

교차하는 두 생존함수의 동일성 검정법에 관한 비교연구[†]

이윤주¹ · 이재원²

¹²고려대학교 통계학과

접수 2015년 1월 27일, 수정 2015년 2월 16일, 게재확정 2015년 4월 6일

요약

두 집단의 생존분포의 비교를 위한 방법으로는 로그-순위 검정법이 널리 쓰이고 있다. 그러나 이는 두 집단의 생존분포 간에 비례위험모형 가정이 성립하는 경우에 적합한 것으로써, 두 집단의 생존분포가 교차하는 상황 하에서는 해당 검정법의 유효성을 장담할 수 없다. 그러나 두 집단의 생존분포가 교차하는 상황은 빈번히 발생하며, 이 같은 상황에서 두 생존분포의 동일성 검정을 위한 여러 연구가 진행되어왔다. 본 논문에서는 두 집단 간 생존분포가 교차하는 상황 하에서의 동일성 검정을 위한 방법들을 고려한다. 나아가 교차하는 상황을 위치에 따라 세분화하고 모의실험을 통해 각 상황에서 다양한 검정법들의 검정력을 비교하였으며, 그 결과를 토대로 주어진 상황에서 적절한 방법의 선택에 유용한 정보를 제공하고자 한다.

주요용어: 교차 생존 분포, 동일성 검정, 로그-순위 검정, 생존분포.

1. 서론

중도절단이 있는 자료에서 두 집단 간 생존분포의 비교는 생존분석에서의 주된 관심사 중 하나이다. 생존분포의 비교를 위해 많은 방법이 고안되어왔으며, 두 집단의 생존분포 비교 시 가장 널리 쓰이는 방법으로는 로그-순위 검정법이 있다 (C. A. Kitabo와 Kim, 2014). 로그-순위 검정법은 비례위험 모형 가정이 성립하는 경우에 검정력이 높은 방법으로, 두 생존분포가 서로 교차하는 경우에는 적합하지 않다. 그러나 실제 두 생존분포가 서로 교차하는 상황은 임상 및 의학 연구를 비롯하여 흔히 관측된다. 가령 암 수술 등 침습적 시술이 필요한 분야를 들 수 있는데, 치료 후 일정시간을 경과한 환자는 처치 직후에 비해 생존 가능성이 높기 때문에 (Lee 등, 2011) 처치 여부에 따른 각 그룹의 생존분포는 교차할 가능성이 있다. 이렇듯 흔히 발생하는 교차하는 두 생존 분포의 동일성 검정을 위해 다양한 검정방법들이 논의 되어왔다.

Gill (1980)은 Renyi type 통계량을 제안하였으며, 이후 Davis와 Xie (2011)는 여러 형태의 가중치들을 이용하여 가중 로그-순위 검정법과 Renyi type 통계량의 비교하고, Renyi type 검정법이 가중 로그-순위 검정보다 검정력이 높음을 보였다. 또한 Schumacher (1984)가 제안한 수정된 Kolmogorov-Smirnov 검정법은 두 경험적 분포의 차이 중 최댓값을 이용한 기존의 Kolmogorov-Smirnov 검정법을 중도절단 자료에 확장하여 적용한 것으로, 두 생존분포 간 차이의 최댓값에 근거한 방법이다. 이와 함께 Schumacher (1984)는 기존의 Cramer-Von Mises 검정법을 두 집단 간 생존분포의 동일성 검정으로 확

[†] 이 논문은 2012년도 정부 (교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업 임 (No. NRF-2012R1A1A2008686).

¹ (136-701) 서울 특별시 성북구 안암로 145, 고려대학교 통계학과, 석사.

² 교신저자: (136-701) 서울 특별시 성북구 안암로 145, 고려대학교 통계학과, 교수.

E-mail: jael@korea.ac.kr

대하여 적용한 바 있다. 이는 두 생존분포 간 차이의 적분제곱에 기반을 둔 방법으로, 두 집단의 생존분포가 교차하는 상황에서 기존의 로그-순위 통계량보다 높은 검정력을 가지는 것으로 알려져 있다. 다음으로 Lee (1996)는 가중 로그-순위 통계량에 대하여 두 생존분포 간 차이의 여러 형태를 반영하는 가중치가 다른 통계량들을 고려하여, 그것들의 최대통계량 (maximum statistic)과 선형결합통계량 (linear combination statistic)에 의한 검정법을 제안하였다. Klein과 Moeschberger (2003)은 그의 저서에서 Cramer-Von Mises 검정법과 함께 가중 카플란-마이어 검정법 (weighted Kaplan-Meier test)과 중위수 검정법 (Median test) 역시 교차하는 두 생존분포의 동질성 검정에 유용하게 사용될 수 있다고 하였으나, Jeong과 Lee (1998)의 연구로부터 두 생존분포가 교차하는 경우 가중 카플란-마이어 검정법과 중위수 검정법은 교차하는 상황 하에서 낮은 검정력을 보였으며, Lee 검정법의 검정력이 좋았음을 확인하였다. Qui와 Sheng (2008)은 교차하는 상황에서 적합한 가중 로그-순위 검정법을 고안하였으며, 나아가 첫 번째 단계에서 로그-순위 검정과 같은 보편적인 방법을 적용하고, 두 번째 단계에서 새로이 고안된 검정법을 적용하는 이단계 (two-stage) 검정법을 제안하였다. 마지막으로 Logan 등 (2008)은 두 생존분포가 교차하는 상황 하에서 분할분석 (split-analysis)를 통해 다양한 검정통계량들을 비교한 바 있다.

