Journal of Korean
Data & Information Science Society
2003, Vol. 14, No.3 pp. 561~570

A Test Procedure for Checking the Proportionality Between Hazard Functions

SeongWon Lee[1)] · JuSeong Kim[2)]

Abstract

We propose a nonparametric test procedure for checking the proportionality assumption between hazard functions using a functional equation. Because of the involvement of censoring distribution function, we consider the large sample case only and obtain the asymptotic normality of the proposeed test statistic. Then we discuss the rationale of the use of the functional equation, give some examples and compare the performances with Andersen's procedure by computing powers through simulations.

Keywords:Censored data; Cumulative hazard function; Initial effect model; Proportionality; Survival function

1. 서론

두 종류의 치료제 또는 치료방법에 의한 생존시간에 대한 통계적 추론을 할 때, 종종 두 위험함수(hazard function)사이에 비례성(proportionality)을 가정하게 되는데 이는 상수 상대위험도(relative risk)로 표현된다. 두 위험함수간에 비례성이 성립할때는 한쪽의 치료제가 다른 치료제보다 전 기간에 걸쳐 우수하다는 것을 나타낸다. 그러나 최초의 초기단계에서는 치료효과가 좋게 나타나지만 시간이 지남에 따라 치료효과가 사라져버리는 경우에는, 비례성의 가정하에서 통계적 추론은 대단히 심각한 결과로 오도될 것이다. 따라서 자료분석의 진단단계에서도 비례성의 가정을 살펴보는 것이 중요한 일이다.

지금까지 비례성의 가정을 검토하기 위한 몇가지 방법들이 제시되었다. 비례위험모형의 제안자인 Cox(1872)는 시간 함수 형태인 공변량을 모형에 포함시켜 비례위험모형에 대한 가정을 검정하였고, Schoenfeld(1980)는 시간과 공변량을 여러개의 구간으

1) Department of Computer Science, Chungbuk National University, Cheongju.
e-mail:rhisw9@daum.net

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

로 분할(partition)한 후, 카이제곱적합도검정법의 개념을 적용하여 비례위험모형의 가정을 검정하였다. 하지만 분할을 임의로하기때문에 검정결과에 영향을 줄 수도 있다. Lee 와 Pirie(1981)는 두 개의 누적위험함수들을 추정한 점들에 의해 위험률의 모형을 그래프를 이용하여 검토 할 수 있는, 소위 경향함수의 성질들을 이용한 그래픽 방법을 고려하였다. 이 방법은 한 눈에 비례위험모형의 가정에 대한 적합성 정도를 알아 볼 수 있지만, 위배여부를 결정할 때 주관적 판단이 개입될 소지가 있다. Andersen (1982)은 새로운 공변량을 포함하였을때에도 비례성의 가정이 지켜지고 있는가를 두 가지방법인, 적합도 검정과 그래픽 방법을 몇몇 응용사례와 함께 논의했는 데, 기저위험함수(baseline hazard funtion)를 상수로 가정하였으며, Wei(1984)는 상대위험도의 개념을 이용, 비정규분포를 이용한 검정절차로 이끌어내는 편우도원리를 이용한 Kolmogorov식 통계량에 기초한 새로운 검정을 제안했다. 또한 Gill과 Schmacher (1987)은 상대위험도에 대한 서로 다른 두 개의 일반화 추정량을 구하여 비교함으로써, 비례성의 가정을 위배하는 경우에 대하여 검정통계량을 제시하였다. 하지만 검정을 하기 위해서는 적합한 가중함수를 선택해야 한다. Grambsch and Fleming(1990)은 반복알고리즘이 요구되는 Wei의 절차를 일반화시켜서, 잔차들을 기초로 한 마팅게일을 이용한 방법을 제안하였다. 우리는 이 논문에서 새로운 방법을 제시하여 이런 난점들을 피하는 절차를 고려하였다.

