응용통계연구 제 11권 1호, 1998년, pp. 151-161

KOZIOL-GREEN 모형에서 생존함수에 대한 붓스트랩 구간추정 *

조 길 호 [†] 정 성 화 [‡] 최 달 우 [§] 채 현 숙 [¶]

요 약

본 논문에서는 Koziol-Green 모형에서 생존함수에 대한 신뢰구간을 붓스트랩 방법을 이용하여 제안하고, 생존함수에 대한 붓스트랩 추정량의 일치성을 밝힌다. 또한 제안된 붓스트랩 신뢰구간들을 기존의 근사적 정규분포를 이용한 신뢰구간과 생존함수에 변수변환을 고려하여 구성한 신뢰구간들과 모의실험을 통하여 비교한 결과 제안된 붓스트랩 신뢰구간이 기존의 방법보다 포함확률 측면에서 더 좋은 결과를 보였고 중도절단율에 덜 민감하다는 것을 보여 주었다.

1. 서 론

$X_1, X_2, \cdots, X_n$은 서로 독립이고 연속인 분포함수 F를 가지는 수명시간을 나타내는 확률변수이고, $Y_1, Y_2, \cdots, Y_n$ 역시 서로 독립이며 연속인 분포함수 G를 가지는 중도절단시간을 나타내는 확률변수라고 하자. 임의중도절단모형을 고려하면 수명시간 X_i는 중도절단시간 Y_i에 의해 오른쪽에서 중도절단되기 때문에 우리가 실제로 관측할 수 있는 변수는 (T_i, δ_i), $i = 1, 2, \cdots, n$이며, 여기서

$$
\begin{aligned}
T_i &= X_i \wedge Y_i, \\
\delta_i &= \begin{cases} 1 & \text{만약 } X_i \le Y_i, \\ 0 & \text{만약 } X_i > Y_i \end{cases}
\end{aligned}
$$

이고, $\wedge$는 X_i와 Y_i 중 작은 값을 나타내는 표시이다. 또한 중도절단시간 Y_i와 수명시간 X_i는 서로 독립이라고 가정한다. 따라서 관측된 T_i의 생존함수는 수명시간 X_i와 중도절단시간 Y_i의 생존함수인 $(1 - F(x))$와 $(1 - G(x))$의 곱으로 표현될 수 있다. 즉, $H(x) = (1 - F(x))(1 - G(x))$.

Koziol과 Green(1976)은 임의중도절단모형에서 중도절단시간의 생존함수 S_G가 수명시간의 생존함수 S_F의 β승으로 표현되는 비례위험모형을 생각하여 적합도 검정을 위한 통계

*이 논문은 1997년도 경북대학교 공모과제 연구비에 의해 지원되었음.
† (702-701) 대구광역시 북구 산격동 1370, 경북대학교 통계학과 부교수
‡ (702-701) 대구광역시 북구 산격동 1370, 경북대학교 통계학과 박사과정
§ (702-701) 대구광역시 북구 산격동 1370, 경북대학교 통계학과 시간강사
¶ (702-701) 대구광역시 북구 산격동 1370, 경북대학교 통계학과 시간강사

량을 제시하고, 이 통계량의 점근적 성질을 유도하였다. Koziol-Green 모형이라 명명되는 이 모형에서 β는 중도절단모수(censoring parameter)로 $E(\delta_i) = \frac{1}{1+\beta} = \alpha$가 되어 α는 중도절단되지 않은 관측치의 기대비율(expected proportion)로 생각할 수 있고, 만약 F나 G가 연속이라면 관측되는 T_i와 δ_i가 서로 독립이 된다는 사실은 쉽게 증명할 수 있다.

Abudushukurov(1984)와 Cheng과 Lin(1984)은 Koziol-Green 모형에서 생존함수에 대한 추정량(이하 ACL 추정량)을 제안하고, 그 추정량에 대한 점근적 성질을 유도하였다. Jie(1990)는 Koziol-Green 모형에서 ACL 추정량이 근사불편추정량(asymptotic unbiased estimator)이 되어 정규분포로 근사됨을 밝히고, 이 근사분포를 이용하여 생존함수에 대한 근사신뢰구간를 구성하였다.

Bogan과 Liestøl(1990)은 생존함수에 대한 로그-로그(log-log)와 아크사인 제곱근(arcsin square root)변환을 고려한 신뢰구간을 구성하여 정규근사신뢰구간과 비교한 결과, 오른쪽에서 중도절단된 생존자료의 경우 표본의 크기가 작을 때(≤ 25), 50%의 중도절단율을 가지더라도 변수변환한 신뢰구간은 정규근사신뢰구간보다 더 만족한 결과를 가져온다고 보고하였다.

