Multiple Ascertainment π가 存在할 때 分離確率母數 θ値의 尤度推定値로서 統計模型의 構成과 遺傳病에 感染된 出生兒의 豫測

申 漢 豊*

序 論

遺傳的 질환이 있는 家系에서 出生하는 子女中에서 遺傳的인 질환을 保有할 수 있는 確率을 豫測하는 방법의 하나로서 尤度推定値(likelihood estimator)를 사용하는 것은 分離分析(segregation analysis)에서 중요한 역활을 하고 있다. Elston과 Stewart(1971)는 이러한 분석방법의 일반적인 통계모형을 定立하였으며 筆者(1974)와 Morton등(1974)은 complex segregation이 될 때에 분석되는 4가지의 統計模型을 주장하였다.

본 연구의 목적은 multiple ascertainment π가 존재하는 경우 分離確率母數(segregation parameter) θ의 尤度推定値를 구하고 둘째로 oligogenic special case에 대한 理論的인 배경을 究明하고자 한다.

理 論

1. 최우추정법(maximum likelihood estimator of θ)

數學的 模型을 最尤推定量(maximum likelihood estimator)으로 檢定하려면 式 (1)을(〈부록 1〉참조)

$$P_{rs}(\theta)=\frac{\binom{s}{r}\theta^r(1-\theta)^{s-r}[1-(1-\pi)^r]}{1-(1-\pi\theta)^s} \tag{1}$$

多項分布(multinomial distribution)로 變換하여

---

* 高麗大學校 統計學科 敎授. 本硏究는 1976年度 産學協同財團 硏究費에 의하여 이루어진 것임을 밝힌다.

$$P_r\{a_{1s},\ a_{2s},\cdots,a_{ss};\theta\}=\frac{n_s!}{a_{1s}!\ a_{2s}!\cdots a_{ss}!}P_{1s}^{a_{1s}}(\theta)\cdots P_{ss}^{a_{ss}}(\theta)$$

$$=n_s!\prod_{r=1}^{s}\frac{P_{rs}^{a_{rs}}(\theta)}{a_{rs}!} \tag{2}$$

式 (2)는 θ의 函數로 되어 尤度(likelihood), $L_s(\theta)$가 된다.

즉 $$L_s(\theta)=\frac{n_s!}{\prod_{r=1}^{s}a_{rs}!}\prod_{r=1}^{s}\left\{[1-(1-\pi)^r]\binom{s}{r}\frac{\theta^r(1-\theta)^{s-r}}{1-(1-\pi\theta)^s}\right\}^{a_{rs}} \tag{3}$$

式 (3)의 $L_s(\theta)$에 log를 取하면

$$\log L_s(\theta)=\text{const}+\sum_{r=1}^{s}\{r\log\theta+(s-r)\log(1-\theta)-\log[1-(1-\pi\theta)^s]\}a_{rs} \tag{4}$$

式 (4)를 미분하여 $L_s(\theta)$를 $U_s(\theta)$로 바꾸면

$$\frac{\partial\log L_s(\theta)}{\partial\theta}=U_s(\theta)=\sum_{r=1}^{s}\left[\frac{r}{\theta}-\frac{(s-r)}{1-\theta}-\frac{\pi s(1-\pi\theta)^{s-1}}{1-(1-\pi\theta)^s}\right]a_{rs}$$

$$=\sum_{r=1}^{s}\left[\frac{r}{\theta(1-\theta)}-\frac{s}{1-\theta}-\frac{\pi s(1-\pi\theta)^{s-1}}{1-(1-\pi\theta)^s}\right]a_{rs}$$

$$=\frac{1}{\theta(1-\theta)}\sum_{r=1}^{s}ra_{rs}-\frac{s}{1-\theta}\sum_{r=1}^{s}a_{rs}-\frac{\pi s(1-\pi\theta)^{s-1}}{1-(1-\pi\theta)^s}\sum_{r=1}^{s}a_{rs}$$

