Estimation of genetic parameters using molecular markers and EM-algorithms

by

K. Emrich and W. Urfer

Abstract

In this paper we present a new method for estimating genetic parameters of an F_2-generation model. Using an iterative algorithm we derive explicit expressions for the Maximum Likelihood estimates of the additive and dominance effects. Finally we calculate the variance covariance matrix of our Maximum Likelihood estimates, which enables us to determine a confidence interval for the location of a quantitative trait locus.

Keywords: molecular genetics, Maximum Likelihood estimation, EM and ECM-algorithms, variance-and covariance matrix of Maximum Likelihood estimates, confidence intervals for genetic parameters

1. Introduction

The recent advent of molecular markers has created a great potential for the understanding of quantitative inheritance. Along with rapid developments in molecular marker technologies, biometrical models have been constructed, refined and generalized for detecting, mapping and estimating the effects of quantitative trait loci (QTLs). The aim of our statistical approach is to find Maximum Likelihood estimates of QTL locations and effects including their estimated standard errors.

During the last 10 years, numerous authors developed a number of statistical models for the so-called interval mapping methods in plant breeding. Lander and Botstein (1989) developed the method of simple interval mapping. This method was the basis of the disclosure of later methods like the composite interval mapping (Zeng, 1994). In interval mapping methods, the parameter estimators for a model are calculated for all possible QTL positions of the genome (e. g. in distances of 5 cM). Using likelihood ratio tests the likelihood ratio (or equivalent the Lod-Score) profile can be calculated to screen the genome for putative QTLs. Obviously, the number and size of the intervals should be considered in determining the threshold value.

Problematic is that during each parameter estimation only a single putative QTL is searched for. Jansen (1996) explains a Monte-Carlo expectation-maximization-algorithm for fitting multiple QTLs to incomplete genetic data. Stephens and Fisch (1998) employ reversible jump Markov-chain Monte-Carlo-methodology to compute posterior densities for the parameters and the number of QTLs. Fisch et al (1996) also developed an approach for generalizing mixture models for the progeny of a biparental cross of inbred lines. This approach allows to use mixture models for the case that the genotypes of a sample of plants are obtained in one generation and the phenotyping takes place in another generation with more individuals to investigate.

Kao and Zeng (1997) presented formulas for deriving Maximum Likelihood estimates using an ECM-Algorithm and explained a way to calculate asymptotic variance-covariance matrices for the estimates. The methods described by Kao and Zeng (1997) can be used for the following design (using Mendel's laws):

We consider experimental populations derived from a cross between two parental inbred lines P_1 and P_2, differing mainly in a quantitative trait of interest. This allows for investigating the realizations of alleles of marker loci of a marker interval that contains a putative QTL (quantitative trait locus). Two flanking markers for an interval, where a putative QTL is being tested, have alleles . If the F_1 individuals are selfed or intermated, an F_2 – population with nine observable marker genotypes is produced.

1.) Parental generation:

A_1	Q_1	B_1		A_2	Q_2	B_2
A_1	Q_1	B_1	$\times$	A_2	Q_2	B_2

2.) F_1-generation

A_1	Q_1	B_1
A_2	Q_2	B_2

3.) F_2-generation:

A_1	Q_1	B_1		A_1	Q_1	B_1		A_2	Q_2	B_2
A_1	Q_1	B_1		A_2	Q_2	B_2		A_2	Q_2	B_2

In relation to 1 : 2 : 1

In addition to these genotypes of the F_2-generation, a small number of recombinants occur by crossing over in the meiosis.

Here, A_1 and A_2 are the possible realizations of the alleles of marker 1 and B_1 and B_2 are the realizations of the alleles of marker 2 in the interval. Obviously, the realizations of the alleles of the putative QTL are not observable, but it is possible to give a table of conditional probabilities of the QTL genotypes given marker genotypes. For a F_2-Generation such a table reads:

Table 1: Conditional probabilities of QTL genotypes given marker genotypes:

No	Marker genotypes	Expected frequencies	QTL-genotypes Q_1/Q_1	Q_1/Q_2	Q_2/Q_2	Sample size n
1	A_1 B_1 / A_1 B_1	$(1-r)^2/4$	1	0	0	n_1
2	A_1 B_1 / A_1 B_2	$r(1-r)/2$	$1-p$	p	0	n_2
3	A_1 B_2 / A_1 B_2	$r^2/4$	$(1-p)^2$	$2p(1-p)$	p^2	n_3
4	A_2 B_1 / A_1 B_1	$r(1-r)/2$	p	$1-p$	0	n_4
5	A_2 B_2 / A_1 B_1	$(1-r)^2/2 + r^2/2$	$cp(1-p)$	$1-2cp(1-p)$	$cp(1-p)$	n_5
6	A_2 B_2 / A_1 B_2	$r(1-r)/2$	0	$1-p$	p	n_6
7	A_2 B_1 / A_2 B_1	$r^2/4$	p^2	$2p(1-p)$	$(1-p)^2$	n_7
8	A_2 B_2 / A_2 B_1	$r(1-r)/2$	0	p	$1-p$	n_8
9	A_2 B_2 / A_2 B_2	$(1-r)^2/4$	0	0	1	n_9

Obviously, there are nine different types of marker combinations.

It should be considered that e. g.

$$\frac{A_2 \quad B_2}{A_1 \quad B_1}$$ is the same combination as $$\frac{A_2 \quad B_1}{A_1 \quad B_2}$$, if no difference can be made between the alleles inherited from the father and the alleles inherited from the mother plant. This is the case in (most) practical applications.

Here p is defined, as r_A/r with r_A is the (e g from earlier investigations known) recombination fraction between the left marker 1 and the putative QTL and r is the (known) recombination fraction between the left marker 1 and the right marker 2.

$c=r^2/[r^2+(1-r)^2]$.

Further assumptions are

-that the possibility of a double recombination event (i. e. crossing over) will be ignored and

-that single crossing over happens independently from one another.

n is the number of plants in the sample F_2 generation and $n_1,....,n_9$ are the number of plants of this generation bearing the combinations of markers 1 to 9.

The conditional probabilities of the QTL genotypes given marker genotypes of table 1 enable us to calculate a matrix $\mathbf{P}=(p_{ji})$ (with dimension $(n\times3)$, j=1,2,...,n and i=1,2,3) for any sample of n plants whose markers are genotyped. This matrix contains the p_{ji} in dependence of the *marker genotype of each plant* of the sample.

2. The models and the likelihood function

It is now possible to define a *deterministic* genetic model for a F_2 population with the above given QTL genotype frequencies (1/4, 1/2, 1/4). The genetic model for one QTL represents the relation between a 'genotypic value' G and some genetic parameters β_0, a and d.

$$G=\begin{bmatrix}G_2\\G_1\\G_0\end{bmatrix}=\begin{bmatrix}1\\1\\1\end{bmatrix}\beta_0+\begin{bmatrix}1 & -\frac{1}{2}\\0 & \frac{1}{2}\\-1 & -\frac{1}{2}\end{bmatrix}\begin{bmatrix}a\\d\end{bmatrix}=\mathbf{1}_{3\times1}\beta_0+\mathbf{DE} \qquad (1)$$

Here, β_0 is a joint value of the genetic model and a and d are additive and dominance effects of QTL in the F_2 population. It is possible to calculate unique solutions of the genetic parameters in dependence of the genotypic values and frequencies of genotypes Q_1/Q_1, Q_1/Q_2 and Q_2/Q_2 of the QTL.

$\mathbf{D}=(D_1, D_2)$, with D_1 represents the status of the additive parameter and D_2 represents the status of the dominance effect.

For the following QTL mapping data

-y_j (j=1,2,...,n) is the investigated trait value of plant j

-$\mathbf{X}_j$ (j=1,2,...,n) is a vector which contains data for the genetic markers and other explanatory variables,

and the following assumptions can be made:

-there is no interaction (that is no epistasis) between QTLs

-there is no interference in crossing over

-there is only one QTL in the testing interval

A statistical composite interval mapping model (CIM, Zeng 1994) can be constructed on the basis of the genetic model:

$$y_j = ax_j^* + dz_j^* + X_j\beta + \varepsilon_j \qquad (2)$$

Here,

y_j is the trait value of the plant j (j=1,2,...,n),

a and d are additive and dominance effects of the putative QTL,

β is a partial regression coefficient vector of dimension k that contains the mean β_0 of the genetic model,

X_j is a subset of $\mathbf{X}_j$ that contains chosen marker and variable information

and $\varepsilon_j \sim N(0,\sigma^2)$.

x_j^* and z_j^* are discrete random effects with

$$x_j^* = \begin{cases} 1 & \\ 0 \ \text{ if the QTL is } & Q_1/Q_1 \\ -1 & \end{cases}$$

$$x_j^* = \begin{cases} 1 & Q_1/Q_1 \\ 0 \text{ if the QTL is } & Q_1/Q_2 \\ -1 & Q_2/Q_2 \end{cases}$$

and $$z_j^* = \begin{cases} 1/2 & \text{if the QTL is } Q_1/Q_2 \\ -1/2 & \text{otherwise} \end{cases}, \text{ j=1,2,...,n.}$$

Of course, the realizations of the putative QTL in plant j are unknown. Thus only the probability distribution of the realizations of the discrete random effects can be given in dependence of the conditional probabilities of the QTL genotypes given marker genotypes for plant j (called p_{ji}, with j=1,2,...,n, i=1,2,3):

$$g_j(x_j^*, z_j^*) = \begin{cases} p_{j1} & \text{if } x_j^* = 1 \quad \text{and } z_j^* = -1/2 \\ p_{j2} & \text{if } x_j^* = 0 \quad \text{and } z_j^* = 1/2 \\ p_{j3} & \text{if } x_j^* = -1 \quad \text{and } z_j^* = -1/2 \end{cases}$$

This is the distribution of the QTL genotype specified by x_j^* and z_j^*.

