# QTL解析の統計モデルと検定の多重性調整

栗木 哲*

**概 要**

個体の形質を規定する遺伝子 (量的形質遺伝子座，QTL) を探索するための統計的手法を QTL 解析という．QTL 解析においては，ロッドスコアとよばれる尤度関数の極大点を探索することによって，QTL の位置が推測される．本稿では，実験交配生物に対する QTL 解析について，その統計モデルと，QTL 探索においてしばしば問題となる検定の多重性調整について解説する．1 章では，QTL 解析が，多重比較・多重検定の一種であることを説明し，問題の枠組みを与える．2 章では QTL 解析のための代表的な統計モデルを，背景となる遺伝知識と併せて可能な限り簡潔に説明する．また各モデルに対して定まるロッドスコアの確率構造を導く．3 章では 2 つの数学的手法 (非線形再生理論およびオイラー標数法) によって，QTL の有無判定のためのロッドスコアの閾値が合理的に設定されることを見る．

## 1 はじめに

### 1.1 QTL 解析と変化点問題

個体のある形質 (形態や性質) が遺伝的な効果によってもたらされると考えられる場合，その原因となる遺伝子を探し出すことは，遺伝学研究の重要な目的の一つとなる．その形質が主として連続量で記述され，また一般には複数の遺伝子と環境要因によって規定されるものである場合，量的形質とよばれる．たとえばマウスの脂肪体重比 (肥満度) は，典型的な量的形質である．量的形質の原因となる遺伝子が，QTL (量的形質遺伝子座，quantitative trait loci) である．

QTL を探索するための統計手法を QTL 解析という．QTL 解析は連鎖とよばれる遺伝現象を積極的に利用した，代表的な連鎖解析である．QTL 解析においては，連鎖と形質発現の双方を確率的な現象と捉えて統計モデルを設定することによって，目的遺伝子の探索が行われる (鵜飼 (2000) , Wu, *et al.* (2007))．

最初にマウスの肥満の原因となる遺伝子を探索するためのデータ解析の例を示す．ここでは肥満度の代用特性である血中アディポネクチン濃度 (単位 $\log_{10}[\mathrm{ng/m}\ell]$) に着目し，それを量的形質としている．また解析対象のマウスは，標準的マウス近交系である B6 と，日本産亜種由来の MSM 系統の雑種 206 個体である．この 2 系統は形質が多くの点で対照的であるため，QTL 解析に適したものである．解析結果は，図 1 のロッドスコア (LOD score) に要約されている．図の横軸はマウスの 20 対の染色体における遺伝子座の位置であり，その点に位置する遺伝子の遺伝子型と，血中アディポネクチン濃度とのある種の連

*統計数理研究所，総合研究大学院大学，〒 106-8569 東京都港区南麻布 4–6–7，Email: kuriki@ism.ac.jp

関の尺度 (ロッド) がプロットされている．この図によると，第 3 および第 16 染色体上に QTL が存在することが示唆される．

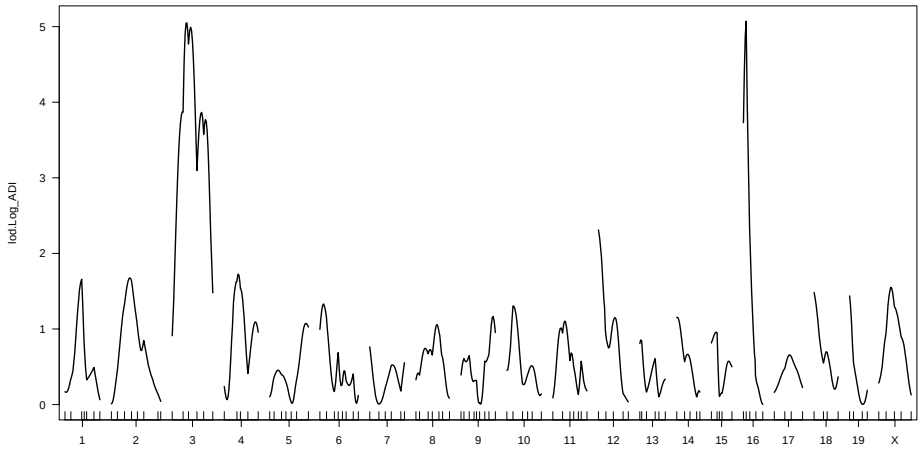


図 1: ロッドスコア

ところでこの解析例のように，観測値として与えられている系列や関数のデータがある時点において変化を示すと考えられるときに，その変化時点を統計的に推測する問題を変化点問題という．一般に変化点問題では (i) 変化点の有無の判断のための統計量の閾値の設定，(ii) 変化点の位置の区間推定，(iii) 複数の変化点が存在する可能性があるときはその個数の推測，などが問題となる．本稿では QTL 解析および関連する連鎖解析において，比較的研究が進んでいる (i) の閾値の設定という問題に焦点を絞り，現時点で知られていることやその背景となる数理について概観する．残念ながら (ii), (iii) については現時点では不十分な結果しか知られておらず，本稿ではほとんど触れることはしない．例えば汎用的手法とされるブートストラップもこれらの解決の役には立たない (Manichaikul, *et al.* (2006))．変化点問題は，より広い立場では特異モデルとよばれるクラスの統計モデルであるが，特異モデルにおいては，正則なモデルで成り立つ種々の漸近的性質が成り立たない (福水，他 (2004))．変化点問題において，ブートストラップやモデル選択規準を，少なくとも正則なモデルと同じ形で用いることはできないのは，このことによる．

ところで，連鎖解析，QTL 解析は，ヒトを対象にするものと実験交配が可能な生物を対象にするものに大別され，その両者では解析の対象とするデータの形が大きく異なる．本稿では後者に対するものを想定する．

## 1.2 多重性調整 (有意水準の調整)

図 1 のようにロッドスコアが明確なピークを持つ場合，その位置付近に QTL が存在すると推測される．しかしながら連鎖や量的形質の発現は確率的な事象であり，それゆえロッドスコアもランダムなグラフである．現われたピークがランダムなゆらぎによる見せかけ

のものでないかどうかを判定するためには，そのための基準，すなわち閾値を合理的に決める必要がある．

ロッドスコアの定義域を$\Gamma$とし，ロッドスコアを$\mathrm{LOD}(\gamma), \gamma \in \Gamma$で表わすとする．$\mathrm{LOD}(\gamma)$は，“QTL が$\gamma$付近に存在しない”という帰無仮説に対する検定統計量であり，その値がある閾値$c$を越えたときに位置$\gamma$の付近に QTL が存在すると判定することができる．ロッドスコアのピークを探索し，その付近に QTL が存在するかどうかを判定することは，全ての$\gamma \in \Gamma$について仮説を同時に検定していると考えられる．そのことから多重検定の考え方によって，閾値$c$を決めることができる．帰無仮説を

$$H_0 : \text{QTL がどの位置にも存在しない}$$

とし，その仮説が真であるにも関わらずロッドが閾値$c$を越え QTL がどこかで発見される (正確には，発見されたと判断される) 事象を偽陽性 (false positive) と定義する．偽陽性確率を$\alpha$ (0.05 あるいは 0.01 など) 以下に調整するためには，閾値$c$を

$$P\Big(\exists\gamma,\ \mathrm{LOD}(\gamma) \geq c_\alpha \,|\, H_0\Big) = \alpha \tag{1.1}$$

となる$c = c_\alpha$とすればよい．この偽陽性確率は，弱い意味での FWER (family-wise error rate in the weak sense) ともよばれる (Hochberg & Tamhane (1987))．上式は

$$P\Big(\max_{\gamma\in\Gamma} \mathrm{LOD}(\gamma) \geq c_\alpha \,|\, H_0\Big) = \alpha \tag{1.2}$$

と書き換えられる．すなわち閾値$c_\alpha$は，確率過程$\mathrm{LOD}(\cdot)$の最大値の上側$\alpha$点である．

ところで$\gamma$を固定すれば，統計量$\mathrm{LOD}(\gamma)$に対する有意点として

$$P\Big(\mathrm{LOD}(\gamma) \geq \tilde{c}_\alpha(\gamma) \,|\, H_0\Big) = \alpha$$

をみたす点$\tilde{c}_\alpha(\gamma)$が定義される．$\mathrm{LOD}(\cdot)$が確率 1 で一定値をとることがない限り，$\max_{\gamma\in\Gamma}\mathrm{LOD}(\gamma)$は$\mathrm{LOD}(\gamma)$よりも確率的に大きな値をとり，$c_\alpha > \tilde{c}_\alpha(\gamma)$ $(0 < \alpha < 1)$ である．逆に，多重検定であることを考慮しないでロッドスコアの棄却点として$\tilde{c}_\alpha(\gamma)$を用いると，偽陽性確率は$\alpha$をこえてしまう．この現象を検定の多重性という．また水準$\alpha$の有意点として，$\tilde{c}_\alpha(\gamma)$の代わりにより値の大きな閾値である$c_\alpha$を用いることを，多重性調整，あるいは有意水準の調整という．

多重検定の多重性調整の方法としてボンフェロニ法が良く知られている．これは，検定の回数 ($\Gamma$の要素数) を$|\Gamma|$とおくとき，全ての$\gamma \in \Gamma$について$\mathrm{LOD}(\gamma)$を水準$\alpha/|\Gamma|$で検定する，すなわち棄却点として$\tilde{c}_{\alpha/|\Gamma|}(\gamma)$を用いる方法である．このとき

$$\begin{aligned} P\Big(\exists\gamma,\ \mathrm{LOD}(\gamma) \geq \tilde{c}_{\alpha/|\Gamma|}(\gamma) \,|\, H_0\Big) &\leq \sum_{\gamma\in\Gamma} P\Big(\mathrm{LOD}(\gamma) \geq \tilde{c}_{\alpha/|\Gamma|}(\gamma) \,|\, H_0\Big) \\ &= \sum_{\gamma\in\Gamma} \alpha/|\Gamma| = \alpha \end{aligned} \tag{1.3}$$

であるので，偽陽性確率は$\alpha$以下に調整される．しかし後で詳しく見るように，$\mathrm{LOD}(\cdot)$は強い相関を持った確率過程であるため，(1.3) の左辺の偽陽性確率は$\alpha$よりも非常に小さな値となり，それにともなって QTL の検出確率 (検出力) も小さなものとなる．とくに

マーカー数が多いとき，あるいはエピスタシスとよばれる複数の QTL による交互作用を検定するときには検定の回数が莫大となりこの傾向が顕著となる．また，2.6 節で説明する区間マッピング法では，ロッドスコアはマーカー間で連続的に補間されるため，$|\Gamma| = \infty$, $\tilde{c}_{\alpha/|\Gamma|}(\gamma) = \infty$ となり，ボンフェロニ法は意味をなさなくなる．以上の理由から，QTL 解析においてはボンフェロニ法を用いることはできない．

次の 2 章では，実験交配における QTL 解析と関連する連鎖解析の統計モデルをいくつか紹介する．さらにそれらのモデルにおいて現われるロッドスコア LOD(·) の確率過程としての構造を調べる．

3 章では，2 章で与えたロッドスコア LOD(·) の構造から，その最大値 $\max_{\gamma\in\Gamma} \mathrm{LOD}(\gamma)$ の上側 $\alpha$ 点 $c_\alpha$ を求める方法を説明する．経験則やシミュレーションに基づく方法に触れた後，理論的な近似法について解説する．確率過程，確率場の最大値の分布については長い研究の歴史がある一方で，近年においても本質的な進展が見られている (Siegmund (1985), Piterbarg (1996), Adler & Taylor (2007), 栗木・竹村 (2007))．本稿では，逐次解析，非線形再生理論を用いる方法と，オイラー標数法とよばれる積分幾何的な手法の 2 通りによって，確率過程としてのロッドスコアの最大値分布の近似を与え，そのことを通して多重性調整が可能であることを見る．

# 2 QTL 解析の統計モデルとロッドスコア

## 2.1 データの形

ここでは QTL 解析が対象とするデータの形と，その背後に想定される確率構造を説明する．交配の実験計画として，BC (戻し交配, backcross) と $\mathrm{F}_2$ の 2 種類を考える．(これらの実験交配については，次節で説明する．)

個体数を $n$ とする．またマーカー遺伝子座の数を $m$ とする．マーカー遺伝子座 (しばしばマーカーと略す) とは，何らかの方法でその遺伝子型が観測できる遺伝子座をいう．個体のそれぞれ $t = 1, \ldots, n$ について，着目している量的形質 (表現型) の測定値 $y^{(t)}$ (スカラ) と $m$ 個のマーカー遺伝子に対する遺伝子型のベクトル $z^{(t)} = (z_1^{(t)}, \ldots, z_m^{(t)})$ が得られている．遺伝子型 $z_i^{(t)}$ は戻し交配の場合は 2 値，$\mathrm{F}_2$ の場合は 3 値をとる．取り扱いの容易さから，

$$z_i^{(t)} = 1, -1 \ \text{(戻し交配の場合)}, \quad 1, 0, -1 \ \text{($\mathrm{F}_2$ の場合)}$$

と表わすことにする (表 1).