본 논문에서는 교차하는 두 집단의 생존분포 비교 시 사용할 수 있는 몇 가지 검정법을 소개하고, 이를 여러 가지 상황 하에서 모의실험을 실시하여 각 검정법들의 검정력을 비교하고자 한다. 2절에서는 각 검정법에 대하여 살펴보고, 3절에서는 모의실험 결과를 정리하였다.

2. 두 생존분포에 대한 동일성 검정법

생존분포의 동일성 검정법 소개에 앞서 비교를 위한 두 표본을 생각하자. 관측치는 i 번째 표본에서 n_i 개의 개체들로부터 얻어지며 ($i = 1, 2$), 임의중도절단 (random censoring)이 발생한다. 이때 (T_{ij}, C_{ij}) 를 각각 i 번째 표본에서의 j 번째 개체로부터 얻어진 실패시간 (failure time)과 중도절단시간 (censoring time)이라고 한다면 ($i = 1, 2; j = 1, \dots, n_i$), 실제 관측되는 것은 (X_{ij}, δ_{ij}) ($i = 1, 2; j = 1, \dots, n_i$)로서, 생존시간 X_{ij} 와 중도절단 지시함수 δ_{ij} 이다. 여기에서 생존시간 $X_{ij} = \min(T_{ij}, C_{ij})$ 이고, δ_{ij} 는 중도절단인 경우 0, 중도절단 되지 않은 경우 1을 갖는다. 이에 근거하여 i 번째 표본의 생존함수 (survival function)는 $S_i(t) = P(X_{ij} > t)$ 으로 정의된다. 우리가 검정하고자 하는 것은 두 생존분포의 동일성이므로, 귀무가설은

$$H_0 : S_1(t) = S_2(t) \quad (2.1)$$

이다. 두 생존분포가 교차하는 경우 로그-순위 검정법은 검정력이 낮은 것으로 잘 알려져 있다. 대안으로는 Gill (1980)이 제안한 Renyi type 검정법을 적용할 수 있으며, 두 생존분포 간의 차이에 기반을 둔 검정법과 가중 로그-순위 검정법의 형태 및 분할분석을 통해 로그-순위 검정 보다 유효한 결과를 얻을 수 있다.

이 가운데 분할분석은 전체 생존분포가 아닌 교차할 것으로 미리 예상 가능한 시점에 대해 해당시점 이후의 생존분포 간 후반부 차이를 대상으로 동일성을 검정하는 방법이므로, 교차하는 시점에 대한 사전 정보가 요구된다 (Bouliotis와 Billingham, 2011). 이에 본 논문에서는 분할분석은 논의에서 제외하고, 전체 생존분포를 대상으로 하는 검정법을 소개하고자 한다.

2.1. Renyi type 검정법

앞서 언급하였듯이, Renyi type 검정법은 두 생존분포가 교차하는 상황에서 유용한 방법으로 잘 알려져 있다 (Gill, 1980). 먼저 표본크기가 각각 n_1, n_2 ($N = n_1 + n_2$)인 두 집단을 고려해보자. 이때 합동표본으로부터의 사건 발생 시간 t_k 에 대하여 $t_1 < t_2 < \dots < t_D$ ($k = 1, \dots, D$)일 때, d_{ik} 와 Y_{ik} 를

각각 i 번째 집단에서 t_k 번째 시간에서의 사망자 수, 위험그룹 인원수라고 하자. $d_k = d_{1k} + d_{2k}$, $Y_k = Y_{1k} + Y_{2k}$ 로 정의하고, W_{1k} 는 가중 함수으로써 로그-순위 검정의 경우 $W_{1k} = 1$, 일반화된 Wilcoxon 검정의 경우 $W_{1k} = Y_{1k} + Y_{2k}$ 을 갖는다. 이외에 가중 로그-순위 검정에 적용 가능한 다양한 가중함수들을 Renyi type 통계량에 동일하게 적용시킬 수 있다.

다음의 통계량

$$Z(t) = \sum_{k=1}^D W_{1k} \left[d_{1k} - Y_{1k} \left(\frac{d_k}{Y_k} \right) \right], \quad k = 1, \dots, D$$

과 이 통계량의 표준 오차(standard error)인 $\sigma(\tau)$ 를 고려하자.

$$\sigma^2(\tau) = \sum_{k=1}^D W_{1k}^2 \left(\frac{Y_{1k}}{Y_k} \right) \left(\frac{Y_{2k}}{Y_k} \right) \left(\frac{Y_k - d_k}{Y_k - 1} \right) (d_k)$$

이때 τ 는 $Y_{1k}, Y_{2k} > 0$ 일 때의 가장 큰 t_k 값이다. 양측 대립 가설(two-sided alternative)를 위한 Renyi type 검정법의 통계량은

$$Q = \sup\{|Z(t), t \leq \tau\} / \sigma(\tau) \tag{2.2}$$

이고, 귀무가설이 참일 때 Q 는 점근적으로 $\sup\{|B(x)|, 0 \leq x \leq 1\}$ 을 따르며, 이때 B 는 표준 브라운 모션 프로세스(standard Brownian motion process)이다.

2.2. 수정된 Kolmogorov-Smirnov 검정법

수정된 Kolmogorov-Smirnov 검정법은 본래 두 표본의 경험적 분포함수에 적용 가능한 Kolmogorov-Smirnov 검정법을 중도절단이 있는 자료에 적용 가능하도록 고안한 것이다 (Schumacher, 1984). 이 방법은 두 위험함수 간 차이의 최대값을 이용한 것으로써, 생존분포 전체에 걸쳐 발생하는 작은 차이보다는 최소 한 시점에서 두 생존분포 간 큰 차이가 있는 경우에 민감한 검정법이다.

