The Korean Journal of Applied Statistics (2012)
25(3), 423–434

DOI: http://dx.doi.org/10.5351/KJAS.2012.25.3.423

A Comparison of Testing Methods for Equality of Survival Distributions with Interval Censored Data

Soo-Hwan Kim[1] · Shin-Jae Lee[2] · Jae Won Lee[3]

[1]Department of Statistics, Korea University
[2]Department of Orthodontics, School of Dentistry and Dental Research Institute, Seoul National University; [3]Department of Statistics, Korea University

(Received January 15, 2012; Revised February 6, 2012; Accepted May 9, 2012)

Abstract

A two-sample test for equality of survival distribution is one of the important issues in survival analysis, especially for clinical and epidemiological research. With interval censored data, some testing methods have been developed. This study introduces some testing methods and compares them under various situations through simulation study. Based on simulation result, it provides some useful information on choosing the most appropriate testing method in a given situation.

Keywords: Interval censored data, log-rank test, IWD test.

1. 서론

구간중도절단자료를 가지는 경우는 사건이 발생한 정확한 시점이 관측되지 않고 두 개의 시점 사이에서 발생하였다는 사실만이 관측되어 있는 경우로 임상 연구 및 의학 연구에서 흔히 발생한다. 예를 들어 특정한 질병이 발생한 시점은 병원에서 진료받은 시각의 사이인 구간으로 관측된다. 하지만 현재 생존분석에 관련된 연구의 대부분은 구간중도절단자료를 가지는 경우가 빈번히 발생함에도 불구하고 우측중도절단자료를 가지는 경우에 대한 것들이 대부분이다. 상대적으로 구간중도절단자료를 가지는 경우의 분석법은 그 수가 그리 많지 않고, 실제로 자료를 분석함에 있어서도 구간중도절단자료를 우측중도절단자료를 가진 경우로 간주하여 분석하는 경우가 많다. 구간중도절단자료를 가지는 경우에 생존함수의 추정법은 Turnbull (1976)이 자기일치법 알고리즘(self-consistency algorithm)을 제안한 바 있고, Gentleman과 Geyer (1994)에서는 자기일치법 알고리즘으로 얻어진 생존함수의 추정치가 비모수적 최우추정치(NPMLE; nonparametric maximum likelihood estiamator)로써 가지는 몇 가지 성질들을 증명하였다.

중도절단이 있는 자료를 분석함에 있어 주요한 관심의 대상들 중 하나는 두 생존분포의 비교이다. 두 생존분포의 비교는 신약의 개발이나 새로운 치료법 개발 등 그 사용이 필수적이니만큼 많은 학자들에

Jae Won Lee was supported by the National Research Foundation(NRF) grant funded by the Korea government(MEST) (2011-0027601). Shin-Jae Lee was supported by the National Research Foundation(NRF) grant funded by the Korea government(MEST) (2011-0026594, 2011-0028067).
[3]Corresponding author: Professor, Department of Statistics, Korea University, Anam-Dong, Seongbuk-Gu, Seoul 136-701, Korea. E-mail: jael@korea.ac.kr

의해 연구가 되어 왔다. 우측중도절단을 가지는 자료에 대해서는 가장 일반적으로 사용되고 있는 검정법인 로그-순위 검정법과 일반화된 Wilcoxon 검정법, 중위수를 이용한 검정법, 가중 Kaplan-Meier 검정법 등이 있다. 두 생존분포의 비교시에는 각 상황마다 높은 검정력을 가지는 검정법을 사용하는 것이 중요한데, 이에 관하여는 Jung과 Lee (1998)의 연구가 있다. Finkelstein (1986)과 Zhao와 Sun (2004)은 우측중도절단자료가 있는 경우의 검정에 사용하는 로그-순위 검정법을 확장하여 구간중도절단자료를 가지는 경우에 적용할 수 있는 로그-순위 검정법을 제안하였다. 다른 검정 방법으로 Petroni와 Wolfe (1994)의 연구에서 제안된 가중적분차(IWD; integrated weighted difference)를 이용하는 검정법이 있다.

본 연구에서는 구간중도절단자료를 가지는 경우에 두 생존분포의 비교에 사용할 수 있는 몇 가지 검정법들을 소개하고, 여러 가지 상황하에서 모의실험을 실시하여 이 검정법들의 유의수준 및 검정력을 비교하였다.

2. 구간중도절단자료를 가지는 경우에 대한 두 생존분포의 동일성 검정법

먼저 비교를 위한 두 표본을 생각하자. 각 표본의 수는 n으로 동일하고, 임의중도절단(random censoring)이 일어난다. T_{il}을 i번째 개체, l번째 표본에서 얻어진 실패시간(failure time)이라 하자 ($i = 1, \ldots, n$; $l = 1, 2$). 실제로 T_{il}는 관측되지 않고, 대신 T_{il}를 포함하고 있는 구간인 $(L_{il}, R_{il}]$만이 알려져 있다($L_{il} < T_{il} \leq R_{il}$). 이에 바탕하여 l번째 표본에서의 생존함수(survival function)는 $S_l(t) = \Pr(T_{il} > t)$으로 정의된다.

우리가 검정하고자 하는 것은 두 생존분포의 동일성이므로 귀무가설은

$$H_0 : S_1(t) = S_2(t) \tag{2.1}$$

이다. 구간중도절단을 가지는 자료에서 식 (2.1)의 가설을 검정하기 위한 여러 가지 방법이 제안되어 왔다. 먼저 각 검정법에 대해 소개하도록 하겠다.

2.1. 로그-순위 검정법

로그-순위 검정법은 생존함수의 동일성을 검정할 때 가장 일반적으로 사용되고 있는 방법 중 하나로 특히 비례위험가정이 성립하는 경우에 검정력이 높은 것으로 알려져 있다. 구간중도절단을 가지는 자료의 경우에 적용될 수 있는 로그-순위 검정법으로는 Finkelstein (1986)과 Zhao와 Sun (2004) 등이 제안한 바 있고, 그 중에서 널리 사용되는 검정법인 Zhao와 Sun (2004)이 제안한 방법을 소개하기로 한다.

