

3 × 3 교차실험의 범수형 자료 분석을 위한 주변확률모형 *

안주선 ¹⁾

요약

세 처리, 세 기간을 갖는 3 × 3 교차실험에서 $c(\geq 3)$ 개의 범주를 가진 자료의 분석에 사용될 수 있는 주변확률모형을 제안한다. 이 모형은 Kenward and Jones(1991)의 결합확률 모형의 대조물(counterpart)로 사용될 수 있고 2항 변수를 갖는 3 × 3 교차실험에서 처리 효과를 분석하기 위한 Balagtas et al(1995)의 일변량주변로지트모형의 일반화이다. 세 종류의 링크변환을 사용하여 주변확률모형방정식이 구성된다. 링크변환행렬과 모형행렬을 구성하는 방법이 주어지고, 모수의 추정이 논의된다. 제안된 모형을 Kenward and Jones 자료의 분석에 응용한다.

주요용어: 교차실험, 주변확률모형, 전이효과, 링크변환.

1. 서론

교차실험(cross-over trials)은 반복측정(repeated measurements)을 위한 실험의 한 종류로서 임상실험연구에서 실험대상자(subjects)가 적을 때 주로 사용된다[Jones and Kenward(1989), Ratkowsky et al(1993)]. 본 논문에서는 세 처리, 세 기간을 갖는 교차실험 자료의 분석을 위한 주변확률모형을 제안한다. 세 종류의 약제 A, B, C의 효과를 조사하고자 할 때 6개의 처리순서(treatment sequences) ABC, ACB, BAC, BCA, CAB, CBA를 사용한다. 실험대상자를 6개의 처리순서 중 하나에 무작위(random)로 배치하고 세 종류의 약 A, B, C를 처리순서에 따라 복용케 한다. 각 약의 효과를 $c(\geq 2)$ 개의 범주(categories) 또는 수준(levels)중 하나로 나타낼 때, 각 실험대상자로부터 세 번의 복용(또는 세 기간)에서 각각 약의 효과가 기록 될 것이다. 이러한 실험을 세 처리, 세 기간을 갖는 3 × 3 교차실험이라 한다. 이 때 반응변수는 c 범주를 취하고, 6개의 $3 \times 3 \times c$ 분할표가 얻어진다. 3절의 표 3.1은 $c = 3$ 일 때의 3 × 3 교차실험 자료의 한 예이다. 교차실험에서는 이전의 처리(또는 약)의 효과가 다음의 처리의 효과에 누적되는 전이효과(carry-over effects)가 존재한다. 실험할 수 있는 기간을 충분히 길게 잡을 수 있다면 실험대상자에게 한가지 약을 복용시킨 후 다른 약을 복용시킬 때까지 충분히 긴 기간(washout period라 함)을 줄 수 있고, 실험대상자의 임상조건(clinical condition)은 처음의 실험상태로 되돌릴 수 있으므로 전이효과를 무시할 수 있을 것이다[Armitage and Hills(1982)]. 그러나 일반적으로 실험기간이 충분히 길지 못하거나 실험기간이 길다고 하더라도 전이효과가 완전히 제거되지 않을 수 있기 때문에 전이

* 이 연구는 98학년도 강릉대학교의 해외과건연구비의 지원에 의하여 이루어졌음.

1) (210-702) 강원도 강릉시 지변동123, 강릉대학교 통계학과 교수

E-mail: jsahn@kangnung.ac.kr

효과를 포함한 모형이 사용되어야 한다. 전이효과에는 첫 번째 기간의 처리효과가 두 번째 기간의 처리효과에 누적되거나 두 번째 기간의 처리효과가 세 번째 기간의 처리효과에 누적되는 일차 전이효과(first order carry-over effects)와 첫 번째 기간의 처리가 세 번째 기간의 처리효과에 누적되는 이차 전이효과(second order carry-over effects)가 있다.

Kenward and Jones(1991)는 Conaway(1989) 방법의 응용으로 실험대상자효과(subject effects) 또는 잠재변수(latent variables)를 포함하는 결합확률모형(joint probability model)을 제안했다. 이들의 모형에서는 이차 전이효과가 고려되지 않았으며 모수를 추정하기 위해 칸의 수만큼 모형방정식들이 필요하다. 또한 실험대상자의 수가 칸(cell)의 수보다 적거나 충분히 크지 않으면 많은 수의 0-칸(zero cell)이 나타날 수 있고, 결합확률 모형으로 모수를 추정할 때 0-칸의 문제가 발생할 수 있게 된다. 일반적으로 0-칸의 문제를 해결하기 위해 주변확률모형을 사용한다. Balagtas et al(1995)은 2항 반응변수를 갖는 3×3 교차실험에서 처리 효과를 조사하기 위해 로지스틱함수 $\ln\left(\frac{\pi}{1-\pi}\right)$ 을 사용한 일변량주변로지트모형(univariate marginal logits model)을 제안했다.

본 논문에서는 Balagtas et al(1995)의 2-범주 경우를 $c(\geq 3)$ -범주로 확장하려고 한다. 2-범주(또는 이항반응변수)의 경우에는 로지스틱함수를 사용할 수 있으나 $c(\geq 3)$ -범주일 때는 다른 링크변환(link transformation) 또는 링크함수(link function)를 사용하게 된다. 제안된 모형은 $c = 2$ 인 경우의 Balagtas et al 모형을 포함하며 3×3 교차실험의 일반화된 주변확률모형이라고 할 수 있다. $c(\geq 3)$ -범주를 갖는 자료의 주변확률모형을 위한 링크변환과 링크변환집합의 행렬표시가 2.1절에서 주어진다. 2.2절에서 주변확률모형방정식을 정의하고 모형행렬이 논의된다. 2.3절에서 효과들의 추정방법을 설명한다. 3절에서 제안된 모형을 사용하여 Kenward and Jones(1991)에 주어진 자료를 분석한다.