$$=\frac{r_s}{\theta(1-\theta)}-\frac{Sn_s}{1-\theta}-\frac{\pi Sn_s(1-\pi\theta)^{s-1}}{1-(1-\pi\theta)^s}$$

$$=\frac{r_s}{\theta(1-\theta)}-\frac{Sn_s[1-(1-\pi\theta)^s-\pi(1-\theta)(1-\pi\theta)^{s-1}]}{(1-\theta)[1-(1-\pi\theta)^s]}$$

$$=\frac{1}{1-\theta}\left[\frac{r_s}{\theta}-\frac{1-(1-\pi)(1-\pi\theta)^{s-1}}{1-(1-\pi\theta)^s}t_s\right] \tag{5}$$

全體 標本에 대한 $U(\theta)$의 값은

$$U(\theta)=\sum_{s=1}^{S}U_s(\theta)=\frac{1}{1-\theta}\left[\frac{R}{\theta}-\sum_{s=1}^{S}\frac{1-(1-\pi)(1-\pi\theta)^{s-1}}{1-(1-\pi\theta)^s}t_s\right] \tag{6}$$

θ의 ML 推定値를 求하면

$$\frac{R}{\theta}=\sum_{s=1}^{S}\frac{1-(1-\pi)(1-\pi\theta)^{s-1}}{1-(1-\pi\theta)^s}t_s \tag{7}$$

$V_{ar}(\hat{\theta})$을 計算하기 위하여 a_{rs}의 期待値를 求하면

$E(a_{rs})=n_sP_{rs}(\theta)$, $r=1,2,\cdots,s$에 式 (1)을 代入

$$E(a_{rs}) = n_s \frac{\binom{s}{r}\theta^r(1-\theta)^{s-r}[1-(1-\pi)^r]}{1-(1-\pi\theta)^s}, \quad r=1,2,\cdots,s \tag{8}$$

r_s의 期待値는

$$E(r_s) = \sum_{r=1}^{s} rE(a_{rs})$$

$$= \frac{n_s}{1-(1-\pi\theta)^s}\left\{\sum_{r=1}^{s} r\binom{s}{r}\theta^r(1-\theta)^{s-r} - \sum_{r=1}^{s} r\binom{s}{r}[(1-\pi)\theta]^r(1-\theta)^{s-r}\right\} \tag{9}$$

式(9)의 대괄호안의 첫째 項이 二項分布의 平均이 된다.

$$\text{즉} \quad \sum_{r=1}^{s} r\binom{s}{r}\theta^r(1-\theta)^{s-r} = s\theta \tag{10}$$

式(9)의 두번째 項을 展開하면

$$r\binom{s}{r} = s\binom{s-1}{r-1}$$

$$\sum_{r=1}^{s} r\binom{s}{r}[(1-\pi)\theta]^r(1-\theta)^{s-r} = s(1-\pi)\theta\sum_{r=1}^{s}\binom{s-1}{r-1}[(1-\pi)\theta]^{r-1}(1-\theta)^{s-r}$$

$$= s\theta(1-\pi)(1-\pi\theta)^{s-1} \tag{11}$$

式(9)에 (10)과 (11)을 代入하면

$$E(r_s) = \frac{s\theta n_s}{1-(1-\pi\theta)^s}[1-(1-\pi)(1-\pi\theta)^{s-1}] \tag{12}$$

期待情報量(the expected amount of information) $I(\theta)$와 分散 $V_{ar}(\hat{\theta})$은

$$\frac{\partial \log L_s(\theta)}{\partial\theta} = U_s(\theta) = \frac{r_s}{\theta(1-\theta)} - \frac{Sn_\cdot}{1-\theta} - \frac{\pi S n_s(1-\pi\theta)^{s-1}}{1-(1-\pi\theta)^s} \tag{13}$$

