遺伝子頻度を推定するための統計的手法について

野　村　哲　郎

1．はじめに

集団の遺伝的構成を把握する上で，遺伝子頻度は最も基礎的なパラメータである。遺伝子頻度は，集団中の全個体について遺伝子型が特定できれば容易に計算できるが，全数調査のできない野外集団を対象とする場合には，標本（サンプル）から遺伝子頻度を推定するのが通例である。その際，対立遺伝子間の優劣関係により表現型から遺伝子型を特定できないときには，遺伝子頻度の推定が複雑になる。筆者は，昆虫集団の遺伝的構成について調査を行っているが，結果をまとめる上で必要とされる遺伝子頻度の推定方法ならびにその統計的性質についてまとめておきたいと思う。以下は，主に安田（1968），Li（1976），Spiess（1989）および Weir（1996）を参考にしてまとめたものである。

2．標本の確率分布

集団遺伝学におけるデータ（標本）の多くは，カテゴリー（遺伝子型や表現型）ごとに個体数をカウントしたものからなる。このようなデータの統計的性質を表現するために，母集団（以下，とくに必要なとき以外は，集団と呼ぶ）の個体は，k 個のカテゴリーのいずれかに属するものとする。ある個体が i 番目のカテゴリーに属する確率を Q_i とする。いま，この集団から大きさ n の標本を抽出するものとしよう。標本において，１～k 番目のカテゴリーに n_1，n_2，…，n_k 個体が含まれる確率は，多項分布の確率関数

$$\Pr\left(n_1,n_2,\cdots,n_k\right)=\frac{n!}{\prod_{i=1}^{k}n_i}\prod_{i=1}^{k}Q_i^{n_i} \qquad (2-1)$$

によって表される。カテゴリーが２つだけのときは，集団において２つのカテゴリーに個体が属する確率を Q および $1-Q$，大きさ n の標本においてそれぞれのカテゴリーに属する個体数を a および $n-a$ とすると，式（2―1）は二項分布の確率関数

$$\Pr\left(a,n-a\right)=\frac{n!}{a!\left(n-a\right)!}Q^a\left(1-Q\right)^{n-a} \qquad (2-2)$$

となる。

つぎに，これらの分布の平均と分散について考えよう。大きさ n の標本中において i 番目のカテゴリーに属する個体数の平均（期待値）は，このカテゴリー以外に属する個体をひとまとめにすることで，二項分布の平均として扱うことができる。すなわち，式（2―2）より

$$E(n_i)=\sum_{r=0}^{n} r\cdot \Pr(n_i=r)$$
$$=\sum_{r=0}^{n} r\frac{n!}{r!(n-r)!}Q_i^r(1-Q_i)^{n-r}=nQ_i.$$

両辺を n で割ることで，i 番目のカテゴリーに属する個体が標本に占める割合（観察頻度）$\tilde{Q}_i=n_i/n$ の期待値は

$$E(\tilde{Q}_i)=Q_i \qquad (2-3)$$

となる。すなわち，$\tilde{Q}_i$ は Q_i の不偏推定量である。

観察度数 n_i の分散は，

$$Var(n_i)=\sum_{r=0}^{n}\left[r-E(n_i)\right]^2\Pr(n_i=r)$$
$$=nQ_i(1-Q_i) \qquad (2-4)$$

観察頻度 $\tilde{Q}_i$ の分散は，

$$Var(\tilde{Q}_i)=\frac{1}{n^2}Var(n_i)=\frac{1}{n}Q_i(1-Q_i) \qquad (2-5)$$

である。標本中で2つのカテゴリーに含まれる観察度数の積 n_in_j の期待値は，

$$E(n_in_j)=\sum_{r=0}^{n}\sum_{s=0}^{n-r} rs\Pr(n_i=r,n_j=s)$$
$$=\sum_{r=0}^{n}\sum_{s=0}^{n-r} rs\frac{n!}{r!s!(n-r-s)!}Q_i^rQ_j^s(1-Q_i-Q_j)^{n-r-s}$$
$$=n(n-1)Q_iQ_j$$

と表される。両辺を n^2で割ることで

$$E(\tilde{Q}_i\tilde{Q}_j)=\frac{n-1}{n}Q_iQ_j$$

が得られる。これらより，n_i と n_j および $\tilde{Q}_i$ と $\tilde{Q}_j$ の共分散は

$$Cov(n_i,n_j)=E(n_in_j)-E(n_i)E(n_j)=-Q_iQ_j$$
$$Cov(\tilde{Q}_i,\tilde{Q}_j)=-\frac{1}{n}Q_iQ_j \qquad (2-6)$$

となる。標本数（n）が一定の場合，一方のカテゴリーに属する個体が多いと，もう一方のカテゴリーに属する個体は少なくなるため，いずれの共分散も負の値をとる。

3．カウント法

この方法は，単純に標本中の問題とする遺伝子の数をカウントして得られた頻度を集団の遺伝子頻度の推定値とする方法である。

⑴対立遺伝子間に優劣関係のない場合

対立遺伝子間に優劣関係がなく，各表現型がそれぞれ1個の遺伝子型に対応する場合について考える。大きさ n の標本中にホモ接合体 A_uA_u が n_{uu} 個体，ヘテロ接合体 A_uA_v が n_{uv} 個体が含まれるなら，A_u 遺伝子の標本中の総数 n_u は

$$n_u = 2n_{uu} + \sum_{u \neq v} n_{uv}$$

である。標本中の総遺伝子数は $2n$ であるから，

$$\begin{aligned} \tilde{p}_u &= \frac{n_{uu}}{n} + \frac{1}{2}\sum_{u \neq v}\frac{n_{uv}}{n} \\ &= \tilde{P}_{uu} + \frac{1}{2}\sum_{u \neq v}\tilde{P}_{uv}. \end{aligned} \qquad (3-1)$$

ここで，$\tilde{P}_{uu}$ および $\tilde{P}_{uv}$ は標本における遺伝子型 A_uA_u および A_uA_v の観察頻度である。式（2―3）より

$$E\left[\tilde{P}_{uu}\right] = P_{uu}$$
$$E\left[\tilde{P}_{uv}\right] = P_{uv}$$

であるから，

$$E\left[\tilde{p}_u\right] = P_{uu} + \frac{1}{2}\sum_{u \neq v} P_{uv} = p_u.$$

したがって，$\tilde{p}_u$ は p_u の不偏推定量である。後で示すように，$\tilde{p}_u$ は p_u の最尤推定値でもある。

n_u の分散は式（2―4）と（2―6）より

$$\begin{aligned} Var\left(n_u\right) &= Var\left(2n_{uu}\right) + \sum_{u \neq v} Var\left(n_{uv}\right) + 2\sum_{u \neq v} Cov\left(2n_{uu}, n_{uv}\right) \\ &= 2n\left(p_u + P_{uu} - 2p_u^2\right) \end{aligned}$$

となる。したがって，$\tilde{p}_u$ の分散は

$$Var\left(\tilde{p}_u\right) = \frac{1}{2n}\left(p_u + P_{uu} - 2p_u^2\right). \qquad (3-2)$$

式（3―2）は，母集団のパラメータ p_u と p_{uu} を含むので，これらを標本からの推定値 $\tilde{p}_u$ と $\tilde{p}_{uu}$ で置き換えることで，$\tilde{p}_u$ の分散の推定値

$$Var\left(\tilde{p}_u\right) = \frac{1}{2n}\left(\tilde{p}_u + \tilde{P}_{uu} - 2\tilde{p}_u^2\right) \qquad (3-3)$$

が得られる。$\tilde{p}_u$ の分布を正規分布で近似すると，$\tilde{p}_u$ の信頼度$100(1-\alpha)$％の信頼区間が

$$\tilde{p}_u \pm Z_{\alpha/2}\sqrt{Var\left(\tilde{p}_u\right)}$$

として設定できる。ここで，$Z_{\alpha/2}$は標準正規分布の上側点 $\alpha/2$ である。

式（3―2）が二項分布を仮定したときの標本分散（式（2―5））と同じ形にならないことには注意を要する。式（3―2）が

$$Var\left(\tilde{p}_u\right)=\frac{1}{2n}p_u\left(1-p_u\right)$$

となるためには，

$$P_{uu}=p_u^2$$
$$P_{uv}=2p_up_v$$

すなわち集団がハーディー・ワインベルグ平衡に達していることが必要である。

数値例 1

表 1 は，香港の中国人1029人について MN 血液型を調べた結果である（Li, 1976）。

表 1．香港の中国人 1029 人の MN 血液型の遺伝子型

遺伝子型	観察度数
MM	342
MN	500
NN	187
	計 1029

M 遺伝子の頻度は，式（3―1）より

$$\tilde{p}_M=\frac{342}{1029}+\frac{1}{2}\frac{500}{1029}=0.5753$$

また $\tilde{p}_M$ の分散は，式（3―3）より

$$Var\left(\tilde{p}_M\right)=\frac{1}{2\times1029}\left(0.5753+\frac{342}{1029}-2\times0.5753^2\right)$$
$$=0.0001194$$

