

## Semi-Markov 모형에 기초한 다중상태 생존자료의 준모수적 분석<sup>1)</sup>

여 성 칠<sup>2)</sup>

### 요 약

병원의 임상연구실험에서 종종 환자들의 치료에 따른 병세의 호전상태를 여러단계로 분류하여 상이한 치료방법에 대한 치료효과간의 차이를 알고자 하는 경우가 있다. 이와같이 다중상태의 생존자료를 분석하기 위해서 본 논문에서는 semi-Markov 모형에 Cox 회귀모형을 적용하여 회귀계수와 기저생존함수를 추정하고 이를 바탕으로 반응확률함수를 추정하였다. 그리고 본 논문의 결과를 실제 임상실험에서 얻어진 자료에 적용하여 분석하였다.

### 1. 서 론

생존분석에 관한 연구는 지난 수십여년간에 걸쳐서 미국이나 유럽 등에서 특히 의학이나 보건학 분야의 임상실험 등에서 얻어진 환자들의 생존자료에 대한 통계적 분석을 위한 기법으로서 현재까지 활발히 발전되어 가고있는 중이다.

그 중에서 환자의 병세에 따른 호전상태를 여러단계로 분류한 소위 다중상태의 생존자료를 분석하기 위한 통계모형들이 최근에 구미 여러나라의 통계학자들에 의해서 제시되고 연구되었다. 이러한 통계모형들 가운데 한가지 예를 들면 암 환자에 대한 임상실험에서 우리는 흔히 초기상태에 있는 환자가 의사로부터 처방을 받은 후에도 계속 병세가 악화되거나 또는 병세가 다소 나아진 상태에 도달되고는 나중에 다시 재발하는 상태로 가는 그림 1과 같은 4단계의 생존도해를 고려해 볼 수 있다

이 경우에 우리는 환자들의 병세가 각 상태  $i$  에 도달되기 까지의 시간  $T_i$  ( $i = 1, 2, 3$ )와 각 상태에서 머무르는 시간에 대해 관심을 갖게된다. 이를 위해 다음과 같이 정의되는 위험함수를 고려한다.

$$\alpha_i(t) = \lim_{h \rightarrow 0} \frac{1}{h} P(t \leq T_i < t+h | T_i \geq t), \quad 0 \leq t < \infty, \quad i = 1, 2, \quad (1.1a)$$

$$\alpha_3(t|s) = \lim_{h \rightarrow 0} \frac{1}{h} P(t \leq T_3 < t+h | T_3 \geq t, T_1 = s), \quad 0 \leq s < t < \infty, \quad (1.1b)$$

1) 본 논문은 1995년 건국대학교 학술진흥연구비 지원에 의해 연구되었음.  
2) (143-701) 서울시 광진구 모진동 93-1, 건국대학교 응용통계학과 교수.

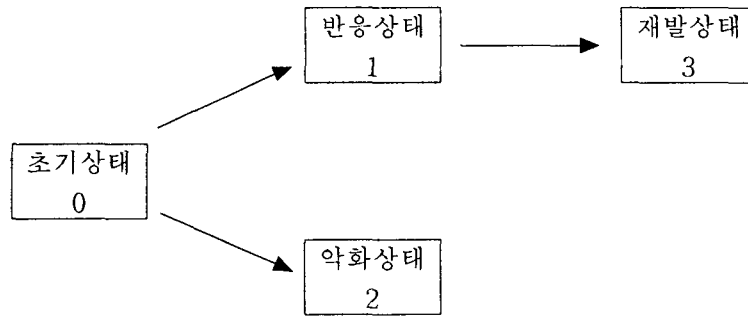


그림 1. 사중상태의 생존도해

즉,  $a_1(t)$ ,  $a_2(t)$ ,  $a_3(t|s)$ 는 환자들의 병세가 각각 처음 초기상태에서 시점  $t$ 에서 반응상태로, 처음 초기상태에서 시점  $t$ 에서 악화상태로, 시점  $s$ 에서 반응상태에 도달한 후 다시 시점  $t$ 에서 재발상태로 이동하는 순간 전이율(위험률)을 나타낸다.

지금까지의 연구들은 이들 전이율에 대해 동형의 Markov 모형, 비동형의 Markov 모형, semi-Markov 모형을 가정하여 연구되었다. 예를들면, Lagakos(1977)는 보조정보가 있을 때에 생존자료의 분석을 위해서 동형의 Markov 모형을 적용하였다. Lagakos등(1978)은 semi-Markov 모형에서 부분적으로 절단된 자료의 분석을 위해서 비모수우도법(nonparametric likelihood method)을 제시하였다. Temkin(1978)은 환자들의 상이한 치료방법의 효과에 대한 차이를 종합적으로 비교하기 위해 반응확률함수(probability-of-being-in-response function)를 제시하고 비동형의 Markov 모형하에서 반응확률함수를 추정하고 이에 대한 적용예를 보였다. Begg와 Larson(1982)은 동형의 Markov 모형하에서 반응확률함수의 성질을 조사하였으며 이 모형하에서 반응확률함수가 환자들의 상이한 치료방법에 따른 치료효과의 차이를 알아보는데 매우 효과적인 척도임을 보였다.

한편 Aalen과 Johansen(1978)은 보다 일반적인 Markov 모형하에서 중도절단자료가 주어졌을 때 전이율에 대한 누적한계추정법을 제시하고 그의 통계적 성질을 조사하였다. Hsieh등(1983)은 다중상태의 생존자료를 분석하기 위해 log-rank 검정법과 관련된 비모수 검정법을 연구하였다. Voelkel과 Crowley(1984)는 셈과정(counting process)기법을 이용하여 semi-Markov 모형하에서 반응확률함수를 추정하고 그에 대한 점근적 성질을 얻어내었다. Tsai등(1986)은 다중상태의 모형에서 이변량생존함수의 비모수추정법을 제시하고 이를 바탕으로 반응확률함수를 추정하였다. Pepe(1991)는 다중상태의 생존모형에서 Temkin(1978)에 의한 비동형의 Markov 모형하에서 반응확률함수의 추정치와 Weibull분포하에서 반응확률함수의 추정치를 시뮬레이션을 통해 비교하였다.

한편, 식 (1.1)은 동형의 Markov 모형 하에서  $a_i(t) = a_i$ ,  $i = 1, 2$ ,  $a_3(t|s) = a_3$  이고, 비동형의 Markov 모형 하에서  $a_3(t|s) = a_3(t)$  이고, semi-Markov 모형에서는  $a_3(t|s) = a_3(t-s)$  이다. 즉, 동형의 Markov 모형에서는 각 상태에 대응하는 위험률이 상수로 표현되고, 비동형의 Markov에서는 위험률이 각 상태에 이르는 시간(absolute time)의 함수로 표현되고, semi-Markov 모형에서는 위험률이 각 상태에서의 체류시간(duration time)의 함수로 표현됨을 의미한다.

