EM アルゴリズム（The Expectation Maximization）

一般にデータは不完全であり、それに基づいて推定しなければならない場合が多い。また、複雑な尤度関数の場合、最尤推定値を求めることは困難な場合がある。EM アルゴリズムは不完全データの問題を完全データのフレームワークで逐次的にパラメーターの最尤推定量を求めてゆく方法で、計算自体より実行し易いアルゴリズムである。ただし、局所的な極大に到達してしまう可能性がある。

完全データ： $X=(Y,Z)$， 不完全データ（観測データ）：Y 、 欠失データ：Z

$Y=t(X)$： X から Y への射影

＜EM アルゴリズムの手順＞

パラメーターベクトル ϑ の下での X の分布密度： $f(X|\vartheta)$

① $Q(\vartheta|\vartheta_n)=E\left[\log f(X|\vartheta)|Y,\vartheta_n\right]$ ：データ Y 、パラメーター $\vartheta=\vartheta_n$ の下での分布密度

$f(X|\vartheta)$ の条件付期待値

② $\vartheta_{n+1}=Arg\,\underset{\vartheta}{Max}\,Q(\vartheta|\vartheta_n)$ ： $Q(\vartheta|\vartheta_n)$ を最大にする ϑ を ϑ_{n+1} とする。

③ $g(Y|\vartheta)$ ：観測データ Y の尤度関数 $\log g(Y|\vartheta_{n+1})\geq\log g(Y|\vartheta_n)$ が成り立つ

（観測データ Y の尤度が増加）

§１．EM アルゴリズムの証明

補助定理１．（Jensen の不等式）

X：確率変数、 $h(x)$ を凸関数（下に凸）とする。

このとき $E[h(X)]\geq h(E[X])$ が成り立つ。厳密に凸な関数 $(\frac{d^2}{dx^2}h(x)>0)$ においては等号は $X=E[X]$ *a.e.* すなわちＸが定数のときのみ成り立つ。

（証明）

$E[X]=\mu$ とする。 $h(x)$ は凸関数なので $\frac{d^2}{dx^2}h(x)\geq 0$ 、テーラーの定理より

$$h(x)=h(\mu)+(x-\mu)h'(\mu)+\frac{1}{2}(x-\mu)^2h''(\theta)\ ,\ (\theta\in(\mu,x))$$

$$\geq h(\mu)+(x-\mu)h'(\mu)$$

x を確率変数 X に置き換えると $h(W) \geq h(\mu) + (X - \mu)h'(\mu)$、期待値を取ると $E[h(X)] \geq h(\mu) + (E[X] - \mu)h'(\mu) = h(\mu)$、すなわち $E[h(X)] \geq h(E[X])$ が成り立つ。

$\frac{d^2}{dx^2}h(x) > 0$ のとき恒等的に $X \equiv \mu$ でなければ、明らかに不等号 $E[h(X)] > h(E[X])$ となる。

補助定理２．（エントロピー不等式）

$f(x),\ g(x)$ を測度 $\mu(x)$ に関する確率分布密度、すなわち $\int f(x)d\mu = \int g(x)d\mu = 1$、

かつ $f(x) > 0, g(x) > 0$ とする。また $E_f[h(x)] = \int h(x) f(x)d\mu$ と定義する。

このとき $E_f[\log f(x)] \geq E_f[\log g(x)]$、等号は $f(x) = g(x)\ \ (a.e.\ \mu)$ のときに限る。

（証明）$-\log x$ は下に凸な関数なので、Jensen の不等式より

$$E_f[\log f(x)] - E_f[\log g(x)] = E_f[-\log \frac{g(x)}{f(x)}] \geq -\log E_f\left[\frac{g(x)}{f(x)}\right]$$

$$= -\log \int \frac{g(x)}{f(x)} f(x)d\mu = -\log \int g(x)d\mu = 0$$

よって $E_f[\log f(x)] \geq E_f[\log g(x)]$。等号は $\frac{g(x)}{f(x)} = E_f\left[\frac{g(x)}{f(x)}\right] = \int g(x)d\mu = 1$ が $a.e.\mu$ に成り立つとき、すなわち $f(x) = g(x)\ \ (a.e.\mu)$ のときに限る。　　（終り）