따라서

$$-\frac{\partial^2 \log L_s(\theta)}{\partial\theta^2} = -\frac{\partial U_s(\theta)}{\partial\theta}$$

$$= -\frac{r_s(2\theta-1)}{\theta^2(1-\theta)^2} + \frac{Sn_\cdot}{(1-\theta)^2}$$

$$-\frac{\pi^2 S n_s(s-1)(1-\pi\theta)^{s-2}[1-(1-\pi\theta)^s]}{[1-(1-\pi\theta)^s]^2}$$

$$-\frac{\pi^2 S s n_s(1-\pi\theta)^{s-1}(1-\pi\theta)^{s-1}}{[1-(1-\pi\theta)^s]^2}$$

$$= -\frac{r_s(2\theta-1)}{\theta^2(1-\theta)^2} + \frac{Sn_\cdot}{(1-\theta)^2}$$

$$-\frac{\pi^2 S n_s(1-\pi\theta)^{s-2}}{[1-(1-\pi\theta)^s]^2}\{s-[1-(1-\pi\theta)^s]\} \tag{14}$$

즉 $I_s(\theta)$, 期待情報量은 式(14)에 (12)을 代入하여 求한다.

$$I_s(\theta)=-\frac{\theta s n_s[1-(1-\pi)(1-\pi\theta)^{s-1}](2\theta-1)}{[1-(1-\pi\theta)^s]\theta^2(1-\theta)^2}$$

$$+\frac{sn_s}{(1-\theta)^2}-\frac{\pi^2 S n_s(1-\pi\theta)^{s-2}}{[1-(1-\pi\theta)^s]^2}\{s-[1-(1-\pi\theta)^s]\} \tag{15}$$

그리고 total expected amount of information, $I(\theta)$는

$$I(\theta)=\sum_{s=1}^{S} I_s(\theta) \tag{16}$$

그러므로 $\hat{\theta}$의 分散인 $V_{ar}(\hat{\theta})$은

$$V_{ar}(\hat{\theta})=\frac{1}{I(\theta)} \tag{17}$$

2. Oligogenic 確率模型에 관한 尤度

i번째 개체에 대하여 관찰된 b_{il}의 尤度는 다음과 같이 표현할 수 있다(〈부록 2〉참조).

$$\alpha(z_i,b_{il})=[\pi_l(z_i)]^{b_{il}}[1-\pi_l(z_i)]^{1-b_{il}}$$

여기서 $\alpha(z_i,b_{il})=\pi_l(z_i)$ $b_{il}=1$이면

$=1-\pi_l(z_i)$ $b_{il}=0$이면

만약 r개의 독립적인 ascertainment방법을 생각하면 다음과 같다.

$$\alpha(\underset{\sim}{z},\ \underset{\sim}{b})=\prod_{l=1}^{r}\alpha(z_0,b_{0l})\prod_{l=1}^{d}\alpha(z_e,b_{el})\prod_{j=1}^{n^e}\alpha(z_{ej},b_{ejl})$$

위의 식을 두 세대(2-generational family)의 尤度로 표현하면

$$\frac{\alpha(\underset{\sim}{z},\underset{\sim}{b})\sum_{s=1}^{k_{h0}}\psi_{hos}g_{hos}(z_o)\prod_{e=1}^{d}\left\{\sum_{t=1}^{k_{he}}\psi_{het/s}g_{het}(z_e)\prod_{j=1}^{n^e}\left[\sum_{u=1}^{k_{hej}}p_{stu}h_{ej}g^{h_{ej}u}(z_{ej})\right]\right\}}{1-\beta(g,\pi)} \tag{18}$$

여기서 $\beta(g,\pi)=$ 임의로 추출한 家系의 확률(proband가 없는 경우)

1) 正規表現型分布(normal phenotypic distribution)