합동표본으로부터 중도절단을 나타내는 지시함수를 δ_{ik} , 시간 t 에서 i 번째 표본에서의 위험그룹인원수를 $Y_i(t)$ 라고 할 때, 관측된 시점 X_{ik} 에 대하여 i 번째 표본의 생존함수 추정량은

$$\widehat{S}_i(t) = \prod_{k: X_{ik} \leq t} \left(\frac{Y_i(X_{ik}) - 1}{Y_i(X_{ik})} \right)^{\delta_{ik}}$$

이고, Nelson (1969)의 누적위험함수 (cumulative hazard function) 추정량은

$$\widehat{\Lambda}_i(t) = \sum_{k: X_{ik} \leq t} \frac{\delta_{ik}}{Y_i(X_{ik})}$$

와 같이 주어진다. 누적위험함수 간의 차이는 $\xi(t) = \Lambda_2(t) - \Lambda_1(t)$ 로서, 그 추정량은 곧 $\widehat{\xi}(t) = \widehat{\Lambda}_2(t) - \widehat{\Lambda}_1(t)$ 이다. Breslow와 Crowley (1974)는 $N^{1/2}(\widehat{\xi}(t) - \xi(t))$ 가 $[0, \tau]$ 에서 평균이 0인 가우스 과정 (gaussian process), $Z_\xi(t)$ 으로 약 수렴함을 증명하였다. 이때 $cov(Z_\xi(s), Z_\xi(t)) = A(s) (0 \leq s \leq t \leq \tau)$ 이고, $\widehat{\psi}_\xi(t) = 1/(\widehat{A}(\tau))^{1/2}$ 이라고 하자. 이로부터 두 표본의 동일성 검정을 위한 수정된 Kolmogorov-Smirnov 통계량이 유도되며, 다음과 같다.

$$Q_{KS\xi} = N^{1/2} \sup_{0 \leq t \leq \tau} \left| \widehat{\xi}(t) \widehat{\psi}_\xi(t) \right| \tag{2.3}$$

귀무가설이 참인 경우 $Q_{KS\xi}$ 는 근사적으로 $\sup\{|W(x)|, 0 \leq x \leq 1\}$ 을 따르며, 이때 W 는 표준 브라운 모션 프로세스를 의미한다 (Aalen, 1976, 1978).

2.3. 수정된 Cramer-Von Mises 검정법

수정된 Cramer-Von Mises 검정법은 기존의 Cramer-Von Mises 검정법을 중도절단이 있는 생존 자료에 적용 가능하도록 수정한 것이다 (Schumacher, 1984). 이 방법은 앞서 살펴본 수정된 Kolmogorov-Smirnov 검정법과 마찬가지로 두 표본의 위험함수 차이에 근거한 방법이다. 모든 기호는 수정된 Kolmogorov-Smirnov 검정법에서 사용한 기호와 동일하며, 수정된 Cramer-Von Mises 통계량은 다음과 같이 주어진다.

$$Q_{CM\xi} = (N/\hat{A}(\tau)) \int_0^\tau (\hat{\xi}(t)\hat{\psi}_\xi(t))^2 d\hat{A}(t) \quad (2.4)$$

이는 두 생존분포 간 차이의 적분제곱에 기반을 둔 방법으로써, 귀무가설 하에서 $Q_{CM\xi}$ 는 $\int_0^1 [W(x)]^2 dx$ 를 따르며, 여기서 W 는 역시 표준 브라운 모션 프로세스를 의미한다.

2.4. Lee 검정법

Lee (1996)가 제안한 이 검정법은 두 생존분포가 어떠한 형태를 가지는가에 대한 사전정보가 없는 경우에 두 생존분포의 동일성 검정을 위한 유용한 방법으로, 다양한 형태의 생존분포 간 차이를 비교적 잘 찾아낼 수 있다.

이제 관측된 시점 X_{ij} 와 중도절단 지시함수 δ_{ij} 에 대하여 t 시점까지의 사망자 수를 $N_{ik}(t) = I(X_{ik} \leq t, \delta_{ik} = 1)$, t 시점에서 생존해 있는 사람의 수를 $Y_{ik}(t) = I(X_{ik} \geq t)$ 라고 하자. 또한 각 그룹 및 전체 사망자 수와 생존해 있는 사람 수에 대해 $N_i = \sum_k N_{ik}, N = \sum_i N_i, Y_i = \sum_k Y_{ik}, Y = \sum_i Y_i$ 로 나타내도록 하자. 이제 두 생존분포의 동일성을 검정하기 위한 가중 로그-순위 검정 통계량은 다음과 같이 표현할 수 있다.

$$W_K = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \int_0^\infty W(t) \frac{Y_1(t)Y_2(t)}{Y_1(t) + Y_2(t)} \left\{ \frac{dN_1(t)}{Y_1(t)} - \frac{dN_2(t)}{Y_2(t)} \right\}$$

위 식에서 $W(t)$ 는 가중함수으로써

$$W(t) = \left\{ \hat{S}(t-) \right\}^\rho \left\{ 1 - \hat{S}(t-) \right\}^\gamma, \quad (\gamma \geq 0, \rho \geq 0)$$

으로 정의되며, 이를 $G^{\rho,\gamma}$ 통계량이라고 부른다 (Flemming과 Harrington, 1991). Lee (1996)는 이로부터 아래의 4가지 통계량을 고려하였다.

