

## 비례위험모형의 적합도 검정법에 관한 연구1)

장 애 방<sup>2)</sup>, 이 재 원<sup>3)</sup>

### 요 약

Cox(1972)가 제안한 비례위험모형은 두 표본의 처리를 비교하거나 공변량의 효과와 생존시간의 관계를 회귀적으로 해석하는 등 다양한 상황에 쓰일 수 있어 널리 이용되고 있다. 따라서 비례위험모형에 대하여 많은 통계 학자들이 연구를 하였는데, 그 중에서도 적합도 검정법에 대하여 여러 편의 논문이 발표되었다. 본 논문에서는 지금까지 제안된 비례위험모형에 대한 적합도 검정법에 관하여 설명하고, 다양한 형태의 자료에 대한 모의실험을 통하여 비례위험모형을 이용하여 생존분석을 실시하려는 통계 분석가들에게 도움이 되도록 각각의 특성에 관하여 논의하였다.

### 1. 서론

통계 분석자들은 일반적으로 모수적인 방법이나 비모수적인 방법을 이용해 자료를 분석하는데 생존자료(survival data)를 분석할 때는 한 가지 방법이 더 있다. 1972년 David Cox가 제안한 비례위험모형(proportional hazards model)이 바로 그것이다. 비례위험모형은 생존시간의 분포에 대한 아무런 가정이 없고 또한 공변량을 모형에 포함시켜, 생존시간에 대한 회귀적인 해석을 할 수 있을 뿐만 아니라 여러 처리간의 효과를 비교할 때에도 적용할 수 있다. 비례위험모형에 대한 적합도 검정법으로는 다음과 같은 방법이 있다.

우선 그림을 이용하여 그 적합성을 검토하는 방법이 있다. 이 방법은 한눈에 비례위험모형의 가정에 대한 적합성 정도를 알아볼 수 있는 장점이 있으나, 위배여부를 결정할 때 주관적 판단이 개입될 소지가 있다. 다음으로 검정통계량을 이용해 비례위험모형의 적합도 검정을 하는 방법이 있다. 비례위험모형의 제안자인 Cox(1972)는 시간의 함수 형태인 공변량을 모형에 포함하여 비례위험모형에 대한 가정을 검정하였고, Schoenfeld(1980)는 시간과 공변량을 여러 개의 구간으로 나눈 후 카이제곱적합도 검정법의 개념을 적용하여 비례위험모형의 가정을 검정하였다. Andersen(1982)은 새로운 공변량을 포함하였을 때에도 비례성의 가정이 지켜지고 있는가를 검정하였는데, 그는 기저위험함수(baseline hazard function)를 상수로 가정하였다. Wei(1984)는 상대위험도(relative risk)의 개념을 이용하여 사망자수의 관측값과 기대값간의 차이를 구하여 적합도 검정을 하는 방법을 제안하였으며, Gill and Schumacher(1987)는 상대위험도에 대한 서로

1) 이 연구는 1994년도 한국과학재단 연구비 지원에 의한 결과임 (과제번호:94-1300-01-01-3)

2) (100-095) 서울시 중구 남대문로 5가 500번지, 제일계당빌딩 6층 경영정보팀

3) (136-701) 서울시 성북구 안암동 5-1, 고려대학교 통계학과 부교수

다른 두 개의 일반화 추정량을 구하여 비교함으로써, 비례성의 가정을 단조증가 또는 단조감소하게 위배하는 경우에 대하여 검정통계량을 제시하였다. Lin and Wei(1991)는 White(1982)가 적합도 검정법에 Fisher의 정보행렬(information matrix)을 이용한 것에 착안한 적합도 검정법을 제시하였다. 우선 비례위험모형의 가정 하에서 근사적 공분산행렬의 역행렬에 관해 두 개의 추정량을 구한다. 이 때 얻어진 두 개의 추정량은 가정한 모형에 근거하여 얻은 것으로 가정이 적합하다면 그 차이가 크지 않으므로 Lin-Wei통계량은 두 개의 추정량의 차이를 이용하여 계산된 것이다.

본 논문에서는 중도절단율과 자료형태에 변화를 주어서 위에서 제시된 6가지 적합도 검정법의 유의수준과 검정력을 비교하였다. 이러한 모의실험 결과는 비례위험모형을 이용해 자료를 분석하려는 통계 분석가들에게 유용한 정보를 제시할 수 있으리라 기대된다.

## 2. 비례위험모형에 대한 적합도 검정법

생존시간(survival time)  $T$ 에 대한 위험함수(hazard function)가

$$\lambda(t) = \lim_{\Delta t \rightarrow 0^+} \frac{\Pr\{t < T < t + \Delta t \mid T > t\}}{\Delta t}$$

으로 정의될 때, 생존함수(survival function)는

$$S(t) = \Pr(T > t) = \exp\left(-\int_0^t \lambda(u) du\right) = \exp(-\Lambda(t))$$

과 같다. 여기서  $\Lambda(t)$ 는 누적위험함수(cumulative hazard function)이다.

Cox(1972)가 제안한 비례위험모형(proportional hazards model)은

$$\lambda(t; \mathbf{z}) = \exp(\boldsymbol{\beta}' \mathbf{z}) \lambda_0(t) \quad (2.1)$$

으로 표현된다. 여기서 공변량  $\mathbf{z} = (z_1, \dots, z_p)'$ 는 각각의 관측치에서 얻은 공변량으로 시간에 따라 변하는 공변량(time-dependent covariate)이 포함될 수도 있고, 기저위험함수  $\lambda_0(t)$ 는  $\mathbf{z} = \mathbf{0}$  일 때의 위험함수로 그 형태에 제약이 없다.  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ 는 모수 벡터로, Cox(1972, 1975)는 편우도(partial likelihood)를 이용해서  $\boldsymbol{\beta}$ 를 추정하는 이론을 소개하였다.  $\lambda(t; \mathbf{z})$ 은  $\mathbf{Z} = \mathbf{z}$  일 때의 위험함수로 기저함수  $\lambda_0(t)$ 의  $\exp(\boldsymbol{\beta}' \mathbf{z})$ 배로 표현된다.

### 2.1 Cox 검정법

Cox(1972)는 시간의 함수형태인 공변량을 모형에 추가하여 비례성의 가정을 검정하는 방법을 제안하였다. 이표본을 예로 들면, 공변량  $z_1$ 이 표본에 따라 0 또는 1의 값을 가지는 지시함수일 때  $z_2 = z_1 t$  인 공변량을 비례위험모형에 포함시킨

$$\lambda(t, \mathbf{z}) = \exp((\beta_1 + \beta_2 t)z_1)\lambda_0(t)$$

으로 분석을 실시한 후에  $z_2$ 의 계수인  $\beta_2$ 에 관한 귀무가설  $H_0: \beta_2=0$ 에 대한 검정을 실시한다. 귀무가설을 기각한다면 위험함수가 시간  $t$ 에 따라 단조변화한다는 결론을 얻게 되어 비례위험모형은 적합하지 않다고 말할 수 있다.

