

## 극단적인 오른쪽 관측중단모형에서 생존함수의 추정<sup>1</sup>

이제만<sup>2</sup>

### 요약

수명시험에서 시험에 장기간 노출된 대상 부품이나 실험 대상자의 수명은 관측되는 경우보다 관측중단이 일어나기가 쉽다. 이와 같은 경우에 임의중단모형에서 생존함수 추정량으로 흔히 이용되는 Kaplan과 Meier의 추정량은 수명분포의 오른쪽 꼬리부분에서 심각한 편의가 발생된다. 이러한 문제점에 대한 대안으로 정상적으로 관측된 최장 수명보다 큰 관측중단수명이 많은 극단적인 오른쪽 관측중단모형에서 새로운 비모수적 생존함수 추정량을 제안하고 그 특성을 몬테칼로 모의실험을 통하여 기존의 추정량과 비교·분석하였다.

주제어: 극단적인 오른쪽 관측중단모형, 생존함수, 수명시험

### 1. 서론

신뢰성분석이나 생존분석에서 임의중단모형으로 관측된 수명자료를 이용한 신뢰도 (Reliability function) 또는 생존함수(Survival function)의 대표적인 비모수적 추정량에는 Kaplan과 Meier(1958)의 추정량을 들 수 있으며 누적위험률(Cumulative hazard rate)의 비도수적추정량은 Nelson(1972)의 추정량을 들 수 있다.

위와 같은 분야에서 이루어지는 수명시험에서는 흔히 시험에 장기간 노출될수록 대상 부품이나 실험 대상자의 수명은 관측되는 경우보다 관측중단이 일어나기가 쉽다. 이러한 사실은 통계적 분석을 어렵게 만든다(Lagakos, 1979). 이러한 경우에 Kaplan과 Meier의 추정량(KM추정량)이나 Nelson의 추정량 등은 수명분포의 오른쪽 꼬리부분에서 심각한 편이가 발생된다. 이러한 신뢰도함수나 생존함수의 꼬리부분에서 발생하는 추정량의 편이는 평균 수명이나 수명분포의 분위수 등의 추정에도 전이가 된다. 이와 같이 정상적으로 관측된 최장수명보다 큰 관측 중단된 관측치가 많은 경우의 임의중단모형을 극단적인 오른쪽 관측중단모형이라고 하자.

<sup>1</sup>이 연구는 안동대학교 1999년 기성회 학술조성비의 지원으로 수행되었음

<sup>2</sup>(760-749)경북 안동시 송천동 안동대학교 자연과학대학 통계학과 부교수

극단적인 오른쪽 관측중단모형에서 Brown, Hollander와 Korwar(1974)는 관측된 최장수명보다 큰 경우의 KM추정량의 값을 지수분포의 생존함수 추정치로 대치한 수정된 추정량(BHK추정량)을 제안하였고, Moeschberger와 Klein(1985)은 관측된 최장수명보다 큰 경우의 KM추정량의 값을 와이블분포의 생존함수 추정치로 대치한 수정된 추정량(WTAIL추정량)을 제안하고 이와 같은 생존함수의 추정량을 이용하여 구성된 평균수명과 수명분포의 90백분위수에 대한 추정량의 특성을 모의실험을 통하여 분석하였다. Moeschberger와 Klein의 분석 결과는 모의실험에서 고려한 대부분의 수명분포에서 BHK추정량, WTAIL추정량을 이용한 경우가 KM추정량을 이용한 경우보다 편이나 제곱평균오차의 관점에서 우수하였다. 그러나 BHK추정량이나 WTAIL추정량은 관측된 최장수명보다 큰 경우에 단순히 지수분포나 와이블분포의 생존함수의 추정값을 대신 사용하여 최장수명보다 큰 관측 중단된 관측치의 정보를 충분히 반영하고 있다고 보기 어렵다.

이러한 측면에서 본 연구에서는 관측중단이 일어난 자료값을 관측중단시점보다 크다는 조건이 주어진 수명의 조건부 기대값으로 대치하여 이용하자는 Buckley와 James(1979)의 제안을 도입하여 극단적인 오른쪽 관측중단모형에서 최장수명이상의 관측 중단된 자료의 정보를 좀더 효과적으로 이용할 수 있을 것으로 기대되는 새로운 생존함수의 추정량을 구성하고 제안된 추정량의 특성을 기존의 추정량과 비교하여 분석한다.