제 2장에서는 우리는 검정통계량을 제시하고, 제시된 검정통계량의 극한분포가 정규분포임을 보이고 실제적 문제에 이용하기 위하여 일치성을 갖는 분산의 추정치를 제시한다. 제 3장에서는 초기효과모형에 대하여 우리가 제시한 검정통계량에 대한 이론적인 판단근거를 제시하고, 실례로 우리가 제시한 검정통계량에 대한 검정절차를 예시하며, 시뮬레이션을 통한 경험적검정력(empirical powers)을 구하여 Andersen 절차와 우리의 검정절차를 비교했다.

2. 검정통계량의 설명과 대표본 근사

표본의 크기가 n_1과 n_2인 두 개의 표본을 고려해보자. 여기서 $n=n_1+n_2$이다.

생존시간에 대한 두 독립표본 $X_{11},\ldots,X_{1n_1}$ 과 $X_{21},\ldots,X_{2n_2}$는 각각 연속생존함수 S_1과 S_2를 가지고, 누적위험함수 Λ_1과 Λ_2를 갖는다고 하자. 또 중도절단된 두 독립표본 $C_{11},\ldots,C_{1n_1}$ 과 $C_{21},\ldots,C_{2n_2}$의 임의의 분포함수는 G_1과 G_2라 하자. 중도절단된 자료를 고려하기위해 $T_{ij}=\min\{X_{ij},C_{ij}\}$와 $\delta_{ij}=I(X_{ij}\le C_{ij})$를 구하였다. 각 $i=1,2$, $j=1,2,\ldots,n_i$이며, 여기서 $I(\cdot)$는 지시함수이다. X_{ij} 와 C_{ij}는 각 i,j에 대해 통계적으로 독립이라고 가정하자.

위 자료에 의하여 다음을 검정하고자 한다.

모든 $t\in[0,\infty)$ 이고 $\theta>0$에 대하여,

$$H_0:\ \Lambda_2(t)=\theta\Lambda_1(t).$$

검정통계량을 구축하기 위해선 아래 두 식을 고려한다.

H_0가 참일 때, 임의의 $\theta>0$ 에 대하여 다음 식을 구할 수 있다.

$$Q_1 = \int_0^{\infty} S_2(t)\, d\Lambda_1(t) = \theta^{-1} \qquad Q_2 = \int_0^{\infty} S_1(t)\, d\Lambda_2(t) = \theta \tag{1}$$

각 i에 대해서 $F_i = 1 - S_i$이며 $d\Lambda_i = dF_i / S_i$ 이다.
따라서 H_0하에서 우리는 다음을 얻을 수 있다.

$$Q = Q_1 Q_2 = \int_0^{\infty} S_2(t)\, d\Lambda_1(t) \int_0^{\infty} S_1(t)\, d\Lambda_2(t) = 1 \tag{2}$$

중도절단되지 않은 자료의 경우는 생존함수들과 누적위험함수들에 대한 경험적 검정통계량의 대입으로 검정통계량을 제시 할수 있다. 그러나 최대관측치가 중도절단된 자료의 경우, 우린 완전한 표본생존함수들을 얻을 수 없다. 하지만 임의의 $\theta > 0$고 $T_i \in (0, \infty)$에 대하여 귀무가설하에서 다음 식이 가능하고,

$$\int_0^{T_1} \frac{S_2(t)}{1 - S_2(T_1)}\, d\Lambda_1(t) = \theta^{-1} \qquad \int_0^{T_2} \frac{S_1(t)}{1 - S_1(T_2)}\, d\Lambda_2(t) = \theta \tag{3}$$

따라서,

$$\int_0^{T_1} \frac{S_2(t)}{1 - S_2(T_1)}\, d\Lambda_1(t) \int_0^{T_2} \frac{S_1(t)}{1 - S_1(T_2)}\, d\Lambda_2(t) = 1 \tag{4}$$

이다.