한편 Efron(1979)에 의해 처음으로 소개된 붓스트랩 방법은 그 후 Efron(1981, 1982, 1987)에 의해 통계량에 대한 붓스트랩 추정량의 표본분포를 사용하여 붓스트랩 근사신뢰구간을 구성하는데 이용되었다.

Csörgö(1988)는 비례위험모형에서 붓스트랩 방법을 이용하여 생존함수에 대한 추정을 연구하였고, Dilta와 Ghorai(1990)는 Csörgö가 고려한 붓스트랩 재표본(bootstrap resampling) 방법과는 달리 Koziol-Green 모형에서는 T_i와 δ_i가 서로 독립이 된다는 사실을 이용한 붓스트랩 재표본 방법을 제안하여 생존함수에 대한 추정문제를 연구하였다.

본 논문에서는 Koziol-Green 모형에서 Dilta와 Ghorai가 제안한 붓스트랩 재표본 방법을 이용하여 추정한 생존함수에 대한 붓스트랩 추정량의 강일치성(strong consistency)을 밝히고, 그 붓스트랩 추정량의 분포함수를 이용하여 구성한 생존함수에 대한 붓스트랩 신뢰구간인 백분위(percentile), 편향정정(Bias corrected : 이하 BC) 그리고 편향정정가속(Bias corrected and accelerated : 이하 BCa) 신뢰구간을 제안한다. 그리고 제안된 붓스트랩 신뢰구간들을 생존함수에 대한 ACL 추정량의 정규근사를 이용하여 구성한 정규근사신뢰구간과 생존함수에 로그-로그와 아크사인 제곱근 변환을 고려하여 구성한 신뢰구간들과 몬테칼로 모의실험을 통하여 비교하고자 한다.

2. 생존함수에 대한 구간추정

Koziol-Green 모형에서 (T_i, δ_i), $i = 1, 2, \cdots, n$가 관측될 때 Abudushukurov(1984)와 Cheng과 Lin(1984)은

$$\alpha_n = \frac{1}{n}\sum_{i=1}^{n}\delta_i$$

과

$$H_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(T_i > t)$$

을 각각 정의하여

$$\hat{S}_n(t) = [H_n(t)]^{\alpha_n}$$

을 생존함수에 대한 추정량으로 제안하였다.

LEMMA 2.1 (Jie(1990)) Koziol-Green 모형에서 생존함수에 대한 ACL 추정량 $\hat{S}_n(t)$는 임의의 $t \geq 0$에 대하여 생존함수 $S_F(t)$에 대한 근사불편추정량이다.

LEMMA 2.2 (Jie(1990)) Koziol-Green 모형에서 F가 연속이라 가정하면, 임의의 $t \geq 0$에 대하여

$$\frac{\sqrt{n}(\hat{S}_n(t) - S_F(t))}{\sigma_n(t)} \xrightarrow{d} N(0,1), \qquad n \to \infty$$

이다. 여기서

$$\sigma_n{}^2(t) = \hat{S_n}^2(t)\left[\alpha_n{}^2\frac{1 - H_n(t)}{H_n(t)} + \hat{S_n}^2(t)\frac{1-\alpha_n}{\alpha_n}\right]$$

이다.

한편 Koziol-Green 모형에서 생존함수에 대한 붓스트랩 추정량은 관측되는 T_i와 δ_i가 서로 독립이 된다는 사실을 이용하여 다음과 같은 절차를 따르는 붓스트랩 재표본 방법에 의해 얻어진다.

I. 붓스트랩 표본 $(T_i{}^*, \delta_i{}^*)$, $i = 1, 2, \cdots, n$ 을 구성

여기서 $T_i{}^*$는 관측된 T_i들의 경험분포함수(empirical distribution function) $1 - H_n$에서 추출된 확률표본들이고, $\delta_i{}^*$는 성공의 확률을 α_n으로 가지는 베르누이분포에서의 확률표본들이다.

II. $\alpha_n{}^*$와 $H_n{}^*(t)$를 계산

$$\alpha_n{}^* = \frac{1}{n}\sum_{i=1}^{n}\delta_i{}^*$$

$$H_n{}^*(t) = \frac{1}{n}\sum_{i=1}^{n} I(T_i{}^* > t)$$

III. 생존함수에 대한 붓스트랩 추정량을 계산

$$\hat{S_n}^*(t) = [H_n{}^*(t)]^{\alpha_n{}^*}$$

정리 2.1 Koziol-Green 모형에서 붓스트랩 재표본에 의해 구성된 $\alpha_n{}^*$와 $H_n{}^*(t)$, $t \geq 0$, 는 각각 확률 1로 α와 $H(t)$에 수렴한다.

