

Estimation of Odds Ratio in Proportional Odds Model

Min-Ja Seo¹⁾ · Ju-Sung Kim²⁾

Abstract

Although the proportional hazards model is the most common approach used for studying the relationship of event times and covariates, alternative models are needed for occasions when it does not fit data. In the two-sample case, proportional odds models are useful for fitting data whose hazard rates converge asymptotically. In this thesis, we propose a new estimator of the relative odds ratio of the proportional odds model when two independent random samples are observed under uncensorship. We prove the asymptotic normality and consistency of the estimator by using martingale-representation. The efficiency of the proposed is assessed through a simulation study.

Keywords : 마팅게일, 비례오즈모형, 생존함수, 오즈비, 정규과정

1. 머리말

환자의 생존시간과 인자들(설명변수) 사이의 관계를 설명하는데 가장 유용한 모형은 Cox의 비례위험모형(proportional hazard model) $\lambda_i(t) = \lambda_0(t) \exp\left(\sum_{j=1}^p \beta_j Z_{ji}\right)$ 이다. 여기서 $\lambda_i(t)$ 는 i 번째 개체에 대한 위험함수이고 $\lambda_0(t)$ 는 기저 위험함수를 나타내며, Z_{ji} 는 i 번째 개체에 관한 j 번째 설명변수의 값이다. 그리고 β_1, \dots, β_p 는 회귀계수이다. 그러나 시점 t 에 대하여, $\frac{\lambda_i(t)}{\lambda_0(t)}$ 가 상수인 비례위험모형은 위험비가 $t \rightarrow \infty$ 일 때 1로 수렴하는 경우를 포함하지 않고 있다. 둘 이상의 집단에 대하여 얻어진 위험함수들의 비가 시간이 지남에 따라 수렴하는 경우가 종종 발생한다. 예를 들면 치료에 대한 초기효과나 진단 시 환자들의 질병상태의 차이들이 시간이 지남에 따라 사라지는 경향이 있을 때 발생하는데, 이런 경우 상수 위험비의 가정은 분명히 부적절하다. 비슷하게 실험군(disease group)의 위험이 대조군(disease-free control group)의 위험에

1) 청주시 흥덕구 개신동 산 48번지 충북대학교 정보통계학과 이학박사

2) (교신저자) 청주시 흥덕구 개신동 산 48번지 충북대학교 정보통계학과 교수
E-mail : kimjs@chungbuk.ac.kr

근접할 때 치료효과에 대한 설명(표현)은 위험비가 1로 수렴함을 요구한다. 치료효과가 사라지는 경우에 대한 실제 예는 Anderson 등(1992)에 의한 악성 흑색종(피부암의 일종) 자료의 분석에서 발견할 수 있다. 일반적으로 위험함수들의 비가 수렴하는 경우, Cox의 비례위험모형을 시간에 영향을 받는 설명변수들을 포함하는 것으로 확장시켜 모형화 했다. 비록 비례위험모형이 사건시간과 공변량의 관계를 연구하는데 사용되어지는 가장 유용한 접근법일지라도, 그것이 자료를 잘못 적합시켰을 경우 다른 모형이 필요하다. 이-표본(two-sample)인 경우에 비례오즈모형이 두 위험률의 비가 수렴하는 경우의 자료를 적합시키는데 유용하다. 만일, 생존시간을 나타내는 두 확률변수 T_1 과 T_2 가 독립이고, 각각 분포함수 F_1 과 F_2 를 가지면, T_1 과 T_2 에 대한 생존함수(survival function)는 각각 $S_1(t) = \Pr\{T_1 > t\} = 1 - F_1(t)$, $S_2(t) = \Pr\{T_2 > t\} = 1 - F_2(t)$ 이고, 생존시간이 t 보다 큰 개체에 대한 오즈(odds)는

각각 $\phi_1(t) = \frac{S_1(t)}{1 - S_1(t)}$, $\phi_2(t) = \frac{S_2(t)}{1 - S_2(t)}$ 이다. 이때, 모든 $t > 0$ 와 어떤 상수 $\theta > 0$

에 대하여, $\phi_2(t) = \theta\phi_1(t)$ 이면 F_1 과 F_2 는 비례오즈모형(proportional odds model)이다. 만일 $\lambda_1(t)$ 와 $\lambda_2(t)$ 가 각각 T_1 과 T_2 에 대한 위험함수이면 비례오즈모형 하에서

