

준모수적 혼합치료 프레일티 모형의 계층가능도 접근법

(Hierarchical-likelihood Approximation of Semi-parametric Mixture Cure Frailty Models)

김준철 (부경대 통계학과), 하일도 (부경대 통계학과)

혼합치료 프레일티 모형 (mixture cure frailty model)

- 충분한 시간이 경과하더라도 일부 환자가 치료가 됨으로 인해 결코 관심사건을 경험하지 않아 생존곡선이 0이 아닌 일정한 상숫값 (즉, 치료율: cure rate)으로 수렴하는 군집된 다변량 생존 자료 (clustered multivariate survival data) (예 : 일부 환자가 여러 번 재발을 경험한 후 치료되는 경우)를 분석하기 위한 계층 가능도 (hierarchical likelihood: h-likelihood: Lee & Nelder, 1996) 접근법을 제안
- 제안된 방법 : 어려운 적분 계산을 요구하지 않음

❖ 혼합치료 프레일티 모형

T_{ij} : i ($i = 1, 2, \dots, q$)번째 환자의 j ($j = 1, 2, \dots, n_i$)번째 생존시간

U_{ij} : i 번째 환자의 j 번째 관측되지 않는 치료상태 (치료되지 않으면 1, 치료되면 0)

V_i : i 번째 환자의 변량효과 (random effect)

(w_{ij}, x_{ij}) : 치료율과 위험률에 영향을 주는 공변량 벡터

$$S(t_{ij}|v_i; w_{ij}, x_{ij}) = 1 - \pi(w_{ij}|v_i) + \pi(w_{ij}|v_i) S_u(t_{ij}|v_i; x_{ij})$$

여기서, $S_u(t|v_i; x_{ij}) = \Pr[T > t|V_i = v_i; x_{ij}]$ 는 $\lim_{t \rightarrow \infty} S_u(t|v_i; x_{ij}) = 0$ 을 만족하는 치료되지 않은

환자의 조건부 생존함수이고 $1 - \pi(w_{ij}|v_i) = \Pr[U = 1|V_i = v_i; w_{ij}]$ 는 조건부 치료율

- 본 논문에서 $\pi(w_{ij}|v_i)$ 와 $S_u(t_{ij}|v_i; x_{ij})$ 를 다음과 같이 모형화함 :

$$\pi(w_{ij}|v_i) = \frac{\exp(\psi_{ij})}{1 + \exp(\psi_{ij})} \text{ and } S_u(t_{ij}|v_i; x_{ij}) = S_{0u}(t_{ij})^{\exp(\eta_{ij})}$$

여기서 $\psi_{ij} = w_{ij}^T \gamma + v_i$ 이고 $\eta_{ij} = x_{ij}^T \beta + \phi v_i$

γ : 공변량 w_{ij} 에 대응하는 절편을 포함하는 $p_1 \times 1$ 회귀계수 벡터

β : 공변량 x_{ij} 에 대응하는 절편을 포함하지 않는 $p_2 \times 1$ 회귀계수 벡터

$\phi \in R$: 두 모형을 연결하는 연관모수 (association parameter)

(여기서 $\phi > 0$ [< 0]이면 두 모형 간 양 [음]의 상관, $\phi = 0$ 이면 두 모형 간 무상관)

예를 들어, 두 모형 간 양의 상관이면 관심사건의 위험률 효과와 발생율 효과가 양의 상관을 의미 (즉, 위험률 효과와 치료율 효과는 음의 상관) (Lai와 Yau, 2008)

- 본 논문에서는 변량효과에 대해 $V_i \sim N(0, \alpha)$ 를 가정

(여기서 $\alpha \geq 0$; $\alpha = 0$ 이면 모든 환자 i 에 대해 변량효과 $v_i = 0$)

계층 가능도 접근법

- $Y_{ij} = \min(T_{ij}, C_{ij})$, $\delta_{ij} = I(T_{ij} \leq C_{ij})$ 여기서 C_{ij} : T_{ij} 에 대응하는 중도절단 지시함수