각 i에 대해서 $\widehat{S}_i$ 를 S_i의 Kaplan-Meier 추정치라 하고, $\widehat{\Lambda}_i$은 Λ_i의 Nelson-Aalen추정치라 하자. T_i를 i번째 표본 중 가장 큰 관측치라고 한다면,

$$\begin{aligned} &\int_0^{T_1} \frac{\widehat{S}_2(t)}{1 - \widehat{S}_2(T_1)}\, d\widehat{\Lambda}_1(t) \int_0^{T_2} \frac{\widehat{S}_1(t)}{1 - \widehat{S}_1(T_2)}\, d\widehat{\Lambda}_2(t) \\ &= \int_0^{T_1} \frac{\widehat{S}_2(t)}{1 - \widehat{S}_2(T_1)} \frac{dN_1(t)}{Y_1(t)} \int_0^{T_2} \frac{\widehat{S}_1(t)}{1 - \widehat{S}_1(T_2)} \frac{dN_2(t)}{Y_2(t)} \end{aligned}$$

은 각 i에서 $N_i(t) = \sum_{j=1}^{n_i} I(T_{ij} \le t,\ \delta_{ij} = 1)$와 $Y_i = \sum_{j=1}^{n_i} I(T_{ij} \ge t)$인 H_0 하에서 1에 가까워질것이다. 그러므로 우리는,

$$Z_n = \sqrt{n}\left\{ \int_0^{T_1} \frac{\widehat{S}_2(t)}{1 - \widehat{S}_2(T_1)} \frac{dN_1(t)}{Y_1(t)} \int_0^{T_2} \frac{\widehat{S}_1(t)}{1 - \widehat{S}_1(T_2)} \frac{dN_2(t)}{Y_2(t)} - 1 \right\} \tag{5}$$

을 H_0를 검정하기 위한 검정통계량으로 제시한다. Z_n에 표기된 모든 추정치들이 일치성을 가지며, H_0하에서 $|Z_n|$의 값이 작아질 것이다. 따라서 검정기준은 $|Z_n|$의 큰 값들에 대하여 H_0를 기각할 수 있다. Cheng(1985) 또한 중도절단되지 않은 자료인 경우 위험함수들의 동일성을 검정하기 위한 통계량을 제시하기 위해 이런 종류의 함수식을 사용했다. 나중에 식(2)와 (4)의 사용에 대해 논의 할 것이다. H_0하에서조차도 Z_n의 분포는 중도절단분포를 포함함으로써 귀무가설하에 Z_n의 정확한 귀무분포를 얻기는 불가능하므로, 대표본근사이론으로 근사분포를 알아보는 것이 타당하다. 우선, 점근적 정규성을 알아보기 위해 몇개의 기호를 소개한다.

각 i에 대해 $1 - H_i = (1 - F_i)(1 - G_i)$라 하자. 또한 각 i에 대해 M_i를 공분산 함수

$\int_0^{\cdot}(1-H_{i-})(1-\Delta\Lambda_i)d\Lambda_i$를 갖는 가우스과정이라 하자. 여기서 함수 H_{i-}는 우방 극한값(right hand limit)을 갖는 H_i의 좌방 연속(left continuous version)이고, $\Delta\Lambda_i(t)=\Lambda_i(t)-\Lambda_i(t_-)$이다. M_i는 마팅게일이다. 아래의 $\Rightarrow$, $\rightarrow^p$, $\rightarrow^d$는 약수렴(weak convergence), 확률수렴(convergence in probability), 분포수렴(convergence in distribution)임을 의미한다. 다음 Lemma는 잘 알려진 결과이다.(cf. Shorack and Wellner 1986).

Lemma 1. 각 i와 $t\in[0,\ T_i]$에 대하여,

$$\sqrt{n_i}\left(\frac{N_i(t)}{Y_i(t)}-\Lambda_i(t)\right)\Rightarrow\int_0^t\frac{dM_i(s)}{1-H_{i-}(s)}.$$

다음은 Gill과 Schmacher(1987)에 의한 델타방법(delta-method)으로서 증명없이 다음 결과를 제시한다.