증명: $|\alpha_n^* - \alpha| \le |\alpha_n^* - \alpha_n| + |\alpha_n - \alpha|$이고, 임의의 $t \ge 0$에 대하여 $|H_n^*(t) - H(t)| \le |H_n^*(t) - H_n(t)| + |H_n(t) - H(t)|$이다. 한편 $E(\delta_i) = \Pr(\delta_i = 1) = \frac{1}{1+\beta} < \infty$이고, $E(I(T_i > t)) = \Pr(T_i > t) = S_F{}^{1+\beta}(t) < \infty$이므로, 강대수법칙(strong law of large number)에 의해 $n \to \infty$일 때, $\alpha_n \xrightarrow{a.s.} \alpha$이고, $H_n(t) \xrightarrow{a.s.} H(t)$이다. 한편 δ_i^*는 성공의 확률을 α_n으로 가지는 베르누이분포의 확률표본들이고, T_i^*는 관측된 T_i들의 경험분포 $1 - H_n(t)$에서 추출된 확률표본들이므로 $E(\delta_i^*) = \alpha_n < \infty$이고, $E(I_{(T_i^* > t)}) = \Pr(T_i^* > t) = H_n(t) < \infty$가 되어, 강대수법칙에 의해 $n \to \infty$일 때, 각각 $\alpha_n^* \xrightarrow{a.s.} \alpha_n$이고, $H_n^*(t) \xrightarrow{a.s.} H_n(t)$이 된다. 따라서 $n \to \infty$일 때, $\alpha_n^* \xrightarrow{a.s.} \alpha$이고, $H_n^*(t) \xrightarrow{a.s.} H(t)$이다. □

정리 2.2 Koziol-Green 모형에서 임의의 $t \ge 0$에 대하여 $\hat{S_n}^*(t)$를 생존함수 $S_F(t)$의 붓스트랩 추정량이라 하면, $\hat{S_n}^*(t)$는 $S_F(t)$에 대한 강일치(strong consistent) 추정량이다.

증명: 임의의 $t \ge 0$에 대하여, $|\hat{S_n}^*(t) - S_F(t)| \le |\hat{S_n}^*(t) - \hat{S}_n(t)| + |\hat{S}_n(t) - S_F(t)|$이고, Lemma 2.1과 정리 2.1에서 $\hat{S}_n(t) \xrightarrow{a.s.} S_F(t)$이고, $\hat{S_n}^*(t) \xrightarrow{a.s.} \hat{S}_n(t)$이라는 사실을 증명할 수 있다. 따라서 $n \to \infty$일 때, $\hat{S_n}^*(t) \xrightarrow{a.s.} S_F(t)$이다. □

2.1. 생존함수의 정규근사 신뢰구간

Koziol-Green 모형에서 ACL 추정량의 정규근사를 이용하면 생존함수 $S_F(t)$에 대한 $100(1-\alpha)\%$ 정규근사 신뢰구간은 다음과 같이 구성될 수 있다.

$$\left(\hat{S}_n(t) - z_{\alpha/2}\frac{\sigma_n(t)}{\sqrt{n}},\ \hat{S}_n(t) + z_{\alpha/2}\frac{\sigma_n(t)}{\sqrt{n}}\right)$$

여기서 $z_{\alpha/2}$는 표준정규분포의 $1 - \frac{\alpha}{2}$ 임계값이다.

한편 소표본에서 더 만족한 결과를 가져올 수 있도록 Bogan과 Liestøl(1990)이 고려한 생존함수에 $\log(-\log(x))$, $\arcsin\sqrt{x}$ 변환을 이용하여 구성한 신뢰구간은 생존함수 추정량의 근사분산(asymptotic variance)과 델타방법(Delta-method)을 이용하여 구성할 수 있다. 각각의 변환을 이용한 생존함수의 $100(1-\alpha)\%$ 신뢰구간은

$$\left(\hat{S_n}(t)^{\exp\left\{z_{\alpha/2}\frac{1}{\hat{S_n}(t)\log\hat{S_n}(t)}\frac{\sigma_n(t)}{\sqrt{n}}\right\}},\ \hat{S_n}(t)^{\exp\left\{-z_{\alpha/2}\frac{1}{\hat{S_n}(t)\log\hat{S_n}(t)}\frac{\sigma_n(t)}{\sqrt{n}}\right\}}\right),$$