- 임의의 중도절단 하에서 공변량 (w_{ij}, x_{ij}) 가 주어졌을 때, 모수 $\Theta = (\gamma^T, \beta^T, \lambda_{0u}^T, v^T, \alpha, \phi)^T$ 에 대한 계층 가능도 (Ha et al, 2001) :

$$h = h(\Theta; \gamma, \delta, u, x, w) = \ell_1 + \ell_2 + \ell_3$$

여기서 $\ell_1 = \ell_1(\gamma; u|v, w) = \sum_{ij} [u_{ij} \psi_{ij} - \log\{1 + \exp(\psi_{ij})\}]$

$$\ell_2 = \ell_2(\beta, \lambda_{0u}, \phi; \gamma, \delta|v, u, x) = \sum_{ij} [\delta_{ij} \log\{\lambda_{0u}(y_{ij}) + \eta_{ij} + \log u_{ij}\} - \Lambda_{0u}(y_{ij}) \exp(\eta_{ij} + \log u_{ij})]$$

$$\ell_3 = \ell_3(\alpha; v) = -(q/2) \log(2\pi\alpha) - (2\alpha)^{-1} \sum_i v_i^2$$

(1) 회귀계수 및 프레일티 (frailty) 추정

- 단면 계층 가능도 (profile h-likelihood: Lee & Nelder, 1996; Ha et al, 2001)를 고려 :

$$h_p = \widetilde{\ell}_1 + \widetilde{\ell}_2 + \ell_3$$

여기서, $\widetilde{\ell}_1$: ℓ_1 에 관측되지 않는 u_{ij} 에 대해, 다음의 \widetilde{u}_{ij} 를 대입한 베르누이 계층 일반화 선형 모형 (hierarchical generalized linear model: HGLM)의 로그-가능도

$\widetilde{\ell}_2^*$: 다음의 Breslow 접근법 (1972, 1974)을 통하여 장애모수 (nuisance parameters) λ_{0uk} 들을 제거한 고정변수 $\log \widetilde{u}_{ij}$ 가 더해진 프레일티 Cox의 비례위험 모형 (Cox's proportional hazard ratio model: PH model)의 부분 (partial) 로그-가능도

$$\widetilde{u}_{ij} = E[U_{ij}|T_{ij} > y_{ij}] = \delta_{ij} + (1 - \delta_{ij}) \frac{\pi(w_{ij}|v_i) S_{0u}(y_{ij})^{\exp(\eta_{ij})}}{1 - \pi(w_{ij}|v_i) + \pi(w_{ij}; \gamma) S_{0u}(y_{ij})^{\exp(\eta_{ij})}}$$

$$\widehat{\Lambda}_{0u}(t) = \sum_{k: Y_{(k)} \leq t} \widehat{\lambda}_{0uk} = \sum_{k: Y_{(k)} \leq t} \frac{d_{(k)}}{\sum_{(i,j) \in R_{(k)}} \exp(\eta_{ij} + \log \widetilde{u}_{ij})}$$

- 본 논문에서는 $\lim_{t \rightarrow \infty} \widehat{S}_u(t|v; x) = 0$ 이 되도록 Lai와 Yau (2008) 의 제안에 따라 $t > \tau$ 일 때,

$\widehat{S}_{0u}(t) = \exp(-\widehat{\lambda}t)$ 를 가정 (Peng, 2003). 여기서 τ 는 중도절단 되지 않은 가장 큰 시간

- 단면 최대 계층 가능도 추정량 (profile MHLE) :

$$\widehat{\theta}^{(k+1)} = \theta^k + [H_p^{-1} S(\theta)]^{(k)} \quad (k = 0, 1, \dots)$$

여기서 $\theta = (\gamma^T, \beta^T, v^T)^T$.

그리고 $S(\theta) = \partial h_p / \partial \theta$ 와 $H_p = -\partial^2 h_p / \partial \theta \partial \theta^T$ 는 다음과 같음 :

$$S(\theta) = \begin{bmatrix} W^T a_1 \\ X^T a_2 \\ W^T a_1 + X^T a_2 - \alpha^{-1} v \end{bmatrix} \text{ and } H_p = \begin{bmatrix} W^T A_1 W & 0 & W^T A_1 Z \\ X^T A_2 X & X^T (\phi A_2) Z & \\ sym & Z^T A_1 Z + Z^T (\phi^2 A_2) Z + Q \end{bmatrix}$$

