Journal of Korean
Data & Information Science Society
2004, Vol. 15, No. 2, pp. 347~353

Simple Estimate of the Relative Risk under the Proportional Hazards Model

Sung Won Lee[1) ・ Ju Sung Kim[2) ・ Jung Sub Park[3)

Abstract

We propose a simple nonparametric estimator of relative risk in the two sample case of the proportional hazards model for complete data. The asymptotic distribution of this estimator is derived using a functional equation. We obtain the asymptotic normality of the proposed estimator and compare with Begun's estimator by confidence interval through simulations.

Keywords : Hazard function; Proportional hazards model; Relative risk; Survival function.

1. 서론

두 종류의 치료제 또는 치료방법에 의한 생존시간에 대한 비교분석을 하는 경우에 종종 두 위험함수(hazard function)사이에 비례성(proportionality)을 가정하게 되는데, 이는 두 위험함수 λ_1과 λ_2 사이의 관계식을 다음과 같이 표현할 수 있다.

$$\lambda_2 = \theta \lambda_1$$

여기서 θ는 양수이며 상대위험도(relative risk)라고 한다. 두 위험함수 간에 비례성이 성립할 경우에는 한 쪽의 치료제가 다른 치료제보다 전 기간에 걸쳐 우수하다는 걸 나타내거나($\theta > 1$) 또는 열등하다는 것을 의미($\theta < 1$)하기도 한다. 이때 $\theta = e^{\beta}$ 라고 표현할 때 β를 회귀계수라고 하고, β에 대한 추정절차로는 Cox(1972)가 제시한 편

---

1) First Author: Department of Computer Science, Chungbuk National University, Cheongju.
E-mail : rhisw9@hanmail.net

2) Professor, Department of Information Statistics, Chungbuk University, Cheongju.

3) Professor, Department of Industrial Information System Engineering, Cheongju University, Cheongju.

우도원리에 의한 최대편우도추정치(Maximum partial likelihood estimate)가 있으며, 이는 비선형식의 근간이 되고, Newton-Raphson방법에 의해 반복적으로 추정된다. Pike(1972)와 Yusuf 등(1985)은 두 개의 간단하고 직관적인 로그순위검정에 기초한 상대위험도(relative risk)를 제안했고, Andersen(1983)은 Cox의 최대편우도추정치의 근사분산과 Begun & Reid(1983)의 추정치의 근사분산이 거의 같음을 보였으며, Cox의 최대편우도추정치는 반복적으로 계산해야 하므로 가중된 선형순위검정(weighted linear rank tests)으로 상대위험도(relative risk)의 클래스를 계수과정식(counting process formulation)으로 추정하였다. 후에 Berry(1991)와 Chi(1994)가 모의실험을 통해 Andersen추정치들의 유한표본성질들을 연구하였다. Begun and Wellner(1982)는 자신들이 제안한 two-step추정치와 Cox의 최대부분우도추정치가 θ의 추정치들 중에 적합한 추정치임을 보였으며, Begun and Reid(1983)는 중도절단된 자료를 이용하여 비례위험모형을 따르는 두 집단의 경우에 Mantel-Haenszel 추정치를 일반화시킨 two-step 추정치를 제안하였다. Begun(1987)은 중도 절단되지 않은 자료를 이용하여 비례위험함수를 갖는 두 집단문제에서 상대위험도를 추정하였는데, Cox모형에서 회귀계수가 하나인 경우에 대하여 상대위험도의 비모수 순위추정량인 two-step추정치를 추정하여 Cox의 편우도추정치와 근사적으로 같아짐을 보였다. Chi & Tseng(2002)는 Peto, Pike와 Andersen에 의해 제안된 우측 중도 절단된 자료에 대한 상대위험도의 추정치들을 수정하여, 구간 중도 절단된 자료에서 상대위험도의 추정절차를 제안하였다. 우리는 이 논문에서 중도절단 되지 않은 자료를 이용하여 단순하게 상대위험도를 추정하는 새로운 방법을 제시하였다.