## 2.2 Schoenfeld 검정법

Schoenfeld(1980)는 적합도 검정에 널리 쓰이는 카이제곱검정법의 개념을 이용한 방법을 제안하였다. 생존시간이 관측되면 사망하였다고 정의하기로 하고, 검정통계량을 계산하기 위해 공변량의 범위를 서로 겹치지 않게  $L$ 개로 분할하고,  $[0, \infty)$ 의 시간도  $R$ 개의 구간으로 나눈다. 이렇게 얻은  $L \times R$ 개 각각의 카티션 곱(Cartesian products)에 대해 사망자수  $d_{lr}$ 를 관측하고 사망자수의 기대값  $e_{lr}$ 를 계산해낸다 ( $l=1, \dots, L$ ;  $r=1, \dots, R$ ). 사망자수에 대한 평균벡터와 분산-공분산 행렬은 Cox(1972)의 조건부평균과 분산-공분산 행렬로 대신하게 된다. 또한 Schoenfeld(1980)는 Cox(1975)의 편우도와 유사한 방법을 이용해 그의 검정통계량이 근사적으로 카이제곱분포를 따른다는 것을 보였다. 모수벡터  $\beta$ 의 값을 알고 있다는 가정 하에서 사망자수의 기대값  $e_{lr}$ 를

$$e_{lr} = \sum_{k \in D_r} \frac{\sum_{i \in R(t_k)} I_i(\mathbf{z}_i) \exp(\beta' \mathbf{z}_i)}{\sum_{i \in R(t_k)} \exp(\beta' \mathbf{z}_i)}$$

와 같이 구한다. 여기서  $D_r$ 은 시간의 구간  $[t_{r-1}, t_r)$ 에서 관측된 사망자들의 집단이며, 위험집단(risk set)  $R(t_k)$ 은  $t_k$ 시점의 직전까지 생존해서 관측이 가능한 환자들의 집단을 의미한다.  $\mathbf{z}_i$ 는  $i$ 번째 관측치의 공변량을 뜻하고,  $I_i(\mathbf{z}_i)$ 은  $l$ 번째 공변량 집단에 대한 지시함수이다. 만약 비례성의 가정이 적합하지 않다면 시간의 어느 한 구간에서 공변량의 영향력이 다른 구간에서보다 더 커지게 된다. Schoenfeld통계량은  $Q = (d - e)' V^{-1} (d - e)$ 으로 자유도는 분산-공분산행렬  $V$ 의 계수(rank)를 따른다.

## 2.3 Andersen 검정법

Andersen(1982)은  $k$ 개의 공변량을 갖는 비례위험모형이  $k+1$ 번째 공변량  $z_{k+1}$ 을 포함시켰을 때에도 역시 비례위험모형을 따른다는 귀무가설을 검정하는 통계량을 고안하였다.  $S$ 는 공변량  $z_{k+1}$ 의 값에 따라 관측대상자를 구분할 때 구분된 집단의 총 수이고,  $n_s$ 를  $s$ 번째 집단에

서의 관측대상자수로 가정한다. 시간은 Schoenfeld(1980)에서와 같이  $R$ 개의 구간으로 나누어 각 구간에서의 기저위험함수를 상수  $\lambda_{sr}$ 로 가정하였다. 상수  $\lambda_{sr}$ 에 대한 최대우도추정량(maximum likelihood estimator)은

$$\hat{\lambda}_{sr} = \frac{d_{sr}}{\sum_{i=1}^{n_r} \exp(\beta' z_{is}) B_{isr}}$$

와 같다 ( $s=1, \dots, S$ ;  $r=1, \dots, R$ ). 여기서  $i$ 번째 관측치( $i=1, \dots, n$ )는  $(i, s)$ 로 표시되며,  $B_{isr}$ 은  $r$ 번째 시간의 구간에서 관측된  $(i, s)$ 의 생존시간이다.  $d_{sr}$ 은  $r$ 번째 시간에서 관측된 집단  $s$ 의 총 사망자수이다. Andersen(1982)이 생각해낸 비례위험모형에 대한 적합도 검정법의 개념은 앞서 소개한 귀무가설을  $\log \hat{\lambda}_{sr}$ 에 대한 선형모형으로 표현하여 검정에 이용하는 것으로서 귀무가설을 선형으로 표현하면

$$\lambda_{s+1,r} = \lambda_{sr} \exp(\alpha_{s+1})$$

이 된다( $s=1, \dots, S$ ;  $r=1, \dots, R$ ). 다시말해서,

$$\log \lambda_{s+1,r} = \log \lambda_{sr} + \alpha_{s+1} = \log \lambda_{1r} + \sum_{l=2}^{s+1} \alpha_l$$

이 성립된다. 귀무가설하에서는  $\log \hat{\Lambda}_s(t)$ 와  $t$ 에 대하여 그림을 그렸을 때, 각각의 그래프가  $\alpha_2, \dots, \alpha_S$ 만큼 떨어져 있다.  $n_s$ 가 클 때는 귀무가설하에서  $\log \hat{\lambda}_{sr}$ 들은 각각 독립이며 평균이  $\log \lambda_{sr}$ 이고 분산이  $[E(d_{sr})]^{-1}$ 인 정규분포를 근사적으로 따른다. 이러한 사실과 더불어  $d_{sr}^{-1}$ 을 분산의 추정량으로 사용하여  $\hat{\alpha}_{s+1}$ 과  $\log \hat{\lambda}_{1r}$ 에 대한 추정을 하면

$$\hat{\alpha}_{s+1} = \frac{\sum_{r=1}^R \frac{\log \hat{\lambda}_{s+1,r} - \log \hat{\lambda}_{sr}}{(d_{s+1,r})^{-1} + (d_{sr})^{-1}}}{\sum_{r=1}^R ((d_{s+1,r})^{-1} + (d_{sr})^{-1})^{-1}},$$

$$\log \hat{\lambda}_{1r} = \log \hat{\lambda}_r^* = \frac{\sum_{s=1}^S d_{sr} (\log \hat{\lambda}_{sr} - \sum_{l=1}^s \hat{\alpha}_l)}{\sum_{s=1}^S d_{sr}}$$

이 된다. 여기서  $\hat{\alpha}_1 = 0$  이다 ( $s=1, \dots, S-1$ ,  $r=1, \dots, R$ ). 위로부터 얻은  $\log \hat{\lambda}_{sr}$ 에 대한 선형모형을 이용하여 구한 우도비검정통계량(likelihood ratio test statistic)

$$Q = \sum_{r=1}^R \sum_{s=1}^S d_{sr} [\log \hat{\lambda}_{sr} - (\log \hat{\lambda}_{1r} + \sum_{i=1}^s \hat{\alpha}_i)]^2$$

는  $n_1, \dots, n_s \rightarrow \infty$  일 때, 자유도가  $(R-1)(S-1)$ 인 카이제곱분포를 근사적으로 따른다.

