$$p_{kj} = Prob(x_j^* = k | M_i, M_{i+1}, \theta) \quad k = 0, 1$$

two markers and one QTL

flanking markers	sample size	QTL genotype QQ (1)	QTL genotype Qq (0)
$\frac{M_iM_{i+1}}{M_iM_{i+1}}$	n_1	$\frac{(1-r_{M_iQ})(1-r_{QM_{i+1}})}{1-r_{M_iM_{i+1}}} \approx 1$	$\frac{r_{M_iQ}r_{QM_{i+1}}}{1-r_{M_iM_{i+1}}} \approx 0$
$\frac{M_im_{i+1}}{M_iM_{i+1}}$	n_2	$\frac{(1-r_{M_iQ})r_{QM_{i+1}}}{r_{M_iM_{i+1}}} \approx 1-\theta$	$\frac{r_{M_iQ}(1-r_{QM_{i+1}})}{r_{M_iM_{i+1}}} \approx \theta$
$\frac{m_iM_{i+1}}{M_iM_{i+1}}$	n_3	$\frac{r_{M_iQ}(1-r_{QM_{i+1}})}{r_{M_iM_{i+1}}} \approx \theta$	$\frac{(1-r_{M_iQ})r_{QM_{i+1}}}{r_{M_iM_{i+1}}} \approx 1-\theta$
$\frac{m_im_{i+1}}{M_iM_{i+1}}$	n_4	$\frac{r_{M_iQ}r_{QM_{i+1}}}{1-r_{M_iM_{i+1}}} \approx 0$	$\frac{(1-r_{M_iQ})(1-r_{QM_{i+1}})}{1-r_{M_iM_{i+1}}} \approx 1$

approximation: ignore double recombination

Backcross Design
mixture model: x_j^* is unobserved

$$f(y_j|x_j^*, \mu, b^*) = \phi\left(\frac{y_j - \mu - b^* x_j^*}{\sigma}\right)$$

$$p_{kj} = Prob(x_j^* = k | M_i, M_{i+1}, \theta)$$

likelihood function

$$L(\mu, b^*, \sigma^2, \theta) = \prod_{j=1}^{n} \sum_{k=0}^{1} p_{kj} f(y_j|k, \mu, b^*)$$

$$= \prod_{j=1}^{n} \left[p_{1j} \phi\left(\frac{y_j - \mu - b^*}{\sigma}\right) + p_{0j} \phi\left(\frac{y_j - \mu}{\sigma}\right) \right]$$

$$\approx \prod_{j=1}^{n_1} \phi\left(\tfrac{y_j-\mu-b^*}{\sigma}\right) \prod_{j=1}^{n_2} \left[(1-\theta)\phi\left(\tfrac{y_j-\mu-b^*}{\sigma}\right) + \theta\phi\left(\tfrac{y_j-\mu}{\sigma}\right)\right]$$
$$\times \prod_{j=1}^{n_3} \left[\theta\phi\left(\tfrac{y_j-\mu-b^*}{\sigma}\right) + (1-\theta)\phi\left(\tfrac{y_j-\mu}{\sigma}\right)\right] \prod_{j=1}^{n_4} \phi\left(\tfrac{y_j-\mu}{\sigma}\right)$$

Backcross Design
maximum likelihood analysis

Textensive literature (*e.g.*, Lander and Botstein 1989; Carbonell and Gerig 1991; Luo and Kearsey 1992; van Ooijen 1992; Carbonell et al. 1992; Luo and Williams 1993; Jansen 1992, 1993, 1994, 1996; Jansen and Stam 1994)

derivation based on EM algorithm

Differentiate log likelihood with respect to b^*
and set to zero to find maximum of L (why?)

$$\frac{\partial \ln L}{\partial b^*} = \sum_{j=1}^{n} P_j \frac{[y_j - \mu - b^*]}{\sigma^2} = 0$$

where

$$P_j = \frac{p_{1j}\phi([y_j - \mu - b^*]/\sigma)}{p_{1j}\phi([y_j - \mu - b^*]/\sigma) + p_{0j}\phi([y_j - \mu]/\sigma)}.$$