과

$$LCL \le S_F(t) \le UCL$$

으로 구성된다. 여기서

$$LCL = \sin^2\left\{\max\left(0, \arcsin\sqrt{\hat{S_n}(t)} - z_{\alpha/2}\frac{\sigma_n(t)}{2\sqrt{n}}\left[\frac{1}{\hat{S_n}(t)(1-\hat{S_n}(t))}\right]^{1/2}\right)\right\}$$

이고,

$$UCL = \sin^2\left\{\min\left(\frac{\pi}{2}, \arcsin\sqrt{\hat{S_n}(t)} + z_{\alpha/2}\frac{\sigma_n(t)}{2\sqrt{n}}\left[\frac{1}{\hat{S_n}(t)(1-\hat{S_n}(t))}\right]^{1/2}\right)\right\}$$

이다.

2.2. 생존함수에 대한 붓스트랩 신뢰구간

붓스트랩 신뢰구간은 생존함수에 대한 붓스트랩 추정량의 경험분포함수를 이용하여 백분위, BC, 그리고 BCa 방법 등으로 구성할 수 있다.

2.2.1. 붓스트랩 백분위 신뢰구간

붓스트랩 백분위 신뢰구간은 붓스트랩 추정량들의 경험분포함수(empirical distribution function)를 이용하여 구성할 수 있다. $\hat{H}$을 생존함수의 붓스트랩 추정량 $\hat{S_n}^*(t)$의 경험분포함수라 하면, $\hat{H}(t_1)$는

$$\hat{H}(t_1) = \frac{1}{B}\sum_{b=1}^{B} I\left(\hat{S_n}^{*b}(t) \le t_1\right)$$

로 구성된다. 여기서 t와 t_1은 임의의 실수값이고 B는 붓스트랩 반복횟수이다. 그리고 $\hat{H}^{-1}(\alpha)$를 $\hat{S_n}^{*b}(t)$의 100α 백분위수 (percentile)라 하면

$$\hat{H}^{-1}(\alpha) = \inf\left\{t_1 : \hat{H}(t_1) \ge \alpha\right\}$$

이다. 즉 $\hat{H}^{-1}(\alpha)$는 붓스트랩 반복을 B번하여 구성한 생존함수의 붓스트랩 추정량을 크기순으로 나열할 때 $B\alpha$번째의 값이 된다. 따라서 붓스트랩 추정량들의 백분위수를 사용하여 구성한 생존함수 $S_F(t)$에 대한 $100(1-\alpha)\%$ 붓스트랩 백분위 신뢰구간은

$$\left(\hat{H}^{-1}\left(\frac{\alpha}{2}\right),\ \hat{H}^{-1}\left(1-\frac{\alpha}{2}\right)\right)$$

으로 계산된다.

2.2.2. BC 방법

BC 방법은 붓스트랩 표본으로 추정한 생존함수에 대한 붓스트랩 추정량이 주어진 자료에서 계산된 통계량에 대하여 편향(bias)되어 있기 때문에 이런 편향을 정정하여 신뢰구간을 구성하는 방법이다. 편향정정상수 $\hat{z}_0$은 정규분포의 누적분포함수를 이용한 중심화(centering)에 의해 다음과 같이 구해진다.

$$\begin{aligned}\hat{z}_0 &= \Phi^{-1}\left\{\hat{H}(\hat{S_n}(t))\right\}\\ &= \Phi^{-1}\left\{\frac{1}{B}\sum_{b=1}^{B} I(\hat{S_n}^{*b} < \hat{S_n}(t))\right\}\end{aligned}$$

여기서 $\Phi^{-1}(\cdot)$는 표준정규분포의 누적분포함수의 역함수이다. 따라서 $\hat{z}_0$은 정규분포상에서 $\hat{S_n}^{*b}(t)$의 중앙값과 $\hat{S_n}(t)$ 사이의 불일치 정도를 나타내는 측도로, $\hat{S_n}^{*b}(t)$들 중에서 정

확히 반이 $\hat{S_n}(t)$보다 작거나 같은 경우에 $\hat{z_0} = 0$이 된다. 그러므로 편향정정상수 $\hat{z_0}$을 이용하여 구성한 생존함수에 $S_F(t)$에 대한 $100(1-\alpha)\%$ BC 신뢰구간은

$$\left(\hat{H}^{-1}(\alpha_1),\ \hat{H}^{-1}(\alpha_2)\right)$$

로 구해진다. 여기서, $\alpha_1 = \Phi\left(2\hat{z_0} + z_{\alpha/2}\right)$이고, $\alpha_2 = \Phi\left(2\hat{z_0} + z_{1-\alpha/2}\right)$이다.