Lemma 2. Y_n 과 $\overline{Y}_n$는 p차원(R^p)의 확률열벡터라 하고 μ는 p차원의 상수벡터라 하자. 함수 h: $R^p\rightarrow R$는 μ부근에서 미분 가능하다고 하고, 도함수 h'은 μ에서 연속이라 하자. $n\rightarrow\infty$ 이며 $a_n\rightarrow\infty$인 경우에 다음을

$$\sqrt{a_n}(Y_n-\overline{Y}_n)\rightarrow^d Y$$

$$\overline{Y}_n\rightarrow^p \mu$$

만족하는 Y에 대하여,

$$\sqrt{a_n}(h(Y_n)-h(\overline{Y}_n))\rightarrow^d h'(\mu)^T Y$$

이다.

다음 Theorem에서 Z_n의 점근적 정규성의 결과를 얻는다.

Theorem. $n_2/n_1\rightarrow\rho$라고 가정하자. 그러면 H_0하에서,

$$Z_n\rightarrow^d Z.$$

여기서 Z는 평균 0과 분산 σ^2을 갖는 정규확률변수이며, 분산 σ^2은

$$\sigma^2=(1+\rho)\theta^2\int_0^{T_1}\left[\frac{S_2(t)}{1-S_2(T_1)}\right]^2\frac{(1-\Delta\Lambda_1(t))d\Lambda_1(t)}{1-H_{1-}(t)}+\frac{1+\rho^{-1}}{\theta^2}\int_0^{T_2}\left[\frac{S_1(t)}{1-S_1(T_2)}\right]^2\frac{(1-\Delta\Lambda_2(t))d\Lambda_2(t)}{1-H_{2-}(t)}.$$

증명: Lemma 2에서 $a_n=n$과 $h(x,y)=xy$ 라 하고,

$$Y_n=\left(\int_0^{T_1}\frac{\widehat{S}_2(t)}{1-\widehat{S}_2(T_1)}\frac{dN_1(t)}{Y_1(t)},\ \int_0^{T_2}\frac{\widehat{S}_1(t)}{1-\widehat{S}_1(T_2)}\frac{dN_2(t)}{Y_2(t)}\right)$$

$$\overline{Y}_n=\left(\int_0^{T_1}\frac{\widehat{S}_2(t)}{1-\widehat{S}_2(T_1)}d\Lambda_1(t),\ \int_0^{T_2}\frac{\widehat{S}_1(t)}{1-\widehat{S}_1(T_2)}d\Lambda_2(t)\right)$$

라 하자.

그러면 Slustky theorem 과 Lemma 1로부터 다음을 얻을수 있다.

$$\sqrt{n}\left\{\int_0^{T_1}\frac{\widehat{S}_2(t)}{1-\widehat{S}_2(T_1)}\frac{dN_1(t)}{Y_1(t)}-\int_0^{T_1}\frac{\widehat{S}_2(t)}{1-\widehat{S}_2(T_1)}d\Lambda_1(t)\right\}$$
$$\Rightarrow\sqrt{1+\rho}\int_0^{T_1}\frac{S_2(t)}{1-S_2(T_1)}\frac{dM_1(t)}{1-H_{1-}(t)}$$

$$\sqrt{n}\left\{\int_0^{T_2}\frac{\widehat{S}_1(t)}{1-\widehat{S}_1(T_2)}\frac{dN_2(t)}{Y_2(t)}-\int_0^{T_2}\frac{\widehat{S}_1(t)}{1-\widehat{S}_1(T_2)}d\Lambda_2(t)\right\}$$
$$\Rightarrow\sqrt{1+\rho^{-1}}\int_0^{T_2}\frac{S_1(t)}{1-S_1(T_2)}\frac{dM_2(t)}{1-H_{2-}(t)}.$$

따라서

$$\overline{Y}_n=\left(\int_0^{T_1}\frac{\widehat{S}_2(t)}{1-\widehat{S}_2(T_1)}d\Lambda_1(t),\ \int_0^{T_2}\frac{\widehat{S}_1(t)}{1-\widehat{S}_1(T_2)}d\Lambda_2(t)\right)$$

는 (θ^{-1},θ)로 확률수렴하므로, Lemma 2로 부터 Z_n이

$$Z=\sqrt{1+\rho}\,\theta\int_0^{T_1}\frac{S_2(t)}{1-S_2(T_1)}\frac{dM_1(t)}{1-H_{1-}(t)}+\frac{\sqrt{1+\rho^{-1}}}{\theta}\int_0^{T_2}\frac{S_1(t)}{1-S_1(T_2)}\frac{dM_2(t)}{1-H_{2-}(t)}$$

로 분포수렴(converges in distribution)한다는 결론을 얻을 수 있다.

