

## 家畜育種の理論とその応用(81)

誌名	畜産の研究 = Animal-husbandry
ISSN	00093874
著者名	林,武司
発行元	養賢堂
巻/号	60巻6号
掲載ページ	p. 716-720
発行年月	2006年6月

農林水産省 農林水産技術会議事務局筑波産学連携支援センター  
Tsukuba Business-Academia Cooperation Support Center, Agriculture, Forestry and Fisheries Research Council  
Secretariat



## 家畜育種の理論とその応用 (81)

— QTLの探索 —

林 武 司\*

### 3-2-4. 最小2乗法を用いたInterval mapping法

これまで、近交系 F2 家系を対象とした最尤法にもとづく Interval mapping 法について見てきた。最尤法による Interval mapping 法では、調べたいゲノム領域上の各部位において QTL 検出のための統計量を算出し QTL の効果等の推定を行うが、そのためには複雑な反復計算が必要となり、解析に多大な時間を要する。実際の QTL 解析では、多数の形質において、全ゲノム領域にわたる多数の部位について QTL の探索を行うことが普通であり、さらに多数回 (1000 回以上) の Permutation test によってしきい値を算出する必要もある。このような規模の大きな解析に対しては、複雑な計算を要する最尤法を用いた Interval mapping 法は効率的ではない。そこで、Haley と Knott (1992) は、最尤法をより計算の簡略な最小 2 乗法で代用した Interval mapping 法を提案した。最小 2 乗法を用いた Interval mapping 法では、反復計算を必要とせず、解析の時間が大幅に短縮される。最小 2 乗法による解析は、多くの場合、最尤法による解析と同様の結果を示し、計算の簡略さからブタやニワトリなどの家畜における他殖性 F2 家系の QTL 解析のための標準的手法として用いられている。この節では、F2 家系における最小 2 乗法を用いた Interval mapping 法について、解説していくことにする。F2 家系としては近交系 F2 家系を考えることにして、家畜における他殖性 F2 家系については、次節で述べることにする。

近交系 F2 家系については、これまでと同様に、2 つの近交系統 P1 と P2 の交雑に由来する F2 家系を意味するものとし、QTL および解析に用いるマーカーについて P1 と P2 では異なるアリルが固定しているものと仮定する。いま、2 つの隣接したマーカー A と B の間のある部位で QTL の探索

を試みる場合を考える。調べたい部位に仮想的な QTL が存在するものとして、その QTL を Q で表す。これまでの表記法に従い、マーカーと QTL (Q) において P1 および P2 に由来するアリルについては、それぞれ添え字 1 と 2 を付加して表すことにする。したがって、A、Q、B において、P1 と P2 の遺伝子型は、それぞれ  $A_1A_1Q_1Q_1B_1B_1$  および  $A_2A_2Q_2Q_2B_2B_2$  と表される。ハプロタイプを考慮した表記では、これらの遺伝子型が  $A_1Q_1B_1/A_1Q_1B_1$  および  $A_2Q_2B_2/A_2Q_2B_2$  と表される。最小 2 乗法を用いた Interval mapping 法の計算過程の説明のために、まず F2 家系における Interval mapping 法のモデル (3.1) を再び確認しておこう。

