

有限差分近似法在數量性狀基因座定位上 最大概似估值變異矩陣估算上之應用

江欣容、劉 清*

國立臺灣大學農藝系

摘要

在數量性狀基因座的定位與分析上，混合模式的最大概似估值可以 EM 法 (expectation maximization, Dempster *et al.* 1977)、ECM 法 (expectation-conditional maximization, Meng and Rubin 1993)、IRLS 法 (iteratively reweighted least squares) 等方法求出，但最大概似估值的漸近變異矩陣因為需要利用概似函數的二次微分式求算，又由於一般混合模式之概似函數的二次微分式相當複雜不易導出，因此如何計算混合模式之最大概似估值的漸近變異矩陣為一相當重要的課題。因為僅有估值卻不知估值的變異，將無法評估根據此估值所做統計推論的可靠性。本研究提出以數值方法上的有限差分近似法 (finite difference approximation method) 來計算概似函數的近似二次微分式及最大概似估值的漸近變異矩陣。為證實利用有限差分近似法所算之漸近變異矩陣結果的正確性，特將模擬之常態分布、二項分布與卜瓦松分布的 F_2 子代的數量性狀資料，分別以概似函數的二次微分式 (analytic derivative) 與有限差分近似法之近似二次微分式 (finite difference approximation of second order derivative) 作計算，並比較兩者之計算結果。模擬結果證實無論數量性狀

資料分布為何，利用概似函數的二次微分式與有限差分近似法的近似二次微分式，兩者所算出的漸近變異矩陣幾乎完全相同。因此在數量性狀基因座的定位與分析上，當各類數學模式之概似函數無已知的二次微分式或不易以解析方法導出時，建議可先以 EM 法、ECM 法、IRLS 法等方法來計算模式之最大概似估值。一旦求得最大概似估值的解析解，則可利用數值方法上的有限差分近似法來計算概似函數的近似二次微分式及最大概似估值的漸近變異矩陣。

關鍵詞：數量性狀基因座、遺傳標識、簡單區間定位法、綜合區間定位法、有限差分近似法、最大概似估值、漸近變異矩陣。

Application of Finite Difference Approximation on the Estimation of Dispersion Matrix of Maximum Likelihood Estimates for Mapping and Analysis of Quantitative Trait Loci (QTL)

Hsin-jung Chiang and Ching Liu*

Department of Agronomy, National Taiwan University, Taipei 10617, Taiwan ROC

ABSTRACT

For mapping and analysis of quantitative trait loci (QTL), the maximum likelihood (ML) estimates of parameters of mixture model can be calculated via EM, ECM, IRLS or other methods, whereas the asymptotic dispersion matrix of ML estimates requires the second order derivative of likelihood function which is generally

* 通信作者, m485@ccms.ntu.edu.tw

投稿日期：2006 年 4 月 5 日

接受日期：2006 年 4 月 23 日

作物、環境與生物資訊 4:201-214 (2007)

Crop, Environment & Bioinformatics 4:201-214 (2007)

189 Chung-Cheng Rd., Wufeng, Taichung Hsien 41301, Taiwan ROC

complicated and not easily derivable. The calculation of asymptotic matrix of ML estimates is important in that it enables us to evaluate the plausibility of our statistical inference based on ML estimates. This study proposed the calculation of the asymptotic dispersion matrix of ML estimates by the finite difference approximation method if the second order derivative of likelihood function was too complicated to derive. To verify the correctness of the asymptotic dispersion matrix calculated by the finite difference approximation method, the asymptotic dispersion matrix of the simulated normal, binary and Poisson distributed F_2 intercross data were calculated by analytical formula and the finite difference approximation of second order derivative. Results from the simulated F_2 intercross data indicated that the asymptotic dispersion matrices calculated by the finite difference approximation method were very close to those of the analytical formula. Therefore, if the second order derivatives of likelihood functions under various kinds of mathematical model settings are too complicated to derive, it is suggested to calculate the ML estimates via EM, ECM, IRLS or other methods which do not require the second order derivative. Once the ML estimates is available, the asymptotic dispersion matrix of ML estimates can be calculated by finite difference approximation method.

Key words: Quantitative trait loci (QTL), Genetic marker, Simple interval mapping (SIM), Composite interval mapping (CIM), Finite difference approximation method, Maximum likelihood estimates, The asymptotic dispersion matrix.

前言

性狀(character or trait)為生物體可被觀察到的特徵，是由基因與環境相互作用所形成的，可分為質量性狀(qualitative traits)與數量性狀(quantitative traits)兩種。質量性狀指個體在表型上可明顯區分為不同類型的性狀，如豌豆花的花色有紫色或白色、人類耳垂有分離臉頰與緊貼臉頰...等，此類性狀的變異在群體內為不連續的，也不易受環境的

影響，大多由一對或少數幾對的主效基因(major genes)或寡基因(oligogenes)所控制。數量性狀是指個體在表型上不能明顯劃分為不同類型的性狀，如作物的產量、倒伏性、抗病性、株高，及人類的身高、體重、智商...等，此類性狀在群體內以連續性變異(continuous variation)的型態存在，易受環境的影響，常由數個微效基因(minor genes)或多基因(polygenes)的累積作用所控制。許多具有重要經濟價值的性狀是以連續性變異的型態存在。早期在育種上是以族群改良(population improvement)的方式來累積優良及有經濟價值的性狀頻度，但此法耗時且易受環境的影響。

隨著分子生物學的快速發展，遺傳學者開始使用分子標識(molecular markers)，如 RFLPs (restriction fragment length polymorphisms)、microsatellite、RAPD (randomly amplified polymorphic DNA)、VNTR (variable number of tandem repeat)等，來定位並分析數量性狀基因座(quantitative trait loci, QTL)的位置與效應，藉著分子標識大多具有共顯性且又不易受環境影響，將控制同一性狀的各個數量基因視為獨立，解析為單一基因，可簡化原本複雜的情形，再利用統計模式分析並找出 QTL 最有可能存在的位置。若能藉此方法辨識出 QTL 的位置，再利用基因轉殖的技術將具有此一優良數量性狀基因座的染色體片段轉殖回親代，便可提高優良性狀的頻度，並能改善傳統方法因易受環境影響而存有不確定性，亦能有效的改進選拔育種的效率。

在定位 QTL 的統計方法上，最初是對所有染色體上標識因子的基因型(marker genotype)做迴歸分析，以偵測在標識因子附近是否有 QTL 的存在(Soller and Brody 1976)。Lander *et al.* (1989)提出區間定位法(interval mapping)，利用目標區間兩翼標識因子(flanking marker)的資訊來定位 QTL，而將 QTL 的定位與分析由點(point)的概念帶