한편, 환자들 이 각 상태에 도달하는 시간이나 각 상태에서 체류하는 시간은 공변량들(예를들면,

환자의 나이, 성별등)에 의해서 영향을 받을 수 있다. 따라서 각 상태에서의 위험률에 Cox(1972, 1975)에 의한 회귀모형을 첨가한 모형들을 고려할 수 있다. 그런데 Yeo(1994)는 그림 1에서 묘사한 다중상태의 생존도해에 대해 비동형의 Markov회귀모형을 적용하여 연구하였다. 따라서 본 논문에서는 semi-Markov회귀모형에 대해 연구한다. 그러나 논의의 단순화를 위해 우리는 공변량  $\mathbf{Z}$ 가 시간  $t$ 에 독립이라고 가정한다. semi-Markov회귀모형의 경우에 식 (1.1)은 다음과 같이 나타낼 수 있다.

$$\alpha_i(t; \mathbf{Z}) = \alpha_{i0}(t) \exp\{\mathbf{Z}\boldsymbol{\beta}_i\}, \quad i = 1, 2, \quad (1.2a)$$

$$\alpha_3(t|s; \mathbf{Z}) = \alpha_{30}(t-s) \exp\{\mathbf{Z}\boldsymbol{\beta}_3\}, \quad (1.2b)$$

여기서  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_p)$ 는  $p$ 개의 설명변수들의 벡터이고  $\boldsymbol{\beta}_i = (\beta_{1i}, \dots, \beta_{pi})'$ 는 각 상태  $i$  ( $i = 1, 2, 3$ )에서 공변량  $\mathbf{Z}$ 에 대응하는 회귀모수들의 벡터이다.

한편 식 (1.2)는 편의상 다음과 같이 표현할 수도 있다.

$$\alpha_i(t; \mathbf{Z}) = \alpha_{i0}(t) \exp\{\mathbf{Z}_i\boldsymbol{\beta}\}, \quad i = 1, 2 \quad (1.3a)$$

$$\alpha_3(t|s; \mathbf{Z}) = \alpha_{30}(t-s) \exp\{\mathbf{Z}_3\boldsymbol{\beta}\}. \quad (1.3b)$$

여기서  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ 는 각  $\boldsymbol{\beta}_i$  ( $i = 1, 2, 3$ )에서 서로 다른 모수들을 모두 포함하는 회귀모수들의 벡터이고  $\mathbf{Z}_i$  ( $i = 1, 2, 3$ )는 각 환자들의 기본 공변량벡터  $\mathbf{Z}$ 에서 상태  $i$ 에서 상태  $j$ 로의 이동에 대응하는 공변량 벡터이다.

본 논문의 제 2절에서는 다중상태의 생존모형에 대해 Andersen 과 Gill(1982), Gill(1984)등에서 연구된 Cox회귀모형에 대한 섹과정 기법을 바탕으로 다중상태의 생존자료에 대해 Cox형 부분우도함수(partial likelihood function)와 비모수우도함수를 제시하고 각 상태에 대응하는 회귀계수 벡터, 누적위험함수와 생존함수를 추정하였다. 제 3절에서는 제 2절에서 얻어진 결과를 바탕으로 반응확률함수를 추정하였다. 제 4절에서는 본 논문에서 얻어진 결과를 실제의 임상실험에서 얻어진 자료에 적용하여 분석하였다. 끝으로 제 5절에서는 본 논문에 대한 간단한 결론과 함께 본 논문과 관련된 앞으로의 연구방향에 대해 제시하였다.

## 2. 다중상태의 생존자료에 대한 우도함수

초기상태에서  $n$  명의 환자가 있다고 하자. 그리고  $T_{ij}$  ( $i = 1, 2, 3; j = 1, 2, \dots, n$ )를  $j$  번째 환자의 병세가 상태  $i$ 로 이동하는 시점이라 하자. 그런데 우리는 흔히 이들 시점들을 모두 관측하지 못하는 경우가 많다. 즉, 우리가 관측할 수 있는 시점들은 중도절단을 내포한 시점  $\hat{T}_{ij}$ 와 중도절단의 여부를 나타내는 지표(censoring indicator)인  $\delta_{ij} = I(\hat{T}_{ij} = T_{ij})$ 이다. 여기서  $I(A)$ 는 집합  $A$ 에 대한 지표함수(indicator function)를 나타낸다.

한편, 그림 1에서 묘사한 다중상태의 생존모형에 대해 다변량 섹과정  $\mathbf{N}(t)$ 를 다음과 같이 정

의한다.

$$N(t) = \{N_{ij}(t) : 0 \leq t < \infty, i = 1, 2, 3; j = 1, 2, \dots, n\},$$

여기서  $N_{ij}(t) = I(\tilde{T}_{ij} \leq t, \delta_{ij} = 1)$ 이다. 즉  $N_{1j}, N_{2j}, N_{3j}$ 는 공변량  $Z_j$ 를 가진  $j$ 번째 환자의 병세가  $t$ 시점 까지 상태 0에서 상태 1로, 상태 0에서 상태 2로, 상태 1에서 상태 3으로 각각 이동하는 것이 관측되면 값이 1이고 그렇지 않으면 0을 나타낸다.

그리고  $Y_{ij}(t)$  ( $i = 0, 1; j = 1, 2, \dots, n$ )를  $j$ 번째 환자가  $t$ 시점 바로 직전까지 중도절단되지 않고 상태  $i$ 에 있는 것으로 관측되면 값이 1이고 (이경우에 우리는  $j$ 번째 환자가 시점  $t$ 에서 위험에 노출되었다고 부른다.) 그렇지 않으면 값이 0이라고 하자. 즉

$$\begin{aligned} Y_{0j}(t) &= I(\tilde{T}_{1j} \geq t, \tilde{T}_{2j} \geq t) \\ &= 1 - N_{1j}(t-) - N_{2j}(t-), \end{aligned} \quad (2.1a)$$

$$\begin{aligned} Y_{1j}(t) &= I(\tilde{T}_{3j} \geq t > \tilde{T}_{1j}, \delta_{1j} = 1) \\ &= N_{1j}(t-) - N_{3j}(t-) \end{aligned} \quad (2.1b)$$

라고 정의한다. 그러면  $Y(t) = \{Y_{ij}(t) : 0 \leq t < \infty, i = 0, 1; j = 1, 2, \dots, n\}$ 는 예측가능한 지표 과정(predictable indicator process)으로 볼 수 있다.

따라서 중도절단의 독립성가정(Andersen 등(1988) 또는 Kalbfleisch 와 Prentice(1980)의 책 pp.119-122 참조)하에서  $N_{ij}(t)$  ( $i = 1, 2; j = 1, 2, \dots, n$ )는 Doob-Meyer 분해정리에 의해서 시그마필드( $\sigma$ -field)  $\mathcal{F}_t^j = \sigma\{N_{1j}(t), N_{2j}(t), Y_{0j}(t+), Z_j\}$ 에 관해서 그의 보상자(compensator)  $A_{ij}(t)$ 와 국소제곱적분가능한 마팅게일(local square integrable martingale)  $M_{ij}(t)$ 의 합으로 나타낼 수 있다. 즉

$$N_{ij}(t) = A_{ij}(t) + M_{ij}(t), \quad i = 1, 2; j = 1, 2, \dots, n \quad (2.2)$$

으로 표현된다.

그리고  $A_{ij}(t)$ 는 어떤 정규성의 가정(Aalen(1978) 3.2절 참조)하에서 절대연속이고 따라서 예측가능한 강도함수(intensity function)  $\lambda_{ij}(t)$ 가 존재해서 다음의 조건을 만족한다.

$$A_{ij}(t) = \int_0^t \lambda_{ij}(u) du \quad (2.3)$$

따라서 식 (2.2)는 다음과 같이 표현할 수 있다.

$$dN_{ij}(t) = \lambda_{ij}(t)dt + dM_{ij}(t). \quad (2.4)$$

더우기  $\lambda_{ij}(t)$ 는 다음의 조건을 만족한다.

$$\lambda_{ij}(t+) = \lim_{h \rightarrow 0} \frac{1}{h} P\{N_{ij}(t+h) - N_{ij}(t) | \mathcal{F}_t\}, \quad (2.5)$$

여기서  $\mathcal{F}_t = \bigvee_{j=1}^n \mathcal{F}_t^j$ , 즉  $\mathcal{F}_t$ 는 모든 환자들  $j=1, 2, \dots, n$ 에 대해서  $\mathcal{F}_t^j$ 를 포함하는 최소의 시그마필드이다. 식 (2.5)를 비공식적으로 표현하면

$$\lambda_{ij}(t)dt = P\{dN_{ij}(t) = 1 | \mathcal{F}_t\} \quad (2.6)$$

이 된다. 여기서  $dN_{ij}(t)$ 는 시점  $t$ 근방에서 간격  $dt$ 인 작은구간  $I_{dt}$ 에서 섹과정  $N_{ij}(t)$ 의 증분을 나타내고  $\mathcal{F}_{t-}$ 는 공변량  $\mathbf{Z}_j$  ( $j=1, 2, \dots, n$ )를 가진 모든 환자들에 대한  $t$  시점 바로 직전까지의 모든 기록을 나타낸다.