Now, it is possible to give a Likelihood function for a sample of n individuals and for the parameter vector $\theta = \left(a, d, \beta, \sigma^2\right)$:

$$L\left(\theta \middle| Y, \mathbf{X}\right) = \prod_{j=1}^{n}\left[\sum_{i=1}^{3} p_{ji} f\left(y_j; \mu_{ji}, \sigma^2\right)\right]$$

with

$$\begin{aligned}\mu_{j1} &= a - d/2 + X_j\beta \\ \mu_{j2} &= d/2 + X_j\beta \\ \mu_{j3} &= -a - d/2 + X_j\beta\end{aligned}$$

and f is the normal density of y_j with expectation value μ_{ji} (i=1,2,3 and j=1,2,..,n) and variance σ^2.

3. Parameter Estimation

3.1 The EM-algorithm

The not observable QTL genotypes can be considered as missing values. Now it is possible to define a data set $Y_{mis}=(y_{(mis,j)})$, (with j=1,2,...,n) of "missing data" for the QTL genotypes, and a data set $Y_{obs}=(y_{(obs,j)})$ (with j=1,2,...,n) for the observed values y_j and the marker information (cofactor vectors X_j , j=1,2,...,n).

It is possible to contemplate a hypothetical complete-data set called $Y_{com}=(Y_{obs}, Y_{mis})$. In such a situation the so-called EM-algorithms (or even the later explained ECM-algorithm) for Maximum Likelihood estimation of the parameters of the statistical model can be used. (see Dempster, Laird, Rubin (1971), Wu (1983) and Meng, Rubin (1993)).

Consider the random variable vector Y_{com} of the complete-data set with density function $f(Y_{com} \mid \theta)$ and $\theta \in \Theta \subseteq \mathbb{R}^d$. If Y_{com} contained only observed values, the objective way to estimate the parameters would be to maximize the complete-data log-likelihood function of θ:

$$l(\theta \mid Y_{com}) \propto \ln f(Y_{com} \mid \theta).$$

Unfortunately, Y_{com} contains the *not observable* missing values Y_{mis}. If we assume that the missing data in Y_{mis} are *missing at random*, than the log-likelihood for θ is:

$$l_{obs}(\theta \mid Y_{obs}) \propto \ln \int f(Y_{com} \mid \theta) dY_{mis}$$

Now in most practical applications (including the here-described situation) it is very complicated to maximize this log-likelihood-function.

The EM-algorithm solves this problem of maximizing l_{obs} by iteratively maximizing $l(\theta \mid Y_{com})$.

For each iteration, the EM-algorithm has two steps, the E-step and the M-step.

Using appropriate starting values for $\theta^{(0)}$,

-the (t+1) E-step finds the conditional *expectation* of the complete data log-likelihood with respect to the conditional distribution of Y_{mis} given Y_{obs} and the parameter $\theta^{(t)}$:

$$Q\left(\theta \middle| \theta^{(t)}\right) = \int l\left(\theta \middle| Y_{com}\right) f\left(Y_{mis} \middle| Y_{obs}, \theta = \theta^{(t)}\right) dY_{mis}$$

This is a function of θ for fixed Y_{obs} and fixed $\theta^{(t)}$.

-The (t+1) st M-step calculates a maximum $\theta^{(t+1)}$ for $Q\left(\theta \middle| \theta^{(t)}\right)$, so that

$$Q\left(\theta^{(t+1)} \middle| \theta^{(t)}\right) \geq Q\left(\theta \middle| \theta^{(t)}\right), \ \forall\, \theta \in \Theta$$

Under certain restrictions (Dempster, Laird, Rubin (1971), Wu (1983)), the sequence of estimates of the iterations steps of the EM-algorithm converges against a (global or local) maximum of l_{obs}.

Obviously, depending on the chosen starting values (and the used restrictions) it is possible that in some applications a stationary value of l_{obs} is found, but very often the EM-algorithm is able to find a maximum.

Today, the EM-algorithm is widely used in different applications. Selinski and Urfer (1998) and Selinski et al (1999) used this method for the estimation of toxicokinetic parameters for the risk assessment of potential harmful chemicals.

In the article of Kao and Zeng (1997), the ECM-algorithm as a subclass of generalized EM-algorithms has been used for the estimation of parameters.

Before this method will be described, the EM-algorithm for the defined Model has to be explained (Emrich (1999)).

For the F_2-Generation situation and the models (1) and (2) it has to be said that the observed data ($y_{(obs,j)}$) given the missing data ($y_{(mis,j)}$) are normally distributed with:

$$f\left(y_{(obs,j)} \middle| \theta, X_j, x_j^*, z_j^*\right) \sim N\left(ax_j^* + dz_j^* + X_j\beta, \sigma^2\right)$$

The conditional density of missing data given specified observations is the above defined density of QTL genotypes $g_j\left(x_j^*, z_j^*\right)$. In accordance with the formula of conditional probability (P(AξB)=P($A \cap B$)/P(B)), the density of the complete data set ($y_{(com,j)}$) can be considered as the likelihood-function and is defined as:

$$L\left(\theta \middle| y_{(com)}\right) = \prod_{j=1}^{n} f\left(y_{(obs,j)} \middle| \theta, X_j, x_j^*, z_j^*\right) g_j\left(x_j^*, z_j^*\right)$$

Now the conditional *expectation* of the complete data log-likelihood with respect to the conditional distribution of Y_{mis} given Y_{obs} and the parameter $\theta^{(t)}$ is in the **E-step of the EM-algorithm** calculated as:

$$\begin{aligned} Q\left(\theta \middle| \theta^{(t)}\right) &= \int \ln L\left(\theta \middle| Y_{com}\right) f\left(Y_{mis} \middle| Y_{obs}, \theta = \theta^{(t)}\right) dY_{mis} \\ &= \int \ln\left[\prod_{j=1}^{n} f\left(y_j; \mu_{ji}, \sigma^2\right) g_j\left(x_j^*, z_j^*\right)\right] f\left(Y_{mis} \middle| Y_{obs}, \theta = \theta^{(t)}\right) dY_{mis} \\ &= \sum_{j=1}^{n} \sum_{i=1}^{3} \ln\left[f\left(y_j; \mu_{ji}, \sigma^2\right) p_{ji}\right] \pi_{ji}^{(t)} \end{aligned}$$

$\pi_{ji}^{(t)} = \dfrac{p_{ji} f\left(y_j; \mu_{ij}^{(t)}, \sigma^{2(t)}\right)}{\sum_{v=1}^{3} p_{jv} f\left(y_j; \mu_{jv}^{(t)}, \sigma^{2(t)}\right)}$ is the *posterior probability of the QTL genotype*.

This follows by using the Bayes formula $P(A_k|B) = \dfrac{P(A_k \cap B)}{P(B)} = \dfrac{P(B|A_k)P(A_k)}{\sum_{s=1}^{l} P(B|A_s)P(A_s)}$

on the conditional distribution of Y_{mis} given Y_{obs}:

$$f\left(y_{(mis,j)} \middle| y_{(obs,j)}, \theta = \theta^{(t)}\right) = \frac{f\left(y_{(obs,j)} \middle| y_{(mis,j)}\right) \cdot f\left(y_{(mis,j)}\right)}{\sum_{l=1}^{3} f\left(y_{(obs,j)} \middle| y_{(mis,j),l\,fix}\right) \cdot f\left(y_{(mis,j),l\,fix}\right)} = \frac{g_j(x_j^*, z_j^*) f\left(y_j; \mu_j, \sigma^2\right)}{\sum_{v=1}^{3} p_{jl} f\left(y_j; \mu_{jv}, \sigma^2\right)}$$

f is the normal density of y_j with expectation μ_{ji} (or $\mu_{ji}^{(t)}$) and variance σ^2 (or $\sigma^{2(t)}$).

For the **M-step of the EM-algorithm** Q should be maximized:

A continuous, two times differentiable multidimensional function of a parameter vector θ has a maximum on the point $\theta = \theta^*$, if

-the partial derivations of the function are zero on the point $\theta = \theta^*$ and

-the matrix of second derivations is negative definite in $\theta = \theta^*$.