入到區間(interval)的概念, 本文稱此法為簡單區間定位法(simple interval mapping, SIM)。Jansen (1992)和 Zeng (1993, 1994)結合簡單區間定位法與染色體上其他標識因子的訊息, 提出綜合區間定位法(composite interval mapping, CIM); 藉由控制目標區間外之其他標識因子的遺傳變異殘差, 對目標區間進行數量性狀基因座的定位分析。在遺傳學上, 若控制單一性狀的兩個或兩個以上的基因座間有交感作用存在, 則稱此交感作用為上位(epistasis); 若一基因座能同時影響多種性狀的表現, 則稱此現象為多效(pleiotropy)。Kao (1995)將綜合區間定位法推廣, 進一步的討論單一性狀且包含上位性效應(epistatic effect)的遺傳模式。Jiang and Zeng (1995)則將綜合區間定位法推廣至單一數量性狀基因座之多效性的遺傳模式。

因上述所有方法之統計模式皆為一種混合模式, 沒有封閉式的解(closed form solution), 因此必須以最大概似法, 如牛頓法(Newton-Raphson method)求解。雖牛頓法能同時求得最大概似估值與最大概似估值漸近變異矩陣(asymptotic dispersion matrix), 但牛頓法必須利用概似函數的二次微分式計算, 因此若概似函數的二次微分式不易導出或公式太複雜, 則只能以其他不需用到概似函數二次微分式的計算方法, 如 EM 法、ECM 法、IRLS 法或其他方法計算, 但最大概似估值的漸近變異矩陣仍需利用概似函數的二次微分式計算, 故以 EM 等方法雖可求算最大概似估值的解, 卻無法再計算最大概似估值的漸近變異矩陣(Liu 1998)。大部分討論 QTL 定位法的文獻多忽略了最大概似估值之漸近變異矩陣的計算, 但僅有估值卻不知估值的變異, 則無法評估根據此估值所做統計推論的可靠性。

本研究提出以數值方法上的有限差分近似法來計算各類數學模式之概似函數二次微分式與最大概似估值的漸近變異矩陣。本文以簡單區間定位法(SIM) (Lander *et al.* 1989)

與綜合區間定位法(CIM) (Zeng 1993)等兩種最常用的數量性狀基因座定位法來建立數量性狀表型值與數量性狀基因型之間的關係, 介紹當數量性狀之機率分布為常態分布、二項分布、卜瓦松分布等指數分布族(exponential family)時, 如何利用有限差分近似法計算混合模式概似函數的近似二次微分式。最後將常態分布、二項分布與卜瓦松分布之 F_2 子代的數量性狀模擬資料, 分別利用 Liu (1997, 1998)所推導之混合模式概似函數的二次解析微分式(analytic derivatives)與有限差分近似法的近似二次微分式, 計算最大概似估值的漸近變異矩陣, 並比較兩者之計算結果。

材料與方法

統計模式

在數量性狀基因座的定位與分析上, 可用於估算混合模式的最大概似估值的方法甚多, 如 EM 法、ECM 法、IRLS 法等計算方法, 但最大概似估值的漸近變異矩陣因為需要利用概似函數的二次微分式求算, 又由於一般混合模式之概似函數的二次微分式相當複雜而不易導出。本文以 F_2 子代為對象, 在不考慮上位性與多效性的情況下, 討論以簡單區間定位法與綜合區間定位法所建構的混合模式之概似函數二次微分式與最大概似估值之漸近變異矩陣的計算; 說明數量性狀之機率分布為常態分布、二項分布、卜瓦松分布等指數分布族時, 各類分布之概似函數二次微分式的推導過程, 並介紹本研究將如何應用數值方法上的有限差分近似法, 計算概似函數之近似二次微分式, 進而求出最大概似估值的漸近變異矩陣。

樣品資料之簡單區間定位法與綜合區間定位法的廣義線性統計模式如下:

$$\begin{aligned} \text{(SIM): } \eta_i &= \eta(\mu_i) = m + a^* x_i^* + d^* z_i^* \\ i &= 1, 2, \dots, n \\ \text{(CIM): } \eta_i &= \eta(\mu_i) = m + a^* x_i^* + d^* z_i^* + \mathbf{b}' \mathbf{x}_i \end{aligned} \quad (1)$$

$$i = 1, 2, \dots, n \quad (2)$$

其中 η_i 為將連續型數量性狀 μ_i 轉換成線性估算元的連結函數。 m 為常數項， a^* 為數量性狀基因座的累加性效應， x_i^* 為累加性效應的指標變數，當 QTL 基因型為 qq、Qq、QQ 時，其值分別為 0、1、2。 d^* 為數量性狀基因座的顯性效應。 z_i^* 為顯性效應的指標變數，當 QTL 基因型為同質型(QQ or qq) 時，其值為 0；當 QTL 基因型為異質型(Qq) 時，其值為 1。 $\mathbf{x}_i$ 為除了兩翼標識因子外的所有標識因子之各項效應指標變數向量，包含累加性效應與顯性效應。 $\mathbf{b}$ 為除了兩翼標識因子外的所有標識因子之各項效應指標變數的偏迴歸係數向量。

觀察(1)、(2)之簡單區間定位法(SIM)與綜合區間定位法(CIM)的統計模式，由於兩種區間定位法的概念皆相同，差別只在於綜合區間定位法(CIM)中多加入了目標區間外所有標識因子的資訊，也就是模式中的 $\mathbf{b}'\mathbf{x}_i$ ，故在不失一般性的原則下，僅詳述利用綜合區間定位法(CIM)估算 QTL 效應之理論推導。而簡單區間定位法(SIM)在計算上與綜合區間定位法(CIM)相似，只需令 $\mathbf{b} = \mathbf{0}$ 即可。

若 μ_{i0} 、 μ_{i1} 、 μ_{i2} 為分別代表第 i 個個體之 QTL 基因型為 qq、Qq、QQ 時，數量性狀期望(平均)值。令 $\boldsymbol{\beta}' = [m \ a^* \ d^* \ \mathbf{b}]$ ，且

$$\begin{bmatrix} \mathbf{x}'_{i0} \\ \mathbf{x}'_{i1} \\ \mathbf{x}'_{i2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \mathbf{x}'_i \\ 1 & 1 & 1 & \mathbf{x}'_i \\ 1 & 2 & 0 & \mathbf{x}'_i \end{bmatrix}$$

則 $\eta(\mu_{i0}) = \mathbf{x}'_{i0}\boldsymbol{\beta}$ ， $\eta(\mu_{i1}) = \mathbf{x}'_{i1}\boldsymbol{\beta}$ ， $\eta(\mu_{i2}) = \mathbf{x}'_{i2}\boldsymbol{\beta}$ 。