따라서 semi-Markov 회귀모형 (1.2)하에서  $\lambda_{ij}(t)$ 는 다음과 같이 표현할 수 있다.

$$\begin{aligned} \lambda_{ij}(t) &= \alpha_{ij}(t; \mathbf{Z}_j) Y_{0j}(t) \\ &= \alpha_{i0}(t) \exp(\mathbf{Z}_j \boldsymbol{\beta}_i) Y_{0j}(t), \quad i=1, 2. \end{aligned} \quad (2.7)$$

또는 모형 (1.3)하에서

$$\lambda_{ij}(t) = \alpha_{i0}(t) \exp(\mathbf{Z}_{ij} \boldsymbol{\beta}) Y_{0j}(t), \quad i=1, 2. \quad (2.8)$$

따라서 식 (2.2)는 다음과 같이 표현할 수 있다.

$$N_{ij}(t) = \int_0^t \alpha_{i0}(x) \exp(\mathbf{Z}_{ij} \boldsymbol{\beta}) Y_{0j}(x) dx + M_{ij}(t), \quad i=1, 2. \quad (2.9)$$

식 (2.9)에서  $N_{ij}(t)$ 를  $j=1, 2, \dots, n$ 에 관해서 모두 합하면 다음의 결과를 얻는다.

$$N_i(t) = \int_0^t \lambda_i(x) dx + M_i(t), \quad i=1, 2 \quad (2.10)$$

$$\text{여기서} \quad N_i(t) = \sum_{j=1}^n N_{ij}(t), \quad i=1, 2, \quad (2.11)$$

$$M_i(t) = \sum_{j=1}^n M_{ij}(t), \quad i=1, 2, \quad (2.12)$$

$$\lambda_i(t) = \alpha_{i0}(t) S_i^{(0)}(\boldsymbol{\beta}, t), \quad i=1, 2, \quad (2.13)$$

$$\text{그리고} \quad S_i^{(0)}(\boldsymbol{\beta}, t) = \sum_{j=1}^n \exp(\mathbf{Z}_{ij} \boldsymbol{\beta}) Y_{0j}(t) \quad i=1, 2 \quad (2.14)$$

이다. 식 (2.11) - (2.14)에서  $N_i(t)$ 는  $t$ 시점 까지 상태  $i$  ( $i=1, 2$ )로 이동하는 것이 관측된 모든 환자들의 수를 나타내고,  $M_i(t)$  ( $i=1, 2$ )는 평균이 0인 마팅계일  $M_{ij}(t)$ 들의 합이므로 역시 평균이 0인 마팅계일이다. 그리고  $S_i^{(0)}(\boldsymbol{\beta}, t)$  ( $i=1, 2$ )는  $t$  시점에 상태 0에서 위험에 노출된 환자들의 집합에 대해 지수선형모형 값  $\exp(\mathbf{Z}_{ij} \boldsymbol{\beta})$ 들의 합을 나타낸다.

한편 semi-Markov모형하에서, 시점  $t$ 에서 상태 1에서 상태 3으로 이동하는  $j$ 번째 환자의 섹과

정  $N_{3j}(t)$ 에 대응하는 강도과정  $\lambda_{3j}(\cdot)$ 는 다음과 같이 주어진다.

$$\lambda_{3j}(t-s) = \alpha_{30}(t-s) \exp(\mathbf{Z}_{3j}\boldsymbol{\beta}) Y_{1j}(t), \quad (2.15)$$

여기서  $s = \inf \{t: N_1(t) = 1\}$  이다.

그런데 식 (2.15)는 그 자체로서는 식 (2.13)에서와 같이 시그마필드  $\mathcal{F}_{t-}$ 에 관해서 승법적 강도 모형으로 표현되지 않는다. 예컨대 시점  $s = 2, 3, 4$  에서  $j = 1, 2, 3$ 인 3명의 환자들이 상태 0에서 상태 1로 이동하는 것이 관측되었다면 시점  $t = 7$ 에서 이들 강도함수들의 합은

$$\begin{aligned} \lambda_3(\cdot) &= \alpha_{30}(5) \exp(\mathbf{Z}_{31}\boldsymbol{\beta}) Y_{11}(7) + \alpha_{30}(4) \exp(\mathbf{Z}_{32}\boldsymbol{\beta}) Y_{12}(7) \\ &\quad + \alpha_{30}(3) \exp(\mathbf{Z}_{33}\boldsymbol{\beta}) Y_{13}(7) \end{aligned} \quad (2.16)$$

이 되고 따라서 식 (2.15)는 승법적 강도모형이 되지않는다. 그러나 Markov 모형의 경우에는 식

(2.16)은  $\lambda_3(7) = \alpha_{30}(7) \sum_{j=1}^3 \exp(\mathbf{Z}_{3j}\boldsymbol{\beta}) Y_{1j}(7)$  가 되어 승법적 강도모형으로 표현된다.

그런데 그림 1에서와 같이 환자들의 병세가 한쪽 방향으로만 진행되는 경우에 일반적으로 semi-Markov 모형의 경우에도 소위 랜덤시간변환(random time transformation)의 기법을 적용해서 승법적 강도모형으로 표현할 수 있다 (Aalen(1975, 5E)과 Voelkel 과 Crowley(1984) 참조). 따라서 식 (2.15)를 승법적 강도모형으로 표현하기 위한 기본적인 아이디어는 상태 1로 들어오는 모든 환자들을 랜덤시간변환에 의해 어떤 적당한 시점까지 가지런히 함께 모아서 새로운 시점에서 상태 3을 향해 다시 노출시키는 것이다. 즉, 각 상태  $i$  ( $i = 0, 1$ )에서  $j$ 번째 환자의 체류시간  $u$ 를 새로운 시간단위로 택해서 기존의 샘플링  $N_{ij}(t)$  ( $i = 1, 2, 3; j = 1, 2, \dots, n$ )와 예측과정  $Y_{ij}(t)$  ( $i = 0, 1$ )에 대응하는 새로운 샘플링  $N_{ij}^*(u)$ 와 예측과정  $Y_{ij}^*(u)$ 를 다음과 같이 정의한다.

$$N_{ij}^*(u) = N_{ij}(u), \quad i = 1, 2; j = 1, 2, \dots, n, \quad (2.17a)$$

$$N_{3j}^*(u) = N_{3j}(u+s), \quad j = 1, 2, \dots, n. \quad (2.17b)$$

그러면  $N_{ij}^*(u)$ 는 시그마필드  $\mathcal{F}_{ju}^* = \mathcal{F}_{uj} \vee \mathcal{G}_{uj}$  (여기서  $\mathcal{G}_{uj} = \sigma\{N_{3j}^*(u), Y_{1j}^*(u), \mathbf{Z}_j\}$ )에 관해서 다음과 같이 분해되어 진다.