제 2장에서는 새로운 점추정치를 소개하였고, 함수식을 이용한 새로운 추정치의 일치성을 보였으며, 또한 추정치의 극한분포가 정규분포임을 보였다. 제 3장에서는 모의실험을 통하여 Begun의 추정치와 새롭게 제안된 추정치의 결과를 표로 나타내어 추정치의 신뢰구간을 비교하였다.

2. 제안된 상대위험도의 점추정치와 근사분포.

각각의 분포함수가 F_1, F_2이며, 표본의 크기가 n_1과 n_2인 독립된 두 개의 표본 $X_{11},\dots,X_{1n_1}$과 $X_{21},\dots,X_{2n_2}$가 있다고 하자. 만약 두 모집단간의 위험함수사이에 비례성이 존재한다고 가정하고, S_1과 S_2는 F_1, F_2에 대한 생존함수이며, 누적위험함수는 Λ_1과 Λ_2로 표현하자.

두 위험함수 사이에 비례성이 성립한다면, 모든 $t \in [0,\infty)$와 $\theta > 0$에 대하여 Λ_1과 Λ_2사이에 다음과 같은 식이 성립한다.

$$\Lambda_2(t) = \theta \Lambda_1(t)$$

이를 생존함수식으로 표현하면 다음과 같다.

$$S_2 = S_1^{\theta}.$$

이와 같은 관계식이 성립할 때, Cox(1972)의 비례위험모형(proportional hazards model)이라 하고, 여기서 θ를 상대위험도라 한다.

두 위험함수 간에 비례성이 성립한다면 상대위험도를 추정하기 위하여 임의의 θ 에 대하여 Lee & Kim(2003)에서 비례위험모형 검정통계량으로 제안된 식으로부터 다음 식을 고려할 수 있다.

$$\int_0^{\infty} S_1(t)\,d\Lambda_2(t) = \theta \tag{1}$$

각 i에 대하여 $F_i = 1 - S_i$이며 $d\Lambda_i = dF_i/S_i$ 이다.

따라서 식 (1)로부터 우리는 각각의 생존함수와 누적분포함수 대신에 경험적인 생존함수와 누적분포함수를 대입한 다음의 $\hat{\theta}$을 θ를 추정하는 추정치라 할 수 있다.

$$\hat{\theta} = \int_0^{\infty} \hat{S_1}(t)\,d\hat{\Lambda}_2(t) = \int_0^{\infty} \hat{S_1}(t)\frac{dN_2(t)}{Y_2(t)} \tag{2}$$

식 (2)에서 $\hat{S}_1$은 $1-\hat{F_1}$이고, $\hat{\Lambda}_2$은 $\dfrac{N_2(t)}{Y_2(t)}$이며, 임의의 $\theta > 0$이고, 각 i에서 $N_i(t) = \sum_{j=1}^{n_i} I(T_{ij} \leq t)$와 $Y_i(t) = \sum_{j=1}^{n_i} I(T_{ij} \geq t)$이다.

식 (2)의 $\hat{\theta}$을 θ의 근사적 성질을 알아보기 위해 가정이 필요하며 다음 정리를 생각하자.

정리 1. $n = n_1 + n_2$이며 각 i, $i = 1, 2$에 대하여 $n \rightarrow \infty$ 함에 따라 $n/n_i \rightarrow \gamma_i \in (1, \infty)$ 임을 가정하면, $\hat{\theta}$는 θ에 대한 일치성을 갖는 추정치이다.

증명: $\hat{S}_1$과 $\hat{\Lambda}_2$는 각각 S_1과 Λ_2에 대한 강일치성을 갖는 추정치임을 이용하면 $\hat{\theta}$가 θ에 대한 일치성을 갖는 추정치를 밝히는 것은 자명하다.

다음 정리는 제안된 추정치의 점근적정규성을 알아보는 것이다.