2. 주변확률 모형

6개의 처리순서 ABC, ACB, BAC, BCA, CAB, CBA에 대응한 6개의 실험대상자들의 그룹을 $G_1, G_2, G_3, G_4, G_5, G_6$ 라 한다. G_1 의 각 실험대상자에게 A, B, C 순으로, G_2 의 각 실험대상자에게 A, C, B 순으로, \dots , G_6 의 각 실험대상자에게 C, B, A 순으로 처리를 실험한다고 가정하고, 그룹 효과들은 세 처리기간동안 변하지 않는다고 가정한다. 그리고 l 째 그룹 내에서 기간반응변수 t_1, t_2, t_3 의 반응이 i, j, k 일 확률을 $Pr(t_1 = i, t_2 = j, t_3 = k | g = l) = \pi_{ijkl}$ ($1 \leq i, j, k \leq c$)라 하면, 각 기간의 주변확률은 다음과 같다.

$$\text{기간 1: } \pi_{r++l} = \sum_{j=1}^c \sum_{k=1}^c Pr(t_1 = r, t_2 = j, t_3 = k | g = l)$$

$$\text{기간 2: } \pi_{+r+l} = \sum_{i=1}^c \sum_{k=1}^c Pr(t_1 = i, t_2 = r, t_3 = k | g = l)$$

$$\text{기간 3: } \pi_{++rl} = \sum_{i=1}^c \sum_{j=1}^c Pr(t_1 = i, t_2 = j, t_3 = r | g = l).$$

단, $\sum_{r=1}^c \pi_{r++l} = \sum_{r=1}^c \pi_{+r+l} = \sum_{r=1}^c \pi_{++rl} = 1$.

처리들 사이의 교호작용(interaction)이 분할표의 칸들과 독립이라고 가정하고, 칸이 변할 때 주변확률벡터(marginal probability vector) $\boldsymbol{\pi} = (\pi_{(1)}^t, \pi_{(2)}^t, \pi_{(3)}^t)^t$ 가 어떻게 영향을 받는가를 조사하는 모형방정식을 정의하고자 한다. 단, $\boldsymbol{\pi}_{(h)} = (\pi_{(h)1}^t, \dots, \pi_{(h)6}^t)^t$ 이고,

$\pi_{(1)l} = (\pi_{1++l} \cdots \pi_{c++l})^t$, $\pi_{(2)l} = (\pi_{+1+l} \cdots \pi_{+c+l})^t$, $\pi_{(3)l} = (\pi_{++1l} \cdots \pi_{++cl})^t$ 이다. t 는 전치행렬을 나타낸다.

2.1. 링크변환

본 절에서 반응변수의 주변종속성(marginal dependence)을 조사하기 위한 모형방정식에 사용할 세 종류의 링크변환집합을 정의한다. 표현의 단순화를 위해 l 그룹의 h 번째 기간에서 반응값 r 을 가질 확률을 π_{hrl} 로 나타낸다. 즉, π_{1rl} 은 π_{r++l} , π_{2rl} 은 π_{+r+l} , π_{3rl} 은 π_{++rl} 을 의미한다. 단, $r = 1, 2, \dots, c$.

2.1.1. 부분로지스틱변환(Partial Logistic Transformation)

$Q_r = \{s | s \in \{1, 2, \dots, c\}, s \neq r\}$ 라 두고, c 개의 부분로지스틱변환들의 집합을

$$PL_r = \left\{ \ln\left(\frac{\pi_{hsl}}{1 - \pi_{hsl}}\right) \mid c \in Q_r \right\}, \quad r = 1, 2, \dots, c$$

로 정의한다. $c=3$ 일 때, 3개의 부분로지스틱변환들의 집합은 다음과 같다.

$$\begin{aligned} PL_1 &= \left\{ \ln\left(\frac{\pi_{h2l}}{1 - \pi_{h2l}}\right), \ln\left(\frac{\pi_{h3l}}{1 - \pi_{h3l}}\right) \right\}, \\ PL_2 &= \left\{ \ln\left(\frac{\pi_{h1l}}{1 - \pi_{h1l}}\right), \ln\left(\frac{\pi_{h3l}}{1 - \pi_{h3l}}\right) \right\}, \\ PL_3 &= \left\{ \ln\left(\frac{\pi_{h1l}}{1 - \pi_{h1l}}\right), \ln\left(\frac{\pi_{h2l}}{1 - \pi_{h2l}}\right) \right\}. \end{aligned}$$

위에서 PL_3 을 사용한 모형은

$$\ln\left(\frac{\pi_{h1l}}{1 - \pi_{h1l}}\right) = f_1(\beta), \quad \ln\left(\frac{\pi_{h2l}}{1 - \pi_{h2l}}\right) = f_2(\beta)$$

이고, 각 확률의 추정량은

$$\pi_{\hat{h}1l} = \frac{\exp[f_1(\hat{\beta})]}{1 + \exp[f_1(\hat{\beta})]}, \quad \pi_{\hat{h}2l} = \frac{\exp[f_2(\hat{\beta})]}{1 + \exp[f_2(\hat{\beta})]}, \quad \pi_{\hat{h}3l} = 1 - \pi_{\hat{h}1l} - \pi_{\hat{h}2l}$$

로 주어진다. 단, $\hat{\beta}$ 는 β 의 추정량이다.