令 $f(y|\boldsymbol{\theta}) = f$ 為數量性狀 y 之機率密度函數， f 代表常態、二項、卜瓦松分布或其他指數分布族的機率密度函數， $\boldsymbol{\theta}$ 代表由模式中所有的參數所構成的向量，包括了數量性狀 y 分布的參數及 QTL 與兩翼標識因子間的互換率。因為 F_2 子代的 QTL 基因型有 qq、Qq 及 QQ 三種可能，以三種基因型可能

出現之機率(Liu 1998)為權重，則數量性狀 y_i (第 i 個個體的數量性狀)的機率密度函數為三個分布給定權重後，相加而得的混合模式(mixture model)，模式如下：

$$\begin{aligned} g(y_i|\boldsymbol{\theta}) &= p_{i0}(\boldsymbol{\theta})f_{i0}(y_i|\boldsymbol{\theta}) + p_{i1}(\boldsymbol{\theta})f_{i1}(y_i|\boldsymbol{\theta}) + p_{i2}(\boldsymbol{\theta})f_{i2}(y_i|\boldsymbol{\theta}) \\ &= \sum_{j=0}^2 p_{ij}(\boldsymbol{\theta})f_{ij}(y_i|\boldsymbol{\theta}) \\ i &= 1, 2, \dots, n \end{aligned} \quad (3)$$

在(3)式中， p_{i0} 、 p_{i1} 、 p_{i2} 是在給定目標區間之兩翼標識因子基因型的條件下，出現各種 QTL 基因型的條件機率，可視為三種基因型(qq、Qq 及 QQ)的事前機率(prior probability)。在實際應用上，通常可透過連續測驗目標區間內的多個位置的方式，如區間內每 2cM 的位置做一次測驗，對每一個測驗位置皆分析該位置若有 QTL 存在的可能性，再比較所有測驗位置的分析結果，則可找出目標區間內最可能有 QTL 存在的位置。由於上述方法對每一個測驗位置而言，QTL 與兩翼標識因子間的距離都是固定的(即互換率之值為已知)，故不需將互換率當成模式的參數，但真實的 QTL 卻可能存在於兩個測驗位置之間，因此若能將互換率也當成未知的參數加入模式中，則可直接估計區間內最可能有 QTL 存在的位置。不過直接估計區間內 QTL 最可能存在的位置，可能無法像連續測驗目標區間內的多個位置的方法，可以看出區間內各個位置存在 QTL 可能性的整體變化趨勢。

通常兩翼標識因子間的互換率 r 若非已知，也可由連鎖分析的方式估計得到，因此只需估計 QTL 與左翼標識因子之間的互換率 r_1 即可，至於 QTL 與右翼標識因子之間的互換率 r_2 則可利用 r 和 r_1 計算：

$$r_2 = \frac{(r - r_1)}{1 - 2r_1}$$

概似函數

所有觀測值之混合模式的概似函數

(likelihood function)如下：

$$L(\theta|y_1, y_2, \dots, y_n) = \prod_{i=1}^n g(y_i|\theta) = \prod_{i=1}^n \left[\sum_{j=0}^2 p_{ij}(\theta) f_{ij}(y_i|\theta) \right] \quad (4)$$

取對數後的對數概似函數 (log likelihood function) 為

$$\log L(\theta) = \sum_{i=1}^n \log g(y_i|\theta) = \sum_{i=1}^n \log \left[\sum_{j=0}^2 p_{ij}(\theta) f_{ij}(y_i|\theta) \right] \quad (5)$$

欲求出參數 θ 的估計值 $\hat{\theta}$ ，可使概似函數的值達到極大，則等同於求出可使對數概似函數一次微分式等於 0 的解：

$$\frac{\partial \log L(\theta)}{\partial \theta} = 0 \quad (6)$$

在(6)式中，向量 $\left[\frac{\partial \log L(\theta)}{\partial \theta} \right]$ 稱為 gradient 向量或 score function。因為混合模式沒有封閉式的解，所以必須以數值方法來求解。若可導出對數概似函數的二次微分式，則可利用牛頓法 (Newton-Raphson method) 遞迴 (iterative) 求解：

設 $\hat{\theta}^0$ 為起始值，則

$$\hat{\theta}^{k+1} = \hat{\theta}^k - \left[\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'} \right]_{\theta=\hat{\theta}^k}^{-1} \cdot \left[\frac{\partial \log L(\theta)}{\partial \theta} \right]_{\theta=\hat{\theta}^k} \quad (7)$$

將(7)反覆遞迴計算，直到符合收斂條件 $|L(\hat{\theta}^{k+1}) - L(\hat{\theta}^k)| < \varepsilon$ 為止 (ε 為給定之可容許誤差值)，而此時所得到之參數估值即為最大概似估值。

若牛頓法無法收斂或混合模式之概似函數的二次微分式相當複雜而不易導出時，則可利用其他不需要用到概似函數二次微分式的計算法，如 EM 法、ECM 法、IRLS 法等方法，計算概似函數的最大概似估值。牛頓法和 EM 法等各種不同的計算方法或許在計算方式的導出與計算效率上相差極大，但最後卻大多能收斂到相同的解。

至於最大概似估值的漸近變異矩陣則可利用(8)求出：

$$\hat{V}(\hat{\theta}) = - \left[\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'} \right]_{\theta=\hat{\theta}}^{-1} \quad (8)$$

在(8)式中，矩陣 $\left[\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'} \right]$ 稱為 Hessian 矩陣 (Hessian 矩陣的負值就是 information 矩陣) 而最大概似估值的漸近變異矩陣即為 Hessian 矩陣之逆矩陣的負值 (或 information 矩陣的逆矩陣)。利用牛頓法(7)來求解最大概似估值的優點，就是可在參數收斂的同時，又能順道求出收斂參數 (即最大概似估值) 的漸近變異矩陣(8)。但若混合模式之概似函數的二次微分式相當複雜或不易導出時，雖可利用其他不需要用到概似函數二次微分式的方法，如 EM 法，計算概似函數的最大概似估值，但最大概似估值的漸近變異矩陣卻仍需要利用概似函數的二次微分式計算。由於在 QTL 的定位分析上，可用於估算混合模式的最大概似估值的方法甚多，故本研究僅討論最大概似估值的漸近變異矩陣的求算。當數量性狀之機率分布為常態分布、二項分布、卜瓦松分布等指數分布族時，各類分布之概似函數的二次微分式請參見 Liu (1998)。