(ρ, γ)	$W(t)$	Distribution of weights	$G^{\rho,\gamma}$
(0, 0)	1(Log-rank)	evenly distributed	W_{K1}
(2, 0)	$\left\{ \hat{S}(t-) \right\}^2$	emphasize early difference	W_{K2}
(0, 2)	$\left\{ 1 - \hat{S}(t-) \right\}^2$	emphasize late difference	W_{K3}
(2, 2)	$\left\{ \hat{S}(t-) \right\}^2 \left\{ 1 - \hat{S}(t-) \right\}^2$	emphasize middle difference	W_{K4}

귀무가설 하에서 $(W_{K1}, W_{K2}, W_{K3}, W_{K4}) = (G^{0,0}, G^{2,0}, G^{0,2}, G^{2,2})$ 는 근사적으로 평균이 0인 다변량 정규분포를 따른다. Lee (1996)는 이 4가지 통계량의 표준화된 통계량 $Z_l^* = W_{Kl} / \sqrt{\widehat{Var}(W_{Kl})}$, ($l = 1, \dots, 4$)를 사용하여 두 가지 검정법을 제안하였다. 그 중 LEE1은 Z_l^* 들의 최대통계량 (maximum of the standardized statistics)으로써

$$LEE1 = \max(Z_1^*, Z_2^*, Z_3^*, Z_4^*) \quad (2.5)$$

이며, LEE2는 Z_l^* 들의 선형결합통계량 (linear combination of the standardized statistics)으로써

$$LEE2 = \sum_{l=1}^4 b_l Z_l^*, \quad (\text{이 때, } \sum_{l=1}^4 b_l = 1) \tag{2.6}$$

이다. 식 (2.5)는 귀무가설 하에서 (Z_1, Z_2, Z_3, Z_4) 가 평균이 0인 다변량 정규분포를 따를 경우 근사적으로 $\max(Z_1, Z_2, Z_3, Z_4)$ 와 같은 분포를 갖는다. 또한 식 (2.6)은 근사적으로 표준정규분포 (standard normal distribution)를 따른다.

2.5. Qui-Sheng 검정법

이 절에서는 Qui-Sheng 검정법들을 소개하고자 한다. Qui와 Sheng (2008)은 가중 로그-순위 검정 형태에 대하여 교차하는 상황 하에서 유용하게 적용 가능한 가중 함수를 제안하였다.

먼저 합동표본으로부터의 사건 발생 시간 $t_1 < t_2 < \dots < t_D$ 를 고려하자. 또한 \hat{L}_1, \hat{L}_2 와 \hat{S} 는 각각 첫 번째 그룹에서의 중도절단, 두 번째 그룹에서의 중도절단과 생존시간에 대한 생존함수의 카플란-마이어 추정량 (Kaplan-Meier estimator)이다.

이제 $0 < \epsilon < 0.5$ 의 작은 수를 고려하자. 또한 $D_\epsilon = [D \cdot \epsilon]$ 와 어떤 $r \in [\epsilon, 1 - \epsilon]$ 에 대하여 $m = [D \cdot r]$ 로 정의하자. 그러면 잠재적인 교차지점 $D_\epsilon \leq m \leq D - D_\epsilon$ 에서 새로운 가중 함수를 다음과 같이 정의한다.

$$w_k^{(m)} = \begin{cases} -1, & k = 1, 2, \dots, m, \\ c_m, & \text{otherwise} \end{cases}$$

여기서 c_m 은 양수이며, 이에 대한 추정량은 아래와 같이 정의된다.

$$\hat{c}_m = \frac{\sum_{k=1}^m \frac{\hat{L}_1(t_k)\hat{L}_2(t_k)}{(n_1/n)\hat{L}_1(t_k) + (n_2/n)\hat{L}_2(t_k)} \cdot \Delta\hat{S}(t_k)}{\sum_{k=m+1}^D \frac{\hat{L}_1(t_k)\hat{L}_2(t_k)}{(n_1/n)\hat{L}_1(t_k) + (n_2/n)\hat{L}_2(t_k)} \cdot \Delta\hat{S}(t_k)}$$

Qui-Sheng 검정 통계량은

$$v = \sup_{D_\epsilon \leq m \leq D - D_\epsilon} V_m \tag{2.7}$$

으로 정의되며, 식 (2.7)의 V_m 은 가중 함수 $w_k^{(m)}$ 의 추정량을 이용하여 아래와 같이 정의된다.

$$V_m = \frac{\sum_{k=1}^D \hat{w}_k^{(m)} \left(d_{1k} - Y_{1k} \frac{d_k}{Y_k} \right)}{\sqrt{\sum_{k=1}^D \left(\hat{w}_k^{(m)} \right)^2 \frac{Y_{1k} Y_{2k} Y_k - d_k}{Y_k Y_k Y_k - 1} d_k}}$$

이는 잠재적인 교차지점 전후로 두 생존분포 간 차이의 부호를 달리할 뿐만 아니라, 서로 다른 가중치를 적용하는 방법이다.

나아가 Qui-Sheng 이단계 (two-stage) 검정법을 적용함에 있어서도 첫 단계에서 시행되는 로그-순위 검정법과 귀무가설 하에서 점근적으로 독립인 특성을 갖는다. 이러한 특성으로부터 이단계 검정법의 적용에 있어 첫 번째와 두 번째 단계의 유의수준을 각각 α_1, α_2 이라고 할 때, 이단계 검정법의 유의수준은 $\alpha_1 + P_{H0}(\text{reject in stage two} \mid \text{fail to reject in stage one})(1 - \alpha_1) = \alpha$ 으로 표현될 수 있으며, 각 두 단계의 검정법은 귀무가설 하에서 독립이므로 유의수준은 점근적으로 $\alpha_1 + \alpha_2(1 - \alpha_1) = \alpha$ 이다. 더불어 p 값은 첫 번째 단계에서 $p_1 \leq \alpha_1$ 일때 p_1 을 가지며, 그렇지 않은 경우에는 $\alpha_1 + p_2(1 - \alpha_1)$ 의 값을 가진다.