## 2. 극단적인 오른쪽 관측중단모형에서 생존함수의 추정량

부품 혹은 체계의 수명시험이나 임상실험에서 실험단위의 수명을 분포 함수가  $F(t)$ 인 확률변수  $T_1, \dots, T_n$ 이라 하고, 수명과 독립적인 관측중단시간을 분포함수가  $G(c)$ 인 확률변수  $C_1, \dots, C_n$ 이라 하자.

여기서 관측되는 수명이 관측중단시간보다 작거나 같으면 관측된 수명을 기록하고 수명이 관측중단시간보다 길어 관측이 중단되면 수명이 관측중단시간보다 크다는 사실을 기록한 수명자료를 얻는다고 하자. 이러한 수명관측모형을 임의중단모형이라고 한다. 이 때  $x_i = \min(T_i, C_i)$ ,  $\delta_i$ 는  $T_i \leq C_i$ 일 때 1,  $T_i > C_i$ 일 때 0의 값을 갖는 지시함수라고 하면 임의중단모형에서 관측되는 자료는  $(x_i, \delta_i)$ ,  $i = 1, \dots, n$ 와 같이 나타낼 수 있다.

여기서  $x_i$ 들의 순서통계량의 값을  $(x_{(1)} \leq \dots \leq x_{(n)})$ 라 하고,  $\delta_i = 1$ 인  $x_i$ 들의 순서통계량의 값을  $(x_{(1)}^* \leq \dots \leq x_{(m)}^*)$ 이라 하고,  $S_j$ 는  $x_{(j)}^*$  직전까지 생존하고있는 실험단위의 수라고 하자. 즉  $m \leq n$ 이고,  $x_{(m)}^*$ 는 관측이 중단됨이 없이 정상적으로 관측된 최장수명이고, 최장수명보다 큰 관측중단된 관측치의 수  $n_C$ 는  $n - m$ 개보다 작거나 같게되고,  $S_j$ 는 수명이  $x_{(j)}^*$ 인 실험단위를 포함한  $x_{(j)}^*$  시점에 관측하에 놓여있는 실험단위의 수이다.

이와 같은 임의중단모형의 자료를 이용한 생존함수  $S(t) = 1 - F(t)$ 의 KM추정량은 다음과 같이 정의된다.

$$\widehat{S}_{KM}(t) = \begin{cases} 1, & t < x_{(1)}^* \text{ 일 때} \\ \prod_{k=1}^j (S_k - 1)/S_k, & x_{(j)}^* \leq t < x_{(j+1)}^*, j = 1, \dots, m \text{ 일 때} \\ 0, & t \geq x_{(m+1)}^* \text{ 일 때} \end{cases}$$

여기서  $x_{(m)}^* < x_{(n)}$ 이면  $x_{(m+1)}^* = x_{(n)}$ 이고,  $x_{(m)}^* = x_{(n)}$ 이면  $x_{(m+1)}^* = \infty$ 이다.

$x_{(m)}^*$ 보다 큰 관측 중단된 관측치가 흔히 나타나는 극단적인 오른쪽 관측중단모형에서 KM 추정량은  $x_{(m)}^*$ 보다 큰  $x$ 에서 상당한 크기의 편의를 포함하게 된다. 생존함수의 추정량이 갖는 이러한 편의는 평균수명이나 수명분포의 분위수 등의 추정에도 전이가 된다. 따라서 극단적인 오른쪽 관측중단모형에서 관측되는 수명자료를 이용한 생존분석에서 원형 그대로의 KM추정량을 사용하여 평균수명 혹은 수명분포의 분위수를 추정하는 것은 한계가 있다.

극단적인 오른쪽 관측중단모형에서 KM추정량이 갖는 이러한 문제점을 해결하기 위하여 Brown, Hollander와 Korwar에 의하여 제안된 BHK추정량과 Moeschberger와 Klein에 의하여 제안된 WTAIL추정량은 다음과 같다.