となる。標準正規分布表より $Z_{0.025}=1.96$ が得られるので，95％信頼区間は，

$$\tilde{p}_M\pm1.96\sqrt{Var\left(\tilde{p}_M\right)}=0.5753\pm0.0214$$

として設定できる。

⑵対立遺伝子間に優劣関係のある場合

対立遺伝子間に優劣関係があり，ヘテロ接合体に表現型でホモ接合体と識別できないものが含まれる場合には，上で示したカウント法を直接に利用することはできない。そこで，ハーディー・ワインベルグ平衡を仮定した反復解法が開発されている（Yasuda and Kimura, 1968; 安田，1968）。

以下，ヒトの ABO 血液型を例にして説明する。遺伝子 A, B, O の頻度をそれぞれ p，q，r とし，表現型―遺伝子型の関係，観察度数およびハーディー・ワインベルグ平衡の下での期待頻度を表 2 に示す。

表２．ＡＢＯ血液型の表現型－遺伝子型の関係、観察度数および期待頻度

表現型	遺伝子型	観察度数	期待頻度
Ａ	ＡＡ、ＡＯ	n_A	p^2+2pr
Ｂ	ＢＢ、ＢＯ	n_B	q^2+2qr
ＡＢ	ＡＢ	n_{AB}	$2pq$
Ｏ	ＯＯ	n_O	r^2
計		n	1

A 遺伝子は AB 型の人に 1 個，A 型では遺伝子型 AA の人に 2 個，遺伝子型 AO の人に 1 個あるが，遺伝子型 AA と AO は区別できない。そこで，A 型で遺伝子型 AA である割合を h_A とすると

$$h_A=\frac{p^2}{p^2+2pr}=\frac{p}{p+2r}$$

これを用いれば，式（３―１）より

$$\begin{aligned}\hat{p}&=\frac{1}{2n}\left\{n_{AB}+2n_Ah_A+n_A\left(1-h_A\right)\right\}\\&=\frac{1}{2n}\left\{n_{AB}+n_A\left(1+h_A\right)\right\}\end{aligned}\qquad(３－４)$$

同様に

$$h_B=\frac{q}{q+2r}$$

とすれば

$$\hat{q}=\frac{1}{2n}\left\{n_{AB}+n_B\left(1+h_B\right)\right\}\qquad(３－５)$$

r の推定値は，$\hat{r}=1-\hat{p}-\hat{q}$ として得られる。計算では，p と q を適当に定め（p_m，q_m とする），h_A と h_B を計算して上の式に代入し，左辺（p_{m+1}，q_{m+1}とする）を求める。さらに，p_{m+1}と q_{m+1}を右辺に代入して，左辺（p_{m+2}と q_{m+2}）を求める。この計算を解が収束するまで繰り返す。後で示すように，この計算手続きは EM アルゴリズムそのものであり，得られる解は最尤推定値である。

４．平方根法

２倍体生物においては，個体の表現型頻度が遺伝子頻度の 2 次関数になることを利用した方法である。この方法は，計算量が少ないのでコンピュータが普及していなかった時代にはよく用いられたが，現在ではほとんど利用されることはない。この方法からは，特殊な場合を除いて最尤推定値は得られないが，最尤推定値を得るための反復解法の初期値を得るためには役立つ方法である。

平方根法は，ヒトの ABO 血液型の遺伝子頻度推定のために開発されたものが多いので，以下では，主に ABO 血液型への適用を説明する。

⑴ Bernstein の方法

表２より，

$$n_O = nr^2$$
$$n_A + n_O = n(p+r)^2$$
$$n_B + n_O = n(q+r)^2$$

これらより，p，q，r の推定値が，

$$\left.\begin{aligned}\hat{p} &= 1-\sqrt{(n_B+n_O)/n}\\ \hat{q} &= 1-\sqrt{(n_A+n_O)/n}\\ \hat{r} &= \sqrt{n_O/n}\end{aligned}\right\} \qquad (4-1)$$

として得られる。この方法から推定される遺伝子頻度は，一般には $\hat{p}+\hat{q}+\hat{r}$ とが１にはならないので，Bernstein は

$$D = 1-\hat{p}-\hat{q}-\hat{r}$$

とおいて，つぎのような補正式を示した。

$$\left.\begin{aligned}\hat{p}' &= \hat{p}(1+D/2)\\ \hat{q}' &= \hat{q}(1+D/2)\\ \hat{r}' &= (\hat{r}+D/2)(1+D/2)\end{aligned}\right\} \qquad (4-2)$$

補正後の遺伝子頻度の和は，$\hat{p}'+\hat{q}'+\hat{r}'=1-D^2/4$ となり，１からの偏差は補正前より小さくなる。この方法では，AB 型の観察度数が考慮されていないので，得られる推定値は明らかに最尤推定値ではない。

数値例 2

以下のデータは，Li（1976）より引用したものである。

表３．雲南省における中国人 6000 人のＡＢＯ血液型

血液型	観察度数
A	$n_A = 1920$
B	$n_B = 1627$
AB	$n_{AB} = 607$
O	$n_O = 1846$
	$n = 6000$

式（4―1）より，

$$\hat{p} = 1-\sqrt{(1627+1846)/6000} = 0.2392$$
$$\hat{q} = 1-\sqrt{(1920+1846)/6000} = 0.2077$$
$$\hat{r} = \sqrt{1846/6000} = 0.5547$$

さらに，$D=1-\hat{p}-\hat{q}-\hat{r}=-0.0016$を用いて，式（4―2）による補正を行えば，

$$\hat{p}'=\hat{p}\left(1+D/2\right)=0.2390$$
$$\hat{q}'=\hat{q}\left(1+D/2\right)=0.2075$$
$$\hat{r}'=\left(\hat{r}+D/2\right)\left(1+D/2\right)=0.5535$$

となる。

⑵ Wiener の方法

Bernstein の方法と同様の考えに基づいて

$$\left.\begin{aligned}\hat{p}&=\sqrt{\left(n_A+n_O\right)/n}-\sqrt{n_O/n}\\ \hat{q}&=\sqrt{\left(n_B+n_O\right)/n}-\sqrt{n_O/n}\\ \hat{r}&=\sqrt{n_O/n}\end{aligned}\right\}\qquad(4-3)$$

として推定する方法である。$W=1-\hat{p}-\hat{q}-\hat{r}=-D$ とおけば，補正後の Bernstein の推定値と同じ値が

$$\hat{p}'=\left(1-W/2\right)\left(\hat{p}+W\right)$$
$$\hat{q}'=\left(1-W/2\right)\left(\hat{q}+W\right)$$
$$\hat{r}'=\left(1-W/2\right)\left(\hat{r}-W/2\right)$$

によって得られる。この方法も AB 型の観察度数が考慮されていないので，Bernstein の方法と同様の問題を持つ。

数値例 3

表 3 の観察度数に，式（4―3）を適用すれば，

$$\hat{p}=\sqrt{\left(1920+1846\right)/6000}-\sqrt{1846/6000}=0.2376$$
$$\hat{q}=\sqrt{\left(1627+1846\right)/6000}-\sqrt{1846/6000}=0.2061$$
$$\hat{r}=\sqrt{1846/6000}=0.5547$$

を得る。また，$W=1-\hat{p}-\hat{q}-\hat{r}=0.0016=-D$ となることが確認できる。

⑶ Yasuda の方法

$$k_A=n_O/\left(n_A+n_O\right)$$
$$k_B=n_O/\left(n_B+n_O\right)$$

とおくと，式（3―4）と（3―5）の h_A および h_B は，

$$h_A=\left(1-\sqrt{k_A}\right)^2\Big/\left(1-k_A\right)$$
$$h_B=\left(1-\sqrt{k_B}\right)^2\Big/\left(1-k_B\right)$$

と書ける。したがって，

$$n_Ah_A=n_A+2n_O-2\sqrt{n_O\left(n_A+n_O\right)}$$
$$n_Bh_B=n_B+2n_O-2\sqrt{n_O\left(n_B+n_O\right)}$$