구간중도절단을 가지는 두 표본을 생각하자. 합동 표본으로부터 Turnbull (1976)의 자기일치법 알고리즘 등을 통해 비모수적 최우추정치인 $\hat{S}_0(t)$을 추정할 수 있다. 합동표본의 $\{0, L_i, R_i\}$에서 중복되지 않는 값을 순서대로 표기한 값을 $s_1 < \cdots < s_m$이라 하고, 계산상의 편의를 위해 $s_{m+1} > s_m$이면서 $t \leq s_{m+1}$인 모든 시점에서 $\hat{S}_0(t) = 0$를 만족하는 s_{m+1}이 존재함을 가정하기로 한다. $i = 1, \ldots, n$, $j = 1, \ldots, m+1$인 (i, j)에서 $\alpha_{ij} = I(s_j \in (L_i, R_i])$라 하고, 식 (2.1)의 검정을 위한 로그-순위 검정통계량을 유도하기 위해 $i = 1, \ldots, n$, $j = 1, \ldots, m$, $l = 1, 2$에서 다음을 정의한다.

$$\delta_i = I(R_i \leq s_m), \qquad \rho_{ij} = I(\delta_i = 0, L_i \geq s_j),$$

$$d'_j = \sum_{i=1}^{n} \delta_i \frac{\alpha_{ij}\left[\hat{S}_0(s_j-) - \hat{S}_0(s_j)\right]}{\sum_{u=1}^{m+1} \alpha_{iu}\left[\hat{S}_0(s_u-) - \hat{S}_0(s_u)\right]},$$

$$n'_j = \sum_{r=j}^{m+1} \sum_{i=1}^{n} \delta_i \frac{\alpha_{ir}\left[\hat{S}_0(s_r-) - \hat{S}_0(s_r)\right]}{\sum\limits_{u=1}^{m+1} \alpha_{iu}\left[\hat{S}_0(s_u-) - \hat{S}_0(s_u)\right]} + \sum_{i=1}^{n} \rho_{ij},$$

$$d'_{jl} = \sum_i{}' \delta_i \frac{\alpha_{ij}\left[\hat{S}_0(s_j-) - \hat{S}_0(s_j)\right]}{\sum\limits_{u=1}^{m+1} \alpha_{iu}\left[\hat{S}_0(s_u-) - \hat{S}_0(s_u)\right]},$$

$$n'_{jl} = \sum_{r=j}^{m+1} \sum_i{}' \delta_i \frac{\alpha_{ir}\left[\hat{S}_0(s_r-) - \hat{S}_0(s_r)\right]}{\sum\limits_{u=1}^{m+1} \alpha_{iu}\left[\hat{S}_0(s_u-) - \hat{S}_0(s_u)\right]} + \sum_i{}' \rho_{ij},$$

여기서, $\sum'$은 모집단 l에서의 모든 i에 대한 합이다.

$$U_l = \sum_{j=1}^{m} \left(d'_{jl} - \frac{n'_{jl} d'_j}{n'_j} \right)$$

으로 얻어진 통계량 $U = (U_1, U_2)'$를 계산할 수 있다. 여기서는 검정을 위해 필요한 U의 분포를 얻기 위해 Fay (1996)가 제안한 permutation 검정방법을 사용하였다.

2.2. Kolmogorov-Smirnov 검정법

Kolmogorov-Smirnov 검정법은 두 분포의 동일성을 검정하기 위한 비모수적 검정방법으로 Fleming 등 (1980)에 의해 우측중도절단자료가 있는 경우 두 생존함수의 동일성 검정법으로 확장된 바 있다. 검정통계량은

$$\text{KS} = \max \left| \hat{S}_1(t) - \hat{S}_2(t) \right| \tag{2.2}$$

으로 주어진다. 이 검정법은 두 생존분포 사이의 차이의 최대값만을 이용하고 있으므로 생존함수 전체에 걸친 작은 차이보다 짧은 기간 동안일지라도 큰 차이가 나는 것에 민감한 검정법이다. 구간중도절단자료를 가지는 경우에서는 Kolmogorov-Smirnov 통계량의 점근분포를 구하는 방법이 복잡하므로 여기서는 붓스트랩방법을 사용한 근사적인 검정을 시행하였다. Pan (2000b)이 제안한 검정방법은 구간중도절단을 가지는 합동표본에서 B개의 붓스트랩 표본을 추출한 후, 이로부터 계산된 B개의 통계량 KS_b와 원 자료에서 계산된 통계량 KS_0로 $\sum_{b=1}^{B} I(\text{KS}_b \geq \text{KS}_0) + 1$를 계산하고, 이 값이 $\alpha(B+1)$보다 작거나 같은 경우 귀무가설을 기각하는 방법이다.