表 1: 個体データ

| 個体番号 | 表現型 | 遺伝子型 |
|---|---|---|
| 1 | $y^{(1)}$ | $z^{(1)} = (z_1^{(1)}, \ldots, z_m^{(1)})$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| $n$ | $y^{(n)}$ | $z^{(n)} = (z_1^{(n)}, \ldots, z_m^{(n)})$ |

これらの個体データとは別に，$m$ 個のマーカー遺伝子 $i = 1, \ldots, m$ のそれぞれについて，それが属する染色体の番号 $c_i$ と，その染色体上での位置 $d_i$ の情報が与えられている．$d_i$ は基準となる点からの遺伝的距離 (単位はモルガン M，またはセンチモルガン cM，これらの意味は次節で説明する) で記述され，同一染色体の中では昇順 ($i < j$ ならば $d_i < d_j$) とする (表 2).

表 2: マーカーデータ

| マーカー番号 | 1 | 2 | $\cdots$ | $m$ |
|---|---|---|---|---|
| マーカー名 | $*$ | $*$ | $\cdots$ | $*$ |
| 染色体番号 | 1 | 1 | $\cdots$ | $c$ |
| 座の位置 (M) | $d_1$ | $d_2$ | $\cdots$ | $d_m$ |

QTL 解析では，個体データである遺伝子型と表現型の組 $(z^{(t)}, y^{(t)})$ を確率変数と考え，モデル化を行う．ただし通常の多変量解析と同様に，個体間では独立と考える．

以下では，$m$ 次元の遺伝子型ベクトルデータ $z^{(t)}$ の周辺分布について説明する．これは連鎖によって引き起こされるものである．遺伝子型 $z^{(t)}$ が与えられたときの表現型 $y^{(t)}$ の分布を表現するための統計モデルについては後の節で説明する．マーカーの位置 $d_i$ の単位はモルガンとする．

マーカー遺伝子の添字 $i = 1, \ldots, m$ に対応させる形で，$\pm 1$ に値をとる確率変数 $\epsilon_i$ ($i = 1, \ldots, m$) を考える．ただしこの列はマルコフ系列で，

$$P(\epsilon_1 = \pm 1) = \frac{1}{2}, \quad P(\epsilon_{i+1} = \pm\epsilon_i \,|\, \epsilon_i) = \begin{cases} \frac{1}{2}\left(1 \pm e^{-2(d_{i+1}-d_i)}\right) & (i, i+1 \text{ は同じ染色体上}), \\ \frac{1}{2} & (i, i+1 \text{ は異なる染色体上}) \end{cases}$$

で定義されるものとする．同時確率分布は

$$P(\epsilon_1, \ldots, \epsilon_m) = \frac{1}{2^m} \prod_{i=1}^{m-1} \left(1 + \epsilon_i \epsilon_{i+1} e^{-2(d_{i+1}-d_i)}\right) \tag{2.1}$$

(ただし座 $i$ と座 $i+1$ が同じ染色体上にないならば $d_{i+1} - d_i = \infty$ とおく) である．任意の $i, j$ について

$$P(\epsilon_i = \pm 1) = \frac{1}{2}, \quad P(\epsilon_j = \pm\epsilon_i \,|\, \epsilon_i) = \begin{cases} \frac{1}{2}\left(1 \pm e^{-2|d_j-d_i|}\right) & (i, j \text{ は同じ染色体上}), \\ \frac{1}{2} & (i, j \text{ は異なる染色体上}) \end{cases}$$

となることに注意する．次に $(\delta_1, \ldots, \delta_m) \in \{-1, 1\}^m$ を $(\epsilon_1, \ldots, \epsilon_m)$ と独立に同じ分布に従うランダムベクトルとする．このとき，遺伝子型 $z^{(t)} = (z_1^{(t)}, \ldots, z_m^{(t)})$ に仮定される確率モデルで最も基本的なものは，

$$\begin{aligned} (z_1^{(t)}, \ldots, z_m^{(t)}) &\stackrel{d}{=} (\epsilon_1, \ldots, \epsilon_m) && \text{(戻し交配の場合)}, \\ &\stackrel{d}{=} \frac{1}{2}(\epsilon_1 + \delta_1, \ldots, \epsilon_m + \delta_m) && (\mathrm{F}_2 \text{ の場合}), \\ & && \text{(各 } t \text{ について独立に)} \end{aligned} \tag{2.2}$$

と表わされる．ここで $\stackrel{d}{=}$ は両辺の分布が等しいことを意味する．

遺伝子型のこのような確率構造は，連鎖により引き起こされるものである．次節ではそのことについて説明する．

## 2.2 実験交配と連鎖

一対の染色体 (相同染色体) の一本は母親由来，一本は父親由来である．各個体の遺伝子型を，記法 $A_1B_1\cdots/A_2B_2\cdots$ によって表わす．これは，一方の親に由来する染色体の遺伝子型が $A_1B_1\cdots$，もう一方の親に由来する染色体の遺伝子型が $A_2B_2\cdots$ であることを意味するものとする．ここで $A_1$ と $A_2$ のペア，あるいは $B_1$ と $B_2$ のペアは同じ座に位置する一対の遺伝子 (対立遺伝子) である．全ての遺伝子座についてその遺伝子型がホモ (すなわち $A_1 = A_2, B_1 = B_2, \ldots$) であるとき，近交系という．

近交系は植物の場合は自殖，動物の場合は兄妹 (けいまい) 交配を繰り返すことによって容易に作り出すことができる．実際，自殖あるいは兄妹交配を，遺伝子型を状態とするマルコフチェインによってモデル化すると，その吸収状態が近交系に対応することが確認できる．(遷移行列の固有値の絶対値で 1 でないものの最大は，自殖あるいは兄妹交配のそれぞれについて 0.5, 0.809 であり，前者の方が収束速度が速い．)

ある近交系の個体 $P_1$ と別の近交系の個体 $P_2$ を掛け合わせた雑種第 1 代を $F_1$ 世代という．$F_1$ 個体とその親 $P_1$ ($P_2$) との掛け合わせを戻し交配 $BC_1$ ($BC_2$)，また $F_1$ 個体同士の自殖，あるいは兄妹交配によって得られる雑種第 2 代を $F_2$ 世代という．

以下では 2 つの遺伝子座に着目する．近交系においては，どの遺伝子座においても遺伝子型はホモであるので，異なる近交系の個体の遺伝子型は

$$P_1 = AB/AB, \qquad P_2 = ab/ab$$

と書くことができる．また $P_1$ と $P_2$ の配偶子 (生殖細胞，すなわち卵，精子) の遺伝子型はそれぞれ AB, ab であるので，雑種第 1 代の遺伝子型は

$$F_1 = AB/ab$$

となる．

$F_1$ 個体の配偶子の遺伝子型としては，次の 4 種類

$$(2.3)\quad F_1\text{の配偶子} = AB,\ ab\ \ (\text{それぞれ確率 } \tfrac{1-r}{2} \text{ で}),\quad Ab,\ aB\ \ (\text{それぞれ確率 } \tfrac{r}{2} \text{ で})$$

が現われる．その理由を述べるために，図 2 に沿って配偶子の生成過程 (減数分裂) を説明しよう．

減数分裂では最初に相同染色体の対が分離し，4 本の染色分体となる．さらに互いに別の親に由来する 2 本の染色分体 (図中で色の異なるもの) は，図のように交差を起こす場合がある．(これは確率的な事象である．) その後 4 本の染色分体は互いに分かれて $F_1$ 個体の配偶子が生成される．いま着目している 2 座の間で，奇数回の交差が起きたとすると，生成される生殖細胞の遺伝子は Ab または aB となる．この現象を組換えという．この結果として，$F_1$ 個体の配偶子の遺伝子型は 4 種類 $AB, ab, Ab, aB$ でその頻度は (2.3) の通

りとなる．なお $r$ は 2 座間で組換えが起きる確率であり，組換え価とよばれる．$\mathrm{P}_1, \mathrm{P}_2$ の各個体においても同じ過程で配偶子が生成されるが，4 つの染色分体の遺伝子型は同じであるので結果として組換えは観察されないことに注意する．

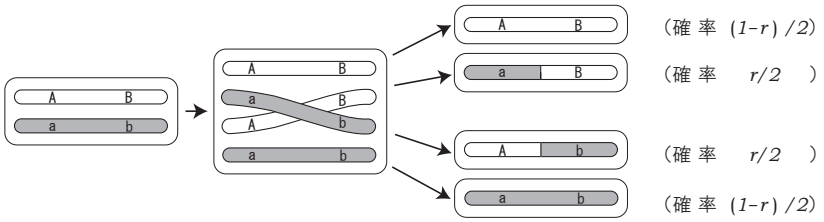


図 2: 減数分裂と交差

さらに $\mathrm{P}_1 = \mathrm{AB/AB}$ と $\mathrm{F}_1 = \mathrm{AB/ab}$ の戻し交配 $\mathrm{BC}_1$ を考えると，その遺伝子型は，

$$\mathrm{BC}_1 = \mathrm{AB/AB},\ \mathrm{AB/ab}\ (\text{それぞれ確率 } \tfrac{1-r}{2} \text{ で}),\quad \mathrm{AB/Ab},\ \mathrm{AB/aB}\ (\text{それぞれ確率 } \tfrac{r}{2} \text{ で})$$

の 4 通りとなる．1 つの遺伝子座に着目した場合，遺伝子型は A/A または A/a の 2 通りであり，一般性を失うことなく 1, −1 で表わすことができる．

また $\mathrm{F}_1$ 個体同士の掛け合わせにより得られる $\mathrm{F}_2$ 個体では，$4 \times 4 = 16$ 通りの遺伝子型が現われる．ただし通常の方法では，2 種類のヘテロ A/a と a/A は識別されない．A/A, A/a, a/a の 3 通りについては，そのマーカー遺伝子が共優性であれば識別することができる．それらを一般性を失うことなく 1, 0, −1 で表わすことにする．2 つの座に着目した場合は，識別できる遺伝子型は $3 \times 3 = 9$ 通りとなる．

交差の確率モデルとして最も基本的なものは，交差の生起をポアソン事象と考えるものである (Haldane (1919))．交差をポアソン事象と考えることにより，染色体上の 2 点間でおこる交差の平均値をもって，その 2 点間の距離と定義することができる．この距離が遺伝的距離で，その単位はモルガン (M) である．この遺伝的距離を遺伝子座の距離と定義することによって，交差の生起事象は強度関数が 1 の定常ポアソン点過程と考えることができる．$x$ (M) 離れた 2 点で，$i$ 回交差がおきる確率は $x^i e^{-x}/i!$ であるので，2 点間の組換え価 $r = r(x)$ は

$$r(x) = P(\text{奇数回交差が起きる}) = \sum_{i:\,\mathrm{odd}} \frac{x^i}{i!} e^{-x} = \frac{1}{2}(1 - e^{-2x}) \tag{2.4}$$

となる．関数 $r(x)$ をホールデンの地図関数という．

$\mathrm{F}_1$ 個体の配偶子の遺伝子型の確率構造を考える．マーカー遺伝子座 $i$ の遺伝子型が $\mathrm{P}_1$ 由来であるとき $\epsilon_i = 1$，$\mathrm{P}_2$ 由来であるとき $\epsilon_i = -1$ とする．ポアソン性の仮定の下で，$(\epsilon_1, \ldots, \epsilon_i)$ と $\epsilon_{i+1} - \epsilon_i$ は独立であるので $\epsilon_1, \ldots, \epsilon_m$ はマルコフ性を持ち，また

$$P(\epsilon_{i+1} = \epsilon_i \,|\, \epsilon_i) = P(\text{座 } i, i+1 \text{ 間で組換えは起らない}) = 1 - r(|d_{i+1} - d_i|),$$

すなわち $\epsilon_1, \ldots, \epsilon_m$ は確率分布 (2.1) に従う $\pm 1$ 列である．

遺伝子型 $z_i$ の定義の仕方より，戻し交配 BC においては $F_1$ 個体の配偶子の $\epsilon_i = 1, -1$ の値が，BC の遺伝子型に一致する．また $F_2$ 個体においては，$\frac{1}{2}(\epsilon_i + \delta_i) = 1, 0, -1$ の値が遺伝子型となる．以上で (2.2) が導出された．

組換え価 $r(x)$ は，$0 \leq r(x) \leq 1/2$ に値をとる単調増加関数である．2 座が同じ遺伝子座 $(x = 0)$ のとき組換えは起らず $(r = 0)$，また 2 つの遺伝子座が非常に離れている，あるいは別の染色体上にあるとき $(x = \infty)$ は確率 $r = 1/2$ で組換えが起きる．2 つの遺伝子座が近くにあり，組換え価が小さな値をとっている状態を，2 座が連鎖するという．染色体の長さは典型的には 100cM (=1M) 程度であり，交差は平均 1 回しか起きない．そのため実験交配の個体の遺伝子型は強い正の相関を持ち，多重性の調整においてはそのことの配慮が必要となる．