$$N_{ij}^*(u) = \Lambda_{ij}^*(u) + M_{ij}^*(u), \quad i = 1, 2, 3; j = 1, 2, \dots, n, \quad (2.18)$$

여기서 
$$\Lambda_{ij}^*(u) = \int_0^u \lambda_{ij}^*(x) dx, \quad (2.19)$$

$$\lambda_{ij}^*(u) = \lambda_{ij}(u), \quad i = 1, 2, \quad (2.20a)$$

$$\lambda_{3j}^*(u) = \alpha_{30}(u) \exp(\mathbf{Z}_{3j}\boldsymbol{\beta}) Y_{1j}^*(u). \quad (2.20b)$$

따라서 식 (2.18)에서  $N_{ij}^*(u)$ 를  $j = 1, 2, \dots, n$  에 관해서 모두 합하면

$$N_i^*(u) = \int_0^u \lambda_i^*(x) dx + M_i^*(u), \quad i = 1, 2, 3 \quad (2.21)$$

이 된다. 여기서

$$N_i^*(u) = \sum_{j=1}^n N_{ij}^*(u), \quad i = 1, 2, 3, \quad (2.22)$$

$$M_i^*(u) = \sum_{j=1}^n M_{ij}^*(u), \quad i = 1, 2, 3, \quad (2.23)$$

$$\lambda_i^*(u) = \alpha_{i0}(u) S_i^{*(0)}(\beta, u), \quad i = 1, 2, 3, \quad (2.24)$$

$$S_i^{*(0)}(\beta, u) = S_i^{(0)}(\beta, u), \quad i = 1, 2, \quad (2.25a)$$

그리고 
$$S_3^{*(0)}(\beta, u) = \sum_{j=1}^n \exp(\mathbf{Z}_{3j} \beta) Y_{1j}^*(u) \quad (2.25b)$$

이다.

한편  $A_{i0}(u)$ 의 추정량은 샘플링  $\mathbf{N}^*(u) = \{N_{ij}^*(u) : 0 \leq u < \infty, i = 1, 2, 3, j = 1, 2, \dots, n\}$ 에 대한 비모수적 우도함수로부터 얻어질 수 있다. 즉 시그마필드  $\mathcal{F}_{u-}^*$ 가 주어졌을 때,  $dN_{ij}^*(u)$ 들은 기저누적위험함수  $A_{i0}(u)$ 가 이산형 또는 연속형인가에 따라 서로 독립인 다항분포 또는 포아송 (Poisson)분포를 따른다고 볼 수 있다. 따라서 중도절단의 비정보성 (noninformative censoring) 하에서 샘플링  $\mathbf{N}$ 에 대한 다항우도(multinomial likelihood)함수는 다음과 같이 주어진다.

$$L = \prod_u \left\{ \prod_{i=1}^2 \prod_{j=1}^n (Y_{0j}(u) dA_{i0}(u) \exp(\mathbf{Z}_{ij} \beta))^{dN_{ij}^*(u)} (Y_{1j}^*(u) dA_{30}(u) \exp(\mathbf{Z}_{3j} \beta))^{dN_{3j}^*(u)} \right. \\ \left. \times \left( 1 - \sum_{i=1}^2 dA_{i0}(u) S_i^{(0)}(\beta, u) \right)^{1 - dN_{\cdot}(u)} \left( 1 - dA_{30}(u) S_3^{*(0)}(\beta, u) \right)^{1 - dN_{3\cdot}^*(u)} \right\}, \quad (2.26)$$

여기서  $\prod_u$ 는 누승적분 (product integral)을 나타낸다 (Gill 과 Johansen(1990) 참조). 그리고

$$N_{\cdot}(u) = \sum_{i=1}^2 \sum_{j=1}^n N_{ij}(u), \quad N_{3\cdot}^*(u) = \sum_{j=1}^n N_{3j}^*(u), \quad dN_{ij}(u) = N_{ij}(u) - N_{ij}(u-),$$

$$dN_{3j}^*(u) = N_{3j}^*(u) - N_{3j}^*(u-) \text{ 이다.}$$

한편  $A_{i0}(u)$ 가 연속형일때 식 (2.26)은 다음과 같이 포아송 우도함수로 표현된다.

$$L = \prod_u \left\{ \prod_{i=1}^2 \prod_{j=1}^n (dA_{i0}(u) \exp(\mathbf{Z}_{ij} \beta))^{dN_{ij}^*(u)} \right\} \left( dA_{30}(u) \exp(\mathbf{Z}_{3j} \beta) \right)^{dN_{3j}^*(u)} \\ \times \exp \left\{ - \sum_{i=1}^2 \int_0^\infty S_{0i}^{(0)}(\beta, u) dA_{i0}(u) \right\} \exp \left\{ - \int_0^\infty S_{13}^{*(0)}(\beta, u) dA_{30}(u) \right\}. \quad (2.27)$$

실제계산의 경우 위의 식 (2.26)과 식 (2.27)에서 누승적분  $\prod_u$  는 센과정  $N_{ij}^*(u)$ 가 점프하는 시점들 즉,  $N_{ij}^*(u) = 1$  인 시점들에서의 누승으로 표현된다. 회귀모수벡터  $\beta$  가 고정되었다고 할 때 우도함수 (2.26) 또는 (2.27)를 최대화 하는  $dA_{i_0}(u)$ 의 추정량은

$$d\hat{A}_{i_0}(u; \beta) = \frac{dN_i(u)}{S_i^{(0)}(\beta, u)}, \quad i = 1, 2, \quad (2.28a)$$

$$d\hat{A}_{3_0}(u; \beta) = \frac{dN_3^*(u)}{S_3^{*(0)}(\beta, u)} \quad (2.28b)$$

이 된다. 따라서 위의 식 (2.28)로 부터  $A_{i_0}(u)$ 의 추정량은 다음과 같이 얻어진다.

$$\hat{A}_{i_0}(u, \beta) = \int_0^u \frac{1}{S_i^{(0)}(\beta, x)} dN_i(x), \quad i = 1, 2, \quad (2.29a)$$

$$\hat{A}_{3_0}(u, \beta) = \int_0^u \frac{1}{S_3^{*(0)}(\beta, x)} dN_3^*(x). \quad (2.29b)$$

그런데 식 (2.29)의 우변항은 어떤 시점  $u$ 에서  $Y_{0j}(u) = 0$  또는  $Y_{1j}^*(u) = 0$  ( $j = 1, 2, \dots, n$ )의 가능성을 배제할 수 없으므로 지표함수  $J_0(u) = I\{Y_0(u) = 0\}$ 와  $J_1(u) = I\{Y_1^*(u) = 0\}$  (여기서  $Y_0(u) = \sum_{j=1}^n Y_{0j}(u)$ ,  $Y_1^*(u) = \sum_{j=1}^n Y_{1j}^*(u)$  임)을 도입해서  $A_{i_0}(u)$  ( $i = 1, 2, 3$ )의 추정량으로 다음과 같이 정의한다.

$$\hat{A}_{i_0}^*(u; \beta) = \int_0^u \frac{J_0(x)}{S_i^{(0)}(\beta, x)} dN_i(x), \quad i = 1, 2 \quad (2.30a)$$

$$\hat{A}_{3_0}^*(u; \beta) = \int_0^u \frac{J_1(x)}{S_3^{*(0)}(\beta, x)} dN_3^*(x). \quad (2.30b)$$