2.3. 가중적분차를 이용한 검정법

가중적분차(IWD)를 이용한 검정법은 추정된 생존함수에 기반한 검정법으로 가중 Kaplan-Meier 검정법의 확장이다. 가중 Kaplan-Meier 검정법은 Pepe와 Fleming (1989)이 제안한 방법으로 로그-순위 검정통계량이 연구종료시점에 가까워질수록 불안정해지는 단점을 보완하기 위해 제안한 방법으로 우측중도절단자료가 있는 경우에 사용할 수 있다. 가중적분차를 이용한 검정법은 Petroni와 Wolfe (1994)가 시간이 이산형이고, 구간중도절단자료를 가지는 경우에 처음으로 적용시킨 바 있고, Fang (2002)이 일반적인 경우에 적용할 수 있는 검정법으로 확장하였다. 검정통계량은

$$\text{IWD} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \int w(u) \left[\hat{S}_1(u) - \hat{S}_2(u) \right] du \tag{2.3}$$

로 주어진다. 식 (2.3)에서 가중치 $w(u)$를 정하는 방법으로는 여러 방법이 제안되어 있지만, 여기서는 문제를 간단하게 하기 위해, 가중치가 1인 경우만 고려하기로 한다. IWD 검정통계량은 귀무가설 하에서 표본수가 커짐에 따라 점근적으로 평균이 0인 정규분포를 따르는 것으로 알려져 있지만, 그 분산을 구하는 과정이 복잡하기 때문에 여기서는 그 대안으로 붓스트랩방법을 사용한 근사적인 검정을 시행하였다. Fang (2002)이 제안한 검정방법은 구간중도절단을 가지는 합동표본에서 B개의 붓스트랩 표본을 추출한 후, 이로부터 계산된 B개의 통계량 IWD_b의 절대값이 원 자료에서 계산된 통계량 IWD_0의 절대값과 비교하여 크거나 같은 횟수가 αB보다 작은 경우 귀무가설을 기각하는 방법이다.

2.4. 단일 대치법을 사용한 검정법

대치법(imputation method)은 자료에 결측치가 있는 경우, 완전한 자료를 얻기 위해 결측을 다른 값으로 대체하는 방법이다. 구간중도절단자료를 가지는 경우에도 대치법을 응용할 수 있는데, 이 경우 구간중도절단을 가지는 자료를 우측중도절단을 가지는 자료의 형태로 바꿀 수 있기 때문에 간단하면서도 유용한 기법이라 할 수 있겠다. 대치법은 단일 대치법(single imputation)과 다중 대치법(multiple imputation)으로 나뉠 수 있으며, 단일 대치법의 종류에는 좌대치법(left imputation), 우대치법(right imputation), 평균대치법(mean imputation) 등이 있다. 여기서는 단일 대치법 중에서 대체적으로 가장 뛰어난 것으로 알려져 있는 평균대치법을 사용하였다.

평균대치법은 구간중도절단이 있는 자료 $(L_{il}, R_{il}]$대신 $R_{il} \neq \infty$인 경우, 구간의 평균인 $X_{il} = (L_{il} + R_{il})/2$를 실패가 일어난 시간으로 간주하고, $R_{il} = \infty$인 경우, $X_{il} = L_{il}$을 우측중도절단이 일어난 시간으로 간주해 분석하는 방법이다. 평균대치법은 구간중도절단을 가지는 자료를 변환하여 우측중도절단을 가지는 자료에 대해 사용할 수 있는 분석방법들을 적용할 수 있다는 장점이 있다. Yun과 Kim (2010)은 구간중도절단자료가 있는 경우의 생존함수의 추정에서 평균대치법과 비모수적 최우추정방법을 모의실험을 통해 비교하였고, 중도절단의 크기가 크지 않은 경우에는 평균대치법이 비모수적 최우추정방법보다 우수한 추정치가 될 수 있음을 보였다. 하지만 평균대치법은 Law와 Brookmeyer (1992)의 연구처럼 생존분포의 동일성 검정에서 유의수준이 보장되지 않는 경우가 있어 그 사용에 주의를 요하는 방법이다. 두 생존분포의 동일성 검정을 위해 본 연구에서는 대치법을 사용해 얻어진 자료에 대해 우측중도절단이 있는 경우에 적용할 수 있는 검정법들인 로그-순위 검정법과 가중 Kaplan-Meier 검정법, Kolmogorov-Smirnov 검정법을 고려하였다. 가중 Kaplan-Meier 검정에는 가중적분차를 이용한 방법에서 붓스트랩 방법을 이용해 분산을 추정한 것과 달리, 풀링하지 않은(unpooled) 분산을 사용하였다.

2.5. 다중 대치법을 사용한 검정법

다중 대치법은 단일 대치법의 단점을 보완한 방법으로, M번의 대치를 시행하여 얻어진 M개의 완전한 자료를 분석한 후, 그 결과를 결합하는 방법이다. Pan (2000a)은 다중 대치법을 사용하여 구간중도절단자료를 가지는 경우에서의 두 생존분포의 동일성 검정을 시행하는 방법을 제안하였다. Pan (2000a)이 제안한 Approximate Bayesian Bootstrap에 기반한 검정방법은 다음의 과정을 거친다.

- $m = 1, \ldots, M$에 대하여

  단계 1: $(L_{il}, R_{il}]$에서 부트스트랩 표본 $(L_{il}^{(m)}, R_{il}^{(m)}]$을 추출한다.

  단계 2: 각 붓스트랩 표본에서 생존함수 $\hat{S}_1^{(m)}, \hat{S}_2^{(m)}$을 추정한다.

  단계 3: i) 유한한 i번째 구간 $(L_{il}, R_{il}]$에 대해서는 $\{L_{il} < X_{il} \leq R_{il}\}$를 조건부로 하는 분포 $\hat{S}_l^{(m)}$에서 표본 X_{il}을 추출하고, $T_{il}^{(m)} = X_{il}$, $\delta_{il}^{(m)} = 1$로 한다. ii) 무한한 i번째 구간 $(L_{il}, \infty]$에 대

해서는 $T_{il}^{(m)} = L_{il}$, $\delta_{il}^{(m)} = 0$으로 한다.

단계 4: 각 $\{T_1^{(m)}, T_2^{(m)}\}$에서 검정통계량 $\hat{G}^{(m)}$과 그의 분산 $\hat{\Sigma}^{(m)}$을 계산한다.