P_j = posterior probability of $x_j^* = 1$
p_{1j} = prior probability (θ, flanking markers)

Backcross Design
solve for linear model parameters

$$\sum_{j=1}^{n} P_j(y_j - \mu - b^*)/\sigma^2 = 0$$

$$\Longrightarrow \hat{b}^* = \sum_{j=1}^{n}(y_j - \mu)P_j / \sum_{j=1}^{n} P_j$$

Differentiate log-likelihood with respect to μ:

$$\frac{\partial \ln L}{\partial \mu} = \sum_{j=1}^{n}[P_j(y_j - \mu - b^*) + (1 - P_j)(y_j - \mu)]/\sigma^2$$

$$\frac{\partial \ln L}{\partial \mu} = 0 \Longrightarrow \hat{\mu} = \sum_{j=1}^{n}(y_j - P_j\hat{b}^*)/n$$

Differentiate log-likelihood with respect to σ^2:

$$\frac{\partial \ln L}{\partial \sigma^2} = \sum_{j=1}^{n}\frac{P_j(y_j - \mu - b^*)^2 + (1 - P_j)(y_j - \mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2}$$

$$\frac{\partial \ln L}{\partial \sigma^2} = 0 \Longrightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}[(y_j - \mu)^2 - P_j b^{*2}].$$

Backcross Design
unknown position θ

$$\begin{aligned}\frac{\partial \ln L}{\partial \theta} &= \sum_{j=1}^{n_2}\left[\frac{-\phi\left(\frac{y_j-\mu-b^*}{\sigma}\right) + \phi\left(\frac{y_j-\mu}{\sigma}\right)}{(1-\theta)\phi\left(\frac{y_j-\mu-b^*}{\sigma}\right) + \theta\phi\left(\frac{y_j-\mu}{\sigma}\right)}\right] \\ &\quad + \sum_{j=1}^{n_3}\left[\frac{\phi\left(\frac{y_j-\mu-b^*}{\sigma}\right) - \phi\left(\frac{y_j-\mu}{\sigma}\right)}{\theta\phi\left(\frac{y_j-\mu-b^*}{\sigma}\right) + (1-\theta)\phi\left(\frac{y_j-\mu}{\sigma}\right)}\right] \\ &= \sum_{j=1}^{n_2}\left[-\frac{P_j}{1-\theta} + \frac{1-P_j}{\theta}\right] + \sum_{j=1}^{n_3}\left[\frac{P_j}{\theta} - \frac{1-P_j}{1-\theta}\right]\end{aligned}$$

$$\frac{\partial \ln L}{\partial \theta} = 0$$

$$\Longrightarrow \hat{\theta} = \frac{\sum_{j=1}^{n_2}[(1 - \hat{P}_j) + \sum_{j=1}^{n_3}\hat{P}_j]}{n_2 + n_3}$$

approximation: ignore double recombination

EM Algorithm

- solutions for μ, b^*, σ^2, θ not in closed form
  each estimate depends on other estimates
  numerical method needed: iterate above equations

- EM (Expectation/Maximization) algorithm
  0. pick initial estimate
  no QTL: $\hat{\mu} = \bar{y}.$, $\hat{b}^* = 0$ or
  least squares estimates of b^* and μ using $x_j^* = p_{1j}$

  1. iterate between E-step and M steps
  E-step: estimate P_j
  M-steps: maximize L for μ, b^*, σ^2, θ

  2. iterate step 1 until convergence of estimates
  criteria: negligible change in estimates