2.1.2. 누적로지스틱변환(Cumulative Logistic Transformation)

Q_r 의 순환적인 순열(cyclic permutation) $(r+1 \cdots c \ 1 \ 2 \ r-1)$, $r = 1, 2, \dots, c$ 을 $(r_1 \ r_2 \ \cdots \ r_{c-1})$ 라 할 때, c 개의 누적로지스틱변환들의 집합은

$$CL_r = \left\{ \ln\left(\frac{\sum_{k=1}^m \pi_{hr_k l}}{1 - \sum_{k=1}^m \pi_{hr_k l}}\right) \mid r_k \in Q_r, m = 1, \dots, c-1 \right\}, \quad r = 1, 2, \dots, c$$

로 정의된다. $c=3$ 일 때, 3개의 누적로지스틱변환들의 집합은 다음과 같다.

$$\begin{aligned}
CL_1 &= \left\{ \ln\left(\frac{\pi_{h2l}}{\pi_{h1l} + \pi_{h3l}}\right), \ln\left(\frac{\pi_{h2l} + \pi_{h3l}}{\pi_{h1l}}\right) \right\}, \\
CL_2 &= \left\{ \ln\left(\frac{\pi_{h3l}}{\pi_{h1l} + \pi_{h2l}}\right), \ln\left(\frac{\pi_{h3l} + \pi_{h1l}}{\pi_{h2l}}\right) \right\}, \\
CL_3 &= \left\{ \ln\left(\frac{\pi_{h1l}}{\pi_{h2l} + \pi_{h3l}}\right), \ln\left(\frac{\pi_{h1l} + \pi_{h2l}}{\pi_{h3l}}\right) \right\}.
\end{aligned}$$

위에서 CL_3 을 사용한 모형은

$$\ln\left(\frac{\pi_{h1l}}{\pi_{h2l} + \pi_{h3l}}\right) = f_1(\beta), \quad \ln\left(\frac{\pi_{h1l} + \pi_{h2l}}{\pi_{h3l}}\right) = f_2(\beta)$$

이고, 각 확률의 추정량은

$$\pi_{\hat{h}1l} = P_1, \quad \pi_{\hat{h}2l} = P_2 - P_1, \quad \pi_{\hat{h}3l} = 1 - P_2$$

이다. 단, $P_i = \frac{\exp[f_i(\hat{\beta})]}{1 + \exp[f_i(\hat{\beta})]}$, $i = 1, 2, 3$.

2.1.3. 베이스라인 변환(Baseline Transformation)

베이스라인이 π_{hrl} 일 때 Q_r 에 대한 베이스라인변환들의 집합을

$$BL_r = \left\{ \ln\left(\frac{\pi_{hsl}}{\pi_{hrl}}\right) \mid s \in Q_r \right\}, \quad r = 1, 2, \dots, c$$

로 정의한다. $c=3$ 일 때, 3개의 베이스라인들의 집합은 다음과 같다.

$$\begin{aligned}
BL_1 &= \left\{ \ln\left(\frac{\pi_{h2l}}{\pi_{h1l}}\right), \ln\left(\frac{\pi_{h3l}}{\pi_{h1l}}\right) \right\}, \\
BL_2 &= \left\{ \ln\left(\frac{\pi_{h1l}}{\pi_{h2l}}\right), \ln\left(\frac{\pi_{h3l}}{\pi_{h2l}}\right) \right\}, \\
BL_3 &= \left\{ \ln\left(\frac{\pi_{h1l}}{\pi_{h3l}}\right), \ln\left(\frac{\pi_{h2l}}{\pi_{h3l}}\right) \right\}.
\end{aligned}$$

위에서 BL_3 을 사용한 모형은

$$\ln\left(\frac{\pi_{h1l}}{\pi_{h3l}}\right) = f_1(\beta), \quad \ln\left(\frac{\pi_{h2l}}{\pi_{h3l}}\right) = f_2(\beta)$$

에서 각 확률의 추정량은

$$\begin{aligned}
\pi_{\hat{h}1l} &= \frac{\exp[f_1(\hat{\beta})]}{1 + \exp[f_1(\hat{\beta})] + \exp[f_2(\hat{\beta})]}, \\
\pi_{\hat{h}2l} &= \frac{\exp[f_2(\hat{\beta})]}{1 + \exp[f_1(\hat{\beta})] + \exp[f_2(\hat{\beta})]}, \\
\pi_{\hat{h}3l} &= \frac{1}{1 + \exp[f_1(\hat{\beta})] + \exp[f_2(\hat{\beta})]}
\end{aligned}$$

이다. 이상에서 정의한 3가지 변환에서 누적로지스틱변환은 처리 A, B, C의 반응값이 순서형일 때, 부분로지스틱변환과 베이스라인 변환은 순서형과 명목형(nominal)일 때 사용할 수 있다.

l 째 그룹의 실험대상자의 수가 n_l 이고, $X_{hil} = n_l \pi_{hil}$ 인 경우, 우도비 통계량

$$G^2 = 2 \sum_{h=1}^3 \sum_{i=1}^c \sum_{l=1}^6 X_{hil} [\ln(X_{hil}) - \ln(n_l \pi_{hil})]$$

는 $\chi^2(df)$ 분포를 갖는다. 모수의 수가 p 일 때 세 변환 모두 자유도 $df = 18(c-1) - p$ 를 갖는다[Anderson(1997), p188 191 참조].