Cox는 그의 비례위험모형에서 기저위험함수에 대해 아무런 가정을 하지 않았으나, Andersen은 기저위험함수를 상수로 가정하여 적합도 검정을 실시하였기 때문에 엄밀하게 말해서 비례위험모형의 적합도 검정에 들어맞는다고 하기는 어렵다.

#### 2.4 Wei 검정법

Wei(1984)는 두 개의 생존함수를 비교할 때, 상대위험도를 이용하여 비례위험가정을 검정하는 방법을 제안하였다. 이표본에서 두 개의 위험함수가 비례한다면, 각각의 생존함수간의 관계는  $S_1 = S_2^\theta$ 로 표현되고  $\theta$ 는 상대위험도로 해석할 수 있다.

이표본에서 각각의 표본의 크기가  $n_1, n_2$ 이고 전체 표본크기를  $n$ 이라 하고  $\Lambda_1, \Lambda_2$ 를 각각에 대한 누적위험함수라 하면, 비례성에 대한 귀무가설은  $\Lambda_1(t) = \theta \Lambda_2(t)$ 로 표현된다. Wei의 검정통계량을 얻는 과정은 아래와 같다. 먼저,

$$\begin{aligned} U_n(\theta; t) &= \int_0^t dN_1(s) - \int_0^t Y_1(s) \frac{\theta}{Y_1(s) + Y_2(s)} \times d[N_1(s) + N_2(s)] \\ &= \sum_{i=1}^{n_1} \delta_{i\cdot} I(t_{i\cdot} \leq t) - \sum_{i=1}^2 \sum_{j=1}^{n_i} \frac{Y_1(t_{ij})\theta}{Y_1(t_{ij})\theta + Y_2(t_{ij})} I(t_{ij} \leq t) \end{aligned}$$

을 구한다. 여기서  $t_{ij}$ 는  $i$ 번째 표본의  $j$ 번째 관측치의 생존시간이고,  $N_i(t)$ 는  $i$ 번째 표본에서 생존시간이  $t$ 보다 작거나 같은 사망자수,  $Y_i(t)$ 는  $i$ 번째 표본에서 생존시간이  $t$ 보다 크거나 같은 관측대상자의 수이다.  $\delta_{ij}$ 는 관측된  $t_{ij}$ 가 사망시간이면 1을 중도절단시간이면 0을 갖는다. 만일 상대위험도  $\theta$ 가 알려져 있고 비례위험모형이 적합하다면, 위에서 구한  $U_n(\theta; t)$ 는 결국  $U_n(\theta; t) = N_1(t) - E_1(t)$ 으로 표현된다. 여기서  $E_i(t)$ 는  $i$ 번째 표본에서 생존시간이  $t$ 보다 작거나 같은 기대사망자수이다. 그러므로 구간  $0 \leq t < \infty$ 에서  $|U_n(\theta; t)|$ 값 중 최대값을 검정통계량으로 이용하는 것이 자연스러운 과정이다. 그러나 상대위험도  $\theta$ 를 모르는 경우에는 편우도법으로 추정하여 검정통계량

$$T_n = (n \hat{\theta} \hat{\eta}(\infty))^{-1/2} \max_{0 \leq t < \infty} |U_n(\hat{\theta}; t)|$$

을 구한다. 여기서

$$\begin{aligned}\hat{\eta}(t) &= n^{-1} \left( \int_0^t \frac{Y_1(s) Y_2(s)}{(\partial Y_1(s) + Y_2(s))^2} d(N_1(s) + N_2(s)) \right) \\ &= n^{-1} \sum_{i=1}^2 \sum_{j=1}^{n_i} \frac{\hat{\delta}_{ij} Y_1(t_{ij}) Y_2(t_{ij})}{(\partial Y_1(t_{ij}) + Y_2(t_{ij}))^2}\end{aligned}$$

이다. 위의 검정통계량에 대한 분포표는 Koziol and Byar(1975)에 나와있다.

## 2.5 Gill-Schumacher 검정법

Gill and Schumacher(1987)는 비례위험(proportional hazards)의 가정 중에서도 특히 귀무가설이  $\lambda_2(t)/\lambda_1(t) = \theta > 0$  인 경우의 적합도 검정에 관심을 가졌다. 즉,

$$H_0 : \text{어떤 양수 } \theta \text{에 대해 } \lambda_2(t)/\lambda_1(t) = \theta,$$

$$H_1 : \text{모든 양수 } \theta \text{에 대해 } \lambda_2(t)/\lambda_1(t) \neq \theta$$

와 같은 가설에 대하여, 가중함수(weight function)  $K_i(t)$  ( $i=1, 2$ )를 이용한 일반화순위추정량(generalized rank estimator)인

$$\begin{aligned}\hat{\theta}_{K_i} &= \int K_i(t) d\hat{\Lambda}_2(t) / \int K_i(t) d\hat{\Lambda}_1(t) \\ &= \hat{R}_2 / \hat{R}_1\end{aligned}$$

으로  $\theta$ 를 추정하였다. 귀무가설  $H_0$ 하에서는 표본크기가 커짐에 따라  $\hat{\theta}_{K_i}$ 는  $\theta$ 에 확률적으로 수렴(converge in probability)한다. 그러므로 표본크기가 클 경우에는 서로 다른 가중함수  $K_1(t)$ 와  $K_2(t)$ 로 구한 추정량  $\hat{\theta}_{K_1}$ 와  $\hat{\theta}_{K_2}$ 간의 차이인  $\hat{R}_{22}/\hat{R}_{21} - \hat{R}_{12}/\hat{R}_{11}$ 는 적어지게 된다. 그러나 대립가설  $H_1$ 하에서는 이 두 추정량의 차이가 커지므로 이 사실에 기초한 검정통계량을 구하여 귀무가설을 검정한다. 계산상의 편의를 위해  $\hat{R}_{22}/\hat{R}_{21} - \hat{R}_{12}/\hat{R}_{11}$ 은

$$Q_{K_1, K_2} = \hat{R}_{11} \hat{R}_{22} - \hat{R}_{21} \hat{R}_{12}$$

로 대체될 수 있으며, 귀무가설 하에서  $Q_{K_1, K_2}$ 는 0에 가까운 값을 갖는다. 귀무가설하에서  $Q_{K_1, K_2}$ 의 분산을

$$\begin{aligned}\widehat{\text{Var}}(Q_{K_1, K_2}) &= \hat{R}_{21} \hat{R}_{22} \hat{V}_{11} - \hat{R}_{21} \hat{R}_{12} \hat{V}_{12} \\ &\quad - \hat{R}_{11} \hat{R}_{22} \hat{V}_{21} + \hat{R}_{11} \hat{R}_{12} \hat{V}_{22}\end{aligned}$$

와 같이 구할 수 있다. 여기서

$$\hat{V}_i = \int K_i(t) K_i(t) [Y_1(t) Y_2(t)]^{-1} d(N_1(t) + N_2(t))$$

으로 표현할 수 있다. 이렇게 추정하여 얻은 Gill-Schumacher 통계량

$$T_{K_1 K_2} = \widehat{\text{Var}}(Q_{K_1 K_2})^{-1/2} Q_{K_1 K_2}$$

은  $n \rightarrow \infty$  일 때 근사적으로 표준정규분포를 따른다. 여기서 쓰인  $N_i(t)$  와  $Y_i(t)$  는 Wei(1984)의 적합도 검정법에서 정의된 것과 같은 의미이다.