정리 2. $\sqrt{n}(\hat{\theta} - \theta)$의 근사분포는 위에서 제시한 가정과 함께 평균이 0, 분산(σ^2)이 다음과 같은 정규분포를 따른다.

$$\sigma^2 = \frac{\gamma_2}{(1-\theta)^2}\int_0^\infty \left(\frac{S_1(t)}{S_2(t)}\right)^2 dF_2(t) + \gamma_1\theta^2\int_0^\infty S_1^2(t)dF_1(t).$$

증명: 먼저 다음 식을 보면,

$$\begin{aligned}\sqrt{n}(\hat{\theta}-\theta) &= \sqrt{n}\left(\int_0^\infty \hat{S_1}(t)\frac{dN_2(t)}{Y_2(t)} - \int_0^\infty S_1(t)d\Lambda_2(t)\right)\\ &= \sqrt{n}\left(\int_0^\infty \hat{S_1}(t)d\hat{\Lambda}_2 - \int_0^\infty S_1(t)d\Lambda_2(t)\right)\\ &= \sqrt{n}\left(\int_0^\infty \hat{S_1}(t)d\hat{\Lambda}_2 - \int_0^\infty S_1(t)d\Lambda_2(t)\right)\\ &= \sqrt{n}\int_0^\infty \hat{S_1}(t)d\left(\hat{\Lambda}_2(t)-\Lambda_2(t)\right) + \sqrt{n}\int_0^\infty \left(\hat{S_1}(t)-S_1(t)\right)d\Lambda_2(t).\end{aligned}$$

따라서

$$\begin{aligned}&\sqrt{n}\int_0^\infty \hat{S_1}(t)d\left(\hat{\Lambda}_2(t)-\Lambda_2(t)\right)\\ &= \sqrt{\frac{n}{n_2}}\int_0^\infty \hat{S_1}(t)d\left[\sqrt{n_2}\left(\hat{\Lambda}_2(t)-\Lambda_2(t)\right)\right]\\ &= \sqrt{\frac{n}{n_2}}\int_0^\infty \hat{S_1}(t)d\left[\sqrt{n_2}\left(\hat{\Lambda}_2(t)-F_2(t)/\hat{S_2}(t)+F_2(t)/\hat{S_2}(t)-\Lambda_2(t)\right)\right]\\ &= \sqrt{\frac{n}{n_2}}\int_0^\infty \hat{S_1}(t)d\left[M_{2n}(t)\frac{S_2(t)}{\hat{S_2}(t)}\right] + \sqrt{\frac{n}{n_2}}\int_0^\infty \hat{S_1}(t)d\sqrt{n_2}\left[F_2(t)\frac{\hat{F_2}(t)-F_2(t)}{\hat{S_2}(t)S_2(t)}\right]\\ &= \sqrt{\frac{n}{n_2}}\int_0^\infty \hat{S_1}(t)d\left[M_{2n}(t)\frac{S_2(t)}{\hat{S_2}(t)}\right] + \sqrt{\frac{n}{n_2}}\int_0^\infty \hat{S_1}(t)d\left[M_{2n}(t)\frac{F_2(t)}{\hat{S_2}(t)}\right]\\ &= \sqrt{\frac{n}{n_2}}\int_0^\infty \hat{S_1}(t)d\left[M_{2n}(t)/\hat{S_2}(t)\right]\\ &= \sqrt{\frac{n}{n_2}}\int_0^\infty \frac{\hat{S_1}(t)}{\hat{S_2}(t)}dM_{2n}(t) - \sqrt{\frac{n}{n_2}}\int_0^\infty \hat{S_1}(t)M_{2n}(t)\frac{d\hat{S_2}(t)}{\hat{S_2^2}(t)}.\end{aligned}$$

