응용통계연구 제 11권 1호, 1998년, pp. 113-127

두 생존분포의 동일성 검정에 관한 비교연구 *

정미남† 이재원‡

요 약

생존자료의 분석에 있어 두 집단간의 생존분포의 비교는 자주 관심의 대상이 되고 있다. 중도절단(censoring)이 존재하는 생존자료에 있어 두 생존분포의 동일성을 검정하는 방법으로 log-rank 통계량과 Gehan의 일반화된 Wilcoxon 통계량에 근거한 검정법이 주로 사용되어 왔다. 그러나 이 두 가지 검정통계량이 어떤 상황에서나 적절한 것은 아니고, 두 생존분포의 여러 가지 형태와 중도절단의 정도에 따라 통계량의 검정력은 크게 달라진다. 따라서 본 논문에서는 두 생존분포의 비교를 위해 제안된 몇가지 검정통계량들을 여러 가지 상황에서 모의실험을 통하여 비교하고, 그 결과를 토대로 주어진 상황에서 적절한 통계량을 선택하는데 대한 유용한 정보를 제공하였다.

1. 서론

의약학 분야에 있어 임상실험(clinical trial)을 통한 새로운 치료법이나 치료약의 개발은 매우 중요한 일이다. 암과 같은 난치병의 치료에 있어 기존의 치료법이나 치료약보다 새로 개발한 치료법이나 치료약이 생존률을 높일 수 있는 것으로 판명된다면 암의 치료나 생존시간의 연장에 큰 도움이 될 것이다. 이와 같이 두 생존분포의 비교는 의약학 분야에서 자주 관심의 대상이 되는데, 가장 일반적으로 사용되고 있는 방법은 Mantel(1966)의 log-rank 검정법과 Gehan(1965)의 일반화된 Wilcoxon 검정법이다. Log-rank 검정법은 비례위험(proportional hazards)의 가정이 성립하는 경우에 검정력이 좋은 반면, 일반화된 Wilcoxon 검정법은 생존률의 초반 차이를 찾아내기에 더 좋은 것으로 알려져 있다. 주어진 상황에 맞는 적절한 검정법을 사용하지 않는다면 검정력이 심각하게 저하될 것이다. 이리하여 두 생존분포를 비교하기 위한 여러 가지 검정통계량이 제안되어 왔는데, Fleming, O'Fallon & O'Brien(1980), Brookmeyer & Crowley (1982), Pepe & Fleming(1989), Lee(1996) 등이 제안한 검정통계량들이 있다.

Fleming, O'Fallon & O'Brien(1980)이 제안한 수정된 Kolmogorov-Smirnov 검정법은 고전적인 Kolmogorov-Smirnov 검정법을 중도절단이 존재하는 생존자료에 적합하도록 수정한 것으로, 두 생존함수의 최대거리(maximum distance)에 근거한 방법이다. Brookmeyer & Crowley(1982)가 제안한 중도절단 자료에 대한 중위수 검정법(median test)은 두 생존곡선의 형태에 민감하기보다는 중위생존시간(median survival time)의 차이에 민감한, 즉 차이

*이 연구는 1994년도 한국과학재단 연구비 지원에 의한 결과임(과제번호:94-1300-01-01-3)
† (156-710) 서울시 동작구 신대방2동 395-68, (주) DACOM System Technologies, 공공사업1팀
‡ (136-701) 서울시 성북구 안암동 5-1, 고려대학교 통계학과 부교수

를 잘 찾아내는, 방법이다. Pepe & Fleming(1989)은 가중 Kaplan-Meier 통계량을 제안하였는데, 이 통계량은 두 생존함수의 차이에 근거한 방법으로 연구종료시점이 가까워짐에 따라 더 작은 가중치를 주는 방법이다. 한편, Lee(1996)는 가중 Log-rank 통계량들 중에서 생존차이의 여러가지 형태를 포괄하는 가중치가 다른 몇가지 통계량들을 고려하고, 그것들의 최대통계량(maximum statistic)과 선형결합통계량(linear combination statistic)에 의한 검정법을 제안하였다.

본 논문에서는 앞에서 언급한 두 생존분포의 동일성을 검정하기 위한 여러 가지 방법들을 소개하고, 여러 가지 상황하에서 모의실험을 실시하여 가장 널리 사용되고 있는 Log-rank 통계량과 일반화된 Wilcoxon 통계량과 이들 검정법의 유의수준과 검정력을 비교하였다.