$Q(\vartheta|\vartheta_n) = E\left[\log f(X|\vartheta)|Y = y, \vartheta_n\right]$　　ϑ_n は現在の ϑ の推定値

$g(Y|\vartheta)$ ：観測データの尤度関数(likelihood)。

補助定理３．

$$Q(\vartheta_n|\vartheta_n) - \log g(y|\vartheta_n) \geq Q(\vartheta|\vartheta_n) - \log g(y|\vartheta)$$

（証明）$\frac{f(x|\vartheta)}{g(y|\vartheta)} = \frac{P(X,\vartheta)}{P(Y,\vartheta)} = P(X,\vartheta|Y,\vartheta)$：不完全データ Y の下での完全データ X の分布

同様に、$\frac{f(x|\vartheta_n)}{g(y|\vartheta_n)} = \frac{P(X,\vartheta_n)}{P(Y,\vartheta_n)} = P(X,\vartheta_n|Y,\vartheta_n)$。

$A = \{x;\ t(x) = y\} = \{X = (y, Z)\}$ すなわち A は空間 $X = (Y, Z)$ の $Y = y$ 断面とすると、

$$\int_A \frac{f(x|\vartheta)}{g(y|\vartheta)}d\mu = \frac{1}{g(y|\vartheta)}\int_A f(x|\vartheta)d\mu = \frac{g(y|\vartheta)}{g(y|\vartheta)} = 1, \qquad \int_A \frac{f(x|\vartheta_n)}{g(y|\vartheta_n)}d\mu = 1$$

A 上で $F = \frac{f(x|\vartheta_n)}{g(y|\vartheta_n)}, G = \frac{f(x|\vartheta)}{g(x|\vartheta)}$ とすると、$Fd\mu, Gd\mu$ は A 上の確率密度である。

補助定理２より $E_F[\log F] \geq E_F[\log G]$、すなわち

$\int_A \left(\log \frac{f(x|\vartheta_n)}{g(x|\vartheta_n)}\right) Fd\mu \geq \int_A \left(\log \frac{f(x|\vartheta)}{g(x|\vartheta)}\right) Gd\mu$。これより

$$E\left[\log f(x|\vartheta_n) \middle| Y = y, \vartheta_n\right] - \log g(y|\vartheta_n) \geq E\left[\log f(x|\vartheta) \middle| Y = y, \vartheta\right] - \log g(y|\vartheta)$$

故に $Q(\vartheta_n|\vartheta_n) - \log g(y|\vartheta_n) \geq Q(\vartheta|\vartheta_n) - \log g(y|\vartheta)$ が成り立つ。　　（証明終り）

これより、任意の ϑ に対して $\log g(y|\vartheta) - \log g(y|\vartheta_n) \geq Q(\vartheta|\vartheta_n) - Q(\vartheta_n|\vartheta_n)$　、

$\vartheta_{n+1} = Arg\,\underset{\vartheta}{Max}\, Q(\vartheta|\vartheta_n)$ とすると、$Q(\vartheta_{n+1}|\vartheta_n) \geq Q(\vartheta_n|\vartheta_n)$ なので、

$\log g(y|\vartheta_{n+1}) - \log g(y|\vartheta_n) \geq Q(\vartheta_{n+1}|\vartheta_n) - Q(\vartheta_n|\vartheta_n) \geq 0$。故に $\log\, g(Y|\vartheta_{n+1}) \geq \log g(Y|\vartheta_n)$ が成り立ち、EM アルゴリズムの手順②、③のステップが証明された。