## 2.6 Lin-Wei 검정법

White(1982)는 정보행렬(information matrix)의 성질을 이용하여 모형을 잘못 선정하는 것(model misspecification)에 관한 검정법을 제안하였다. Lin and Wei(1991)는 White의 방법을 편우도를 이용한 비례위험모형에 적용하였다.

비례위험모형에 대한 편우도함수(partial likelihood function)

$$L(\beta) = \prod_{i=1}^n \left[ \frac{\exp(\beta' z_i)}{\sum_{j \in R} \exp(\beta' z_j)} \right]$$

에 로그를 취하면

$$\begin{aligned} l(\beta) &= \log L(\beta) = \sum l_i(\beta) \\ &= \sum [\beta' z_i - \log(\sum_{j \in R} \exp(\beta' z_j))] \end{aligned}$$

와 같다. 모형(2.1)에 대한 최대편우도추정량(maximum partial likelihood estimator)  $\hat{\beta}$ 은 근사적으로 정규분포를 따른다는 사실이 알려져 있다. 모형 (2.1)하에서의  $n^{1/2}(\hat{\beta} - \beta_0)$ 의 근사적 공분산행렬의 역행렬을  $\Omega$ 라 하면,  $\Omega$ 는 두 개의 일치추정량인

$$\begin{aligned} A_n(\hat{\beta}) &= -n^{-1} \partial^2 l(\beta) / \partial \beta^2 \Big|_{\beta = \hat{\beta}}, \\ B_n(\hat{\beta}) &= n^{-1} \sum (\partial l_i(\beta) / \partial \beta) (\partial l_i(\beta) / \partial \beta)' \Big|_{\beta = \hat{\beta}} \end{aligned}$$

으로 추정할 수 있다.  $D_n(\hat{\beta}) = A_n(\hat{\beta}) - B_n(\hat{\beta})$ 이고  $\text{vec}(H)$ 를  $((i-1)m+j)$ 번째 요소가  $m \times n$ 행렬  $H$ 의  $(i, j)$ 번째 요소와 같은  $mn$ 벡터라 가정하면, 모형(2.1)하에서  $n^{1/2} \text{vec}(D_n(\hat{\beta}))$ 가 평균이 0이고 공분산 행렬이  $Q(\beta_0)$ 인 정규분포에 근사한다는 것이 증명되었다.  $Q(\beta_0)$ 는  $Q_n(\hat{\beta})$ 으로 추정하고  $p(p+1)/2$  벡터  $d_n(\hat{\beta})$ 는  $D_n(\hat{\beta})$ 의 상삼각위치

(upper triangular) 요소라 할 때, 왈드통계량(Wald statistic)

$$W_n = n d_n(\hat{\beta})' Q_n(\hat{\beta})^{-1} d_n(\hat{\beta})$$

은 모형(2.1)하에서 자유도  $p(p+1)/2$ 의 카이제곱분포를 따르는 변수로 그 분포가 수렴한다.

Lin and Wei(1991)는 모형(2.1)의 가정을 다음과 같이 세 가지로 간주하였다. 첫 번째 가정은 관련이 있는 모든 공변량은 모형에 포함되었다는 것이고, 두 번째로는 공변량에 대한 위험함수의 회귀형태는 지수적이란 것, 마지막으로 위험함수는 공변량의 회귀함수와 기저위험함수의 곱의 형태로 나타낸다는 것이다. 이와 같이 이 통계량은 비례성의 가정외에도 다른 가정을 검정할 수 있도록 고안된 것이다.

### 3. 모의실험의 구조 및 결과

비례위험모형에 대한 적합도 검정법을 모의실험을 통해 좀 더 구체적으로 살펴보기로 한다. 모의실험에 이용한 통계량은 2장에서 소개한 Cox통계량, Schoenfeld통계량, Andersen통계량, Wei통계량, Gill-Schumacher통계량과 Lin-Wei통계량이다. Schoenfeld통계량과 Andersen통계량은 6개의 구간으로 시간을 나누어 계산하였다. Gill-Schumacher통계량에 이용한 두 개의 가중함수는 Gehan 함수와 로그순위(log-rank)함수이다. 두 함수를 간단히 소개하면 Gehan함수는 가중치가 남은 환자수에 비례하고 로그순위(log-rank)함수는 가중치가 일정하다.

모의실험은 실제로 생존분석시 흔히 볼 수 있는 이표본 자료에 대해 이루어졌으며, 공변량은 한 개로 표본에 관한 지시함수이다. 검정력을 알아보기 위하여 여러 가지 형태의 자료를 생성하였다. 자료는 <그림 1>에서 나타난 것처럼 두 표본 모두 완치되지 않는 4가지 경우와 <그림 2>에서처럼 두 표본 모두 완치되는 6가지 경우로 나누었다. 완치되지 않을 때의 자료형태는 비례위험모형의 가정을 따르는 자료(I)와 두 처리간의 차이가 초반(II), 후반(III), 그리고 중반(IV)에 나타나는 경우이다. 완치될 때의 자료는 생존시간이 각각  $\exp(1)$ 이나  $\exp(2)$ 를 따르면서 치료율이 0.2 혹은 0.4인 6가지 경우이다. 이 10가지 형태의 자료는 중도절단 유무와 두 가지의 표본크기에 따라 각각 생성되었다. 중도절단이 있을 시에는 생존시간과 독립적으로 균일 분포  $U(0, 2)$ 를 따른다. 표본크기는 50과 100의 두 가지이다. 위의 과정을 통해 얻은 각각의 자료를 이용해 유의수준 0.05와 0.01에 대해 1000번씩 모의실험을 시행하였다.

각 표의 제목 옆에 있는 자료에 대한 위험율비(hazard ratio)는 다음과 같이 표현된 것이다. 즉,  $\theta = \lambda_2(t)/\lambda_1(t)$ 라 할 때, 비례성의 가정하에서는  $\theta$  값이 시간에 관계없이 일정하나 대립가설하에서는 시간이 지남에 따라  $\theta$ 의 값이 수시로 변한다. 표에서  $\theta = a \rightarrow \dots \rightarrow d$ 는 시간에 따른 위험율비의 변화를 나타내는 것이다. 또한  $\theta$ 가 단조변화하는 경우에 대해서도 언급하였다.