2. 두 생존분포의 동일성 검정법

먼저 임의중도절단(random censoring)이 일어나는 두 표본을 가정하자. 관측치는 i번째 표본에서 N_i개의 개체들로부터 얻어진다($i = 1, 2$). 이 때 (T_{ij}, C_{ij})를 표본 i의 j번째 개체로부터 얻어진 실패시간(failure time)과 중도절단시간(censoring time)이라고 한다면($i = 1, 2; j = 1, \cdots, N_i$), 실제 관측되는 것은 생존시간 X_{ij}와 중도절단 지시함수 δ_{ij}이고 그 관계는 다음과 같다.

$$X_{ij} = \min(T_{ij}, C_{ij}), \delta_{ij} = I\{T_{ij} \le C_{ij}\}(i = 1, 2; j = 1, \cdots, N_i)$$

한편 $N_i(t)$를 시점 t에서 표본 i에 생존하고 있는 개체의 수라 하고, $D_i(t)$는 시점 t에서의 사망개체의 수라 하자. i번째 표본의 생존함수(survival function)는

$$S_i(t) = \Pr(X_{ij} > t) = \exp(-\Lambda_i(t))$$

로 표현되고, 여기서 $\Lambda_i(t)$는 누적위험함수(cumulative hazard function)이다.

지금 우리가 검정하고자 하는 것은 두 생존분포의 동일성이므로 귀무가설은

$$H_0 : S_1 = S_2 \tag{2.1}$$

이다. 생존분석에 있어 식 (2.1)의 가설을 검정하기 위한 통계적 방법으로는 log-rank 검정법과 일반화된 Wilcoxon 검정법이 가장 널리 사용되고 있다. 그러나 이 두 가지 검정법이 모든 상황에서 적절한 것은 아니다. 여기에서는 이 두 가지 방법과 비교될 수 있는 다섯가지 검정법을 소개하도록 하겠다.

2.1. 수정된 Kolmogorov-Smirnov 검정법

이 검정법은 두 표본의 경험적 분포함수(empirical distribution)의 최대거리(maximum distance)에 근거한 고전적인 Kolmogorov-Smirnov 검정법을 중도절단이 있는 생존자료에 대해 수정한 것으로, 경험적 분포함수 대신 생존함수를 사용하여 유도된 방법이다.

혼합표본으로부터 사망 혹은 중도절단으로 관측된 시점을 $t_j(j=1,\cdots,m)$라 하고, 그 중 사망시점을 $T_j(j=1,\cdots,d)$, 중도절단시점을 $\tau_j(j=1,\cdots,c)$라고 하자. 이 때 $m \le d+c$이다. Nelson(1969)의 누적위험함수(cumulative hazard function)를 $\beta_i(t) = -\ln S_i(t)$라 하면, 누적위험추정량은 $\hat{\beta}_i(t) = \sum_{T_j \le t} \sum_{k=0}^{D_i(T_j)-1} \{N_i(T_j) - k\}^{-1}$와 같이 주어진다. 이 때 두 표본의 동일성 검정을 위한 Kolmogorov-Smirnov 통계량의 수정된 형태가 유도되는데, 그 형태는

$$MKS = \sup_{0 \le t \le T} Y_{N_1 N_2}(t) \tag{2.2}$$

와 같다(cf. Schumacher, 1984). 여기서

$$\begin{aligned} Y_{N_1N_2} = \; & \frac{1}{2}\{\hat{S}_1(t) + \hat{S}_2(t)\} \int_0^t \left\{ \frac{N_1\hat{C}_1(s-)N_2\hat{C}_2(s-)}{N_1\hat{C}_1(s-) + N_2\hat{C}_2(s-)} \right\}^{\frac{1}{2}} \\ & \times\{N_1(s)N_2(s) \ge 0\} d\{\hat{\beta}_1(s) - \hat{\beta}_2(s)\} \end{aligned}$$

이고, $T = \max\{t_j : N_1(t_j)N_2(t_j) > 0\}$이다. 이 때 $\hat{C}_i(\cdots)$는 중도절단생존함수(censoring survival function)의 Kaplan-Meier 추정량이고, $\hat{C}_i(s-)$는 아주 작은 $\varepsilon \ge 0$에 대하여 $\hat{C}_i(s-\varepsilon)$을 의미한다. 한편 MSK는 $H_0 : S_1 = S_2$가 참일 때 근사적으로 평균이 0인 정규분포를 따르는 것이 증명되었다.