이제 회귀모수벡터  $\beta$ 의 추정량을 구하기 위해 식 (2.28)을 식 (2.26) 또는 식 (2.27)에 대입하면 다음과 같이 부분적으로 최대화된 우도함수를 얻는다.

$$L^* = L_1(\beta) \cdot \prod_u \left\{ \prod_{i=1}^3 (dN_i^*(u))^{dN_i^*(u)} (1 - dN_i^*(u))^{1 - dN_i^*(u)} \right\} \quad (2.31)$$

또는

$$L^* = L_1(\beta) \cdot \prod_u \prod_{i=1}^3 dN_i^*(u)^{dN_i^*(u)} \exp(-N_i^*(\infty)) \quad (2.32)$$

여기서



$$L_1(\beta) = \prod_u \prod_{j=1}^n \left\{ \prod_{i=1}^2 \left( \frac{Y_{0j}(u) \exp(Z_{ij}\beta)}{S_i^{(0)}(\beta, u)} \right)^{dN_{ij}(u)} \left( \frac{Y_{1j}^*(u) \exp(Z_{3j}\beta)}{S_3^{*(0)}(\beta, u)} \right)^{dN_{1j}^*(u)} \right\} \quad (2.33)$$

이다. 위의 식 (2.33)은 그림 1에서와 같은 다중상태의 부분우도(partial likelihood)함수라고 간주할 수 있다. 식 (2.33)의 양변에 로그를 취하면

$$\log L_1(\beta) = \sum_{i=1}^3 \left\{ \sum_{j=1}^n \int_0^\infty Z_{ij}\beta dN_{ij}^*(x) - \int_0^\infty \log S_i^{*(0)}(\beta, x) dN_i^*(x) \right\} \quad (2.34)$$

을 얻는다.

식 (2.33) 또는 식 (2.34)를 최대로 하는  $\beta$ 의 값은 방정식  $U(\beta) = \frac{\partial}{\partial \beta} \log L_1(\beta) = 0$ 의

해로 구해진다. 따라서 위의 방정식의 해를  $\hat{\beta}$  라고 두고 이 값을 식 (2.30)에 대입하면  $A_{i0}(u)$ 의 추정량으로  $\hat{A}_{i0}^*(u, \hat{\beta}), i=1, 2, 3$  을 얻는다.

이제 위에서 얻어진 회귀모수벡터  $\beta$ 와 기저누적위험함수  $A_{i0}(u)$ 의 추정량들의 점근분포를 살펴보기로 한다. 그런데 semi-Markov 회귀모형하에서 우도함수 (2.26) 또는 (2.27)과 부분우도함수 (2.33)은 Markov 회귀모형의 경우와 비교할 때 기본 시간단위가 달라졌을 뿐 형식적으로 같은 형태를 취한다 (Yeo(1994)참조). 따라서 이들의 추정량  $\hat{\beta}$  와  $\hat{A}_{i0}^*(u, \hat{\beta})$  는 Markov 회귀모형의 경우와 동일한 점근분포를 따르게 된다. 그리고 이들 추정량들의 점근분포를 나타내기 위해서 Andersen과 Gill(1982)에서와 같은 기호를 사용하기로 한다. 즉.

$$S_i^{*(1)}(\beta, u) = \frac{\partial}{\partial \beta} S_i^{*(0)}(\beta, u), \quad i=1, 2, 3, \quad (2.35)$$

$$S_i^{*(2)}(\beta, u) = \frac{\partial^2}{\partial \beta^2} S_i^{*(0)}(\beta, u), \quad i=1, 2, 3 \quad (2.36)$$

이라고 두면

$$\begin{aligned} S_i^{*(1)}(\beta, u) &= S_i^{(1)}(\beta, u) \\ &= \sum_{j=1}^n Z_{ij} \exp(Z_{ij}\beta) Y_{0j}(u), \quad i=1, 2, \end{aligned} \quad (2.37a)$$

$$S_3^{*(1)}(\beta, u) = \sum_{j=1}^n Z_{3j} \exp(Z_{3j}\beta) Y_{1j}^*(u) \quad (2.37b)$$

와 
$$S_i^{*(2)}(\beta, u) = \sum_{j=1}^n Z_{ij}^{\otimes 2} \exp(Z_{ij}\beta) Y_{0j}(u), \quad i=1, 2, \quad (2.38a)$$

$$S_3^{*(2)}(\beta, u) = \sum_{j=1}^n Z_{3j}^{\otimes 2} \exp(Z_{3j}\beta) Y_{1j}^*(u) \quad (2.38b)$$

가 된다(여기서 임의의 벡터  $a$ 에 대해서  $a^{\otimes 2} = a a'$ 라고 표기함). 그리고

$$\begin{aligned} \mathbf{E}_i^*(\boldsymbol{\beta}, u) &= (E_{i1}^*(\boldsymbol{\beta}, u), \dots, E_{ip}^*(\boldsymbol{\beta}, u)) \\ &= \frac{S_i^{*(1)}(\boldsymbol{\beta}, u)}{S_i^{*(0)}(\boldsymbol{\beta}, u)} \end{aligned} \quad (2.39)$$

와

$$\begin{aligned} \mathbf{V}_i^*(\boldsymbol{\beta}, u) &= \{V_{ikl}^*(\boldsymbol{\beta}, u) : k, l = 1, \dots, p\} \\ &= \frac{S_i^{*(2)}(\boldsymbol{\beta}, u)}{S_i^{*(0)}(\boldsymbol{\beta}, u)} - \mathbf{E}_i^*(\boldsymbol{\beta}, u)^{\otimes 2} \end{aligned} \quad (2.40)$$

라고 두면 점수통계량(score statistic)은

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^3 \left\{ \sum_{j=1}^n \int_0^\infty \mathbf{Z}_{ij} dN_{ij}^*(x) - \int_0^\infty \mathbf{E}_i^*(\boldsymbol{\beta}, u) dN_{ij}^*(x) \right\} \quad (2.41)$$

이고 표본정보행렬(sample information matrix)은

$$\mathbf{I}(\boldsymbol{\beta}) = \sum_{i=1}^3 \int_0^\infty \mathbf{V}_i^*(\boldsymbol{\beta}, x) dN_i^*(x) \quad (2.42)$$

으로 주어진다.

그러면 어떤 정규성의 조건하에서 (Andersen과 Gill(1982) p.1105, p.1110 참조)  $\frac{1}{\sqrt{n}} \mathbf{U}(\boldsymbol{\beta})$  는 마팅게일 중심극한정리에 의해서 표본의 크기  $n$ 이 클 때 근사적으로 평균  $\mathbf{0}$ 이고 추정된 분산-공분산 행렬  $\frac{1}{n} \mathbf{I}(\hat{\boldsymbol{\beta}})$ 을 갖는  $p$ -변량 정규분포를 따른다는 것을 알 수 있다. 또한  $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ 는  $\mathbf{U}(\hat{\boldsymbol{\beta}})$ 의  $\boldsymbol{\beta}_0$  근방(여기서  $\boldsymbol{\beta}_0$ 는 회귀모수벡터  $\boldsymbol{\beta}$ 의 참값을 나타냄)에서 테일러(Taylor)급수확장을 이용해서, 표본의 크기  $n$ 이 클 때 근사적으로 평균  $\mathbf{0}$ 이고 추정된 분산-공분산 행렬  $n\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}})$ 인  $p$ -변량 정규분포를 따른다는 것을 알 수 있다. 따라서 예컨대, 귀무가설  $H_0: \boldsymbol{\beta} = \boldsymbol{\beta}_0$ 을 검정하기 위해서 위의 결과를 적용하면 아래의 검정통계량을 얻는다.

$$Q_1 = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \mathbf{I}(\hat{\boldsymbol{\beta}}) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \quad (2.43)$$

$$Q_2 = \mathbf{U}(\boldsymbol{\beta}_0)' \mathbf{I}^{-1}(\boldsymbol{\beta}_0) \mathbf{U}(\boldsymbol{\beta}_0), \quad (2.44)$$

$$Q_3 = -2(\log L_1(\boldsymbol{\beta}_0) - \log L_1(\hat{\boldsymbol{\beta}})), \quad (2.45)$$

여기서 검정통계량  $Q_i$  ( $i=1, 2, 3$ )는 모두 귀무가설  $H_0$ 하에서 근사적으로 자유도  $p$ 인  $\chi^2$ 분포를 따른다.