これらを式（3―4）と（3―5）に代入すると

$$\left.\begin{aligned}\hat{p}&=\frac{1}{n}\left[\frac{1}{2}n_{AB}+n_A+n_O-\sqrt{n_O\left(n_A+n_O\right)}\right]\\ \hat{q}&=\frac{1}{n}\left[\frac{1}{2}n_{AB}+n_B+n_O-\sqrt{n_O\left(n_B+n_O\right)}\right]\\ \hat{r}&=1-\hat{p}-\hat{q}\end{aligned}\right\}\qquad(4-4)$$

として推定値が得られる（Yasuda, 1984）。この推定値は，最尤推定値ではないが，AB 型の観察度数も考慮されている点で，先の 2 つの方法よりも優れている。

数値例 4

表 3 の観察度数に式（4―4）を適用すれば，推定値として

$$\hat{p}=\frac{1}{n}\left[\frac{1}{2}n_{AB}+n_A+n_O-\sqrt{n_O\left(n_A+n_O\right)}\right]=0.2388$$
$$\hat{q}=\frac{1}{n}\left[\frac{1}{2}n_{AB}+n_B+n_O-\sqrt{n_O\left(n_B+n_O\right)}\right]=0.2074$$
$$\hat{r}=1-\hat{p}-\hat{q}=0.5538$$

が得られる。

⑷対立遺伝子間に特殊な優劣関係のある場合

m 個の対立遺伝子があり，それらの間に $A_1>A_2>\cdots>A_m$ なる優劣関係のある場合を考える。対立遺伝子 A_i の頻度を p_i とし，ハーディー・ワインベルグ平衡を仮定すると，各表現型の期待頻度は表 4 に示すように与えられる。

表４．表現型－遺伝子型の関係、観察度数および期待頻度

表現型	遺伝子型	観察度数	期待頻度
P_1	$A_1A_1, A_1A_2, \ldots, A_1A_m$	n_1	$1-q_2^2$
P_2	$A_2A_2, A_2A_3, \ldots, A_2A_m$	n_2	$q_2^2-q_3^2$
⋮	⋮	⋮	⋮
P_m	A_mA_m	n_m	q_m^2
計		n	1

$q_i = p_i + p_{i+1} + \cdots + p_m$

表現型の期待頻度は，つぎのようにして得られる。$(p_1+p_2+\cdots+p_m)^2$ を展開して，

$$
\begin{aligned}
&p_1^2 + 2p_1p_2 + 2p_1p_3 + \cdots + 2p_1p_m && S_1\\
&+p_2^2 + 2p_2p_3 + 2p_2p_4 + \cdots + 2p_2p_m && S_2\\
&\qquad\vdots\\
&+p_{m-1}^2 + 2p_{m-1}p_m && S_{m-1}\\
&+p_m^2 && S_m
\end{aligned}
$$

と書いて，ハーディー・ワインベルグの法則を適用すると第 i 行の和 S_i が表現型 P_i の期待頻度となる。一方，第 i 行から第 m 行までの和は，

$$\sum_{j=i}^{m} S_j = \left(p_i + p_{i+1} + \cdots + p_m\right)^2$$

と書けるので，

$$
\begin{aligned}
S_i &= \sum_{j=i}^{m} S_j - \sum_{j=i+1}^{m} S_j\\
&= \left(p_i + p_{i+1} + \cdots + p_m\right)^2 - \left(p_{i+1} + p_{i+2} + \cdots + p_m\right)^2\\
&= q_i^2 - q_{i+1}^2
\end{aligned}
$$

が得られる。ここで，$q_i = p_i + p_{i+1} + \cdots + p_m$ である。

$$N_i = n_i + n_{i+1} + \cdots + n_m$$

とすると，表現型の期待頻度から遺伝子頻度が

$$
\left.
\begin{aligned}
&\hat{p}_1 = 1 - \sqrt{N_2/n}\\
&\hat{p}_i = \sqrt{N_i/n} - \sqrt{N_{i+1}/n} \qquad i = 1, 2, \cdots, m\\
&\hat{p}_m = \sqrt{n_m/n}
\end{aligned}
\right\}
\qquad (4-5)
$$

として推定できる。後で示すように，この推定値は最尤推定値である。

数値例 5

ナミテントウの斑紋は，二紋型，四紋型，斑型および紅型の４つに大別できる（写真１）。

写真1. ナミテントウの斑紋型

これらの斑紋は，４つの対立遺伝子 h^c，h^{SP}，h^a および h によって決定されている。４つの対立遺伝子間には $h^c > h^{SP} > h^a > h$ なる優劣関係があり，４つの斑紋型（表現型）と遺伝子型の関係は表５に示すようになる。表５には，筆者が2002～2004年に京都産業大学構内およびその周辺で採集した636個体を斑紋型別に分類した個体数も合わせて示してある。

表５．ナミテントウの斑紋型ー遺伝子型の関係および京都産業大学周辺での観察度数

斑紋型	遺伝子型	観察度数
二紋型	h^ch^c, h^ch^{SP}, h^ch^a, h^ch	431
四紋型	$h^{SP}h^{SP}$, $h^{SP}h^a$, $h^{SP}h$	91
斑型	h^ah^a, h^ah	29
紅型	hh	85
計		636

h^c，h^{SP}，h^a および h の頻度をそれぞれ p，q，r，s とすると，式（４―５）より

$$\hat{p} = 1 - \sqrt{\frac{91+29+85}{636}} = 0.43226147$$

$$\hat{q} = \sqrt{\frac{91+29+85}{636}} - \sqrt{\frac{29+85}{636}} = 0.144436484$$

$$\hat{r} = \sqrt{\frac{29+85}{636}} - \sqrt{\frac{85}{636}} = 0.05779497$$

$$\hat{s} = \sqrt{\frac{85}{636}} = 0.36557872$$

が最尤推定値として得られる。

５．最尤法

⑴最尤法について

標本から母集団のパラメータ（母数）を推定する際，「得られた標本は母集団を代表するもの」と信じて分析を行うのが通常である。この信念を究極まで推し進めた推定法が Fisher によって考案さ

れた最尤法である。最尤法は，得られたデータに対して「もっともらしさ」が最大になるようなパラメータを母集団のパラメータの推定値とする，という指針に基づく。Fisher は，この「もっともらしさ」を，母集団に仮定した分布の下で，手にした構成を持つデータ（標本）が得られる確率（尤度）として定義した。

いま，優劣関係のない 3 つの対立遺伝子 A_1，A_2および A_3が，それぞれ頻度 p_1，p_2 および p_3（$=1-p_1-p_2$）で分離しているハーディー・ワインベルグ平衡に達している集団を考える。大きさ n の標本を抽出したところ，各遺伝子型の観察度数が

A_1A_1	A_1A_2	A_1A_3	A_2A_2	A_2A_3	A_3A_3
n_{11}	n_{12}	n_{13}	n_{22}	n_{23}	n_{33}

であったとしよう。多項分布の確率関数（式（2―1））より，このような標本が得られる確率は

$$\begin{aligned}&\Pr\left(n_{11},n_{12},n_{13},n_{22},n_{23},n_{33}\right)\\&=C\left(p_1^2\right)^{n_{11}}\left(2p_1p_2\right)^{n_{12}}\left[2p_1\left(1-p_1-p_2\right)\right]^{n_{13}}\\&\quad\times\left(p_2^2\right)^{n_{22}}\left[2p_2\left(1-p_1-p_2\right)\right]^{n_{23}}\left[\left(1-p_1-p_2\right)^2\right]^{n_{33}}\end{aligned}$$

である。ここで，$C=n!\Big/\prod_{i\le j}^{3}n_{ij}!$である。これを母数（$p_1$，$p_2$）の関数 $L(p_1,\ p_2)$ とみなしたものを尤度関数という。得られた標本を最も出現しやすくさせる母数，すなわち $L(p_1,\ p_2)$ を最大にする p_1 と p_2 を母数の推定値としたものが最尤推定値である。最尤推定値は，推定量として多くの望ましい性質を持つ（くわしくは，稲垣（1990）などを参照）。

尤度関数の最大化には，一般にその対数をとった対数尤度関数を最大化すればよい。いまの場合，対数尤度関数は

$$\begin{aligned}\ln L\left(p_1,p_2\right)=C&+n_{11}\log p_1^2+n_{12}\log\left(2p_1p_2\right)\\&+n_{13}\log\left[2p_1\left(1-p_1-p_2\right)\right]+n_{22}\log p_2^2\\&+n_{23}\log\left[2p_2\left(1-p_1-p_2\right)\right]+n_{33}\log\left(1-p_1-p_2\right)^2\end{aligned}$$