2.2. 중위수 검정법(MEDIAN TEST)

이 절에서 설명하는 중위수 검정법은 고전적인 중위수 검정법(cf. Westenberg, 1948)을 Brookmeyer & Crowley(1982)가 중도절단이 있는 생존자료에 적용한 것이다.

i번째$(i=1,2)$ 표본으로부터의 생존함수의 Kaplan-Meier 추정량은 $\hat{S}_i(t) = 1 - \hat{F}_i(t)$인데, 이 때 가중 Kaplan-Meier 추정량을

$$\hat{F}_w(t) = 1 - \hat{S}_w(t) = \sum_{i=1}^{2} \lambda_i^N \hat{F}_i(t)$$

와 같이 정의한다. 여기서 $\lambda_i^N = N_i/N$이고 $N = \sum_{i=1}^{2} N_i$이다. 이 때 혼합표본(pooled sample)의 중위수를

$$\hat{M} = inf\{t : \hat{F}_w(t) \ge \frac{1}{2}\} = \hat{F}_w^{-1}(\frac{1}{2})$$

로 정의할 경우, $\hat{F}_i$은 계단함수(step function)이므로 $\hat{F}_w(\hat{M})$이 정확하게 $\frac{1}{2}$의 값을 가지지 못한다. 이러한 이유로 계단함수 $\hat{F}_i$를 각각의 계단점을 연결한 연속형함수 $\hat{F}_i^*$로 변환하고, 이 때 $\hat{F}_w^*(\hat{M}^*) = \frac{1}{2}$을 만족하는 $\hat{M}^*$를 얻게 된다.

이로부터 Brookmeyer & Crowley(1982)는 다음과 같은 검정통계량을 제안하였는데, 이 통계량은 근사적으로 $\chi^2(1)$을 따르는 것으로 증명되었다.

$$T_M = N\left\{\frac{\hat{F}_1^*(\hat{M}^*) - \frac{1}{2}}{\sigma_0}\right\}^2$$

여기서 σ_0^2는

$$\hat{\sigma}_0^2 = \frac{\lambda_2}{\lambda_1}\{\lambda_2\hat{V}_1(\hat{M}^*) + \lambda_1\hat{V}_2(\hat{M}^*)\}$$

에 의해 추정되고,

$$\hat{V}_i(\hat{M}^*) = \{\hat{S}_i^*(\hat{M}^*)\}^2 \sum_{\{j:X_{ij}\le M^*\}} \frac{N_iD_i(X_{ij})}{N_i(X_{ij})(N_i(X_{ij}) - D_i(X_{ij}))}$$

이다.

2.3. 가중 Kaplan-Meier 통계량에 의한 검정법

일반적으로 중도절단이 존재하는 생존자료에 있어 i번째($i = 1, 2$) 표본으로부터 Kaplan-Meier 추정량 $\hat{S}_i(t)$는 t가 연구종료시점에 가까워질수록 중도절단의 가능성이 높아지기 때문에 안정적이지 못한 경향이 있다. Pepe & Fleming(1989)의 가중 Kaplan-Meier 통계량은 이러한 단점을 보완한 것으로 t가 연구종료시점에 가까워짐에 따라 $\hat{S}_1(t) - \hat{S}_2(t)$에 더 작은 가중치를 주는 방법인데, 그 형태는

$$WMK = \sqrt{\frac{N_1N_2}{N}}\int_0^{T_c}\hat{w}(t)\{\hat{S}_1(t) - \hat{S}_2(t)\}dt \tag{2.3}$$

와 같다. 여기서 가중함수(weight function) $\hat{w}(t)$는

$$\hat{w}(t) = \frac{\hat{C}_1^-(t)\hat{C}_2^-(t)}{\hat{p}_1\hat{C}_1^-(t) + \hat{p}_2\hat{C}_2^-(t)}$$

이고, $T_c = \sup\{t; \hat{C}_1(t)\wedge\hat{C}_2(t) > 0\}$, $\hat{C}_i(\cdot)$는 중도절단생존함수의 Kaplan-Meier 추정량이고, $\hat{C}_i^-(t)$는 $\hat{C}_i^{-1}(t)$을 의미한다. 한편 O'Sullivan & Fleming(1986)은 어떤 양의 상수 Γ와 δ에 대하여, $|w(t)| \le \Gamma\{C_i^-(t)\}^{(1/2)+\delta}$, $|\hat{w}(t)| \le \Gamma\{\hat{C}_i^-(t)\}^{(1/2)+\delta}$이 만족되면 식 (2.3)이 근사적으로 분산이 σ^2인 정규분포를 따름을 보였다. 이 때 분산의 추정치로는 혼합표본으로부터 구한

$$\hat{\sigma}_p^2 = -\int_0^{T_c}\frac{\{\int_t^{T_c}\hat{w}(u)\hat{S}(u)du\}^2}{\hat{S}(t)\hat{S}^-(t)}\frac{\hat{p}_1\hat{C}_1^-(t) + \hat{p}_2\hat{C}_2^-(t)}{\hat{C}_1^-(t)\hat{C}_2^-(t)}d\hat{S}(t)$$