1) BHK추정량:

$$\widehat{S}_{BHK}(t) = \begin{cases} \widehat{S}_{KM}(t), & t \leq x_{(m)}^* \text{ 일 때} \\ \exp(-t/\theta^*), & t > x_{(m)}^* \text{ 일 때,} \end{cases}$$

여기서  $\theta^*$ 는  $\widehat{S}_{BHK}(x_{(m)}^*) = \exp(-x_{(m)}^*/\theta^*)$ 를 만족하는 값이다.

2) WTAIL추정량:

$$\widehat{S}_{WTAAIL}(t) = \begin{cases} \widehat{S}_{KM}(t), & t \leq x_{(m)}^* \text{ 일 때} \\ \exp(-\frac{t}{\widehat{\theta}}), & t > x_{(m)}^* \text{ 일 때} \end{cases}$$

여기서  $\widehat{k}, \widehat{\theta}$ 는 생존함수를  $S(t) = \exp(-t^k/\theta)$ 라고 가정할 때  $k, \theta$ 의 최우추정치이다.

앞에서 고려한 이러한 생존함수의 추정량들은  $x_{(m)}^*$ 보다 큰  $n_C$ 개의 관측 중단된 관측치의 정보를 효과적으로 이용하고 있다고 보기 어렵다. 관측 중단된 관측치의 효과적인 이용과 관련하여 Buckley와 James(1979)는 임의중단자료의 선형회귀에서 관측이 중단된 자료 값을 수명이 관측중단시간보다 크다는 조건하에서 수명의 조건부 기대값으로 대체한 수정된 자료를 이용하여 회귀모수를 추정할 것을 제안하였다.

이와같은 Buckley와 James의 생각과 생존함수의 BHK추정량과 WTAIL추정량을 이용하여 임의 중단자료  $(x_i, \delta_i), i = 1, \dots, n$ 의 관측 중단된 관측치를 수명이 관측중단시간보다 크다는 사실을 조건하에서 수명의 조건부 기대값의 추정치로 수정한 자료는  $(t_1^*, \dots, t_n^*)$ 와  $(t_1^{\sim}, \dots, t_n^{\sim})$  나타낼 수 있다.

$$t_i^* = x_i \delta_i + \left[ x_i + \frac{\int_{x_i}^{\infty} (1 - \hat{F}_{BHK}(t)) dt}{1 - \hat{F}_{BHK}(x_i)} \right] (1 - \delta_i)$$

혹은

$$t_i^{\sim} = x_i \delta_i + \left[ x_i + \frac{\int_{x_i}^{\infty} (1 - \hat{F}_{WTAI L}(t)) dt}{1 - \hat{F}_{WTAI L}(x_i)} \right] (1 - \delta_i),$$

여기서  $\hat{F}_{WTAI L}(x_i) = 1 - \hat{S}_{WTAI L}(x_i)$ ,  $\hat{F}_{BHK}(x_i) = 1 - \hat{S}_{BHK}(x_i)$ .

이제 이와 같이 수정된 수명자료 ( $t_1^*, \dots, t_n^*$ )와 ( $t_1^{\sim}, \dots, t_n^{\sim}$ )의 경험분포함수를 이용하여 생존함수  $S(t) = 1 - F(t)$ 의 비모수적 추정량을 각각 구성하면 다음과 같다.

$$\hat{S}_{BJ 1}(x) = 1 - \frac{1}{n} \sum_{i=1}^n I(t_i^* \leq t),$$

$$\hat{S}_{BJ 2}(x) = 1 - \frac{1}{n} \sum_{i=1}^n I(t_i^{\sim} \leq t),$$

여기서  $I(A)$ 는  $A$ 이면 1,  $A^c$ 이면 0인 지시함수이다.

제안된 추정량의 복잡한 구성으로 추정량의 점근적 분포와 소표본 특성에 대한 해석학적인 분석은 어렵기 때문에 추정량의 특성을 분석하기 위한 실질적인 방안의 하나로 오른쪽 관측중단률, 수명분포의 형태를 요인으로 하는 몬테칼로 모의실험을 통하여 생존함수에 대한 기존의 추정량과 제안된 추정량을 이용한 각각의 평균수명과 수명분포의 90백분위수의 추정량의 편의와 평균제곱오차의 행태를 분석하기로 한다.