Likelihood ratio test statistic

The LOD score is test statistic

$$\text{LOD}(\theta) = \text{LOD} = -\log_{10}\frac{L(\hat{\hat{\mu}}, b^* = 0, \hat{\hat{\sigma}}^2)}{L(\hat{\mu}, \hat{b^*}, \hat{\sigma}^2)}$$

for the hypotheses

$$H_0: \ b^* = 0 \ \text{ and } \ H_1: \ b^* \neq 0$$

assume putative QTL located at position (θ)
$\hat{\mu}$, $\hat{b^*}$, $\hat{\sigma}^2$ are maximum likelihood estimates under H_1
$\hat{\hat{\mu}}$, $\hat{\hat{\sigma}}^2$ are ML estimates under H_0 with $b^* = 0$

LOD score test and the usual likelihood ratio test

$$\text{LR} = -2\ln\frac{L(\hat{\hat{\mu}}, b^* = 0, \hat{\hat{\sigma}}^2)}{L(\hat{\mu}, \hat{b^*}, \hat{\sigma}^2)}$$

$$\text{LOD} = \frac{1}{2}\log_{10}(e)\text{LR} = 0.217\text{LR}$$

$$\text{LR} = 2\log(10)\text{LR} = 4.605\text{LOD}$$

genome scan: profile likelihood

- systematic strategy
  search for QTL by scanning positions on genome
  scan each interval between flanking markers
  graphic display: plot LOD versus map position
- profile likelihood
  profile one parameter (position)
  while maximizing over all others
  plot likelihood ratio statistic versus map position
- critical threshold
  LOD score exceeds pre-defined threshold
  QTL is indicated near maximum LOD
  width: 1–2 LOD support interval (Lander and Botstein 1989)
- properties
  estimates are asymptotically unbiased
  if assumption of at most one QTL on a chromosome is true

EM Algorithm Steps

- Expectation (E-step) $E(x^*)$

$$P_j = \frac{p_{1j}\phi([y_j - \mu - b^*]/\sigma)}{p_{1j}\phi([y_j - \mu - b^*]/\sigma) + p_{0j}\phi([y_j - \mu]/\sigma)}.$$

- Maximization (M-steps) of likelihood

$$\widehat{b}^* = \sum_{j=1}^{n}(y_j - \mu)P_j / \sum_{j=1}^{n} P_j$$

$$\widehat{\mu} = \sum_{j=1}^{n}(y_j - P_j\widehat{b}^*)/n$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}[(y_j - \mu)^2 - P_j b^{*2}]$$

$$\widehat{\theta} = \frac{\sum_{j=1}^{n_2}[(1 - \hat{P}_j) + \sum_{j=1}^{n_3}\hat{P}_j]}{n_2 + n_3}$$

Haley-Knott regression approximation

- simplified approximation (Haley Knott 1992; Martinez Curnow 1992)
- regress y_j on expected value of x_j^*

  $E(x_j^*) = p_{1j} = Prob(x_j^* = 1|M_i, M_{i+1}, \theta)$

  least squares estimate of b^*

$$y_j = \mu + b^* p_{1j} + e_j \quad j = 1, 2, \cdots, n$$

- straightforward statistical analysis
  has correct mean: $E(y_j) = \mu + b^* p_{1j}$
  assume normal with equal variance, not mixture of normals
- very good approximation of LOD profile
  Haley Knott (1992); Rebai et al. (1995)
  overestimates variance (correction in Xu 1995)

Threshold determination

- large sample theory for LR (or LOD) test statistic
  asymptotically chi-square distributed
  under null hypothesis

$$4.605\text{LOD}(\theta) = \text{LR}(\theta) \sim \chi_1^2$$

- one degree of freedom for BC design
  two degrees of freedom for F_2 design
  (Lander Botstein 1989; van Ooijen 1992; Zeng 1994)

multiple testing problem

- test usually performed on whole genome
  what is distribution of maximum LOD score
  over the whole genome under null hypothesis?