である。これを p_1 と p_2 で偏微分して得られる偏導関数をゼロとおいた 2 つの方程式（尤度方程式）

$$\begin{aligned}\frac{\partial\ln L}{\partial p_1}&=\frac{2n_{11}+n_{12}+n_{13}}{p_1}-\frac{2n_{33}+n_{13}+n_{23}}{1-p_1-p_2}=0\\\frac{\partial\ln L}{\partial p_2}&=\frac{2n_{22}+n_{12}+n_{23}}{p_2}-\frac{2n_{33}+n_{13}+n_{23}}{1-p_1-p_2}=0\end{aligned}$$

の解として，p_1 および p_2 の最尤推定値が得られる。すなわち，最尤推定値は

$$\hat{p}_1 = \frac{2n_{11} + n_{12} + n_{13}}{n}$$
$$\hat{p}_2 = \frac{2n_{22} + n_{12} + n_{23}}{n}$$

となり，この場合にはカウント法による推定値（式（3―1））と一致する。なお，$\hat{p}_3 = 1 - \hat{p}_2 - \hat{p}_3$である。

対数尤度関数の2階偏導関数に－1を乗じたものはFisherの情報量と呼ばれる。いまの場合，

$$-\frac{\partial^2 \ln L}{\partial p_1^2} = \frac{2n_{11} + n_{12} + n_{13}}{p_1^2} + \frac{2n_{33} + n_{13} + n_{23}}{(1 - p_1 - p_2)^2}$$
$$-\frac{\partial^2 \ln L}{\partial p_2^2} = \frac{2n_{22} + n_{12} + n_{23}}{p_2^2} + \frac{2n_{33} + n_{13} + n_{23}}{(1 - p_1 - p_2)^2}$$
$$-\frac{\partial^2 \ln L}{\partial p_1 \partial p_2} = \frac{2n_{33} + n_{13} + n_{23}}{(1 - p_1 - p_2)^2}$$

である。これらの期待値を要素に持つ行列

$$E(\mathbf{I}) = 2n \begin{bmatrix} \frac{1}{p_1} + \frac{1}{(1 - p_1 - p_2)} & \frac{1}{(1 - p_1 - p_2)} \\ \frac{1}{(1 - p_1 - p_2)} & \frac{1}{p_2} + \frac{1}{(1 - p_1 - p_2)} \end{bmatrix}$$

は，情報行列と呼ばれる。この行列の逆行列は，推定値の分散共分散行列になる。いまの場合，

$$\left[E(\mathbf{I})\right]^{-1} = \begin{bmatrix} Var(\hat{p}_1) & Cov(\hat{p}_1, \hat{p}_2) \\ Cov(\hat{p}_1, \hat{p}_2) & Var(\hat{p}_2) \end{bmatrix} = \begin{bmatrix} \frac{p_1(1 - p_1)}{2n} & -\frac{p_1 p_2}{2n} \\ -\frac{p_1 p_2}{2n} & \frac{p_1(1 - p_1)}{2n} \end{bmatrix}$$

となり，式（2―5）および（2―6）と一致する。

⑵ Baileyの方法

尤度方程式の形が複雑になり，解（最尤推定値）を得ることが困難な場合が生じる。しかし，特殊な場合には簡単に最尤推定値が得られる方法が開発されている。ここでは，独立な母数の数とデータの自由度が一致する場合に適用できるBaileyの方法を示す。

Bailey（1951）は，母数の自由度とデータの自由度が一致する場合には，観察値が母数によって表される期待値 m_i と等しいとおいた方程式の解が，最尤推定値を与えることを証明した（簡潔な証明は，Weir（1996）を参照）。たとえば，優劣関係のない2つの対立遺伝子 A_1 と A_2 が頻度 p_1 と p_2 で分離し，近交係数（厳密には，WrightのF統計量 F_{IS}）が f である集団を考えよう。各遺伝子型の期待頻度は

遺伝子型	期待頻度
A_1A_1	$p_1^2 + p_1p_2f$
A_1A_2	$2p_1p_2(1-f)$
A_2A_2	$p_2^2 + p_1p_2f$

となる。この集団から得た大きさ n の標本中に遺伝子型 A_1A_1，A_1A_2 および A_2A_2 の個体が，それぞれ n_{11}，n_{12} および n_{22} 個体含まれたとする。$p_1+p_2=1$ であるから独立な母数は，p_1 と f の 2 つであり，一方 $n_{11}+n_{12}+n_{22}=n$ であるからデータの自由度も 2 である。Bailey の方法を適用すると，p_1 と f の最尤推定値は，

$$n\left[\hat{p}_1^2 + \hat{p}_1(1-\hat{p}_1)\hat{f}\right] = n_{11}$$
$$n\left[2\hat{p}_1(1-\hat{p})\left(1-\hat{f}\right)\right] = n_{12}$$

の解

$$\hat{p}_1 = \frac{1}{2n}(2n_{11}+n_{12})$$
$$\hat{f} = 1-\frac{n_{12}}{2n\hat{p}_1(1-\hat{p}_1)}$$

として得られる。$\hat{p}_1$ はカウント法の推定値と一致する。また，Bailey の方法を適用すると，式（4—5）が最尤推定値を与えることも明らかである。

⑶ EM アルゴリズム

Bailey の方法が遺伝学のデータに適用できる場面は限られている。とくに，対立遺伝子間に優劣関係のある場合，表現型が遺伝子型と一対一に対応しないため，Bailey の方法を適用できないことが多い。このような場合に，推定のために必要なすべての情報を持つ完全なデータが得られたという仮想的な状況で尤度関数を設定し，反復解法で最尤推定値を計算するアルゴリズムが EM アルゴリズムである。

集団中に 2 つの対立遺伝子 A_1と A_2が頻度 p_1と $1-p_1$ で分離しているものとしよう。A_1 は A_2 に対して優性であるとすると，識別できるのは $A_1A_1+A_1A_2$ と A_2A_2 の 2 つのクラスだけである。標本中でのそれぞれの個体数を $n-n_{22}$ および n_{22} とする。いま，A_1の頻度として，任意に p_1'を設定すると，期待される A_1A_2 の標本中での度数は

$$n_{12}^{*} = \left(\frac{2p_1'(1-p_1')}{1-p_1'^2}\right)\times(n-n_{12})$$

である（E ステップ）。これを既知の観察度数のように扱い，尤度関数を最大化するように尤度方程式を作ると（M ステップ），p_1 の推定値が

$$p_1'' = \frac{1}{2n}\left(n_{12}^{*} + 2n_{22}\right)$$
$$= \frac{1}{2n}\left[\frac{2p_1'\left(1-p_1'\right)}{1-p_1'^2}\left(n-n_{12}\right)+2n_{22}\right]$$

として得られる。この手続きを反復することで，p_1 の最尤推定値が得られる。なお，この場合には，

$\hat{p}_1 = 1-\sqrt{n_{22}/n}$が収束解（最尤推定値）であることは，解析的に明らかである。

また，安田（1968）の反復解法（式（3―4）と（3―5））は，EM アルゴリズムそのものであることも明らかである。

数値例 6

雲南省における中国人6000人の ABO 血液型に関するデータ（表 3 ）を，EM アルゴリズムの数値例として用いる。A 遺伝子と B 遺伝子の頻度をそれぞれととすると，2 つの式

$$p'' = \frac{1}{2n}\left[n_{AB} + n_A\left\{1+\frac{p'}{p'+2\left(1-p'-q'\right)}\right\}\right]$$
$$p'' = \frac{1}{2n}\left[n_{AB} + n_B\left\{1+\frac{q'}{q'+2\left(1-p'-q'\right)}\right\}\right]$$

による反復計算の収束解として最尤推定値（$\hat{p}$ および $\hat{q}$）が得られる。O 遺伝子の頻度の最尤推定値は $\hat{r}=1-\hat{p}-\hat{q}$ である。

反復計算の初期値として Bernstein の近似解（数値例 2 ）$p'=0.2392$ と $q'=0.2077$ を用いた場合の計算結果を表 6 に示す。この例では，解は 8 回程度の反復でほぼ収束していることがわかる。なお，$p'=q'=1/3$ を初期値とした場合には，解の収束までに18回程度の反復が必要になる。