먼저 비례성의 가정이 성립하는 경우의 모의실험의 결과를 살펴보자. <표 1>에서 보는 바와 같이 위험율비는  $\theta = 0.5$ 로 시간에 관계없이 일정하다. Lin-Wei통계량은 중도절단유무에 따라 그 값이 큰 차이를 보인다. 특히 중도절단이 없는 경우는 유의수준보다 매우 큰 값을 갖



는다. Gill-Schumacher통계량은 유의수준보다 약간 높은 값이 구해졌으며, Wei통계량등 다른 통계량들은 유의수준보다 낮거나 비슷한 값이 얻어졌다.

두 표본의 생존분포의 차이가 초반에 나타날 때의 모의실험 결과는 <표 2>에 있다. 이 때 위험율비는 불규칙하게 변하고 있다. Lin-Wei통계량은 다른 통계량에 비해 검정력이 낮다. 중도절단유무에 관계없이 가장 높은 검정력을 보이는 통계량이 Wei통계량이다. 단조변화하는 대립가설에 대해 민감하다고 알려져있는 Cox통계량이나 Gill-Schumacher통계량도 대립가설(II)에서는 다른 검정통계량과 비슷하거나 조금 낮은 검정력을 보이고 있다. 즉 전체적으로 보아 위험율비가 불규칙하게 변할 때에는 이 두 통계량의 검정력이 다소 떨어진다.

<표 3>은 후반에 생존분포의 차이가 나타나는 자료의 모의실험 결과를 나타낸 것으로 이 대립가설은 위험율비가 중앙시간(median time)을 기준으로 1에서 0.1로 감소한 후 다른 변화가 없어서 단조감소하는 경우와 비교할 만하다. 따라서 Lin and Wei(1991)이나 Song and Lee(1994)에서 보인 단조변화하는 대립가설들과 비교할 수 있다. 여기에서는 모든 통계량들이 대체로 만족할만한 검정력을 보인다. 특히 대립가설(II)에서 상대적으로 가장 낮은 검정력을 보인 Cox통계량이 대립가설(III)에서는 다른 어느 통계량보다도 큰 검정력을 보인다. 이 결과는 Cox통계량이 단조변화에 민감하다는 사실을 보여준다. 또한, Lin and Wei(1991)나 Song and Lee(1994)에서 보인 것처럼 단조변화하는 대립가설하에서는 대체로 모든 통계량이 만족할만한 검정력을 갖는다는 사실과 일치한다. 그러나 Lin-Wei통계량은 대립가설 (II)에서와 마찬가지로 다른 통계량보다 낮은 검정력을 보일 뿐 아니라, 귀무가설(I)이나 대립가설(II)에서와 같이 중도절단유무에 따른 검정력의 차이가 두드러진다. 이는 Lin and Wei(1991)에서도 언급된 것처럼 가정된 모형에 근거하여 얻은 공분산행렬의 추정량들이 잔차와 생존시간의 관계에 대하여 정보를 많이 줄 수 없기 때문이다.

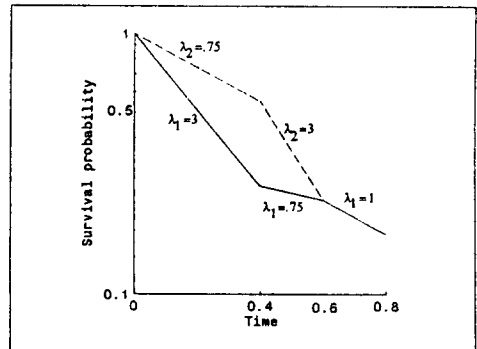
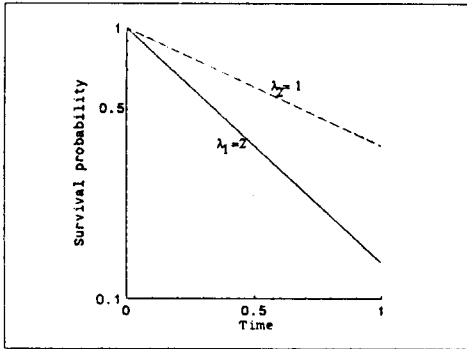
생존분포의 차이가 중반에 나타날 때의 자료에 대한 모의실험 결과는 <표 4>에 있다. 이 때의 위험율비는 다른 어느 대립가설하에서보다도 불규칙하다. 단조변화에는 비교적 민감하다고 알려진 Cox통계량과 Gill-Schumacher통계량이 이 대립가설에서는 검정력이 몹시 낮다. 이런 상황은 앞에서 여러 번 강조된 것처럼 두 통계량이 단조변화의 경우외에는 검정력이 약하다는 사실을 단적으로 보여준다. 또한 Lin-Wei통계량도 역시 검정력이 낮다. Lin-Wei통계량은 비례위험모형에 대한 가정을 2.6절에서 언급한 것처럼 좀더 일반적으로 해석하여 만들어진 통계량 이므로, 위험율비가 상수라는 특정한 가정에는 좋은 검정력을 가지지 못하는 듯하다. 이 대립가설에서 다른 통계량에 비하여 검정력이 나은 것은 Schoenfeld통계량, Andersen통계량과 Wei통계량이다. Schoenfeld통계량이나 Andersen통계량은 구간을 나누어 계산된 것이므로 각 구간에서의 위배여부에 더 민감하게 반응할 수 있기 때문에 이런 상황의 대립가설하에서는 다른 통계량보다 검정력이 더 좋다.

다음으로 완치되는 경우에 대해 알아보겠다. 완치되는 경우의 위험율비는 비례성의 가정이 적합하는 경우를 제외하곤 모두 단조변화하고 있다. 그러므로 모의실험 결과는 다음의 두 가지 형태의 자료에 대해서만 보이기로 한다.

<표 5>는 생존시간이 각각  $\exp(1)$ 과  $\exp(2)$ 를 따르고 완치율이 각각 0.2, 0.4를 따르는 경우의 모의실험 결과이다. 이 때의 위험율비는 단조감소한다. Cox통계량, Wei통계량과 Gill-Schumacher통계량의 검정력이 다른 검정통계량보다 크며, Lin-Wei통계량의 검정력이 다른 통계량에 비해 작다. 표본크기가 100으로 커지면 검정력이 더 좋아지는데, 이 때 Andersen

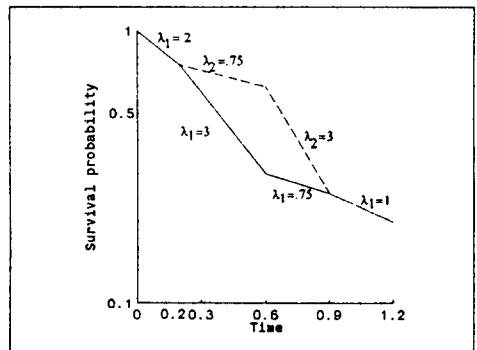
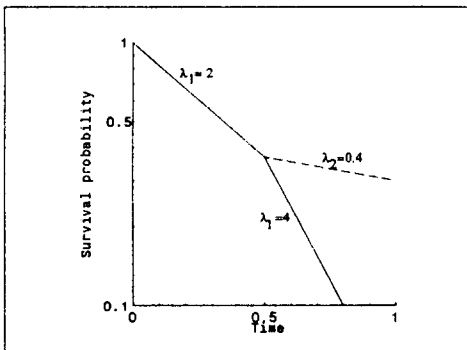
통계량의 검정력이 다른 통계량의 변화에 비해 더 크게 변하고 있다.