除常態分布、二項分布與卜瓦松分布，其他指數分布族 (如多項分布、珈瑪分布等) 也可利用相同的方式推導出對數概似函數的二次微分式，而能更進一步計算最大概似估值的漸近變異矩陣。以上述方法雖能推導出各類指數分布族之對數概似函數的二次微分式，但推導的過程非常繁複，所推導出之一次與二次微分式亦隨數量性狀分布的不同而改變，且若模式中的參數有所變動時，也必須再重新推導各微分式。

概似函數之有限差分近似二次微分式

利用數值方法上的有限差分近似法計算，可避免複雜的微分式推導過程，只需使用概似函數即可計算概似函數的近似二次微分式，而能更進一步的求出最大概似估值之

漸近變異矩陣。

有限差分近似法可包括向前差分(forward-difference approximation)、中央差分(central-difference approximation)、向後差分(backward-difference approximation)等不同的計算方式(Abramowitz and Stegun 1972)；上述的三種計算方式，以中央差分所求之近似解較為準確，故本研究使用有限差分近似法之中央差分的計算方式來計算概似函數的二次微分式的近似解。

若 f 為 $\mathbf{x}' = [x_1 \ x_2 \ \cdots \ x_v]$ 之 v 個參數所構成的目標函數，則利用有限差分近似法的中央差分計算方式，可計算目標函數 f 對第 i 個參數 x_i 的近似一次微分式：

$$g_i = \frac{\partial f}{\partial x_i} = \frac{f(\mathbf{x} + h_i \mathbf{u}_i) - f(\mathbf{x} - h_i \mathbf{u}_i)}{2h_i} \quad (9)$$

其中 $\mathbf{u}_i$ 為 $v \times 1$ 的單位向量(i^{th} unit vector)，除第 i 列位置之值 1 外，其餘皆為 0；

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{v \times 1}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{v \times 1}, \quad \dots, \quad \mathbf{u}_v = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}_{v \times 1}$$

$h_i = \sqrt[3]{\rho_i(1 + |x_i|)}$ ， ρ_i 為控制計算準確度的微調值，本研究將 $\sqrt[3]{\rho_i}$ 之值設為 10^{-5} 。

而目標函數 f 之近似二次微分式則可利用下式計算(Abramowitz and Stegun, 1972)：

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{-f(\mathbf{x} + 2h_i \mathbf{u}_i) + 16 \cdot f(\mathbf{x} + h_i \mathbf{u}_i) - 30 \cdot f(\mathbf{x})}{12h_i^2} + \frac{16 \cdot f(\mathbf{x} - h_i \mathbf{u}_i) - f(\mathbf{x} - 2h_i \mathbf{u}_i)}{12h_i^2} \quad (10)$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{f(\mathbf{x} + h_i \mathbf{u}_i + h_j \mathbf{u}_j) - f(\mathbf{x} + h_i \mathbf{u}_i - h_j \mathbf{u}_j)}{4h_i h_j} - \frac{f(\mathbf{x} - h_i \mathbf{u}_i + h_j \mathbf{u}_j)}{4h_i h_j} + \frac{f(\mathbf{x} - h_i \mathbf{u}_i - h_j \mathbf{u}_j)}{4h_i h_j} \quad (11)$$

其中 $\mathbf{u}_i$ 和 h_i 之定義與(9)相同。

有限差分近似法是使用目標函數 f 來計算 f 的近似二次微分式，若將(10)、(11)之目標函數 f 代入本研究之對數概似函數 $\log L(\boldsymbol{\theta})$ ，且令 $\mathbf{x}' = \boldsymbol{\theta}' = [\theta_1 \ \theta_2 \ \cdots \ \theta_v]$ 為概似函數 $L(\boldsymbol{\theta})$ 中由 v 個參數所構成的向量，則可直接使用對數概似函數 $\log L(\boldsymbol{\theta})$ 來計算對數概似函數的近似二次微分式與最大概似估值的漸近變異矩陣：若

$$\mathbf{H} = \left[\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} \equiv \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1v} \\ c_{21} & c_{22} & \cdots & c_{2v} \\ \vdots & & \ddots & \vdots \\ c_{v1} & c_{v2} & \cdots & c_{vv} \end{bmatrix}_{v \times v}$$

為最大概似估值 $\hat{\boldsymbol{\theta}}$ 的 Hessian 矩陣，則

$$c_{ii} = \left[\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_i^2} \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} = \frac{-\log L(\hat{\boldsymbol{\theta}} + 2h_i \mathbf{u}_i) + 16 \cdot \log L(\hat{\boldsymbol{\theta}} + h_i \mathbf{u}_i) - 30 \cdot \log L(\hat{\boldsymbol{\theta}})}{12h_i^2} + \frac{16 \cdot \log L(\hat{\boldsymbol{\theta}} - h_i \mathbf{u}_i) - \log L(\hat{\boldsymbol{\theta}} - 2h_i \mathbf{u}_i)}{12h_i^2} \quad (12)$$

$$c_{ij} = \left[\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} = \frac{\log L(\hat{\boldsymbol{\theta}} + h_i \mathbf{u}_i + h_j \mathbf{u}_j) - \log L(\hat{\boldsymbol{\theta}} + h_i \mathbf{u}_i - h_j \mathbf{u}_j)}{4h_i h_j} - \frac{\log L(\hat{\boldsymbol{\theta}} - h_i \mathbf{u}_i + h_j \mathbf{u}_j)}{4h_i h_j} + \frac{\log L(\hat{\boldsymbol{\theta}} - h_i \mathbf{u}_i - h_j \mathbf{u}_j)}{4h_i h_j} \quad (13)$$

因此可求出最大概似估值的漸近變異矩陣 $\hat{\mathbf{V}}(\hat{\boldsymbol{\theta}}) = -\mathbf{H}^{-1}$ 。

有限差分近似法是一種數值微分法，雖然也可使用有限差分近似法之數值微分公式來遞迴求解概似函數的最大概似估值，但這種求解方式的計算量大又費時，且當遞迴參數的起始值不夠好的時候，會發生參數無法收斂而求不出解的情況。由於在 QTL 的定位分析上，可用於估算混合模式的最大概似估值的方法甚多，如 EM 法、ECM 法、IRLS 法等，故本研究不建議直接使用有限差分近似法來求解概似函數的最大概似估值。但若

已由 EM 法等計算方法求出最大概似估值，卻因為概似函數的二次微分式相當複雜或不易以解析方法導出，而無法計算漸近變異矩陣，則可使用(12)、(13)之有限差分近似法的近似二次微分式來直接計算最大概似估值的 Hessian 矩陣。而此 Hessian 矩陣之逆矩陣的負值，就是最大概似估值的漸近變異矩陣。