表６．EM アルゴリズムによる遺伝子頻度の計算結果

反復	$\hat{p}$	$\hat{q}$	$\hat{r}=1-\hat{p}-\hat{q}$
0	0.2392	0.2077	0.5531
1	0.23896229	0.20755219	0.55306514
2	0.23900812	0.20758784	0.55348552
3	0.23899484	0.20757723	0.55340404
4	0.23899698	0.20757897	0.55342792
5	0.23899619	0.20757834	0.55342405
6	0.23899627	0.20757841	0.55342548
7	0.23899622	0.20757837	0.55342532
8	0.23899622	0.20757837	0.55342541
9	0.23899622	0.20757837	0.55342541
10	0.23899622	0.20757837	0.55342541

⑷ Newton–Raphson 法

母数 ϕ の最尤推定値を $\hat{\phi}$，ϕ に関する尤度方程式を

$$S_{\phi}=\frac{\partial \ln L}{\partial \phi}$$

とすると，定義より $S_{\hat{\phi}}=0$ である。そこで，$S_{\hat{\phi}}=0$ を任意の値 ϕ' のまわりにテーラー展開し，2 次以上の項を無視すると

$$S_{\hat{\phi}}=0=S_{\phi'}+\left(\hat{\phi}-\phi'\right)\left[\frac{\partial S_{\phi}}{\partial \phi}\right]_{\phi=\phi'}$$

となる。これより，$\hat{\phi}$ の近似値 ϕ'' が

$$\begin{aligned}\phi''&=\phi'-S_{\phi'}\Big/\left[\frac{\partial S_{\phi}}{\partial \phi}\right]_{\phi=\phi'}\\&=\phi'+S_{\phi'}/I\left(\phi'\right)\end{aligned}$$

として得られる（式の右辺第 2 項の分母は，情報量であることに注意されたい）。この式を用いて反復計算を行えば，収束解として最尤推定値 $\hat{\phi}$ が得られる。

推定すべき母数が複数あるときには，情報行列 $\mathrm{I}(\varphi)$ を用いて，以下の式で反復計算を行う。

$$\boldsymbol{\varphi}''=\boldsymbol{\varphi}'+\mathbf{I}^{-1}\left(\boldsymbol{\varphi}'\right)S_{(\varphi')} \qquad (5-1)$$

ここで，φ は ϕ のベクトルである。この方法は，計算の副産物として推定値の分散共分散行列 $\mathrm{I}^{-1}(\varphi)$ が得られるという利点がある。

ヒトの ABO 血液型への応用を考えてみよう。A および B 遺伝子の頻度をそれぞれ p_A および p_B として表 2 の結果を適用すると，期待頻度は表 7 に示すように表せる。

表 7．ABO 血液型の期待頻度と観察度数

血液型	遺伝子型	期待頻度	観察度数
A	AA, AO	$p_A\left(2-p_A-2p_B\right)$	n_A
B	BB, BO	$p_B\left(2-2p_A-p_B\right)$	n_B
AB	AB	$2p_Ap_B$	n_{AB}
O	OO	$\left(1-p_A-p_B\right)^2$	n_O

尤度関数は

$$\begin{aligned}L\propto&\left[p_A\left(2-p_A-2p_B\right)\right]^{n_A}\left[p_B\left(2-2p_A-p_B\right)\right]^{n_B}\\&\times\left[2p_Ap_B\right]^{n_{AB}}\left[\left(1-p_A-p_B\right)^2\right]^{n_O}\end{aligned}$$

であるから，対数尤度関数は

$$\begin{aligned}\ln L=&\left(n_A+n_{AB}\right)\log p_A+n_A\log\left(2-p_A-2p_B\right)\\&+\left(n_B+n_{AB}\right)\log p_B+n_B\log\left(2-2p_A-p_B\right)\\&+2n_O\log\left(1-p_A-p_B\right)+\log 2\end{aligned}$$

となる。したがって，尤度方程式が

$$\frac{\partial \ln L}{\partial p_A}=\frac{n_A+n_{AB}}{p_A}-\frac{n_A}{2-p_A-2p_B}-\frac{2n_B}{2-2p_A-p_b}-\frac{2n_O}{1-p_A-p_B}=0$$
$$\frac{\partial \ln L}{\partial p_B}=\frac{n_B+n_{AB}}{p_B}-\frac{2n_A}{2-p_A-2p_b}-\frac{n_B}{2-2p_A-p_b}-\frac{2n_O}{1-p_A-p_B}=0$$

として得られる。また，Fisher の情報量は

$$-\frac{\partial^2 \ln L}{\partial p_A^2}=\frac{n_A+n_{AB}}{p_A^2}+\frac{n_A}{(2-p_A-2p_B)^2}+\frac{2n_B}{(2-2p_A-p_B)^2}+\frac{2n_O}{(1-p_A-p_B)^2}$$
$$-\frac{\partial^2 \ln L}{\partial p_B^2}=\frac{n_B+n_{AB}}{p_B^2}+\frac{2n_A}{(2-p_A-2p_B)^2}+\frac{n_B}{(2-2p_A-p_B)^2}+\frac{2n_O}{(1-p_A-p_B)^2}$$
$$-\frac{\partial^2 \ln L}{\partial p_A \partial p_B}=\frac{2n_A}{(2-p_A-2p_B)^2}+\frac{2n_B}{(2-2p_A-p_B)^2}+\frac{2n_O}{(1-p_A-p_B)^2}$$

である。これらを，式（5—1）に代入して得られる

$$\begin{bmatrix} p_A'' \\ p_B'' \end{bmatrix}=\begin{bmatrix} p_A' \\ p_B' \end{bmatrix}+\left[\begin{matrix} -\dfrac{\partial^2 \ln L}{\partial p_A^2} & -\dfrac{\partial^2 \ln L}{\partial p_A \partial p_B} \\ -\dfrac{\partial^2 \ln L}{\partial p_A \partial p_B} & -\dfrac{\partial^2 \ln L}{\partial p_B^2} \end{matrix}\right]^{-1}_{\begin{pmatrix} p_A=p_A' \\ p_B=p_B' \end{pmatrix}}\left[\begin{matrix} \dfrac{\partial \ln L}{\partial p_A} \\ \dfrac{\partial \ln L}{\partial p_B} \end{matrix}\right]_{\begin{pmatrix} p_A=p_A' \\ p_B=p_B' \end{pmatrix}} \qquad (5-2)$$

の収束解が最尤推定値となる。

数値例 7

数値例 6 で用いた ABO 血液型のデータに対して，Bernstein の近似解 $p'=0.2392$ と $q'=0.2077$ を初期値として，式（5—2）で反復計算を行った結果を表 8 に示す。収束解が EM アルゴリズムの結果（表 6）と一致することに注目されたい。

表 8．Newton-Raphson による遺伝子頻度の計算結果

反復	$\hat{p}_A$	$\hat{p}_B$	$\hat{p}_O$	$Var(\hat{p}_A)$	$Var(\hat{p}_B)$	$Cov(\hat{p}_A,\hat{p}_B)$
0	0.2392	0.2077	0.5531			
1	0.23899114	0.20757422	0.55343463	0.00001808	0.00001599	-.00000434
2	0.23899635	0.20757847	0.55342518	0.00001805	0.00001597	-.00000433
3	0.23899621	0.20757837	0.55342542	0.00001805	0.00001597	-.00000433
4	0.23899622	0.20757837	0.55342541	0.00001805	0.00001597	-.00000433
5	0.23899622	0.20757837	0.55342541	0.00001805	0.00001597	-.00000433

O 遺伝子の推定頻度 $\hat{p}_0$ の分散は，

$$\begin{aligned} Var\left(\hat{p}_O\right) &= Var\left(1-\hat{p}_A-\hat{p}_B\right) \\ &= Var\left(\hat{p}_A\right)+Var\left(\hat{p}_B\right)+2Cov\left(\hat{p}_A,\hat{p}_B\right) \\ &= 0.00001805+0.00001597-2\times 0.00000433 \\ &= 0.00002536 \end{aligned}$$

である。

6．ベイズ法

これまでの推定法は，母集団の分布を仮定し，その分布を規定する真の母数（パラメータ）が存在することを前提としている。これに対してベイズ法では，母集団の分布という概念を捨て去り，データが得られたという条件の下でのパラメータの確率分布である事後確率をもとに議論する。この方法は，ある事象Bの事前確率 $\Pr(B)$ を用いて，事象Aが起こった下での事象Bの起こる条件付確率（事後確率）$\Pr(B|A)$ が

$$\Pr\left(B\middle|A\right)=\frac{\Pr\left(A\middle|B\right)}{\Pr\left(A\right)}\Pr\left(B\right)$$

なるベイズの定理により得られることに基づいたものである。これは，「過去の経験に基づく知識＝事前確率」を「新たな経験」により修正して「新たな知識＝事後確率」とする，われわれの「学習」を定式化したものと言える。