**注 2.1** ここでは交差をポアソン点過程によりモデル化しているが，より一般には再生過程を用いてモデル化することができる (Karlin & Liberman (1983))．

## 2.3 単一マーカー分析

本節では QTL 解析の統計モデルとして基本となる単一マーカー分析 (single marker analysis) を説明する．とくに断らない限りは $F_2$ 集団を扱う．

QTL 解析が対象とするデータは，表 1，表 2 の形であった．単一マーカー分析とは，マーカー遺伝子座 $i$ の遺伝子型で集団を 3 群に分け，その 3 群で形質 (表現型) の平均が等しいという仮説に対する分散分析統計量，あるいは尤度比検定統計量を $T(i)$ とおき，その統計量を最も大きくする遺伝子座 $\hat{i} = \operatorname{argmax} T(i)$ の付近に QTL が存在すると判断するという手順である．

この手順に対応する統計モデルは次のようなものである．遺伝子型 $z_i^{(t)} = 1, 0, -1$ に対応した 3 つの群の形質の平均を 3 つのパラメータによって

$$\mu + \alpha + \delta, \quad \mu - \delta, \quad \mu - \alpha + \delta$$

と表現する．QTL は 1 つだけどこかの座 ($i$ とおく) に存在すると仮定すると，モデルは以下のようになる．

(2.5) 1 つの $i \in \{1, \ldots, m\}$ が存在し，

$$y^{(t)} = \mu + \alpha z_i^{(t)} + \delta w(z_i^{(t)}) + \varepsilon^{(t)}, \quad \varepsilon^{(t)} \sim N(0, \sigma^2) \quad (t = 1, \ldots, n).$$

ただし

$$w(z) = \begin{cases} 1 & (z = \pm 1), \\ -1 & (z = 0) \end{cases}$$

とおいた．

このモデル (2.5) に含まれる未知パラメータは $(i, \mu, \alpha, \delta, \sigma^2)$ である．パラメータ $\alpha, \delta$ はそれぞれ QTL の加法効果，優性効果と解釈されている．QTL の効果がない $(\alpha = \delta = 0)$ ときは，QTL の位置パラメータ $i$ は推測不能な量となる．この例のようにパラメータが特別な値をとるときにモデルの識別性が崩れるモデルを特異モデルという (福水，他 (2004))．

QTL の位置 $i$ が既知であるという仮定をおくと，モデルは正則となる．その仮定の下で，パラメータ $(\mu, \alpha, \delta, \sigma^2)$ の最尤推定量を $(\hat{\mu}(i), \hat{\alpha}(i), \hat{\delta}(i), \hat{\sigma}^2(i))$，また QTL が存在しない (QTL の効果がない) という帰無仮説 $\alpha = 0, \delta = 0$ の下での最尤推定量を $(\tilde{\mu}, 0, 0, \tilde{\sigma}^2)$ とおく．サンプルサイズ $n$ を明示する形で，尤度関数を $L_n$ と書くとき，座 $i$ が QTL であるという仮定の下でのロッドスコア $\mathrm{LOD}_n(i)$ および尤度比検定統計量 $\mathrm{LRT}_n(i)$ は

$$\mathrm{LOD}_n(i) = \log_{10} \frac{L_n(\hat{\mu}(i), \hat{\alpha}(i), \hat{\delta}(i), \hat{\sigma}^2(i))}{L_n(\tilde{\mu}, 0, 0, \tilde{\sigma}^2)} = 0.217\, \mathrm{LRT}_n(i)$$

である．(ロッドスコアは尤度比の常用対数として定義される．すなわち尤度比検定統計量の $(2 \log 10)^{-1} = 0.217$ 倍である．) QTL の位置 $i$ も未知とするモデル (2.5) の下で，最尤推定量は $(\hat{\mu}(\hat{i}), \hat{\alpha}(\hat{i}), \hat{\delta}(\hat{i}), \hat{\sigma}^2(\hat{i}))$，ただし

$$\hat{i} = \mathrm{argmax}\, \mathrm{LOD}_n(i)$$

であり，単一マーカー分析における QTL の推測手順がモデル (2.5) の下での $i$ の最尤推定量を与えることが分かる．

以降では，多重性調整のための $\max_{1 \le i \le m} \mathrm{LRT}_n(i)$ の分布計算の準備として，QTL が存在しないという帰無仮説の下での $\mathrm{LRT}_n(i)$ $(i = 1, \ldots, m)$ の同時漸近分布を与える．尤度比検定の一般論より，帰無仮説の下では個体数 $n$ についての漸近的性質として，各 $i$ に対して $\mathrm{LRT}_n(i)$ は漸近的に自由度 2 のカイ 2 乗分布に従う．しかしそれらは独立ではない．カイ 2 乗確率変数の間の相関構造は以下のように表わされる．

**命題 2.1** 座間 $i, j$ の組み換え価を $\frac{1}{2}(1 - \rho_{ij})$ とおく．QTL が存在しないという帰無仮説 $H_0$ の下で，$i = 1, \ldots, m$ の同時分布の意味で分布収束

$$\mathrm{LRT}_n(i) \;\Rightarrow\; T_i = U_i^2 + V_i^2 \quad (n \to \infty) \tag{2.6}$$

が成り立つ．ただし $(U_1, V_1, \ldots, U_m, V_m)$ は平均 0 の $2m$ 次元正規分布ベクトルで

$$\mathrm{Cov}(U_i, U_j) = \rho_{ij}, \quad \mathrm{Cov}(V_i, V_j) = \rho_{ij}^2, \quad \mathrm{Cov}(U_i, V_j) = 0$$

をみたすものである．とくに各 $i$ について $T_i$ は自由度 2 のカイ 2 乗分布に従う．

**証明** しばしば個体を識別する添字 $(t)$ を省略する．$\epsilon_i^{(t)} = \epsilon_i = \pm 1$ を母由来の相同染色体の第 $i$ 座の遺伝子型，$\delta_i^{(t)} = \delta_i = \pm 1$ を父由来のそれとする．このとき

$$z_i^{(t)} = z_i = \frac{1}{2}(\epsilon_i + \delta_i), \quad w(z_i^{(t)}) = w_i = \epsilon_i \delta_i$$

と表わすことができる．

2 つの $m$ 次元ベクトル $(\epsilon_1, \ldots, \epsilon_m), (\delta_1, \ldots, \delta_m) \in \{-1, 1\}^m$ は独立で，それぞれの各成分は平均 0，分散 1，また組換えに由来する相関構造

$$\begin{aligned} E[\epsilon_i \epsilon_j] &= E[\delta_i \delta_j] \\ &= 1 \times P(\epsilon_i = \epsilon_j) + (-1) \times P(\epsilon_i \neq \epsilon_j) \\ &= \frac{1}{2}(1 + \rho_{ij}) - \frac{1}{2}(1 - \rho_{ij}) = \rho_{ij} \end{aligned}$$

を持っていた．$z_i$, $w_i$ の 1, 2 次モーメントは $E[z_i]=0$, $E[w_i]=0$,

$$\begin{aligned}\mathrm{Cov}(z_i,z_j) &= (E[\epsilon_i\epsilon_j]+E[\delta_i\delta_j])/4=\rho_{ij}/2,\\ \mathrm{Cov}(w_i,w_j) &= E[\epsilon_i\delta_i\epsilon_j\delta_j]=E[\epsilon_i\epsilon_j]E[\delta_i\delta_j]=\rho_{ij}^2,\\ \mathrm{Cov}(z_i,w_j) &= (E[\epsilon_i\epsilon_j\delta_j]+E[\delta_i\epsilon_j\delta_j])/2=0\end{aligned}$$

である．

一般性を失うことなくパラメータの真値を $\mu=0$, $\sigma^2=1$ とおく．

$$y=\begin{pmatrix}\vdots\\ y^{(t)}\\ \vdots\end{pmatrix}_{1\le t\le n}\qquad X_i=\begin{pmatrix}\vdots & \vdots\\ z_i^{(t)} & w(z_i^{(t)})\\ \vdots & \vdots\end{pmatrix}_{1\le t\le n}$$

とおく．テイラー展開より

$$\mathrm{LRT}_n(i)\ \approx\ y^TQX_i^T(X_i^TQX_i)^{-1}X_i^TQy,\quad Q=I_n-\mathbf{1}_n\mathbf{1}_n^T/n,\ \ \mathbf{1}_n=(1,\dots,1)^T.$$

ここで $\approx$ は両辺の差が $o_p(1)$ であること意味する．さらに

$$\frac{1}{n}X_i^TQX_i=\frac{1}{n}\sum_{t=1}^n\begin{pmatrix}z_i^{(t)}-\bar z_i\\ w_i^{(t)}-\bar w_i\end{pmatrix}(z_i^{(t)}-\bar z_i,w_i^{(t)}-\bar w_i)\approx\begin{pmatrix}\frac12 & 0\\ 0 & 1\end{pmatrix}$$

に注意すると

$$\mathrm{LRT}_n(i)\ \approx\ \frac{2}{n}\Big\{\sum_{t=1}^n(y^{(t)}-\bar y)z_i^{(t)}\Big\}^2+\frac{1}{n}\Big\{\sum_{t=1}^n(y^{(t)}-\bar y)w_i^{(t)}\Big\}^2\ \approx\ u_i^2+v_i^2,$$

ただし

$$u_i=\sqrt{\frac{2}{n}}\sum_{t=1}^n y^{(t)}z_i^{(t)}\ \Big(\approx\sqrt{\frac{n}{2}}\,\hat\alpha(i)\Big),\qquad v_i=\frac{1}{\sqrt n}\sum_{t=1}^n y^{(t)}w_i^{(t)}\ \Big(\approx\sqrt n\,\hat\delta(i)\Big)$$

である．ここで $u_i$, $v_i$ の平均は 0, 共分散関数は

$$\mathrm{Cov}(u_i,u_j)=2\,\mathrm{Var}(y)\,\mathrm{Cov}(z_i,z_j)=\rho_{ij},\quad \mathrm{Cov}(v_i,v_j)=\mathrm{Var}(y)\,\mathrm{Cov}(w_i,w_j)=\rho_{ij}^2,$$

$$\mathrm{Cov}(u_i,v_j)=\sqrt2\,\mathrm{Var}(y)\,\mathrm{Cov}(z_i,w_j)=0.$$

あとは中心極限定理による． ∎

**注 2.2** 戻し交配の場合は，$i=1,\dots,m$ の同時分布の意味で

$$\mathrm{LRT}_n(i)\ \Rightarrow\ U_i^2\quad(n\to\infty).$$

**注 2.3** 命題 2.1 は，組換え価がホールデンの地図関数 (2.4) であることは仮定していない．(2.4) の仮定の下では，$i,j$ 座間の遺伝的距離を $r_{ij}$ とするとき，$\rho_{ij}=e^{-2r_{ij}}$ であるので

$$\rho_{ij}=\rho_{i,i+1}\rho_{i+1,i+2}\cdots\rho_{j-1,j}.$$

すなわち $(U_i)_{i\ge1}$ および $(V_i)_{i\ge1}$ はマルコフ性を持った正規分布変数列となる．

## 2.4 分離比の検定

ここで述べる分離比の検定は，QTL 解析ではないが，数理的には前節の単一マーカー分析と非常に似た構造を持つ連鎖解析である．

$F_2$ 個体の 1 つの遺伝子座に着目する．対立遺伝子を A, a とする．(2.2) によると，それが共優性である限り，遺伝子型は分離比の期待比率

$$\mathrm{A/A} : \mathrm{A/a} : \mathrm{a/a} = 1 : 2 : 1$$

を持つはずである．(A/a と a/A は区別されていない．) しかし実際の観測値は，理論値である 1:2:1 の比率を持った母集団からのサンプルとはみなされないことがある．このような現象は分離のゆがみとよばれる．そのような現象が起こる理由として，その近くに致死遺伝子 (生殖隔離障壁) が存在することが考えられる．生殖隔離障壁とは，それが特定の遺伝子型をとったときに生殖率，稔性が下がるような遺伝子をいう．Harushima, *et al.* (2001) は，イネの $F_2$ 集団を用いて，そのような分離のゆがみを伴う生殖隔離障壁を検出している．

いま，分離のゆがみの検出のために，表 1，表 2 のデータが利用可能であるとする．ただし表 1 の表現型のデータ $y^{(t)}$ はここでは不要である．

分離比が理論値に従っているかどうかを検定するためには，多項分布のカイ 2 乗適合度検定を用いることができる．各遺伝子型の個体数の観測度数を $n_{\mathrm{A/A}}, n_{\mathrm{A/a}}, n_{\mathrm{a/a}}$ とおく．分離比の検定統計量

$$T_n = \frac{(n_{\mathrm{A/A}} - n/4)^2}{n/4} + \frac{(n_{\mathrm{A/a}} - n/2)^2}{n/2} + \frac{(n_{\mathrm{a/a}} - n/4)^2}{n/4} \qquad (n = n_{\mathrm{A/A}} + n_{\mathrm{A/a}} + n_{\mathrm{a/a}})$$