asymptotic theory for maximum LOD

- based on an correlated normals
- dense map: infinite markers
  Ornstein-Uhlenbeck diffusion process
  correlation depends on r in simple way
  determine genome-wide critical values
  Lander Botstein (1989), Feingold et al. (1993)
  Lander Schork (1994); Dupuis Siegmund (1999)
- sparse map: loosely linked markers
  Zeng (1994) and Rebai et al (1994)
- guideline of 2-3 LOD
  to ensure a 5% overall false positive rate
  Lander Botstein (1989); Lander Kruglyak (1995)

Backcross Design
chi-square approximation

at a marker: squared t test statistic

$$t = \frac{\tilde{\mu}_1 - \tilde{\mu}_0}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_0}\right)}} \longrightarrow N(0,1)$$

under null hypothesis $\mu_1 = \mu_0$ as $n \to \infty$
likelihood principle: LR $\longrightarrow \chi^2_1$ as $n \to \infty$
fact: $N(0,1)^2 = \chi^2_1$

at a position between flanking marker (fixed θ)

$$\text{LR} \longrightarrow \chi^2_1 \quad \text{as} \quad n \to \infty$$

LR threshold approaches chi-square critical value

$$\text{LR}_\alpha \longrightarrow \chi^2_{1,\alpha} \quad \text{as} \quad n \to \infty$$

Backcross Design
genome-wide threshold

maximum LR for a marker interval

$$\max[\text{LR}(\theta, 0 < \theta < 1)]$$

distribution is between χ^2_1 and χ^2_2
closer to χ^2_1 for small interval (say $<$ 10cM)

$$\chi^2_{1,\alpha} < \text{LR}_\alpha < \chi^2_{2,\alpha}$$

why? two parameters: b^* and θ

maximum LR for whole genome

$$\max[\text{LR}(\lambda, \lambda \in \text{genome})]$$

called experimental-wise threshold
find critical value T_α under null hypothesis
null: no QTL in the whole genome
chance max[LR] exceeds T_α *somewhere* in genome

$$\alpha = Prob\{\max[\text{LR}(\lambda, \lambda \in \text{genome})] > T_\alpha | H_0\}$$

BC genome-wide threshold
sparse-map case

markers sparse and widely separately
marker intervalsapproximately independent
consider M "independent" intervals (Bonferroni)

$$\begin{aligned} 1-\alpha &= Prob(\text{no error in the genome}) \\ &= \prod_{i=1}^{M} Prob(\text{no error in interval } i) \\ &= \prod_{i=1}^{M} (1-p) = (1-p)^M \approx 1 - Mp \end{aligned}$$

$$\Longrightarrow p \approx \alpha/M$$

$$\chi^2_{1,\alpha/M} < T_\alpha < \chi^2_{2,\alpha/M}$$

BC genome-wide threshold
dense-map case

markers everywhere (continuous)
λ = position on genome
what is relationship between λ and θ?

$$LR(\lambda) \to z^2(\lambda) \quad \text{with } z \sim N(0,1)$$

consider two (linked) positions λ_1, λ_2
LR(λ_1) and LR(λ_2) are correlated

$$Corr(z(\lambda_1), z(\lambda_2)) = 1 - 2r$$

with r the recombination between these loci

Lander and Botstein (1989)'s
Proposition for Backcross Design

Consider an organism with C chromosomes and genetic length G, measured in Morgans. When no QTL are present, the probability that the LOD score exceeds a high level T is approximately $(C + 2Gt)\chi^2[t]$, where $T = (2\ln 10)^{-1}t$ and $\chi^2[t]$ denotes the inverse cumulative distribution function of the χ^2 distribution with 1 d.f. In order to make the probability less than α that a false positive occurs somewhere in the genome, the appropriate LOD threshold is thus approximately $T_\alpha = (2\ln 10)^{-1}t_\alpha$ where t_α solves the equation $\alpha = (C + 2Gt_\alpha)\chi^2[t_\alpha]$.

Bottom line: For many organisms, LOD threshold is between 2 and 3. For mouse with 20 chromosomes, it is close to 3. For maize with 10 chromosomes, it is close 2.7. The threshold for F_2 design should be higher than that for backcross design.