＜補足説明＞この証明で、不等式 $\int_A \left(\log \frac{f(x|\vartheta_n)}{g(x|\vartheta_n)}\right) Fd\mu \geq \int_A \left(\log \frac{f(x|\vartheta)}{g(x|\vartheta)}\right) Gd\mu$ が本質的な部分である。両辺は領域 A 上の積分であり、積分結果は (ϑ_n, y) および (ϑ, y) の関数である。

§２．例

＜例１＞　くじ引きの当りの確率

　くじ引きで、くじを最大２回引くチャンスがあるとする。ただし１回目に当らない場合はくじを戻してもう一度引くことにする。N 人がこの方法でくじを引いて k 人が当ったとき、１回くじを引いたとき当る確率 ϑ を最尤推定する。各人の結果を当たりは A、はずれは O で表す。

不完全データ Y:　　$N(A) = k,\ N(O) = n - k$　　２回のチャンスで当たりの人数とはずれの人数

完全データ：$n(A), n(OA), n(OO) = N(O)$　１回目に当った人数、２回目に当った人数
２回ともはずれた人数

パラメーター：　当る確率ϑ

完全データ X の尤度関数

$$f(X|\vartheta) = \frac{n!}{n(A)!n(OA)!n(OO)!}\vartheta^{n(A)}\{(1-\vartheta)\vartheta\}^{n(OA)}\{(1-\vartheta)^2\}^{n(OO)}$$　（多項分布）

ϑ_m　：　第ｍステップ目のϑの推定値

$$\begin{aligned}Q(\vartheta|\vartheta_m) &= E\left[\log f(X|\vartheta)\,\middle|\, Y, \vartheta_m\right] \\ &= E[n(A)\log\vartheta + n(OA)\log\left((1-\vartheta)\vartheta\right) + 2n(OO)\log(1-\vartheta) \\ &\qquad + \log\left(\frac{n!}{n(A)!n(OA)!n(OO)!}\right)|Y, \vartheta_m]\end{aligned}$$

(Y, ϑ_m)が与えられた条件下で$n(OO) = N(O)$より$E[n(OO)|Y, \vartheta_m] = N(O)$。

$f(X|\vartheta)$は多項分布で、$n(A) + n(OA) = N(A)$より、(Y, ϑ_m)の条件下で$n(A)$は二項分布

$B(N(A), \frac{\vartheta_m}{\vartheta_m + (1-\vartheta_m)\vartheta_m})$に従う。よって

$$n_m(A) = E\left[n(A)\middle|Y, \vartheta_m\right] = N(A) \times \frac{\vartheta_m}{\vartheta_m + (1-\vartheta_m)\vartheta_m} = \frac{k\vartheta_m}{\vartheta_m + (1-\vartheta_m)\vartheta_m}$$

$$n_m(OA) = E\left[n(OA)\middle|Y, \vartheta_m\right] = N(A) \times \frac{(1-\vartheta_m)\vartheta_m}{\vartheta_m + (1-\vartheta_m)\vartheta_m} = \frac{k(1-\vartheta_m)\vartheta_m}{\vartheta_m + (1-\vartheta_m)\vartheta_m}$$

以上より

$$\begin{aligned}Q(\vartheta|\vartheta_m) &= n_m(A)\log\vartheta + n_m(OA)\log\left((1-\vartheta)\vartheta\right) + 2N(O)\log(1-\vartheta) \\ &\qquad + E\left[\log\left(\frac{n!}{n(A)!n(OA)!n(OO)!}\right)\middle|\, Y, \vartheta_m\right]\end{aligned}$$

最後の項$E[\log(\cdots)|\, Y, \vartheta_m]$にはパラメーター$\vartheta$は含まれないことに注意しよう。

$Q(\vartheta|\vartheta_m)$が最大になるϑを求める。

$\frac{\partial Q}{\partial\vartheta} = \frac{n_m(A)}{\vartheta} + n_m(OA)\left(-\frac{1}{1-\vartheta} + \frac{1}{\vartheta}\right) - 2N(O)\frac{1}{1-\vartheta} = 0$ より