上の式の事象Bをパラメータ ϕ，事象Aをデータ $\{n\}$，事前および事後確率の密度を $\pi(\phi)$ および $\pi(\phi|\{n\})$ とすれば，

$$\pi\left(\phi\middle|\{n\}\right)=\frac{\Pr\left(\{n\}\middle|\phi\right)\pi\left(\phi\right)}{\int\Pr\left(\{n\}\middle|\phi\right)\pi\left(\phi\right)d\phi} \qquad (6-1)$$

と書ける。

ヒトのMN血液型への応用として，以下のデータが得られたものとしよう。

血液型	遺伝子型	観察度数
M	MM	n_{MM}
MN	MN	n_{MN}
N	NN	n_{NN}
		計 n

このデータが得られたという条件の下でのM遺伝子の頻度 p_M の事後分布を考える。まず，事前分布は取り扱いが便利なように，Gunnel and Wearden（1995）にしたがって，パラメータ α と β を持つベータ分布で近似する。すなわち，

$$\pi\left(p_M\right)=\frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}p_M^{\alpha-1}\left(1-p_M\right)^{\beta-1} \qquad (6-2)$$

また，与えられた p_M の下でのデータ中のM遺伝子の数（$n_M=2n_{MM}+n_{MN}$）の確率分布は，式（2―

1）より，

$$\Pr\left(n_M \middle| p_M\right) = \frac{(2n)!}{n_M!(2n-n_M)!} p_M^{n_M} \left(1-p_M\right)^{2n-n_M}$$

である。これらを式（6―1）に代入して整理すると，事後分布として

$$\pi\left(p_M \middle| n_M\right) = \frac{\Gamma(\alpha+\beta+2n)}{\Gamma(\alpha+n_M)\Gamma(\beta+2n-n_M)} p_M^{\alpha+n_M-1}\left(1-p_M\right)^{\beta+2n-n_M} \qquad (6-3)$$

が得られる。

数値例 8

あるヒト集団において，MN 血液型についての過去の調査結果から事前分布は $\alpha=26$，$\beta=14$ のベータ分布で近似できるものとする。新たな調査で，つぎのような結果が得られたとしよう。

遺伝子型	観察度数
MM	$n_{MM}=30$
MN	$n_{MN}=20$
NN	$n_{NN}=10$
	$n=60$

式（6―2）および（6―3）から求めた事前分布と事後分布を図1に示した。事前分の平均は

$$E\left[p_M\right] = \frac{\alpha}{\alpha+\beta} = 0.6500$$

事後分布の平均は

$$E\left[p_M \middle| n_M\right] = \frac{\alpha+n_M}{\alpha+\beta+2n} = 0.6625$$

である。

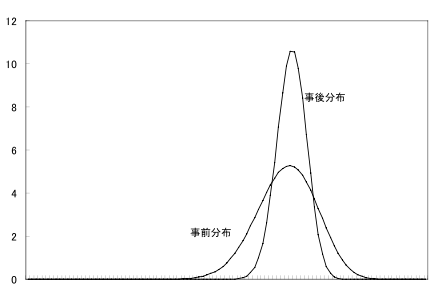

図1．事前分布 $\pi\left(p_M\right)$ と事後分布 $\pi\left(p_M \middle| n_M\right)$

7．昆虫集団への適用例

⑴ナミテントウの鞘翅斑紋に関わる遺伝子の最尤法による頻度推定

ナミテントウの鞘翅斑紋遺伝子の頻度の推定法については，すでに数値例5で示した。推定のための式（式（4―5））が，最尤推定値を与えることもBaileyの公式から明らかである。ここでは，Newton–Raphson法による最尤推定を試みる。もちろん推定される遺伝子頻度は，式（4―5）から得られる値に等しいが，Newton–Raphson法による反復解法は，推定値の分散を同時に与えるというが利点である。

二紋型遺伝子（h^c），四紋型遺伝子（h^{SP}）および斑型遺伝子（h^a）の頻度を，それぞれp，qおよびrとする。表9には，4つの斑紋型の遺伝子型，期待頻度，観察度数を示した。

表9．ナミテントウの斑紋型の遺伝子型、期待頻度および観察度数

斑紋型	遺伝子型	期待頻度	観察度数
二紋型	h^ch^c, h^ch^{SP}, h^ch^a, h^ch	$p^2+2pq+2pr+2p(1-p-q-r)$ $=p(2-p))$	n_1
四紋型	$h^{SP}h^{SP}$, $h^{SP}h^a$, $h^{SP}h$	$q^2+2qr+2q(1-p-q-r)$ $=q(2-2q-q)$	n_2
斑型	h^ah^a, h^ah	$r^2+2r(1-p-q-r)$ $=r(2-2p-2q-r)$	n_3
紅型	hh	$(1-p-q-r)^2$	n_4
計		1	n

期待頻度と観察度数から，尤度関数は

$$L\propto\left[p(2-p)\right]^{n_1}\left[q(2-2p-q)\right]^{n_2}\left[r(2-2p-2q-r)\right]^{n_3}\left[(1-p-r-r)^2\right]^{n_4}$$

したがって，対数尤度関数は

$$\begin{aligned}\ln L=&n_1\left\{\log p+\log(2-p)\right\}+n_2\left\{\log q+\log(2-2p-q)\right\}\\&+n_3\left\{\log r+\log(2-2p-2q-r)\right\}+2n_4\log(1-p-q-r)\end{aligned}$$

となる。対数尤度関数より，以下の尤度方程式が得られる。

$$\frac{\partial\ln L}{\partial p}=n_1\left(\frac{1}{p}-\frac{1}{2-p}\right)-\frac{2n_2}{2-2p-q}-\frac{2n_3}{2-2p-2q-r}-\frac{2n_4}{1-p-r-r}=0$$

$$\frac{\partial\ln L}{\partial q}=n_2\left(\frac{1}{q}-\frac{1}{2-2p-q}\right)-\frac{2n_3}{2-2p-2q-r}-\frac{2n_4}{1-p-r-r}=0$$

$$\frac{\partial\ln L}{\partial r}=n_3\left(\frac{1}{r}-\frac{1}{2-2p-2q-r}\right)-\frac{2n_4}{1-p-r-r}=0$$

さらに尤度方程式を p，q あるいは r で偏微分し，－1 を乗じて以下の情報量を得る。

$$-\frac{\partial^2 \ln L}{\partial p^2} = n_1\left\{\frac{1}{p^2}+\frac{1}{(2-p)^2}\right\}+\frac{4n_2}{(2-2p-q)^2}+\frac{4n_3}{(2-2p-2q-r)^2}+\frac{2n_4}{(1-p-q-r)^2}$$

$$-\frac{\partial^2 \ln L}{\partial q^2} = n_2\left\{\frac{1}{q^2}+\frac{1}{(2-2p-q)^2}\right\}+\frac{4n_3}{(2-2p-2q-r)^2}+\frac{2n_4}{(1-p-q-r)^2}$$

$$-\frac{\partial^2 \ln L}{\partial r^2} = n_3\left\{\frac{1}{r^2}+\frac{1}{(2-2p-2q-r)^2}\right\}+\frac{2n_4}{(1-p-q-r)^2}$$

$$-\frac{\partial^2 \ln L}{\partial p\partial q} = \frac{2n_2}{(2-2p-q)^2}+\frac{4n_3}{(2-2p-2q-r)^2}+\frac{2n_4}{(1-p-q-r)^2}$$

$$-\frac{\partial^2 \ln L}{\partial p\partial r} = \frac{2n_3}{(2-2p-2q-r)^2}+\frac{2n_4}{(1-p-q-r)^2}$$

$$-\frac{\partial^2 \ln L}{\partial q\partial r} = \frac{2n_3}{(2-2p-2q-r)^2}+\frac{2n_4}{(1-p-q-r)^2}$$

これらを以下の式に代入し，反復計算による収束解として最尤推定値 $\hat{p}$，$\hat{q}$ および $\hat{r}$ が得られる。

$$\begin{bmatrix} p'' \\ q'' \\ r'' \end{bmatrix} = \begin{bmatrix} p' \\ q' \\ r' \end{bmatrix} + \begin{bmatrix} -\frac{\partial^2 \ln L}{\partial p^2} & -\frac{\partial^2 \ln L}{\partial p\partial q} & -\frac{\partial^2 \ln L}{\partial p\partial r} \\ -\frac{\partial^2 \ln L}{\partial p\partial q} & -\frac{\partial^2 \ln L}{\partial q^2} & -\frac{\partial^2 \ln L}{\partial q\partial r} \\ -\frac{\partial^2 \ln L}{\partial p\partial r} & -\frac{\partial^2 \ln L}{\partial q\partial r} & -\frac{\partial^2 \ln L}{\partial r^2} \end{bmatrix}^{-1}_{\substack{p=p' \\ q=q' \\ r=r'}} \begin{bmatrix} \frac{\partial \ln L}{\partial p} \\ \frac{\partial \ln L}{\partial q} \\ \frac{\partial \ln L}{\partial r} \end{bmatrix}_{\substack{p=p' \\ q=q' \\ r=r'}}$$