을 사용하고, $\hat{p}_i = N_i/N$이다. 한편, 중도절단이 존재하지 않는 경우에 $\hat{w}(t)$는 1의 값을 가지게 되므로 이 검정법은 중도절단 자료에 대한 Z-검정의 일반화로 간주될 수 있다.

2.4. LEE 검정법

시점 t까지의 사망자수를 $N_{ij}(t) = I\{X_{ij} \le t, \delta_{ij} = 1\}$라 하고, 시점 t에서 살아있는 사람의 수를 $Y_{ij}(t) = I\{X_{ij} \ge t\}$로 나타내자. 또한 $\bar{N}_i = \sum_j N_{ij}$, $\bar{N} = \sum_i \bar{N}_i$, $\bar{Y}_i = \sum_j Y_{ij}$, $\bar{Y} = \sum_i \bar{Y}_i$로 놓기로 하자. 이때 두 생존분포의 동일성을 검정하기 위한 가중 Log-rank 통계량을

$$W_K = \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \int_0^\infty W(t) \frac{\bar{Y}_1(t)\bar{Y}_2(t)}{\bar{Y}_1(t) + \bar{Y}_2(t)} \left\{ \frac{d\bar{N}_1(t)}{\bar{Y}_1(t)} - \frac{d\bar{N}_2(t)}{\bar{Y}_2(t)} \right\}$$

와 같은 형태로 표현할 수 있다. 여기서 가중함수는

$$W(t) = \{\hat{S}(t-)\}^\rho \{1 - \hat{S}(t-)\}^\gamma (\gamma \ge 0, \rho \ge 0)$$

로 정의하고, 이 때 W_K를 $G^{\rho,\gamma}$통계량이라 부른다. Lee(1996)는 이로부터 다음의 네 가지 통계량을 고려하였다.

(ρ, γ)	$W(t)$	가중함수에 따른 특징	$G^{\rho,\gamma}$
$(0,0)$	1 (Log-rank)	비례위험의 가정 만족	W_{K_1}
$(2,0)$	$\{\hat{S}(t-)\}^2$	초반 차이의 강조	W_{K_2}
$(0,2)$	$\{1-\hat{S}(t-)\}^2$	후반 차이의 강조	W_{K_3}
$(2,2)$	$\{\hat{S}(t-)\}^2\{1-\hat{S}(t-)\}^2$	중반 차이의 강조	W_{K_4}

$H_0 : S_1 = S_2$하에서

$$(W_{K_1}, W_{K_2}, W_{K_3}, W_{K_4}) = (G^{0,0}, G^{2,0}, G^{0,2}, G^{2,2})$$

는 근사적으로 평균이 0인 다변량 정규분포를 따른다. Lee(1996)는 이러한 네 가지 통계량을 동시에 사용하는 두 가지 유용한 검정법을 제안하였는데, 표준화된 통계량인 $\{Z_l^* = W_{K_l}/\sqrt{\hat{Var}(W_{K_l})}, l = 1, \cdots, 4\}$을 사용하였다. 그 중 한 방법은 Z_l^*들의 최대통계량(maximum statistic)인

$$Lee1 = \max\{Z_1^*, Z_2^*, Z_3^*, Z_4^*\} \tag{2.4}$$

을 사용하는 방법이고, 또 다른 방법은 Z_l^*들의 선형결합통계량(linear combination statistic)인

$$Lee2 = \sum_{l=1}^4 b_l Z_l^* (\text{이 때 } \sum_{l=1}^4 b_l = 1) \tag{2.5}$$

을 사용하는 것이다. 식 (2.4)는 (Z_1, Z_2, Z_3, Z_4)가 평균이 0인 4변량 정규분포를 따를 때 $\max\{Z_1, Z_2, Z_3, Z_4\}$와 근사적으로 같은 분포를 가진다. 또한 식 (2.5)는 근사적으로 표준정규분포를 따르므로 이 두 통계량에 의해 $H_0 : S_1 = S_2$을 검정할 수 있다.