$n_m(A) + n_m(OA) - \{n_m(A) + 2n_m(OA) + 2N(O)\}\vartheta = 0$

これから　$\vartheta_{m+1}=\dfrac{n_m(A)+n_m(OA)}{n_m(A)+2n_m(OA)+2N(O)}=\dfrac{N(A)}{2n-n_m(A)}$

分母の$n_m(A)+2n_m(OA)+2N(O)$は n 人が引いたくじの総数のｍステップ目の推定値であり、分子は引いた当たりくじの数である。$n_m(A)$を代入すると次の漸化式を得る

$\vartheta_{m+1}=\dfrac{k(2-\vartheta_m)}{(4n-k)-2n\vartheta_m}$。極限値$\lim_{m\to\infty}\vartheta_m=\vartheta$とすると、$\vartheta=\dfrac{k(2-\vartheta)}{(4n-k)-2n\vartheta}$。$0\le\vartheta\le1$より、$\vartheta=1-\sqrt{1-\dfrac{k}{n}}$となる。

（注意）　この問題では２回のチャンスで当る確率は$\vartheta+(1-\vartheta)\vartheta$、２回ともはずれの確率は$(1-\vartheta)^2$なので、データの尤度関数は　$g(Y|\vartheta)={}_nC_k\left(\vartheta+(1-\vartheta)\vartheta\right)^k\left((1-\vartheta)^2\right)^{n-k}$となる。

$\log g(Y|\vartheta)=\log{}_nC_k+k\log\vartheta+k\log(2-\vartheta)+2(n-k)\log(1-\vartheta)$、　極値は

$\dfrac{\partial}{\partial\vartheta}\log g(Y|\vartheta)=\dfrac{k}{\vartheta}+\dfrac{k}{\vartheta-2}+\dfrac{2(n-k)}{\vartheta-1}=0$より$n\vartheta^2-2n\vartheta+k=0,\ (0\le\vartheta\le1)$。故に

$\vartheta=1-\sqrt{1-\dfrac{k}{n}}$と最尤推定値が直接もとまる。このような簡単な尤度関数の場合は直接求めることができるが、複雑な尤度関数の場合にも、逐次的に最尤推定値に近づけて行くのがEM アルゴリズムである。

＜例２＞　遺伝子頻度の推定

仮定：Hardy-Weinberg の法則は成立していると仮定する。

ABO 血液型

不完全（観測）データY：各血液型のサンプル数　$n(A),n(B),n(AB),n(O)$

$$n(A)+n(B)+n(AB)+n(O)=n$$

完全データX：遺伝子型のサンプル数　$n(AA),n(AO),n(BB),n(BO),n(AB),n(OO)$

$$n(AA)+n(AO)+n(BB)+n(BO)+n(AB)+n(OO)=n$$

パラメーター：遺伝子頻度　$\vartheta=(p(A),p(B),p(O)),\quad p(A)+p(B)+p(O)=1$

完全データXの尤度関数(likelihood)

$$f(X|\vartheta)=\frac{n!}{n(AA)!n(AO)!n(BB)!n(BO)!n(AB)!n(OO)!}(p(A)^2)^{n(AA)}(P(B)^2)^{n(BB)}(p(O)^2)^{n(OO)}$$
$$\times(2p(A)p(O))^{n(AO)}(2p(B)p(O))^{n(BO)}(2p(A)p(B))^{n(AB)}$$

$\vartheta_m=(p_m(A),p_m(B),p_m(O))$：第ｍステップでのパラメーターの値

$$Q(\vartheta|\vartheta_m) = E\left[\log f(X|\vartheta)\middle|Y=(n(A),n(B),n(AB),n(O)),\vartheta_m\right]$$
$$= E\left[\begin{array}{l} 2n(AA)\log p(A) + 2n(BB)\log p(B) + 2n(OO)\log p(O) + n(AO)\log(2p(A)p(O)) \\ + n(BO)\log(2p(B)p(O)) + n(AB)\log(2p(A)p(B)) + \log\left(\dfrac{n!}{n(AA)!\cdots n(OO)!}\right) \end{array} \middle| Y,\vartheta_m\right]$$