$$y_i = \mu + z_{1i}a + z_{2i}d + z_{3i}(-a) + e_i \quad (3.1)$$

ここで、 $y_i$  は第  $i$  番目の F2 個体の表現型値、 $a$  と  $d$  は調べたい部位に想定した QTL である Q の相加効果および優性効果、さらに、 $z_{1i}$ 、 $z_{2i}$ 、 $z_{3i}$  は、遺伝子型  $Q_1Q_1$ 、 $Q_1Q_2$ 、 $Q_2Q_2$  それぞれに対して  $(z_{1i}, z_{2i}, z_{3i}) = (1, 0, 0)$ 、 $(0, 1, 0)$ 、 $(0, 0, 1)$  という値をとる QTL の遺伝子型を表す変数である。また、F2 個体数は  $n$  であるとする。Q の位置が A あるいは B の位置と一致しないときには、 $z_{1i}$ 、 $z_{2i}$ 、 $z_{3i}$  の値は未知であるが、最尤法を用いた Interval mapping 法では、マーカー A と B の遺伝子型の情報を利用してそれらの値を反復計算により推測する作業を行う。そのために多大な計算量、計算時間を必要とすることになる。そこで、 $z_{ki}$  ( $k=1, 2, 3$ ) の値をマーカー A と B の遺伝子型が観測されたもとの条件付き期待値で置き換えることを考えてみよう。この条件付き期待値を  $E(z_{ki}|A, B)$  と表すことにする。この期待値については、マーカー A と B の遺伝子型が観測されたもとの、Q の遺伝子型がそれぞれ  $Q_1Q_1$ 、 $Q_1Q_2$ 、 $Q_2Q_2$  となる条件付き確率  $p_{1i}$ 、 $p_{2i}$ 、 $p_{3i}$  により

\* 独立行政法人・農業生物資源研究所 ゲノム研究グループ・家畜ゲノム研究チーム (Takeshi Hayashi)

$$E(z_{1i}|A,B) = 1 \times p_{1i} + 0 \times p_{2i} + 0 \times p_{3i} = p_{1i}$$

$$E(z_{2i}|A,B) = 0 \times p_{1i} + 1 \times p_{2i} + 0 \times p_{3i} = p_{2i}$$

$$E(z_{3i}|A,B) = 0 \times p_{1i} + 0 \times p_{2i} + 1 \times p_{3i} = p_{3i}$$

と計算することができる。QTLの遺伝子型の(条件付き)確率 $p_{1i}, p_{2i}, p_{3i}$ については、すでに3-2-2-2節において詳しく説明してある。これらの確率のAとQの間の組換え価, $r_{AQ}$ とQとBの間の組換え価, $r_{QB}$ を用いた表現式については、表3.2を参照してほしい。

モデル(3.1)は, $z_{ki}(k=1,2,3)$ を上で得られた $E(z_{ki}|A,B)$ で置き換えると

$$y_i = \mu + (p_{1i} - p_{3i})a + p_{2i}d + \varepsilon_i \quad (3.26)$$

と書き表すことができる。(3.26)における $\varepsilon_i$ は残差項を表すが、変数 $z_{ki}(k=1,2,3)$ の $E(z_{ki}|A,B)$ からの変動も含むために、 $\varepsilon_i$ は(3.1)における残差項 $e_i$ とは異なっており、それを考慮して $e_i$ とは異なる記号を用いている。これについて、多少詳しく見ていこう。 $z_{ki}(k=1,2,3)$ を $E(z_{ki}|A,B) (= p_{ki})$ とそれ以外の部分 $c_{ki}$ に分けると

$$z_{ki} = E(z_{ki}|A,B) + c_{ki} = p_{ki} + c_{ki} \quad (k=1,2,3) \quad (3.27)$$

と表される。 $c_{ki}$ は, $z_{ki}$ の $E(z_{ki}|A,B)$ からのずれを表す項で平均は0であり、マーカーA,Bの遺伝子型が既知のもとでの条件付き分散 $V(c_{ki}|A,B)$ は、ベルヌーイ分布の分散を考えて、 $V(c_{ki}|A,B) = p_{ki}(1-p_{ki})$ となる。(3.27)を(3.1)に代入すると、

$$y_i = \mu + (p_{1i} - p_{3i})a + p_{2i}d + ac_{1i} + dc_{2i} - ac_{3i} + e_i \quad (3.28)$$