μ_{hiu}가 정규표현형분포의 평균이고 분산이 σ^2인 경우

$$g_{hiu}(z_i)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{1}{2}\left(\frac{z_i-\mu_{hiu}}{\sigma}\right)^2\right]=\phi(z_i-\mu_{hiu},\sigma^2) \tag{19}$$

z_i대신에 만약 i번째의 개체에 대하여 $z_{i_1},z_{i_2},\cdots z_{i_v}$의 관측치 벡타를 갖는다면 $P_{n_i}(z_{ij})$가 polynomial인 $Z_i=P_{n_1}(z_{i_1})+P_{n_2}(z_{i_2})+\cdots+P_{n_v}(z_{i_v})$가 된다. 따라서 $g_{hiu}(Z_r)$도 式(19)와 같이 一變數正規分布(univariate normal distribution)가 된다. 특별히 나이를 관측치로 하는 경우에는 polynomial regression에 의한 age correction이 된다.

2) 베르누이 表現型分布(Bernoulli phenotypic distribution)

正常(normal)과 非正常(affected)을 0과 1로 놓으면

$$g_{hiu}(z_i)=(f_{hiu})^{z_i}(1-f_{hiu})^{(1-z_i)} \tag{20}$$

여기서 f_{hiu}는 sex h_i와 因子型(genotype) u가 갖는 表現型(phenotype) z_i의 확률이다.

3) 정규분포의 평균과 분산이 因子型에 의하여 구해지므로 密度函數 $\phi(a_i-\mu_{hiu},\sigma^2{}_{hiu})$로 표현할 수 있다. 非正常인 경우에 a_i는 age of onset이 되고 a_i'은 current age가 된다. 그러면 $g_{hiu}(z_i,a_i',a_i)$인 表現型分布는 다음과 같이 표시된다.

$$g_{hiu}(1,a_i',a_i)=\gamma_{hiu}\ \phi(a-\mu_{hiu},\sigma^2{}_{hiu})$$

$$g_{hiu}(0,a_i',a_i)=1-\ \gamma_{hiu}\ \phi\left(\frac{a_i-\mu_{hiu}}{\sigma_{hiu}}\right), \tag{21}$$

여기서 $$\phi(a)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{t^2}{2}\right)dt \tag{22}$$

위의 식에서 母數 γ는 f와 類似하며 f가 penetrance인데 비하여 γ는 susceptibility를 측정하는 것이 된다.

3) Ascertainment 函數

asertammont 函數 α가 $\pi_l(z_i)$의 확률로서 계산되고. $\pi_l(z_i)$에 대하여 3가지의 확률모형이 있다.

a) Threshold Model

$$\pi_l(z_i)=\pi_l \quad z_i\geqq T \text{ 이면}$$
$$=0 \quad z_i<T \text{ 이면} \tag{23}$$

여기서 z_i가 Bernoulli variable이며 $T=1$이면 ascertainment모형이 된다. 이 경우에 $\pi\to 0$이면 single selection, $0<\pi<1$이면 multiple selection, $\pi=1$이면 complete selection이 된다.

$B_{lT}=\sum b_{il}$과 $B_{lL}=\sum(1-b_{il})$로 놓으면 먼저 정의한 $\alpha(\underset{\sim}{z},\underset{\sim}{b})$를 다음과 같이 표현할 수 있다.

$$\alpha(\underset{\sim}{z},\underset{\sim}{b})=\prod_{l=1}^{r}\pi_l^{B_{lT}}(1-\pi_l)^{B_{lL}}$$

이 식은 계산이 편리한 점이 있다.

b) Exponential Model

연속형의 자료에 대한 모형으로서 다음과 같이 定義된다.