は，分離比が 1 : 2 : 1 であるという帰無仮説の下で，自由度 2 のカイ 2 乗分布を漸近分布に持つ．

分離のゆがみを検出するためには，$m$ 個のマーカー遺伝子 $(i = 1, \ldots, m)$ について，このカイ 2 乗検定を同時に行う必要がある．そのために，ここでも多重性の調整が必要となる．第 $i$ 座における分離比の検定統計量を $T_{n,i}$ とする．最大値 $\max_{1 \le i \le m} T_{n,i}$ の分布を近似する準備として，$T_{n,i}$ $(i = 1, \ldots, m)$ の同時漸近分布を求めよう．

次のカイ 2 乗統計量の分解に注意する．

$$T_n = \frac{(n_{\mathrm{A/A}} - n/4)^2}{n/4} + \frac{(n_{\mathrm{A/a}} - n/2)^2}{n/2} + \frac{(n_{\mathrm{a/a}} - n/4)^2}{n/4} = U_n^2 + V_n^2,$$

ただし

$$U_n = \sqrt{\frac{2}{n}}(n_{\mathrm{A/A}} - n_{\mathrm{a/a}}), \qquad V_n = \frac{1}{\sqrt{n}}(n_{\mathrm{A/A}} + n_{\mathrm{a/a}} - n_{\mathrm{A/a}}).$$

ここで，もし A/a と a/A が識別可能 (相が既知) とすると，頻度のデータは表 3 の形に集計される．また $V_n = \frac{1}{\sqrt{n}}(n_{\mathrm{A/A}} + n_{\mathrm{a/a}} - n_{\mathrm{A/a}} - n_{\mathrm{a/A}})$ である．1 つの個体 ($t$ 番目とする) の，この表に対する寄与を考える．4 つのセルのうちの 1 か所で 1 回カウントされているはずである．相同染色体の母方染色体の第 $i$ 座の遺伝子型 $\epsilon_i^{(t)} = \pm 1$ と，父方の対応する遺伝子型 $\delta_i^{(t)} = \pm 1$ に立ち返ると，この個体の，表 3 への寄与は表 4 となる．表 4 において，1 つのセルは 1，他の 3 つのセルは 0 である．

表 3: 相が既知の場合

| | A | a | |
|---|---|---|---|
| A | $n_{\mathrm{A/A}}$ | $n_{\mathrm{A/a}}$ | |
| a | $n_{\mathrm{a/A}}$ | $n_{\mathrm{a/a}}$ | |
| | | | $n$ |

表 4: 個体 $t$ の表 3 への寄与

| | A | a | |
|---|---|---|---|
| A | $\frac{1}{4}(1+\epsilon_i^{(t)})(1+\delta_i^{(t)})$ | $\frac{1}{4}(1+\epsilon_i^{(t)})(1-\delta_i^{(t)})$ | $\frac{1}{2}(1+\epsilon_i^{(t)})$ |
| a | $\frac{1}{4}(1-\epsilon_i^{(t)})(1+\delta_i^{(t)})$ | $\frac{1}{4}(1-\epsilon_i^{(t)})(1-\delta_i^{(t)})$ | $\frac{1}{2}(1-\epsilon_i^{(t)})$ |
| | $\frac{1}{2}(1+\delta_i^{(t)})$ | $\frac{1}{2}(1-\delta_i^{(t)})$ | 1 |

このことから，第 $i$ 座の分離比の統計量は $T_{n,i} = U_{n,i}^2 + V_{n,i}^2$，ただし

$$\begin{aligned}
U_{n,i} &= \sqrt{\frac{2}{n}}\Big\{\sum_{t=1}^n \frac{1}{4}(1+\epsilon_i^{(t)})(1+\delta_i^{(t)}) - \sum_{t=1}^n \frac{1}{4}(1-\epsilon_i^{(t)})(1-\delta_i^{(t)})\Big\} \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^n u_i^{(t)}, \qquad u_i^{(t)} = \frac{1}{\sqrt{2}}(\epsilon_i^{(t)} + \delta_i^{(t)}), \\
V_{n,i} &= \frac{1}{\sqrt{n}}\Big\{\sum_{t=1}^n \frac{1}{4}(1+\epsilon_i^{(t)})(1+\delta_i^{(t)}) + \sum_{t=1}^n \frac{1}{4}(1-\epsilon_i^{(t)})(1-\delta_i^{(t)}) \\
&\qquad - \sum_{t=1}^n \frac{1}{4}(1+\epsilon_i^{(t)})(1-\delta_i^{(t)}) - \sum_{t=1}^n \frac{1}{4}(1-\epsilon_i^{(t)})(1+\delta_i^{(t)})\Big\} \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^n v_i^{(t)}, \qquad v_i^{(t)} = \epsilon_i^{(t)}\delta_i^{(t)}
\end{aligned}$$

と分解できる．

2 つの座 $i, j$ に着目する．座間の組み換え価を $\frac{1}{2}(1-\rho_{ij})$ とすると，前節ですでに計算したように，$E[\epsilon_i^{(t)}\epsilon_j^{(t)}] = E[\delta_i^{(t)}\delta_j^{(t)}] = \rho_{ij}$．したがって，$E[u_i^{(t)}] = 0$, $E[v_i^{(t)}] = 0$,

$$\mathrm{Cov}(u_i^{(t)}, u_j^{(t)}) = \rho_{ij}, \qquad \mathrm{Cov}(v_i^{(t)}, v_j^{(t)}) = \rho_{ij}^2, \qquad \mathrm{Cov}(u_i^{(t)}, v_j^{(t)}) = 0$$

が成り立つ．以上から中心極限定理によって，次が従う．

**命題 2.2** 座間 $i, j$ の組換え価を $\frac{1}{2}(1-\rho_{ij})$ とおく．$i = 1, \ldots, m$ の同時分布の意味で，分布収束

$$T_{n,i} \;\Rightarrow\; T_i \quad (n \to \infty)$$

が成り立つ．ただし $T_i$ は命題 2.1 の (2.6) で定義したものである．

つまり単一マーカー分析におけるロッドスコアの同時漸近分布と全く同じものが現われる．

## 2.5 エピスタシス，遺伝子座相互作用の検出

今までは，QTL はたかだか 1 つ存在するというモデルを扱ってきた．しかし量的形質は，複数の QTL によって引き起こされるものと考えられているため，そのようなモデルでは不十分である．とくにエピスタシスとよばれる QTL 間の交互作用を検出するためには，複数の QTL の存在を仮定した統計モデルを使う必要がある．

たとえば 2 つの QTL の存在を仮定した場合，単一マーカー分析のモデル (2.5) に対応するものとして，次のモデルが考えられる．

$$
\begin{aligned}
&i, j \in \{1, \dots, m\} \text{ が存在し，} \\
&\quad y^{(t)} = \mu + \alpha_1 z_i^{(t)} + \delta_1 w_i^{(t)} + \alpha_2 z_j^{(t)} + \delta_2 w_j^{(t)} \\
&\qquad\quad + \beta_1 z_i^{(t)} z_j^{(t)} + \beta_2 z_i^{(t)} w_j^{(t)} + \beta_3 w_i^{(t)} z_j^{(t)} + \beta_4 w_i^{(t)} w_j^{(t)} \\
&\qquad\quad + \varepsilon^{(t)}, \quad \varepsilon^{(t)} \sim N(0, \sigma^2) \quad (t = 1, \dots, n),
\end{aligned}
$$

ただし

$$
w_i^{(t)} = w(z_i^{(t)}) = \begin{cases} 1 & (z_i^{(t)} = \pm 1), \\ -1 & (z_i^{(t)} = 0). \end{cases}
$$

座 $i$ と座 $j$ が QTL であるという仮定の下で，エピスタシスが存在しないという帰無仮説 $\beta_1 = \cdots = \beta_4 = 0$ の尤度比検定を考える．サンプルサイズ $n$ に対するロッドスコア (尤度比検定統計量) を，$\mathrm{LRT}_n(i, j)$ と書く．多重性調整のため，エピスタシスが存在しないという帰無仮説の下での $\mathrm{LRT}_n(i, j)$ $(i, j = 1, \dots, m)$ の同時漸近分布を与えたい．尤度比検定の一般論より，帰無仮説の下では個体数 $n$ についての漸近的性質として，各 $i, j$ に対して $\mathrm{LRT}_n(i, j)$ は自由度 4 のカイ 2 乗分布に従う．しかしそれらは独立ではない．その相関構造は複雑なものとなるが，2 座 $i, j$ が同じ染色体の上にない場合は，次のような直積型構造であることが示される．

**命題 2.3** 座間 $i, j$ の組換え価を $\frac{1}{2}(1 - \rho_{ij})$ とおく．座 $i$ と座 $j$ が同じ染色体上にある (ない) ことを $i \sim j$ $(i \nsim j)$ と書く．エピスタシスが存在しないという帰無仮説の下で，全ての

$$
(i, j) \in \{(i, j) \mid 1 \le i, j \le m,\ i \nsim j\}
$$

についての同時分布の意味で，分布収束

$$
\mathrm{LRT}_n(i, j) \ \Rightarrow\ T_{ij} = U_{1,ij}^2 + \cdots + U_{4,ij}^2 \quad (n \to \infty)
$$

が成り立つ．ただし $(U_{k,ij})$ は $k$ が異なると互いに独立な平均 0 の正規分布の配列で

$$
\mathrm{Cov}(U_{k,ij}, U_{k,i'j'}) = \rho_{ii'}\rho_{jj'}\ (k = 1),\ \rho_{ii'}^2\rho_{jj'}\ (k = 2),\ \rho_{ii'}\rho_{jj'}^2\ (k = 3),\ \rho_{ii'}^2\rho_{jj'}^2\ (k = 4)
$$

をみたすものである．

春島，他 (2006) は，遺伝子座相互作用によって引き起こされる生殖隔離障壁を検出するために，2 座 $i, j$ の遺伝子型の組合せとして得られる $3 \times 3$ 表の独立性検定を行い，相互作用を独立性の乖離として検出することを試みた．その検定統計量の同時分布は帰無仮説の下で命題 2.3 と同じであることを示すことができる．

## 2.6 区間マッピング法と Haley-Knott の回帰分析

単一マーカー分析では，各マーカー遺伝子座についてロッドスコアが計算された．ここで説明する 2 つの方法は，ロッドスコアを補完によってマーカー遺伝子座の位置以外でも定義するものであり，マーカー間隔が密でない場合に有効である．

ここでも $F_2$ 集団で考える．QTL がある位置に 1 つ存在してそれが形質に影響を与えるというモデルを考える．このような仮想 QTL のことを putative QTL という．個体 $t$ の QTL の遺伝子型を $z_*^{(t)}$ とおく．これは $1, 0, -1$ の値をとる潜在変数である．さらにこれは，相同染色体上の母方，父方の遺伝子型 $\epsilon_*^{(t)}, \delta_*^{(t)}$ $(= \pm 1)$ によって，$z_*^{(t)} = \frac{1}{2}(\epsilon_*^{(t)} + \delta_*^{(t)})$ と表される．

以下では混乱のない限り個体の添字 $(t)$ を省略する．仮想 QTL の位置を $\gamma$ とする．いま $d_i \le \gamma \le d_{i+1}$，つまり QTL はマーカー遺伝子座 $i$ と $i+1$ の間に存在するとする．$\epsilon = (\epsilon_1, \ldots, \epsilon_m)$ と $\epsilon_*$ の同時分布，$\delta = (\delta_1, \ldots, \delta_m)$ と $\delta_*$ の同時分布は，(2.1) と同様にマルコフ性によって $\gamma$ の関数として陽に書き下すことができる．このことから，

$$z = (z_1, \ldots, z_m) = \frac{1}{2}(\epsilon + \delta) = \frac{1}{2}(\epsilon_1 + \delta_1, \ldots, \epsilon_m + \delta_m)$$

が与えられたときの，QTL の遺伝子型 $z_*$ の条件付き分布が以下のように与えられる．

$$P(z_* \,|\, z\,; \gamma) = \frac{P(z, z_*\,; \gamma)}{P(z)}, \tag{2.7}$$

ただし

$$\begin{aligned} P(z, z_*\,; \gamma) = & \sum_{z=(\epsilon+\delta)/2,\ z_*=(\epsilon_*+\delta_*)/2} \\ & \frac{1}{2^{m+1}} \prod_{j=1,\ j\neq i}^{m-1} \Big(1 + \epsilon_j \epsilon_{j+1} e^{-2(d_{j+1}-d_j)}\Big)\Big(1 + \delta_j \delta_{j+1} e^{-2(d_{j+1}-d_j)}\Big) \\ & \times \Big(1 + \epsilon_i \epsilon_* e^{-2(\gamma - d_i)}\Big)\Big(1 + \epsilon_* \epsilon_{i+1} e^{-2(d_{i+1}-\gamma)}\Big) \\ & \times \Big(1 + \delta_i \delta_* e^{-2(\gamma - d_i)}\Big)\Big(1 + \delta_* \delta_{i+1} e^{-2(d_{i+1}-\gamma)}\Big), \qquad d_i \le \gamma \le d_{i+1}, \end{aligned}$$

$$P(z) = \sum_{z=(\epsilon+\delta)/2} \frac{1}{2^m} \prod_{j=1}^{m-1} \Big(1 + \epsilon_j \epsilon_{j+1} e^{-2(d_{j+1}-d_j)}\Big)\Big(1 + \delta_j \delta_{j+1} e^{-2(d_{j+1}-d_j)}\Big). \tag{2.8}$$