실제문제들에서 제안된 검정절차를 응용하기위해서 분산 σ^2의 일치추정량을 구해야한다. 다음은 분산 σ^2의 하나의 일치추정량으로 제시할수 있다.

$$\begin{aligned}\widehat{\sigma^2}&=n\left[\int_0^{T_2}\frac{\widehat{S}_1(t)}{1-\widehat{S}_1(T_2)}d\widehat{\Lambda}_2(t)\right]^2\int_0^{T_1}\left[\frac{\widehat{S}_2(t)}{1-\widehat{S}_2(T_1)}\right]^2\left(1-\frac{\Delta N_1(t)-1}{Y_1(t)-1}\right)\frac{dN_1(t)}{Y_1(t)}\\&\quad+n\left[\int_0^{T_1}\frac{\widehat{S}_2(t)}{1-\widehat{S}_2(T_1)}d\widehat{\Lambda}_1(t)\right]^2\int_0^{T_2}\left[\frac{\widehat{S}_1(t)}{1-\widehat{S}_1(T_2)}\right]^2\left(1-\frac{\Delta N_2(t)-1}{Y_2(t)-1}\right)\frac{dN_2(t)}{Y_2(t)}\\&=n\left[\int_0^{T_2}\frac{\widehat{S}_1(t)}{1-\widehat{S}_1(T_2)}\frac{dN_2(t)}{Y_2(t)}\right]^2\int_0^{T_1}\left[\frac{\widehat{S}_2(t)}{1-\widehat{S}_2(T_1)}\right]^2\left(1-\frac{\Delta N_1(t)-1}{Y_1(t)-1}\right)\frac{dN_1(t)}{Y_1(t)}\\&\quad+n\left[\int_0^{T_1}\frac{\widehat{S}_2(t)}{1-\widehat{S}_2(T_1)}\frac{dN_1(t)}{Y_1(t)}\right]^2\int_0^{T_2}\left[\frac{\widehat{S}_1(t)}{1-\widehat{S}_1(T_2)}\right]^2\left(1-\frac{\Delta N_2(t)-1}{Y_2(t)-1}\right)\frac{dN_2(t)}{Y_2(t)}\end{aligned}$$

일치성에 대한 증명은 Gill(1980)의 결과를 이용하면 쉽게 보일수 있다.

3. 토의와 예제

우선, 모든 초기효과모형들을 포함하고 있는 H_1하에서 Q값의 변화에 대한 연구를 통하여, 우리의 검정절차가 초기효과의 여부를 알아보는데 사용될 수 있는지 알아보겠다.

먼저 다음을 가정하자.

$$\Lambda_2(t) = \begin{cases} \theta_1\Lambda_1(t) \ , & 0 \le t < t_p \\ \theta_2\Lambda_1(t) \ , & t \ge t_p \end{cases} \tag{6}$$

임의의 $p \in (0,1)$에 대하여 $t_p \in [0, \infty)$는 $F_1(t_p) = p$가 되는 점이며 $\theta_1 > 0$과 $\theta_2 > 0$는 $\theta_1 \ne \theta_2$를 만족한다.

식 (6)에서 $\theta_1 = \theta_2$이거나, $p=0$또는 $p=1$일 때 비례위험모형임을 주목하자. 또한 (6)에 대한 생존함수들은 다음과 같다.

$$S_2(t) = \begin{cases} S_1^{\theta_1}(t), & 0 \le t < t_p \\ S_1^{\theta_1-\theta_2}(t_p) S_1^{\theta_2}(t), & t \ge t_p \end{cases} \tag{7}$$

식 (6)과 (7)로부터 다음을 구할 수 있다.