단계 5: 다음과 같이 정의된 $\hat{G}$과 $\hat{\Sigma}$을 계산한다.

$$\hat{G} = \frac{1}{M}\sum_{m=1}^{M}\hat{G}^{(m)}, \qquad \hat{\Sigma} = \frac{1}{M}\sum_{m=1}^{M}\hat{\Sigma}^{(m)} + \left(1+\frac{1}{M}\right)\mathrm{var}\left(\hat{G}^{(1)},\ldots,\hat{G}^{(M)}\right).$$

단계 6: 귀무가설 하에서 $\hat{G} \sim N(0,\hat{\Sigma})$이고, 이를 바탕으로 p-값을 계산한다.

여기서는 다중 대치법에 의해 얻어진 우측중도절단을 가지는 자료들에 대해 로그-순위 검정법과 풀링하지 않은 분산을 사용한 가중 Kaplan-Meier 검정법을 고려하였다.

2.6. 커널 방식으로 생존분포를 추정한 후의 검정법

우측중도절단자료에서 비모수적 최우추정치인 Kaplan-Meier 추정치를 사용해 생존함수를 추정하는 경우, 실패가 발생한 시점에서만 생존율에 차이가 생기기 때문에 계단모양의 함수로 생존함수가 추정된다. 이런 문제를 보완하기 위해 제안된 방법 중 하나는 커널(kernel) 방식의 함수추정 기법을 사용해 생존분포를 부드러운 곡선의 형태로 추정하는 것이다. 우측중도절단 및 구간중도절단자료를 가지는 경우의 생존함수의 추정에 대한 몇 가지 연구에서 이 같은 방법이 제안된 바 있다. 먼저 중도절단이 없는 경우에 대한 커널 방식 추정량은 다음과 같다.

$$\hat{f}_c(t;h) = E_{\hat{F}_n}\left[\frac{1}{h}K\left(\frac{t-T_i}{h}\right)\right] \tag{2.4}$$

식 (2.4)의 K는 커널로 사용되는 함수이고, h는 띠너비(bandwidth), $\hat{F_n}$은 비모수적 최우추정치이다. Braun 등 (2005)이 제안한 방법은 구간중도절단이 있는 경우에 적용할 수 있는 커널 방식 추정량으로서 식 (2.4)를 다음과 같이 확장하였다.

$$\hat{f}_j(t;h) = E_{\hat{f}_{j-1}}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{t-T_i}{h}\right)\middle| T_i \in (L_i, R_i]\right]. \tag{2.5}$$

구간중도절단이 있는 경우에 대한 커널 방식 추정량은 초기치 $\hat{f}_0$에서 시작하여 식 (2.5)가 수렴할 때까지 반복하여 얻어진다. 커널 함수추정에서 중요하게 다루어지는 문제 중 하나는 적절한 띠너비를 선택하는 방법에 관한 것인데, cross-validation을 이용한 방법이 주로 사용된다. 여기서도 같은 방법을 사용해 띠너비를 추정해 사용하기로 한다. 커널 함수로는 정규분포를 사용하였다.

커널 방식으로 추정된 두 생존분포의 동일성 검정을 실시하기 위해 두 가지 방법을 고려하였다. 첫 번째는 앞서 설명한 Kolmogorov-Smirnov 검정이고, 두 번째는 가중적분차를 이용한 검정이다. Kolmogorov-Smirnov 검정과 가중적분차를 이용한 검정에서는 앞의 경우와 마찬가지로 붓스트랩방법을 통한 근사적인 검정을 시행하였다.

3. 모의실험의 구조 및 결과

3.1. 모의실험 구조

구간중도절단자료를 가지는 경우 두 생존분포의 동일성 검정을 위한 몇 가지 검정법들을 주어진 상황에 따라 모의실험을 통해 검정력을 비교해 보고자 한다. 고려한 검정법들은 총 10가지로 각각의 검정법들을 다음과 같이 표기하기로 하자.

1. log-rank(NPMLE): 비모수적 최우추정방법으로 생존함수 추정 후, 로그-순위 검정법
2. KS(NPMLE): 비모수적 최우추정방법으로 생존함수 추정 후, Kolmogorov-Smirnov 검정법
3. IWD(NPMLE): 비모수적 최우추정방법으로 생존함수 추정 후, 가중적분차를 이용한 검정법
4. log-rank(SI): 평균대치법 사용 후, 로그-순위 검정법
5. KS(SI): 평균대치법 사용 후, Kolmogorov-Smirnov 검정법
6. IWD(SI): 평균대치법 사용 후, 가중 Kaplan-Meier 검정법
7. log-rank(MI): 다중 대치법 사용 후, 로그-순위 검정법
8. IWD(MI): 다중 대치법 사용 후, 가중 Kaplan-Meier 검정법
9. KS(kernel): 커널 방법으로 생존함수 추정 후, Kolmogorov-Smirnov 검정법
10. IWD(kernel): 커널 방법으로 생존함수 추정 후, 가중적분차를 이용한 검정법

통계량의 비교를 위한 상황설정은 Petroni와 Wolfe (1994)를 참고하여 다음의 5가지 상황을 고려하였다. 각 상황 중, 두 생존분포가 동일한 경우와 비례위험 가정이 성립하는 경우는 각각 3가지 경우를 고려하였고, 위험함수가 교차하는 경우는 2가지 경우를 고려하였다. 비례위험 가정이 성립하는 경우의 두 집단의 위험률의 비는 각각 1.73, 2, 2.25가 되도록 설정하였다.

1. 두 생존분포가 동일한 경우(EH1, EH2, EH3).
2. 비례위험 가정이 성립하는 경우(PH1, PH2, PH3).
3. 초반에 차이가 나는 경우(early).
4. 후반에 차이가 나는 경우(late).
5. 위험함수가 교차하는 경우(CH1, CH2).

생존분포가 동일한 경우과 비례위험 가정이 성립하는 경우의 표본추출에는 와이블분포를 사용하였고, 나머지 경우는 지수분포를 사용하였다. 각 상황에 사용된 분포들은 Table 3.1에 정리하였다.

구간중도절단을 가지는 각 자료의 생성에는 다음과 같은 방법을 사용하였다.