위험비는 $\frac{\lambda_2(t)}{\lambda_1(t)} = \frac{1}{1 - (\theta - 1)S_1(t)}$ 이다. 따라서 $t \rightarrow \infty$ 일 때, $\theta > 1$ 이면 위험비 $\frac{\lambda_2(t)}{\lambda_1(t)}$

는 증가하면서 1로 수렴하고, $\theta < 1$ 이면 감소하면서 1로 수렴한다. Bennett(1983a)에 의하여 생존분석에 도입된 비례오즈모형은 비례위험모형의 대안으로 각광을 받고 있으며, 지금까지 많은 연구가 이루어져 왔다. Bennett(1983a)는 여러 공변량 값이 각 구간 상에서 측정되어져 있거나, 개체들의 집단이 두개 이상일 때 비례오즈회귀모형이 적합될 수 있음을 보여 주었다. Dabrowska와 Doksum(1988)은 이-표본 일반화 오즈비모형에서, 스코어함수를 이용하여 상대오즈비를 추정하였으며, Wu(1995)는 이-표본 비례오즈모형에서 초기 \sqrt{n} -일치추정치를 사용한 1단계(one-step)추정치를 기본으로 하는 상대오즈비에 대한 추정방법을 제시하였다. Rossini와 Tsiatis(1996)는 구간중도절단된 자료의 분석에서 비례오즈모형을 사용하였고, Huang과 Rossini(1997)는 구간중도절단을 갖는 비례오즈실패-시간회귀모형에 대하여 체(sieve) 최우추정량을 사용하여 회귀계수를 추정하였다. Murphy(1997)는 Bennett(1983b)의 준모수적 최우추정법이 회귀계수에 대한 효율적인 추정량을 제공한다는 사실을 보였으며, 우측중도절단자료를 갖는 비례오즈모형에서 모수에 대한 최우추정법을 고려하였다. Song Yang과 Prentice(1999)는 우측 중도절단자료에 대하여 비례오즈회귀모형을 적합시키기 위하여, 가중경험오즈함수들을 사용하였다. Kirmani와 Gupta(2001)는 비례오즈모형에 대한 구조, 관계, 특성 등에 대하여 설명하였다.

본 논문의 목적은 관심 있는 두 군에 대한 위험함수들이 시간에 대하여 비례하는 것이 아닌 그들의 비가 1로 가는 경향이 있을 때, 중도절단자료를 갖지 않는 경우에 대한 상대오즈비를 추정하는 새로운 방법을 제시하는 것이다. 본 논문의 2절에서는 새로운 추정량을 제안하고 이 제안된 추정량의 일치성과 근사적 정규성을 마팅게일 방법을 이용하여 보인다. 또한 3절에서는 우리가 제안한 추정량에 대한 유효성을 로그-로지스틱분포를 사용하여 모의실험을 통해 제안한 추정량과 Dabrowska-Doksum의 추정량에 대한 각각의 MSE(Mean Square Error)를 비교하고 근사적 정규성에 대한 시각적 확인을 위해 히스토그램을 그려본다.

2. 추정량의 제안과 정규 근사

생존시간을 나타내는 독립된 양의 확률변수들 T_{ij} ($i = 1, 2, j = 1, \dots, n_i$) 가 연속생존함수 $S_i(t) = pr(T_{ij} > t)$, 분포함수 $F_i(t) = 1 - S_i(t)$, 비례오즈함수 $\phi_i(t) = \frac{S_i(t)}{1 - S_i(t)}$ 누적위험함수 $\Lambda_i(t) = -\log S_i(t)$ ($i = 1, 2$)를 갖는다고 하자. 또한 생존시간 T_{ij} 는 중도절단 되지 않았다고 가정하자. 만일 두 군이 비례오즈모형을 만족한다면 모든 $t > 0$ 와 어떤 $\theta > 0$ 에 대해서 식 $\phi_2(t) = \theta\phi_1(t)$ 이 성립한다. 따라서

$$\frac{S_2(t)}{S_2(t) - 1} = \theta \frac{S_1(t)}{S_1(t) - 1} \tag{2-1}$$

이다. 한편, $\int_0^\infty -dS_1(t) = 1$ 이므로 식 (2-1)로부터 다음의 식 (2-2)가 만족된다.

$$\theta = \int_0^\infty [1 - S_1(t)]^2 d\left[\frac{S_2(t)}{S_2(t) - 1}\right] \tag{2-2}$$

식 (2-2)에서 우리는 $S_i(t)$ ($i = 1, 2$) 대신에 경험생존함수 $S_{n_i}(t) = \frac{1}{n_i} \sum_{j=1}^{n_i} I_{\{T_{ij} > t\}}$ 를 대입한 다음의 식 (2-3)을 식 (2-2)에 대한 추정량으로 제안한다.

$$\hat{\theta} = \int_0^\infty [1 - S_{n_1}(t)]^2 d\left[\frac{S_{n_2}(t)}{S_{n_2}(t) - 1}\right] \tag{2-3}$$

제안된 추정량은 다음의 중요한 특성을 갖는다.