$$Q_1(p, \theta_1, \theta_2) = \int_0^\infty S_2(t)\,d\Lambda_1(t) = \frac{1}{\theta_1}(1-(1-p)^{\theta_1}) + \frac{1}{\theta_2}(1-p)^{\theta_2}$$

$$Q_2(p, \theta_1, \theta_2) = \int_0^\infty S_1(t)\,d\Lambda_2(t) = \theta_1 p + \theta_2(1-p).$$

$q=1-p$라고 표기하자. 그러면

$$D(p, \theta_1, \theta_2) = Q_1(p, \theta_1, \theta_2)Q_2(p, \theta_1, \theta_2) - 1 = q(1-q)^{\theta_1}\left(\frac{\theta_2}{\theta_1} - 1\right) + pq^{\theta_1}\left(\frac{\theta_1}{\theta_2} - 1\right).$$

만약 $\theta_1 = \theta_2$이면 $D(p, \theta_1, \theta_2) = 0$이 되고, $\theta_1,\ \theta_2$의 임의의 쌍에 대해서도 $D(0, \theta_1, \theta_2) = D(1, \theta_1, \theta_2) = 0$을 주목하자. 이것은 두 누적위험함수들 Λ_1과 Λ_2가 그런 경우에 비례성을 따르기 때문이다. 따라서 우리의 목표는 아래 4가지 경우에 대해, $D(p, \theta_1, \theta_2)$가 초기효과모형(6) 아래 임의의 $p \in (0,\ 1)$와 $\theta_1 \ne \theta_2$에서 결코 0을 얻을 수 없음을 보이는 것이다.

(i) $1 \le \theta_1 < \theta_2$ (ii) $0 \le \theta_2 < \theta_1 \le 1$ (iii) $1 \le \theta_2 < \theta_1$ (iv) $0 < \theta_1 < \theta_2 \le 1$

(i)과 (ii)에선 $D(p, \theta_1, \theta_2) > 0$을, 그리고 (iii)과 (iv)에선 $D(p, \theta_1, \theta_2) < 0$ 을 보일 것이다. (i)의 경우, $q^{\theta_1} \le q$, $\theta_2/\theta_1 > 1$과 $0 < \theta_1/\theta_2 < 1$을 주목하면 다음을 알 수 있다.

$$D(p,\theta_1,\theta_2)=q(1-q^{\theta_1})\left(\frac{\theta_2}{\theta_1}-1\right)+pq^{\theta_1}\left(\frac{\theta_1}{\theta_2}-1\right)$$
$$\geq pq\left(\frac{\theta_2}{\theta_1}-1\right)+pq\left(\frac{\theta_1}{\theta_2}-1\right)$$
$$=\frac{pq}{\theta_1\theta_2}(\theta_1-\theta_2)^2>0\,.$$

또한 (i)에서 쓰여진 같은 방법으로 (ii)에서도 $q^{\theta_1}\geq q,\ 0<\theta_2/\theta_1<1,\ \theta_1/\theta_2>1$라는 사실들로 다음을 쉽게 보일 수 있다.

$$D(p,\theta_1,\theta_2)=\frac{pq}{\theta_1\theta_2}(\theta_1-\theta_2)^2>0.$$

(iii)의 경우엔 $D(p,\theta_1,\theta_2)<0$을 보이기 위해 다음 Lemma를 이용한다.

Lemma 3. 임의의 고정된 y >1 에 대하여,

$$g(x)=1-yx^{\,y-1}+(y-1)x^y$$

라 하자. 그러면 모든 $x\in(0,1)$에 대하여 $g(x)>0$이다.

증명: $g(0)=1$과 $g(1)=0$이라 하면, g는 (0, 1)에서 단조(monotone)임을 보여주면 된다.

$$\frac{dg(x)}{dx}=y(y-1)x^{\,y-2}(x-1)<0 \qquad \text{모든 } x\in(0,1),$$

이므로 Lemma의 결과는 자명하다.

이제 $D(p,\theta_1,\theta_2)$를 다시 써 보면,

$$D(p,\theta_1,\theta_2)=(\theta_2-\theta_1)\left(\frac{q}{\theta_1}(1-q^{\theta_1})-\frac{p}{\theta_2}q^{\theta_1}\right)$$

이다.