단계 1: 특정분포로부터 생존시간 T를 생성한다.

단계 2: 주어진 상수 a에 대해 균일분포 $U(0,a)$로부터 U_1과 U_2를 생성하고 이를 바탕으로 $L = \max(0, T-U_1)$과 $R = T + U_2$를 생성한다.

단계 3: $U(0,k)$에서 중도절단시간 C를 생성한다.

단계 4: $T > C$인 경우 중도절단이 일어난 것으로 간주하고, $L = \max(0, C-U_1)$, $R = \infty$로 변환한다.

설정된 각 상황에 대해 표본의 크기는, 현실적인 가정은 아니나 계산상의 편의를 위해 집단 1과 집단 2에서 모두 동일하게 50으로 정하였다. k는 생존자료에서의 중도절단율을 결정하는 모수이고, 각 상황에서 k는 통합 표본에서의 중도절단율이 0%(NC), 40%(C)가 되도록 설정하였다. 구간중도절단 자료의 생성에서도 구간의 최대길이를 결정하는 a를 0.15인 경우와 0.4인 경우에 대해 실험하였다. Kolmogorov-Smirnov 검정과 가중적분차를 이용한 검정에서의 붓스트랩 반복은 200회로 하였고, 다중 대치법에서는 30회의 대치를 시행하였다. 모든 검정은 유의수준 0.05하에서 실시되었고, 각각 1000번씩의 반복을 거쳐 기각한 횟수로 검정력을 측정하였다.

Table 3.1. Configurations of the distributions

상황	집단1	집단2
EH1	Weibull$(0.5, 1)$	Weibull$(0.5, 1)$
EH2	Weibull$(1, 1)$	Weibull$(1, 1)$
EH3	Weibull$(2, 1)$	Weibull$(2, 1)$
PH1	Weibull$(0.5, 1)$	Weibull$(0.5, 3)$
PH2	Weibull$(1, 1)$	Weibull$(1, 2)$
PH3	Weibull$(2, 1)$	Weibull$(2, 1.5)$
early	Exp$(\lambda_1(t))$,	Exp$(\lambda_2(t))$,
	$\lambda_1(t) = 0.25I(t \leq 0.75) + 0.5I(t > 0.75)$	$\lambda_2(t) = 0.75I(t \leq 0.75) + 0.5I(t > 0.75)$
late	Exp$(\lambda_1(t))$,	Exp$(\lambda_2(t))$,
	$\lambda_1(t) = 1.0I(t \geq 0)$	$\lambda_2(t) = 1.0I(t \leq 0.5) + 2.0I(t > 0.5)$
CH1	Exp$(\lambda_1(t))$,	Exp$(\lambda_2(t))$,
	$\lambda_1(t) = 0.5I(t \leq 0.75) + 1.5I(0.75 < t \leq 1.5)$	$\lambda_2(t) = 1.5I(t \leq 0.75) + 0.5I(0.75 < t \leq 1.5)$
	$+1.0I(t > 1.5)$	$+1.0I(t > 1.5)$
CH2	Exp$(\lambda_1(t))$,	Exp$(\lambda_2(t))$,
	$\lambda_1(t) = 0.5I(t \leq 0.5) + 0.1I(0.5 < t \leq 1.25)$	$\lambda_2(t) = 1.5I(t \leq 0.5) + 0.1I(0.5 < t \leq 1.25)$
	$+1.5I(1.25 < t \leq 1.75) + 1.0I(t > 1.75)$	$+0.5I(1.25 < t \leq 1.75) + 1.0I(t > 1.75)$

Table 3.2. Sizes of the tests at the nominal level of 0.05

검정법	구간최대길이	EH1		EH2		EH3	
		NC	C	NC	C	NC	C
log-rank(NPMLE)	0.3	0.050	0.059	0.051	0.054	0.055	0.049
	0.8	0.052	0.059	0.054	0.051	0.061	0.048
KS(NPMLE)	0.3	0.024	0.014	0.042	0.016	0.032	0.026
	0.8	0.027	0.013	0.024	0.021	0.028	0.012
IWD(NPMLE)	0.3	0.051	0.068	0.067	0.060	0.056	0.056
	0.8	0.058	0.063	0.056	0.058	0.059	0.049
log-rank(SI)	0.3	0.060	0.063	0.058	0.055	0.057	0.058
	0.8	0.055	0.060	0.056	0.057	0.060	0.056
KS(SI)	0.3	0.050	0.052	0.049	0.048	0.048	0.045
	0.8	0.049	0.049	0.050	0.047	0.054	0.041
IWD(SI)	0.3	0.050	0.092	0.057	0.074	0.058	0.073
	0.8	0.052	0.093	0.059	0.070	0.057	0.070
log-rank(MI)	0.3	0.057	0.061	0.056	0.054	0.050	0.053
	0.8	0.055	0.053	0.055	0.057	0.049	0.036
IWD(MI)	0.3	0.050	0.090	0.057	0.072	0.056	0.065
	0.8	0.051	0.088	0.055	0.072	0.050	0.048
KS(kernel)	0.3	0.058	0.045	0.060	0.049	0.058	0.043
	0.8	0.041	0.046	0.053	0.045	0.046	0.034
IWD(kernel)	0.3	0.037	0.053	0.054	0.053	0.051	0.050
	0.8	0.044	0.055	0.049	0.062	0.057	0.047

3.2. 모의실험 결과

두 생존분포가 동일한 경우 (Table 3.2) 로그-순위 검정법은 비모수적 최우추정방법으로 생존함수를 추정한 경우뿐 아니라 단일 대치법 및 다중 대치법을 사용하여 검정을 실시한 경우 모두 유의수준인