The Theorem of Hurwitz says that a real, symmetric Matrix $S=(s_{ij})$ is positive definite, if

$$\det\begin{pmatrix} s_{11} & \cdots & s_{1k} \\ \vdots & \ddots & \vdots \\ s_{k1} & \cdots & s_{kk} \end{pmatrix} > 0 \text{, for k=1,2,...,n.}$$

Thus the steps for calculating the maximum of Q are,

-to build the partial derivations and set them equal to zero

-and it would be advantageous, if the system of partial derivatives (=0) could be brought in a form like

$$\mathbf{A}\theta=b$$

because:

-then the zero point of the system of partial derivatives is

$$\theta^*=\mathbf{A}^{-1}b,$$

-for such a case the matrix of second derivatives could be calculated by differentiating ($b-\mathbf{A}\theta$ (=0)), and it follows, that the matrix of second derivatives is ($-\mathbf{A}$) and this matrix is negative definite, if **A** is positive definite.

The conditional *expectation* of the complete data log-likelihood with respect to the conditional distribution of Y_{mis} given Y_{obs} and the parameter $\theta^{(t)}$ can be brought to the form:

$$Q(\theta|\theta^{(t)}) = \sum_{j=1}^{n}\sum_{i=1}^{3}\left[-\ln(\sqrt{2\pi}) - \ln(\sigma) - \frac{1}{2}\left(\frac{y_j - \mu_{ji}}{\sigma}\right)^2 + \ln(p_{ji})\right]\cdot \pi_{ji}^{(t)}$$

Using this transformation, it is easy to build the partial derivatives of Q for the parameters a, d, $\beta_0,...,\beta_{k-1}$ and σ^2 (for the (t+1) th iteration step of the EM-algorithm), to equal them with zero and to transform the equations in a desirable form:

$$\frac{\partial Q}{\partial a} = \frac{1}{\sigma^2}\sum_{j=1}^{n}\left[\left(y_j - a^{(t+1)} + \left(\frac{d}{2}\right)^{(t+1)} - X_j\beta^{(t+1)}\right)\pi_{j1}^{(t)} - \left(y_j + a^{(t+1)} + \left(\frac{d}{2}\right)^{(t+1)} - X_j\beta^{(t+1)}\right)\pi_{j3}^{(t)}\right] = 0$$

$$\frac{\partial Q}{\partial a} = 0 \Rightarrow \sum_{j=1}^{n}\left(\pi_{j1}^{(t)} + \pi_{j3}^{(t)}\right)a^{(t+1)} + \frac{1}{2}\sum_{j=1}^{n}\left(\pi_{j3}^{(t)} - \pi_{j1}^{(t)}\right)d^{(t+1)} + \sum_{j=1}^{n}\left(\pi_{j1}^{(t)} - \pi_{j3}^{(t)}\right)X_{j0}\beta_0^{(t+1)} + ... + \sum_{j=1}^{n}\left(\pi_{j1}^{(t)} - \pi_{j3}^{(t)}\right)X_{jk}\beta_{k-1}^{(t+1)} = \sum_{j=2}^{n}\left(\pi_{j1}^{(t)} - \pi_{j3}^{(t)}\right)y_j$$

$$\frac{\partial Q}{\partial d} = \frac{1}{\sigma^2}\sum_{j=1}^{n}\left[-\left(y_j - a^{(t+1)} + \left(\frac{d}{2}\right)^{(t+1)} - X_j\beta^{(t+1)}\right)\pi_{j1}^{(t)} + \left(y_j - \left(\frac{d}{2}\right)^{(t+1)} - X_j\beta^{(t+1)}\right)\pi_{j2}^{(t)} - \left(y_j + a^{(t+1)} + \left(\frac{d}{2}\right)^{(t+1)} - X_j\beta^{(t+1)}\right)\pi_{j3}^{(t)}\right] = 0$$

$$\frac{\partial Q}{\partial d} = 0 \Rightarrow \frac{1}{2}\sum_{j=1}^{n}\left(\pi_{j3}^{(t)} - \pi_{j1}^{(t)}\right)a^{(t+1)} + \frac{1}{4}\sum_{j=1}^{n}\left(\pi_{j1}^{(t)} + \pi_{j2}^{(t)} + \pi_{j3}^{(t)}\right)d^{(t+1)} + \frac{1}{2}\sum_{j=1}^{n}\left(-\pi_{j1}^{(t)} + \pi_{j2}^{(t)} - \pi_{j3}^{(t)}\right)X_{j1}\beta_0^{(t+1)} + ... + \frac{1}{2}\sum_{j=1}^{n}\left(-\pi_{j1}^{(t)} + \pi_{j2}^{(t)} - \pi_{j3}^{(t)}\right)X_{jk}\beta_{k-1}^{(t+1)} = \frac{1}{2}\sum_{j=1}^{n}\left(-\pi_{j1}^{(t)} + \pi_{j1}^{(t)} - \pi_{j3}^{(t)}\right)y_j$$

$$\frac{\partial Q}{\partial \beta_{m-1}} = \frac{1}{\sigma^2}\sum_{j=1}^{n}\left[\left(y_j - a^{(t+1)} + \left(\frac{d}{2}\right)^{(t+1)} - X_j\beta^{(t+1)}\right)\pi_{j1}^{(t)}X_{jm} + \left(y_j - \left(\frac{d}{2}\right)^{(t+1)} - X_j\beta^{(t+1)}\right)\pi_{j2}^{(t)}X_{jm} + \left(y_j + a^{(t+1)} + \left(\frac{d}{2}\right)^{(t+1)} - X_j\beta^{(t+1)}\right)\pi_{j3}^{(t)}X_{jm}\right] = 0,\ m = 1,2,...,k$$

$$\frac{\partial Q}{\partial \beta_{m-1}} = 0 \Rightarrow \sum_{j=1}^{n}\left(\pi_{j1}^{(t)} - \pi_{j3}^{(t)}\right)X_{jm}a^{(t+1)} + \frac{1}{2}\sum_{j=1}^{n}\left(-\pi_{j1}^{(t)} + \pi_{j2}^{(t)} - \pi_{j3}^{(t)}\right)X_{jm}d^{(t+1)} + \sum_{j=1}^{n}X_{j1}X_{jm}\beta_0^{(t+1)} + ... + \sum_{j=1}^{n}X_{jk}X_{jm}\beta_{k-1}^{(t+1)} = \sum_{j=1}^{n}X_{jm}y_j$$

$$\frac{\partial Q}{\partial \sigma} = -\sum_{j=1}^{n}\frac{1}{\sigma^{(t+1)}} + \frac{1}{\sigma^{3(t+1)}}\sum_{j=1}^{n}\sum_{i=1}^{3}\left(y_j - \mu_{ji}^{(t+1)}\right)^2\pi_{ji}^{(t)} \stackrel{!}{=} 0$$

$$\frac{\partial Q}{\partial \sigma} = 0 \Rightarrow \sigma^{2(t+1)} = \frac{1}{n}\sum_{j=1}^{n}\left[\left(y_j - \mu_{j1}{}^{(t+1)}\right)^2\pi_{j1}^{(t)} + \left(y_j - \mu_{j2}{}^{(t+1)}\right)^2\pi_{j2}^{(t)} + \left(y_j - \mu_{j3}{}^{(t+1)}\right)^2\pi_{j3}^{(t)}\right]$$

Now, the equations of partial derivatives of a, d, $\beta_0,\dots,\beta_{k-1}$ can be written as a matrix equation:

$\mathbf{A}\theta^{*(t+1)} = b$, with $\theta^{*(t+1)} = \left(a^{(t+1)}, d^{(t+1)}, \beta_0^{(t+1)},\dots,\beta_{k-1}^{(t+1)}\right)^T$.