3. 모의실험

3.1. 모의실험 구조

앞에서 언급한 검정법들과 더불어 두 생존분포의 동일성 검정에 일반적으로 사용되고 있는 로그-순위 검정법 및 일반화된 Wilcoxon 검정법을 모의실험에서 비교하고자 한다. 모의실험에 사용된 통계량은 전체 생존분포에 대해 적용하는 검정법인 Renyi-type 검정통계량, 수정된 Kolmogorov-Smirnov 검정법, Cramer-Von Mises 검정법, Lee (1996)의 최대통계량 및 선형통계량, Qui와 Sheng (2008)의 검정법과 이단계 검정법이다. 2.1절에서 언급한 분할분석은 전체 생존분포가 아닌 특정 시점 이후의 구간만을 고려하므로 비교 대상에서 제외하였고, Renyi type 검정통계량의 경우 로그-순위와 일반화된 Wilcoxon 검정법 버전을 고려하였다. 비교에 사용된 각각의 검정법을 다음과 같이 표시하고자 한다.

비교에 사용된 각각의 검정법을 다음과 같이 표시하고자 한다.

1. LR : 로그-순위 검정법
2. GH : 일반화된 Wilcoxon 검정법
3. Renyi_LR : Renyi type 로그-순위 검정법
4. Renyi_GH : Renyi type 일반화된 Wilcoxon 검정법
5. MKS : 수정된 Kolmogorov-Smirnov 검정법
6. CVM : Cramer-Von Mises 검정법
7. LEE1 : Lee (1996)의 최대통계량
8. LEE2 : Lee (1996)의 선형결합통계량
9. QS1 : Qui와 Sheng (2008)의 검정법
10. QS2 : Qui와 Sheng (2008)의 이단계 검정법

통계량의 비교를 위하여 다음의 5가지 상황을 설정하여 자료를 생성하였다.

- I. 두 생존분포가 동일한 경우.
- II. 비례위험모형 가정이 성립하는 경우.
- III. 초반에 교차하는 경우.
- IV. 중반에 교차하는 경우.
- V. 후반에 교차하는 경우.

각 상황들에 대한 두 집단의 분포를 Table 3.1와 Figure 3.1에 정리하였다. Table 3.1의 5가지 상황의 자료에 있어 두 집단의 표본크기는 각각 100으로 정하였으며, 중도절단율은 각 상황에서 0%, 25%와 50%로 설정하였다. LEE2 선형결합통계량에서 Z_i^* 들의 계수로는 $b_1 = b_2 = b_3 = b_4 = 1/4$ 를 사용하였고, Qui와 Sheng (2008)의 검정법에서 $\epsilon = 0.1$ 로 하였다. 모든 검정은 유의수준이 0.05와 0.01인 경우에 대해 시행되었으며, 1000번의 반복을 거쳐 기각한 횟수로 검정력을 측정하였다.

Table 3.1 Survival distribution for each group

	group 1	group 2
I. Equal	Weibull(1,1)	Weibull(1,1)
II. Under the proportional hazard assumption	Weibull(1,1)	Weibull(2,1)
III. Crossing at early times	Weibull(1,1)	Weibull(1,0.5)
IV. Crossing at middle times	Weibull(1,1)	Weibull(0.6,0.5)
V. Crossing at late times	Weibull(1,1)	Weibull(0.5,0.6)

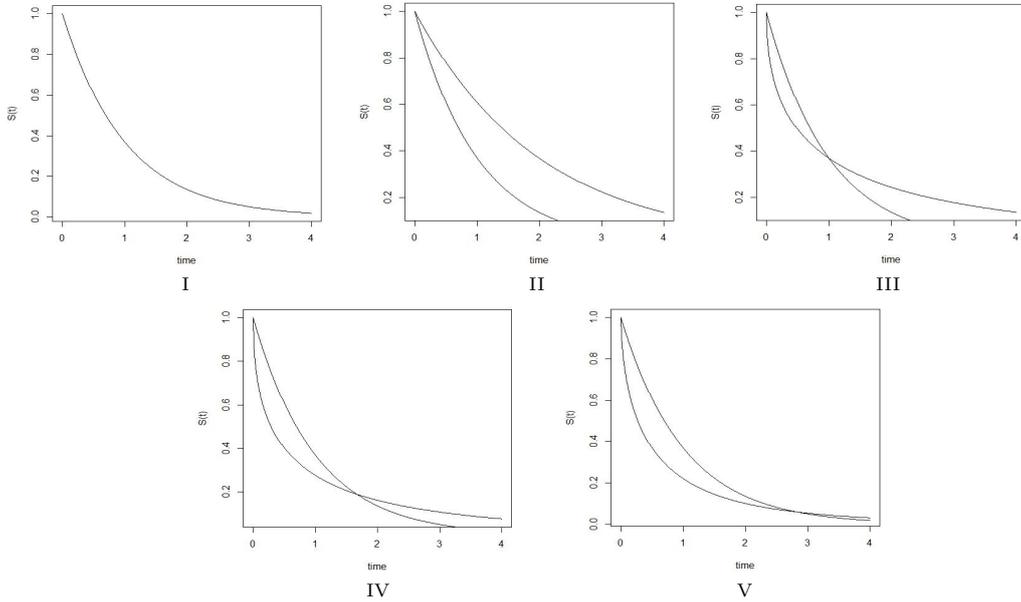


Figure 3.1 Survival distribution for each group; I. Equal, II. Under the proportional hazard assumption, III. Crossing at early times, IV. Crossing at middle times, V. Crossing at late times