AB 型と O 型の血液型の数は観測データ Y から明らかに

$$n_m(AB) = E\left[n(AB)\middle|Y,\vartheta_m\right] = n(AB), \quad n_m(OO) = E\left[n(OO)\middle|Y,\vartheta_m\right] = n(O)$$

完全データ X の尤度は多項分布なので、Hardy-Weinberg の法則は成り立つという仮定の下で、条件Y,ϑ_mの下で、遺伝子型 AA および AO の数の条件付期待値は

$$n_m(AA) = E\left[n(AA)\middle|Y,\vartheta_m\right] = n(A)\times\frac{Freq\ of\ AA}{Freq\ of\ Atype} = n(A)\times\frac{(p_m(A))^2}{(p_m(A))^2 + 2p_m(A)p_m(O)}$$

$$n_m(AO) = E\left[n(AO)\middle|Y,\vartheta_m\right] = n(A)\times\frac{2p_m(A)p_m(O)}{(p_m(A))^2 + 2p_m(A)p_m(O)}$$

A を B に置き換えれば同様に$n_m(BB), n_m(BO)$も求められる。

$$\begin{aligned} Q(\vartheta|\vartheta_m) &= 2n_m(AA)\log p(A) + 2n_m(BB)\log p(B) + 2n_m(OO)\log p(O) \\ &\quad + n_m(AO)(\log 2 + \log p(A) + \log p(O)) + n_m(BO)(\log 2 + \log p(B) + \log p(O)) \\ &\quad + n_m(AB)(\log 2 + \log p(A) + \log p(B)) + E\left[\log\left(\frac{n!}{n(AA)!\cdots n(OO)!}\right)\middle|Y,\vartheta_m\right] \end{aligned}$$

最後の項の期待値はパラメーターϑを含まないので次のステップのϑの推定に関係しない。

$Q(\vartheta|\vartheta_m)$を最大にするϑの値がϑ_{m+1}となる。ラグランジュ未定係数法によって求める。

$H(\vartheta,\lambda) = Q(\vartheta|\vartheta_m) + \lambda(p(A)+p(B)+p(O)-1)$　とする。

$$\frac{\partial}{\partial p(A)}H = \frac{2n_m(AA)}{p(A)} + \frac{n_m(AO)}{p(A)} + \frac{n_m(AB)}{p(A)} + \lambda = 0 \quad \cdots(1)$$

$$\frac{\partial}{\partial p(B)}H = \frac{2n_m(BB)}{p(B)} + \frac{n_m(BO)}{p(B)} + \frac{n_m(AB)}{p(B)} + \lambda = 0 \quad \cdots(2)$$

$$\frac{\partial}{\partial p(O)}H = \frac{2n_m(OO)}{p(O)} + \frac{n_m(AO)}{p(O)} + \frac{n_m(BO)}{p(O)} + \lambda = 0 \quad \cdots(3)$$

$$\frac{\partial}{\partial \lambda}H = p(A)+p(B)+p(O)-1 = 0 \quad \cdots(4)$$

(1),(2),(3)より

$$p(A)=(2n_m(AA)+n_m(AO)+n_m(AB))/(-\lambda),\ \ p(B)=(2n_m(BB)+n_m(BO)+n_m(AB))/(-\lambda)$$
$$p(O)=(2n_m(OO)+n_m(AO)+n_m(BO))/(-\lambda)$$

(4)に代入して　$\lambda=-2(n_m(OO)+n_m(AA)+\cdots)=-2n$。

これより

$$p_{m+1}(A)=\frac{2n_m(AA)+n_m(AO)+n_m(AB)}{2n},\ \ p_{m+1}(B)=\frac{2n_m(BB)+n_m(BO)+n_m(AB)}{2n}$$
$$p_{m+1}(O)=\frac{2n_m(OO)+n_m(AO)+n_m(BO)}{2n}$$