Table 3.3. Powers under the proportional hazards assumption

검정법	구간최대길이	PH1		PH2		PH3	
		NC	C	NC	C	NC	C
log-rank(NPMLE)	0.3	0.712	0.536	0.893	0.734	0.968	0.840
	0.8	0.730	0.513	0.925	0.722	0.976	0.817
KS(NPMLE)	0.3	0.717	0.165	0.581	0.463	0.757	0.549
	0.8	0.117	0.105	0.340	0.230	0.508	0.322
IWD(NPMLE)	0.3	0.666	0.502	0.897	0.706	0.946	0.810
	0.8	0.667	0.462	0.911	0.667	0.934	0.723
log-rank(SI)	0.3	0.730	0.538	0.908	0.733	0.972	0.837
	0.8	0.720	0.507	0.899	0.689	0.969	0.803
KS(SI)	0.3	0.672	0.501	0.856	0.700	0.958	0.798
	0.8	0.670	0.461	0.854	0.666	0.945	0.766
IWD(SI)	0.3	0.642	0.599	0.904	0.770	0.965	0.864
	0.8	0.633	0.556	0.897	0.727	0.959	0.822
log-rank(MI)	0.3	0.723	0.527	0.903	0.731	0.969	0.840
	0.8	0.672	0.425	0.923	0.742	0.972	0.810
IWD(MI)	0.3	0.645	0.594	0.898	0.767	0.961	0.863
	0.8	0.623	0.538	0.930	0.758	0.971	0.823
KS(kernel)	0.3	0.466	0.465	0.795	0.675	0.898	0.777
	0.8	0.400	0.441	0.808	0.631	0.929	0.750
IWD(kernel)	0.3	0.455	0.513	0.868	0.723	0.945	0.845
	0.8	0.353	0.475	0.912	0.718	0.954	0.830

0.05 근처의 확률로 기각이 일어남을 확인할 수 있었다. Kolmogorov-Smirnov 검정법은 비모수적 최우추정방법으로 생존함수를 추정한 후 검정을 시행하였을 때 검정력이 유의수준보다 낮은 보수적인 결과를 보여주었다. 단일 대치법을 사용한 후 Kolmogorov-Smirnov 검정을 시행하였을 때와 커널 방법으로 생존함수를 추정한 후에 Kolmogorov-Smirnov 검정을 시행한 경우에는 유의수준 정도로 기각이 일어났다. 가중적분차를 이용한 검정법의 경우 중도절단이 일어나지 않은 상황에서는 유의수준 정도로 기각이 일어났으나, 중도절단이 일어난 상황에서는 단일 대치법과 다중 대치법을 사용한 경우에 유의수준보다 약간 높은 정도로 기각이 일어났다.

비례위험 가정이 성립하는 경우 (Table 3.3) 로그-순위 검정법을 사용하는 경우는 비모수적 최우추정방법으로 생존함수를 추정한 경우와 단일 대치법 및 다중 대치법을 사용하여 검정을 실시한 경우의 차이는 크지 않았고, 이러한 로그-순위 검정법이 비례위험 가정이 성립하는 상황으로 고려한 세 가지 경우(PH1, PH2, PH3)에서 모두 여타 검정법들에 비해 대체적으로 높은 검정력을 가지는 것을 확인 할 수 있었다. 비례위험 가정이 성립하는 경우는 두 생존분포간에 큰 차이가 나타나기 보다는 긴 시점 동안 작은 차이가 유지되고 있는 형태에 가까우므로, Kolmogorov-Smirnov 검정법은 다른 검정법에 비해서 다소 낮은 검정력을 보여주었다. 특히 비모수적 최우추정방법으로 생존함수를 추정한 후 검정법을 적용하였을 때는 구간최대길이가 긴 경우가 짧은 경우에 비해 검정력이 다소 낮은 것을 알 수 있었다. 가중적분차를 이용한 검정법은 로그-순위 검정법과 Kolmogorov-Smirnov 검정법의 중간 정도의 검정력을 가지고 있는 것으로 확인되었다. PH1에서는 커널 방법으로 생존함수를 추정한 후 가중적분차를 이용한 검정을 시행한 경우 나머지 방법을 사용하여 생존함수를 추정한 후 검정을 시행한 때보다 검정력이 현저히 낮았다.

Table 3.4. Powers in the case of early difference

검정법	구간최대길이	NC	C
log-rank(NPMLE)	0.3	0.356	0.432
	0.8	0.401	0.488
KS(NPMLE)	0.3	0.578	0.392
	0.8	0.510	0.334
IWD(NPMLE)	0.3	0.445	0.420
	0.8	0.454	0.475
log-rank(SI)	0.3	0.356	0.428
	0.8	0.351	0.411
KS(SI)	0.3	0.458	0.562
	0.8	0.461	0.549
IWD(SI)	0.3	0.444	0.446
	0.8	0.425	0.438
log-rank(MI)	0.3	0.351	0.425
	0.8	0.381	0.455
IWD(MI)	0.3	0.435	0.441
	0.8	0.450	0.476
KS(kernel)	0.3	0.598	0.483
	0.8	0.605	0.532
IWD(kernel)	0.3	0.408	0.378
	0.8	0.431	0.440

초반에 차이가 나는 경우 (Table 3.4) 초반에 차이가 나는 경우는 단일 대치법을 사용한 경우와 커널 방법으로 생존함수를 추정한 후의 Kolmogorov-Smirnov 검정법이 우수한 성능을 보여주었다. 로그-순위 검정법은 검정력이 가장 낮았고, 가중적분차를 이용한 검정은 중간 정도의 검정력을 가지는 것으로 확인할 수 있었다. 고려된 각 검정법별로 생존함수의 추정 방법에 따른 차이는 크지 않았다. 실제 사용에서는 로그-순위 검정법이나 가중적분차를 이용한 검정시에 Gehan의 검정법 등과 같이 생존함수의 앞쪽에 가중치를 두는 방법을 사용하여 검정을 시행한다면 유효한 차이를 검정하는 데 유리할 것이다.