With:

$$
\begin{bmatrix}
\sum_{j=1}^{n}\left(\pi_{j1}^{(t)}+\pi_{j3}^{(t)}\right) & \frac{1}{2}\sum_{j=1}^{n}\left(\pi_{j3}^{(t)}-\pi_{j1}^{(t)}\right) & \sum_{j=1}^{n}\left(\pi_{j1}^{(t)}-\pi_{j3}^{(t)}\right)X_{j1} & \dots & \sum_{j=1}^{n}\left(\pi_{j1}^{(t)}-\pi_{j3}^{(t)}\right)X_{jk} \\
\frac{1}{2}\sum_{j=1}^{n}\left(\pi_{j3}^{(t)}-\pi_{j1}^{(t)}\right) & \frac{1}{4}\sum_{j=1}^{n}\left(\pi_{j1}^{(t)}+\pi_{j2}^{(t)}+\pi_{j3}^{(t)}\right) & \frac{1}{2}\sum_{j=1}^{n}\left(-\pi_{j1}^{(t)}+\pi_{j2}^{(t)}-\pi_{j3}^{(t)}\right)X_{j1} & \dots & \frac{1}{2}\sum_{j=1}^{n}\left(-\pi_{j1}^{(t)}+\pi_{j2}^{(t)}-\pi_{j3}^{(t)}\right)X_{jk} \\
\sum_{j=1}^{n}\left(\pi_{j1}^{(t)}-\pi_{j3}^{(t)}\right)X_{j1} & \frac{1}{2}\sum_{j=1}^{n}\left(-\pi_{j1}^{(t)}+\pi_{j2}^{(t)}-\pi_{j3}^{(t)}\right)X_{j1} & \sum_{j=1}^{n}X_{j1}X_{j1} & \dots & \sum_{j=1}^{n}X_{jk}X_{j1} \\
\vdots & \vdots & \vdots & & \vdots \\
\sum_{j=1}^{n}\left(\pi_{j1}^{(t)}-\pi_{j3}^{(t)}\right)X_{jk} & \frac{1}{2}\sum_{j=1}^{n}\left(-\pi_{j1}^{(t)}+\pi_{j2}^{(t)}-\pi_{j3}^{(t)}\right)X_{jk} & \sum_{j=1}^{n}X_{j1}X_{jk} & \dots & \sum_{j=1}^{n}X_{jk}X_{jk}
\end{bmatrix}
\begin{pmatrix}
a^{(t+1)} \\ d^{(t+1)} \\ \beta_0^{(t+1)} \\ \vdots \\ \beta_{k-1}^{(t+1)}
\end{pmatrix}
=
$$

$$
\begin{pmatrix}
\sum_{j=2}^{n}\left(\pi_{j1}^{(t)}-\pi_{j3}^{(t)}\right)y_j \\
\frac{1}{2}\sum_{j=1}^{n}\left(-\pi_{j1}^{(t)}+\pi_{j1}^{(t)}-\pi_{j3}^{(t)}\right)y_j \\
\sum_{j=1}^{n}X_{j1}y_j \\
\vdots \\
\sum_{j=1}^{n}X_{jk}y_j
\end{pmatrix}
$$

-The partial derivative for σ is dependent from the other parameters a, d, $\beta_0,\dots,\beta_{k-1}$, but the other partial derivatives (=0) do not contain σ^2.

-Now it is possible to calculate the parameter estimator of the first (k+1) parameter by solving $\mathbf{A}\theta^{*(t+1)} = b$ and to use these estimators for calculating $\sigma^{2(t+1)}$ with

$$\sigma^{2(t+1)} = \frac{1}{n}\sum_{j=1}^{n}\left[\left(y_j - \mu_{j1}{}^{(t+1)}\right)^2 \pi_{j1}^{(t)} + \left(y_j - \mu_{j2}{}^{(t+1)}\right)^2 \pi_{j2}^{(t)} + \left(y_j - \mu_{j3}{}^{(t+1)}\right)^2 \pi_{j3}^{(t)}\right].$$

Now the parameter estimator vector of the (t+1)-th iteration step of the EM-algorithm is

$$\theta^{(t+1)} = \left(a^{(t+1)}, d^{(t+1)}, \beta_0^{(t+1)},\dots,\beta_{k-1}^{(t+1)}, \sigma^{(t+1)}\right)^T.$$

3.2 The ECM-algorithm

In the article of Kao and Zeng (1997) a generalization of the EM-algorithm, called the ECM (Expectation/ Conditional Maximization) has been used for parameter estimation (Meng, Rubin, (1993)). The ECM-algorithm should be preferred when the complete-data Maximum Likelihood estimation is complicated. For this case, the calculation of parameter estimators using the ECM-algorithm is slightly easier than with the EM-algorithm. For more complicated models the ECM-algorithm could more considerably simplify the calculation of parameter estimators. On the other hand, the ECM-algorithm needs further restrictions to converge against a Maximum Likelihood estimator.

The ECM-algorithm replaces the original M-step of the EM-algorithm by a number of S conditional maximization (CM-) steps. These CM-steps are computationally simpler to calculate because each maximization is conditional on a function g_s (s=1,...,S). Therefore, in each CM-step a maximization of $Q\left(\theta^{(t+s/S)}\right)$ takes place.

The (t+1) th iteration step is analogous to the E-step of the EM-Algorithm. Therefore, the calculation of the conditional *expectation* of the complete data log-likelihood with respect to the conditional distribution of Y_{mis} given Y_{obs} and the parameter $\theta^{(t)}$ computes as:

$$Q(\theta|\theta^{(t)}) = \sum_{j=1}^{n}\sum_{i=1}^{3}\left[-\ln(\sqrt{2\pi}) - \ln(\sigma) - \frac{1}{2}\left(\frac{y_j - \mu_{ji}}{\sigma}\right)^2 + \ln(p_{ji})\right]\cdot \pi_{ji}^{(t)}$$

Then four CM-steps are calculated:

-The first CM-step is the maximization of Q under the condition that $d = d^{(t)}$, $\beta = \beta^{(t)}$ and $\sigma = \sigma^{(t)}$ are fixed values, which were calculated in the (t)-th step of the ECM-algorithm. So the partial derivative $\partial Q/\partial a = 0$ is calculated and can be solved to:

$$a^{(t+1)} = \frac{\sum_{j=1}^{n}\left[\left(\pi_{j1}^{(t)} - \pi_{j3}^{(t)}\right)\left(y_j - x_j\beta^{(t)}\right) - \frac{1}{2}\left(\pi_{j3}^{(t)} - \pi_{j1}^{(t)}\right)d^{(t)}\right]}{\sum_{j=1}^{n}\left(\pi_{j1}^{(t)} + \pi_{j3}^{(t)}\right)}.$$

-The second CM-step contains the maximization of Q under the condition of fixed parameter estimator values of earlier steps, i. e. $a^{(t+1)}, \beta^{(t)}$ and $\sigma^{(t)}$ are fixed values and the partial derivative $\partial Q/\partial d = 0$ can be transformed to:

$$d^{(t+1)} = \frac{\sum_{j=1}^{n}\frac{1}{2}\left[\left(-\pi_{j1}^{(t)} + \pi_{j2}^{(t)} - \pi_{j3}^{(t)}\right)\left(y_j - x_j\beta^{(t)}\right) - \left(\pi_{j3}^{(t)} - \pi_{j1}^{(t)}\right)a^{(t+1)}\right]}{\frac{1}{4}\sum_{j=1}^{n}\left(\pi_{j1}^{(t)} + \pi_{j2}^{(t)} + \pi_{j3}^{(t)}\right)}.$$

-The third CM-step uses the fixed parameter estimators of the last two earlier CM-steps, so that $a^{(t+1)}, d^{(t+1)}$ and $\sigma^{(t)}$ are fixed:

$$\beta^{(t+1)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^{T}\left[Y - \Pi^{(t)}\,\mathbf{D}E^{(t+1)}\right]$$

Here, $\mathbf{D} = (D_1, D_2)$ is the design matrix of the genetic model,

$$Y = \left(y_1,\ldots,y_n\right)^T,\ \Pi^{(t)} = \left(\pi_{ji}^{(t)}\right),\ E^{(t+1)} = \left(a^{(t+1)}, d^{(t+1)}\right)^T \text{ and } X = \left(X_1,\ldots,X_n\right)^T.$$

-The fourth CM-step now uses the calculated parameter estimators of the last three CM-steps as fixed values.

$$\sigma^{2(t+1)} = \frac{1}{n}\left[\left(Y - X\beta^{(t+1)}\right)^T\left(Y - X\beta^{(t+1)}\right) - 2\left(Y - X\beta^{(t+1)}\right)^T\Pi^{(t)}\,DE^{(t+1)}\right.$$

The parameter estimator vector of the (t+1)-th iteration step of the ECM-algorithm can be written as

$$\theta^{(t+1)} = \left(a^{(t+1)}, d^{(t+1)}, \beta_0^{(t+1)},\ldots,\beta_{k-1}^{(t+1)}, \sigma^{(t+1)}\right)^T.$$

Each iteration step of the ECM-algorithm computes a conditional maximum rather than the (unconditioned) maximum of the EM-algorithm, that is to say the ECM-algorithm needs additional conditions in order to converge against a Maximum Likelihood estimator. An important condition is that the set of constraint functions is space filling.

Moreover it is of interest to discuss the cases where both the algorithms converge against the identical Maximum Likelihood estimators. Further discussion of this point is planned in later work.

4. The Asymptotic Variance-Covariance Matrix

Kao and Zeng (1997) described a method to calculate an asymptotic variance-covariance matrix for the specified model. The result of the iterative algorithm to estimate the parameters of the model is $\theta = \left(p, a, d, \sigma^2, \beta_0, ..., \beta_{k-1}\right)^T$. The parameter p is set as earlier estimated (or "known", e. g. from a lod score analysis). The posteriori-probabilities π_{ji} of the parameter estimation are known from the EM-algorithm as well.

For cases where the EM-algorithm is used, Louis (1982) derived a method to acquire the asymptotic variance-covariance matrix. The obtaining of the asymptotic variance-covariance matrix is equivalent to extracting the observed information of the incomplete problem.