정리 군 $i = 1, 2$ 에 대하여 각각의 표본의 크기를 n_i ($i = 1, 2$)라고 할 때, $n = n_1 + n_2$ 라 하고, $\lim_{n_i \rightarrow \infty} \frac{n}{n_i} = \rho_i$ 이라고 하자. 그러면

- (1) $\hat{\theta}$ 는 θ 의 일치추정량이다.
- (2) $\sqrt{n}(\hat{\theta} - \theta)$ 의 근사분포는 평균이 0, 분산 σ^2 이 아래와 같은 정규분포를 따른다.

$$\sigma^2 = 4\rho_1\theta^2 \int_0^\infty S_1(t)^2 dF_1(t) + \rho_2 \int_0^\infty \left(\frac{S_1(t) - 1}{S_2(t) - 1}\right)^4 [2S_2(t) + 1] dF_2(t).$$

증명: (1) 군 $i = 1, 2$ 에 대하여 각각의 표본의 크기를 n_i ($i = 1, 2$)라 할 때 $n = n_1 + n_2$ 라 하고 $\lim_{n_i \rightarrow \infty} \frac{n}{n_i} = \rho_i$ 이라고 가정하자. 그리고 $F_{n_i}(t) = \frac{1}{n_i} \sum_{j=1}^{n_i} I_{\{T_{ij} \leq t\}}$ 을 군

i 에 대한 확률표본 T_{i1}, \dots, T_{in_i} 의 경험분포함수라 하고 $S_{n_i}(t) = \frac{1}{n_i} \sum_{j=1}^{n_i} I_{\{T_{ij} > t\}}$ 를 경험생존함수라고 하자.

(1)의 증명

$S_{n_i}(t)$ 는 $S_i(t)$ 의 일치추정량이라는 사실로부터 $\hat{\theta}$ 이 θ 의 일치추정량이다.[Serfling, 1980 pp.24 Theorem].

(2)의 증명

$i=1, 2$ 에 대하여, $N_i(t) = \sum_{j=1}^{n_i} I_{\{T_{ij} \leq t\}}$ 를 군 i 에서의 시점 t 또는 그 이전의 사망자의 총수라고 하고, $Y_i(t) = \sum_{j=1}^{n_i} I_{\{T_{ij} \geq t\}}$ 를 군 i 에서의 시점 t 에서 위험과정 상태의 총수라 하자. 그러면, 가정에 의하여 $\frac{N_i(t)}{n_i} = F_{n_i}(t)$ 이고, $\frac{Y_i(t-)}{n_i} = 1 - F_{n_i}(t) = S_{n_i}(t)$ 이다. 과정 $M_{n_i}(t) = \sqrt{n_i} \left[F_{n_i}(t) - \int_0^t [1 - F_{n_i}(y)] d\Lambda_i(y) \right]$ 는 이차적분가능 마팅계일이며 평균 0이고 분산이 $V_{M_i}(t) = \int_0^t [1 - \Delta\Lambda_i(y)] dF_i(y)$ 인 $M_{n_i}(t)$ 의 정규극한과정 $M_i(t)$ 가 존재한다. 여기서 Λ_i 는 군 i 에서의 누적위험함수를 표시 한다. 기호 \xrightarrow{p} 와 \xrightarrow{d} 는 확률수렴과 분포수렴을 각각 나타낸다.

이제, $Z_n = \sqrt{n}(\hat{\theta} - \theta)$ 를 살펴보면 다음과 같다.

$$\begin{aligned} Z_n &= \sqrt{n}(\hat{\theta} - \theta) \\ &= \sqrt{n} \left[\int_0^\infty [1 - S_{n_1}(t)]^2 d\left(\frac{S_{n_2}(t)}{S_{n_2}(t) - 1}\right) - \int_0^\infty [1 - S_1(t)]^2 d\left(\frac{S_2(t)}{S_2(t) - 1}\right) \right] \\ &= \sqrt{n} \int_0^\infty \left[(1 - S_{n_1}(t))^2 - (1 - S_1(t))^2 \right] d\left(\frac{S_2(t)}{S_2(t) - 1}\right) \\ &\quad + \sqrt{n} \int_0^\infty [1 - S_{n_1}(t)]^2 d\left(\frac{S_2(t)}{1 - S_2(t)} - \frac{S_{n_2}(t)}{1 - S_{n_2}(t)}\right) \end{aligned}$$