3.2. 모의실험 결과

3.2.1. 두 생존분포가 동일한 경우

Table 3.2는 유의수준의 달성정도를 나타내는 결과이다. 유의수준 0.05와 0.01에 대해 95% 신뢰구간은 각각 [0.036,0.064]와 [0.004, 0.016]으로써, 대부분의 검정법에서 이 범위를 크게 벗어나지 않음을 알 수 있다. 로그-순위 검정법과 로그-순위 버전의 Renyi type 검정법, Lee의 검정법 및 Qui-Sheng 이단계 검정법은 중도절단이 있는 경우와 비교하여 중도절단이 없는 경우 그 값이 다소 오르는 경향을 보였고, Cramer-Von Mises 검정법은 이와 반대로 값이 다소 떨어지는 경향을 보였다. 그러나 이들 검정법들은 두 상황 모두에서 유의수준 근처의 확률로 기각하고 있다.

Table 3.2 Sizes of test statistics

Test statistics	$\alpha=0.05$			$\alpha=0.01$		
	censoring rate			censoring rate		
	0%	25%	50%	0%	25%	50%
LR	0.041	0.064	0.051	0.006	0.019	0.015
GH	0.044	0.060	0.043	0.008	0.016	0.010
Renyi_LR	0.033	0.056	0.045	0.004	0.013	0.010
Renyi_GH	0.050	0.057	0.046	0.011	0.006	0.010
MKS	0.037	0.037	0.044	0.006	0.010	0.004
CVM	0.041	0.037	0.033	0.006	0.010	0.005
LEE1	0.047	0.058	0.058	0.009	0.014	0.011
LEE2	0.041	0.062	0.058	0.003	0.012	0.010
QS1	0.043	0.035	0.044	0.005	0.004	0.007
QS2	0.037	0.041	0.059	0.005	0.011	0.015

3.2.2. 비례위험모형 가정이 성립하는 경우

Table 3.3의 비례위험모형 가정이 성립하는 경우에는 모든 검정법의 검정력이 높았다. 특히 로그-순위 검정법, 로그-순위 버전의 Renyi type 검정법을 비롯하여 수정된 Kolmogorov-Smirnov 검정

법과 Lee의 두 가지 검정법, Qui-Sheng 이단계 검정법에서 검정력이 매우 높게 나타났다. 일반화된 Wilcoxon 검정법, 일반화된 Wilcoxon 버전의 Renyi type 검정법과 수정된 Cramer-Von Mises 검정법의 검정력은 앞선 검정법들에 비해 낮지만, 역시 상당한 값을 보였다. 또한 Renyi type의 검정통계량은 본래 가중 로그-순위 검정법에 근거한 검정법들에 비해 약간 낮은 검정력을 보였다. 주목할 만한 것은 Qui-sheng 검정법으로써, 이는 본래 두 생존분포가 교차하는 상황에 착안하여 고안된 방법이기 때문에 위험모형 가정이 성립하는 경우 다른 검정법들에 비해 현격히 낮은 검정력을 보였다.

Table 3.3 Power of test statistics under the proportional hazard assumption

Test statistics	$\alpha=0.05$			$\alpha=0.01$		
	censoring rate			censoring rate		
	0%	25%	50%	0%	25%	50%
LR	0.999	0.994	0.894	0.985	0.946	0.757
GH	0.985	0.948	0.878	0.950	0.862	0.707
Renyi_LR	0.994	0.971	0.882	0.985	0.905	0.725
Renyi_GH	0.977	0.935	0.823	0.922	0.838	0.650
MKS	0.998	0.981	0.870	0.971	0.917	0.664
CVM	0.987	0.954	0.799	0.943	0.837	0.565
LEE1	0.998	0.980	0.869	0.975	0.910	0.717
LEE2	0.996	0.980	0.901	0.973	0.898	0.713
QS1	0.153	0.135	0.091	0.046	0.052	0.025
QS2	0.998	0.972	0.850	0.974	0.912	0.688

3.2.3. 초반에 교차하는 경우

Table 3.4의 두 생존분포가 초반에 교차하는 경우에는 알려진 바와 같이 로그-순위 검정법과 일반화된 Wilcoxon 검정법의 검정력이 낮았다. Renyi type의 두 검정통계량과 Lee의 선형결합통계량에 의한 검정법은 앞선 두 검정법 보다는 높은 값을 가졌으나, 검정력이 아주 높지는 않았다. 수정된 Kolmogorov-Smirnov 검정법과 수정된 Cramer-Von Mises 검정법은 매우 유사한 검정력을 보였다. 이들은 로그-순위 검정법과 일반화된 Wilcoxon 검정법 보다는 검정력이 높았지만, 초반에 교차하는 상황이 오랜 시점동안 두 생존분포 간 작은 차이가 유지되는 형태에 가깝기 때문에 다른 검정법들에 비해 낮은 값을 가졌다. 반면 Lee의 최대통계량에 의한 검정법과 Qui-Sheng 검정법과 이단계 검정법은 매우 높은 검정력을 보였다.