3. 모의실험의 구조 및 결과

3.1. 모의실험의 구조

두 생존분포의 동일성 검정을 위한 여러 가지 통계량들을 가장 일반적으로 사용되는 log-rank 통계량, 일반화된 Wilcoxon 통계량과 함께 여러 가지 상황 하에서 모의실험을 통해 비교해 보고자 한다. 모의실험을 위한 통계량은 2장에서 소개한 수정된 Kolmogorov-Smirnov 통계량, 중위수 검정법에서의 통계량, 가중 Kaplan-Meier 통계량, Lee의 최대통계량 및 선형결합통계량이다. 이 때 각각의 통계량을 다음과 같이 표시하기로 한다.

1. MKS (수정된 Kolmogorov-Smirnov 통계량)
2. T_M(중위수 검정법에서의 통계량)
3. WKM (가중 Kaplan-Meier 통계량)
4. Lee 1 (Lee의 최대통계량)
5. Lee 2 (Lee의 선형결합통계량)
6. log-rank (log-rank 통계량)
7. Gehan (일반화된 Wilcoxon 통계량)

귀무가설 하에서 자료는 두 집단이 각각 exp(1)을 따르도록 생성하였고 (표 3.2), 대립가설 하에서는 7가지 상황을 설정하여 자료를 생성하였다 (표 3.3 - 표 3.9). 그러한 여러가지 상황에 대한 두 집단의 분포를 다음의 표 3.1에 정리하였다.

8가지 형태의 자료에 있어 두 집단의 각각에 대한 표본의 크기는 50으로 정하였고, 유의수준이 0.05와 0.01인 경우에 대해 1000번의 모의실험을 실시하였다. 한편 자료는 중도절단이 없는 경우(NC)와 중도절단이 있는 경우(C)에 대해서 생성되었는데, 중도절단이 있는 경우에는 중도절단시간(censoring time)이 $U(0,2)$의 분포를 따른다. 중도절단시간이 $U(0,2)$을 따르고, 생존시간이 exp(1)을 따르면 중도절단율이 43%가 된다.

3.2. 모의실험의 결과

이 절에서는 두 생존분포의 동일성 검정을 위한 7가지 통계량에 관한 모의실험 결과를 제공하기로 한다. $\alpha = 0.05$인 경우와 $\alpha = 0.01$인 경우가 비슷한 결과를 가지므로 $\alpha = 0.05$인 경우만 살펴보도록 하겠다. 여기에서 Lee의 선형결합통계량의 사용에 있어 Z_l^*들의 계수로는 $b_1 = b_2 = b_3 = b_4 = \frac{1}{4}$을 사용하였다.

3.2.1. 두 생존분포가 동일한 경우 (표 3.2)

이 경우 유의수준의 달성정도는 검정통계량의 종류에 따라 약간 다른 양상을 띠고 있음을 알 수 있다. Log-rank 검정법과 중위수 검정법은 유의수준을 약간 초과하고 있는 반

표 3.1: 각각의 상황에서 설정된 두 집단의 분포

	집단 1	집단 2
표 3.2	exp(1)	exp(1)
표 3.3	exp(1)	exp(2)
표 3.4	exp(0.75)→ exp(3)→ exp(1)	exp(3)→ exp(0.75)→ exp(1)
표 3.5	exp(2)→ exp(0.5)	exp(2)→ exp(4)
표 3.6	exp(2)→ exp(0.25)→ exp(4)→ exp(1)	exp(2)→ exp(4)→ exp(0.25)→ exp(1)
표 3.7	Weibull(1,1.5)	Weibull(1.5,1)
표 3.8	Weibull(2,1)	Weibull(1,0.5)
표 3.9	Weibull(1,0.5)	Weibull(1,0.5)+0.5

면, 수정된 Kolmogorov-Smirnov 검정법과 Lee의 두 검정법은 유의수준에 다소 못미치는 경향이 있다. 이 중 특히 Lee의 최대통계량에 의한 검정법은 중도절단이 없을 때와 비교하여 중도절단이 있는 경우에 값이 다소 떨어졌으며, 일반화된 Wilcoxon 검정법과 가중 Kaplan-Meier 검정법은 그과 반대로 값이 오르는 경향을 보였다.

3.2.2. 비례위험의 가정이 성립하는 경우 (표 3.3)

먼저 중도절단이 없는 경우를 살펴보면 log-rank 검정법과 수정된 Kolmogorov-Smirnov 검정법, Lee의 두 가지 검정법의 검정력이 좋았고, 중위수 검정법의 검정력은 다소 떨어지는 경향을 보였다. 중도절단이 있는 경우에 있어서는 먼저 log-rank 검정법과 Lee의 두 가지 검정법의 검정력이 좋았고, 수정된 Kolmogorov-Smirnov 검정법의 검정력은 중도절단이 없는 경우와는 달리 매우 낮았다. 이로부터 수정된 Kolmogorov -Smirnov 검정법의 검정력은 비례위험의 가정이 성립하는 경우에 있어 중도절단의 유무에 크게 영향을 받는 것으로 볼 수 있다.