The likelihood function of the (hypothetically) complete data set can be found by making the following considerations:

-The complete data problem can be regarded as a two stage hierarchical model.

-The QTL genotypes are the realizations of a two dimensional random variable $\left(x_j^*, z_j^*\right)$ (j=1,2,...,n), which are randomized from a trinomial experiment.

-Each realization of the random variable is assigned to one of the QTL genotypes $Q_1/Q_1, Q_1/Q_2$ or Q_2/Q_2 .

-The observations are normally distributed with a mean depends on the realization of the QTL genotype.

Then the likelihood function of the (hypothetical) complete data set is:

$$\lambda\left(Y_{com} \middle| p, a, d, \sigma^2, \beta\right) = \prod_{j=1}^{n} \left[p_{j1} f\left(y_j; \mu_{j1}, \sigma^2\right)^{-\frac{1}{2}\left(x_j^*+1\right)\left(z_j^*-\frac{1}{2}\right)} \times p_{j2} f\left(y_j; \mu_{j2}, \sigma^2\right)^{\left(x_j^*+1\right)\left(z_j^*+\frac{1}{2}\right)} \times p_{j3} f\left(y_j; \mu_{j3}, \sigma^2\right)^{\frac{1}{2}\left(x_j^*-1\right)\left(z_j^*-\frac{1}{2}\right)} \right]$$

with

$$-\frac{1}{2}(x_j^* + 1)(z_j^* - \frac{1}{2}) = \begin{cases} 1 \text{ for a QTL genotyp of } Q_1/Q_1 \\ 0 \text{ otherwise} \end{cases}$$

$$\left(x_j^* + 1\right)\left(z_j^* + \frac{1}{2}\right) = \begin{cases} 1 \text{ for a QTL genotyp of } Q_1/Q_2 \\ 0 \text{ otherwise} \end{cases}$$

$$\frac{1}{2}\left(x_j^* - 1\right)\left(z_j^* - \frac{1}{2}\right) = \begin{cases} 1 \text{ for a QTL genotyp of } Q_2/Q_2 \\ 0 \text{ otherwise} \end{cases}$$

Now, for a *fixed* sample the log-likelihood-function can be described as:

$$\ln(\lambda) = \sum_{j=1}^{n}\sum_{i=1}^{3}\left\{ \ln p_{ji} - \ln\left(\sqrt{2\pi}\right) - \ln(\sigma) - \frac{1}{2}\left(\frac{y_j - \mu_{ji}}{\sigma}\right)^2 \right\}$$

For independent observations the information matrices can be calculated as:

$$I_{obs}\left(\theta \,|\, \mathbf{Y}_{obs}\right) = I_{com} - I_{mis}$$

$$I_{com} = \sum_{j=1}^{n} E\left[-\frac{\partial^2 \ln\lambda\left(y_{(com,j)} \,|\, \theta\right)}{\partial\theta^2} \,\Big|\, y_{(obs,j)}, \theta\right]_{\theta}$$

$$I_{mis} = \sum_{j=1}^{n} E\left\{\left[\frac{\partial \ln\lambda\left(y_{(com,j)} \,|\, \theta\right)}{\partial\theta}\right]\left[\frac{\partial \ln\lambda\left(y_{(com,j)} \,|\, \theta\right)}{\partial\theta}\right]^{T} \,\Big|\, y_{(obs,j)}, \theta\right\}_{\theta}$$

$$+ \sum_{i\neq j}^{n} E\left\{\frac{\partial \ln\lambda\left(y_{(com,i)} \,|\, \theta\right)}{\partial\theta} \,\Big|\, y_{(obs,i)}, \theta\right\}_{\theta} E\left\{\left[\frac{\partial \ln\lambda\left(y_{(com,j)} \,|\, \theta\right)}{\partial\theta} \,\Big|\, y_{(obs,j)}, \theta\right]^{T}\right\}_{\theta}$$

And the asymptotic variance covariance matrix is calculated as the inverse of the observed information matrix:

$$\text{Cov}(\theta) = (I_{obs})^{-1}.$$

Obviously, to calculate these information matrices the first and second derivatives of $\ln(\lambda)$ have to be calculated. On the following page the tables 2 and 3 show the first and second derivatives of conditional probabilities of QTL genotypes given marker genotypes p_{ji} (j=1,2,...,n and i=1,2,3) for the possible marker genotypes of table 1. These matrices enable us to calculate matrices of first and second derivatives $P^{(1)} = \left(p_{ji}^{(1)}\right)$ and $P^{(2)} = \left(p_{ji}^{(2)}\right)$ in dependence of the *marker genotypes of a sample of n plants* (equivalent to matrix P in chapter 1). Then the information matrices I_{com} and I_{mis} can be calculated using the tables 4 and 5. Therefore I_{obs} and the desired asymptotic variance-covariance matrix can be calculated.

Table 2 : First derivatives of the conditional probabilities of QTL genotypes given marker genotypes:

no. of genotype of the marker	Q_1/Q_1	Q_1/Q_2	Q_2/Q_2
1	0	0	0
2	$-\frac{1}{1-p}$	$\frac{1}{p}$	0
3	$-\frac{2}{1-p}$	$\frac{1-2p}{p(1-p)}$	$\frac{2}{p}$
4	$\frac{1}{p}$	$-\frac{1}{1-p}$	0
5	$\frac{1-2p}{p(1-p)}$	$\frac{-2c(1-2p)}{1-2cp(1-p)}$	$\frac{1-2p}{p(1-p)}$
6	0	$-\frac{1}{1-p}$	$\frac{1}{p}$
7	$\frac{2}{p}$	$\frac{1-2p}{p(1-p)}$	$-\frac{2}{1-p}$
8	0	$\frac{1}{p}$	$-\frac{1}{1-p}$
9	0	0	0

Table 3: Second derivatives of the conditional probabilities of QTL genotypes given marker genotypes:

no. of genotype of the marker	Q_1/Q_1	Q_1/Q_2	Q_2/Q_2
1	0	0	0
2	$-\frac{1}{(1-p)^2}$	$-\frac{1}{p^2}$	0
3	$-\frac{2}{(1-p)^2}$	$\frac{-2p^2+2p-1}{p^2(1-p)^2}$	$-\frac{2}{p^2}$
4	$-\frac{1}{p^2}$	$-\frac{1}{(1-p)^2}$	0
5	$\frac{-2p^2+2p-1}{p^2(1-p)^2}$	$\frac{4c[1+c(-2p^2+2p-1)]}{[1-2cp(1-p)]^2}$	$\frac{-2p^2+2p-1}{p^2(1-p)^2}$
6	0	$-\frac{1}{(1-p)^2}$	$-\frac{1}{p^2}$
7	$-\frac{2}{p^2}$	$\frac{-2p^2+2p-1}{p^2(1-p)^2}$	$-\frac{2}{(1-p)^2}$
8	0	$-\frac{1}{p^2}$	$-\frac{1}{(1-p)^2}$
9	0	0	0

with $c=r_A^2/(r_A^2+(1-r_A^2))$

Table 4: $I_{com} = \sum_{j=1}^{n} E\left[-\frac{\partial^2 \ln \lambda(y_{(com,j)} \mid \theta)}{\partial\theta_s \partial\theta_t}\right]$ with $\theta_s, \theta_t \in \theta = (p, a, d, \sigma^2, \beta_0, \dots, \beta_{m-1})'$