(2.7) の $P(z_* \,|\, z\,; \gamma)$ は，$\gamma \to d_i$ または $d_{i+1}$ のとき $z_i$ または $z_{i+1}$ に確率 1 で値をとる一点分布となる．$P(z_* \,|\, z\,; \gamma)$ は $\gamma$ の連続関数であるが，マーカー点で滑らかではない．

仮想 QTL の遺伝子型の情報は，$m$ 個のマーカーのなかで，とくに QTL に隣接するマーカー (flanking marker) $i$ と $i+1$ が多く持っていると考えられる．そのため QTL の遺伝

子型の予測のために (2.7) の代わりに $P(z_* \,|\, z_i,\, z_{i+1}\,;\gamma)$ を考えることもできる．これは簡便法であるが，戻し交配の場合のように，マーカー遺伝子型自体にマルコフ性が成り立つばあいには，(2.7) と正確に一致する．

Lander & Botstein (1989) の区間マッピング法 (interval mapping) とは，次の統計モデルを仮定した解析法である．

$$z_*^{(t)} \;\sim\; P(z_*^{(t)} \,|\, z^{(t)};\gamma), \tag{2.9}$$

$$y^{(t)} \;=\; \mu + \alpha z_*^{(t)} + \delta w(z_*^{(t)}) + \varepsilon^{(t)}, \quad \varepsilon^{(t)} \sim N(0,\sigma^2) \tag{2.10}$$

$(t = 1, \ldots, n)$，ただし

$$w(k) = \begin{cases} 1 & (k = \pm 1), \\ -1 & (k = 0). \end{cases}$$

また Haley & Knott (1992) は，区間マッピング法の簡便法として，潜在変数をその期待値で置きかえた次の回帰分析を提案した．

$$y^{(t)} \;=\; \mu + \alpha E[z_*^{(t)} \,|\, z^{(t)}\,;\gamma] + \delta E[w(z_*^{(t)}) \,|\, z^{(t)}\,;\gamma] + \varepsilon^{(t)}, \tag{2.11}$$
$$\varepsilon^{(t)} \sim N(0,\sigma^2) \quad (t = 1, \ldots, n).$$

両者において，仮想 QTL の位置 $\gamma$ の関数として尤度関数が定義されるので，連続な曲線としてロッドスコアが定義される．

また両者ともに論文で提案されたオリジナルの形は，QTL の遺伝子型の分布として隣接マーカーの情報だけを用いた $P(z_* \,|\, z_i, z_{i+1}\,;\gamma)$ を仮定するものであるが，本稿では全マーカーの情報を用いた $P(z_* \,|\, z\,;\gamma)$ で考えることにする．

以下では，区間マッピング法の尤度関数の形を書き下し，ロッドスコアの確率過程としての構造を調べていくことにする．Haley-Knott の回帰分析の場合は最後に触れる．

マーカーの遺伝子型 $z^{(t)}$ と QTL の位置 $\gamma$ が与えられたとき $z_*^{(t)}$ は 3 値の離散分布に従う．記法を簡単にするために，その確率を (2.7) を使って

$$\pi_k^{(t)}(\gamma) = P(z_*^{(t)} = k \,|\, z^{(t)}\,;\gamma), \quad k = -1, 0, 1$$

とおくと，$z^{(t)}$ が所与のときの $y^{(t)}$ の分布は，コンポーネント数が 3 の正規分布の有限混合分布

$$\sum_{k=-1}^{1} \pi_k^{(t)}(\gamma) f_{k,\theta}(y^{(t)})$$

となる．ただし正規分布 $N(\mu + \alpha k + \delta w(k), \sigma^2)$ の密度関数を $f_{k,\theta}(\cdot)$, $\theta = (\alpha, \delta, \mu, \sigma^2)$ とおいた．したがって，$(z^t, y^{(t)})$ $(t = 1, \ldots, n)$ の同時密度関数は，(2.8) を用いて

$$\prod_{t=1}^{n} \Big\{ \sum_{k=-1}^{1} \pi_k^{(t)}(\gamma) f_{k,\theta}(y^{(t)}) P(z^{(t)}) \Big\}$$

と書ける．

ロッドスコアを描くためには，各 $\gamma$ についてこの尤度を $\theta$ について最大化する必要がある．そのために EM アルゴリズム (Wu (1983) など) を用いることができる．ここで扱う

モデルは，混合確率は個体 $t$ に依存すること，また ($\gamma$ を所与とすると) 混合確率には推測対象の未知パラメータが含まれないこと，の 2 点で通常の有限混合モデルとはやや異なっている．

$z_*^{(t)}$ と一対一に対応するダミー変数ベクトル

$$e^{(t)} = (e_{-1}^{(t)}, e_0^{(t)}, e_1^{(t)}), \quad e_k^{(t)} = \begin{cases} 1 & (z^{(t)} = k) \\ 0 & (z^{(t)} \neq k) \end{cases}$$

を導入する．マーカー遺伝子の遺伝子型，表現型，ならびに潜在変数 $(z^{(t)}, y^{(t)}, e^{(t)})$ $(t = 1, \ldots, n)$ の同時分布は

$$\prod_{t=1}^{n} \Big[ \prod_{k=-1}^{1} \Big\{ \pi_k^{(t)}(\gamma) f_{k,\theta}(y^{(t)}) \Big\}^{e_k^{(t)}} P(z^{(t)}) \Big]$$

と書けるので

$$E_\theta \Big[ e_k^{(t)} \mid (z^{(t)}, y^{(t)})_{t=1,\ldots,n} \Big] = \frac{\pi_k^{(t)}(\gamma) f_{k,\theta}(y^{(t)})}{\sum_{k=-1}^{1} \pi_k^{(t)}(\gamma) f_{k,\theta}(y^{(t)})}$$

となる．これを用いて，EM アルゴリズムは次のようにまとめられる．

1. 初期値
$$\hat{e}_k^{(t)} := \pi_k^{(t)}(\gamma) \quad (k = -1, 0, 1;\ t = 1, \ldots, n).$$

2. 以下を $\hat{\theta}$ と $\hat{e}_k^{(t)}$ が収束するまで繰りかえす．
$$\hat{\theta} := \operatorname*{argmax}_\theta \prod_{t=1}^{n} \prod_{k=-1}^{1} \Big\{ \pi_k^{(t)}(\gamma) f_{k,\theta}(y^{(t)}) \Big\}^{\hat{e}_k^{(t)}},$$
$$\hat{e}_k^{(t)} := \frac{\pi_k^{(t)}(\gamma) f_{k,\hat{\theta}}(y^{(t)})}{\sum_{k=-1}^{1} \pi_k^{(t)}(\gamma) f_{k,\hat{\theta}}(y^{(t)})} \quad (k = -1, 0, 1;\ t = 1, \ldots, n).$$

ステップ 2 において $\hat{\theta}$ は，

$$Q_n(\theta\,;\gamma) = \sum_{t=1}^{n} \sum_{k=-1}^{1} \hat{e}_k^{(t)} \Big\{ \frac{1}{\sigma^2} (y^{(t)} - \mu - \alpha k - \delta w(k))^2 + \log \sigma^2 + (\text{定数}) \Big\}$$

を $\theta = (\alpha, \delta, \mu, \sigma^2)$ について最小化する操作 (重み付き最小 2 乗法) で陽に求めることができる．

1 章の図 1 は区間マッピング法により描いたものである．同じものを第 3 染色体について拡大して描いたものが図 3 である．ロッドスコア曲線が，マーカー間を補完していることが分かる．

**注 2.4** ここで区間マッピング法におけるモデリングを振りかえってみる．統計モデルは，連鎖を記述する部分 (2.9) と，形質発現を記述する部分 (2.10) で構成されていた．後者の形質発現モデルを工夫することによって，QTL が複数ある場合や，表現型が非正規分布や離散分布，あるいは多変量分布に従う場合などを扱うことができる．このような統一的な扱いは，Sen & Churchill (2001) により提案され，R/qtl (Broman, *et al.* (2003)) として実装されている．

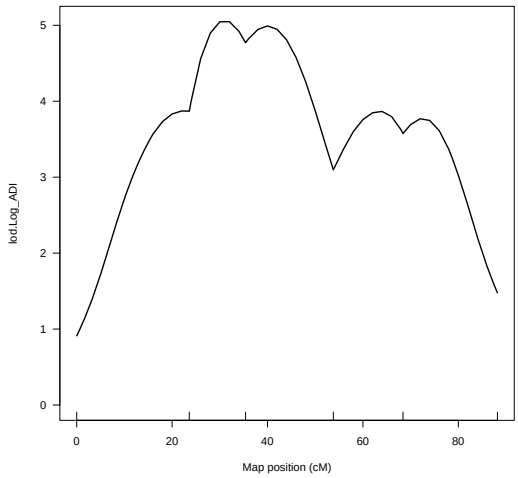


図 3: 区間マッピング法によるロッドスコア (第 3 染色体)

## 2.7 区間マッピング法のロッドスコア

区間マッピング法のロッドスコア (の定数倍である) $\mathrm{LRT}_n(\gamma)$ は，各 $\gamma$ について尤度比検定統計量であり，QTL が存在しないという帰無仮説の下で漸近的に自由度 2 のカイ 2 乗分布に従う．ここでは，QTL が存在しないという帰無仮説の下で，$\mathrm{LRT}_n(\cdot)$ を確率過程 (漸近的カイ 2 乗確率過程) とみなしたときの相関構造を確定する．これは多重性調整のために必要である．

対数尤度は

$$L_n^{(\gamma)}(\theta) = \sum_{t=1}^{n} \log \sum_{k=-1}^{1} \pi_k^{(t)}(\gamma) f_{k,\theta}(y^{(t)}) + (\theta \text{ を含まない項}), \quad \theta = (\alpha, \delta, \mu, \sigma^2)$$

であった．

パラメータ $\theta$ を $\theta = (\theta_1, \theta_2)$，ただし $\theta_1 = (\alpha, \delta)$, $\theta_2 = (\mu, \sigma^2)$ と分割表現する．帰無仮説は $H_0 : \theta_1 = 0$ である．一般性を失わずに，真値を $\theta_0 = (\theta_{10}, \theta_{20}) = (0, 0, 0, 1)$ とする．

$\gamma$ を固定し，$\theta$ に関するフィッシャー情報行列を

$$I(\theta_0; \gamma) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix} = \mathrm{Var}_{\theta_0}\Big( \frac{\partial L_n^{(\gamma)}}{\partial \theta}\Big|_{\theta_0} \Big) \tag{2.12}$$

とおく．尤度比検定の一般論から，真値 $\theta_0$ の下で

$$\mathrm{LRT}_n(\gamma) \;\approx\; \left\{ \Big( \frac{\partial L_n^{(\gamma)}}{\partial \theta_1}, \frac{\partial L_n^{(\gamma)}}{\partial \theta_2} \Big) \begin{pmatrix} I \\ -I_{22}^{-1} I_{21} \end{pmatrix} I_{11\cdot 2}^{-1} \Big( I, -I_{12} I_{22}^{-1} \Big) \begin{pmatrix} \frac{\partial L_n^{(\gamma)}}{\partial \theta_1} \\ \frac{\partial L_n^{(\gamma)}}{\partial \theta_2} \end{pmatrix} \right\}_{(\theta_{10}, \theta_{20})}$$

ただし $I_{11\cdot 2} = I_{11} - I_{12} I_{22}^{-1} I_{21}$ である．

帰無仮説 $\alpha=\delta=0$ のとき，$f_{k,\theta}$ は $k$ によらない．このことから，スコア関数を真値で評価すると

$$
\begin{aligned}
\frac{\partial L_n^{(\gamma)}}{\partial\theta}\Big|_{\theta_0} &= \sum_{t=1}^{n}\frac{\sum_k \pi_k^{(t)}(\gamma) f_{k,\theta}(y^{(t)})\frac{\partial}{\partial\theta}\log f_{k,\theta}(y^{(t)})}{\sum_k \pi_k^{(t)}(\gamma) f_{k,\theta}(y^{(t)})}\Big|_{\theta_0} \\
&= \sum_{t=1}^{n}\sum_{k=-1}^{1}\pi_k^{(t)}(\gamma)\frac{\partial}{\partial\theta}\log f_{k,\theta}(y^{(t)})\Big|_{\theta_0} \\
&= \sum_{t=1}^{n}\begin{pmatrix} y^{(t)}\sum_k k\pi_k^{(t)}(\gamma) \\ y^{(t)}\sum_k w(k)\pi_k^{(t)}(\gamma) \\ y^{(t)} \\ (y^{(t)2}-1)/2 \end{pmatrix}
\end{aligned}
\tag{2.13}
$$