식 Z_n 의 전반부를 Z_1 , 후반부를 Z_2 라고 하면, Z_1 은 다음의 식과 같다.

$$\begin{aligned}
 Z_1 &= \sqrt{n} \int_0^\infty [(1 - S_{n_1}(t))^2 - (1 - S_1(t))^2] d\left(\frac{S_2(t)}{S_2(t) - 1}\right) \\
 &= \sqrt{n} \int_0^\infty [S_1(t) - S_{n_1}(t)] [2 - S_1(t) - S_{n_1}(t)] d\left(\frac{S_2(t)}{S_2(t) - 1}\right) \\
 &= \sqrt{\frac{n}{n_1}} \int_0^\infty \frac{\sqrt{n_1} [F_{n_1}(t) - F_1(t)]}{1 - F_1(t)} [1 - F_1(t)] [F_{n_1}(t) + F_1(t)] d\left(\frac{S_2(t)}{S_2(t) - 1}\right) \\
 &= \sqrt{\frac{n}{n_1}} \int_0^\infty \left[\int_0^t \frac{1}{1 - F_1(y)} dM_{n_1}(y) \right] [1 - F_1(t)] [F_{n_1}(t) + F_1(t)] d\left(\frac{S_2(t)}{S_2(t) - 1}\right) \\
 &= \sqrt{\frac{n}{n_1}} \int_0^\infty \frac{S_2(t)}{1 - S_2(t)} d\left[1 - F_1(t)\right] [F_{n_1}(t) + F_1(t)] \int_0^t \frac{1}{1 - F_1(y)} dM_{n_1}(y)
 \end{aligned}$$

마팅계일 $M_{n_1}(t)$ 는 평균 0, 분산이 $V_{M_1}(t) = \int_0^t [1 - \Delta A_1(y)] dF_1(y)$ 인 정규과정 $M_1(t)$ 에 수렴한다. 그리고 $F_{n_1}(t)$ 가 $F(t)$ 의 경험분포함수이므로 $F_{n_1}(t) \xrightarrow{P} F_1(t)$ 이다. 또한 가정에 의하여 $\lim_{n_1 \rightarrow \infty} \frac{n}{n_1} = \rho_1$ 이다. 따라서 Mann-Wald 정리와 Slutsky 정리에 의하여 다음의 관계가 성립한다.

$$Z_1 \xrightarrow{d} 2 \sqrt{\rho_1} \int_0^\infty \frac{S_2(t)}{1 - S_2(t)} d\left[[1 - F_1(t)] F_1(t) \int_0^t \frac{1}{1 - F_1(y)} dM_1(y) \right].$$

그리고 Z_1 의 극한분포는 평균이 0이고, 분산 σ^2 이 다음과 같은 정규분포이다.

$$\begin{aligned}
 \sigma_1^2 &= (2 \sqrt{\rho_1})^2 \int_0^\infty \left(\frac{S_2(t)}{1 - S_2(t)}\right)^2 d\left[[1 - F_1(t)]^2 F_1^2(t) \int_0^t \left(\frac{1}{1 - F_1(y)}\right)^2 dV_{M_1}(y) \right] \\
 &= 4\rho_1 \int_0^\infty \left(\frac{S_2(t)}{1 - S_2(t)}\right)^2 d\left[[1 - F_1(t)]^2 F_1^2(t) \int_0^t \left(\frac{1}{1 - F_1(y)}\right)^2 d \int_0^y \frac{1 - F_1(u)}{1 - F_1(u-)} dF_1(u) \right] \\
 &= 4\rho_1 \int_0^\infty \left(\frac{S_2(t)}{1 - S_2(t)}\right)^2 d\left[[1 - F_1(t)]^2 F_1^2(t) \int_0^t d\left(\frac{1}{1 - F_1(y)}\right) \right] \\
 &= 4\rho_1 \int_0^\infty \left(\frac{S_2(t)}{1 - S_2(t)}\right) \\
 &= 4\rho_1 \theta^2 \int_0^\infty S_1^2(t) [4S_1(t) - 1] dF_1(t). \tag{2-4}
 \end{aligned}$$

Z_2 에 대한 다음의 식을 생각해보자.