### 3. 제안된 추정량의 특성 비교

#### 3.1 평균수명을 추정할 경우

제안된 생존함수의 추정량  $\hat{S}_{BJ 1}$ ,  $\hat{S}_{BJ 2}$ 를 이용하여 평균수명을 추정할 경우 이들 추정량의 특성을 기존의 생존함수의 추정량  $\hat{S}_{KM}$ ,  $\hat{S}_{BHK}$ ,  $\hat{S}_{WTAI L}$ 을 이용하는 경우와 비교 분석하기 위하여 이들 추정량을 이용하여 평균수명  $\mu$ 의 추정량을 각각 다음과 같이 구성하였다.

$$\hat{\mu}_- = \int_0^{\infty} \hat{S}_-(t) dt,$$

여기서,  $\hat{S}_- = \hat{S}_{BJ 1}, \hat{S}_{BJ 2}, \hat{S}_{KM}, \hat{S}_{BHK}, \hat{S}_{WTAI L}$ .

제안된 생존함수의 추정량  $\hat{S}_{BJ 1}$ ,  $\hat{S}_{BJ 2}$ 를 이용한 평균수명의 추정량  $\hat{\mu}_{BJ 1}$ ,  $\hat{\mu}_{BJ 2}$ 의 특성을 몬테칼로방법으로 편의와 평균제곱오차의 관점에서  $\hat{\mu}_{KM}$ ,  $\hat{\mu}_{BHK}$ ,  $\hat{\mu}_{WTAI L}$ 의 특성과 비교하기 위한 모의실험은 Moeschberger와 Klein(1985)에 소개된 미국의 국립독물학연구센터(National Center for Toxicological Research)의 동물 대상의 수명관측실험을 모형으로 하

여 수명분포가 평균수명이 400, 500, 600이고, 형상모수가 0.5, 1.0, 4.0인 와이블분포일 때, 표본의 크기는 48이고, 관측중단시간이 560인 관측중단 모형을 고려하였다. 각 수명분포로부터 수명  $T_i$ 를 크기  $n = 48$ 의 난수를 생성하고, 560에서 관측중단이 된다고 가정하여 관측 중단자료를 이용하여  $\hat{\mu}_{BJ 1}, \hat{\mu}_{BJ 2}, \hat{\mu}_{KM}, \hat{\mu}_{BHK}, \hat{\mu}_{WTAIL}$ 의 값을 구하는 과정을 10000회 반복 시행하여 구한 각 추정량의 평균제곱오차와 편의의 추정치를 표3.1에 정리하였다.

표3.1 평균수명의 추정량들에 대한 편의/100과 평균제곱오차/100<sup>2</sup>의 추정치

수명분포 weibull	평균	560에서 평균중단률(%)	KM	BHK	WTAIL	BJ1	BJ2
$\alpha = 0.5$	400	18.87	-2.104	-1.526	0.224	-1.406	0.199
			4.5218	2.6569	2.0179	2.3035	2.0526
	500	22.51	-2.893	-2.130	0.306	-1.993	0.281
$\alpha = 1.0$	400	24.78	-1.091	-0.068	0.074	0.009	0.065
			1.2724	0.4533	0.49547	0.4569	0.4847
	500	32.71	-1.736	-0.088	0.109	0.009	0.099
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004
			0.0371	0.0708	0.0297	0.0739	0.0274
	500	34.67	-0.383	1.571	0.016	1.598	0.014
$\alpha = 4.0$	400	7.56	-0.113	0.132	0.004	0.144	0.004</

$\hat{\mu}_{BHK}$ 가 가장 작았다.

(3) 위험률이 증가(increasing failure rate)하는 수명분포에서 평균수명을 추정할 때 고려된 추정량 가운데 추정량  $\hat{\mu}_{BJ 2}$ 는 편의와 평균제곱오차가 가장 작았다.