한편  $\sqrt{n}(\hat{A}_{i0}(u, \hat{\boldsymbol{\beta}}) - A_{i0}(u))$ 는 표본의 크기  $n$ 이 클 때 근사적으로 평균  $0$ 이고 추정된

분산함수

$$n \left\{ \int_0^u \frac{dN_i^*(x)}{S_i^{*(0)}(\hat{\beta}, x)^2} + H_i^*(\hat{\beta}, x)' I^{-1}(\hat{\beta}) H_i^*(\hat{\beta}, x) \right\} \quad (2.46)$$

인 독립증분을 갖는 가우시안(Gaussian)과정을 따르게 된다. 여기서

$$H_{jk}(\hat{\beta}, t) = - \int_0^t \frac{S_{jk}^{(1)}(\hat{\beta}, x)}{\{S_{jk}^{(0)}(\hat{\beta}, x)\}^2} dN_{jk}(x) \quad (2.47)$$

이다.

한편 공변량 벡터  $\mathbf{Z}$ 가 주어졌을 때 각 상태  $i$  ( $i=1, 2, 3$ )에 이르는 전이시간  $T_i$ 에 대응하는 기저생존함수와 생존함수는 각각  $S_{i0}(t) = P(T_i > t; \mathbf{Z} = \mathbf{0})$ 와  $S_i(t; \mathbf{Z}) = P(T_i > t; \mathbf{Z})$  이다. 그리고 이들의 추정량을 위험함수의 누승적분으로 나타내면 각각

$$\hat{S}_{i0}(t) = \prod_{0 \leq x \leq t} (1 - d\hat{A}_{i0}(x, \hat{\beta})), \quad i = 1, 2, 3, \quad (2.48)$$

$$\begin{aligned} \hat{S}_i(t; \mathbf{Z}) &= \prod_{0 \leq x \leq t} (1 - d\hat{A}_i(x; \mathbf{Z})) \\ &= \prod_{0 \leq x \leq t} (1 - \exp(\mathbf{Z}\hat{\beta}_i) d\hat{A}_{i0}(x, \hat{\beta})), \quad i = 1, 2, 3. \end{aligned} \quad (2.49)$$

이 되고 이들 추정량들 역시 표본의 크기  $n$ 이 클 때 근사적으로 가우시안 과정을 따르게 됨을 알 수 있다.

### 3. 반응확률함수

앞 절에서 얻어진 결과를 바탕으로 우리는 이제 반응확률함수를 추정한다. 제 1절에서 언급한 바와 같이 Temkin(1978)에 의한 반응확률함수는 공변량 벡터  $\mathbf{Z}$ 가 주어졌을 경우에 다음과 같이 정의된다.

$$\begin{aligned} P(t; \mathbf{Z}) &= P(T_1 \leq t < T_4; \mathbf{Z}) \\ &= \int_0^t S_1(s-; \mathbf{Z}) S_2(s-; \mathbf{Z}) S_3(t|s; \mathbf{Z}) dA_1(s; \mathbf{Z}) \end{aligned} \quad (3.1)$$

여기서  $S_3(t|s; \mathbf{Z}) = \exp\left(-\int_s^t \alpha_3(x|s; \mathbf{Z}) dx\right)$  이고  $T_4$ 는 상태 2 또는 상태 3으로의 전이시간을 나타낸다.

Semi-Markov 회귀모형 하에서 이 반응확률함수의 추정량은 식 (2.32a)와 식 (2.51)을 식 (3.1)에 대입함으로써

$$\hat{P}(t; \mathbf{Z}) = \int_0^t \widehat{S}_1(s-; \mathbf{Z}) \widehat{S}_2(s-; \mathbf{Z}) \widehat{S}_3(t-s; \mathbf{Z}) d\widehat{A}_1(s; \mathbf{Z}) \quad (3.2)$$

으로 주어진다.

그런데 식 (3.2)에서  $\hat{P}(t; \mathbf{Z})$ 는 관측된 각 전이시간  $t_{ij}$  에서 값을 갖는 스텝함수(step function) 임을 알 수 있다. 따라서  $\hat{P}(t; \mathbf{Z})$ 는 실제계산을 위해 다음과 같이 표현할 수 있다.

시점  $t$  까지 관측된 모든 전이시점들  $t_{ij} (i=1, 2, 3; j=1, \dots, k_i)$ 에 대해서 이들 시점들을 크기 순으로 배열한 것을 예컨대,  $0 \equiv t_0 < t_1 < \dots < t_k \leq t$  라 두면

$$\begin{aligned} \hat{P}(t; \mathbf{Z}) = & \sum_{i=1}^k \left\{ \prod_{h=1}^{i-1} \left( 1 - \frac{\exp(\mathbf{Z}_1 \hat{\beta}) \Delta N_1(t_h)}{S_1^{(0)}(\hat{\beta}, t_h)} - \frac{\exp(\mathbf{Z}_2 \hat{\beta}) \Delta N_2(t_h)}{S_2^{(0)}(\hat{\beta}, t_h)} \right) \right. \\ & \left. \times \frac{\exp(\mathbf{Z}_1 \hat{\beta}) \Delta N_1(t_i)}{S_1^{(0)}(\hat{\beta}, t_i)} \prod_{h=i+1}^m \left( 1 - \frac{\exp(\mathbf{Z}_3 \hat{\beta}) \Delta N_3^*(t_h - t_i)}{S_3^{*(0)}(\hat{\beta}, t_h - t_i)} \right) \right\} \quad (3.3) \end{aligned}$$

이다.

한편  $P(t; \mathbf{Z})$ 의 대표본 성질은 Aalen과 Johansen(1978)과 Andersen et al.(1991)에서 연구된 전이확률함수의 대표본 성질에 대한 결과를 바탕으로 알 수 있다. 즉 표본의 크기  $n$ 이 클 때  $\sqrt{n}(\hat{P}(t; \mathbf{Z}) - P(t; \mathbf{Z}))$ 는 근사적으로 평균이 0이고 추정된 분산함수  $\hat{V}_1 + \hat{V}_2$  를 갖는 가우시안 과정을 따른다. 여기서

$$\hat{V}_1 = \mathbf{Q}'(t, \hat{\beta}; \mathbf{Z}) \mathbf{I}^{-1}(\hat{\beta}) \mathbf{Q}(t, \hat{\beta}; \mathbf{Z}), \quad (3.4)$$

$$\begin{aligned} \hat{V}_2 = & \int_0^t \{ \widehat{S}_1(s; \mathbf{Z}) \}^2 \{ \widehat{S}_2(s; \mathbf{Z}) \}^2 \{ \widehat{S}_3(t-s; \mathbf{Z}) - \hat{P}(t-s; \mathbf{Z}) \}^2 \\ & \times \frac{\exp(\mathbf{Z}_1 \hat{\beta}) J_0(s)}{\{ S_1^{(0)}(\hat{\beta}, s) \}^2} dN_1(s) \quad (3.5) \end{aligned}$$