위에서 M_{2n}은 마팅게일(martingale)이며 정규과정(normal process) M_2에로의 약수렴은 잘 알려져 있다. 따라서 $\sqrt{n}\int_0^\infty \hat{S_1}(t)d\left(\hat{\Lambda}_2(t)-\Lambda_2(t)\right)$는 부분적분과 Slutsky 정리를 이용하면

$$\frac{\sqrt{\gamma_2}}{1-\theta}\int_0^\infty \frac{S_1(t)}{S_2(t)}dM_2(t)$$

으로 약수렴 하는 것을 확인할 수 있다.
마찬가지로

$$\begin{aligned}&\sqrt{n}\int_0^\infty (\hat{S_1}(t)-S_1(t))d\Lambda_2(t)\\ &=\sqrt{\frac{n}{n_1}}\int_0^\infty \sqrt{n_1}(\hat{S_1}(t)-S_1(t))d\Lambda_2(t)\\ &\doteq-\sqrt{\frac{n}{n_1}}\int_0^\infty M_{1n}(t)S_1(t)d\Lambda_2(t)\end{aligned}$$

따라서

$\sqrt{n}\int_0^\infty (\hat{S_1}(t)-S_1(t))d\Lambda_2(t)$는

$$\sqrt{\frac{n}{n_1}}\theta\int_0^\infty S_1(t)dM_1(t)$$

으로 약수렴 한다. 따라서 $M_j(t)$는 마팅게일(martingale)이므로 마팅게일 중심극한정리에 의하면 결과를 얻을 수 있다.

3. 모의실험과 결과

생존함수와 위험함수가 간단한 형태를 갖고 있어 가장 많이 사용되는 지수분포(Exponential distribution)와 와이블분포(Weibull distribution)로 모의실험의 난수를 발생시키기 위한 분포로 가정하였다.

먼저 지수분포는 위험함수 $h(t)$가 상수로 주어지는데,

$$h(t)=\lambda,\quad \lambda>0$$

이다. 따라서 생존함수와 확률밀도함수는

$$\begin{aligned}S(t) &= \exp\left\{-\int_0^t h(u)du\right\} = e^{-\lambda t},\ t>0\\ f(t) &= \frac{-dS(t)}{dt} = \lambda e^{-\lambda t}, \qquad t>0\end{aligned}$$

이고, 생존시간의 t의 기대값과 분산은 각각 $E(T)=1/\lambda$, $Var(T)=1/\lambda^2$이 된다.

와이블분포($Weibull(\gamma, \lambda)$)는 지수분포의 일반화된 형태로 가장 많이 쓰이는 생존분포모형으로 위험함수와 생존함수, 확률밀도함수는 다음과 같이 주어진다.

$$
\begin{aligned}
h(t) &= \lambda\gamma t^{\gamma-1} \quad \gamma>0,\ \lambda>0,\ t>0, \\
H(t) &= \lambda t^{\gamma} \\
S(t) &= e^{-\lambda t^{\gamma}}, \\
f(t) &= \lambda\gamma t^{\gamma-1}e^{-\lambda t^{\gamma}}.
\end{aligned}
$$

와이블분포는 평균과 분산이 명확한 수식으로 주어지진 않지만, 모수의 값에 따라 여러 가지 형태의 함수로 표현할 수 있다. 분포의 모양은 γ값에 의존하므로 γ를 형태모수(shape parameter)라 부르고, λ는 척도모수(scale parameter)로서 수평축의 스케일을 변화시킨다. 일반적으로 γ 값이 1에서 3사이 값을 취하는 것이 적합하다고 알려져 있고, γ값이 클수록 위험률은 시간이 경과함에 따라 누증적으로 증가하게된다. 앞의 지수분포는 γ값이 1인 와이블분포이다.