と表すことができ、(3.26)と(3.28)を比較すると(3.26)の残差項 $\varepsilon_i$ は、

$$\varepsilon_i = ac_{1i} + dc_{2i} - ac_{3i} + e_i \quad (3.29)$$

と書くことができる。したがって、(3.26)の残差

項 $\varepsilon_i$ は(3.1)の残差項 $e_i$ と $z_{ki}$ の変動に関係したそれ以外の項を含むことになる。(3.29)より $\varepsilon_i$ のマーカーA,Bの遺伝子型が既知のもとでの条件付き分散 $V(\varepsilon_i|A,B)$ は、

$$V(\varepsilon_i|A,B) = p_{2i}(1-p_{2i})(a+d)^2 + \sigma_e^2$$

となる。ここで、 $\sigma_e^2$ は $e_i$ の分散である。 $V(\varepsilon_i|A,B)$ は $i$ に依存して異なる値となり、(3.26)は残差分散が個体によって異なる不均一なモデルとなっている。

最小2乗法では、(3.26)の残差項についての平方和(残差平方和)が最も小さくなるようにパラメータ $\mu, a, d$ の値を定め、そのときの残差平方和が小さいほどモデルの適合度が高いとみなす。モデルの適合度が有意に高い場合に、調べている部位にQTLが存在するものと判断することになる。具体的には、(3.26)より残差平方和 $S_1^2$ を

$$S_1^2 = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \{y_i - \mu - (p_{1i} - p_{3i})a - p_{2i}d\}^2$$

として求め、 $S_1^2$ を最小とするようなパラメータ $\mu, a, d$ の値を見いだす。このためには、 $S_1^2$ をこれらのパラメータで偏微分して0とおいて得られる、以下のような $\mu, a, d$ についての連立方程式を解けばよい。

$$\frac{\partial S_1^2}{\partial \mu} = -2 \sum_{i=1}^n \{y_i - \mu - (p_{1i} - p_{3i})a - p_{2i}d\} = 0$$

$$\frac{\partial S_1^2}{\partial a} = -2 \sum_{i=1}^n (p_{1i} - p_{3i}) \{y_i - \mu - (p_{1i} - p_{3i})a - p_{2i}d\} = 0$$

$$\frac{\partial S_1^2}{\partial d} = -2 \sum_{i=1}^n p_{2i} \{y_i - \mu - (p_{1i} - p_{3i})a - p_{2i}d\} = 0$$

これらの連立方程式を解くことは容易であり、得られた解を $\hat{\mu}, \hat{a}, \hat{d}$ とおく。そのとき $S_1^2$ の値を $\min S_1^2$ と表すと

$$\min S_1^2 = \sum_{i=1}^n \{y_i - \hat{\mu} - (p_{1i} - p_{3i})\hat{a} - p_{2i}\hat{d}\}^2$$

となる。さらにQTLが存在しないもとでのモデルを考えると、(3.26)で $a=d=0$ として、

$$y_i = \mu_0 + \varepsilon_{0i} \quad (3.30)$$

となるが、このモデルについて残差 $\varepsilon_{0i}$ の平方和を $S_0^2$ とする。 $S_0^2 = \sum_{i=1}^n (y_i - \mu_0)^2$ より $S_0^2$ は $\mu_0^2 = \sum_{i=1}^n y_i/n$ のときに最小となり、その最小値を $\min S_0^2$ とおく。QTLの検出については、最尤法を用いたInterval mapping法の場合と同様に、QTLが存在しないとする仮説 $H_0\{a=d=0\}$ とQTLが存在するものとする対立仮説 $H_1\{a \neq 0 \text{ または } d \neq 0\}$ の仮説検定を行うことになるが、最小2乗法を用いたInterval mapping法においては、その検定のための統計量は、 $H_0$ と $H_1$ にそれぞれ対応するモデル(3.30)と(3.26)における残差平方和の最小値 $\min S_0^2$ と $\min S_1^2$ から構成することになる。

最小2乗法によるInterval mapping法を提案したHaleyとKnott(1992)は、QTLの検出のために、

$$LS = n \log \left[ \frac{\min S_0^2}{\min S_1^2} \right] \quad (3.31)$$

という形の統計量を考えた。この統計量 $LS$ は、多くの場合、最尤法によるInterval mapping法の統計量 $LRT$ と同様の値を示す。このことは、3-2-1節の(3.11)式で与えた $LRT$ の表現

$$LRT = n \log \left( \frac{\hat{\sigma}_0^2}{\hat{\sigma}_e^2} \right) \quad (3.11)$$