表 5 で示した京都産業大学周辺のデータを用い，初期値を $p'=0.4$，$q'=0.1$ および $r'=0.1$ とした計算結果を表10に示す。

表１０．最尤法による遺伝子頻度の推定結果

反復	$\hat{p}$	$\hat{q}$	$\hat{r}$	$Var(\hat{p})$	$Var(\hat{q})$	$Var(\hat{r})$	$Cov(\hat{p},\hat{q})$	$Cov(\hat{q},\hat{r})$	$Cov(\hat{q},\hat{r})$
0	0.4000	0.1000	0.1000						
1	0.44086701	0.13409663	0.03691351	0.00025377	0.00010209	0.00027454	-.00002610	-.00006458	-.00002156
2	0.43454297	0.14488096	0.05056414	0.00027096	0.00016991	0.00004547	-.00006044	-.00001140	-.00000593
3	0.43248306	0.14450693	0.05694944	0.00026744	0.00019275	0.00008277	-.00006895	-.00002097	-.00001308
4	0.43226451	0.14436665	0.05778362	0.00026648	0.00019201	0.00010318	-.00006794	-.00002636	-.00001647
5	0.43226147	0.14436484	0.05779497	0.00026638	0.00019172	0.00010597	-.00006774	-.00002711	-.00001692
6	0.43226147	0.14436484	0.05779497	0.00026638	0.00019171	0.00010601	-.00006774	-.00002712	-.00001692
7	0.43226147	0.14436484	0.05779497	0.00026638	0.00019171	0.00010601	-.00006774	-.00002712	-.00001692
8	0.43226147	0.14436484	0.05779497	0.00026638	0.00019171	0.00010601	-.00006774	-.00002712	-.00001692

紅型遺伝子の頻度の最尤推定値 $\hat{s}$ は,

$$\hat{s} = 1 - \hat{p} - \hat{q} - \hat{r} = 0.36557872$$

推定値の分散は

$$\begin{aligned} Var(\hat{s}) &= Var(1 - \hat{p} - \hat{q} - \hat{r}) \\ &= Var(\hat{p}) + Var(\hat{q}) + Var(\hat{r}) + 2Cov(\hat{p},\hat{q}) + 2Cov(\hat{p},\hat{r}) + 2Cov(\hat{q},\hat{r}) \\ &= 0.00034055 \end{aligned}$$

として得られる。

$100\times(1-\alpha)$％信頼区間は，最尤推定値の漸近正規性を利用して，たとえば $\hat{p}$ について

$$\hat{p} \pm Z_{\alpha/2}\sqrt{Var(\hat{p})}$$

として設定できる。他の推定値についても同様に，信頼区間が設定できる。表11に結果をまとめておく。

表１１．京都産業大学周辺(2002-04 年)のナミテントウ集団における斑紋遺伝子の頻度の推定値と９５%信頼区間

遺伝子	最尤推定値	95%信頼区間
二紋型	0.4323	[0.4003, 0.4643]
四紋型	0.1444	[0.1172, 0.1715]
斑型	0.0578	[0.0376, 0.0780]
紅型	0.3656	[0.3294, 0.4017]

駒井（1956）は，1940–43年に京都市内で2494個体のナミテントウを採集し，斑紋型別割合について報告している。彼の記録を最尤法によって分析した結果を表12に示す。

表１２．1940-43 年の京都市における記録から推定した斑紋遺伝子の頻度と 2002-04 の頻度の差

遺伝子	遺伝子頻度	差
二紋型	0.3979	0.0344*
四紋型	0.1498	-0.0540^{ns}
斑型	0.0608	-0.0030^{ns}
紅型	0.3914	-0.0258^{ns}

差：（2002-04 年の頻度）－（1940-43 年の頻度）

＊：$p<0.05$

つぎに，1940–43年の遺伝子頻度と2002–04年の遺伝子頻度との差について統計的に検定する。以下，二紋型遺伝子を例とする。2002–04年および1940–43年の頻度をそれぞれ $\hat{p}_1$ および $\hat{p}_2$ として，差

$$d = \hat{p}_1 - \hat{p}_2$$

がゼロと有意に異なるかどうかを検定する。検定の帰無仮説（H_0）と対立仮説（H_1）は

$$H_0 : d = 0$$
$$H_1 : d \neq 0$$

である。2002–04年および1940–43年の総標本数をそれぞれ n_1および n_2，平均の遺伝子頻度 $\bar{p}$ を

$$\overline{p} = \frac{2n_1p_1 + 2n_2p_2}{2n_1 + 2n_2} = \frac{n_1p_1 + n_2p_2}{n_1 + n_2}$$

とし，帰無仮説の下でのの分布を

$$\text{平均：} E\left[d\right] = 0$$

$$\text{分散：} \sigma_d^2 = \left(\frac{1}{2n_1} + \frac{1}{2n_2}\right)\overline{p}\left(1 - \overline{p}\right)$$

の正規分布で近似する。したがって，

$$u = \frac{\hat{p}_1 - \hat{p}_1}{\sigma_d}$$

は標準正規分布にしたがう。$n_1 = 636$，$n_2 = 2492$ および表11と12に示した遺伝子頻度の推定値を代入すると，$u = 2.23$ を得る。標準正規分布表の値 $Z_{0.05/2} = 1.96$，$Z_{0.01/2} = 2.58$ と比較すると，$u > Z_{0.05/2}$ であるから，2002–04年と1940–43年の二紋型遺伝子の頻度の差は，5％水準（$p < 0.05$）で有意である。他の遺伝子について同様の検定を行った結果も合わせて表12に示す。

⑵ミドリシジミの雌の斑紋の遺伝様式に関する統計的推論

ミドリシジミは，ハンノキ林に生息する美しい蝶である。この蝶は翅表面に著しい性的二型を示す。すなわち，雄は金属性に輝く緑色なしているのに対して（写真２），雌は暗褐色を呈する。この雌にはＡ型，Ｂ型，ＡＢ型およびＯ型の４つの型があることが知られている（写真２）。Ａ型は前翅第３および４室に橙色が入り，Ｂ型は前翅 I_b および中室に青色の斑がある。ＡＢ方はＡ型とＢ型の特徴をあわせ持っており，Ｏ型はいずれの特徴も持たず，一面に暗褐色を呈する。

雌に見られる多型は，その呼び方から察せられるように，ヒトのABO血液型と同じ遺伝様式にしたがうと考えられてきた。駒井（1956）は，全国数ヶ所から得られた標本集団について，雌の各型の出現度数を調べ，それらがABO血液型と同様に１遺伝子座上の３つの対立遺伝子を仮定したとき（ABO仮説）のハーディー・ワインベルグ頻度と適合すると報告している。しかし，その結果を詳細に見ると，必ずしもABO仮説の適合度は高いものとは言えず，後で述べるように系統的とも思われる誤差が含まれている。

このような点から，ABO仮説を疑問視する報告が数人の研究者および愛好家から提出されている（長谷川，1961；西田，1981；鎌田，1987）。彼らはいずれも，Ａ斑とＢ斑は独立した（連鎖のない）別の遺伝子座によって支配を受け，それぞれの遺伝子座上の優劣関係のある２つの対立遺伝子が

(1)雄

(2)雌 A 型

(3)雌 B 型

(4)雌 AB 型

(5)雌 O 型

写真 2．ミドリシジミの翅表面

4 つの型を決定する（A–B 仮説）と主張している。2 つの仮説の正当性については，いまだに結論が出されていない。そこで，この問題について最尤法による検討を試みた。

材料

材料としては，長谷川（1961）の報告から，2 地域のデータを選んで用いた。それぞれの地域の 4 つの型の観察度数は，表13に示すとおりである。

表１３．ミドリシジミの標本集団

型	長野県諏訪	大阪府能勢
A 型	28	31
B 型	17	99
AB 型	17	26
O 型	41	106
計	103	262

ABO 仮説

A 斑の遺伝子頻度を p_A，B 斑の遺伝子頻度を p_B として，表14に ABO 仮説に基づく各型の遺伝子型および期待頻度を示した。

表１４．ABO 仮説に基づく各型の期待頻度

型	遺伝子型	期待頻度	観察度数
A 型	AA, AO	$p_A(2-p_A-2p_B)$	n_A
B 型	BB, BO	$p_B(2-2p_A-p_B)$	n_B
AB 型	AB	$2p_Ap_B$	n_{AB}
O 型	OO	$(1-p_A-p_B)^2$	n_O
計		1	n