생존시간이 각각  $\exp(2)$  와  $\exp(1)$ 을 따르고 완치율이 0.2로 같을 때의 자료에 대한 모의실험 결과는 <표 6>에 실려 있다. 이 때의 위험율비는 단조증가하는데 앞서 <표 5>에서와 마찬가지로 Cox통계량, Wei통계량과 Gill-Schumacher통계량의 검정력이 다른 통계량의 검정력보다 크며 Schoenfeld통계량의 검정력이 다소 낮아진 것을 볼 수 있다. 여기서도 표본크기가 커지면 검정력이 커지는 자연스런 결과를 볼 수 있다.



I. 비례위험모형의 가정이 적합하는 경우

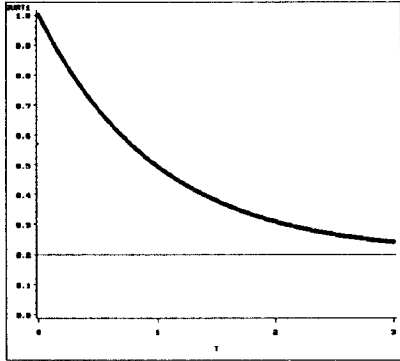
II. 초반에 차이가 나는 경우



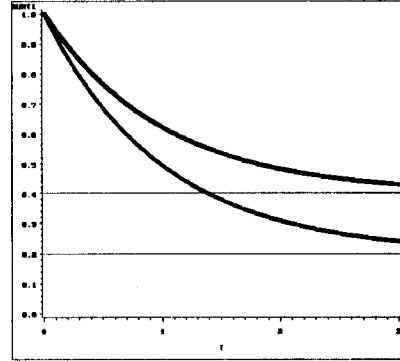
III. 후반에 차이가 나는 경우

IV. 중반에 차이가 나는 경우

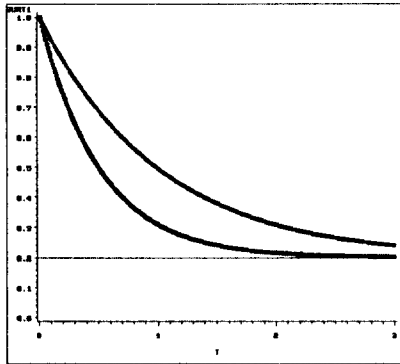
<그림 1> 완치되지 않는 경우 4 가지



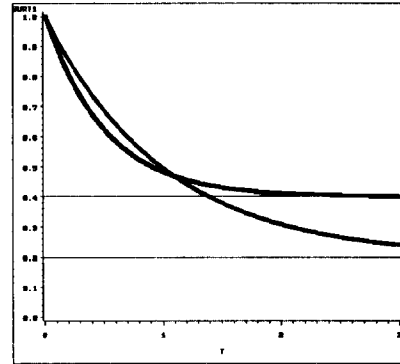
집단1: exp(1), 완치율=0.2  
 집단2: exp(1), 완치율=0.2



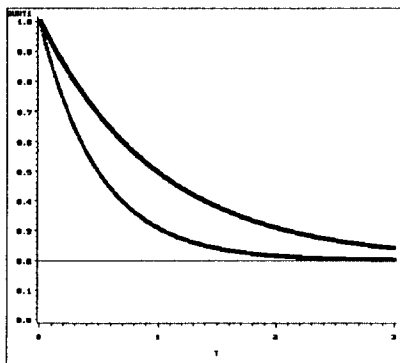
집단1: exp(1), 완치율=0.2  
 집단2: exp(1), 완치율=0.4



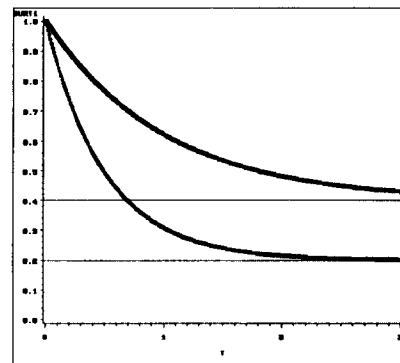
집단1: exp(1), 완치율=0.2  
 집단2: exp(2), 완치율=0.2



집단1: exp(1), 완치율=0.2  
 집단2: exp(2), 완치율=0.4



집단1: exp(2), 완치율=0.2  
 집단2: exp(1), 완치율=0.2



집단1: exp(2), 완치율=0.2  
 집단2: exp(1), 완치율=0.4

<그림 2> 완치되는 경우 6가지

<표 1> 비례위험모형의 가정이 적합하는 경우 (configuration I :  $\theta=0.5$ )

표본크기	검정통계량	유의수준		
		$\alpha=0.01$	$\alpha=0.05$	
50	Cox	0.009	0.053	
		0.008	0.052	
	Shoenfeld	0.008	0.045	
		0.007	0.037	
	Andersen	0.007	0.042	
		0.005	0.030	
	Wei	0.008	0.041	
		0.005	0.033	
	Gill-Schumacher	0.016	0.062	
		0.016	0.063	
	Lin-Wei	0.027	0.086	
		0.007	0.033	
	100	Cox	0.009	0.045
			0.011	0.046
Shoenfeld		0.007	0.035	
		0.014	0.053	
Andersen		0.009	0.033	
		0.009	0.044	
Wei		0.009	0.053	
		0.014	0.058	
Gill-Schumacher		0.011	0.061	
		0.011	0.064	
Lin-Wei		0.017	0.062	
		0.001	0.021	

첫째 줄 : 중도절단 없음

둘째 줄 :  $c_i \sim U(0, 2)$

<표 2> 초반에 차이가 나는 경우 (configuration II :  $\theta = 0.25 \rightarrow 4 \rightarrow 1$ )