と(3.31)を比べることにより容易に予想できるであろう。

HaleyとKnott(1992)により提案された最小2乗法によるInterval mapping法は、その後Haleyら(1994)によりブタなどの家畜におけるF2家系や戻し交配家系のQTL解析に応用された。ここでは、QTLの検出のための統計量として、分散分析で用いられる $H_0$ と $H_1$ の残差分散比にもとづく $F$ 値が $\min S_0^2$ と $\min S_1^2$ から、次のように構成された。

$$F = \frac{(\min S_0^2 - \min S_1^2) / (df_0 - df_1)}{\min S_1^2 / df_1} \quad (3.32)$$

ここで、 $df_0$ および $df_1$ は、それぞれ $H_0$ と $H_1$ の残差

分散の自由度であり、モデル(3.30)と(3.26)に対応して、 $n-1$ および $n-3$ となる。家畜における最小2乗法を用いたQTL解析では、 $F$ 値が統計量として用いられることが多い。 $F$ 値については、ある1つの部位に注目した場合、 $H_0$ のもとのその分布は第1自由度が $df_0 - df_1$ 、第2自由度が $df_1$ のF分布に従う。しかしながら、あるゲノム領域にわたってQTLの探索を試みた場合には、QTL検出の基準となるしきい値の算出のためには、その領域にわたって得られた $F$ 値の最大値の分布を考える必要がある。 $F$ 値のしきい値については、最尤法の場合と同様に、あらかじめ設定した有意水準に対応したしきい値をPermutation testにより定めることができる。

$LS$ と $F$ 値の関係であるが、十分に0に近い $x$ について $\log(1+x)$ と $x$ が同様の値となることを利用すると

$$LS = n \log \left[ \frac{\min S_0^2}{\min S_1^2} \right] = n \log \left[ 1 + \frac{\min S_0^2 - \min S_1^2}{\min S_1^2} \right] = n \frac{\min S_0^2 - \min S_1^2}{\min S_1^2}$$

となり、(3.32)から

$$LS = n \frac{df_0 - df_1}{df_1} F = \frac{2n}{n-3} F$$

となる。 $n$ が十分に大きければ、 $LS = 2F$ となる。F2個体数が100頭以上あれば、この関係が近似的に成り立つと考えて良い。上で述べたように、 $LS$ は最尤法による $LRT$ と同様の値を示すので、 $LRT = 2F$ 、すなわちF2家系の解析では $F$ 値は $LRT$ のおよそ半分くらいの値になると思えば良い。

最小2乗法によるInterval mapping法でのQTLの探索は、ゲノム上の各部位で統計量 $LS$ あるいは $F$ 値を計算し、最尤法の場合と同様に統計量のピークが有意な値を示したときにQTLが検出されたものと判断し、ピークの位置をQTLの位置とみなす。また、そのときの、 $\hat{a}$ 、 $\hat{d}$ が検出されたQTLの相加効果および優性効果の推定値(最小2乗推定値)となる。最小2乗法を用いたInterval mapping法は、最尤法のような複雑な反復計算を必要とせず、多数の形質についてのQTLを全ゲノ

ム上で探索する場合でも、短時間で容易に結果を出すことができる。さらに、計算時間が短縮されるために、パラメータを増やしたモデルについても対応が容易であり、モデル (3.26) に性の効果やその他の非遺伝的な効果を含めた解析が可能となる。次節では、家畜の他殖性 F2 家系に対する Interval mapping 法について述べるが、家畜ではマウスなどの実験動物と異なり、各個体が受ける環境効果をほぼ均一にするような十分な管理のもとでの飼育は困難である。したがって、家畜における肉量や肉質などの形質は、遺伝的要因以外に飼育環境などのさまざまな効果の影響を受ける。したがって、家畜の他殖性 F2 家系においては、非遺伝的な効果をいくつか含めたモデルを用いることが普通であり、最小 2 乗法にもとづく Interval mapping 法が標準的な手法となっている。