각 변환들은 $k \times k$ 단위행렬, $I_k, \mathbf{1}_k = (1 \ 1 \ \dots \ 1)^t$, 적당히 정의된 행렬들을 사용하여 $L \ln(TM\pi)$ 인 형으로 나타낼 수 있다. 예를 들어 PL_3, CL_3, BL_3 는 $M_{54 \times 162} = (M_1 \ M_2 \ M_3)^t$, $L_{36 \times 72} = I_{18} \otimes [I_2 \otimes [1 \ -1]] = I_{36} \otimes [1 \ -1]$,

$$PL_3 \text{ 경우 : } T_{72 \times 36} = I_{18} \otimes \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad CL_3 \text{ 경우 : } T_{72 \times 36} = I_{18} \otimes \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$BL_3 \text{ 경우 : } T_{72 \times 36} = I_{18} \otimes \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

라 두면 $L \ln(TM\pi)$ 로 표현된다. $L \ln(TM\pi)$ 은 주변확률이 모두 영(zero)이 아닐 때 미분가능한 단조함수(monotonic differentiable function)임을 알 수 있다. 단, $M_1 = I_{18} \otimes [1_3^t \otimes 1_3^t]$, $M_2 = I_6 \otimes [I_3^t \otimes 1_3^t]$, $M_3 = I_6 \otimes [1_3^t \otimes I_3]$ 이고, \otimes 은 직적(direct product)을 나타낸다.

교차실험에서 얻어진 자료를 분석하고자할 때, 변환집합의 선택방법으로 $\ln()$ 함수가 정의되지 않은 경우가 가장 적은 것, 영의 수정 값의 변화에 덜 민감한 것, 또는 모형의 적합도 검정을 통해 적합도를 높게 만드는 것을 선택하여 사용할 수 있다. 예를 들어 $\ln()$ 함수가 정의되지 않은 경우가 가장 적은 것은 반응값 1에서 0-칸이 많을 때 PL_1 과 CL_2 , 반응값 2에서 0-칸이 많을 때 PL_2 와 CL_3 , 그리고 반응값 3에서 0-칸이 많을 때 PL_3 과 CL_1 이다.

2.2. 주변확률모형과 모형행렬

본 절에서 3×3 교차실험에서 그룹 효과, 기간 효과, 기간과 처리의 교호작용, 일차와 이차 전이효과 등이 있을 때 이들을 분리하여 의미를 부여하고 처리 효과를 추정할 수 있는 모형방정식을 정의한다. 실험대상 각각의 효과는 고려하지 않고, 반응과 관련된 부그룹 효과 $g_{1l}, g_{2l}, \dots, g_{cl}, l = 1, 2, \dots, 6$ 만을 모형에 포함시킨다. 이 때 $\sum_{i=1}^c g_{il}$ 은 l 째 그룹의 순서효과(sequence effect)에 대응된다. 기간 효과들을 반응과 관련시킬 때 그룹효과와 처리효과가 교락되기 때문에 세 개의 기간효과 p_1, p_2, p_3 만을 고려한다. 처리와 기간사이의 교호작용(τp), 첫 번째 기간에서 두 번째 기간으로의 일차 전이효과(기간 1에서 일차 전

이효과; $c12$), 두 번째 기간에서 세 번째 기간으로의 일차 전이효과(기간 2에서 일차 전이효과; $c23$)와 첫 번째 기간에서 세 번째 기간으로의 이차 전이효과(기간 1에서 이차 전이효과; $c13$)를 갖는 모형을 생각한다. 이상에서 언급한 효과들을 추정할 주변확률모형방정식을 2.1절에 주어진 링크함수 ($L(h, r, l)$ 로 표시)를 사용하여 다음과 같이 정의한다.

$$\begin{aligned} L(1, r, l) &= \alpha + g_{rl} + \tau_{t[1,l]r} + p_1 + (\tau p)_{t[1,l]r,1}, \\ L(2, r, l) &= \alpha + g_{rl} + \tau_{t[2,l]r} + p_2 + (\tau p)_{t[2,l]r,2} + c12_{t[1,l]r}, \\ L(3, r, l) &= \alpha + g_{rl} + \tau_{t[3,l]r} + p_3 + (\tau p)_{t[3,l]r,3} + c23_{t[2,l]r} + c13_{t[1,l]r}, \\ & r = 1, 2, \dots, c-1, l = 1, 2, \dots, 6 \end{aligned}$$

단, h 는 기간, r 은 반응치, l 은 그룹을 나타내고, α 는 위치상수(location constant), $t[h, l]$ 은 그룹 l 에서 기간 h 의 처리, $\tau_{t[h,l]r}$ 은 반응치 r 을 갖는 처리 $t[h, l]$ 의 효과, $c12_{t[1,l]r}$ 은 반응치 r 을 가질 때 기간 1에서 $t[1, l]$ 의 일차 전이효과, $c23_{t[2,l]r}$ 은 반응치 r 을 가질 때 기간 2에서 $t[2, l]$ 의 일차 전이효과, $c13_{t[1,l]r}$ 은 반응치 r 을 가질 때 기간 1에서 $t[1, l]$ 의 이차 전이효과를 나타낸다.

위의 방정식은 모형행렬(또는 계획행렬) X 와 모수(또는 효과)들의 벡터 β 를 사용하여 $L \ln(TM\pi) = X\beta$ 로 표현할 수 있다. 모형행렬 X 는 $18(c-1)$ 개의 행벡터와 추정할 모수의 개수와 같은 수의 열벡터를 갖는다. X 는 적당한 가장자리제약식(corner restrictions) (예, $g_{11} = 0, p_1 = 0, \tau_A = 0, \dots$ 등)을 사용하여 만들어진다. 제약식 $g_{11} = 0$ 대신 $\alpha = 0$ 을 사용하여 위치상수를 그룹효과에 포함시킬 수 있다. 고려할 수 있는 모형행렬 X 의 열계수(column rank)가 표 2.1에서 조사된다.