3.2.3. 초반에 차이가 나는 경우 (표 3.4)

중도절단이 없는 경우에 있어서는 수정된 Kolmogorov-Smirnov 검정법의 검정력이 가장 좋았고, 그 다음으로 Lee의 최대통계량에 의한 검정법과 일반화된 Wilcoxon 검정법, 가중 Kaplan-Meier 검정법의 검정력이 좋았으나, Lee의 선형결합통계량에 의한 검정법의 검정력은 다소 떨어졌다. 중도절단이 있는 경우에는 가중 Kaplan-Meier 검정법, 중위수 검정법, 일반화된 Wilcoxon 검정법의 검정력이 가장 좋았고, 수정된 Kolmogorov-Smirnov 검정법과 Lee의 최대통계량에 의한 검정법의 검정력이 그 다음으로 좋았다. 그러나 이 경우 log-rank 검정법과 Lee의 선형결합통계량에 의한 검정법의 검정력은 약간 떨어지는 경향을 보였다.

3.2.4. 후반에 차이가 나는 경우 (표 3.5)

먼저 중도절단이 없는 경우에서는 수정된 Kolmogorov-Smirnov 검정법, Lee의 두 가지 검정법, log-rank 검정법의 검정력이 좋은 반면, 일반화된 Wilcoxon 검정법, 가중 Kaplan-Meier 검정법, 중위수 검정법의 검정력은 상당히 떨어졌다. 이에 반해 중도절단이 있는 경우에서는 Lee의 최대통계량의 검정력이 가장 좋았고, 그 다음으로 Lee의 선형결합통계량과 수정된 Kolmogorov-Smirnov 검정법, log-rank 검정법의 검정력이 좋았으며, 나머지 검정법들의 검정력은 매우 낮았다. 주목할 만한 점은 이 경우에서도 수정된 Kolmogorov-Smirnov 검정법은 비례위험의 가정이 성립하는 경우와 마찬가지로 중도절단이 존재하는 경우에 검정력이 크게 떨어졌는데, 이로부터 수정된 Kolmogorov-Smirnov 검정법은 중도절단의 유무에 의해 크게 영향을 받음을 알 수 있다.

3.2.5. 중반에 차이가 나는 경우 (표 3.6)

중도절단이 없는 경우를 살펴보면 Lee의 최대통계량에 의한 검정법이 가장 높은 검정력을 보여 주었고, 그 다음으로 Lee의 선형결합통계량에 의한 검정법과 가중 Kaplan-Meier 검정법, 일반화된 Wilcoxon 검정법의 검정력이 높았으며, 수정된 Kolmogorov-Smirnov 검정법의 검정력이 가장 낮았다. 중도절단이 있는 경우에는 Lee의 최대통계량에 의한 검정법이 가장 높은 검정력을 보여 주었고, Lee의 선형결합통계량에 의한 검정법과 중위수 검정법, log-rank 검정법의 검정력이 그 다음으로 좋았다. 한편 이 경우에서도 수정된 Kolmogorov-Smirnov 검정법의 검정력이 가장 낮았다.

3.2.6. 두 생존분포의 차이가 서서히 커지다가 후반에 다시 줄어드는 경우 (표 3.7)

중도절단이 없는 경우를 보면 가중 Kaplan-Meier 검정법과 일반화된 Wilcoxon 검정법, Lee의 최대통계량에 의한 검정법의 검정력이 높았으며, Lee의 선형결합통계량에 의한 검정법과 log-rank 검정법의 검정력이 낮았다. 중도절단이 있는 경우에서는 가중 Kaplan-Meier 검정법과 일반화된 Wilcoxon 검정법의 검정력이 높았으며, Lee의 선형결합통계량에 의한 검정력이 가장 떨어졌다.