由於概似函數的二次微分式與有限差分的近似二次微分式，兩式在形態上差異很大，為證實利用有限差分近似法所算之漸近變異矩陣的正確性，本文將以模擬之常態分布、二項分布與卜瓦松分布等三組數量性狀資料，分別利用概似函數的二次微分式與概似函數的有限差分法近似二次微分式來計算最大概似函數的漸近變異矩陣，並比較兩者之計算結果。

資料模擬

本研究首先利用電腦程式模擬產生三組 n 個個體的 F_2 子代資料，分別用於討論三種不同資料型態之數量性狀觀測值(表型值)的情況：第(1)組為常態分布性狀值資料，觀測值數目 $n=250$ ；第(2)組為二項分布性狀值資料，觀測值數目 $n=500$ ；第(3)組為卜瓦松分布性狀值資料，觀測值數目 $n=500$ 。由於二項分布或卜瓦松分布表型值資料是先模擬出常態分布性狀值再做適當的轉換，因轉換過程可能會損失一些訊息，故在此模擬產生較多的 F_2 子代個體資料。每個個體分別紀錄一個數量性狀 y 及 36 個遺傳標識，標識代號

A(qq)及 B(QQ)分別為同質的親本基因型，標識代號 H(Qq)為異質的雜交組合型，這 36 個遺傳標識分別屬於 6 條染色體(6 個連鎖群)，每條染色體的長度為 100cM 並各具有 6 個遺傳標識，遺傳標識間之距離為 20cM，第一條染色體之遺傳標識為 1~6，第二條染色體之遺傳標識為 7~12，其餘類推；第一至第五條染色體各含有一個 QTL，分別位於自左端算起第 70、49、27、8、30 cM 處，第六條染色體則不具有 QTL。

第(1)組之常態分布性狀值資料所設定的五個 QTL 累加性效應分別為 1.5、1.25、1.0、0.75、0.5，顯性效應分別為 0.5、0.4、0.3、0.2、0.1(如 Table 1)。第(2)組之二項分布與第(3)組之卜瓦松分布性狀值資料所設定的五個 QTL 累加性效應分別為 3、2.5、1.0、0.75、0.5，顯性效應分別為 1、0.8、0.3、0.2、0.1(如 Table 2)。由於二項分布或卜瓦松分布表型值資料是先模擬出常態分布性狀值再做適當的轉換，因轉換過程可能會損失一些訊息，故模擬資料所設定之累加性效應與顯性效應亦略大於常態分布的表型值資料。數量性狀 y 模擬為同時受到五個 QTL 的作用。

模擬每一個個體資料的過程皆可分為兩個階段來進行：第一階段為模擬產生各標識因子的基因型，此時每一個 QTL 亦可視為一個標識因子；第二階段則模擬產生各條染色體的性狀值，並將其加總後得到數量性狀的表型值。至此第(1)組資料即正式模擬完成。因為第(2)組資料欲模擬得到的是二項分布的

Table 1. The number of markers, number of QTL, QTL position, and QTL effects on each chromosome for normally distributed simulation data.

Chromosome	Marker	QTL Number	QTL Position		Additive effect (a)	Dominant effect (d)
			(cM)	Interval		
1	1~6	1	70	4	1.5	0.5
2	7~12	1	49	3	1.25	0.4
3	13~18	1	27	2	1.0	0.3
4	19~24	1	8	1	0.75	0.2
5	25~30	1	30	2	0.5	0.1
6	31~36	0	—	—	—	—

Table 2. The number of markers, number of QTL, QTL position, and QTL effects on each chromosome for binary and Poisson distributed simulation data.

Chromosome	Marker	QTL Number	QTL Position		Additive effect	Dominant effect
			(cM)	Interval	(a)	(d)
1	1~6	1	70	4	3	1
2	7~12	1	49	3	2.5	0.8
3	13~18	1	27	2	1.0	0.3
4	19~24	1	8	1	0.75	0.2
5	25~30	1	30	2	0.5	0.1
6	31~36	0	—	—	—	—

數量性狀表型值，故需再將服從常態分布的連續性狀值轉換成二項分布的資料，其轉換之步驟如下：先計算 500 個個體數量性狀平均值，再將性狀值大於或等於平均值的個體令其觀測值為 1，小於平均值的個體則令其觀測值為 0，即可將連續的性狀值轉換為二項分布的資料。因為第(3)組資料欲模擬得到的是卜瓦松分布的數量性狀表型值，故必須再將服從常態分布的連續性狀值轉換成卜瓦松分布的資料，其轉換之步驟如下：若 y_i 為第 i 個個體($i=1, 2, \dots, 500$)服從常態分布的連續型性狀值，隨機產生一個服從卜瓦松分布之數值 o_i ，即可將連續的性狀值轉換為卜瓦松分布類型的資料。

依上述過程可模擬得到的觀測值資料包括各條染色體上每一個標識因子的基因型及標識因子間的距離，與觀測到的數量性狀值的資料。

結果

一般可用於求解模式之最大概似估值之解析方法甚多，如牛頓法、EM 法等，雖然各種不同方法在計算方式之導出與計算效率上可能相差極大，但最後大多能收斂到相同的解。本研究 F_2 子代的模擬資料，將利用推導之概似函數的一次與二次微分式 (Liu 1998)，以牛頓法遞迴求解最大概似估值與最大概似估值漸近變異矩陣；另一方面，也將牛頓法求得的最大概似估值代入概似函數之有限差分近似二次微分式來計算漸近變異矩

陣。最後，再比較概似函數二次微分式之導式與有限差分近似法之近似微分式兩者的漸近變異矩陣計算結果。

本研究使用 Fortran 程式語言撰寫牛頓法程式(簡稱: Fortran NR)與有限差分近似法程式(簡稱: Fortran FD)；為驗證所撰寫之牛頓法程式的正確性，本研究亦使用統計分析軟體 SAS/OR 內之 NLP (NonLinear Programming) 程式(簡稱: SAS NLP)作計算。本研究之資料模擬使用簡單區間定位法與綜合區間定位法，利用廣義線性模式建立數量性狀基因的表型值與各標識因子的關係，選定第一條染色體上由標識4為左翼標識因子，標識5為右翼標識因子所形成的目標區間，計算此區間內之數量性狀基因座最可能的存在位置，與此位置之數量性狀基因的各項效應。第一條染色體上由標識4為左翼標識因子，標識5為右翼標識因子所形成的目標區間，即為第一條染色體上設定之數量性狀基因座的所在區間。