$= \frac{1}{\sigma^2}\{\text{table}\}$

θ_s,θ_t	p	a	d	σ^2	β_0	...	β_{m-1}
p	$-\sigma^2 \sum_{j=1}^{n}\sum_{i=1}^{3} p_{ji}^{(2)} \pi_{ji}$	0	0	0	0	...	0
a	0	$\sum_{j=1}^{n} \pi_{j1} + \pi_{j3}$	$\sum_{j=1}^{n} \frac{1}{2}(-\pi_{j1} + \pi_{j3})$	$\frac{1}{\sigma^2}\sum_{j=1}^{n}\begin{pmatrix}(y_j - \mu_{j1})\pi_{j1} \\ -(y_j - \mu_{j'3})\pi_{j3}\end{pmatrix}$	$\sum_{j=1}^{n}(\pi_{j1} - \pi_{j3})X_{j1}$	...	$\sum_{j=1}^{n}(\pi_{j1} - \pi_{j3})X_{jk}$
d	0	$\sum_{j=1}^{n} \frac{1}{2}(-\pi_{j1} + \pi_{j3})$	$\frac{n}{4}$	$\frac{1}{2\sigma^2}\sum_{j=1}^{n}\begin{pmatrix}-(y_j - \mu_{j1})\pi_{j1} \\ +(y_j - \mu_{j2})\pi_{j2} \\ -(y_j - \mu_{j'3})\pi_{j3}\end{pmatrix}$	$\frac{1}{2}\sum_{j=1}^{n}(-\pi_{j1} + \pi_{j2} - \pi_{j3})X_{j1}$	...	$\frac{1}{2}\sum_{j=1}^{n}(-\pi_{j1} + \pi_{j2} - \pi_{j3})X_{jk}$
σ^2	0	$\frac{1}{\sigma^2}\sum_{j=1}^{n}\begin{pmatrix}(y_j - \mu_{j1})\pi_{j1} \\ -(y_j - \mu_{j'3})\pi_{j3}\end{pmatrix}$	$\frac{1}{2\sigma^2}\sum_{j=1}^{n}\begin{pmatrix}-(y_j - \mu_{j1})\pi_{j1} \\ +(y_j - \mu_{j2})\pi_{j2} \\ -(y_j - \mu_{j'3})\pi_{j3}\end{pmatrix}$	$\frac{n}{2\sigma^2}$	$\frac{1}{\sigma^2}\sum_{j=1}^{n}\sum_{i=1}^{3}(y_j - \mu_{j1})X_{j1}\pi_{ji}$	...	$\frac{1}{\sigma^2}\sum_{j=1}^{n}\sum_{i=1}^{3}(y_j - \mu_{j1})X_{jk}\pi_{ji}$
β_0	0	$\sum_{j=1}^{n}(\pi_{j1} - \pi_{j3})X_{j1}$	$\frac{1}{2}\sum_{j=1}^{n}(-\pi_{j1} + \pi_{j2} - \pi_{j3})X_{jk}$	$\frac{1}{\sigma^2}\sum_{j=1}^{n}\sum_{i=1}^{3}(y_j - \mu_{j1})X_{j1}\pi_{ji}$	$\sum_{j=1}^{n}\sum_{i=1}^{3} X_{j1}X_{j1}$	...	$\sum_{j=1}^{n} X_{jk}X_{j1}$
...	...	...	...	...	...	...	...
β_{k-1}	0	$\sum_{j=1}^{n}(\pi_{j1} - \pi_{j3})X_{jk}$	$\frac{1}{2}\sum_{j=1}^{n}(-\pi_{j1} + \pi_{j2} - \pi_{j3})X_{jk}$	$\frac{1}{\sigma^2}\sum_{j=1}^{n}\sum_{i=1}^{3}(y_j - \mu_{j1})X_{jk}\pi_{ji}$	$\sum_{j=1}^{n} X_{j1}X_{jk}$	...	$\sum_{j=1}^{n} X_{jk}X_{jk}$

Table 5: $I_{mis} = \sum_{j=1}^{n} E\left\{\left[\frac{\partial \ln \lambda(y_{(com,j)} \mid \theta)}{\partial \theta_s}\right]\left[\frac{\partial \ln \lambda(y_{(com,j)} \mid \theta)}{\partial \theta_t}\right]' \middle| y_{(obs,j)}, \theta\right\} + \sum_{j=1}^{n} E\left\{\left[\frac{\partial \ln \lambda(y_{(com,j)} \mid \theta)}{\partial \theta_s}\right] \middle| y_{(obs,j)}, \theta\right\}_{\theta=\theta^*} E\left\{\left[\frac{\partial \ln \lambda(y_{(com,j)} \mid \theta)}{\partial \theta_t}\right] \middle| y_{(obs,j)}, \theta\right\} = \frac{1}{\sigma^2}\{\text{table}\}$

θ_s, θ_t	p	a	d
p	$\sum_{j=1}^{n}\sum_{i=1}^{3}\left(p_{ji}^{(1)}\right)^2 \pi_{ji} +$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3} p_{mi}^{(1)}\pi_{mi}\right]\left[\sum_{i=1}^{3} p_{ji}^{(1)}\pi_{ji}\right]$	$\sum_{j=1}^{n} p_{j1}^{(1)}\pi_{j1}(y_j-\mu_{j1}) - p_{j3}^{(1)}\pi_{j3}(y_j-\mu_{j3}) +$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3} p_{ji}^{(1)}\pi_{ji}\right][(y_m-\mu_{m1})\pi_{m1} - (y_m-\mu_{m3})\pi_{m3}]$	$\frac{1}{2}\sum_{j=1}^{n}\left(-(y_j-\mu_{j1})\pi_{j1}p_{j1}^{(1)} + (y_j-\mu_{j2})\pi_{j2}p_{j2}^{(1)} - (y_j-\mu_{j3})\pi_{j3}p_{j3}^{(1)}\right) +$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3} p_{ji}^{(1)}\pi_{ji}\right]\frac{1}{2}[-(y_m-\mu_{m1})\pi_{m1} + (y_m-\mu_{m2})\pi_{m2} - (y_m-\mu_{m3})\pi_{m3}]$
a	$\sum_{j=1}^{n} p_{j1}^{(1)}\pi_{j1}(y_j-\mu_{j1}) - p_{j3}^{(1)}\pi_{j3}(y_j-\mu_{j3}) +$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3} p_{ji}^{(1)}\pi_{ji}\right][(y_m-\mu_{m1})\pi_{m1} - (y_m-\mu_{m3})\pi_{m3}]$	$\sum_{j=1}^{n}\left((y_j-\mu_{j1})^2\pi_{j1} + (y_j-\mu_{j3})^2\pi_{j3}\right) +$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}[(y_j-\mu_{j1})\pi_{j1} - (y_j-\mu_{j3})\pi_{j3}][(y_m-\mu_{m1})\pi_{m1} - (y_m-\mu_{m3})\pi_{m3}]$	$\frac{1}{2}\sum_{j=1}^{n}\left(-(y_j-\mu_{j1})^2\pi_{j1} + (y_j-\mu_{j3})^2\pi_{j3}\right) +$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}[(y_j-\mu_{j1})\pi_{j1} - (y_j-\mu_{j3})\pi_{j3}]\frac{1}{2}\begin{bmatrix}-(y_m-\mu_{m1})\pi_{m1}\\ +(y_m-\mu_{m2})\pi_{m2} - (y_m-\mu_{m3})\pi_{m3}\end{bmatrix}$
d	$\frac{1}{2}\sum_{j=1}^{n}\left(-(y_j-\mu_{j1})\pi_{j1}p_{j1}^{(1)} + (y_j-\mu_{j2})\pi_{j2}p_{j2}^{(1)} - (y_j-\mu_{j3})\pi_{j3}p_{j3}^{(1)}\right)$ $+\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3} p_{ji}^{(1)}\pi_{ji}\right]\frac{1}{2}\begin{bmatrix}-(y_m-\mu_{m1})\pi_{m1} +\\ (y_m-\mu_{m2})\pi_{m2} - (y_m-\mu_{m3})\pi_{m3}\end{bmatrix}$	$\frac{1}{2}\sum_{j=1}^{n}\left(-(y_j-\mu_{j1})^2\pi_{j1} + (y_j-\mu_{j3})^2\pi_{j3}\right) +$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\begin{bmatrix}(y_j-\mu_{j1})\pi_{j1}\\ -(y_j-\mu_{j3})\pi_{j3}\end{bmatrix}\frac{1}{2}\begin{bmatrix}-(y_m-\mu_{m1})\pi_{k1} +\\ (y_m-\mu_{m2})\pi_{m2} - (y_m-\mu_{m3})\pi_{m3}\end{bmatrix}$	$\frac{1}{4}\sum_{j=1}^{n}(y_j-\mu_{ji})^2\pi_{ji} +$ $\frac{1}{4}\sum_{\substack{j,m=1\\ m\neq j}}^{n}[-(y_j-\mu_{j1})\pi_{j1} + (y_j-\mu_{j2})\pi_{j2} - (y_j-\mu_{j3})\pi_{j3}]\begin{bmatrix}-(y_m-\mu_{m1})\pi_{m1}\\ +(y_m-\mu_{m2})\pi_{m2} - (y_m-\mu_{m3})\pi_{m3}\end{bmatrix}$
σ^2	$\sum_{j=1}^{n}\left(\sum_{i=1}^{3}\left(\frac{(y_j-\mu_{ji})^2}{2\sigma^2} - \frac{1}{2}\right)p_{ji}^{(1)}\pi_{ji}\right) +$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3} p_{mi}^{(1)}\pi_{mi}\right]\left[\sum_{i=1}^{3}\left(\frac{(y_j-\mu_{ji})^2}{2\sigma^2} - \frac{1}{2}\right)\pi_{ji}\right]$	$\sum_{j=1}^{n}\left(\frac{(y_j-\mu_{j1})^2}{2\sigma^2} - \frac{1}{2}\right)(y_j-\mu_{j1})\pi_{j1} - \left(\frac{(y_j-\mu_{j3})^2}{2\sigma^2} - \frac{1}{2}\right)(y_j-\mu_{j3})\pi_{j3}$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}[(y_m-\mu_{m1})\pi_{m1} - (y_m-\mu_{m3})\pi_{m3}]\left[\sum_{i=1}^{3}\left(\frac{(y_l-\mu_{li})^2}{2\sigma^2} - \frac{1}{2}\right)\pi_{li}\right]$	$\frac{1}{2}\sum_{j=1}^{n} -\left(\frac{(y_j-\mu_{j1})^2}{2\sigma^2} - \frac{1}{2}\right)(y_j-\mu_{j1})\pi_{j1} + \left(\frac{(y_j-\mu_{j2})^2}{2\sigma^2} - \frac{1}{2}\right)(y_j-\mu_{j2})\pi_{j2}$ $-\left(\frac{(y_j-\mu_{j3})^2}{2\sigma^2} - \frac{1}{2}\right)(y_j-\mu_{j3})\pi_{j3}$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\frac{1}{2}[-(y_m-\mu_{m1})\pi_{m1} + (y_m-\mu_{m2})\pi_{m2} - (y_m-\mu_{m3})\pi_{m3}]\left[\sum_{i=1}^{3}\left(\frac{(y_l-\mu_{li})^2}{2\sigma^2} - \frac{1}{2}\right)\pi_{ji}\right]$
β_0	$\sum_{j=1}^{n}\left(\sum_{i=1}^{3}(y_j-\mu_{ji})X_{j1}p_{ji}^{(1)}\pi_{ji}\right) +$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3} p_{mi}^{(1)}\pi_{mi}\right]\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{j1}\pi_{ji}\right]$	$\sum_{j=1}^{n}(y_j-\mu_{j1})^2X_{j1}\pi_{j1} - (y_j-\mu_{j3})^2X_{j1}\pi_{j3}$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}[(y_m-\mu_{m1})\pi_{m1} - (y_m-\mu_{m3})\pi_{m3}]\left[\sum_{i=1}^{3}(y_l-\mu_{li})X_{l1}\pi_{li}\right]$	$\frac{1}{2}\sum_{j=1}^{n} -(y_j-\mu_{j1})^2X_{j1}\pi_{j1} + (y_j-\mu_{j2})^2X_{j1}\pi_{j2} - (y_j-\mu_{j3})^2X_{j3}\pi_{j3}$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\frac{1}{2}[-(y_m-\mu_{m1})\pi_{m1} + (y_m-\mu_{m2})\pi_{m2} - (y_m-\mu_{m3})\pi_{m3}]\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{j1}\pi_{ji}\right]$
…	…	…	…
β_{k-1}	$\sum_{j=1}^{n}\left(\sum_{i=1}^{3}(y_j-\mu_{ji})X_{jk}p_{ji}^{(1)}\pi_{ji}\right) +$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3} p_{mi}^{(1)}\pi_{mi}\right]\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{jk}\pi_{ji}\right]$	$\sum_{j=1}^{n}(y_j-\mu_{j1})^2X_{jk}\pi_{j1} - (y_j-\mu_{j3})^2X_{jk}\pi_{j3}$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}[(y_j-\mu_{j1})\pi_{j1} - (y_j-\mu_{j3})\pi_{j3}]\left[\sum_{i=1}^{3}(y_m-\mu_{mi})X_{km}\pi_{mi}\right]$	$\frac{1}{2}\sum_{j=1}^{n} -(y_j-\mu_{j1})^2X_{jk}\pi_{j1} + (y_j-\mu_{j2})^2X_{jk}\pi_{j2} - (y_j-\mu_{j3})^2X_{jk}\pi_{j3}$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\frac{1}{2}[-(y_m-\mu_{m1})\pi_{mi} + (y_m-\mu_{m2})\pi_{m2} - (y_m-\mu_{m3})\pi_{m3}]\left[\sum_{i=1}^{3}(y_l-\mu_{ji})X_{jk}\pi_{ji}\right]$