### 3.2 수명분포의 90백분위수를 추정할 경우

제안된 생존함수의 추정량  $\hat{S}_{BJ 1}, \hat{S}_{BJ 2}$ 를 이용하여 수명분포의 90백분위수를 추정할 경우 이들 추정량의 특성을 기존의 생존함수의 추정량  $\hat{S}_{KM}, \hat{S}_{BHK}, \hat{S}_{WTAILE}$ 를 이용하는 경우와 비교·분석하기 위하여 이들 추정량을 이용하여 수명분포의 90백분위수  $F^{-1}(0.9)$ 의 추정량을 각각 다음과 같이 구성하였다.

$$\hat{F}^{-1}(0.9) = \inf\{t | \hat{F}_-(t) \geq 0.9\},$$

여기서  $\hat{F}_-(t) = 1.0 - \hat{S}_-(t)$ 이고,  $\hat{S}_-$ 은  $\hat{S}_{BJ 1}, \hat{S}_{BJ 2}, \hat{S}_{KM}, \hat{S}_{BHK}$  혹은  $\hat{S}_{WTAILE}$ 이다.

제안된 생존함수의 추정량  $\hat{S}_{BJ 1}, \hat{S}_{BJ 2}$ 를 이용한  $F^{-1}(0.9)$ 의 추정량  $\hat{F}_{BJ 1}^{-1}, \hat{F}_{BJ 2}^{-1}$ 의 특성을 몬테칼로방법으로 편의와 평균제곱오차의 관점에서  $\hat{F}_{KM}^{-1}, \hat{F}_{BHK}^{-1}$ 의  $\hat{F}_{WTAILE}^{-1}$ 의 특성과 비교하기 위한 모의실험은 평균수명의 추정을 위한 모의실험에서와 같이 수명분포가 평균수명이 400, 500, 600이고, 형상모수가 0.5, 1.0, 4.0인 와이블분포일 때, 표본의 크기는 48로 하되, 수명분포의 80백분위수보다 큰 수명중 3/4은 85백분위수에서, 1/4은 80백분위수에서 관측이 중단된 관측중단모형을 고려하였다.

각 수명분포로부터 수명  $T_i$ 를 크기  $n = 48$ 의 난수를 생성하고, 편의상 각 수명분포의 85백분위수보다 큰 수명은 85백분위수에서 관측이 중단되고 남은 수명은 80백분위수에서 관측이 중단된 것으로 가정한 관측중단자료를 이용하여 추정량  $\hat{F}_{BJ 1}^{-1}, \hat{F}_{BJ 2}^{-1}, \hat{F}_{KM}^{-1}, \hat{F}_{BHK}^{-1}, \hat{F}_{WTAILE}^{-1}$ 의 값을 구하는 과정을 10000회 반복 시행하여 구한 각 추정량의 평균제곱오차와 편의의 추정치를 표3.2에 정리하였다.

수명분포의 90백분위수의 추정량들에 대한 모의실험의 결과를 통하여 다음과 같은 사실을 알 수 있었다.

(1) 위험률이 감소(decreasing failure rate)하는 수명분포에서 고려된 추정량 가운데 편의는  $\hat{F}_{WTAILE}^{-1}$ 이 가장 작게 나타났으며 평균제곱오차는  $\hat{F}_{BJ 1}^{-1}$ 이 가장 작았다.

(2) 위험률이 상수(constant failure rate)인 수명분포에서 고려된 추정량 가운데 편의는  $\hat{F}_{BJ 1}^{-1}$ 이 가장 작게 나타났으며 평균제곱오차는  $\hat{F}_{BHK}^{-1}$ 가 가장 작았다.

(3) 위험률이 증가(increasing failure rate)하는 수명분포에서 고려된 추정량 가운데  $\hat{F}_{BJ 2}^{-1}$ 는 편의가 가장 작게 나타났으며  $\hat{F}_{WTAILE}^{-1}$ 은 평균제곱오차가 가장 작게 나타났다.

결론적으로 극단적인 오른쪽관측중단 모형에서 Kaplan과 Meier의 추정량을 수정하여 Brown, Hollander와 Korwar(1974)와 Moeschberger와 Klein(1985)이 제안한 생존함수에 대한 추정량은 관측중단이 일어난 자료값을 관측중단시점보다 크다는 조건이 주어진 수명의 조건부 기대값으로 대체하여 이용하는 Buckley와 James 형 추정량을 이용하여 수정함

로써 수 명분포의 위험률 유형에 따라 평균수명이나 수명분포의 백분위수를 추정하는 경우에 편의와 평균제곱오차를 줄일 수 있다는 사실을 확인 할 수 있었다.