である．$L_n^{(\gamma)}(\theta_0)$ は $\gamma$ に依存しない (特異モデルであるため) が，$\partial L_n^{(\gamma)}/\partial\theta|_{\theta_0}$ は $\gamma$ に依存することに注意する．

スコアベクトルの分散共分散関数を

$$
R(\gamma,\tilde\gamma)=\frac{1}{n}\mathrm{Cov}_{\theta_0}\Big(\frac{\partial L_n^{(\gamma)}}{\partial\theta}\Big|_{\theta_0},\frac{\partial L_n^{(\tilde\gamma)}}{\partial\theta}\Big|_{\theta_0}\Big)
$$

とおく．

$$
\sum_{k=-1}^{1}k\pi_k^{(t)}(\gamma)=E\big[z_*^{(t)}\,|\,z^{(t)};\gamma\big],\qquad \sum_{k=-1}^{1}w(k)\pi_k^{(t)}(\gamma)=E\big[w(z_*^{(t)})\,|\,z^{(t)};\gamma\big]
$$

に注意すると，簡単な計算により

$$
R(\gamma,\tilde\gamma)=\begin{pmatrix} R_{11}(\gamma,\tilde\gamma) & & O \\ & 1 & 0 \\ O & 0 & 1/2 \end{pmatrix}_{4\times4}
$$

ただし

$$
\begin{aligned}
R_{11}(\gamma,\tilde\gamma) &= E\left[\begin{pmatrix}\sum_k k\pi_k^{(t)}(\gamma)\\ \sum_k w(k)\pi_k^{(t)}(\gamma)\end{pmatrix}\Big(\sum_k k\pi_k^{(t)}(\tilde\gamma),\sum_k w(k)\pi_k^{(t)}(\tilde\gamma)\Big)\right] \\
&= E\left[\begin{pmatrix}E[z_*\,|\,z;\gamma]\\ E[w(z_*)\,|\,z;\gamma]\end{pmatrix}\Big(E[z_*\,|\,z;\tilde\gamma],E[w(z_*)\,|\,z;\tilde\gamma]\Big)\right]
\end{aligned}
\tag{2.14}
$$

が分かる．ここで外側の期待値はマーカーの遺伝子型 $z$ についてとる．フィッシャー情報行列 (2.12) は $I(\theta_0;\gamma)=nR(\gamma,\gamma)$ であり，ブロック対角行列 ($I_{12}=I_{21}^T=0$) となる．

$C(\gamma)$ を各成分が $\gamma$ について滑らかな $2\times2$ 行列で $C(\gamma)R_{11}(\gamma,\gamma)C(\gamma)^T=I_2$ をみたすものとする．中心極限定理より以下が示される．

**命題 2.4** 有限個の $\gamma$ の，有限次元周辺分布の分布収束の意味で

$$
\mathrm{LRT}_n(\gamma)\ \Rightarrow\ T(\gamma)=U(\gamma)^2+V(\gamma)^2\quad(n\to\infty).
$$

ただし，$U(\cdot), V(\cdot)$ は平均 0 の正規過程で，その共分散関数は (2.14) の $R_{11}(\gamma, \tilde{\gamma})$ を用いて

$$R^{UV}(\gamma, \tilde{\gamma}) = \mathrm{Cov}\left(\begin{pmatrix} U(\gamma) \\ V(\gamma) \end{pmatrix}, \begin{pmatrix} U(\tilde{\gamma}) \\ V(\tilde{\gamma}) \end{pmatrix}\right) = C(\gamma) R_{11}(\gamma, \tilde{\gamma}) C(\tilde{\gamma})^T \tag{2.15}$$

として記述される．とくに固定した $\gamma$ について $\mathrm{LRT}_n(\gamma)$ は漸近的に自由度 2 のカイ 2 乗分布に従う．

**注 2.5** 上の命題は，確率過程としての収束は述べていない．確率過程としての収束は，尤度比確率場の弱収束に関する Ibragimov & Has'minskii (1981) の方法で示すことができる (Yoshida (2006), Introduction を参照).

**注 2.6** Haley-Knott の回帰分析 (2.11) のスコア関数を真値 $\theta_0$ で評価すると，(2.13) と同じ形となる．ロッドスコアは区間マッピング法と同じ極限分布を持つ.

# 3 多重性調整のための方法

## 3.1 経験的方法とシミュレーション

前章ではいろいろな QTL 解析のモデルについて，QTL が存在しないという帰無仮説のもとで，ロッドスコアの確率構造を確定した．その結果を出発点として，本章では多重性調整のために必要な，ロッドスコアの最大値の分布の近似法について説明する．そのために利用可能な数学的な方法として，非線形再生理論やオイラー標数法がある．それらについては，後の 3.2, 3.3 節で説明することとし，本節では QTL 解析の多重性調整のために行われている経験則とシミュレーションによる方法を紹介する.

Lander & Kruglyak (1995) は，ロッドスコアの多重性調整の目安として，多重性未調整の $p$ 値を表 5 に従って解釈することを提唱している．しかしゲノムワイドの多重性調整は，染色体長のみならずマーカーの密度に大きく依存する．(3.2 節，表 4 の数値実験を参照．) そのため，機械的にこの表を用いることは行われてはいない (石川 (2006) など).

表 5: 多重性未調整 $p$ 値の解釈

| 実験交配の手法 | suggestive | significant |
|---|---|---|
| BC (1 d.f.) | $3.4 \times 10^{-3}$ | $1.0 \times 10^{-4}$ |
| $F_2$ (1 d.f., 加法効果) | $3.4 \times 10^{-3}$ | $1.0 \times 10^{-4}$ |
| $F_2$ (1 d.f., 優性効果) | $2.4 \times 10^{-3}$ | $7.2 \times 10^{-4}$ |
| $F_2$ (2 d.f) | $1.6 \times 10^{-3}$ | $5.2 \times 10^{-5}$ |

多重性調整の方法として現在広く用いられている方法は，Churchill & Doerge (1994) の提案による並べ替え検定である．その手順は次のようなものである.

1. 個体数を $n$ とし，集合 $\{1, \ldots, n\}$ の置換の全体を $\Pi_n$ とおく.

2. $N$ を十分大きな数とし，以下の手順を $N$ 回繰りかえす．その繰り返しを $k = 1, \ldots, N$ とする．

   (i) ランダムに $\pi \in \Pi_n$ を選ぶ．

   (ii) 表現型 $(y^{(t)})_{t=1,\ldots,n}$ を，置換 $\pi$ によって入れ替えたデータセット (表 6) について QTL 解析を行い，ロッドスコアの最大値を数値的に探索する．それを $\mathrm{MaxLOD}_k$ とおく．

3. $\{\mathrm{MaxLOD}_k\}_{k=1,\ldots,N}$ の経験分布をロッドスコアの最大値の経験分布とみなして $p$ 値の推定値を

$$\widehat{p\text{値}} = \frac{\#\{k \mid \mathrm{MaxLOD}_k \geq \mathrm{MaxLOD}\ (\text{実現値})\}}{N}$$

   と計算する．

表 6: 並べ替え検定のためのデータセット

| 個体番号 | 表現型 | 遺伝子型 |
|---|---|---|
| 1 | $y^{(\pi(1))}$ | $z^{(1)} = (z_1^{(1)}, \ldots, z_m^{(1)})$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| $n$ | $y^{(\pi(n))}$ | $z^{(n)} = (z_1^{(n)}, \ldots, z_m^{(n)})$ |

この並べ替え検定は直感的に分かりやすい方法であり，R/qtl などの多くのプログラムに実装されている．しかし次のような問題点がある．

最初の問題点は，並べ替え検定の一般論に関わる問題である．並べ替え検定が生成する $\{\mathrm{MaxLOD}_k\}_{k=1,\ldots,N}$ の経験分布は，帰無仮説が成り立つ場合とそうでない場合とでは当然異なるものである．そのために，帰無仮説が正しい場合には閾値を正しく推定できたとしても，帰無仮説が正しくない場合には閾値を過大評価する可能性がある．そのときは，ピークの検出確率 (検出力) の低下を招くことになる．

もうひとつの問題点は，計算量の問題である．説明した並べ替え検定の手順では，ロッドスコアの最大値を数値的に探索する必要があるため，計算時間がかかる．QTL を 1 つしか想定しない場合は問題ないが，エピスタシスの検定などで複数の QTL を想定する場合は，探索すべき組合せ数が莫大となり，計算が実行可能でなくなる場合がある．

並べ替え検定は汎用的でしばしば用いられる手法であるが，これらの理由のためその解釈や利用に注意を払う必要がある．可能な限り，他の多重性調整の方法の結果と併用するのがよいと思われる．問題を単一マーカー分析あるいは分離比の検定における多重性調整に限定すると，AR モデルを用いたモンテカルロシミュレーションも利用可能である．

1. $\epsilon_1, \ldots, \epsilon_m, \delta_1, \ldots, \delta_m$ を標準正規分布 $N(0,1)$ に従う i.i.d. 列とする．$U_1 = \epsilon_1, V_1 = \delta_1$ とおく．

2. $i = 2, \ldots, m$ について以下を計算する．

$$\begin{aligned} U_i &= \alpha_i U_{i-1} + \sqrt{1-\alpha_i^2}\epsilon_i \quad (\alpha_i = e^{-2(d_i - d_{i-1})}), \\ V_i &= \beta_i V_{i-1} + \sqrt{1-\beta_i^2}\delta_i \quad (\beta_i = e^{-4(d_i - d_{i-1})}), \\ T_i &= U_i^2 + V_i^2. \end{aligned}$$

3. $\max_{1\le i\le m} T_i$ を計算する．

上の手続きを十分な回数繰り返すことによって，命題 2.1, 2.2 の $\max_{1\le i\le m} T_i$ の経験分布を求めることができる．しかしながら，並べ替え検定のときと同様，複数の QTL を想定した場合には計算量が膨大となる．

## 3.2 非線形再生理論による近似

本節以降では，ロッドスコアの最大値の分布を理論的に求める (近似する) ための方法を 2 つ紹介する．最初に述べる方法は，逐次解析，非線形再生理論を用いるものであり，単一マーカー分析や分離のゆがみの検定において，マーカーが密で間隔が等間隔に近い場合 ($|d_{i+1} - d_i| \approx \Delta \ll 1$) に有効な方法である (Dupuis & Siegmund (1999))．

まず一般的な形で問題設定を行う．$Z_k(t),\ t \in I \subset \mathbb{R}$ $(k = 1, 2)$ を互いに独立な正規過程で，平均 0，共分散関数が

$$\mathrm{Cov}(Z_k(t), Z_k(\tilde{t})) = e^{-\rho_k |t-\tilde{t}|} \quad (\rho_k > 0)$$

であるもの (Ornstein-Uhlenbeck 過程) とする．自由度 2 のカイ 2 乗確率過程の平方根を

$$Y(t) = \sqrt{Z_1(t)^2 + Z_2(t)^2} \quad (t \in I)$$

で定義する．

**命題 3.1** 格子間隔を $\Delta > 0$ とし，$J = \{j \in \mathbb{Z} \mid j\Delta \in I\}$ とおく．$b \to \infty$, $\Delta \to 0$, $b\sqrt{\Delta} \to c\ (> 0)$ のとき

$$P\Big(\max_{j\in J} Y(j\Delta) \ge b\Big) \ \sim\ |I|\, b^2 e^{-b^2/2} \int_0^{2\pi} \frac{d\theta}{2\pi}\, \bar{\rho}(\theta)\, \nu(c\sqrt{2\bar{\rho}(\theta)}). \tag{3.1}$$

ここで $|\cdot|$ は集合のルベーグ測度，$\bar{\rho}(\theta) = \rho_1 \cos^2\theta + \rho_2 \sin^2\theta$．また $\Phi(\cdot)$ を標準正規分布の分布関数とするとき

$$\nu(x) = \begin{cases} 2x^{-2} \exp\Big\{-2\displaystyle\sum_{n=1}^{\infty} n^{-1} \Phi\Big(-\frac{1}{2}x\sqrt{n}\Big)\Big\} & (x > 0), \\ 1 & (x = 0). \end{cases}$$

**注 3.1** $x \le 2$ くらいでは $\nu(x) \approx \exp(-0.583x)$ と近似できる．実用的にはこの範囲で足りることが多い．

**注 3.2** 形式的に $\Delta = 0$, $c = 0$ とおいて得られる式

$$P\Big(\sup_{t\in I} Y(t) \geq b\Big) \sim |I| K b^2 e^{-b^2/2}, \quad K = \int_0^{2\pi} \frac{d\theta}{2\pi} \bar{\rho}(\theta) = \frac{1}{2}(\rho_1 + \rho_2) \quad (b \to \infty)$$

も成り立つ (Piterbarg (1996), Corollary 7.1 より).