(iii)에서는 $\theta_2-\theta_1<0$ 일 때, $\frac{q}{\theta_1}(1-q^{\theta_1})-\frac{p}{\theta_2}q^{\theta_1}>0$을 확인 할 수 있으며, Lemma 3를 이용하여 다음 결과를 얻었다.

$$\frac{q}{\theta_1}(1-q^{\theta_1})-\frac{p}{\theta_2}q^{\theta_1}\geq\frac{q}{\theta_1}(1-q^{\theta_1})-pq^{\theta_1}=\frac{q}{\theta_1}(1-\theta_1q^{\,\theta_1-1}+(\theta_1-1)q^{\theta_1}).$$

마지막으로, (iii)과 같은 방법으로 (iv)도 확인 할 수 있다.

$$D(p,\theta_1,\theta_2)<0\,.$$

모든 초기효과모형들의 클래스를 포함하는 H_1하에서 $D(p,\theta_1,\theta_2)\neq 0$ 임을 보였다. 또한, H_1하에서 확률적으로 $Z_n/\sqrt{n}\rightarrow D(p,\theta_1,\theta_2)$라면, H_1하에서확률적으로 $|Z_n|\rightarrow\infty$라고 할수있다. 이것은 검정통계량의 수열 $\{Z_n\}$이 H_1하에서 일치성을 갖는다는 의미이다.

하지만, 위험함수들이 서로간에 교차되는 경우엔 이 검정방법을 적용할수 없다. 예

를들어, $\theta_1=1/2$, $\theta_2=2$라 하면 $Q(8/9, 1/2, 2)=1$이 된다. 이런 경우엔 위험함수들이 비례가 아니더라도 위험함수들이 서로 교차하는 경우에대해 $Q(p, \theta_1, \theta_2)=1$이됨을 의미한다. 자궁암자료를 이용하여 제시된 절차를 예시하도록 하자(Prentice 1978). 그림 1은 자료의 두 경험적 누적 위험 함수들을 점도표로 찍은 것이다(점선은 1그룹, 실선은 그룹2). 수직축은 누적위험값이고, 수평축은 시간을 나타낸다. 이 예제는 초기효과 모형에 따른다고 생각할수 있다. $|Z_n|=1.739$, $\widehat{\sigma^2}=0.729$이고 p-값은 0.05보다 작다.

또한 Gill과 Schumacher(1986)에서 난소암 자료(Fleming. 1980)에 대한 경험적 누적 위험함수들의 그래프는 위험함수들 사이 non-proportionality의 강한 증거를 보여준다. 그러나 p-값은 0.25가량이며, 이것은 중도절단된 자료가 너무 많은 경우이기 때문에 분산이 과대추정된 것으로 보인다. 우리의 제안된 절차는 중도절단된 자료가 많은 경우에 대하여 민감하지 않을 것이다.

아래표들은 $\theta=\theta_1/\theta_2$인 지수분포에 기초한 초기효과모형(6)에 대해 p 와 θ의 다양한 값들에 의한 Andersen의 검정력과 우리가 제시한 검정절차의 검정력에 관한 모의실험한 결과들이다. 구간 $[0, \infty)$은 Andersen검정 경우에 각 p로 $[0, t_p)$와 $[t_p, \infty)$으로 나누었다. 그 결과들은 표본 크기를 15와 20으로 1000번씩 모의실험을 기초로 했다. 표 1은 우리가 제시한 검정에 대한 경험적 검정력을 요약했고, 표2는 Andersen의 검정에 대한 경험적 검정력을 보여준다. 자료는 중도절단 되지 않은 경우를 고려했다. 표에서 p =1일 경우 비례위험가정을 따르는 것을 알수있다. 표에서 보는것처럼 , 치료기간이 짧으면 우리가 제시한 검정이 더 효과적인 것 같으며, Andersen의 검정은 치료기간이 길 때 더 효과적인 것 같이 보인다.