2.2.3. BCa 방법

BCa 방법은 $\hat{S_n}$의 편향과 표준오차를 동시에 정정하는 방법으로, BCa 신뢰구간을 구성하기 위해서는 편향정정상수 $\hat{z_0}$과 가속(acceleration)상수 $\hat{a}$을 계산하여야 한다. $\hat{z_0}$은 BC 방법에서와 같이 계산할 수 있고 가속상수 $\hat{a}$은 여러 가지 방법으로 계산할 수 있으나 본 논문에서는 잭나이프 통계량을 이용하여 $\hat{a}$을 계산하기로 한다. $T_1, T_2, \cdots, T_n$의 관측값이 주어질 때 생존함수 $S_F(t)$의 의사잭나이프(pseudo-jackknife) 통계량을 $\hat{S_{n(i)}}(t)$라 하면, 가속상수 $\hat{a}$은 다음과 같이 계산할 수 있다.

$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\hat{S_{n(\cdot)}}(t) - \hat{S_{n(i)}}(t)\right)^3}{6\left\{\sum_{i=1}^{n}\left(\hat{S_{n(\cdot)}}(t) - \hat{S_{n(i)}}(t)\right)^2\right\}^{3/2}}$$

여기서 $\hat{S_{n(\cdot)}}(t) = \frac{1}{n}\sum_{i=1}^{n}\hat{S_{n(i)}}(t)$ 이다. 따라서 편향정정상수 $\hat{z_0}$과 가속상수 $\hat{a}$을 이용하여 구성한 생존함수에 $S_F(t)$에 대한 $100(1-\alpha)\%$ BCa 신뢰구간은 다음과 같다.

$$\left(\hat{H}^{-1}(\alpha_3),\ \hat{H}^{-1}(\alpha_4)\right)$$

여기서,

$$\alpha_3 = \Phi\left(\hat{z_0} + \frac{\hat{z_0} + z_{\alpha/2}}{1 - \hat{a}(\hat{z_0} + z_{\alpha/2})}\right)$$

$$\alpha_4 = \Phi\left(\hat{z_0} + \frac{\hat{z_0} + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z_0} + z_{1-\alpha/2})}\right)$$

이다. 만약 $\hat{z_0}$과 $\hat{a}$이 0이라면, $\alpha_3 = z_{\alpha/2}$이고, $\alpha_4 = z_{1-\alpha/2}$가 되어 BCa 신뢰구간은 붓스트랩 백분위 신뢰구간과 일치하게 된다.

3. 모의실험

제안된 생존함수에 대한 신뢰구간들의 정도(precision)를 포함확률(coverage probability)과 신뢰구간의 평균길이(average length) 측면에서 비교하기 위해 몬테칼로 모의실험을 시행하였다. Koziol-Green 모형을 가정하기 위하여 수명시간 X_i와 중단시간 Y_i는 지수분포(exponential distribution)에서 추출된 임의표본들로 표본의 크기가 20, 30, 50 그리고 100인 경우에, 지수분포의 모수를 조정하여 중단율이 10%, 30% 그리고 50%가 되도록 각각 구성하였다. 일반적으로 생존함수의 추정량들은 생존함수의 평균 부근에서는 잘 적합되

나 끝부분으로 갈수록 민감하다고 알려져 있어, 생존함수의 신뢰구간을 생존함수의 25백분위수, 50백분위수 그리고 75백분위수에서 각각 구성하여 비교하였다. 붓스트랩 신뢰구간을 구성하기 위하여 붓스트랩 반복을 2000번하였고, 포함확률과 신뢰구간의 평균길이를 구하기 위하여 모의실험 반복을 1000번씩 시행하였다. 그리고 신뢰계수(confidence level) $1-\alpha$는 0.90을 고려하였다.