第一條染色體上設定之 QTL 的位置為 70 cM，與標識 4(左翼標識因子)之距離為 10 cM (0.1 M)，因此在單點分析中，

$$r_1 = \frac{1 - \exp[-2(0.1)]}{2} = 0.090634623$$

；而區間分析欲估計之標識 4 與標識 5 間，QTL 最可能的所在位置 r_1 和 r_1 估值的信賴區間，應相當接近或包含 0.090634623。

為更明確的了解利用概似函數的二次微分式與概似函數的有限差分二次近似微分

式，兩者之漸近變異矩陣計算結果的準確程度，本模擬之計算結果報表將列出到小數點以下較多位數的數值。

簡單區間定位法(SIM)之QTL效應的計算結果

(1)常態分布性狀值資料:

由區間分析的計算結果可知，標識4與標識5間最有可能存在QTL的位置為 $\hat{r}_1=0.1033724743$ ， $SE(\hat{r}_1)=0.0321186285$ ，因此95%的信賴區間為(0.0391352173, 0.1676097313)，包含了原始資料設定QTL的位置0.090634623。

(2)二項分布性狀值資料:

由區間分析的計算結果可知，標識4與標識5間最有可能存在QTL的位置為 $\hat{r}_1=0.0821865944$ ， $SE(\hat{r}_1)=0.0137376152$ ，因此95%的信賴區間為(0.0547113640, 0.1096618248)，包含了原始資料設定QTL的

位置0.090634623。

(3)卜瓦松分布性狀值資料:

由區間分析的計算結果可知，標識4與標識5間最有可能存在QTL的位置為 $\hat{r}_1=0.07910399$ ， $SE(\hat{r}_1)=0.01132901$ ，因此95%的信賴區間為(0.05644597, 0.10176201)，包含了原始資料設定QTL的位置0.090634623。

綜合區間定位法(CIM)之QTL效應的計算結果

由於本文所模擬之每類資料皆有36個標識因子，又因綜合區間定位法是以目標區間外之所有標識因子做為共變數，因此參數數目增加了34個標識因子之累加性效應與34個標識因子之顯性效應。加入目標區間外之所有標識因子做為共變數之目的在消除其他區間內數量性狀基因座的影響以提高定位之準確度，而本文主要目的在比較概似函數的

Table 3. Estimates and standard errors of QTL effects by simple interval mapping for normally distributed simulation data from two analytic derivatives (Fortran NR and SAS NLP) and one finite difference approximation (Fortran FD, standard error only).

Parameter	Fortran NR (analytic derivatives)		SAS NLP (analytic derivatives)		Fortran FD
	Estimate	Approx Std Err	Estimate	Approx Std Err	Approx Std Err
r_1	0.1033724743	0.0321186285	0.103372	0.032119	0.0321186226
m	14.3442898189	0.2472879233	14.344290	0.247288	0.2472878338
a^*	1.2673281773	0.1749617022	1.267328	0.174962	0.1749616824
d^*	0.0291557183	0.2562767363	0.029156	0.256277	0.2562765324
s^2	3.0189030628	0.2818107323	3.018903	0.281811	0.2818108971

Table 4. Estimates and standard errors of QTL effects by simple interval mapping for binary distributed simulation data from two analytic derivatives (Fortran NR and SAS NLP) and one finite difference approximation (Fortran FD, standard error only).

Parameter	Fortran NR (analytic derivatives)		SAS NLP (analytic derivatives)		Fortran FD
	Estimate	Approx Std Err	Estimate	Approx Std Err	Approx Std Err
r_1	0.0821865944	0.0137376152	0.082187	0.013738	0.0137376155
m	4.1981539641	1.2642277158	-4.198150	1.264223	1.2642035972
a^*	-2.9386203497	0.6462376693	2.938618	0.646235	0.6462265647
d^*	-1.6600799139	0.6801748182	1.660078	0.680173	0.6801614814

Table 5. Estimates and standard errors of QTL effects by simple interval mapping for Poisson distributed simulation data from two analytic derivatives (Fortran NR and SAS NLP) and one finite difference approximation (Fortran FD, standard error only).

Parameter	Fortran NR (analytic derivatives)		SAS NLP (analytic derivatives)		Fortran FD
	Estimate	Approx Std Err	Estimate	Approx Std Err	Approx Std Err
r_1	0.07910399	0.01132901	0.07910	0.01133	0.0113290048
m	1.56378800	0.05238321	1.56379	0.05238	0.0523832689
a^*	0.41785997	0.03029599	0.41786	0.03030	0.0302960074
d^*	0.23660175	0.04285275	0.23660	0.04285	0.0428528149

Table 6. Estimates and standard errors of QTL effects by composite interval mapping for normally distributed simulation data from two analytic derivatives (Fortran NR and SAS NLP) and one finite difference approximation (Fortran FD, standard error only).

Parameter	Fortran NR (analytic derivatives)		SAS NLP (analytic derivatives)		Fortran FD
	Estimate	Approx Std Err	Estimate	Approx Std Err	Approx Std Err
r_1	0.1231287488	0.0196117990	0.123129	0.019612	0.0196118092
m	8.6148404372	0.5023998330	8.614842	0.502400	0.5024000061
a^*	1.5965514460	0.1870942538	1.596551	0.187094	0.1870943872
d^*	0.1777568817	0.1950055936	0.177757	0.195006	0.1950060128
s^2	0.9455191746	0.1192609827	0.945519	0.119261	0.1192610218

解析二次微分式與有限差分近似二次微分式之差異大小，為節省篇幅，本綜合區間定位法結果並未列出共變數部分之效應估值。

(1)常態分布性狀值資料：

由區間分析的計算結果可知，標識4與標識5間最有可能存在QTL的位置為 $\hat{r}_1=0.1231287488$ ， $SE(\hat{r}_1)=0.0196117990$ ，因此95%的信賴區間為(0.0839051508, 0.1623523468)，包含了原始資料設定QTL的位置0.090634623。

(2)二項分布性狀值資料：

由區間分析的計算結果可知，標識4與標識5間最有可能存在QTL的位置為 $\hat{r}_1=0.0873130722$ ， $SE(\hat{r}_1)=0.0123980477$ ，因此95%的信賴區間為(0.0625169768, 0.1121091676)，包含了原始資料設定QTL的位置0.090634623。

(3)卜瓦松分布性狀值資料：

由區間分析的計算結果可知，標識4與標識

5間最有可能存在QTL的位置為 $\hat{r}_1=0.06941064$ ， $SE(\hat{r}_1)=0.01489874$ ，因此95%的信賴區間為(0.03961316, 0.09920812)，包含了原始資料設定QTL的位置0.090634623。本研究之區間分析的計算結果，都可準確估計標識4與標識5間最有可能存在QTL的位置。

SAS/OR內之PROC NLP程式(Sas Institute Inc. 1993)為一個可作一般性最佳化(general purpose optimization)的程式，因此可用於求解最大概似估值。在PROC NLP程式中，先定義一個需極大化或極小化的連續函數(目標函數)，再指定一些參數的起始值，通常如此即可遞迴求解出最大概似估值，且所指定之起始值並不需太精確。PROC NLP之缺點為執行相當費時，且當模式內參數過多時，非常不易計算Hessian矩陣，即不易求得最大概似估計值之漸近變異矩陣，因此PROC NLP並不適合作為QTL區間定位這種計算極為繁複的計算工具。在此以SAS軟

Table 7. Estimates and standard errors of QTL effects by composite interval mapping for binary distributed simulation data from two analytic derivatives (Fortran NR and SAS NLP) and one finite difference approximation (Fortran FD, standard error only).