$$\pi_l(z_i)=\min[\exp(k_0+k_1z_i+k_2z_i^2),1] \tag{24}$$

여기서 $k_2=0$이거나 또는 $k_2<0$ 그리고 $k_1^2\leqq 4k_0k_2$

c) Age of Onset-Dependent Threshold Model

proband가 되는 확률을 $\pi_l(z_i,a_i)$로 놓으면

$$\pi_l(z_i,a_i)=\min[\exp(k_0+k_1a_i+k_2a_i^2),1]$$

if $z_i=1$ and either $k_2=0$, or $k_2<0$ and $k_1^2\leqslant 4k_0k_2$ (25)

4) Sampling Correction

式(18)의 尤度의 z_i에 대한 重積分 또는 合計는 β로서 sampling correction이다. 계산의 편의상 순서를 바꾸면

$$\sum_{s=1}^{k_{h0}}\psi_{h_0s}\zeta_{h_0s}\prod_{e=1}^{d}\left[\sum_{t=1}^{k_{he}}\psi_{h_e t's}\zeta_{h_e t}\prod_{j=1}^{n_e}\left(\sum_{u=1}^{k_{he'}}p_{stu}{}^{he'j}\zeta_{he'ju}\right)\right]$$

여기서 ζ_{hiu}는 모든 z_i에 대한 積分 또는 合計임.

$$\prod_{l=1}^{r}\alpha(z_i,0)g_{hiu}(z_i)=\prod_{l=1}^{r}[1-\pi_l(z_i)]g_{hiu}(z_i) \tag{26}$$

ζ_{hiu}는 因子型 u를 갖는 h_i의 확률로서 proband가 되지않고 또한 이 확률은 g와 π가 결정됨에 따라 값이 결정된다. 이것을 誘導하여 보면 다음과 같다.

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{a'}\exp\left[-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2+k_0+k_1z+k_2z^2\right]dz\\
&=\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{a'}\exp\left\{-\frac{1}{2\sigma^2}[(1-2k_2\sigma^2)z^2-2(\mu+k_1\sigma^2)z+\mu^2-2k_0\sigma^2]\right\}dz\\
&=\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{a'}\exp\left[-\frac{1-2k_2\sigma^2}{2\sigma^2}\left(z-\frac{\mu+k_1\sigma^2}{1-2k_2\sigma^2}\right)^2\right]dz\cdot\exp\left[k_0-\frac{\mu^2}{2\sigma^2}+\frac{(\mu+k_1\sigma^2)^2}{2\sigma^2(1-2k_2\sigma^2)}\right]\\
&=(1-2k_2\sigma^2)^{-\frac{1}{2}}\phi\left[\frac{1}{\sigma}\left(a'\sqrt{1-2k_2\sigma^2}-\frac{\mu+k_1\sigma^2}{\sqrt{1-2k_2\sigma^2}}\right)\right]\\
&\qquad\cdot\exp\left\{\frac{1}{2\sigma^2}\left[2k_0\sigma^2-\mu^2+\frac{(\mu+k_1\sigma^2)^2}{1-2k_2\sigma^2}\right]\right\}
\end{aligned} \tag{27}$$

여기서 $k_2=0$이면

$$\phi\left(\frac{a'-\mu-k_1\sigma^2}{\sigma}\right)\exp\left(k_0+\mu k_1+\frac{1}{2}k_1{}^2\sigma^2\right) \tag{28}$$

표현형 분포(phenotypic distribution)와 ascertainment함수에서 논의하였던 g와 π에 관한것을 ζ로서 유도하여 보면

a) Normal g and Threshold Model for π

式(26), (19), (23)을 사용하여

$$\zeta_{hiu}=\int_{-\infty}^{\infty}\prod_{l=1}^{r}[1-\pi_l(z)]g_{hiu}(z)dz$$

$$=\int_{T}^{\infty}\prod_{l=1}^{r}(1-\pi_l)\phi(z-\mu_{hiu},\sigma^2)dz+\int_{-\infty}^{T}\prod_{l=1}^{r}(1)\phi(z-\mu_{hiu},\sigma^2)dz$$