이고,

$$\mathbf{Q} = \int_0^t \widehat{S}_0(s; \mathbf{Z}) \widehat{S}_1(s; \mathbf{Z}) d\mathbf{W}(s) \widehat{S}_3(t-s; \mathbf{Z}), \quad (3.6)$$

$$\mathbf{W}(t) = \exp(\mathbf{Z}_1 \hat{\beta}) \int_0^t \frac{(\mathbf{Z}_1 - \mathbf{E}_1^*(\hat{\beta}, s)) J_0(s)}{S_1^{(0)}(\hat{\beta}, s)} dN_1(s) \quad (3.7)$$

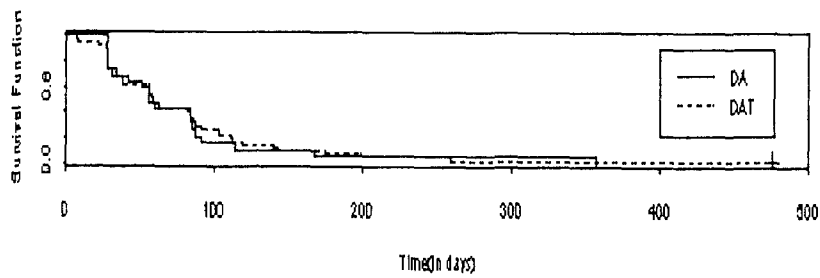
이다.

### 4. 적용 사례

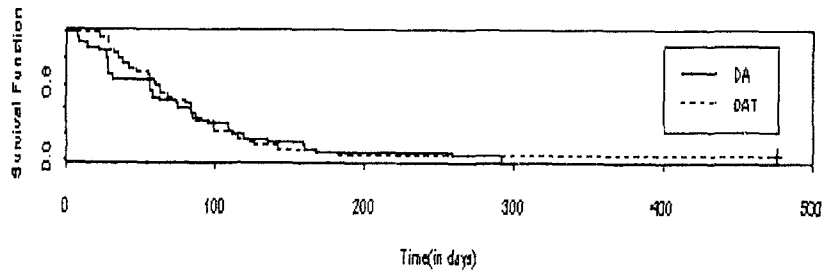
본 절에서는 앞에서 얻어진 결과를 Hsieh(1980)에서 주어진 실제자료에 적용하는 예를 보이고자 한다. 이 자료는 미국 위스콘신주에 있는 어떤 암 연구 센터에서 135명의 유방암환자들을 대

상으로 임상실험을 수행한 결과이다. 이 임상실험의 목적은 이들 유방암환자들에 대해서 두가지 처방약, 즉 Adriamycin과 Dibromodulcitol를 사용한 약(DA)과 이 약에 Tamoxifen이라는 약품을 첨가시킨 약(DAT)에 대해 치료효과의 차이를 조사하는 것이다.

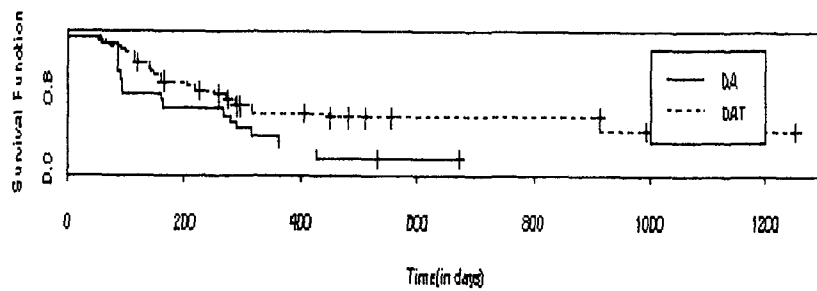
본 연구에서 환자들의 반응상태에 이르는 기준으로 ECOG(Eastern Cooperative Oncology Group)기준을 적용하였다. 135명의 환자들 중에서 임의로 55명과 80명으로 나누어서 각각 DA와 DAT를 처방하였다. DA를 처방받은 55명의 환자들 가운데 18명이 병세가 호전상태, 즉 반응상태에 이른 후 다시 재발되었고 2명은 반응을 보인 후로 다시 재발되지 않았으며 35명은 반응을 보



(a) 반응상태



(b) 악화상태



(c) 재발상태

그림 2. 각 상태에서의 생존함수

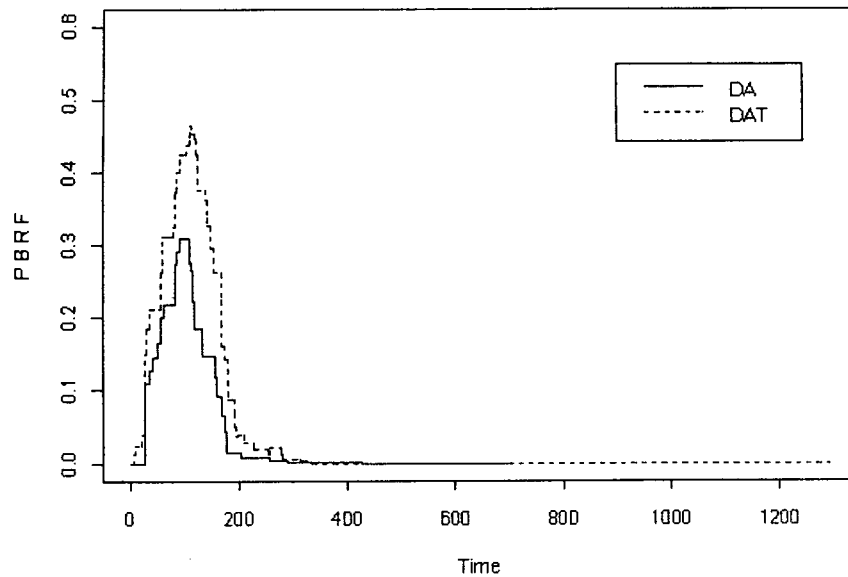


그림 3. 반응확률함수

이지 않고 악화되었다. 그리고 DAT를 처방받은 80명의 환자들 중에서 26명이 반응을 보인 후 다시 재발되었고 18명이 반응을 보인 후로 재발되지 않았고 35명이 반응없이 악화되었고 1명은 반응이나 악화되지를 않았다.

본 논문의 연구결과를 적용하기 위해서 DA와 DAT를 처방받은 이들 135명의 환자를 하나의 표본집단으로 간주한다. 그래서 식 (1.3)과 (1.4)에서 공변량  $Z$  를 이들 두 그룹을 구별하는 가변수로서 DA처방을 받은 환자들인 경우  $Z=0$ , DAT처방을 받은 환자들인 경우에는  $Z=1$ 이라 둔다. 그러면 식 (1.3)과 (1.4)는  $\alpha_i(t; Z) = \alpha_{i0}(t)e^{\beta_i Z}$  ( $i=1,2$ ),  $\alpha_3(t|s) = \alpha_{30}(t-s)e^{\beta_3 Z}$  으로 표현된다.

Hsieh(1980)에서 주어진 135명의 유방암 환자들에 대해 식 (2.49)를 이용해서 각 상태에 대응하는 생존함수를 추정한 그래프가 그림 2에서 보여준다. 이 그림의 (a)와 (b)에서 보는 것처럼 반응상태와 악화상태에 도달하는 경우에 DA와 DAT를 처방받은 두 그룹간에 별다른 차이가 없음이 보여진다. 그러나 이 그림의 (c)에서 보듯이 반응상태에서 재발상태로 옮겨가는 경우에 DAT처방을 받은 그룹의 환자들의 생존확률이 DA처방을 받은 그룹의 환자들 보다 높다는 것이 보여진다.