표 2.1: 고려할 수 있는 모형에서 모수의 수와 X 의 열계수($c = 3$ 인 경우)

모형 ⁽¹⁾	*	C12	C23	C13	C12,C23	C12,C13	C23,C13	C12,C23,C13
모수의 수	16	21	21	21	26	26	26	31
열계수	16	21	21	21	26	26	25	30
모형 ⁽²⁾	*	C12	C23	C13	C12,C23	C12,C13	C23,C13	C12,C23,C13
모수의 수	18	23	23	23	28	28	28	33
열계수	18	23	23	23	28	28	27	32
모형 ⁽³⁾	*	C12	C23	C13	C12,C23	C12,C13	C23,C13	C12,C23,C13
모수의 수	26	31	31	31	36	36	36	41
열계수	26	31	31	31	32	32	31	32

주) 모형⁽¹⁾, 모형⁽²⁾, 모형⁽³⁾의 행은 각각 모형 {Gr, Tr}, {Gr, Tr, Pe}, {Gr, Tr, Pe, TrPe}에 추가된 변인들이다. “*”는 전이효과를 갖지 않은 경우이다.

표 2.1에서 Gr은 그룹, Tr은 처리, Pe는 기간, TrPe는 처리와 기간의 교호작용(interaction), C12, C23은 기간 1, 기간 2에서 일차 전이, C13은 기간 1에서 이차 전이를 나타낸다. 모수의 수와 열계수가 같은 경우는 X 의 열벡터들이 선형독립이고 최대열계수(full

column rank)를 갖음을 뜻한다. 모형 {Gr, Tr, C23, C13}, ..., {Gr, Tr, Pe, TrPe, C12, C23, C13}에서는 모수의 수와 열계수가 서로 다르다. 이로부터 (C23, C13), (C12, C23, TrPe), (C12, C13, TrPe), (C23, C13, TrPe)에서 괄호 속의 변인들이 서로 교락되어 있음을 알 수 있다. 여기서 {Gr, Tr, C23, C13}은 그룹(Gr), 처리(Tr), 기간 2에서 일차 전이(C23), 기간 1에서 이차 전이(C13)를 고려한 모형을 나타낸다.

2.3. 모수의 추정

범주형 자료의 모형방정식은 미지의 칸 확률(cell probability)들의 벡터 π 를 모수(또는 효과)들의 벡터 β 와 관련지은 것이다. Haber(1985)는 일반화된 로그선형모형 $A \ln(\pi) = X\beta$ 을 제안했고 π 와 β 의 최우추정량들(maximum likelihood estimators, MLE)과 $Var(\hat{\beta})$ 의 대표본 추정량을 유도했고, Balagtas et al (1995)은 모형 $A \ln(B\pi) = X\beta$ 의 추정량을 논의했다. 2.2절에서 정의된 $L \ln(TM\pi) = X\beta$ 는 일반화된 주변확률 로그선형모형이고, Balagtas et al (1995)에서 논의된 방법으로 β 의 MLE와 $Var(\hat{\beta})$ 의 대표본 추정량을 얻을 수 있다. β 의 MLE는 Newton-Raphson방법으로부터 우도(likelihood)를 최대로 한 것이고, X 가 최대열계수인 모형행렬일 때,

$$\hat{\beta} = (X^t X)^{-1} X^t [L \ln(TM\hat{\pi})]$$

로 주어진다. X 가 최대열계수를 갖지 않은 경우에는 일반화역행렬(generalized inverse)을 사용하여 $\hat{\beta} = (X^t X)^- X^t [L \ln(TM\hat{\pi})]$ 로 추정한다. $\hat{\pi}$ 는 π 의 MLE이고 포아송 로그우도(Poisson log-likelihood)의 커널(kernel), $K(\mu) = \mathbf{n}^t \ln(\mu) - \mathbf{1}^t \mu$, 을 최대화함으로써 얻어진다 [Balagtas et al (1995)참조]. 주변확률벡터의 MLE, $M\hat{\pi}$ (M 은 2.1절에서 정의된 행렬임), 는 각 그룹 내에서 각 기간에 대한 주변상대도수들의 벡터임을 알 수 있다.

l 째 그룹에서 각 기간에 대한 주변도수(marginal frequency)들의 합은 일정 ($\sum_{i=1}^c N_{i++l} = \sum_{j=1}^c N_{+j+l} = \sum_{k=1}^c N_{++kl} = n_{+++l}$) 하므로 세 기간에 대한 주변도수들의 분포는

$$\begin{aligned} (N_{1++l}, \dots, N_{c++l}|l) &\sim Multi(n_{+++l}; \pi_{1++l}, \dots, \pi_{c++l}), \\ (N_{+1+l}, \dots, N_{+c+l}|l) &\sim Multi(n_{+++l}; \pi_{+1+l}, \dots, \pi_{+c+l}), \\ (N_{++1l}, \dots, N_{++cl}|l) &\sim Multi(n_{+++l}; \pi_{++1l}, \dots, \pi_{++cl}) \end{aligned}$$

이고, $G^2 = 2 \sum_{i=1}^c \sum_{l=1}^6 [n_{i++l} \ln(\frac{n_{i++l}}{\hat{m}_{i++l}}) + n_{+i+l} \ln(\frac{n_{+i+l}}{\hat{m}_{+i+l}}) + n_{++il} \ln(\frac{n_{++il}}{\hat{m}_{++il}})]$ 는 근사적으로 χ^2 분포를 따른다. 이 때 자유도는 2.1절에서와 같이 $18(c-1) - p$ 이다. 단, $\hat{m}_{i++l} = n_{+++l}\pi_{i++l}$ 이고, p 는 모형에 포함된 모수의 수이다.