Parameter	Fortran NR (analytic derivatives)		SAS NLP (analytic derivatives)		Fortran FD
	Estimate	Approx Std Err	Estimate	Approx Std Err	Approx Std Err
r_1	0.0873130722	0.0123980477	0.08731	0.01240	0.0123980608
m	21.8068593076	5.0251756022	21.80686	5.02518	5.0252245377
a^*	-7.3782093790	1.7702714542	-7.37821	1.77027	1.7702892510
d^*	-4.6310371276	1.2333122095	-4.63104	1.23331	1.2333226499

Table 8. Estimates and standard errors of QTL effects by composite interval mapping for Poisson distributed simulation data from two analytic derivatives (Fortran NR and SAS NLP) and one finite difference approximation (Fortran FD, standard error only).

Parameter	Fortran NR (analytic derivatives)		SAS NLP (analytic derivatives)		Fortran FD
	Estimate	Approx Std Err	Estimate	Approx Std Err	Approx Std Err
r_1	0.06941064	0.01489874	0.069403	0.014898	0.0148987268
m	0.98655288	0.11758918	0.986580	0.117589	0.1175889841
a^*	0.43190675	0.03980539	0.431931	0.039806	0.0398053393
d^*	0.15401118	0.04726772	0.154017	0.047268	0.0472676553

體計算的目的只是為了驗證本研究以 Fortran 程式語言所撰寫之牛頓法程式的正確性。

由上述之模擬資料的計算結果可知，利用概似函數的二次微分式與概似函數的有限差分近似二次微分式，兩種計算方法所算出之模式參數的近似標準誤 (asymptotic standard error) 都可準確到小數點以下三到六位；由於 Fortran NR 之最大概似估值的計算結果，幾乎都包含了許多位的小數，而本模擬是直接將 Fortran NR 報表之最大概似估值的計算結果代入 Fortran FD 計算，如此可能造成 Fortran FD 近似標準誤之計算結果的誤差，若能改進 Fortran FD 之最大概似估值的資料讀取方式，則利用概似函數之有限差分近似二次微分式的近似標準誤計算結果應能再準確到小數點以下更多位數。本研究為更明確的了解利用概似函數的二次微分式與概似函數的有限差分近似二次微分式，兩者

之漸近變異矩陣計算結果的準確程度，因此列出小數點以下較多位數的近似標準誤計算結果，對於實際數量性狀資料之分析，則毋須計算至小數點以下如此多位的數值，故兩種方法之漸近變異矩陣計算結果，幾乎可說是完全相同。

利用有限差分近似法的近似二次微分式計算最大概似估值的漸近變異矩陣，不但方法簡單，在使用上也不需先做概似函數微分式的繁複推導，且對本研究模擬之各類數量性狀資料(常態分布、二項分布與卜瓦松分布資料)，或由不同的 QTL 定位法(SIM 與 CIM)和連結函數(Identity、Logit、Log)所組成的統計模式與概似函數，利用概似函數之有限差分近似二次微分式與利用概似函數二次微分式的漸近變異矩陣計算結果幾乎完全相同。因此可證實本研究提出利用有限差分近似法來計算最大概似估值之漸近變異矩陣的實用性與正確性。

討論

在數量性狀基因座的定位分析上，可用於估算混合模式的最大概似估值的方法甚多，如 EM 法、ECM 法、IRLS 法等，但最大概似估值的漸近變異矩陣因為需要利用概似函數的二次微分式求算，又由於一般混合模式之概似函數的二次微分式相當複雜不易導出，因此本研究提出以數值方法上的有限差分近似法來計算概似函數的近似二次微分式與最大概似估值的漸近變異矩陣。

本研究主要使用簡單區間定位法與綜合區間定位法建構 QTL 定位分析的混合模式，對於常態分布、二項分布與卜瓦松分布等各類型的數量性狀資料，簡單區間定位法對於染色體上僅具有 1 個 QTL 資料的檢定力較高，但若染色體上有兩個以上的 QTL 時，則以綜合區間定位法的結果較佳 (Liu and Chiang 1998, Liu and Chen 1999, Liu and Chang 2000)，由於實際資料並無法事先得知各染色體上 QTL 的個數，所以都會同時使用兩種模式來判斷及比較。無論是使用簡單區間定位法或綜合區間定位法的模式來作分析，主要目的仍是判斷染色體上是否有 QTL 存在，與染色體上 QTL 最有可能的存在位置，但若能在計算混合模式之最大概似估值的同時，又能提供估值的變異，則可再更進一步的評估根據此估值所作之統計推論的可靠性。

由於二項分布或卜瓦松分布表型值資料是先模擬出常態分布性狀值再做適當的轉換，因轉換過程可能會損失一些訊息，故本研究模擬產生較多二項分布與卜瓦松分布的個體資料，而資料模擬所設定之數量性狀基因的累加性效應與顯性效應亦略大於常態分布的表型值資料。本研究之資料模擬僅計算第一條染色體上標識 4 與標識 5 之間最可能有數量性狀基因座的存在位置與此位置之數量性狀基因的各項效應，而計算結果也都可準確估計標識 4 與標識 5 間最有可能存在 QTL 的

位置。透過相同的方法可掃描 F_2 子代的所有染色體，計算染色體上的每一個目標區間內最可能有數量性狀基因座的存在位置和此位置之數量性狀基因的各項效應。

基本上，當數量性狀資料之機率分布為指數分布族時，皆可利用本文所整理的方法推導對數概似函數二次微分式，但微分式推導的過程既複雜，且若原本的 QTL 定位的模式中加入了其他新的訊息，如時間、地點或年份做為共變數時，又必須再重新推導各微分式。利用本研究之有限差分近似法計算概似函數的近似二次微分式，不但方法簡單，且當數量性狀資料之分布改變或 QTL 的定位模式改變時，仍可利用相同的公式計算各模式之對數概似函數的近似二次微分式。而由本研究之模擬結果可知，對常態分布、二項分布及卜瓦松分布三種類型之 F_2 子代的數量性狀資料，利用概似函數的二次微分式與利用概似函數的有限差分近似二次微分式，兩者所算出之最大概似估值的漸近變異矩陣幾乎都能完全相同。