표본크기	검정통계량	유의수준		
		$\alpha = 0.01$	$\alpha = 0.05$	
50	Cox	0.498	0.414	
		0.346	0.626	
	Shoenfeld	0.550	0.773	
		0.510	0.804	
	Andersen	0.462	0.746	
		0.275	0.694	
	Wei	0.843	0.955	
		0.648	0.861	
	Gill-Schumacher	0.675	0.866	
		0.592	0.773	
	Lin-Wei	0.369	0.547	
		0.163	0.345	
	100	Cox	0.899	0.785
			0.911	0.903
Shoenfeld		0.954	0.992	
		0.971	0.997	
Andersen		0.942	0.991	
		0.949	0.994	
Wei		0.999	1.000	
		0.980	0.996	
Gill-Schumacher		0.949	0.992	
		0.912	0.986	
Lin-Wei		0.513	0.675	
		0.345	0.546	

첫째 줄 : 중도절단 없음

둘째 줄 :  $c_i \sim U(0, 2)$

<표 3> 후반에 차이가 나는 경우 (configuration III:  $\theta = 1 \rightarrow 0.1$ )

표본크기	검정통계량	유의수준		
		$\alpha = 0.01$	$\alpha = 0.05$	
50	Cox	0.893	0.984	
		0.571	0.805	
	Shoenfeld	0.782	0.929	
		0.302	0.596	
	Andersen	0.905	0.970	
		0.353	0.665	
	Wei	0.846	0.954	
		0.334	0.623	
	Gill-Schumacher	0.763	0.910	
		0.413	0.637	
	Lin-Wei	0.116	0.767	
		0.013	0.177	
	100	Cox	0.942	1.000
			0.869	0.974
Shoenfeld		0.999	1.000	
		0.838	0.943	
Andersen		1.000	1.000	
		0.884	0.961	
Wei		1.000	1.000	
		0.860	0.965	
Gill-Schumacher		0.983	1.000	
		0.773	0.908	
Lin-Wei		0.979	0.999	
		0.397	0.787	

첫째 줄 : 중도절단 없음

둘째 줄 :  $c_i \sim U(0, 2)$

&lt;표 4&gt; 증반에 차이가 나는 경우

(configuration IV:  $\theta = 1 \rightarrow 0.25 \rightarrow 4 \rightarrow 1$ )

표본크기	검정통계량	유의수준		
		$\alpha = 0.01$	$\alpha = 0.05$	
50	Cox	0.030	0.090	
		0.021	0.070	
	Shoenfeld	0.118	0.335	
		0.138	0.397	
	Andersen	0.074	0.286	
		0.114	0.358	
	Wei	0.090	0.291	
		0.135	0.135	
	Gill-Schumacher	0.020	0.072	
		0.020	0.072	
	Lin-Wei	0.005	0.020	
		0.002	0.028	
	100	Cox	0.057	0.154
			0.042	0.094
Shoenfeld		0.409	0.654	
		0.549	0.791	
Andersen		0.389	0.643	
		0.530	0.490	
Wei		0.638	0.637	
		0.319	0.319	
Gill-Schumacher		0.024	0.108	
		0.014	0.067	
Lin-Wei		0.000	0.002	
		0.002	0.010	

첫째 줄 : 증도절단 없음

둘째 줄 :  $c_i \sim U(0, 2)$

&lt;표 5&gt; 완치율이 다른 경우 (집단1: exp(1) 완치율=0.2, 집단2: exp(2) 완치율=0.4)

표본크기	검정통계량	유의수준		
		$\alpha = 0.01$	$\alpha = 0.05$	
50	Cox	0.576	0.877	
		0.079	0.305	
	Shoenfeld	0.116	0.306	
		0.005	0.033	
	Andersen	0.282	0.541	
		0.023	0.091	
	Wei	0.449	0.714	
		0.059	0.180	
	Gill-Schumacher	0.647	0.818	
		0.158	0.341	
	Lin-Wei	0.001	0.019	
		0.001	0.015	
	100	Cox	0.986	0.988
			0.479	0.573
Shoenfeld		0.566	0.776	
		0.039	0.119	
Andersen		0.842	0.942	
		0.181	0.394	
Wei		0.880	0.968	
		0.179	0.411	
Gill-Schumacher		0.946	0.983	
		0.322	0.570	
Lin-Wei		0.010	0.183	
		0.019	0.113	

첫째 줄 : 중도절단 없음

둘째 줄 :  $c_i \sim U(0, 2)$



&lt;표 6&gt; 완치율이 같은 경우 (집단1: exp(2) 완치율=0.2, 집단2: exp(1) 완치율=0.2)

표본크기	검정통계량	유의수준		
		$\alpha=0.01$	$\alpha=0.05$	
50	Cox	0.316	0.667	
		0.030	0.144	
	Shoenfeld	0.037	0.117	
		0.004	0.011	
	Andersen	0.081	0.334	
		0.018	0.054	
	Wei	0.184	0.417	
		0.019	0.088	
	Gill-Schumacher	0.330	0.549	
		0.059	0.152	
	Lin-Wei	0.072	0.226	
		0.001	0.025	
	100	Cox	0.597	0.903
			0.079	0.197
Shoenfeld		0.156	0.330	
		0.004	0.024	
Andersen		0.469	0.665	
		0.055	0.133	
Wei		0.513	0.759	
		0.041	0.153	
Gill-Schumacher		0.633	0.814	
		0.085	0.230	
Lin-Wei		0.100	0.290	
		0.003	0.007	

첫째 줄 : 중도절단 없음

둘째 줄 :  $c_i \sim U(0, 2)$

#### 4. 맺음말

모형을 선택하여 자료를 분석하기에 앞서 적합도 검정법을 실시하여 가정을 검토하여야 한다. 특히 널리 쓰이는 모형의 경우에는 그만큼 모형 선택에 오류가 있을 수도 있다고 생각하고 더 세심한 주의를 기울여 적합여부를 판별해야 할 것이다.

위의 모의실험 결과를 정리하면 다음과 같다. 위험율이 비례성의 가정을 따르는 경우에는 대부분의 통계량이 유의수준과 비슷한 값을 가지나 Lin-Wei통계량이 중도절단에 민감하게 반응하여 중도절단이 없을 시에는 유의수준보다 약간 큰 값을 가진다. 비례성의 가정을 따르지 않고 위험율이 단조변화할 때는 Cox통계량과 Gill-Schumacher통계량의 검정력이 좋았으며, Wei통계량의 검정력도 다른 통계량에 비해 큰 값을 가지고 있었다. 그러나 위험율이 불규칙하게 변화할 시에는 시간을 여러 개의 구간으로 나누어 얻는 Schoenfeld통계량이나 Andersen 통계량의 검정력이 뛰어났다. 특히 완치되지 않는 경우인 대립가설 (IV)에서처럼 위험율의 변화가 심할 때에는 시간을 관측갯수에 따라 균등하게 나누기 보다는 서로 다른 위험율을 갖도록 나누었을 때 Schoenfeld통계량과 Andersen통계량의 검정력이 훨씬 좋았다. 대립가설 (IV)에서처럼 위험율이 불규칙하게 변할 때에는 Cox통계량이나 Gill-Schumacher통계량의 검정력이 현저히 떨어졌는데, 이는 Cox통계량과 Gill-Schumacher통계량이 단조변화하는 대립가설에 특히 민감하도록 고안되었다는 것을 단적으로 보여주고 있다. 이 결과는 Lin and Wei(1991)이나 Song and Lee(1994)에서 보여진 것과 같은 결과이다.