Table 3.4 Power of test statistics when two survival distributions cross at early times

Test statistics	$\alpha=0.05$			$\alpha=0.01$		
	censoring rate			censoring rate		
	0%	25%	50%	0%	25%	50%
LR	0.248	0.215	0.154	0.082	0.071	0.060
GH	0.206	0.112	0.061	0.066	0.034	0.012
Renyi_LR	0.520	0.347	0.203	0.151	0.091	0.053
Renyi_GH	0.792	0.623	0.327	0.549	0.323	0.101
MKS	0.433	0.280	0.184	0.087	0.061	0.039
CVM	0.435	0.285	0.198	0.114	0.061	0.040
LEE1	0.999	0.967	0.828	0.979	0.860	0.647
LEE2	0.400	0.456	0.413	0.133	0.197	0.175
QS1	0.959	0.928	0.847	0.880	0.795	0.642
QS2	0.998	0.965	0.838	0.900	0.776	0.590

3.2.4. 중반에 교차하는 경우

중반에 교차하는 경우에는 Figure 3.1에서 알 수 있듯이, 초반에 교차하는 경우에 비해 교차지점을 중심으로 앞쪽에서 두 생존함수 간의 차이가 크게 나타나는 형태이다 (Table 3.5). 일반화된 Wilcoxon 검

정법은 초기 사망에 더 큰 비중을 두어 두 집단의 생존 확률을 비교하고자 할 경우 각 사망시점에서 두 군의 총 대상자 수로 가중합하여 그 통계량이 계산되는 방법이다. 이에 두 생존분포 간 초반의 차이에 많은 비중을 두는 일반화된 Wilcoxon 검정법의 검정력이 향상되었다 (Park 등, 2010). 이와 관련하여 일반화된 Wilcoxon 검정법 버전의 Renyi type 검정통계량에서 검정력이 높았다. 반면 중반에 교차하는 경우에서 현저히 낮은 검정력을 보인 로그-순위 검정법에 비해 로그-순위 버전의 Renyi type 검정통계량은 높은 값을 가졌다. Lee의 선형결합통계량에 의한 검정법은 가장 낮은 검정력을 보였다. 이밖에, 수정된 Cramer-Von Mises 검정법과 Lee의 최대통계량에 의한 검정법, Qui-Sheng 검정법 및 이단계 검정법의 검정력은 매우 좋았다. 특히 Qui-Sheng 검정법과 이단계 검정법은 중도절단율의 변화에도 그 검정력이 크게 변하지 않았다.

Table 3.5 Power of test statistics when two survival distributions cross at middle times

Test statistics	$\alpha=0.05$			$\alpha=0.01$		
	censoring rate			censoring rate		
	0%	25%	50%	0%	25%	50%
LR	0.123	0.105	0.090	0.035	0.038	0.028
GH	0.897	0.648	0.383	0.716	0.413	0.172
Renyi_LR	0.869	0.705	0.463	0.579	0.387	0.166
Renyi_GH	0.990	0.938	0.724	0.946	0.789	0.443
MKS	0.869	0.750	0.488	0.556	0.343	0.199
CVM	0.924	0.832	0.615	0.739	0.556	0.338
LEE1	0.995	0.955	0.754	0.957	0.803	0.474
LEE2	0.104	0.073	0.061	0.028	0.014	0.011
QS1	0.999	0.997	0.963	0.995	0.970	0.870
QS2	0.999	0.993	0.932	0.994	0.953	0.829

3.2.5. 후반에 교차하는 경우

두 생존분포가 후반에 교차하는 경우는 두 생존분포 간 차이가 앞쪽에서 크게 나는 형태에 가깝다 (Table 3.6). 따라서 초반의 차이에 가중치를 크게 두는 일반화된 Wilcoxon 검정법과 이에 근거한 Renyi type 검정법의 검정력이 높게 나타났다. 더불어 이러한 형태는 짧은 기간 내에 큰 차이가 나타나므로, 수정된 Kolmogorov-Smirnov 검정법과 Cramer-Von Mises 검정법에서 높은 검정력을 보였다. 로그-순위 버전의 Renyi type 검정법과 Lee의 최대통계량에 의한 검정법 역시 검정력이 좋았다. 앞서 언급한 검정법들 가운데 로그-순위 버전의 Renyi type 검정법을 제외하고 중도절단이 없는 경우 유의수준 0.05와 0.01 하에서 모두 높은 값을 가졌으며, 여기서 일반화된 Wilcoxon 검정법을 제외하면 유의수준 0.05 하에서 중도절단율이 25%인 경우에도 좋은 검정력을 보였다. 마지막으로 Qui-Sheng 검정법과 이단계 검정법 또한 그 검정력이 매우 높았다.

Table 3.6 Power of test statistics when two survival distributions cross at late times

Test statistics	$\alpha=0.05$			$\alpha=0.01$		
	censoring rate			censoring rate		
	0%	25%	50%	0%	25%	50%
LR	0.627	0.482	0.403	0.406	0.277	0.208
GH	0.981	0.894	0.679	0.931	0.727	0.460
Renyi_LR	0.924	0.826	0.626	0.759	0.588	0.335
Renyi_GH	0.993	0.948	0.813	0.963	0.841	0.592
MKS	0.946	0.858	0.659	0.759	0.591	0.366
CVM	0.975	0.923	0.784	0.905	0.773	0.527
LEE1	0.981	0.938	0.740	0.943	0.830	0.498
LEE2	0.583	0.382	0.235	0.291	0.143	0.063
QS1	0.984	0.928	0.819	0.914	0.806	0.591
QS2	0.980	0.928	0.807	0.908	0.791	0.582

4. 결론

두 생존분포의 동일성 검정은 생존분석에서 주된 관심사이며, 동시에 두 생존분포가 교차하는 상황은 실제 의학연구에서 빈번히 발생하고 있기에 이에 관한 많은 연구가 진행되어왔다. 특히 본 논문에서는 교차하는 두 집단의 생존분포 상황 하에서 교차지점에 대한 사전 정보가 없이 전체 생존분포를 대상으로 하는 검정법들에 대해 다루었다. 생존분포 간의 교차가 발생하는 경우 개발된 방법 외에 교차되지 않은 경우 개발되었으나 선행 연구로부터 교차하는 상황 하에서도 높은 검정력을 보인 검정법들을 탐구하고, 나아가 모의실험을 통해 교차하는 지점과 중도절단율에 따른 다양한 상황에서 로그-순위 검정법, 일반화된 Wilcoxon 검정법과의 검정력의 비교를 수행하였다. 그 결과를 종합하면 다음과 같다.