3. 응용

표 3.1은 Kenward and Jones(1991)에 주어진 자료이다. 이 자료는 본태성월경곤란증(essential dysmenorrhea)환자의 고통 경감과 세 가지 치료약(위약(placebo), A, 약한 진통

표 3.1: 본태성월경근란증환자의 고통 경감의 도수자료

반응	G_1	G_2	G_3	G_4	G_5	G_6	반응	G_1	G_2	G_3	G_4	G_5	G_6
1 1 1	0	2	0	0	3	1	1 1 2	1	0	0	1	0	0
1 1 3	1	0	1	0	0	0	1 2 1	2	0	0	0	0	0
1 2 2	3	0	1	0	0	0	1 2 3	4	3	1	0	2	0
1 3 1	0	0	1	1	0	0	1 3 2	0	2	0	0	0	0
1 3 3	2	4	1	0	0	1	2 1 1	0	1	1	0	0	3
2 1 2	0	0	2	0	1	1	2 1 3	0	0	1	0	0	0
2 2 1	1	0	0	6	1	1	2 2 2	0	2	1	0	0	0
2 2 3	1	0	0	0	0	0	2 3 1	0	0	0	1	0	2
2 3 2	0	0	0	0	0	0	2 3 3	0	2	0	0	1	0
3 1 1	0	0	0	1	0	2	3 1 2	0	0	2	0	2	1
3 1 3	0	0	3	0	4	1	3 2 1	0	0	0	1	0	0
3 2 2	0	0	0	1	0	0	3 2 3	0	0	0	0	0	0
3 3 1	0	0	0	0	0	1	3 3 2	0	0	0	0	0	0
3 3 3	0	0	0	0	0	0							

제 (analgesic at low dose), B, 강한 진통제(analgesic at high dose),C,와의 관계를 조사하기 위해 다음 실험방법으로 교차실험을 하여 얻은 결과이다.

실험방법: 86명의 실험대상자를 무작위로 6개의 순서그룹 $G_1(ABC)$, $G_2(ACB)$, $G_3(BAC)$, $G_4(BCA)$, $G_5(CAB)$, $G_6(CBA)$ 에 각각 15, 16, 15, 12, 14, 14명씩 할당(allocate)하고, 각 그룹의 실험대상자에게 처리순서대로 약을 복용하게 한 후 각 처리기간 끝에서 경감(relief)을 조사한다. 경감의 수준(반응)은 1: 차이 없음(none), 2: 약간 좋아짐(moderate), 3: 아주 좋음(complete)의 3개이다.

표 3.1에서 반응이 “ $i j k$ ”이고 첫 번째 순서그룹 “ G_1 ”에서의 도수 f_1 은 첫 번째 기간에서 복용한 약 A의 반응이 i 이고, 두 번째 기간에서 복용한 약 B의 반응이 j 이고, 세 번째 기간에서 복용한 약 C의 반응이 k 인 실험자가 f_1 이었음을 나타낸다. 다른 순서그룹에서도 같은 방법으로 도수가 기록된다.

표 3.1의 분할표자료는 162 칸 중 112 칸의 도수가 0이므로 0-칸의 문제를 줄이기 위해 주변확률모형을 사용하는 것이 바람직하고, 반응값 $r = 3$ 에서 0 또는 0에 가까운 주변상대 도수를 가지는 경향이 있으므로 2.1절에서 정의된 PL_3 또는 CL_1 을 사용하여 분석할 수 있다. 어떤 자료를 로그선형모형에 적합시킬 때 영은 적당한 값으로 수정되어야한다.

표 3.2는 표 3.1자료를 최대열계수를 갖는 모형 {Gr, Tr, Pe, C12, C13}에 변환집합 PL_3 , CL_1 과 BL_1 을 사용할 때 영의 수정 값에 따른 모형의 우도비(likelihood ratio, G^2)의 변화를 나타낸 표이다. 이 표는 PL_3 과 CL_1 은 수정값의 변화에 민감하지 않으나 BL_1 은 수정값의 변화에 민감함을 보이고 있다. 따라서 0-칸의 문제를 줄이고 적합한 모형을 찾기 위

표 3.2: 세 변환집합에서 영의 수정 값의 변화에 따른 우도비

수정 값	PL_3	CL_1	BL_1	수정 값	PL_3	CL_1	BL_1
0.1	21.0815	21.0888	21.7583	1E-6	21.2332	21.2332	128.547
0.01	21.1397	21.1405	31.8624	1E-7	21.2333	21.2333	162.215
0.001	21.2149	21.2150	48.6748	1E-8	21.2333	21.2333	198.260
1E-4	21.2305	21.2305	70.9812	1E-9	21.2333	21.2333	236.177
1E-5	21.2329	21.2329	97.8857	1E-10	21.2333	21.2333	275.574

해 적당한 변환이 PL_3 과 CL_1 이라고 볼 수 있다. 본 논문에서는 주변상대도수가 영인 경우 10^{-9} 으로 수정하고 PL_3 를 사용한 모델에 적합시킬 것이다.

표 3.3은 표 3.1자료에 대해 부분로지스틱변환 PL_3 을 사용하여 표 2.1에서 최대행계수를 갖는 모형들의 우도비(G^2), 자유도(df)와 p-값을 SAS/IML을 사용한 프로그램으로 구한 것이다.

표 3.3: 최대 열계수를 갖는 모형들의 우도비, 자유도와 p-값

모형 ⁽¹⁾	*	C12	C23	C13	C12,C23	C12,C13
G^2	33.648	29.938	25.970	23.969	23.184	25.394
df	20	15	15	15	10	10
p-값	0.029	0.012	0.038	0.066	0.010	0.005
모형 ⁽²⁾	*	C12	C23	C13	C12,C23	C12,C13
G^2	27.719	22.166	25.551	23.332	21.233	19.976
df	18	13	13	13	8	8
p-값	0.067	0.053	0.020	0.038	0.007	0.010
모형 ⁽³⁾	*	C12	C23	C13		
G^2	23.203	15.253	19.146	19.146		
df	10	5	5	5		
p-값	0.010	0.009	0.002	0.002		

주) 모형⁽¹⁾, 모형⁽²⁾, 모형⁽³⁾, “*” 는 표 2.1과 같다.