후반에 차이가 나는 경우 (Table 3.5) 후반에 차이가 나는 경우는 대체로 모든 검정법의 검정력이 낮았다. 특히 중도절단이 있는 경우는 중도절단이 없는 경우와 비교했을 때 검정력이 큰 폭으로 하락하는 것을 확인할 수 있는데 이는 중도절단된 관측치가 다수 포함되어 있는 경우 생존함수의 뒷부분으로 갈수록 생존함수의 추정이 불안정하게 되어 생존함수 후반부에 두 분포의 차이가 있음에도 불구하고 그 차이를 검정하지 못하는 경우가 발생하기 쉽기 때문이다. Kolmogorov-Smirnov 검정법의 검정력이 다른 두 검정법에 비해 낮게 나타났고, 중도절단이 없는 경우에는 로그-순위 검정법과 가중적분차를 이용한 검정법이 비슷한 검정력을 보여주었다. 중도절단이 있는 경우에는 가중적분차를 이용한 검정법이 로그-순위 검정법에 비해 약간 높은 검정력을 가지는 것으로 보인다. 생존함수의 추정 방법에 따른 차이는 각 검정법 별로 크지는 않았으나 가중적분차를 이용한 검정법에 대해서는 커널 방법으로 추정한 후 검정한 경우가 나머지 경우에 비해 검정력이 낮았다. 반대로 Kolmogorov-Smirnov 검정법은 중도절단이 있을 때 커널 방법으로 생존함수를 추정한 후의 검정이 나머지 방법보다 우수하였다. 구간최대길이가 긴 경우에는 구간최대길이가 짧은 경우에 비해 Kolmogorov-Smirnov 검정법은 검정력이 하락하였고, 나머지 두 검정법은 검정력에 큰 차이를 보이지 않았다.

Table 3.5. Powers in the case of late difference

검정법	구간최대길이	NC	C
log-rank(NPMLE)	0.3	0.436	0.171
	0.8	0.488	0.172
KS(NPMLE)	0.3	0.142	0.111
	0.8	0.062	0.047
IWD(NPMLE)	0.3	0.457	0.198
	0.8	0.441	0.146
log-rank(SI)	0.3	0.483	0.185
	0.8	0.481	0.176
KS(SI)	0.3	0.375	0.143
	0.8	0.375	0.139
IWD(SI)	0.3	0.481	0.252
	0.8	0.494	0.220
log-rank(MI)	0.3	0.471	0.187
	0.8	0.459	0.173
IWD(MI)	0.3	0.476	0.261
	0.8	0.470	0.222
KS(kernel)	0.3	0.309	0.213
	0.8	0.305	0.172
IWD(kernel)	0.3	0.419	0.214
	0.8	0.430	0.196

Table 3.6. Powers in the case of crossing hazard rates

검정법	구간최대길이	CH1		CH2	
		NC	C	NC	C
log-rank(NPMLE)	0.3	0.091	0.136	0.689	0.799
	0.8	0.112	0.162	0.692	0.809
KS(NPMLE)	0.3	0.126	0.069	0.824	0.672
	0.8	0.095	0.064	0.691	0.565
IWD(NPMLE)	0.3	0.179	0.172	0.823	0.807
	0.8	0.185	0.195	0.831	0.806
log-rank(SI)	0.3	0.094	0.135	0.687	0.792
	0.8	0.094	0.135	0.686	0.793
KS(SI)	0.3	0.102	0.139	0.810	0.851
	0.8	0.106	0.143	0.813	0.845
IWD(SI)	0.3	0.174	0.192	0.826	0.819
	0.8	0.169	0.193	0.818	0.818
log-rank(MI)	0.3	0.092	0.132	0.687	0.794
	0.8	0.104	0.199	0.680	0.798
IWD(MI)	0.3	0.168	0.191	0.826	0.818
	0.8	0.192	0.203	0.837	0.822
KS(kernel)	0.3	0.190	0.122	0.883	0.812
	0.8	0.174	0.137	0.885	0.835
IWD(kernel)	0.3	0.150	0.149	0.802	0.785
	0.8	0.171	0.177	0.818	0.789

위험함수가 교차하는 경우 (Table 3.6) 위험함수가 작은 폭으로 교차하는 경우(CH1)에는 중도절단이 없는 상황에서는 커널 방법으로 생존함수를 추정한 후의 Kolmogorov-Smirnov 검정법이 우수하였고, 중도절단이 있는 상황에서는 가중적분차를 이용한 검정법이 나머지에 비해 우수하였다. 대체적으로 모든 검정법들의 검정력이 0.2 이하로 낮았고, 중도절단율 및 구간최대길이의 변화에 따른 검정력의 변동폭이 크지 않았다. 이 경우 역시 초반에 차이가 나는 경우와 마찬가지로 로그-순위 검정 또는 가중적분차를 이용한 검정법을 시행할 때 생존함수의 앞쪽에 가중치를 주는 방법을 사용함으로써 검정력의 향상을 가져올 수 있을 것이다. 위험함수가 큰 폭으로 교차하는 경우(CH2)에는 가중적분차를 이용한 검정법과 커널 방식으로 생존함수를 추정한 후의 Kolmogorov-Smirnov 검정법이 우수한 검정력을 보여주었다. 로그-순위 검정법은 이 경우, 다른 두 검정법에 비해 다소 낮은 검정력을 보여주었다. 가중적분차를 이용한 검정에서는 커널 방법으로 생존함수를 추정한 경우가 다른 방법으로 생존함수를 추정한 경우보다 검정력이 낮았고, 다중대치법을 사용한 후의 검정이 나머지에 비해 우수하였다. 구간최대길이에 따른 검정력의 차이는 비모수적 최우추정방법을 사용한 후의 Kolmogorov-Smirnov 검정법을 제외한 나머지 검정법 모두에서 크지 않았다. Kolmogorov-Smirnov 검정법과 가중적분차를 이용한 검정법은 중도절단이 있는 경우가 중도절단이 없는 경우에 비해 검정력이 낮게 나타났으나, 로그-순위 검정법은 위험함수가 교차하는 두 경우 모두 중도절단이 있는 경우가 그렇지 않은 경우에 비해 검정력이 높았다.