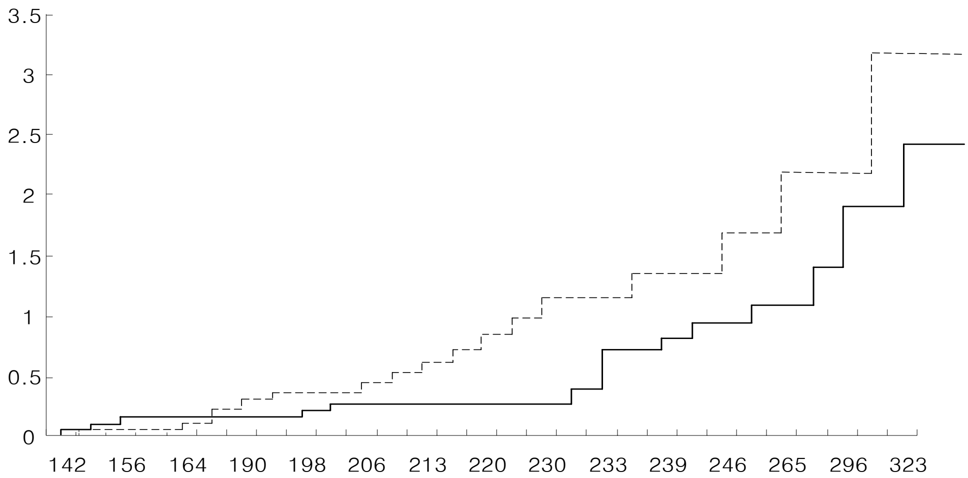

Figure 1. Plots of empirical cumulative hazard functions

θ \ p	1/8	2/8	3/8	4/8	6/8	1
1.5	0.705	0.699	0.479	0.398	0.149	0.054
2.0	0.775	0.627	0.503	0.374	0.159	0.046
3.0	0.770	0.610	0.552	0.356	0.152	0.052

table 1. 제안된 검정을 한 경험적 검정력

θ \ p	1/8	2/8	3/8	4/8	6/8	1
1.5	0.253	0.352	0.555	0.662	0.563	0.056
2.0	0.261	0.383	0.606	0.722	0.590	0.050
3.0	0.289	0.336	0.657	0.746	0.587	0.063

table 2. Andersen의 검정을 한 경험적 검정력

참고문헌

1. Aalen, O.(1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics* 6, 701－726
2. Andersen, P. K. (1982). Testing goodness of fit of Cox's regression model. *Biometrics* 38, 67－77.
3. Begun, J. M. and Reid, N(1983). Estimating the relative risk with censored data.
Journal of the American Statistical Association 78, 339－341
4. Cheng, K. F.(1985). Tests for the equality of failure rates.
Biometrika 72, 211－215
5. Fleming, T. R., O'Fallon, J. R., O'Brien, P. C. and Harrington, D. P.(1980). Modified Kolmogrorov-Smirnov test procedures with application to arbitrarily censored data. *Biometrics* 36, 607－625
6. Gill, R. D.(1980). *Censoring and stochastic integrals* (Mathematicals centre, Amsterdam, Mathematical centre tracts 124).
7. Gill, R. D. and Schmacher, M.(1987). A simple test of the proportional hazards assumption. *Biometrika* 74, 289－300.
8. Lee, L. and Pirie, W. R. (1981).A graphical method for comparing trends in series of events. *Communications in Statistics-Theory and Methods A* 10,827-848
9. Prentice, R. (1978). Linear rank tests with right censored data. *Biometrika* 63, 291-298
10. Schoenfeld, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika* 67 145-153.

11. Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics* (Wiley, New York).
12. Therneau, T. M., Grambsch, P. M. and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* 77, 147-160.
13. Wei, L. J.(1984). Testing goodness of fit for the proportional hazards model with censored observations. *Journal of the American Statistical Association* 79, 649-652
14. 장애방,이재원(1997) 비례위험모형의 적합도검정법에 관한 연구. 응용통계연구 제10권 1호. 85-104.

[2003년 3월 접수, 2003년 8월 채택]