표 3.3로부터 다음과 같은 세가지 자료의 특성을 볼 수 있다. (ㄱ) 기간효과를 포함하지 않은 모형에서 “C13” 를 제외하면 전이효과가 있다는 증거를 볼 수 없다 ($G^2_{\{Gr,Tr\}} - G^2_{\{Gr,Tr,C13\}} = 9.679$, $Pr(\chi^2(5) > 9.679) = 0.085$). (ㄴ) “*”, “C12”, “C12, C13”에서 모형⁽¹⁾(Pe를 포함하지 않음)과 모형⁽²⁾(Pe를 포함)의 G^2 의 차이가 각각 5.929, 7.772, 5.418이고 $Pr(\chi^2(2) > 5.418) = 0.067$ 임으로 기간효과를 무시할 수 없고, “C23”, “C13”, “C12,

표 3.4: 모형 { Gr, Tr, Pe, C12 }에서 효과와 표준오차의 추정

그룹	g_{11}	g_{21}	g_{12}	g_{22}	g_{13}	g_{23}	g_{14}	g_{24}	g_{15}	g_{25}	g_{16}	g_{26}
효과	1.364	-1.234	1.111	-1.502	1.122	-1.247	1.210	-0.612	1.238	-1.377	1.338	-1.530
ASE	0.287	0.287	0.263	0.263	0.262	0.220	0.268	0.269	0.224	0.203	0.253	0.224
처리												
효과				(τ_A)			τ_B	τ_C				
ASE				.			0.414	0.407				
기간												
효과				$(p1)$			$p2$	$p3$				
ASE				.			0.337	0.731				
전이												
효과			$(A1)$	$A2$	$(A3)$	$B1$	$B2$	$(B3)$	$C1$	$C2$	$(C3)$	
ASE			.	0.181	.	0.190	0.208	.	0.177	0.147	.	
G^2												
				G^2	df	p -값	$(\chi^2$		p -값)			
				22.166	13	0.053	(21.443		0.065)			

C23"에서 G^2 의 차이는 모두 1.951이하이고 $Pr(\chi^2(2) > 1.951) = 0.377$ 로 기간(Pe)효과를 모델 포함시킬 수 없다. 이는 모델에 포함된 전이효과의 종류에 따라 기간효과가 유의적일 수도 있고 유의적이지 아닐 수도 있다는 것을 말한다. (c) 모형⁽²⁾(TrPe를 포함하지 않음)과 모형⁽³⁾(TrPe를 포함)의 경우 G^2 의 차이는 모두 6.913이하이고 $Pr(\chi^2(8) > 6.913) = 0.546$ 이므로 처리와 기간사이의 교호작용(TrPe)은 없는 것으로 볼 수 있다.

실험대상자 수(86)가 칸의 수(162)에 비해 절반정도로 적기 때문에 적합한 모형을 찾는 것이 쉽지 않으나 위의 세 가지 자료의 특성으로부터 적합한 모형은 {Gr, Tr, Pe}로 볼 수 있다. 분석자의 목적에 따라 p-값이 0.05보다 큰 {Gr, Tr, C13}나 {Gr, Tr, Pe, C12}를 선택할 수도 있다. 본 논문에서는 교차실험에서 가장 많이 고려해야 하는 기간 1에서 일차 전이 효과(C12)를 포함한 모형 {Gr, Tr, Pe, C12}로부터 효과들을 추정하고자 한다. 처리효과의 추정결과는 {Gr, Tr, Pe}에서와 거의 같음을 볼 수 있었고, 모형 {Gr, Tr, Pe, C12}에 포함된 모수들의 추정치(효과)와 점근표준오차(ASE : Balagtas et al(1995)참조)가 표 3.4에 주어진다. 한편 전이효과 C23과 C13은 두 효과가 교락관계에 있으므로(표 2.1참조) 동일한 모형에서 추정이 가능하지 않고, 두 모형 {Gr, Tr, Pe, C12, C23}과 {Gr, Tr, Pe, C12, C13}이 모두 적합한 모형일 때, 두 효과의 차 C23-C12와 C13-C12를 추정하여 비교 할 수 있다.

표 3.4에서 "그룹"의 g_{rl} 은 l 째 그룹에서 반응치 r 을 갖는 부그룹 효과를 나타내고, 모델에서는 " $g_{3l} = 0$ "을 제약조건으로 사용한다. 부그룹 효과는 반응치 1에서 양의 값을 반응치 2에서 음의 값으로 추정되고 있다. 그러나 부그룹 효과의 합들 ($g_{1l} + g_{2l}$, $l=1, 2, 3, 4, 5, 6$)은 강한 진통제 C를 둘째 기간에 복용한 그룹 중 약한 진통제 B를 첫째기간에 복용한 그룹 4(BCA)에서 가장 큰 값(0.598)을 갖고, 위약 A를 첫째기간 복용한 그룹 2(ACB)에서 가

장 작은 값(-0.391)을 갖음을 볼 수 있다. “처리”에서 약한 진통제 B의 효과와 위약 A의 효과(제약조건: $\tau_A = 0$)의 차이(1.954)가 강한 진통제 C의 효과와 약한 진통제 B의 효과의 차이(0.278)보다 크음을 알 수 있다. 이는 진통제가 고통을 줄이는 데 도움을 주나 진통제 효과의 강약에는 큰 차이가 없음을 의미한다. 이 사실은 Kenward and Jones(1991)의 결과와 일치한 것이다. “기간”의 경우 첫 번째 기간(제약조건: $p_1=0$)과 두 번째 기간(p_2)의 차이가 두 번째 기간과 세 번째 기간(p_3)의 차이보다 크며, $p_1 > p_3 > p_2$ 로 첫 번째 기간의 효과가 가장 크고 두 번째 기간의 효과가 가장 작게 나타난 것을 볼 수 있다. 이는 첫 번째 복용한 약이 가장 영향을 많이 준다는 것을 의미한다. 끝으로 “전이(c12)”의 경우 위약에서 반응값 2(A2)를 가질 때 가장 큰 값(1.464)을 보이고, 강한 진통제에서 반응값 2(C2)를 가질 때 가장 작은 값(-0.132)을 보이고 있는 것으로 보아 약의 효과와 관계없이 약을 복용할 때 심리적 안정감을 잘 찾는 환자의 경우 전이효과가 큰 것으로 생각된다. 전이효과의 추정을 위해 모델에서 고려한 가장자리 제약조건은 $A_1=A_3=B_3=C_3=0$ 이다.