$$Z_2 = \sqrt{n} \int_0^\infty [1 - S_{n_1}(t)]^2 d\left(\frac{S_2(t)}{1 - S_2(t)} - \frac{S_{n_2}(t)}{1 - S_{n_2}(t)}\right).$$

먼저, $\frac{S_2(t)}{1-S_2(t)} - \frac{S_{n_2}(t)}{1-S_{n_2}(t)}$ 를 계산해 보면 다음과 같다.

$$\begin{aligned} \frac{S_2(t)}{1-S_2(t)} - \frac{S_{n_2}(t)}{1-S_{n_2}(t)} &= \frac{S_2(t)}{[1-S_2(t)][1-S_{n_2}(t)]} \frac{S_2(t) - S_{n_2}(t)}{S_2(t)} \\ &= \frac{S_2(t)}{[1-S_2(t)][1-S_{n_2}(t)]} \frac{[F_{n_2}(t) - F_2(t)]}{1-F_2(t)} \end{aligned} \quad (2-5)$$

따라서, Z_2 는 다음의 식을 만족한다.

$$\begin{aligned} Z_2 &= \sqrt{\frac{n}{n_2}} \int_0^\infty [1-S_{n_1}(t)]^2 d \left[\frac{S_2(t)}{[1-S_2(t)][1-S_{n_2}(t)]} \frac{\sqrt{n_2} [F_{n_2}(t) - F_2(t)]}{1-F_2(t)} \right] \\ &= \sqrt{\frac{n}{n_2}} \int_0^\infty [1-S_{n_1}(t)]^2 d \left[\frac{S_2(t)}{[1-S_2(t)][1-S_{n_2}(t)]} \int_0^t \frac{1}{1-F_2(y)} dM_{n_2}(y) \right]. \end{aligned} \quad (2-6)$$

마팅계일 $M_{n_2}(t)$ 는 평균 0, 분산 $V_{M_2}(t)$ 가 $V_{M_2}(t) = \int_0^t [1 - \Delta A_2(y)] dF_2(y)$ 인 정규극한과정 $M_2(t)$ 를 갖는다. 그리고 $S_{n_i}(t)$ ($i=1, 2$)가 $S_i(t)$ 의 경험생존함수이므로, $1-S_{n_i}(t) \xrightarrow{p} 1-S_i(t)$ 이다. 또한 가정에 의하여 $\lim_{n_2 \rightarrow \infty} \frac{n}{n_2} = \rho_2$ 이다. 따라서 Mann-Wald 정리와 Slutsky 정리에 의하여 식 (2-6)은 다음의 관계를 만족한다.

$$Z_2 \xrightarrow{d} \sqrt{\rho_2} \int_0^\infty [1-S_1(t)]^2 d \left[\frac{S_2(t)}{[1-S_2(t)]^2} \int_0^t \frac{1}{1-F_2(y)} dM_2(y) \right].$$

$M_2(t)$ 의 평균이 0이므로, Z_2 의 극한분포는 평균 0, σ_2^2 이 다음과 같은 정규분포이다.

$$\begin{aligned} \sigma_2^2 &= (\sqrt{\rho_2})^2 \int_0^\infty [1-S_1(t)]^4 d \left[\left(\frac{S_2(t)}{[1-S_2(t)]^2} \right)^2 \int_0^t \left(\frac{1}{1-F_2(y)} \right)^2 dV_{M_2}(y) \right] \\ &= \rho_2 \int_0^\infty [1-S_1(t)]^4 d \left[\left(\frac{S_2(t)}{[1-S_2(t)]^2} \right)^2 \int_0^t \frac{1}{[1-F_2(y)]^2} \frac{1-F_2(y)}{1-F_2(y-)} dF_2(y) \right] \\ &= \rho_2 \int_0^\infty [1-S_1(t)]^4 d \left[\left(\frac{S_2(t)}{[1-S_2(t)]^2} \right)^2 \int_0^t d \left(\frac{1}{1-F_2(y)} \right) \right] \\ &= \rho_2 \int_0^\infty \left[\frac{1-S_1(t)}{1-S_2(t)} \right]^4 [2S_2(t) + 1] dF_2(t) \end{aligned} \quad (2-7)$$

그런데 $M_1(t)$ 와 $M_2(t)$ 이 서로 독립이므로, $\sqrt{n}(\hat{\theta} - \theta) = Z_1 + Z_2$ 의 극한분포는 평균 0, 식 (2-4)와식 (2-7)로부터 분산 $\sigma^2 = \sigma_1^2 + \sigma_2^2$ 이 다음과 같은 정규분포이다.