표3.2 90백분위수의 추정량들에 대한 편의/100과 평균제곱오차/ 100<sup>2</sup>의 추정치

수명분포 weibull	평균	KM	BHK	WTAILE	BJ1	BJ2
$\alpha = 0.5$	400	-6.004	-3.950	1.840	-3.167	4.089
		36.3051	17.5059	23.5881	12.5381	50.1567
	500	-7.525	-4.950	2.332	-3.982	5.111
		57.0227	27.4753	37.2002	19.7324	78.2168
	600	-9.006	-5.925	2.756	-4.752	6.127
		81.6865	39.3883	53.0200	28.2266	112.8057
$\alpha = 1$	400	-3.152	-0.448	0.614	0.280	0.648
		10.0578	2.8569	3.4686	3.6612	4.7722
	500	-3.940	-0.560	0.758	0.358	0.815
		15.7153	4.4639	5.3914	5.6938	7.4153
	600	-4.736	-0.671	0.923	0.418	0.972
		22.7028	6.4307	7.8117	8.1987	10.6950
$\alpha = 4$	400	-0.543	1.648	0.072	2.053	0.009
		0.3001	4.3272	0.0673	6.2389	0.0893
	500	-0.680	2.069	0.090	2.577	0.013
		0.4710	6.8033	0.1049	9.7811	0.1382
	600	-0.817	2.488	0.110	3.096	0.015
		0.6799	9.8273	0.1519	14.1029	0.1990

### 참고문헌

1. Breslow, N. E. and Crowley, J. (1974). A Large Sample Study of the Life Table and Product Limit Estimators under Random Censorship, *The Annals of Statistics*, Vcl. 2, 435-453.
2. Brown, B. W., Jr., Hollander, M., and Korwar, R. M. (1974). Nonparametric Tests of Independence for Censored Data, with Applications to Heart Transplant Studies. *In Reliability and Biometry: Statistical Analysis of Lifelength*, F. Proschan and R. J. Serfling (eds), 291-302. Philadelphia: Society for Industrial and Applied Mathematics.

3. Buckley, J. and James, I. (1979). Linear Regression with Censored Data, *Biometrika*, Vol. 66, 429-436.
4. Fleming, T. R. and Harrington, D. P. (1984). Nonparametric Estimation of the Survival Distribution in Censored Data, *Communications in Statistics-Theory and Methods*, Vol. 13, 2469-2486.
5. Kaplan, E. L. and Meier, P.(1958). Nonparametric Estimation from Incomplete Sample, *Journal of the American Statistical Association*, Vol.53, 457-481.
6. Lagakos, S. W. (1979) General Right Censoring and its Impact on the Analysis of Survival Data, *Biometrics*, 35, 139-156.
7. Miller, R. G.(1981). *Survival Analysis*, John Wiley & Sons, Inc., New York.
8. Moeschberger, M. L. and Klein, J. P. (1985). A Comparison of Several Methods of Estimating the Survival Function when There Is Extreme Right Censoring, *Biometrics*, Vol. 41, 253-259.
9. Nelson, W. B.(1972). Theory and Applications of Hazard Plotting for Censored Failure Data, *Technometrics*, Vol.14, 945-996.



## Estimation of the Survival Function under Extreme Right Censoring Model <sup>3</sup>

Jae Man Lee <sup>4</sup>

### Abstract

In life-testing experiments, in which the longest time an experimental unit is on test is not a failure time, but rather a censored observation. For the situation the Kaplan-Meier estimator is known to be a biased estimator of the survival function. Several modifications of the Kaplan-Meier estimator are examined and compared with bias and mean squared error.

*Key Words and Phrases:* Extreme right censoring model, Survival function, Life-Testing.

---

<sup>3</sup>This work was supported by Andong National University Grant, 1999

<sup>4</sup>Associate Professor, Dept. of Statistics, Andong National University, Kyungbuk 760-749