θ_s,θ_t	σ^2	β_0	…	β_{k-1}
p	$\sum_{j=1}^{n}\left(\sum_{i=1}^{3}\left(\frac{(y_j-\mu_{ji})^2}{2\sigma^2}-\frac{1}{2}\right)p_{ji}^{(1)}\pi_{ji}\right)+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3}p_{mi}^{(1)}\pi_{mi}\right]\left[\sum_{i=1}^{3}\left(\frac{(y_j-\mu_{ji})^2}{2\sigma^2}-\frac{1}{2}\right)\pi_{ji}\right]$	$\sum_{j=1}^{n}\left(\sum_{i=1}^{3}(y_j-\mu_{ji})X_{j1}p_{ji}^{(1)}\pi_{ji}\right)+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3}p_{mi}^{(1)}\pi_{mi}\right]\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{j1}\pi_{ji}\right]$	…	$\sum_{j=1}^{n}\left(\sum_{i=1}^{3}(y_j-\mu_{ji})X_{jk}p_{ji}^{(1)}\pi_{ji}\right)+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3}p_{mi}^{(1)}\pi_{mi}\right]\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{jk}\pi_{ji}\right]$
a	$\sum_{j=1}^{n}\left(\frac{(y_j-\mu_{j1})^2}{2\sigma^2}-\frac{1}{2}\right)(y_j-\mu_{j1})\pi_{j1}-\left(\frac{(y_j-\mu_{j3})^2}{2\sigma^2}-\frac{1}{2}\right)(y_j-\mu_{j3})\pi_{j3}$ $+\sum_{\substack{j,m=1\\ m\neq j}}^{n}[(y_m-\mu_{m1})\pi_{m1}-(y_m-\mu_{m3})\pi_{m3}]\left[\sum_{i=1}^{3}\left(\frac{(y_j-\mu_{ji})^2}{2\sigma^2}-\frac{1}{2}\right)\pi_{ji}\right]$	$\sum_{j=1}^{n}(y_j-\mu_{j1})^2X_{j1}\pi_{j1}-(y_j-\mu_{j3})^2X_{j1}\pi_{j3}+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}[(y_m-\mu_{m1})\pi_{m1}-(y_m-\mu_{m3})\pi_{m3}]\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{j1}\pi_{ji}\right]$	…	$\sum_{j=1}^{n}(y_j-\mu_{j1})^2X_{jk}\pi_{j1}-(y_j-\mu_{j3})^2X_{jk}\pi_{j3}$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}[(y_j-\mu_{j1})\pi_{j1}-(y_j-\mu_{j3})\pi_{j3}]\left[\sum_{i=1}^{3}(y_m-\mu_{mi})X_{km}\pi_{mi}\right]$
d	$\frac{1}{2}\left(\sum_{j=1}^{n}-\left(\frac{(y_j-\mu_{j1})^2}{2\sigma^2}-\frac{1}{2}\right)(y_j-\mu_{j1})\pi_{j1}+\left(\frac{(y_j-\mu_{j2})^2}{2\sigma^2}-\frac{1}{2}\right)(y_j-\mu_{j2})\pi_{j2}-\left(\frac{(y_j-\mu_{j3})^2}{2\sigma^2}-\frac{1}{2}\right)(y_j-\mu_{j3})\pi_{j3}\right)$ $+\sum_{\substack{j,m=1\\ m\neq j}}^{n}\frac{1}{2}\begin{bmatrix}-(y_m-\mu_{m1})\pi_{m1}\\ +(y_m-\mu_{m2})\pi_{m2}-(y_m-\mu_{m3})\pi_{m3}\end{bmatrix}\cdot\left[\sum_{i=1}^{3}\left(\frac{(y_j-\mu_{ji})^2}{2\sigma^2}-\frac{1}{2}\right)\pi_{ji}\right]$	$\frac{1}{2}\sum_{j=1}^{n}-(y_j-\mu_{j1})^2X_{j1}\pi_{j1}+(y_j-\mu_{j2})^2X_{j1}\pi_{j2}$ $-(y_j-\mu_{j3})^2X_{j3}\pi_{j3}$ $+\sum_{\substack{j,m=1\\ m\neq j}}^{n}\frac{1}{2}\begin{bmatrix}-(y_j-\mu_{j1})\pi_{j1}\\ +(y_j-\mu_{j2})\pi_{j2}\\ -(y_j-\mu_{j3})\pi_{j3}\end{bmatrix}\cdot\left[\sum_{i=1}^{3}(y_m-\mu_{mi})X_{j1}\pi_{mi}\right]$	…	$\frac{1}{2}\sum_{j=1}^{n}-(y_j-\mu_{j1})^2X_{jk}\pi_{j1}+(y_j-\mu_{j2})^2X_{jk}\pi_{j2}-(y_j-\mu_{j3})^2X_{jk}\pi_{j3}$ $\sum_{\substack{j,k=1\\ k\neq j}}^{n}\frac{1}{2}[-(y_m-\mu_{m1})\pi_{mi}+(y_m-\mu_{m2})\pi_{m2}-(y_m-\mu_{m3})\pi_{m3}]$ $\cdot\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{jk}\pi_{ji}\right]$
σ^2	$\sum_{j=1}^{n}\left[\sum_{i=1}^{3}\left(\frac{(y_j-\mu_{ji})^2}{2\sigma^2}-\frac{1}{2}\right)^2\pi_{ji}\right]+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3}\left(\frac{(y_m-\mu_{mi})^2}{2\sigma^2}-\frac{1}{2}\right)\pi_{mi}\right]\left[\sum_{i=1}^{3}\left(\frac{(y_j-\mu_{ji})^2}{2\sigma^2}-\frac{1}{2}\right)\pi_{ji}\right]$	$\sum_{j=1}^{n}\sum_{i=1}^{3}\left(\frac{(y_j-\mu_{ji})^2}{2\sigma^2}-\frac{1}{2}\right)^2(y_j-\mu_{ji})X_{j1}\pi_{ji}+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3}\left(\frac{(y_m-\mu_{mi})^2}{2\sigma^2}-\frac{1}{2}\right)\pi_{mi}\right]\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{j1}\pi_{ji}\right]$	…	$\sum_{j=1}^{n}\sum_{i=1}^{3}\left(\frac{(y_j-\mu_{ji})^2}{2\sigma^2}-\frac{1}{2}\right)^2(y_j-\mu_{ji})X_{jk}\pi_{ji}+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3}\left(\frac{(y_m-\mu_{mi})^2}{2\sigma^2}-\frac{1}{2}\right)\pi_{mi}\right]\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{jk}\pi_{ji}\right]$
β_0	$\sum_{j=1}^{n}\left[\sum_{i=1}^{3}\left(\frac{(y_j-\mu_{ji})^2}{2\sigma^2}-\frac{1}{2}\right)^2\pi_{ji}\right](y_j-\mu_{ji})X_{j1}\pi_{ji}+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3}\left(\frac{(y_m-\mu_{mi})^2}{2\sigma^2}-\frac{1}{2}\right)\pi_{mi}\right]\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{j1}\pi_{ji}\right]$	$\sum_{j=1}^{n}(y_j-\mu_{ji})^2X_{j1}X_{j1}\pi_{ji}+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{j1}\pi_{ji}\right]\left[\sum_{i=1}^{3}(y_m-\mu_{mi})X_{m1}\pi_{mi}\right]$	…	$\sum_{j=1}^{n}(y_j-\mu_{ji})^2X_{j1}X_{jm}\pi_{ji}+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{j1}\pi_{ji}\right]\left[\sum_{i=1}^{3}(y_m-\mu_{mi})X_{km}\pi_{mi}\right]$
…			…	
β_{k-1}	$\sum_{j=1}^{n}\left[\sum_{i=1}^{3}\left(\frac{(y_j-\mu_{ji})^2}{2\sigma^2}-\frac{1}{2}\right)^2\pi_{ji}\right](y_j-\mu_{ji})X_{jk}\pi_{ji}+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3}\left(\frac{(y_m-\mu_{mi})^2}{2\sigma^2}-\frac{1}{2}\right)\pi_{mi}\right]\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{jk}\pi_{ji}\right]$	$\sum_{j=1}^{n}(y_j-\mu_{ji})^2X_{j1}X_{jk}\pi_{ji}+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{j1}\pi_{ji}\right]\left[\sum_{i=1}^{3}(y_m-\mu_{mi})X_{km}\pi_{mi}\right]$	…	$\sum_{j=1}^{n}(y_j-\mu_{ji})^2X_{jk}X_{jk}\pi_{ji}+$ $\sum_{\substack{j,m=1\\ m\neq j}}^{n}\left[\sum_{i=1}^{3}(y_j-\mu_{ji})X_{jk}\pi_{ji}\right]\left[\sum_{i=1}^{3}(y_m-\mu_{mi})X_{km}\pi_{mi}\right]$