$I = [0, 1]$, $\Delta = 0.01$ (染色体長 1M, マーカー数 $m = 100$ に相当) の場合の数値例を図 4 に示す．一点鎖線 $(-\cdot-)$ と実線 (—) はそれぞれ格子点上の最大値の上側確率 $P\big(\max_{j\in J} Y(j\Delta) \geq b\big)$ の命題 3.1 による近似値とシミュレーションによる推定値を表している．すなわち図の縦軸は $p$ 値に対応する．ただしここでは最大値分布の近似として (3.1) の右辺をそのまま用いるのではなく，それと漸近的に同等な $1 - \exp\{-(3.1)$ 右辺 $\}$ を用いている．この両者は $p$ 値が 0.5 程度より小さい範囲で非常に精度良く一致していることが見て取れる．

また破線 $(- -)$ は $\Delta = 0$，すなわち連続集合 $I$ 上の最大値分布 $P\big(\sup_{t\in I} Y(t) \geq b\big)$ である．Lander & Botstein (1989) は区間マッピング法を提案するとともに，Ornstein-Uhlenbeck 過程の最大値の分布を用いてゲノムワイドな多重性調整を行うことを提唱している．それがこの場合に対応するが，マーカー間隔 10cM ($\Delta = 0.01$) という密な場合であってもその近似はあまり良くない．

点線 $(\cdots)$ は自由度 2 のカイ 2 乗分布の上側確率である．これは多重調整を行わない場合の $p$ 値に対応する．

命題 3.1 の証明は，章末で与える．なおここで与えた近似は，マーカー間距離が常に一定値 $\Delta$ であるというモデルに基づくものである．実際問題への適用の際には，$\Delta$ にマーカー間距離の平均値を代入して用いることになる．Dupuis & Siegmund (1999) は類似の問題設定において，そのような場合であってもよい近似を与えることを数値計算によって確認している．

また命題 2.3 で与えた 2 次元格子上のカイ 2 乗確率場の最大値についても，同様の近似を与えることができる (栗木 (2007)).

## 3.3 ランダム関数の零点の個数の期待値

次に，滑らかなサンプルパスを持つ確率場の最大値の上側裾確率を近似するためのオイラー標数法 (Euler characteristic heuristic) を，添字集合が 1 次元という特殊な場合について説明する．この方法は，信号処理の分野でライスの公式 (Rice's formula) として知られているものと同等である．区間マッピング法のロッドスコアは，漸近的には区分的に滑らかなサンプルパスを持つカイ 2 乗確率過程であったので，オイラー標数法によってその最大値の分布を近似することができる．オイラー標数法の一般論については，栗木・竹村 (2007) を参照のこと．

$Z(t)$, $t \in I \subset \mathbb{R}$ は，実数に値をとるランダムな $C^1$ 関数で，各 $t$ について $(Z(t), \dot{Z}(t))$ $(\dot{Z}(t) = dZ(t)/dt)$ が縮退しない分布を持つとする．$Z(t) = u$ となる $t$ (方程式 $Z(t) - u = 0$ の零点) の個数を

$$N_u = \#\{t \in I \mid Z(t) = u\}$$

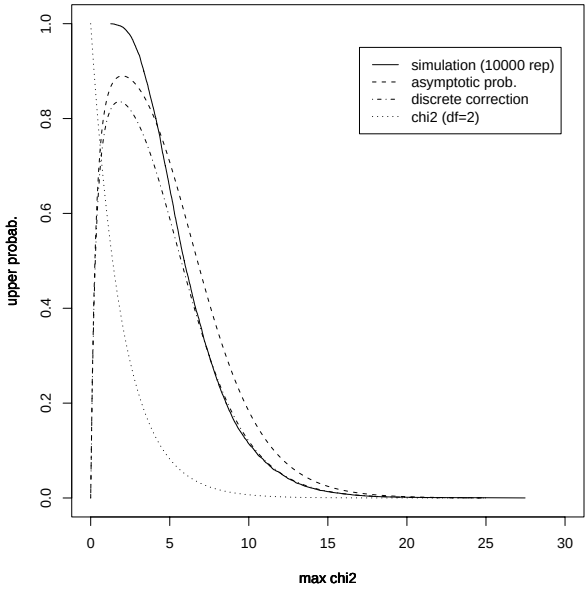


図 4: 格子点上の最大値の分布

( $-\cdot-$：命題 3.1 による近似，—：シミュレーション，$--$：連続近似，$\cdots$：$\chi^2_2$ 分布)

とおく．

ところで，任意の連続関数 $f$ に対して

$$\int |\dot{Z}(t)|\, f(Z(t))\, dt = \int N_u f(u)\, du \tag{3.2}$$

が成り立つことが容易に分かる (図 5)．

(3.2) の両辺について，$Z(\cdot)$ の期待値をとると

$$\begin{aligned} \int E[N_u] f(u) du &= \int E\big[|\dot{Z}(t)|\, f(Z(t))\big]\, dt \\ &= \int\int E\big[|\dot{Z}(t)| \mid Z(t) = u\big] f(u) p_{Z(t)}(u)\, du\, dt \end{aligned}$$

を得る．最右辺において，$p_{Z(t)}$ は $Z(t)$ の周辺密度である．これが任意の連続関数 $f$ で成り立つので，

$$E[N_u] = \int E\big[|\dot{Z}(t)| \mid Z(t) = u\big]\, p_{Z(t)}(u)\, dt \quad \text{a.s.} \tag{3.3}$$

である (Azaïs & Wschebor (2005))．

いま閾値 $u$ が大きいとする．このとき，関数 $Z(t)$ は閾値 $u$ を超えることは稀であり，もし一度超えてもまたすぐに閾値を下回ることが予想される．つまり $N_u$ が 0 または 2 以外

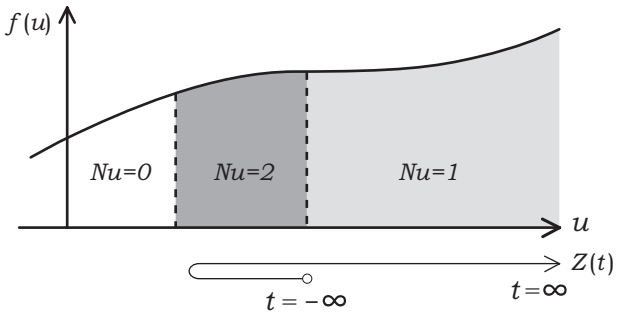


図 5: (3.2) の図による説明

の値をとる確率が小さいことが期待される．そのような状況では

$$\begin{aligned}\frac{1}{2}E[N_u] &\approx P(N_u = 2)\\ &\approx P\big(\exists t \in I,\ Z(t) \geq u\big) \ = \ P\Big(\sup_{t\in I} Z(t) \geq u\Big)\end{aligned}$$

である．オイラー標数近似とは，最大値分布の近似式として

$$P\Big(\sup_{t\in I} Z(t) \geq u\Big) \approx \frac{1}{2}E[N_u] \quad (u \text{ が大きいとき}) \tag{3.4}$$

とおき，(3.4) の右辺として，(3.3) を用いるというものである．ここではこのような直感的な議論にとどめるが，正規確率場，カイ 2 乗確率場に対するオイラー標数法は，正則条件の下で非常に良い近似を与えることが知られている．(栗木・竹村 (2007) の参考文献を参照．)

このオイラー標数法によって，$\mathrm{F}_2$ 集団の区間マッピング法によるロッドスコアの最大値の分布を近似しよう．仮想 QTL の位置を $\gamma$ とする．区間マッピング法で補間されたロッドスコア (尤度比検定統計量) $\mathrm{LRT}_n(\gamma)$, $\gamma \in \Gamma$ の，QTL が存在しないという帰無仮説のもとでの漸近分布は，連続かつ隣り合うマーカーの間では滑らかなパスを持つカイ 2 乗確率過程 $T(\gamma)$, $\gamma \in \Gamma$ であった．全ての隣り合うマーカー間について $E[N_u]/2$ を計算し，それを全て足しあわせることによって最大値分布の近似が可能である (Rebaï, *et al.* (1994))．

**命題 3.2** $R^{UV}(\gamma, \tilde{\gamma})$ を，命題 2.4 で与えた共分散関数 (2.15) とする．

$$A(\gamma) = \frac{\partial}{\partial\tilde{\gamma}} R^{UV}(\gamma,\tilde{\gamma})\Big|_{\tilde{\gamma}=\gamma}, \qquad B(\gamma) = \frac{\partial^2}{\partial\gamma\partial\tilde{\gamma}} R^{UV}(\gamma,\tilde{\gamma})\Big|_{\tilde{\gamma}=\gamma}$$

とおく．$h = (h_1, h_2)^T$ を円周上の一様分布 $\mathrm{Unif}(\mathbb{S}^1)$ とする．ロッドスコアの極限過程 $T(\cdot)$ の，$\Gamma = [d_1, d_m]$ 上最大値のオイラー標数近似は

$$P\Big(\sup_{\gamma\in\Gamma} T(\gamma) \geq u\Big) \approx \frac{1}{\sqrt{2\pi}} u^{1/2} e^{-u/2} \sum_{i=1}^{m-1} \int_{d_i}^{d_{i+1}} E\Big[\sqrt{h^T(B(\gamma) - A(\gamma)^T A(\gamma))h}\Big]\, d\gamma \tag{3.5}$$

で与えられる．

この命題の証明も，本章の章末で与える．その導出は Davies (1987), Theorem A.1 と本質的に同じである．

## 3.4 命題の証明

### 3.4.1 命題 3.1 の証明

ここで与える証明は，Kim & Siegmund (1989) と Siegmund (1992) を組合せたものである．$t$ を固定し，$j \in \mathbb{Z}$ を添字とみなした確率変数列

$$\widetilde{Y}_j = b(Y(t+j\Delta) - y)\Big|_{(Z_1(t),Z_2(t))=(y\cos\theta, y\sin\theta)}$$

を定義する．$Y(t) = \sqrt{Z_1(t)^2 + Z_2(t)^2} = y$ に注意する．$b, y \to \infty$, $b \sim y$, $\Delta \to 0$, $b\sqrt{\Delta} \to c\ (>0)$ の極限を考える．$Y(t+j\Delta)$ を $t$ のまわりでテイラー展開し，また正規分布の条件付分布の公式より，$(Y_j)_{j\in\mathbb{Z}}$ の有限次元の周辺分布は正規分布で，その平均，分散は

$$\begin{aligned} E[\widetilde{Y}_j] &= -\bar{\rho}(\theta)c^2|j|, \\ \mathrm{Cov}(\widetilde{Y}_j, \widetilde{Y}_{\tilde{j}}) &= \bar{\rho}(\theta)c^2(|j| + |\tilde{j}| - |j-\tilde{j}|) \\ &= \begin{cases} 2\bar{\rho}(\theta)c^2 \min(|j|, |\tilde{j}|) & (j \text{ と } \tilde{j} \text{ は同符号}), \\ 0 & (\text{異符号}) \end{cases} \end{aligned}$$

であることを確認できる．これは $X_i\ (i \in \mathbb{Z})$ を独立な正規分布 $N(-\bar{\rho}(\theta)c^2, 2\bar{\rho}(\theta)c^2)$ の列としたときの，両側ランダムウォーク

$$S_j = \begin{cases} X_1 + \cdots + X_j & (j > 0), \\ 0 & (j = 0), \\ X_{-1} + \cdots + X_{-|j|} & (j < 0) \end{cases}$$

の分布に等しい．すなわち

$$(\ldots, \widetilde{Y}_{-1}, \widetilde{Y}_0, \widetilde{Y}_1, \ldots) \ \Rightarrow\ (\ldots, S_{-1}, S_0, S_1, \ldots) \tag{3.6}$$

である．

条件 $(Z_1(j^0\Delta), Z_2(j^0\Delta)) = (y\cos\theta, y\sin\theta)$ を与えた条件付確率測度を $\widetilde{P}$ で表わす．$-x = b(b-y)$ (すなわち $y = b + x/b$) とおく．まず (3.6) より

$$\begin{aligned} \widetilde{P}\Big(\max_{j>j^0} Y(j\Delta) < b\Big) &= \widetilde{P}\Big(\max_{j>j^0} b(Y(j\Delta) - y) < -x\Big) \\ &= \widetilde{P}\Big(\max_{j>0} \widetilde{Y}_j < -x\Big) \ \to\ P\Big(\max_{j>0} S_j < -x\Big) \end{aligned}$$

に注意する．

これからが証明の本体である．$Y(j\Delta) \geq b$ をみたすような点 $j$ の最大値を $j^0$ とおく．事象 $\{\max_{j\in J} Y(j\Delta) \geq b\}$ は $j^0$ の値によって排反に分割される．

$$P\Big(\max_{j\in J} Y(j\Delta) \geq b\Big) = \sum_{j^0\in J} P\Big(\max_{j>j^0} Y(j\Delta) < b,\ Y(j^0\Delta) \geq b\Big). \tag{3.7}$$