在一般最佳化問題的求解上，只要目標函數可微分(本研究之目標函數為對數概似函數 $\log L(\theta)$)，即可利用有限差分近似法來遞迴求解目標函數之各參數估值，但因有限差分近似法在遞迴求解的過程上收斂相當緩慢，有時也可能會發生無法收斂的情況，除非沒有其他更好的計算方法，通常不建議使用有限差分近似法來遞迴求解。一般可用於求解 QTL 定位分析之概似函數最大概似估值的方法甚多，故不建議利用有限差分近似法來遞迴求解概似函數的最大概似估值。本研究僅使用有限差分近似法計算目標函數(對數概似函數 $\log L(\theta)$)之近似二次微分式，配合以其他計算方法(如 EM 法)所求出之最大概似估值，計算最大概似估值的漸近變異矩陣，因為是以最大概似估值為起使值代入計算，因此無需遞迴求解，僅需一次計算迴路。此外，除了本研究所模擬之常態分布、二項分布及卜瓦松分布的數量性狀資料之外，對

其他分布的數量性狀資料，理論上只要混合模式的對數概似函數 $\log L(\theta)$ 可微分，都可利用相同之計算概似函數近似二次微分式的方法來求算最大概似估值的漸近變異矩陣。在數值分析中，有限差分近似法可計算目標函數之近似二次微分式的方法很多，各種不同計算近似二次微分式的方法在計算時間與計算結果的準確度上有所差異。本研究使用中央差分法計算，而模擬資料的漸近變異矩陣計算結果也證實本研究之有限差分近似法的正確性。

本研究僅討論最大概似估值之漸近變異矩陣的計算，由於本研究在資料模擬時是利用牛頓法所求出的最大概似估值，代入有限差分近似法計算估值的漸近變異矩陣；因為牛頓法在求出最大概似估值的同時，也順道求出了最大概似估值的漸近變異矩陣，但有限差分近似法只計算最大概似估值的漸近變異矩陣，故本研究在模擬結果中並沒有特別比較兩種方法的計算時間。由於牛頓法之遞迴求解的過程未必能收斂且在計算上必須使用概似函數的二次微分式，若牛頓法無法收斂或混合模式之概似函數的二次微分式相當複雜而不易導出時，雖可利用其他的計算方法，如 EM 法，計算概似函數的最大概似估值，但最大概似估值的漸近變異矩陣卻仍需要利用概似函數的二次微分式計算，因此本研究提出利用有限差分近似法來計算概似函數的近似二次微分式與最大概似估值的漸近變異矩陣。牛頓法和 EM 法等不同的計算方法或許在計算方式的導出與計算效率上相差極大，但最後卻大多能收斂到相同的解。在實際的應用上，本研究之有限差分近似法的使用，必須配合其他的計算方法(如 EM 法、ECM 法、IRLS 法等)所求出之最大概似估值，才可計算最大概似估值的漸近變異矩陣，故若以求解概似函數之最大概似估值及其漸近變異矩陣的整體計算時間來看，牛頓法所需的計算時間應仍小於其他的計算方法(如 EM 法)搭配有限差分近似法所需的計算

時間。

近年來數量性狀基因座定位法的發展，有些定位法的統計模式內加入了環境效應之類的隨機因子(random effect)，或考慮到模式內因子間的交通作用等，複雜的混雜混合模式(mixed mixture model)在求解上更加困難，而最大概似估值的變異矩陣在計算上也更為不易，若能利用本研究之有限差分近似法計算，應可有效解決求解模式上所需的複雜計算，關於這方面情況可能需要再進一步的研究與探討。在數量性狀基因座的定位與分析上，當各類數學模式之概似函數無已知的二次微分式或不易以解析方法導出時，本研究建議可先以 EM 法、ECM 法、IRLS 等方法來計算模式之最大概似估值。一但求得最大概似估值的解析解，則可利用本研究之有限差分近似法來計算概似函數的近似二次微分式及最大概似估值的漸近變異矩陣。

引用文獻

- Abramowitz M, IA Stegun (1972) Handbook of Mathematical Functions. Dover Publications, Inc., New York. 884pp.
- Dempster AP, NM Laird, DB Rubin (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. Ser B*, 39:1-38.
- Jansen RC (1992) A general mixture model for mapping quantitative trait loci by using molecular markers. *Theor. Appl. Gener.* 85:252-260.
- Jiang CJ, ZB Zeng (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* 140:1111-1127.
- Kao CH (1995) Statistical methods for locating positions and analyzing epistasis of multiple quantitative trait genes using molecular marker information. PhD thesis, North Carolina State University, NC, USA.
- Lander ES, DB Botstein (1989) Mapping Mendelian factors under-lying quantitative traits using RFLP linkage maps. *Genetics* 121:185-199.

- Liu C (1997) Calculation of maximum likelihood estimates and asymptotic dispersion matrix for mapping and analysis of quantitative trait loci. **Chinese Agron. J.** 7: 213-243.
- Liu C (1998) Maximum likelihood estimation of a general mixture model by Newton-Raphson method for mapping and analysis of quantitative trait loci. **Chinese Agron. J.** 8: 69-82.
- Liu C, HK Chang (2000) Mapping and analysis of quantitative trait loci (QTL) for ordinal data. **Chinese Agron. J.** 10: 279-295.
- Liu C, PY Chen (1999) Mapping and Analysis of quantitative trait loci (QTL) for Poisson distributed data. **Chinese Agron. J.** 9: 127-141.
- Liu C, CM Chiang (1998) Mapping and analysis of quantitative trait loci (QTL) for binary distributed data. **Chinese Agron J.** 8: 213-224.
- Meng XL, DB Rubin (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. **Biometrika** 80:267-268.
- Sas Institute Inc. (1993) Nonlinear programming extended user's guide. SAS Institute Inc., Cary, NC. 155pp.
- Soller M, T Brody (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. **Theor. Appl. Genet.** 47:35-39.
- Zeng ZB (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. **Proc. Natl. Acad. Sci. USA** 90:10972-10976.
- Zeng ZB (1994) Precision mapping of quantitative trait loci. **Genetics** 136:1457-1468.