$$\sigma^2 = 4\rho_1\theta^2 \int_0^\infty S_1^2(t)[4S_1(t) - 1] dF_1(t) + \rho_2 \int_0^\infty \left(\frac{1 - S_1(t)}{1 - S_2(t)}\right)^4 [2S_2(t) + 1] dF_2(t).$$

3. 모의실험결과

이 절에서 우리는 제안된 추정량의 근사적 결과에 대한 평가를 위하여 모의실험을 실시하였다. 모의실험은 로그-로지스틱 분포에서 이루어졌는데, 그 이유는 이 모형이 와이블 모형과 지수모형처럼 생존함수와 위험함수에 대한 용이한 대수적 표현을 갖는 장점 때문이다. 로그-로지스틱분포에서, 생존함수와 분포함수 그리고 오즈함수와 위험함수는 다음과 같다.

$$\begin{aligned} S(t) &= \left[1 + \left(\frac{t}{\alpha}\right)^\beta\right]^{-1} && \alpha > 0, \quad \beta > 0, \quad t > 0. \\ F(t) &= \left(\frac{t}{\alpha}\right)^\beta \left[1 + \left(\frac{t}{\alpha}\right)^\beta\right]^{-1} \\ \phi(t) &= \left(\frac{t}{\alpha}\right)^{-\beta} \\ \lambda(t) &= \frac{\beta t^{\beta-1}}{\alpha^\beta + t^\beta} \end{aligned}$$

여기서, β 는 형태모수(shape parameter)이고, α 는 위치모수(location parameter)이다.

만일 $\beta \leq 1$ 이면 위험률은 시간이 경과함에 따라 감소하고, $\beta > 1$ 이면 $\lambda(0) = 0$ 이고 $\lim_{t \rightarrow \infty} \lambda(t) = 0$ 이다. 두 표본과 관련된 모의실험을 위하여, β 는 동일한 것으로 간주할

것이다. 표본1에 대한 위험함수는 $\lambda_1(t) = \frac{\beta t^{\beta-1}}{\alpha_1^\beta + t^\beta}$ 이고, 표본2에 대한 위험함수는

$$\lambda_2(t) = \frac{\beta t^{\beta-1}}{\alpha_2^\beta + t^\beta} \text{이므로, 두 표본에 대한 위험비는 } \frac{\lambda_2(t)}{\lambda_1(t)} = \frac{\beta t^{\beta-1}}{\alpha_1^\beta + t^\beta} \frac{\alpha_2^\beta + t^\beta}{\beta t^{\beta-1}} = \frac{\alpha_2^\beta + t^\beta}{\alpha_1^\beta + t^\beta}$$

이다. 따라서 $\lim_{t \rightarrow \infty} \frac{\lambda_2(t)}{\lambda_1(t)} = 1$ 이다. 다시 말하면, 시간이 커짐에 따라 1로 수렴한다. 그

러므로 비례오즈모형을 따른다는 가설 하에서 로그-로지스틱분포에서 모의실험을 실행할 수 있다.

먼저 두 모집단의 분포를 각각 로그-로지스틱분포에서, 모수 $\theta = \frac{\phi_2(t)}{\phi_1(t)} = \left(\frac{\alpha_2}{\alpha_1}\right)^\beta$ 로부터 $\theta = 0.25, \theta = 0.5, \theta = 0.75, \theta = 1, \theta = 1.25, \theta = 1.5$ 가 되도록 α_1 과 α_2 를 정하여, 표본1의 자료수 n_1 과 표본2의 자료수 n_2 로 한 표본 (n_1, n_2) 를 (10, 10), (10, 20), (30, 30), (30, 45), (30, 60), (100, 120)씩 5000번 발생시켜 몬테카를로 방법을 사용하여 모수를 추정하였다. θ 에 대한 각각의 추정치 $\hat{\theta}$ 에 대한 MSE (mean

squared error)를 제안한 추정치와 Dabrowska-Doksum의 추정치에 대하여 구하고 <표 1>에 나타내어 비교하였다. 그리고 생성된 자료로부터 정리에서 증명한 근사적으로 평균 0을 갖는 정규성을 검토하기 위하여, $\theta = 0.25$ 일 때 $(\hat{\theta} - \theta)$ 의 값 5000개 대한 평균과 표준오차를 계산하여 <표 2>에, 히스토그램을 <그림 1>에 각각 나타내었다. 다음의 <표 1>을 보면 우리가 제안한 추정치에 대한 MSE가 Dabrowska-Doksum의 추정치에 대한 MSE보다 모두 작은 것으로 나타난다. 이것은 우리의 추정량이 더 효율적임을 의미한다. 그리고 θ 의 값이 클수록 MSE의 값이 두 추정량 모두 커짐을 알 수 있다. 그런데 우리가 제안한 추정치에 대한 MSE보다 Dabrowska-Doksum의 추정치에 대한 MSE가 더 커짐을 알 수 있다. 또한, 같은 모수에 대하여 자료의 수가 클수록 MSE가 작아짐을 보여준다.