さらに事象を $\tan^{-1}\frac{Z_2(j^0\Delta)}{Z_1(j^0\Delta)}$ の値によって排反に分割する．

$$\begin{aligned}(3.7)\text{ 右辺} &= \int_0^{2\pi} \sum_{j^0\in J} P\Big(\max_{j>j^0} Y(j\Delta) < b,\ Y(j^0\Delta) \geq b,\ \tan^{-1}\frac{Z_2(j^0\Delta)}{Z_1(j^0\Delta)} \in d\theta\Big) \\ &= \int_{y\geq b}\int_0^{2\pi} \sum_{j^0\in J} P\Big(\max_{j>j^0} Y(j\Delta) < b,\ Y(j^0\Delta) \in dy,\ \tan^{-1}\frac{Z_2(j^0\Delta)}{Z_1(j^0\Delta)} \in d\theta\Big),\end{aligned}$$

ただし $d\theta = (\theta, \theta + d\theta)$, $dy = (y, y + dy)$ である．

$$Y(j^0\Delta) = y, \quad \tan^{-1}\frac{Z_2(j^0\Delta)}{Z_1(j^0\Delta)} = \theta$$

と値を与えることは $(Z_1(j^0\Delta), Z_2(j^0\Delta)) = (y\cos\theta, y\sin\theta)$ と値を与えることと等価なので，

$$\begin{aligned}&P\Big(\max_{j>j^0} Y(j\Delta) < b,\ Y(j^0\Delta) \in dy,\ \tan^{-1}\frac{Z_2(j^0\Delta)}{Z_1(j^0\Delta)} \in d\theta\Big) \\ &= \widetilde{P}\Big(\max_{j>j^0} Y(j\Delta) < b\Big) \times P\Big(Y(j^0\Delta) \in dy,\ \tan^{-1}\frac{Z_2(j^0\Delta)}{Z_1(j^0\Delta)} \in d\theta\Big) \\ &\to P\Big(\max_{j>0} S_j < -x\Big) \times P(Y(j^0\Delta) \in dy) \times \frac{d\theta}{2\pi}\end{aligned}$$

である．ここで $y = b + x/b$ と変数変換する．$\int_{y>b} = \int_{x>0}$ ならびに

$$P(Y(j^0\Delta) \in dy) = ye^{-y^2/2}dy \sim e^{-b^2/2}e^{-x}dx \quad (b\to\infty)$$

に注意する．

さらに逐次解析で知られた関係式

$$\int_0^\infty e^{-x} P\Big(\max_{j>0} S_j < -x\Big)dx = \bar\rho(\theta)c^2\nu(c\sqrt{2\bar\rho(\theta)})$$

および

$$\Delta^{-1}c^2 \sim b^2, \quad \Delta\sum_{j^0\in J} \sim \int_I dt \quad (\Delta\to 0)$$

を組合せると (3.1) を得る． ∎

### 3.4.2 命題 3.2 の証明

仮想 QTL の位置を $\gamma \in [d_i, d_{i+1}]$ とする．命題 2.4 の共分散関数 $R^{UV}(\gamma,\tilde\gamma)$ (2.15) は引数について十分に滑らかである．とくに

$$\Big(\frac{\partial}{\partial\gamma}\Big)^3\Big(\frac{\partial}{\partial\tilde\gamma}\Big)^3 R^{UV}(\gamma,\tilde\gamma)$$

が $\gamma = \tilde{\gamma}$ の近傍で存在し有界なので，$U(\cdot)$, $V(\cdot)$ は確率 1 で連続微分可能で，また微分過程 $\dot{U}(\cdot)$, $\dot{V}(\cdot)$ も連続な正規過程となる (Adler & Taylor (2007), Theorem 1.4.2).

微分過程 $\dot{U}(\cdot)$, $\dot{V}(\cdot)$ も平均 0 で，その共分散関数は

$$\mathrm{Cov}\begin{pmatrix} U(\gamma) \\ V(\gamma) \\ \dot{U}(\gamma) \\ \dot{V}(\gamma) \end{pmatrix} = \begin{pmatrix} I_2 & A(\gamma) \\ A(\gamma)^T & B(\gamma) \end{pmatrix}$$

である．ここで

$$\begin{aligned} 0 &= \frac{d}{d\gamma}\mathrm{Cov}(U(\gamma), U(\gamma)) = 2\mathrm{Cov}(\dot{U}(\gamma), U(\gamma)), \\ 0 &= \frac{d}{d\gamma}\mathrm{Cov}(U(\gamma), V(\gamma)) = \mathrm{Cov}(\dot{U}(\gamma), V(\gamma)) + \mathrm{Cov}(U(\gamma), \dot{V}(\gamma)) \end{aligned}$$

であるので $A(\gamma)^T = -A(\gamma)$ (交代行列) である．

以下では引数の $\gamma$ は省略する．$(\dot{U}, \dot{V})$ の，$(U, V)$ を与えた条件付き分布は

$$\begin{pmatrix} \dot{U} \\ \dot{V} \end{pmatrix}\Big|_{(U,V)} \sim N\Big(A^T\begin{pmatrix} U \\ V \end{pmatrix}, B - A^T A\Big)$$

であり，また

$$T = (U, V)\begin{pmatrix} U \\ V \end{pmatrix}, \qquad \dot{T} = 2(U, V)\begin{pmatrix} \dot{U} \\ \dot{V} \end{pmatrix}$$

より，

$$\begin{aligned} \dot{T}\Big|_{(U,V)} &\sim N\Big(2(U,V)A^T\begin{pmatrix} U \\ V \end{pmatrix}, 4(U,V)(B - A^T A)\begin{pmatrix} U \\ V \end{pmatrix}\Big) \\ &= N\Big(0, 4(U,V)(B - A^T A)\begin{pmatrix} U \\ V \end{pmatrix}\Big) \end{aligned}$$

である．したがって

$$\frac{\dot{T}}{\sqrt{4(U,V)(B - A^T A)\binom{U}{V}}}\Big|_{(U,V)} = \xi \sim N(0, 1)$$

となり，$\xi$ の条件付き分布は条件 $(U, V)$ によらないことが分かるので，

$$\dot{T} = \xi\sqrt{4(U,V)(B - A^T A)\begin{pmatrix} U \\ V \end{pmatrix}} \qquad (\xi \sim N(0,1) \text{ は } (U,V) \text{ と独立})$$

と表わすことができる．

さらに

$$h = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \frac{1}{\sqrt{T}}\begin{pmatrix} U \\ V \end{pmatrix}$$

とおくと，$\dot{T} = 2\xi\sqrt{T}\sqrt{h^T(B - A^T A)h}$ であり，$h$ は $\xi, T$ と独立に，円周上の一様分布 $\mathrm{Unif}(\mathbb{S}^1)$ に従う．

以上より

$$\begin{aligned} E\big[|\dot{T}| \mid T=u\big]\, p_T(u) &= E\big[|\dot{T}|\, 1_{\{T\in(u,u+du)\}}\big]/du \\ &= 2E[|\xi|]\, E\big[\sqrt{T}\, 1_{\{T\in(u,u+du)\}}\big]/du \times E\Big[\sqrt{h^T(B-A^TA)h}\Big] \\ &= 2\sqrt{\frac{2}{\pi}}u^{1/2}\frac{1}{2}e^{-u/2}E\Big[\sqrt{h^T(B-A^TA)h}\Big]. \end{aligned}$$

これを $\gamma$ で積分して,

$$\frac{1}{2}E[N_u] = \frac{1}{\sqrt{2\pi}}u^{1/2}e^{-u/2}\int_{d_i}^{d_{i+1}} E\Big[\sqrt{h^T(B(\gamma)-A(\gamma)^TA(\gamma))h}\Big]\, d\gamma,$$

ただし右辺の期待値は $h \sim \mathrm{Unif}(\mathbb{S}^1)$ についてとる.

これを全ての隣り合うマーカーについて足しあわせることによって (3.5) が得られる. ∎

## 謝辞

1 章で紹介したマウスの QTL 解析例 (図 1, 3) は，国立遺伝学研究所の城石研究室 (前野哲輝氏，城石俊彦氏) によるものです．また同研究所の春島嘉章氏，倉田のり氏からは，本稿の内容について有益なコメントをいただきました．本稿は，数理科学総合セミナー II (2006 年度前期，東京大学数理科学研究科）の講義録に加筆したものです．同大学の吉田朋広氏には，注 2.5 をご指摘いただいたきました．これらの方々に感謝いたします.

## 参考文献

[1] Adler, R. J. and Taylor, J. E. (2007). *Random Fields and their Geometry*, Springer.

[2] Azaïs, J-M. and Wschebor, M. (2005). On the distribution of the maximum of a Gaussian field with $d$ parameters, *Ann. Appl. Probab.*, **15** (1A), 254–278.

[3] Broman, K. W., Wu, H., Sen, Ś. and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses, *Bioinformatics*, **19** (7), 889–890.

[4] Churchill, G. A. and Doerge, R. W. (1994). Empirical threshold values for quantitative triat mapping, *Genetics*, **138** (3), 963–971.

[5] Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternatives, *Biometrika*, **74** (1), 33–43.

[6] Dupuis, J. and Siegmund, D. (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers, *Genetics*, **151** (1), 373–386.

[7] 福水健次, 栗木哲, 竹内啓, 赤平昌文 (2004). 「特異モデルの統計学 — 未解決問題への新しい視点」, 統計科学のフロンティア 7, 岩波書店.

[8] Haldane, J. B. S. (1919). The combination of linkage values and the calculation of distances between the loci of linked factors, *J. Genetics*, **8**, 299-309.

[9] Haley, C. S. and Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers, *Heredity*, **69**, 315–324.

[10] 春島嘉章, 栗木哲, 水多陽子, 藤澤洋徳, 倉田のり (2006). 配偶体内または接合体内の異なる遺伝子座間の相互作用による生殖的隔離障壁の検出, 育種学研究, **8**, 別冊 2 号, 157.

[11] Harushima, Y., Nakagahra, M., Yano, M., Sasaki, T. and Kurata, N. (2001). A genome-wide survey of reproductive barriers in an intraspecific hybrid, *Genetics*, **159** (2), 883–892.

[12] Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures*, Wiley.

[13] Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*, Springer.

[14] 石川明 (2006). 動物モデルによる多因子性疾患の QTL 解析：その基礎的理論と解析方法 (改訂版), NAGOYA Repository, `http://hdl.handle.net/2237/6779`

[15] Karlin, S. and Liberman, U. (1983). Measuring interference in the chiasma renewal formation process, *Adv. Appl. Probab.*, **15** (3), 471–487.

[16] Kim, H.-J. and Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression, *Biometrika*, **76** (3), 409–423.

[17] 栗木哲 (2007). 直積型の相関構造を持つカイ 2 乗確率場の最大値の分布, 日本数学会 2007 年度年会 統計数学分科会講演アブストラクト, 89–90.

[18] 栗木哲, 竹村彰通 (2008). チューブの体積と正規確率場の最大値の分布, 数学, **60** (2), 134–155.

[19] Lander, E. S. and Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps, *Genetics*, **121** (1), 185–199.

[20] Lander, E. S. and Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results, *Nature Genetics*, **11** (3), 241–247.

[21] Manichaikul, A., Dupuis, J., Sen, S. and Broman, K. W. (2006). Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus, *Genetics*, **174** (1), 481–489.

[22] Piterbarg, V. I. (1996). *Asymptotic Methods in the Theory of Gaussian Processes and Fields*, Translations of Mathematical Monographs, 148, AMS.

[23] Rebaï, A., Goffinet, B. and Mangin, B. (1994). Approximate thresholds of interval mapping tests for QTL detection, *Genetics*, **138** (1), 235–240.

[24] Sen, S. and Churchill, G. A. (2001). A statistical framework for quantitative trait mapping, *Genetics*, **159** (1), 371–387.

[25] Siegmund, D. (1985). *Sequential Analysis*, Springer.

[26] Siegmund, D. O. (1992). Tail approximations for maxima of random fields, in *Probability Theory*, L. H. Y. Chen, K. P. Choi, K. Hu and J-H. Lou (eds.), Walter de Gruyter, 147–158.

[27] 鵜飼保雄 (2000). 「ゲノムレベルの遺伝解析 — MAP と QTL」, 東京大学出版会.

[28] Wu, C. F. J. (1983). On the convergence properties of the EM algorithm, *Ann. Statist.*, **11** (1), 95–103.

[29] Wu, R., Ma, C.-X., and Casella, G. (2007). *Statistical Genetics of Quantitative Traits*, Springer.

[30] Yoshida, N. (2006). Polynomial type large deviation inequalities and convergence of statistical random fields, ISM Research Memorandum No. 1021, to appear in *Ann. Inst. Statist. Math.*