그림 2에서 나타난 결과를 수치적으로 살펴보기 위해서 식 (2.41)에서 주어진 방정식  $U_i(\beta_i) = 0$ 을 풀면  $\beta_i$  ( $i=1,2,3$ )의 추정치로서  $\hat{\beta}_1 = -0.0583$ ,  $\hat{\beta}_2 = -0.0743$ ,  $\hat{\beta}_3 = -0.692$  를 얻는다. 그리고 식 (2.42)로 부터  $\hat{\beta}_i$  ( $i=1,2,3$ )의 추정된 표준편차, 즉  $\widehat{s.d.}(\hat{\beta}_i) = I_i^{-\frac{1}{2}}(\hat{\beta}_i)$  는 각각 0.276, 0.238, 0.321 이 된다. 따라서, 예컨대 가설  $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ 을 검정하기 위해서 식 (2.43)을 이용하면 검정통계량의 값이  $Q_1 = 4.7894$  이 되고 따라서  $p$ 값이 0.1879가 되어서 DAT에 의한 처방이 DA처방에 비해 그다지 효과가 있

다고는 보여지지 않는다. 그런데 각 상태  $i=1,2,3$  에 대해서 가설  $H_0: \beta_i = 0$ ,  $H_1: \beta_i < 0$  에 대한  $p$ 값은 각각 0.4161, 0.3775, 0.0158 가 된다. 이 결과로 부터 Adriamycin과 Dibromodulcitol 에 Tamoxifen을 첨가한 것이 환자들의 반응상태로의 진행을 증가시키지는 않으나, 일단 반응상태에 도달된 후에 재발상태로의 진행은 억제시키는 효과가 있음을 보게된다.

한편 DA와 DAT를 처방받은 환자들을 종합적으로 비교하기 위한 방안으로 식 (3.3)을 이용하여 이들 두 그룹의 반응확률함수를 추정한 그래프가 그림 3에서 보여진다. 이 그림에서 보는 바와 같이, DAT처방을 받은 환자들의 반응확률이 DA처방을 받은 환자들보다 높음을 알 수 있다.

## 5. 결 론

본 논문에서 우리는 다중상태의 생존자료를 분석하기 위해서 semi-Markov 모형에 기초하여 Cox회귀모형을 적용하였다. 우리는 환자들에 대한 서로 다른 처방의 효과를 알아보기 위해 환자들의 병세의 호전상태를 여러 단계로 분류시킨 후 종합적으로 비교하는 것이 효과적임을 알 수 있다.

한편 본 논문에서 연구된 결과를 바탕으로 앞으로의 연구방향으로서 다음과 같은 경우를 고려할 수 있을 것이다. 첫째, 본 논문에서는 논의의 간편성을 위해 공변량  $Z$  가 전이시간  $t$  에 독립적임을 가정하였다. 그런데 공변량  $Z$  가 전이시간  $t$  에 의존하는 경우, 즉  $Z = Z(t)$ 인 경우에 대해서 큰 어려움이 없이 확장시킬 수 있을 것으로 본다. 둘째, 각 상태에 이르는 전이시간  $T_i$  ( $i=1,2,3$ )들이 관측되지 아니한 확률인자, 예컨대  $W$  에 관련되고  $W = w$  가 주어졌을 때 이들 전이시간  $T_i$  들이 조건부독립인 경우에 대한 모형을 고려할 수 있을 것이다. 최근에 이러한 소위 'Frailty모형'에 대한 연구가 활발히 진행되고 있음에 비추어 볼 때 이에 대한 연구는 나름대로 가치가 있을 것으로 생각된다.

## 참 고 문 헌

- [1] Aalen, O. O. (1975). Statistical Inference for a Family of Counting Processes, Ph.D. Dissertation, Department of Statistics, University of California, Berkeley.
- [2] Aalen, O. O. (1978). Non-parametric Inference for a Family of Counting Processes, *Annals of Statistics*, Vol. 6, 701-726.
- [3] Aalen, O. and Johansen, S.(1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian J. Statist.*, Vol. 5, 141-150.
- [4] Andersen, P. K. and Gill, R. D.(1982). Cox's Regression Model for Counting Processes : a Large Sample Study, *Annals of Statistics*, Vol. 10, 1100-1120.
- [5] Andersen, P. K. et al. (1988). Censoring, truncation and filtering in statistical models based on counting processes, *Contemp. Math.* Vol. 80, 19-60.

- [6] Andersen, P. K. et al. (1991). Non- and semi-parametric estimation of transition probabilities from censored observations of a non-homogeneous Markov process, *Scandinavian Journal of Statistics*. Vol. 18, 153-167.
- [7] Begg, C. B. and Larson, M.(1982). A Study of the Use of the Probability-of-Being-in-Response Function as a Summary of Tumor Response Data, *Biometrics*, Vol. 38, 59-66.
- [8] Cox, D. R.(1972). Regression Models and Life Tables (with discussion). *Journal of the Royal Statistical Society, B*, Vol. 34, 187-220.
- [9] Cox, D. R.(1975). Partial Likelihood, *Biometrika*, Vol. 62, 269-276.
- [10] Gill, R. D. (1984). Understanding Cox's Regression Model: a Martingale Approach, *Journal of the American Statistical Association*, Vol. 79, 441-447.
- [11] Gill, R. D. and Johansen, S. (1990). A survey of product-integration with a view towards application in survival analysis, *Annals of Statistics*, Vol. 18, 1501-1555.
- [12] Hsieh, F. Y.(1980). Multistate Survival Analysis, Ph.D. Thesis, Department of Statistics, University of Wisconsin-Madison.
- [13] Hsieh, F. Y., Crowley, J. and Tormey, D. C.(1983). Some Test Statistics for Use in Multistate Survival Analysis, *Biometrika*, Vol. 70, 111-119.
- [14] Johansen, S.(1978). The Product Limit Estimate as a Maximum Likelihood Estimate, *Scandinavian Journal of Statistics*, Vol. 5, 195-199.
- [15] Kalbfleisch, J. D. and Prentice, R. L.(1973). Marginal Likelihoods Based on Cox's Regression and life model. *Biometrika*, Vol. 60, 267-278.
- [16] Kalbfleisch, J. D. and Prentice, R. L.(1980). *The Statistical Analysis of Failure Time Data*, Wiley, New York.
- [17] Lagakos, S. W.(1977). Using Auxiliary Variables for Improved Estimates of Survival Time, *Biometrics*, Vol. 33, 399-404.
- [18] Lagakos, S. W., Sommer, C. J. and Zelen, M.(1978). Semi-Markov Models for Partially Censored Data, *Biometrika*, Vol. 65, 311-317.
- [19] Pepe, M. S.(1991). Inference for Events with Dependent Risks in Multiple Endpoint Studies, *Journal of the American Statistical Association*, Vol. 86, 770-778.
- [20] Temkin, N. R.(1978). An Analysis for Transient States with Application to Tumor Shrinkage, *Biometrics*, Vol. 34, 571-580.
- [21] Tsai, W. Y., Leurgans, S. and Crowley, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring, *Annals of Statistics*, Vol. 14, 1351-1365.
- [22] Voelkel, J. G. and Crowley, J.(1984). Non-parametric Inference for a Class of Semi-Markov Processes with Censored Observations, *Annals of Statistics*, Vol. 12, 142-160.
- [23] Yeo, S.(1994). On the Analysis of Multistate Survival Data using Cox's Regression Model, *The Korean Journal of Applied Statistics*, Vol. 7, 53-77.