5. Discussion and outlook

In this work we have described two methods for the estimation of genetic parameters and a way to calculate an asymptotic variance-covariance matrix. In a forthcoming paper examples will be published showing that the EM-algorithm and the ECM-algorithm as described by Kao and Zeng may or may not converge against different maxima. In other cases the ECM-algorithm converges against stationary values that are not maxima, while the EM-algorithm converges against a maximum.

The calculation of the asymptotic variance-covariance-matrix is similar for both parameter estimation methods and enables us to calculate confidence intervals. Today, a range of widely applicable software programs for QTL-analysis exists. PLABQTL (Utz and Melchinger (1996)) used a multiple regression approach with flanking markers according to the procedure described by Haley and Knott (1992). Basten et al (1999) have implemented the ECM-algorithm as described by Kao and Zeng (1997) for QTL-analysis.

Melchinger et al. (1998) evaluated testcross progenies of 344 F_3 lines in combination with two unrelated testers plus additional testcross progenies from an independent but smaller sample of 107 F_3 lines from the same cross in combination with the same two testers for grain yield and four other important agronomic traits. For a more detailed statistical analysis of this data set A.E. Melchinger and H.F. Utz from the Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, provided plant height measurements of an F_2 –population of maize, which were genotyped for a total of 89 marker loci. This data set is (like other practical applications) based on so called adjusted means. Urfer et al (1999) presented several ways to calculate adjusted means in α-designs.

The aim of our further statistical approach is to find Maximum Likelihood estimates for QTL locations and effects including their estimated standard errors using the described and further methods. Recently, Kao, Zeng and Teasdale (1999) presented a new statistical approach for interval mapping, called multiple marker interval mapping (MIM). It uses multiple marker intervals simultaneously to fit multiple putative QTLs directly in the model for mapping QTLs. Here the ECM-algorithm is used as well. To integrate our estimation procedure in this approach seems to be a promising field for further statistical research.

Another area of challenging statistical problems is the mapping of QTLs for cellular defects in glucose and fatty acid metabolism. Al-Majali et al (1999) constructed a radiation hybrid map of the proximal region of rat chromosome 4. This map will facilitate identification of genes underlying cardiovascular and metabolic QTLs and also provides an interesting comparison of synteny relationships between the rats, mouse and human genomes.

Acknowledgement

We would like to thank the German Research Foundation (DFG) for financial support of the Graduate College and the Collaborative Research Centre at our Department of Statistics.

References:

Al-Majali, K. M., Glazier, A. M., Norsworthy, P. J., Wahid, F. N., Cooper, L. D., Wallace, C. A., Scott, J., Lausen, B. and Aitman, T. J. 1999: A high-resolution radiation hybrid map to the proximal region of rat chromosome 4. *Mammalian Genome* **19**, 471-476

Basten, C.J., Weir, B.S. and Zeng, Z.-B., 1999: QTL Cartographer, Version 1.13. *Department of Statistics, North Carolina State University,* Raleigh, NC.

Dempster, A. P., Laird, N. M. and Rubin, D. B. 1971: Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Statist. Soc.* **B 39**, 1-38.

Emrich, K., 1999: Estimation of effects of genes in plant breeding. **Master Thesis.** *Department of Statistics, University of Dortmund.*

Fisch, R. D., Ragot, M. and Gay, G. 1996: A Generalization of the Mixture Model in the Mapping of Quantitative Trait Loci for Progeny From a Biparental Cross of Inbred Lines, *Genetics* **143**, S. 571-577

Haley, C. S., and S. A. Knott, 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315-234.

Jansen, R.C., 1996: A general Monte Carlo method for mapping multiple quantitative trait loci. *Genetics* **142**, 305-311.

Kao, C.-H., and Zeng, Z.-B., 1997: General formulas for obtaining the MLE's and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. *Biometrics* **53**. 653-665.

Kao, C.-H., Zeng, Z.-B. and Teasdale, Robert D., 1999: Multiple Interval Mapping for Quantitative Trait Loci. *Genetics* **152**, 1203-1216.

Lander, E. S. and Botstein, D. 1989: Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps, *Genetics* **121**, S. 185-199.

Louis, T. A. 1982: Finding the observed information matrix when using the EM-algorithm. . *J. Roy. Statist. Soc., Ser. B* **44**, 226-233

Melchinger, A.E., Utz, H.F. and Schön, C.C., 1998: Quantitative trait locus (QTL) mapping using different testers and independent populaton samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* **149**, 383-403.

Meng, X.-L. and D. B. Rubin, 1993. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* **80**: 267–268.

Selinski, S. and Urfer, W., 1998: Interindividual and interoccasion variability of toxicokinetic parameters in population models. *Technical Report* **38/1998**, University of Dortmund. Availabe from the world wide web: http://www.statistik.uni-dortmund.dc/sfb475/sfblit.htm.

Selinski, S., Golka, K., Bolt, H. M., Urfer, W. 1999: Estimation of toxicokinetic parameters in population models for inhalation studies with ethylene. Submitted to *Environmetrics*.

Stephens, D.A. and Fisch, R.D., 1998: Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. *Biometrics* **54**, 1334-1347.

Urfer, W. Mejza, S. and Hering, F. 1999: Quantitative trait loci mapping in plant genetics by α-design experiments and molecular genetic marker systems. *Technical Report* **34/1999**, University of Dortmund. Availabe from the world wide web: http://www.statistik.uni-dortmund.dc/sfb475/sfblit.htm.

Utz, H.F. and Melchinger, A.E., 1996: PLABQTL: A program for composite interval map ping of QTL. *J. Quant. Trait Loci*. **2**. 1-5.

Wu, C. F. J. 1983: On the convergence properties of the EM algoritm. *Ann. Statist.* **11**, 95-103.

Zeng, Z.-B. 1994: Precision Mapping of Quantitative Trait Loci. *Genetics* **136**, S. 1457-1468.