非遺伝的な効果をいくつか含めることによりパラメータの個数が増加し、残差平方和  $S_1^2$  等の表現式が煩雑になってしまう傾向がある。そこで、次節の準備として、いくつかの非遺伝的な効果を含むモデルを用いた最小 2 乗法による Interval mapping 法で得られる  $\min S_1^2$  あるいは  $LS$  や  $F$  値などの統計量に対する行列とベクトルを用いた表記法を述べておこう。

F2 世代における  $n$  個体の表現型値のデータを  $n$  次元ベクトル  $y=(y_1, y_2, \dots, y_n)'$  で表し、さらに  $n \times 3$  ( $n$  行 3 列) の行列  $X$  と 3 次元のベクトル  $b$ 、 $n$  次元のベクトル  $\epsilon$  を次のように定義する。

$$X = \begin{bmatrix} 1 & p_{11} - p_{31} & p_{21} \\ 1 & p_{12} - p_{32} & p_{22} \\ \vdots & \vdots & \vdots \\ 1 & p_{1n} - p_{3n} & p_{2n} \end{bmatrix}, b = \begin{bmatrix} \mu \\ a \\ d \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

このとき、モデル (3.26) は

$$y = Xb + \epsilon \tag{3.33}$$

と簡略に表すことができ、残差平方和  $S_1^2$  は

$$S_1^2 = (y - Xb)'(y - Xb) \tag{3.34}$$

となる。ここで、もう一度確認しておく、 $[\ ]'$  は

ベクトルや行列の転置 (行と列を入れ替える操作) を表わす記号である。 $S_1^2$  を最小にするパラメータのベクトル  $b$  を求めるためには、 $S_1^2$  を  $b$  で微分して  $0=(0,0,0)'$  とおいた方程式を解けば良く、(3.34) より

$$\frac{\partial S_1^2}{\partial b} = -2X'(y - Xb) = 0$$

となり、これを解くことにより  $S_1^2$  を最小にする  $b$  は

$$b = (X'X)^{-1}X'y$$

と表すことができる。これより、 $S_1^2$  の最小値  $\min S_1^2$  は、

$$\min S_1^2 = y'(I - X(X'X)^{-1}X')y$$

となることが示される。ここで一方、 $I$  は  $n \times n$  の単位行列である。(3.33) と同様の表記によりモデル (3.30) は、

$$y = X_0 b_0 + \epsilon_0$$

と表すことができる。ただし、

$$X_0 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, b_0 = \mu_0, \epsilon_0 = \begin{bmatrix} \epsilon_{01} \\ \epsilon_{02} \\ \vdots \\ \epsilon_{0n} \end{bmatrix}$$

である。QTL が存在しないという仮説  $H_0$  に対応するモデル (3.30) については、パラメータは  $\mu_0$  だけであり、これをベクトル表記するメリットはないが、性や季節等の非遺伝的な効果を考慮する場合には、 $H_0$  のモデルでも複数のパラメータを含むことになり、行列やベクトルを用いた表記が有用である。モデル (3.30) の残差平方和  $S_0^2$  の最小値  $\min S_0^2$  は、

$$\min S_0^2 = y'(I - X_0(X_0'X_0)^{-1}X_0')y$$

となる。したがって、QTL 検出のための統計量  $LS$  と  $F$  値については、それぞれ、

$$LS = n \log \left[ \frac{y' \{ I - X_0 (X_0' X_0)^{-1} X_0' \} y}{y' \{ I - X (X' X)^{-1} X' \} y} \right]$$

および

$$F = \frac{y' \{ X (X' X)^{-1} X' - X_0 (X_0' X_0)^{-1} X_0' \} y / (df_0 - df_1)}{y' \{ I - X (X' X)^{-1} X' \} y / df_1}$$