Gene counting

＜まとめ＞

血液型　　観測データ：　$Y=\{n(A),n(B),n(AB),n(O)\}$

　　パラメータ：　$\vartheta=\{p(A),p(B),p(O)\}$　遺伝子頻度

初期推定値：　$\vartheta(0)=\{p_0(A),p_0(B),p_0(O)\}$　任意の推定値

第ｍステップ：　$\vartheta_m=\{p_m(A),p_m(B),p_m(O)\}$

$$n_m(AA)=n(A)\times\frac{(p_m(A))^2}{(p_m(A))^2+2p_m(A)p_m(O)},$$
$$n_m(AO)=n(A)\times\frac{2p_m(A)p_m(O)}{(p_m(A))^2+2p_m(A)p_m(O)}$$

などにより各遺伝子型の数を推定

第ｍ＋１ステップ：　$\vartheta_{m+1}=\{p_{m+1}(A),p_{m+1}(B),p_{m+1}(O)\}$

$$p_{m+1}(A)=\frac{2n_m(AA)+n_m(AO)+n_m(AB)}{2n}$$

など Gene counting により

$p_{m+1}(B),\ p_{m+1}(O)$ も求める。

＜例３＞　Haplotype 頻度の推定

　二つの遺伝子座　：　対立遺伝子　遺伝子座１$\{A,a\}$、　遺伝子座２$\{B,b\}$

Hardy-Weinberg 平衡を仮定、連鎖平衡は仮定しない。

Haploid 頻度：$P(AB),P(Ab),P(aB),P(ab)$　を推定する。

サンプル個体数$=n$

不完全（観測）データ Y：$n(AABB),\ n(AABb),n(AAbb),n(AaBB),n(AaBb),$
$n(Aabb),n(aaBB),n(aaBb),n(aabb)$

完全データ X:　$n\left(\frac{AB}{AB}\right),n\left(\frac{AB}{Ab}\right),n\left(\frac{AB}{aB}\right),n\left(\frac{AB}{ab}\right),n\left(\frac{Ab}{aB}\right),n\left(\frac{Ab}{Ab}\right),n\left(\frac{Ab}{ab}\right),n\left(\frac{aB}{aB}\right)$
$n\left(\frac{aB}{ab}\right),n\left(\frac{ab}{ab}\right)$

不完全データでは$n(AaBb)$と表示される両遺伝子座ヘテロ接合体の個体が完全データでは$n\left(\frac{Ab}{aB}\right), n\left(\frac{AB}{ab}\right)$と相が区別されている。

パラメーター　$\vartheta = \left(P(AB), P(Ab), P(aB), P(ab)\right)$

完全データの尤度関数(likelihood)　多項分布により

$$f(X|\vartheta) = \frac{n!}{n\left(\frac{AB}{AB}\right)!\cdots n\left(\frac{ab}{ab}\right)!}\left(P^2(AB)\right)^{n\left(\frac{AB}{AB}\right)}\left(2P(AB)P(Ab)\right)^{n\left(\frac{AB}{Ab}\right)}\cdots\left(P^2(ab)\right)^{n\left(\frac{ab}{ab}\right)}$$

$$Q\left(\vartheta|\vartheta_m\right) = E\left[\log f(X|\vartheta)\,|Y,\vartheta_m\right] = E\left[\begin{array}{l} 2n\left(\frac{AB}{AB}\right)\log P(AB) + n\left(\frac{AB}{Ab}\right)\log\left(2P(AB)P(Ab)\right) + \\ \cdots + 2n\left(\frac{ab}{ab}\right)\log P(ab) + \log\frac{n!}{n\left(\frac{AB}{AB}\right)!\cdots n\left(\frac{ab}{ab}\right)!}\,|\,Y,\vartheta_m \end{array}\right]$$