この仮説の下での尤度関数は

$$L=\frac{n!}{n_A!n_B!n_{AB}!n_O!}\left[p_A(2-p_A-2p_B)\right]^{n_A}\left[p_B(2-2p_A-p_B)\right]^{n_B} \times\left[2p_Ap_B\right]^{n_{AB}}\left[(1-p_A-p_B)^2\right]^{n_O} \qquad (7-1)$$

すでに述べた手続きにより，遺伝子頻度の最尤推定値は尤度方程式

$$\frac{\partial \ln L}{\partial p_A}=\frac{n_A+n_{AB}}{p_A}-\frac{n_A}{2-p_A-2p_B}-\frac{2n_B}{2-2p_A-p_b}-\frac{2n_O}{1-p_A-p_B}=0$$
$$\frac{\partial \ln L}{\partial p_B}=\frac{n_B+n_{AB}}{p_B}-\frac{2n_A}{2-p_A-2p_b}-\frac{n_B}{2-2p_A-p_b}-\frac{2n_O}{1-p_A-p_B}=0$$

の解として得られる。

表11に示したデータから求めた遺伝子頻度の最尤推定値を表15に示す。

表１５．ABO 仮説の下での遺伝子頻度の最尤推定値

	$\hat{p}_A$	$\hat{p}_B$	$\hat{p}_O(=1-\hat{p}_A-\hat{p}_B)$
長野県諏訪	0.2418	0.1759	0.5823
大阪府能勢	0.1138	0.2729	0.6133

A–B 仮説

A 斑は遺伝子座 A 上の優性遺伝子 A（その対立遺伝子は a とする）によって発現し，B 斑は遺伝子座 A とは独立した遺伝子座 B 上の優性遺伝子 B（その対立遺伝子は b とする）によって発現するとする。A および B 遺伝子の頻度をそれぞれ p_A および p_B とし，ハーディー・ワインベルグ平衡を仮定すると，各斑紋の出現頻度の期待値は表16に示すようになる。

表１６．A-B 仮説に基づく各型の期待頻度

型	遺伝子型	期待頻度	観察度数
A 型	AAbb, Aabb	$\left[p_A^2+2p_A(1-p_A)\right](1-p_B)^2$	n_A
B 型	aaBB, aaBb	$(1-p_A)^2\left[p_B^2+2p_B(1-p_B)\right]$	n_B
AB 型	AABB, AABb, AaBB, AaBb	$\left[p_A^2+2p_A(1-p_A)\right]\left[p_B^2+2p_B(1-p_B)\right]$	n_{AB}
O 型	aabb	$(1-p_A)^2(1-p_B)^2$	n_O
計		1	n

したがって，尤度関数は

$$
\begin{aligned}
L=\frac{n!}{n_A!n_B!n_{AB}!n_O!}&\left\{\left[p_A^2+2p_A(1-p_A)\right](1-p_B)^2\right\}^{n_A}\\
&\times\left\{(1-p_A)^2\left[p_B^2+2p_B(1-p_B)\right]\right\}^{n_B}\\
&\times\left\{\left[p_A^2+2p_A(1-p_A)\right]\left[p_B^2+2p_B(1-p_B)\right]\right\}^{n_{AB}}\\
&\times\left\{(1-p_A)^2(1-p_B)^2\right\}^{n_O}
\end{aligned}
\qquad (7-2)
$$

となる。また，尤度方程式は

$$
\frac{\partial \ln L}{\partial p_A}=(n_A+n_{AB})\left(\frac{1}{p_A}-\frac{1}{2-p_A}\right)-(n_B+n_O)\left(\frac{1}{1-p_A}\right)=0
$$

$$
\frac{\partial \ln L}{\partial p_B}=(n_B+n_{AB})\left(\frac{1}{p_B}-\frac{1}{2-p_B}\right)-(n_A+n_O)\left(\frac{1}{1-p_B}\right)=0
$$

となり，最尤推定値は

$$
n\hat{p}_A^2-2n\hat{p}_A+n_A+n_{AB}=0
$$

$$
n\hat{p}_B^2-2n\hat{p}_A+n_B+n_{AB}=0
$$

の解として得られる

表15に示したデータから求めた遺伝子頻度の最尤推定値を表17に示す。

表１７．A-B 仮説の下での遺伝子頻度の最尤推定値

	遺伝子座 A		遺伝子座 B	
	$\hat{p}_A$	$\hat{p}_a(=1-\hat{p}_A)$	$\hat{p}_B$	$\hat{p}_b(=1-\hat{p}_B)$
長野県諏訪	0.2496	0.7504	0.1815	0.8185
大阪府能勢	0.1154	0.8846	0.2769	0.7231

2つの仮説の比較

表18に，2つの地域における観察度数と両仮説の下での遺伝子頻度の推定値から期待される度数を示した。また，仮説の適合度を調べるために行った χ^2 検定の結果（自由度はいずれの仮説でも1）および式（7—1）と（7—2）から計算した尤度もあわせて示してある。

表１８．諏訪および能勢における観察度数と両仮説の下での期待度数、x^2 検定の結果ならびに尤度

地域		A	B	AB	O	χ^2 (Prob)	尤度
諏訪	観察度数	28	17	17	41		
	期待値（ABO）	35.02	24.29	8.76	34.92	12.40 (<0.01)	0.0446×10^{-4}
	期待値（A-B）	30.15	19.15	14.85	38.85	0.82 (0.3-0.5)	7.3256×10^{-4}
能勢	観察度数	31	99	26	106		
	期待値（ABO）	39.96	107.21	16.27	98.55	9.02 (<0.01)	0.0536×10^{-4}
	期待値（A-B）	29.81	97.81	27.19	107.19	0.13 (0.7-0.8)	3.2902×10^{-4}

この結果から，A–B 仮説のほうが ABO 仮説よりも観察度数にはるかによく当てはまることがわかる。さらに，ABO 仮説では，いずれの地域においても A 型と B 型の頻度は過大に，逆に AB 型と O 型の頻度は過小に推定されている。駒井（1956）は，他のいくつかの地域についても表18と同様に観察度数と ABO 仮説の下での期待値を比較しているが，このような系統的と思われる誤差は全地域において認められる。駒井（1956）はこの点に関して，遺伝子型間の適応度の違いを原因としているが，表18の結果は，むしろ ABO 仮説自体を誤りとし，A–B 仮説を受け入れるほうがはるかに合理的な説明ができることを示唆している。

本種は，かつて累代飼育が困難とされていたが，愛好家の長年にわたる努力により現在では飼育技術が確立されている。今後は，交配実験などにより上記の推論を検証すべきであろう。

参考文献

Bailey, N. T. J.（1951）Testing the solubility of maximum likelihood equations in the routine application of scoring methods. Biometrics 7: 268–274.

Gunel, E. and Wearden, S.（1995）Bayesian estimation and testing of gene frequencies. Theor. Appl. Genet. 91: 534–543.

長谷川順一（1961）ミドリシジミ雌４型の遺伝・遺伝。15：10月号61–63.

稲垣宣生（1990）数理統計学・裳華房・東京。

鎌田邦彦（1987）ミドリシジミ類♀の４型について・蝶研フィールド・2: No. 6, 23.

駒井　卓（1956）蝶２種の集団遺伝学・“集団遺伝学”79–83. 培風館。

Li, C. C.（1976）First Course in Population Genetics. Boxwood, California.

西田眞也（1981）アイノミドリシジミの遺伝について・月刊むし・123: 3–12.

Spiess, E. B.（1989）Genes in Populations. 2nd ed. John Wiley & Sons, New York.

Weir, B. S.（1996）Genetic Data Analysis II. Sinauer, Massachusetts.

安田徳一（1968）カウント法による遺伝子頻度の推定・人類遺伝学雑誌・12：226–245.

Yasuda, N.（1984）A note on gene frequency estimation in the ABO and ABO–like system. Jpn. J. Hum. Genet. 29：371–380.

Yasuda, N. and Kimura, M.（1968）A gene–counting method of maximum likelihood for estimating gene frequencies in ABO and ABO–like systems. Ann. Human Genet. 31：409–420.