<표1 > 제안된 추정치와 Dabrowska-Doksum의 추정치에 대한 MSE 비교

모수 (θ)	(n_1, n_2)	MSE		모수 (θ)	(n_1, n_2)	MSE	
		제안된 추정치	Doksum 추정치			제안된 추정치	Doksum 추정치
0.25	(10,10)	0.269	0.421	1	(10,10)	5.322	12.470
	(10,20)	0.230	0.408		(10,20)	5.320	11.430
	(30,30)	0.222	0.396		(30,30)	5.168	11.060
	(30,45)	0.211	0.387		(30,45)	5.164	10.574
	(30,60)	0.189	0.384		(30,60)	4.907	9.904
	(100,120)	0.154	0.382		(100,120)	1.941	4.441
0.5	(10,10)	1.035	2.434	1.25	(10,10)	9.592	19.510
	(10,20)	1.030	2.360		(10,20)	9.438	17.910
	(30,30)	0.989	2.280		(30,30)	9.337	17.220
	(30,45)	0.914	2.256		(30,45)	9.300	16.360
	(30,60)	0.866	2.238		(30,60)	8.916	14.990
	(100,120)	0.742	1.719		(100,120)	2.687	5.801
0.75	(10,10)	2.602	6.594	1.5	(10,10)	16.020	27.540
	(10,20)	2.585	6.112		(10,20)	15.550	25.130
	(30,30)	2.541	5.947		(30,30)	15.530	23.920
	(30,45)	2.439	5.757		(30,45)	15.480	22.650
	(30,60)	2.374	5.515		(30,60)	15.070	20.550
	(100,120)	1.303	3.064		(100,120)	3.519	7.087

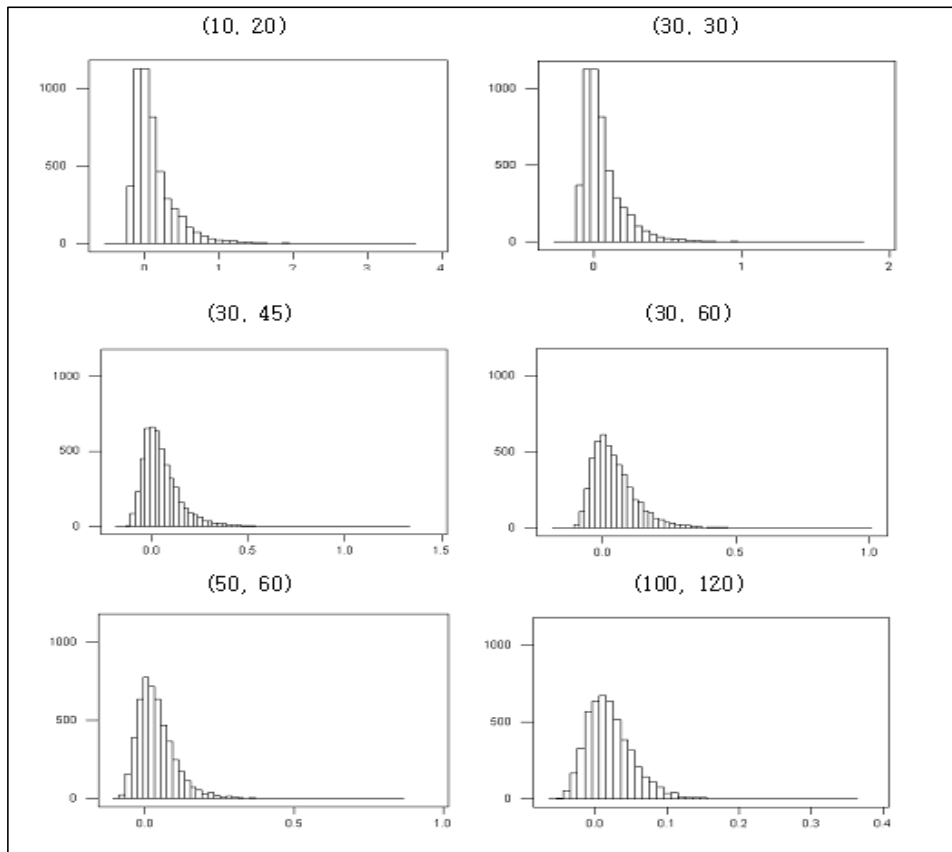
다음의 <표 2>를 보면 자료의 수가 커질수록 평균은 0값에 근사함을 알 수 있다.