Permutation test

data-driven empirical critical values for mapping QTL based on concept of permutation test (Fisher)
Churchill Doerge (1994); Doerge Churchill (1996)

1. Randomly permute responses y while leaving markers M intact (simulates null hypothesis of no association between genotype and phenotype).

2. Perform IM analysis on permuted sample.

3. Repeat many times to obtain empirical distribution of test statistic. Determine 95% significance value to test null hypothesis.

4. Declare QTL if original test statistic is higher in a genomic region.

advantages of permutation critical value

critical values reflect specifics of experiment
sample size, experimental design, missing data, etc.
flexible and extensible to other mapping methods

permutation test details

consider data from a population for QTL mapping
phenotypic trait value $y_j, j = 1, \cdots, n$
t marker genotypes $\mathbf{M}_j = \{M_{1j}, M_{2j}, \cdots, M_{tj}\}$
for convenience define the pair

$$X_j = \{y_j, \mathbf{M}_j\}$$

$$\mathbf{X} = \{X_1, X_2, \cdots, X_n\}$$

permuted sample of size n
permutation = random sample without replacement of all n items (reordering)
$h(1), h(2), \cdots, h(n)$ is permutation of $1, 2, \cdots, n$

$$\mathbf{X}' = \{X_1', X_2', \cdots, X_n'\}$$

permute phenotypes only y: $y_j' = y_{h(j)}$

$$X_j' = \{y_j', \mathbf{M}_j\}$$

In permuted samples $\mathbf{X}'$, y' and $\mathbf{M}$ do not have any intrinsic relationship, thus simulating the null hypothesis of no QTL.

empirical distribution of permutation sample

"collect" N permuted samples
compute maximum likelihood test statistic
–for position λ
–maximum value for a marker interval
–maximum value for genome

$$LR'_1, LR'_2, \cdots, LR'_N$$

$\alpha \times 100\%$ threshold

estimate test threshold under null hypothesis

$\hat{T}_\alpha = \alpha \times 100$ percentile of $\{LR'_1, LR'_2, \cdots, LR'_N\}$

compare this value with original test statistic
to declare the mapping of a QTL.

how large should N be?

Churchill and Doerge (1994) suggestions:
$\alpha = 0.05$: $N \geq 1000$
$\alpha = 0.01$: $N \geq 5000$

confidence intervals and standard errors

usual idea: estimate $\pm 2\times$ standard error

$$\bar{y}_\cdot \pm 1.96 s_y/\sqrt{n}$$

more general idea: likelihood region
use contours of likelihood surface
log likelihood has quadratic drop away from peak
ready interpretation for nonlinear problems
such as QTL position

one-LOD support interval

want confidence interval for QTL position
Lander Botstein (1989) proposed 1- LOD interval
(standard procedure in human genetic analysis)
based on asymptotical χ^2 distribution of LR statistic
center on ML estimate of QTL position

approximate interval for small effects

Mangin, Goffinet and Rebai (1994)
1-LOD method appropriate when QTL effect is large
test statistic not χ^2 when QTL effect is small
1-LOD interval underestimates confidence interval
approximate method to account for QTL effect

Fisher's information matrix

Kao and Zeng (1997)
estimate sampling variances of $\tilde{\theta}, \tilde{b}^*$
using Fisher's information of QTL
(opposite of second derivative of log L)
under the mixture model framework
interval mapping, composite interval mapping

Bootstrap estimate of sampling variance

Bootstrap samples are generally created by sampling with replacement n individual observations. An observation consists of a trait phenotype and marker genotypes. At each bootstrap sample, n observations are sampled with replacement out of the pool of the n original observations. Some records can appear more than once in a bootstrap sample, while others are not included at all. Estimation on the parameters is performed for each bootstrap sample, and the mean, standard error, confidence interval of estimates can be estimated from N bootstrap samples. For example, the empirical central 95% confidence interval of a QTL position can be determined by ordering the N estimates and taking the bottom and top 2.5th percentile of the N bootstrap estimates, respectively. Visscher, Thompson and Haley (1996) studied the use of bootstrap samples to estimate confidence interval of QTL position.