4. 결론

두 생존분포의 동일성 검정은 생존분석에서 큰 관심의 대상인 만큼 관련된 연구가 활발히 진행되어 왔다. 구간중도절단자료가 있는 경우에도 두 생존분포의 동일성을 검정하기 위해 제안된 몇 가지 검정통계량들이 있는데, 각 검정법들은 비모수적 최우추정법으로 생존함수를 추정한 경우뿐 아니라 단일 대치법, 다중 대치법, 커널 방식의 추정법등을 사용하여 생존함수를 추정한 경우에도 적용시킬 수 있다. 본 논문에서는 생존함수를 추정하는 여러 가지 방식에 대해 로그-순위 검정법, Kolmogorov-Smirnov 검정법, 가중적분차를 이용한 검정법을 적용하여 보았고, 두 그룹의 생존분포의 형태 및 중도절단율, 중도절단구간의 길이를 달리한 몇 가지 상황을 설정하여 각각의 상황 하에서 모의실험을 통해 각 검정법들의 검정력을 비교하여 보았다.

그 결과, 비례위험의 가정이 성립하는 경우에는 로그-순위 검정법이 대체로 우수한 검정력을 보여주었고, 초반에 차이가 나는 경우와 큰 폭으로 위험함수가 교차하는 경우에는 Kolmogorov-Smirnov 검정법이, 후반에 차이가 나는 경우와 작은 폭으로 위험함수가 교차하는 경우에는 가중적분차를 이용한 검정법이 각각 높은 검정력을 가지는 것으로 확인되었다. 덧붙여 로그-순위 검정법은 다른 검정법들에 비해 중도절단율이 커짐에 따라 검정력의 하락폭이 상대적으로 크지 않음을 알 수 있었고, 구간최대길이의 변화에도 마찬가지로 검정력이 크게 변하지 않았다. 단일 대치법 또는 다중 대치법을 사용하여 구간중도절단이 있는 자료를 우측중도절단이 있는 자료로 변환한 후 검정을 시행했을 때는 많은 경우에 비모수적 최우추정법을 사용해 생존함수를 추정한 경우와 비교해 검정력에 큰 차이가 없는 것으로 보이므로 붓스트랩 등의 복잡한 계산 없이도 검정을 수행할 수 있다고 할 수 있다. 커널 방식으로 사용해 생존함수를 추정한 경우는 Kolmogorov-Smirnov 검정법의 사용에서 비모수적 최우추정법으로 생존함수를 추정한 경우에 비해 고려한 대부분의 상황에서 현저히 좋은 검정력을 보여주었고, 중도절단구간의 최대길이나 중도절단율의 변화에 대해서도 비모수적 최우추정법으로 추정한 후 검정법을 적용시킨 경우보다 검정력이 덜 민감하게 변하였다.

모의실험 결과로 보아 각 검정법들은 검정통계량이 가지는 특성에 따라 어떤 상황에서는 높은 검정력을 보여주지만 다른 상황에서는 두 생존분포에 차이가 있음에도 불구하고 차이를 검정하는 데 실패하기도

한다. 본 연구는 모의실험을 통해 주어진 상황에서 적절한 검정법을 선택하는데 유용한 정보를 제공하였다. 이러한 정보는 의학연구 및 역학연구등에서 빈번하게 사용되는 생존분포의 비교에 도움이 될 것이다.

References

Braun, J., Duchesne, T. and Stafford, J. E. (2005). Local likelihood density estimation for interval censored data, *Canadian Journal of Statistics*, **33**, 39–60.

Fang, H. (2002). Nonparametric survival comparison for interval-censored continuous data, *Statistica Sinica*, **12**, 1073–1083.

Fay, M. P. (1996). Rank invariant tests for interval censored data under the grouped continuous model, *Biometrics*, **52**, 811–822.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data, *Biometrics*, **42**, 845–854.

Fleming, T. R., O'Fallon, J. R. and O'Brien, P. C. (1980). Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right-censored data, *Biometrics*, **36**, 607–625.

Gentleman, R. and Geyer, C. J. (1994). Maximum likelihood for interval censored data: Consistency and computation, *Biometrika*, **81**, 618–623.

Jung, M. N. and Lee, J. W. (1998). A comparison of the statistical methods for testing the equality of two survival distributions, *The Korean Journal of Applied Statistics*, **11**, 113–127.

Law, C. G. and Brookmeyer, R. (1992). Effects of mid-point imputation on the analysis of doubly censored data, *Statistics in Medicine*, **11**, 1569–1578.

Pan, W. (2000a). A two-sample test with interval censored data via multiple imputation, *Statistics in Medicine*, **19**, 1–11.

Pan, W. (2000b). Smooth estimation of the survival function for interval censored data, *Statistics in Medicine*, **19**, 2611–2624.

Pepe, M. S. and Fleming, T. R. (1989). Weighted Kaplan-Meier statistics: A class of distance tests for censored survival data, *Biometrics*, **45**, 497–507.

Petroni, G. and Wolfe, R. (1994). A two-sample test for stochastic ordering with interval-censored data, *Biometrics*, **50**, 77–87.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data, *Journal of the Royal Statistical Society, Series B*, **38**, 290–295.

Yun, E. and Kim, C. (2010). Estimation of survival function and median survival time in interval-censored data, *The Korean Journal of Applied Statistics*, **23**, 521–531.

Zhao, Q. and Sun, J. (2004). Generalized log-rank test for mixed interval-censored failure time data, *Statistics in Medicine*, **23**, 1621–1629.