완치가 되는 경우에는 공교롭게도 비례성의 가정에 적합하는 경우를 제외하고는 모두 위험율이 증가하거나 감소하면서 단조변화하는 모습을 보였다. 위험율이 단조변화할 시에는 Cox통계량, Gill-Schumacher통계량과 Wei통계량의 검정력이 돋보였다. 또한 Schoenfeld통계량, Andersen통계량과 Lin-Wei통계량의 검정력은 예상대로 저조했다. 이와 같이 단조변화할 시에는 Cox통계량과 Gill-Schumacher통계량이 돋보인다는 것을 확인할 수 있었다.

지금까지의 결과를 볼 때, 단조변화할 때에는 Cox통계량, Wei통계량과 Gill-Schumacher통계량의 검정력이 좋고 불규칙하게 변할 시에는 Cox통계량이나 Gill-Schumacher통계량보다는 Schoenfeld통계량과 Andersen통계량이 돋보인다. 그러므로 자료의 형태에 대해 아무런 사전지식을 갖고 있지 않은 경우에는 두 통계량을 이용하여 적합도 검정법을 하는 것이 바람직하다. Lin-Wei통계량은 같은 가설하에서도 중도절단 여부에 따라 검정력의 차이가 두드러졌을 뿐 아니라 대부분의 가설에 대해 다른 검정통계량보다 검정력이 낮았다. 이것은 Lin-Wei통계량이 2.6절에서 언급된 것처럼 여러 가정에 대하여 검정할 수 있도록 고안된 것이기 때문이다. 만일 비례위험모형이 어떻게 부적절한가를 자세하게 알고 싶다면, 그림을 이용한 검정법등 다른 검정법을 이용하는 것도 좋을 것이다 (cf. Chappell, 1992). Schoenfeld통계량과 Andersen통계량은 구간의 수에 따라 다른 결과를 낳기도 하지만 단조변화가 아닌 경우의 대립가설에 대하여 다른 통계량보다 검정력이 좋다. Wei통계량은 모든 가설에 대하여 적절한 수치를 보여 고안자가 의도한 대로 총괄적인(omnibus) 검정법임을 보였다. 또한 검정력이 표본크기가 커지고 유의수준이 커질수록 좋았으나 검정력의 순위는 자료형태에 따라 달랐으므로 검정력의 순위가 유의수준이나 표본크기의 영향보다는 자료형태에 좌우된다는 사실을 확인할 수 있었다.

모의실험 결과에 의하면 자료에 따라 특정한 검정법의 검정력이 뛰어나다는 것을 알 수 있다. 시간에 따라 위험율이 단조변화하는 것 같은 자료에 대하여는 어느 검정법을 써도 무방

하나, 자료의 형태에 대하여 아무런 사전지식이 없을 경우에는 Cox통계량, Gill-Schumacher통계량이나 Lin-Wei통계량의 사용을 자제하고 Schoenfeld통계량이나 Andersen통계량을 이용하여 적합도 검정법을 실시하는 것이 좋다.

앞에서 소개한 6가지의 비례위험모형의 적합성 검정법들 중에서, Cox, Wei, Gill-Schumacher의 검정법은 공변량이 지시변수 하나만 존재할 때 적용할 수 있는 반면에 Lin-Wei, Schoenfeld, Andersen의 검정법은 공변량의 형태나 갯수에 관계없이 적용할 수 있다. 이 중 Schoenfeld와 Andersen의 검정법은 공변량의 범위와 시간에 따라 주관적으로 구간을 나누어야 하는 단점이 있지만 Lin-Wei의 검정법은 보다 일반적으로 적용할 수 있는 장점이 있다.

## 참 고 문 헌

- [1] Andersen, P. K. (1982). Testing goodness of fit of Cox's regression and life model, *Biometrics*, 38, 67-77.
- [2] Chappell, R. (1992). Proportional hazards diagnostics and their implementation: a brief summary, *Technical Report No. 64, Dept. of Biostatistics, University of Wisconsin-Madison*.
- [3] Cox, D. R. (1972). Regression models and life tables (with discussion), *Journal of Royal Statistical Society Series B*, 34, 187-220.
- [4] Cox, D. R. (1975). Partial likelihood, *Biometrika*, 62, 269-276.
- [5] Gill, R. D. and Schumacher, M. (1987). A simple test of the proportional hazards assumption, *Biometrika*, 74, 289-300.
- [6] Haara, P. (1987). A note on the asymptotic behavior of the empirical score in Cox's regression model for counting processes, *Proceedings of the 1st World Congress of the Bernoulli Society*, 139-142, VNU Science Press, Tashkent.
- [7] Kay, R. (1977). Proportional hazards regression models and the analysis of censored survival data, *Applied Statistics*, 26, 227-237.
- [8] Koziol, J. and Byar, D. (1975). Percentage points of the asymptotic distribution of one and two sample K-S statistics for truncated or censored data, *Technometrics*, 17, 507-510.
- [9] Lin, D. Y. and Wei, L. J. (1991). Goodness-of-fit tests for the general Cox regression model, *Statistica Sinica*, 1, 1-17.
- [10] Schoenfeld, D. (1980). Chi-Squared goodness-of-fit tests for the proportional hazards regression model, *Biometrika*, 67, 145-153.
- [11] Song, H. H. and Lee, S. H. (1994). Goodness of fit tests of Cox's proportional hazards model, *Journal of the Korean Statistical Society*, 23, No. 2, 379-402.
- [12] Wei, L. J. (1984). Testing goodness of fit for proportional hazards model with censored observations, *Journal of the American Statistical Association*, 79, 649-652.
- [13] White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica* 50, 1-25.

## A study on the goodness-of-fit tests for proportional hazards model<sup>4)</sup>

Ae Bang Jang<sup>5)</sup> , Jae Won Lee<sup>6)</sup>

### Abstract

Proportional hazards model has been widely used for analyzing survival data. This article reviews some well-known goodness-of-fit tests for proportional hazards model. Simulation studies also provide some insights into the properties of these test statistics across several types of survival distributions and degrees of censorship.

---

4) This is research was supported by Korea Science and Engineering Foundation grant.

5) Management Information Team, Cheil C&C Inc., 6F Cheil Bldg. 500, 5-ga, Namdaemoon-no, Chung-ku, Seoul 100-095.

6) Associate Professor, Department of Statistics, Korea University, 5-1 Anam-dong, Sungbuk-gu, Seoul 136-701.