式(22)에서

$$\zeta_{hiu}=\left[1-\phi\left(\frac{T-\mu_{hiu}}{\sigma}\right)\right]\prod_{l=1}^{r}(1-\pi_l)+\phi\left(\frac{T-\mu_{hiu}}{\sigma}\right)$$

b) Normal g and Exponential Model for π $(r=1)$

式(26), (19) 및 (24)에서 $\exp(k_0+k_1z)=1$이 $z=k_0/k_1$과 equivalent하므로

i) If $k_2=0$ and $k_1>0$,

$$\zeta_{hiu}=\int_{-\infty}^{-k_0/k_1}\phi(z-\mu_{hiu},\sigma^2)[1-\exp(k_0+k_1z)]dz$$

ii) If $k_2=0$ and $k_1<0$,

$$\zeta_{hiu}=\int_{-k_0/k_1}^{\infty}\phi(z-\mu_{hiu},\sigma^2)[1-\exp(k_0+k_1z)]dz$$

iii) If $k_2<0$ and $k_1\leqq 4k_0k_2$,

式(22), (27), (28)에 의하여

i) $$\zeta_{hiu}=\Phi\left(-\frac{k_0/k_1+\mu_{hiu}}{\sigma}\right)-\exp\left(k_0+\mu_{hiu}k_1+\frac{1}{2}\sigma^2k_1^2\right)\cdot\Phi\left(-\frac{k_0/k_1+\mu_{hiu}+\sigma^2k_1}{\sigma}\right),$$

ii) $$\zeta_{hiu}=\Phi\left(\frac{k_0/k_1+\mu_{hiu}}{\sigma}\right)-\exp\left(k_0+\mu_{hiu}k_1+\frac{1}{2}\sigma^2k_1^2\right)\cdot\Phi\left(\frac{k_0/k_1+\mu_{hiu}+\sigma^2k_1}{\sigma}\right),$$

iii) $$\zeta_{hiu}=1-(1-2k_2\sigma^2)^{-\frac{1}{2}}\exp\left\{\frac{1}{2\sigma^2}\left[2k_0\sigma^2-\mu^2{}_{hiu}+\frac{(\mu_{hiu}+k_1\sigma^2)}{1-2k_2\sigma^2}\right]\right\}$$

c) Bernoulli g and Threshold Model for π

式(26)에서 $z=1$과 $z=0$으로 놓고 式 (20)과 (23)을 $T=1$로 놓으면

$$\zeta_{hiu}=f_{hiu}\prod_{l=1}^{r}(1-\pi_l)+1-f_{hiu}$$

3. 討議 및 結論

筆者(1974)와 Morton 및 Maclean(1974)이 유전적인 가설검정을 위한 방법으로 3가지의 통계모형을 제시하였고 Elston과 Stewart(1971)가 또한 Ascertainment 函數에서 3종류의 확률모형을 제시함과 동시에 수집된 家系의 자료에 응용한 바 있다. North Carolina 대학의 Elston등의 GENSEG 컴퓨터 프로그램은 하와이 대학 Morton 등의 COMSEG프로그램보다 缺測値 (missing value)를 補完하고 資料를 다시 배열하는 데에는 장점이 있으나 CONSEG에 비하여 각 模型에 관한 separate program의 效率(efficiency)이 낮다. 두 프로그램의 공통점은 Newton-Raphson의 반복법(iteration)을 사용하지 않고 最尤副프로그램(maximum-likelihood subroutine)을 이용하여 尤度를 극대화한 점이다.

one-locus model에서의 分離比(segregation ratios)의 검정은 확율의 벡타 $\underset{\sim}{T}'=(T_1, T_2, T_3)$의 값이 $(1\ \frac{1}{2}\ 0)$으로 제한되었을 때 maximun likelihood를 비교하게 된다. 동일한 가설에 대한 검