4. 결론

본 논문에서는 세 처리, 세 기간을 갖는 교차실험에서 반응변수가 세 개 이상의 범주를 갖는 자료의 분석을 위한 주변확률모형을 연구하였다. 교차실험은 임상실험에서 실험대상자수가 충분히 크지 않을 때 주로 사용하므로 교차실험에서 얻은 범주형 자료에는 0-칸이 많이 나타날 수 있다. 이러한 자료의 분석을 위해 세 종류의 링크변환을 사용한 주변확률모형을 제안했다. 각 변환에서는 반응변수가 취하는 범주 수에 따라 변환집합이 만들어진다.

교차실험에서 얻어진 자료를 분석하고자할 때, 변환집합의 선택방법으로 각 반응범주에 대응한 0-칸의 빈도수에 따라 변환집합을 선택하여 사용할 수도 있고, 모형의 적합도 검정을 통해 적합도를 높게 만드는 것을 선택할 수도 있으나, $\ln()$ 함수가 정의되지 않은 경우가 가장 적은 집합을 선택하여 사용하는 것이 0-칸의 문제를 줄이는 편리한 방법으로 생각된다. 또한 교차실험에서 처리효과 이외에 고려해야하는 처리순서, 처리시기, 처리효과와 전이, 처리와 기간과의 교호작용 등을 내포한 모형방정식이 주어졌다. 이 모형방정식에 대응한 모형행렬로부터 변인(처리, 기간, 전이)들의 교락관계를 조사했다. 둘째 기간에서의 일차 전이효과와 첫째 기간에서의 이차 전이효과는 제안된 모형에서 교락되어 있으므로 하나의 모형에서 비교할 수 없고, 첫째 기간에서의 일차 전이효과와 각 각을 포함한 두 개의 적합된 모형에서 두 전이효과의 차이로 비교할 수 있다.

끝으로 본 논문에서 제안된 주변확률모형을 사용하여 Kenward and Jones의 자료를 분석했다. 이 분석에서 그룹효과는 강한 진통제를 복용한 시기와 가장 밀접하고, 처리효과는 진통효과의 유무가 진통효과의 강약보다 더 큰 영향을 준 것을 볼 수 있었다. 기간 효과에서 첫 번째 복용한 약이 영향을 가장 많이 주고, 전이효과에서 약의 효과와 관계없이 약을 복용할 때 심리적 안정감을 잘 찾는 환자의 경우 전이효과가 크다는 것을 볼 수 있었다.

감사의 글

본 논문에 대하여 여러 가지 귀중한 조언을 주신 두 분의 심사위원님들께 감사를 드립니다.

참고문헌

- [1] Anderson, E.B. (1997). *Introduction to the Statistical Analysis of Categorical Data*. Springer, New York.
- [2] Armitage, P and Hills, M.P. (1982). The Two-Period Crossover Trial. *The Statistician* Vol. 3. 119-1311.
- [3] Balagtas, C.C. , Becker, M.P. and Lang, J.B. (1995). Marginal Modeling of Categorical Data from Crossover Experiments. *Appl. Statist.* Vol. 44. 63-77.
- [4] Conaway, M.R. (1989). Analysis of repeated categorical measurements with conditional likelihood methods. *JASA.* Vol. 84, 1094-1060.
- [5] Harber, H. (1985). Maximum likelihood methods for linear and log-linear models in categorical data. *Computational Statistics & Data Analysis.* Vol. 3, 1-10, North-Holland.
- [6] Jones, B. and Kenward, M.G. (1989). *The Design and Analysis of Cross-over Trials.* Chapman and Hall, London.
- [7] Kenward, M.G. and Jones, B. (1991). The Analysis of Categorical Data from Cross-over Trials using a Latent Variable Model. *Statistics in Medicine.* Vol.10, 1607-1619.
- [8] Ratkowsky, D.A., Evans, M.A., and Alldredge, J.R. (1993). *Cross-over Experiments.* Marcel Dekker, New York.

[2000년 3월 접수, 2000년 10월 채택]

The Marginal Model for Categorical Data Analysis of 3 × 3 Cross-over Trials *

Ju Sun Ahn ¹⁾

ABSTRACT

The marginal model is proposed for the analysis of data which have $c(\geq 3)$ categories in the 3×3 cross-over trials with three periods and three treatments. This model could be used for the counterpart of the Kenward-Jones' joint probability one and should be the generalization of Balagtas et al's univariate marginal logits one, which analyze the treatment effects in the 3×3 cross-over trials with binary response variables [Kenward and Jones(1991), Balagtas et al(1995)]. The model equations for the marginal probability are constructed by the three types of link functions. The methods would be given for making of the link function matrices and model ones, and the estimation of parameters shall be discussed. The proposed model is applied to the analysis of Kenward and Jones' data.

Keywords: Cross-over trials; Marginal probability model; Carry-over effects; Link transformation.

* This work was supported by '98, Fund for Research Abroad, Kangnung National University.

1) Professor, Department of Statistics, Kangnung National University.

E-mail: jsahn@kangnung.ac.kr