3.2.7. 두 생존분포가 서로 교차하는 경우 (표 3.8)

중도절단이 없는 경우를 살펴보면 수정된 Kolmogorov-Smirnov 검정법의 검정력이 가장 높고, Lee의 두 가지 검정법의 검정력이 그 다음으로 좋았다. log-rank 검정법과 일반화된 Wilcoxon 검정법, 가중 Kaplan-Meier 검정법의 검정력은 조금 떨어졌으며 중위수 검정법은 가장 낮은 검정력을 보여 주었다. 다음으로 중도절단이 있는 경우를 보면 역시 수정된 Kolmogorov-Smirnov 검정법의 검정력이 가장 높았으며, Lee의 두 가지 검정법의 검정력이 그 다음으로 좋았다. 그러나 log-rank 검정법의 검정력은 다소 떨어졌으며, 나머지 검정법들의 검정력은 매우 낮았다.

3.2.8. 이동(shift)에 의한 두 생존분포의 차이가 발생하는 경우 (표 3.9)

먼저 중도절단이 없는 경우를 보면 가중 Kaplan-Meier 검정법과 일반화된 Wilcoxon 검정법의 검정력이 가장 좋았고, 그 다음으로는 log-rank 검정법과 Lee의 두 가지 검정법의 검정력이 좋았다. 그러나 수정된 Kolmogorov-Smirnov 검정법과 중위수 검정법의 검정력은 조금 낮았다. 한편 중도절단이 있는 경우에는 가중 Kaplan-Meier 검정법과 일반화된 Wilcoxon 검정법의 검정력이 가장 좋았고, log-rank 검정법과 수정된 Kolmogorov-Smirnov 검정법의 검정력이 그 다음으로 좋았다. 이 경우에서 주목할 만한 사실은 중도절단의 유무와 관계없이 가중 Kaplan-Meier 검정법이 가장 높다는 것인데, 이것은 log-rank 통계량이나 수정된 Kolmogorov-Smirnov 통계량과 같이 생존시간의 순위에 근거한 순위통계량(rank-based statistic)이 이동(shift)에 의한 생존분포의 차이를 잘 발견해내지 못하는데 비해 가중 Kaplan-Meier 검정법은 이러한 상황에 민감하다는 특징 때문이다.

전체적으로 살펴보면 Lee의 최대통계량(maximum statistic)이 모든 상황에 있어 다른 통계량들에 비해 대체로 높은 검정력을 보였고, 이로부터 두 생존분포의 형태를 미리 알 수 없는 상황에서 최대통계량에 의한 검정이 여러 가지 형태의 생존차이를 비교적 잘 찾아낼 수 있는 검정임을 알 수 있었다.

표 3.2: 두 생존분포가 동일한 경우의 검정력 비교

검정통계량	유의수준($\alpha = 0.05$)	
	NC	C
MKS	0.035	0.036
T_M	0.073	0.071
WKM	0.043	0.057
Lee 1	0.046	0.037
Lee 2	0.034	0.038
log-rank	0.060	0.060
Gehan	0.044	0.057

표 3.3: 비례위험의 가정이 성립하는 경우의 검정력 비교

검정통계량	유의수준($\alpha = 0.05$)	
	NC	C
MKS	0.993	0.351
T_M	0.609	0.647
WKM	0.848	0.687
Lee 1	0.936	0.780
Lee 2	0.944	0.815
log-rank	0.940	0.805
Gehan	0.856	0.687

표 3.4: 초반에 차이가 나는 경우의 검정력 비교

검정통계량	유의수준($\alpha = 0.05$)	
	NC	C
MKS	0.998	0.935
T_M	0.890	0.960
WKM	0.939	0.966
Lee 1	0.938	0.938
Lee 2	0.599	0.621
log-rank	0.646	0.754
Gehan	0.936	0.963

표 3.5: 후반에 차이가 나는 경우의 검정력 비교

검정통계량	유의수준($\alpha = 0.05$)	
	NC	C
MKS	0.981	0.341
T_M	0.088	0.053
WKM	0.146	0.065
Lee 1	0.993	0.807
Lee 2	0.805	0.463
log-rank	0.735	0.335
Gehan	0.166	0.064

표 3.6: 중반에 차이가 나는 경우의 검정력 비교

검정통계량	유의수준($\alpha = 0.05$)	
	NC	C
MKS	0.137	0.080
T_M	0.203	0.227
WKM	0.245	0.147
Lee 1	0.472	0.427
Lee 2	0.264	0.263
log-rank	0.209	0.204
Gehan	0.247	0.152

표 3.7: 두 생존분포의 차이가 서서히 커지다가 후반에 다시 줄어드는 경우의 검정력 비교

검정통계량	유의수준($\alpha = 0.05$)	
	NC	C
MKS	0.641	0.592
T_M	0.536	0.552
WKM	0.743	0.741
Lee 1	0.719	0.665
Lee 2	0.365	0.454
log-rank	0.384	0.541
Gehan	0.747	0.737