정으로서 본연구는 $\underset{\sim}{T}, \underset{\sim}{T}_0' = (1\ \frac{1}{2}\ 0)$의 最尤推定値인 $\hat{\underset{\sim}{T}}$를 다음과 같은 통계량$(\hat{\underset{\sim}{T}} - \underset{\sim}{T}_0)' V^{-1} (\hat{\underset{\sim}{T}} - \underset{\sim}{T}_0)$으로서 계산을 시도하였으며 컴퓨터 프로그램을 變形하였다. 여기서 V는 $\hat{\underset{\sim}{T}}$의 분산-공분산행렬(variance-covariance matrix)의 一致推定値를 의미하며 이 통계량은 尤度比判定(likelihood-ratio criterion)에 漸近的으로 同一(asymtotically identical)하고 동일한 수의 자유도를 갖는 χ^2분포와 비교될수 있다. Hardy-Weinberg equilibrium를 검정하기 위하여 Elandt-Johnson(1970, 1971)과 Murphy(1967)등이 연구한 분리분석(segregation analysis)의 일반적인 방법에 대한 특별한 경우가 된다. 이 방법은 母數(parameter)가 많을 때 동시에 추정하기 어려운 경우에 V값과 $\hat{\underset{\sim}{T}}$값을 쉽게 구할 수 있는 것이다.

본 연구에서 논의한 模型은 oligogenic한 경우를 이론적으로 검토하면서 특히 multiple ascertainment가 존재할 때 最尤推定値에 관한 이론과 oligogenic model에 관하여 Elston등(1974), Morton등(1974)과 筆者(1974)의 연구를 이론적으로 종합 검토 하였다. 결과적으로 분리분석에 관한 다수의 이론적 모형이 발표되었으나 대개의 경우 일반적인 방법에 대한 특수한 경우인 것이다. 두 방법 모두 장단점이 있으며 장차의 분리분석에 관한 연구는 monogenic, polygenenic, oligogenic, megaphenic, environmental effect등에 관한 각 要素(factor)의 연관성을 구명하는것이 중요한 연구과제가 될 것이다.

産學協同財團에 제출한 연구계획에는 家系資料(family-pedigree data)를 분석키로 되었으나, 불행하게도 현재 한국에 있는 종합병원에 자료가 없으므로 실제 분석을 시도 못한것이 유감스럽다.

〈부록 1〉 上記式(1)—(17)의 부호 설명

s=the family size(즉 the size of the sibship)

S=the maximal family size in the sample

r=the number of affected children in the sibship of size s $(r=1, 2, \cdots, s)$

a_{rs}=the recorded number of families of size s, each with r affected children

$n_s=\sum_{r=1}^{s} a_{rs}$=the total number of recorded families of size s.(n_s가 fixed 되었다고 가정)

$N=\sum_{s=1}^{S} n_s=\sum_{r=1}^{S}\sum_{r=1}^{s} a_{rs}$=the total number of recorded families of all sizes, that is, the sample size which is fixed, together with fixed S

$r_s=\sum_{r=1}^{s} r a_{rs}$=the number of affected children in all n_s families

$R=\sum_{s=1}^{S} r_s=\sum_{s=1}^{S}\sum_{r=1}^{s} r a_{rs}$=the total number of affected children in the whole sample

$t_s=Sn_s$=the number of recorded children in all n_s families

$T=\sum_{s=1}^{S} t_s=\sum_{s=1}^{S} s n_s$=the total number of recorded children in the whole sample

〈부록 2〉 上記式(18)—(28)의 부호 설명(oligogenic 確率模型)

h_i=i번째의 性(sex)

h_0=동일한 부모의 性

h_e=재혼한 경우 남편이나 아내의 性

Z=phenotypic trait의 측도(測度)

Z_0=common parent의 측도

Z_e=e번째의 배우자의 측도

Z_{ej}=e번째 배우자에 의한 j번째 자녀의 측도

h_i=1이면 i번째 자녀가 남성을 의미하고

h_i=2이면 i번째 자녀가 여성을 의미한다.