bootstrap sample specifics

bootstrap sample is random sample of size n

$$\mathbf{X}^* = \{X_1^*, X_2^*, \cdots, X_n^*\}$$

drawn with replacement from original sample $\mathbf{X}$
X_i^* equals any of original $X_j, j = 1, 2, \cdots, n$ with probability $1/n$

1. Select N independent bootstrap samples, $\mathbf{X}_1^*, \mathbf{X}_2^*, \cdots, \mathbf{X}_N^*$, each of size n drawn with replacement from $\mathbf{X}$.

2. Estimate parameter θ from the bth bootstrap sample

$$\widehat{\theta}_b^* = S(\mathbf{X}_b^*) \quad \text{for} \quad b = 1, 2, \cdots, N$$

3. Estimate the standard error, $SE(\widehat{\theta})$, of $\widehat{\theta}$ by the sample standard deviation of the bootstrap

estimates, $\widehat{\theta}_b^*$, from the N replications

$$SE(\widehat{\theta})_B = \left[\sum_{b=1}^{N} [\widehat{\theta}_b^* - \widehat{\overline{\theta}}^*]^2/(N-1)\right]^{1/2}$$

where $\widehat{\overline{\theta}}^* = \sum_{b=1}^{N} \widehat{\theta}_b^*/N$.

4. Estimate the 95% confidence interval, $CI(\widehat{\theta})$, of $\widehat{\theta}$ by

$$CI(\widehat{\theta})_B = \text{2.5th and 97.5th percentiles of}$$

$$\{\widehat{\theta}_1^*, \widehat{\theta}_2^*, \cdots, \widehat{\theta}_N^*\}$$

Variance explained by QTL

Sometimes the magnitude of a QTL is also reported as the proportion of the variance explained by the QTL (σ^2_{exp}). This is usually estimated as

$$\widehat{\sigma}^2_{exp} = \frac{\widehat{\sigma}^2_{tot} - \widehat{\sigma}^2_{reduce}}{\widehat{\sigma}^2_{tot}}$$

where $\widehat{\sigma}^2_{tot}$ is an estimate of the total phenotypic variance ($\widehat{\bar{\sigma}}^2$ at the null hypothesis) and $\widehat{\sigma}^2_{reduce}$ is an estimate of the residual variance of the IM model ($\widehat{\sigma}^2$ at the alternative hypothesis).

However, this estimate is not additive for multiple QTL indicated by the analysis and usually overestimates the variance explained by a QTL, because the mapping analysis is not independent for different regions of the genome. Sometimes, $\widehat{\sigma}^2_{exp}$ for different QTL can add up more than 100%. A more appropriate estimate of variances and covariances explained by different QTL can be obtained through multiple interval mapping discussed below.

IM Advantages and disadvantages
advantages over single marker analysis

1. The probable position of the QTL can be inferred by the support interval;

2. The estimated locations and effects of QTL tend to be asymptotically unbiased if there is only one segregating QTL on a chromosome;

3. The method requires fewer individuals than one marker analysis for the detection of QTL.

IM Advantages and disadvantages
problems with interval mapping

1. test is not an interval test
   depends on effects of QTL outside an interval
   can have significant LOD with no QTL in interval
   single QTL on chromosome near maximum LOD?
   but number of QTL on chromosome unknown

2. if more than one QTL on chromosome
   test statistic at a position affected by all QTL
   estimated positions and effects of "QTL" biased

3. not efficient to use only two markers
   ignores information from other markers

Lander and Botstein (1989) proposed extension
map multiple QTL simultaneously with more b^*, x_j^*
multiple interval mapping (MIM)
many technical issues to be resolved