여기서 $a_1 = \widetilde{u} - \pi$, $a_2 = \delta - \mu^*$ 이고 $\mu^* = \widehat{\Lambda}_{0u}(y) \exp(\eta + \log \widetilde{u})$

$$A_1 = \text{diag}\{\pi_{ij}(1 - \pi_{ij})\}, A_2 = PB - PMDM^T P \text{이고 } Q = \alpha^{-1} q$$

$$P = \text{diag}\{\exp(\eta_{ij} + \log \widetilde{u}_{ij})\}, B = \text{diag}\{\widehat{\Lambda}_{0u}(y_{ij})\} \text{이고 } D = \text{diag}\{\widehat{\lambda}_{0uk}^2 / d_{(k)}\}$$

(2) 프레일티 분산 추정

- 단면 제한 가능도 (profile restricted likelihood: Ha et al, 2001, 2017)를 고려 :

$$p_\alpha = \left[h_p - \frac{1}{2} \log \det\{H_p / (2\pi)\} \right]_{\theta = \widehat{\theta}}$$

- α 의 추정방정식 :

$$\frac{\partial p_\alpha}{\partial \alpha} = \frac{\partial \widehat{h}_p}{\partial \alpha} - \frac{1}{2} \text{trace} \left\{ \widehat{H}_p \frac{\partial \widehat{H}_p}{\partial \alpha} \right\} = 0$$

여기서

$$\frac{\partial \widehat{H}_p}{\partial \alpha} = \frac{\partial H_p}{\partial \alpha} \Big|_{\theta = \widehat{\theta}} = \begin{bmatrix} W^T A'_1 W & 0 & W^T A'_1 Z \\ X^T A'_2 X & X^T (\phi A'_2) Z & \\ sym & Z^T A'_1 Z + Z^T (\phi^2 A'_2) Z + Q' \end{bmatrix}$$

$$A'_r = \left\{ \frac{\partial A_r}{\partial \alpha} + \frac{\partial A_r}{\partial \alpha} \frac{\partial v}{\partial \alpha} \right\} \Big|_{v = \widehat{v}} \quad (r = 1, 2) \quad \text{where} \quad \frac{\partial \widehat{v}}{\partial \alpha} = - \left(- \frac{\partial^2 h_p}{\partial v \partial v^T} \right) \left(- \frac{\partial^2 h_p}{\partial v \partial \alpha} \right) \Big|_{\theta = \widehat{\theta}}$$

(3) 연관 모수 추정

- grid search 방법으로 p_α 를 최대화 하는 $\widehat{\phi}$ 를 찾음

실제 자료 적합

- 스페인의 전향적 코호트 연구 (prospective cohort study)에서 사용된 403명의 대장암 (colorectal cancer) 환자에 대한 자료 (Gonzalez et al, 2005).
- 동일한 환자에 대하여 여러 번의 재발로 인한 총 861건의 재입원 사건 기록
- Tawiah (2020)에 따라 치료율을 수반하지 못하는 환자가 포함되지 않게 79번 환자 (3건) 제외
- 재발을 경험하지 못하게 하는 경쟁사건 (competing event)을 경험 (환자가 사망)한 112건 제외
- 총 364명의 환자에 대한 747건의 재입원 자료를 이용
- 생존시간은 대장암 수술 후 재입원하기까지의 시간 (중도절단 비율: 약 39.22%)
- Chemo 변수 : 1 (화학치료군), 0 (방사선치료군)
- R 패키지 “frailtypack”의 자료명 readmission에서 사용 가능.