수렴정도는 약하지만 평균이 근사적으로 0이 된다고 말할 수 있다.

<표 2> 모수가 $\theta = 0.25$ 일 때 $(\hat{\theta} - \theta)$ 에 대한 평균과 표준오차

(n_1, n_2)	(10, 20)	(30, 30)	(30, 45)	(30, 60)	(50, 60)	(100, 120)
평균	0.128	0.064	0.051	0.046	0.036	0.02095
표준오차	0.004	0.002	0.001	0.001	0.00099	0.00049

근사적 정규성에 대한 검토를 위한 다음 <그림 1>을 보면 모두 0을 중심으로 오른쪽 꼬리가 긴 모양을 나타낸다. 그런데 두 집단의 자료의 총수가 커질수록 오른쪽 꼬리가 점점 짧아지고 모양이 0을 중심으로 좌우 대칭인 경향을 보인다. 따라서 그림의 결과는 평균 0을 갖는 근사적 정규성을 나타낸다고 말할 수 있다.



<그림 1> 모수가 $\theta = 0.25$ 일 때 $(\hat{\theta} - \theta)$ 에 대한 히스토그램

앞으로 비례오즈모형 하에서 완비자료에 대한 상대오즈 비 추정에서와 마찬가지로 중도 절단된 자료의 경우에 대한 상대오즈 비 추정과 관련된 연구도 대단히 유용할 것

으로 생각한다.

참고문헌

1. Aalen, Odd (1978), Nonparametric Inference for Family of Counting Processes, *The Annals of Statistics*, 6, No. 4, 701-726.
2. Andersen, P. K. and Ø. Borgan and N. Keiding and R. D. Gill (1992), *Statistical Models Based on Counting Processes*, Springer series in Statistics, 476-591.
3. Bennett, S. (1983a). Log-Logistic Regression Models for Survival data, *Applied. statist.*, 32, No. 2, 165-171.
4. Bennett, S. (1983b). Analysis of Survival Data by the Proportional Odds Model, *Statistics in Medicine*, 2. 273-277.
5. Brown, B. M. (1971), Martingale Central Limit Theorems, *The Annals of Mathematical Statistics*, 42, No.1, 59-66.
6. Dorota M. Daborowska and Kjell A Doksum (1988). Estimation and Testing in a Two-sample Generalized odds-Rate Model, *Journal of the American Statistical Association*, 83, No. 403, 744-749.
7. Galen R. Shorack. and Jon A. Weller (1986). *Empirical Processes with Applications to Statistics*, John Willey & Sons, Inc. 1-150, 258-333.
8. Jean-Yves Dauxois and Syed N. U. A. Kirmani (2003). Testing the Proportional Odds Model under Random Censoring, *Biometrika*, 90, 4, 913-922.
9. Jian Huang and A. J. Rossini (1997). Sieve Estimation for the Proportional-Odds Failure-Time Regression Model with The Interval Censoring, *Journal of the American Statistical Association*, 92. 960-967.
10. Per Krag Andersen, Ørnulf Borgan, Niels Keiding and Richard D. Gill (1992). *Statistical Models Based on Counting Processes*, Springer Series in Statistics.
11. S. A. Murphy, A. J. Rossini and A. W Van Der Vaart (1997). Maximum Likelihood Estimation in the Proportional Odds Model, *Journal of the American Statistical Association*, 92, 968-976.
12. S. N. U. A. Kirmani and Ramesh C. Gupta (2001). On the Proportional Odds Model in Survival Analysis, *Ann. Inst. Statist. Math.* 53, 2, 203-216.
13. Song Yong and Ross L. Prentice (1999). Semiparametric Inference in the Proportional Odds Regression Model, *Journal of the American Statistical Association*, 94, 125-136.
14. Robert J. Serfling (1980). *Approximation theorems of mathematical statistics*, John Willy & Sons Inc. 1-54.

[2006년 7월 접수, 2006년 9월 채택]