と書くことができる。

【お知らせ】

## アグリネイチャー初級スチュワード ビジネススクール

— 2006年度開催案内 —

アグリネイチャー活動は単に化学農薬や化学肥料や土壌改良剤を使わないというだけの概念ではありません。むしろ人間が積極的に自然との共生をはかり、農業の多面的機能に価値を見出そうとしています。自然と人間は対立概念ではなく、互いに依存しあう共生概念なのです。

- 生き物の多様性を損なわない農業を自覚的に行なう農業者(農業自然の番人)の育成
- そうして生まれた農産物に、「高い自然価値」と「農業の環境サービス機能」を見出すことができる社会の実現

この2つのテーマを追求してゆくのがアグリネイチャー活動です。

アグリネイチャースチュワードとは、「農業自然の番人」としての役割を実践し、社会から正当な報酬を得ることをいいます。アグリネイチャースチュワード協会は、そのための人材育成を目的とした団体です。

アグリネイチャースチュワード養成講座では、初級チュワードコース(基礎コース:アグリネイチャーの価値観と技術を学ぶ)とスチュワードコース(ビジネスコース:アグリネイチャーのビジネス開発について学ぶ)の2つのコースがございますが、初級チュワードを取得された方のみスチュワードコースを受講することができます。今回は、この初級チュワードコースのご案内です。

\*スチュワードとは「委ねられたものを無私の心で管理する人」を指します。

### 2泊3日合宿集中コース募集要項

日時:2006年7月21日(金) - 23日(日)

場所:アグリネイチャーいじま

(長野県上伊那郡・飯島町)

定員10~30名(応募者が満たない場合は延期)

経費:合計63,500円

授業料 5万円 + 5%消費税 = 52,500円

宿泊料 一泊5,500円 × 2泊 = 11,000円

なお交通費は現地まで各人負担です

日程

7月21日(金) 夜 現地集合

注:同日午後1時~5時 特別講演・実習

「イノシシの被害対策」(麻布大専任講師

江口祐輔)にも参加出来ます。

7月22日(土)午前8時~夜10時

1. 午前8 - 10時【講義・実習】水田の虫見板利用技術(宇根 豊)
2. 午前10 - 12時【実習】水田・里山の生き物調査実習(伊藤一幸+米山富和・FM 研究所)
3. 午後1 - 3時【講義】アグリネイチャー『農業自然』の基本構想と経営戦略(松木洋一)
4. 午後3 - 5時【講義】自然共生栽培の生産技術体系(梅津哲一)
5. 午後5 - 7時【講義】農場に生きる野生生物の知識とマネジメント①  
「農場雑草管理技術・里山植物の多様性保全技術」(伊藤一幸+中野千賀・FM 研究所)
6. 午後8 - 10時【講義】農場に生きる野生生物の知識とマネジメント②  
「農場昆虫管理技術・ビオトープ管理技術」(新里達也)

7月23日(日)午前8時~午後4時

7. 午前8 - 12時【実習】ビオトープ建設実習(新里+小林元邦+上原俊憲・FM)
8. 午後1 - 2時【ワークショップ】消費者・食品企業家とのビジネス開発(矢崎栄司+梅津+松木+中野・FM)
9. 午後2時 - 4時 資格試験
10. 4時 現地解散

詳しいことのお問い合わせは以下まで  
アグリネイチャースチュワード協会

<http://www.agrature.jp/index.html>

東京校

〒150-0043 東京都渋谷区道玄坂1-22-9

國分渋谷ビル10F

Tel) 03-5728-8718 Fax) 03-5459-8186

tokyo@agrature.jp

飯島校

〒399-3702 長野県上伊那郡飯島町3907-1636 番地

Tel/Fax) 0265-86-6072 ijima@agrature.jp

または

日本獣医生命科学大学

応用生命科学部 食料自然共生経済学教室

松木洋一 教授

〒180-8602 東京都武蔵野市境南町1-7-1

Tel/Fax) 0422-32-4346