비례위험모형 가정이 성립하는 경우 대부분의 검정법에서 검정력이 높았으나, 교차하는 상황 하에 유용하도록 고안된 Qui-Sheng 검정법은 매우 낮은 검정력을 보였다. 초반에 교차하는 경우 Lee의 최대통계량에 의한 검정법과 Qui-Sheng 검정법 및 이단계 검정법에서 검정력이 높았다. 중반에 교차하는 경우에는 일반화된 Wilcoxon 버전의 Renyi type 검정통계량과 수정된 Cramer-Von Mises 검정법, Lee의 최대통계량에 의한 검정법, Qui-Sheng 검정법 및 이단계 검정법에서 높은 검정력을 보였으며, 특히 Qui-Sheng 검정법과 이단계 검정법은 중도절단율의 변화에도 검정력의 차이가 크게 나타나지 않았다. 후반에 교차하는 경우에는 초반에서 큰 차이가 나타나는 두 생존분포의 형태에 따라 일반화된 Wilcoxon 검정법 및 그에 상응하는 Renyi type 검정통계량, 수정된 Kolomogorov-Smirnov 검정법, 수정된 Cramer-Von Mises 검정법, Lee의 최대통계량에 의한 검정법, Qui-Sheng 검정법, 이단계 검정법에서 높은 검정력을 보였다. 특히 Lee의 최대통계량과 Qui-Sheng 이단계 검정법은 두 생존분포의 형태에 대한 사전정보가 없을 때 유용하다는 특징으로 인해 교차하는 상황에서도 역시 다른 통계량들에 비해 좋은 검정력을 보였고, 교차하는 상황임을 고려한다면 Qui-Sheng 검정법 역시 두 생존분포의 차이를 잘 찾아낼 수 있음을 알 수 있었다.

본 논문의 결과를 토대로 주어진 상황 하에서 적합한 동일성 검정법을 선택하는데 유용한 정보를 얻을 수 있었으며, 여기서의 정보가 의학연구 및 임상연구에서 빈번하게 발생하는 교차하는 형태의 두 생존분포 비교 시 도움이 될 수 있을 것으로 기대하는 바이다.

References

- Aalen, O. O. (1976). Nonparametric inference in connection with multiple decrement models. *Scandinavian Journal of Statistics*, **3**, 15-27.
- Aalen, O. O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. *The Annals of Statistics*, **6**, 534-545.
- Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, **2**, 437-453.
- Kitabo, C. A. and Kim, J. T. (2014). Survival analysis of bank loan repayment rate for customers of Hawassa commercial bank of Ethiopia. *Journal of the Korean Data & Information Science Society*, **25**, 1591-1598.
- Davis, M. and Xie, S. X. (2011). *Caution: Hazards crossing! using the renyi test statistic in survival analysis*, PharmaSUG 2011, Tennessee.
- Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis*, Wiley, New York.
- Gill, R. D. (1980) *Censoring and stochastic integrals*, The Mathematical Centre, Netherlands.
- 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
- Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data*, 2nd Ed., Springer, Berlin.
- Lee, J. W. (1996). Some versatile tests based on the simultaneous use of weighted log-rank statistics. *Biometrics*, **52**, 721-725.

- Lee, W. K., Song, M. U. and Jeong, S. H. (2011). Estimation of conditional mean residual life function with random censored data. *Journal of the Korean Data & Information Science Society*, **22**, 89-97.
- Logan, B. R., Klein, J. P. and Zhang, M. J. (2008). Comparing treatment in the Presence of crossing survival curves: An application to bone marrow transplantation. *Biometrics*, **64**, 733-740.
- Nelson, W. B. (1969). Hazard plotting for incomplete failure data. *Journal of Quality Technology*, **1**, 27-52.
- Park, J. K., Oh, K. H. and Kim, M. S. (2010). A default-rate comparison of the construction and other industries using survival method. *Journal of the Korean Data & Information Science Society*, **21**, 747-756.
- Qui, P. and Sheng, J. (2008). A two-stage procedure for comparing hazard rate functions. *Journal of the Royal Statistical Society B*, **70**, 101-208.
- Schumacher, M. (1984). Two-sample tests of cramer-von mises- and kolmogorov-smirnov-type for randomly censored data. *International Statistical Review*, **52**, 263-281.

A comparison of the statistical methods for testing the equality of crossing survival functions[†]

Youn Ju Lee¹ · Jae Won Lee²

^{1,2}Department of Statistics, Korea University

Received 27 January 2015, revised 16 February 2015, accepted 6 April 2015

Abstract

Log-rank is widely used for testing equality of two survival functions, and this method is efficient only under the proportional hazard assumption. However, crossing survival functions are common in practice. Therefore, many approaches have been suggested to test equality of them. This study considered several methods; Renyi type test, modified Kolmogorov-Smirnov and Cramer-von Mises test, and weighted Log-rank test, which can be applied when the survival functions cross, and simulated power of those methods. Based on the simulation results, we provide the useful information to choose a suitable approach in a given situation.

Keywords: Crossing survival functions, equality of survival, log-rank test, survival function.

[†] This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A2008686).

¹ Master, Department of Statistics, Korea University, Seoul 136-701, Korea.

² Corresponding author: Professor, Department of Statistics, Korea University, Seoul 136-701, Korea.
E-mail: jael@korea.ac.kr