표 3.8: 두 생존분포가 서로 교차하는 경우의 검정력 비교

검정통계량	유의수준($\alpha = 0.05$)	
	NC	C
MKS	0.971	0.864
T_M	0.274	0.121
WKM	0.371	0.059
Lee 1	0.903	0.731
Lee 2	0.812	0.527
log-rank	0.631	0.372
Gehan	0.483	0.059

표 3.9: 이동(shift)에 의한 차이가 발생하는 경우의 검정력 비교

검정통계량	유의수준($\alpha = 0.05$)	
	NC	C
MKS	0.697	0.610
T_M	0.512	0.334
WKM	0.914	0.987
Lee 1	0.811	0.420
Lee 2	0.774	0.309
log-rank	0.842	0.743
Gehan	0.991	0.986

4. 결론

생존자료의 분석에 있어 두 생존분포의 동일성 검정은 자주 관심의 대상이 되는 만큼 검정법의 사용에 있어 주의를 기울여야 한다. 이미 널리 사용되고 있는 log-rank 검정법과 일반화된 Wilcoxon 검정법은 각각의 검정법의 특성에 따라 어떤 상황에서는 생존분포의 차이를 잘 찾아내는 반면, 다른 어떤 상황에서는 차이가 기대됨에도 불구하고 그 차이를 발견하지 못하는 오류를 범하기도 한다. 이러한 오류는 인간의 생명과 관련있는 의학분야에서는 치명적인 결과를 낳을 수 있다. 이러한 이유로 본 논문에서는 생존분포의 동일성 검정을 위한 7가지 검정통계량을 여러가지 상황하에서 비교하여 보았다. 이로부터 주어진 상황에서 적절한 통계량을 선택하는데 대한 유용한 정보를 얻을 수 있었는데, 이러한 정보는 매우 빈번하게 행해지는 생존분포의 비교에 도움이 될 것이다.

참고문헌

[1] Barr, D. R. and Davidson, T. G. (1973). A Kolmogorov-Smirnov test for censored examples, *Technometrics*, 15, 739-757.

[2] Brookmeyer, R. and Crowley, J. (1982). A k-Sample Median Test for Censored Data, *Journal of the American Statistical Association*, Vol. 77, No. 378, 433-440.

[3] Fleming, T. R., Harrington, D. P. and O'Sullivan, M. (1987). Supremum versions of the log-rank and generalized Wilcoxon statistics, *Journal of the American Statistical Association*, 82, 312-320.

[4] 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.

[5] Kaplan, E. L., and Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, 53, 457-481.

[6] Lee, J. W. (1996). Some Versatile Tests Based on the Simultaneous Use of Weighted Log-Rank Statistics, *Biometrics*, 52, 468-472.

[7] Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemotheraphy Report*, 50, 163-170.

[8] O'Sullivan, M. and Fleming, T. R. (1986). Technical Report No. 163, Center for Stochastic Processes, University of North Carolina.

[9] 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.

[10] Pepe, M. S. and Fleming, T. R. (1991). Weighted Kaplan-Meier Statistics: Large Sample and Optimality Considerations, *Journal of the Royal Statistical Society, Series B*, Vol. 53, No. 2, 341-352.

[11] Reid, N. (1981). Estimating the median survival time, *Biometrika*, 68, 601-608.

[12] Schey, H. M. (1977). The asymptotic distribution of the one-sided Kolmogorov-Smirnov statistic for truncated data, *Communications in Statistics, A6*, 1361-1365.

[13] Schmacher, M. (1984). Two-sample tests of the Cramer-von Mises and Kolmogorov-Smirnov type for randomly censored data, *International Statistical Review, 52*, 263-281.

[1997년 4월 접수, 1997년 12월 최종수정]

A Comparison of the Statistical Methods for Testing the Equality of Two Survival Distributions *

Mi Nam Jung †, Jae Won Lee ‡

Abstract

There have been a great deal of interests in comparing two survival distributions in clinical trials. This paper compares some well-known statistical methods for testing the equality of two survival distributions. Simulation studies also provide some insights into the properties of these test statistics across several types of survival distributions and degrees of censorship.

*This research was supported by Korea Science and Engineering Foundation (KOSEF) grant (No. 94-1300-01-01-3) in 1994.
† DACOM System Technologies, 395-68 Sindaebang 2-dong Dongjak-gu, Seoul 156-710.
‡ Department of Statistics, Korea University, 5-1 Anam-dong Sungbuk-gu, Seoul 136-701.