모의실험을 통하여 자료를 각각 지수분포와 와이블분포로 난수를 발생시켜 새로이 제안된 추정량과 Begun의 추정량을 비교하였다. 먼저 비례위험모형을 따른다는 가설에 준하여, 두 모집단의 분포를 각각 지수분포 $Exp(1)$, $Exp(2)$와 $Weibull(\gamma,\lambda)$분포에서 형태모수 γ는 같게 하고 척도모수 λ를 각각 비례를 따르게 하기위하여 $Weibull(2,1)$, $Weibull(2,0.5)$에서 자료를 발생시켰다. 첫 번째 집단의 자료 n_1와 두 번째 집단의 자료 n_2를 15,15 30,30 50,50 100,100개씩 1000번 발생시키고, 다음에는 첫 번째 집단과 두 번째 집단의 자료를 1:2 1:3 및 1:4의 비율로 1000번 발생시켜 상대위험도를 추정하였다. 각각에 따라 추정된 θ의 95% 신뢰구간을 구한 결과, 제안된 추정치가 Begun의 추정치보다 모수 θ에 더 가까운 값을 보였다. 다음 표에 새롭게 제안된 추정치와 Begun의 추정치의 결과를 보면 각각의 경우에 대하여 제안된 추정치가 Begun의 추정치보다 모수 θ에 접근했음을 확인 할 수 있다. 본 연구에 이어 중도절단된 자료에 대한 추정은 향후 과제로 남기겠다.

<표 1> θ의 95% 신뢰구간

	$Exp(1)$, $Exp(2)$		$Weibull(2,1)$, $Weibull(2,0.5)$	
n_1, n_2	제안된 추정치	Begun 추정치	제안된 추정치	Begun 추정치
5,10	(0.477202,0.503818)	(0.442363,0.468157)	(0.477811,0.504389)	(0.443153,0.468947)
5,15	(0.481030,0.507490)	(0.48144, 0.496016)	(0.482092,0.508788)	(0.469317,0.497423)
5,20	(0.484487,0.511653)	(0.483524,0.513316)	(0.485821,0.513379)	(0.485250,0.515670)
15,15	(0.492401,0.502319)	(0.452382,0.461398)	(0.489172,0.499168)	(0.449702,0.458718)
30,30	(0.493610,0.500470)	(0.473257,0.479843)	(0.492820,0.499720)	(0.472967,0.479553)
50,50	(0.495104,0.500436)	(0.482273,0.487487)	(0.494394,0.499726)	(0.481893,0.487107)
100,100	(0.496238,0.500002)	(0.489788,0.493512)	(0.495378,0.499142)	(0.489278,0.493002)

참고문헌

1. Andersen, P.K.(1983) Comparing survival distribution via hazard ratio estimates. Scandinavian Journal of statistics 10,77-85
2. Begun, J.M.(1987) Estimate of relative risk. Metrika, 34: 65-87
3. Begun, J.M., and Wellner, J.A.(1982) Asymptotic efficiency of relative risk estimates. in Johnson Festschrift, ed. P.P. Sen
4. Begun, J.M., Reid, N.(1983) Estimating the relative risk with censored data. Journal Amer Statist Assoc 78:337-341
5. Berry, G., Kitchen, R.M., Mock, P.A.(1991) A comparison of two simple hazard ratio estimators based on the logrank test. Statistics in Medicine 10, 749-755
6. Berstein, L., Anderson, J., and Pike, M.C.(1981) Estimation of the proportional hazard in two treatment group clinical trials. Biometrics, 37, 513-520.
7. Chi, Yunchan.(1994) Improving partial likelihood estimator of relative risk and comparisons among other estimators of relative risk. Technical report, Dept of statistics, National Cheng-Kung University, NSC 82-0208-M 006-021
8. Chi, Yunchan., Tseng, Chea-Hui.(2002) Comparison of several relative risk estimators with interval-censored data. Biometrical Journal. 44, no 2, 197-212
9. Cox, D.R.(1975) Partial likelihood. Biometrika 62:269-276
10. Crowley, J.(1975) Estimation of relative risk in survival studies. Technical Report No 423, Department of Statistics, University of Wisconsin
11. Pike(1972) Asymptotically efficient rank invariant test procedures. Journal of the Royal statistical society, series A 135, 201-203
12. Lee, S.W., Kim, J.S.(2003) A test procedure for checking the proportionality between hazard functions, Data&Information Science Society, 14, no 3, 561-570
13. Yusuf, S., Peto, R., Lewis, J., Collins, R., Slieght, P.(1985) Beta blockade during and after myocardial infarction: an overview of the randomized clinical trials. Progress in cardiovascular diseases 27, 335-371

[2003년 12월 접수, 2004년 4월 채택]