$g_{hiu}(z)$=因子型 u와 性 h_i를 가진 子女에 대한 表現型에 관한 확률밀도 함수

k_{hi}=性 h_i에 대한 다른 因子型의 數

$P_{stu}{}^{hi}$=父母의 因子型이 s와 t일때 性 h_i가 갖는 因子型 u의 확률

ϕ_{hos}=h_0가 갖는 확률(父系의 集團의 因子型이 s일때)

ϕ_{hetfs}=h_e가 갖는 확률(h_e의 因子型은 t이고 배우자의 因子型은 s일때)

d=결혼의 회수($e=1, 2, \cdots, d$)

n_e=n번째 결혼에서 생기는 자녀수($j=1, 2, \cdots, n_e$)

b_{il}=l번째 ascertainment방법에 의해 결정된 i번째에 관한 proband의 상태

$\alpha(\underset{\sim}{z}, \underset{\sim}{b})$=全體家系에 관한 ascertainment 함수

參 考 文 獻

[1] Elandt-Johnson, R. C., "Segregation Analysis for Complex Modes of Inheritance," *Am. J. Hum. Genet.*, Vol. 22(1970), pp.129-140.

[2] Elandt-Johnson, R. C., "Complex Segregation Analysis II, Multiple Classification," *Am. J. Hum. Genet.* Vol., 23(1971), pp.17-32.

[3] Elston, R. C. and Stewart, J., "A General Model for the Genetic Analysis of Pedigree Data," *Hum. Hered.*, Vol. 23(1971), pp.523-542.

[4] Kaplan, E. B. and Elston, R. C., "A Subroutine Package for Maximum-likelihood Estimation (MAXLIK)," Univ. of North Carolina, Institute of Statistics, Mimeo Series, No. 823, 1972.

[5] Morton, N. E., "Segregation Analysis," in *Computer Application in Genetics,* edited by N. E., Morton, Univ. of Hawaii Press, 1969.

[6] Morton, N. E. and Maclean C. J., "Analysis of Family Resemblance III, Complex Segregation of Quantitative Traits," *Am. J. Hum., Genet.* Vol. 26 (1974), pp.489～503.

[7] Murphy, E.A. and Bolling, D.R., "Testing of Single-locus hypothesis where There is Incomplete Separation of the Phenotypes," *Am. J. Hum., Genet.*, Vol. 19(1967), pp.322-344.

[8] Shin, Han P., "Complex Segregation Analysis," *Jour. Korean Stat. Soc.*, Vol. 3, No. 2(1974), pp.103-115.

〈ABSTRACT〉

Maximum Likelihood Estimator of the Segregation Parameter under Multiple Ascertainment ($0<\pi<1$) with Known π

Han Poong Shin*

Theoretical details are considered of maximum likelihood estimator of segregation parameter and of several oligogenic models for segregation analysis.

For a one-locus model, the maximum likelihood obtained when the vector of transmission probabilities $\underset{\sim}{T}'=(T_1\ T_2\ T_3)$ is estimated is compared to the maximum likelihood obtained under the restriction $\underset{\sim}{T}'=(1\ \frac{1}{2}\ 0)$. Another method of testing the same hypothesis is to compute the statistic $(\hat{\underset{\sim}{T}}-T_0)'V^{-1}(\hat{\underset{\sim}{T}}-T_0)$, where $\hat{\underset{\sim}{T}}$ is the maximum-likelihood estimate of $\underset{\sim}{T}$, $T_0'=(1\ \frac{1}{2}\ 0)$, and V is a consistent estimate of the variance-covariance matrix of $\hat{\underset{\sim}{T}}$, this statistic is asymtotically identical to the likelihood-ratio criterion and so is compared to a χ^2 distribution with the same number of degrees of freedom.

* Professor, Department of Statistics, Korea University.