Double hetero $n\left(\frac{AB}{ab}\right), n\left(\frac{Ab}{aB}\right)$以外の遺伝子頻度は不完全データと完全データの違いはないので、例えば$n\left(\frac{AB}{AB}\right) = n(AABB),\ n\left(\frac{AB}{Ab}\right) = n(AABb)$など全て観測データから求まる。

故に第ｍステップでの Double hetero $n\left(\frac{AB}{ab}\right), n\left(\frac{Ab}{aB}\right)$以外の遺伝子型の数については

$n_m\left(\frac{AB}{AB}\right) = E\left[n\left(\frac{AB}{AB}\right)|Y,\vartheta_m\right] = n(AABB)$等となる。

Double hetero $n\left(\frac{AB}{ab}\right), n\left(\frac{Ab}{aB}\right)$については多項分布の性質より

$$n_m\left(\frac{AB}{ab}\right) = E\left[n\left(\frac{AB}{ab}\right)|\,Y,\vartheta_m\right] = n(AaBb)\times\frac{2P_m(AB)P_m(ab)}{2P_m(AB)P_m(ab)+2P_m(Ab)P_m(aB)}$$

$$n_m\left(\frac{Ab}{aB}\right) = E\left[n\left(\frac{Ab}{aB}\right)Y,\vartheta_m\right] = n(AaBb)\times\frac{2P_m(Ab)P_m(aB)}{2P_m(AB)P_m(ab)+2P_m(Ab)P_m(aB)}$$

例２と同様にしてラグランジュの未定乗数法により、各ハプロイド頻度の推定値は

$$P_{m+1}(AB) = \frac{1}{2n}\left\{2n(AABB) + n(AABb) + n(AaBB) + n_m\left(\frac{AB}{ab}\right)\right\}$$

$$P_{m+1}(Ab) = \frac{1}{2n}\left\{n(AABb) + 2n(AAbb) + n_m\left(\frac{Ab}{aB}\right) + n(Aabb)\right\}$$

$$P_{m+1}(aB)=\frac{1}{2n}\left\{n(AaBB)+n_m\left(\frac{aB}{Ab}\right)+2n(aaBB)+n(aaBb)\right\}$$

$$P_{m+1}(ab)=\frac{1}{2n}\left\{n_m\left(\frac{ab}{AB}\right)+n(Aabb)+n(aaBb)+2n(aabb)\right\}$$

また A,B の遺伝子頻度は

$$P_m(A)=P_m(AB)+P_m(Ab)$$
$$=\frac{1}{2n}\{2n(AABB)+2n(AABb)+2n(AAbb)+n(AaBB)+n(AaBb)+n(Aabb)\}$$
$$P_m(B)=P_m(AB)+P_m(aB)$$
$$=\frac{1}{2n}\{2n(AABB)+2n(AaBB)+2n(aaBB)+n(AABb)+n(AaBb)+n(aaBb)\}$$

これらは観測データから決まるのでmに依存しない定数である。

$P_m(AA)+P_m(Ab)+P_m(aB)+P_m(ab)=1$なので、３変数を推定すれば十分である。

＜参考＞　ラグランジュの未定乗数法

二つのC^1級関数$F(x,y),G(x,y)$に対して、$H(x,y,\lambda)=F(x,y)-\lambda G(x,y)$と置く。条件$G(x,y)=0$の下で、関数$F(x,y)$は点$(x_0,y_0)$で極値を取るとする。さらに、

$G_x(x_0,y_0),\ G_y(x_0,y_0)$の少なくとも一つが０でないならば、

$$H_x(x_0,y_0,\lambda)=F_x(x_0,y_0)-\lambda G_x(x_0,y_0)=0,$$
$$H_y(x_0,y_0,\lambda)=F_y(x_0,y_0)-\lambda G_y(x_0,y_0)=0$$