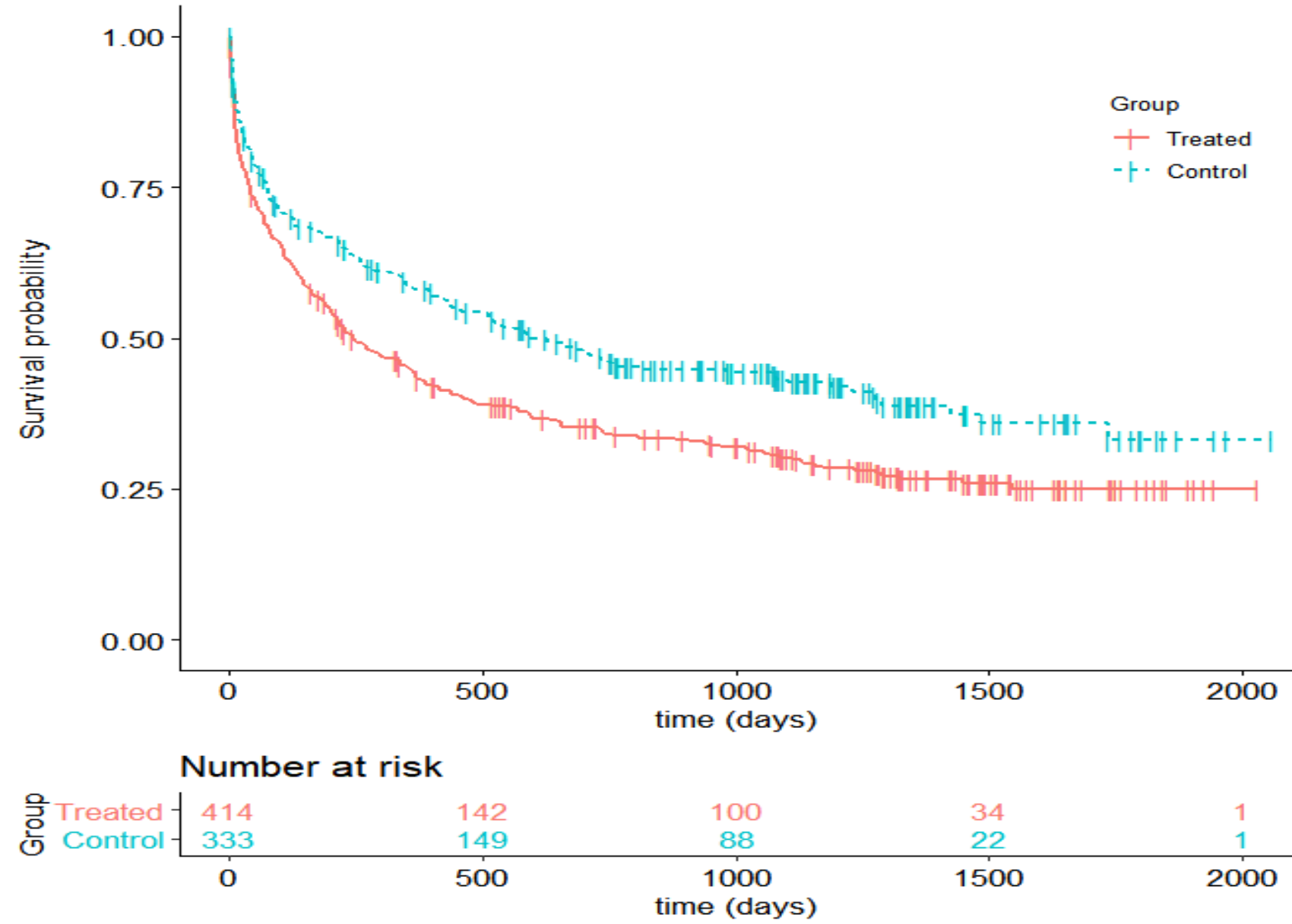


Figure. Kaplan-Meier survival curve of patients of the number of 364 for Treated (Chemo=1) or Control (Chemo=0)

Table. Estimation results of fitting three models with a single covariate (Chemo)

Para.	Proposed	PHMC	PH
	Est. (SE)	Est. (SE)*	Est. (SE)
γ_0	0.704 (0.159)**	1.222 (0.246)**	- -
γ_1	-0.426 (0.223)*	-0.388 (1.063)	- -
β_1	-0.172 (0.112)	-0.295 (0.161)**	-0.341 (0.097)**
α	1.650	- -	- -
ϕ	0.311	- -	- -

NOTE) Para. : parameters, PHMC: mixture cure model without frailty, PH : standard PH model, *: SD for bootstrap samples of the number of 1,000, **: p<0.05; *, p<0.1

Main references

- González, J. R., et al. (2005). Sex differences in hospital readmission among colorectal cancer patients. Journal of Epidemiology & Community Health, 59, 506-511.
- Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. Journal of the Royal Statistical Society: Series B (Methodological), 58, 619-656.
- Ha, I. D., Lee, Y. and Song, J. K. (2001). Hierarchical likelihood approach for frailty models. Biometrika, 88, 233-233.
- Lai, X. and Yau, K. K. (2008). Long-term survivor model with bivariate random effects: Applications to bone marrow transplant and carcinoma study data. Statistics in Medicine, 27, 5692-5708