を満たすλが存在する。

（証明）条件より$G_x(x_0,y_0),\ G_y(x_0,y_0)$の少なくとも一つが０でないので$G_y(x_0,y_0)\neq 0$

の場合を示す。このとき陰関数の定理より点x_0の近傍でC^1級関数$y=f(x)$で

$G(x,f(x))=0$および$f(x_0)=y_0$を満たすものが存在する。$\varphi(x)=F(x,f(x))$とすると、

$\varphi(x)$は点x_0で極値を取るので$\varphi'(x_0)=0$、すなわち

$F_x(x_0,f(x_0))+F_y(x_0,f(x_0))f'(x_0)=0,\quad G(x,f(x))=0$の両辺をｘで微分すると

$G_x(x_0,f(x_0))+G_y(x_0,f(x_0))f'(x_0)=0$。二つの式から$f'(x_0)$を消去すると

$F_x(x_0, f(x_0)) - F_y(x_0, f(x_0))\dfrac{G_x(x_0, y_0)}{G_y(x_0, y_0)} = 0$。 $\dfrac{F_y(x_0, y_0)}{G_y(x_0, y_0)} = \dfrac{F_x(x_0, y_0)}{G_x(x_0, y_0)} = \lambda$ と置くと、

$F_x(x_0, f(x_0)) - \lambda G_x(x_0, f(x_0)) = 0, \quad F_y(x_0, f(x_0)) - \lambda G_y(x_0, f(x_0)) = 0$ が成り立つ。

λ を未定乗数と言い、この方法をラグランジュの未定乗数法と言う。（証明終わり）

例題： 条件 $x^2 + y^2 = 1$ の下で、関数 $x^2 y^3$ の最大値、最小値を求めよ。

（解） $H(x, y, \lambda) = F(x, y) - \lambda G(x, y) = x^2 y^3 - \lambda(x^2 + y^2 - 1)$ とする。

条件 $G(x, y) = x^2 + y^2 - 1 = 0$ と $G_x(x, y) = 2x = 0,\ G_y(x, y) = 2y = 0$ を同時に満たす (x, y) はないので、

$$H_x(x_0, y_0, \lambda) = 2x_0 {y_0}^3 - 2\lambda x_0 = 0,$$
$$H_y(x_0, y_0, \lambda) = 3{x_0}^2 {y_0}^2 - 2\lambda y_0 = 0$$

を満たす λ が存在する。

$x_0 \neq 0,\ y_0 \neq 0$ のとき、$2\lambda = 2{y_0}^3 = 3{x_0}^2 y_0$, 故に $2{y_0}^2 = 3{x_0}^2$。 ${x_0}^2 + {y_0}^2 = 1$ なので

${x_0}^2 = \dfrac{2}{5},\ {y_0}^2 = \dfrac{3}{5}$ すなわち $(x_0, y_0) = \left(\sqrt{\dfrac{2}{5}}, \pm\sqrt{\dfrac{3}{5}}\right), \left(-\sqrt{\dfrac{2}{5}}, \pm\sqrt{\dfrac{3}{5}}\right)$、

このとき、$F\left(\pm\sqrt{\dfrac{2}{5}}, \sqrt{\dfrac{3}{5}}\right) = \dfrac{6}{25}\sqrt{\dfrac{3}{5}}, \quad F\left(\pm\sqrt{\dfrac{2}{5}}, -\sqrt{\dfrac{3}{5}}\right) = -\dfrac{6}{25}\sqrt{\dfrac{3}{5}}$ 。

$x_0 = 0$ または $y_0 = 0$ のとき $F(x_0, y_0) = 0$、よって

最大値 $F\left(\pm\sqrt{\dfrac{2}{5}}, \sqrt{\dfrac{3}{5}}\right) = \dfrac{6}{25}\sqrt{\dfrac{3}{5}}$, 最小値 $F\left(\pm\sqrt{\dfrac{2}{5}}, -\sqrt{\dfrac{3}{5}}\right) = -\dfrac{6}{25}\sqrt{\dfrac{3}{5}}$ 。