모의실험 결과를 그림 1 - 그림 6으로 나타내었는데 이를 살펴보면 다음의 몇 가지 사실을 알 수 있다. 전체적으로 중단율이 높을수록 생존함수에 대한 신뢰구간들의 포함확률이 신뢰계수 0.90에서 점점 멀어지는 경향이 있는데, 이런 경향은 붓스트랩 신뢰구간들 보다 정규근사신뢰구간과 변수변환을 통하여 구성한 신뢰구간에서 훨씬 더 두드러지게 나타난다. 그리고 생존함수의 50백분위수에서의 신뢰구간들은 정규근사신뢰구간과 변수변환을 이용한 신뢰구간 그리고 붓스트랩 신뢰구간들이 포함확률 측면에서 거의 차이가 없으나, 25백분위수나 75백분위수에서 구성한 신뢰구간들을 비교해 보면 붓스트랩 신뢰구간들이 포함확률 측면에서 다른 신뢰구간들 보다 나은 정도를 보여주고 있다. 그러나 신뢰구간의 평균길이 측면에서는 대부분의 경우에서 붓스트랩 방법을 이용한 신뢰구간들이 다른 신뢰구간들 보다 조금 넓게 구성되는 경향을 볼 수 있으나 그 차이가 큰 것은 아니었다. 한편 표본의 크기가 증가할수록 모든 신뢰구간들이 신뢰계수 0.90에 수렴하고, 신뢰구간의 평균길이는 점점 짧아지는 경향을 나타내어 제안한 방법들의 타당성을 뒷받침 해주었다.

4. 결 론

본 논문에서는 Koziol-Green 모형에서 생존함수에 대한 붓스트랩 구간추정 문제를 고려하였다. 일반적으로 정규근사신뢰구간은 중단율이 크거나 표본의 크기가 작으면 만족할 만한 결과를 가져오지 못한다. 이런 문제를 해결하기 위하여 생존함수에 변수변환을 고려하여 신뢰구간을 구성하는 방법이 연구되었다. 본 논문에서는 다른 해결책의 하나로 붓스트랩을 이용하여 구성한 붓스트랩 신뢰구간을 제안하고, 이들 붓스트랩 신뢰구간의 정도를 다른 신뢰구간들과 비교하였으며, 생존함수에 대한 붓스트랩 추정량의 강일치성을 밝혔다.

몬테칼로 모의실험으로 제안한 붓스트랩 신뢰구간과 정규근사신뢰구간 그리고 생존함수에 변수변환하여 구성한 신뢰구간들을 비교한 결과, 표본의 크기가 작고 중단율이 큰 경우에, 정규근사신뢰구간과 변수변환을 이용하여 구성한 신뢰구간보다 붓스트랩 방법을 이용하여 구성한 신뢰구간이 더 좋은 결과를 가져다주었고, 붓스트랩 신뢰구간이 다른 신뢰구간들에 비해 중단율에 덜 민감하다는 것을 볼 수 있었다. 특히 모든 경우에서 BC 와 BCa 신뢰구간이 잘 적합되었다. 한편 신뢰구간의 평균길이 측면에서는 대부분의 경우에서 정규근사신뢰구간이 붓스트랩 방법을 이용한 신뢰구간 보다 조금 짧게 구성되었으나 그 차이는 미미하였다.

참고문헌

[1] Abdushukurov, A. A. (1984). On some estimates of the distribution function under random censorship, In : Conference of Young Scientists, Math. Inst. Acad. Sci. Uzbeck SSR, Tashkent. VINITI No. 8756-V (in Russian).

[2] Borgan, Ø. and Liestøl, K. (1990). A note on confidence intervals and bands for the survival curve based on transformations, *Scandinavian Journal of Statistics*, **17**, 35-41.

[3] Cheng, P. E. and Lin, G. D. (1984). Maximum likelihood estimation of survival function under the Koziol-Green proportional hazard model, Technical Report B-34-5, Institute of Statistics, Academia Sinica, Taipei, Taiwan.

[4] Csörgö, S. (1988). Estimation in the proportional hazards model of random censorship, *Statistics*, **19**, 437-463.

[5] Dilta, G. and Ghorai, J. K. (1990). Bootstrap approximation with censored data under the proportional hazard model, *Communication in statistics - Theory and Methods*, **19**, 573-581.

[6] Efron, B. (1979). Bootstrap methods : Another look at the jackknife, *The Annals of Statistics*, **7**, 1-26.

[7] Efron, B. (1981). Nonparametric standard errors and confidence intervals(with discussion), *Canadian Journal of Statistics*, **9**, 139-172.

[8] Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM.

[9] Efron, B. (1987). Better bootstrap confidence intervals(with discussion), *Journal of the American Statistical Association*, **82**, 171-200.

[10] Koziol, J. A. and Green, S. B. (1976). A Cramer-von Mises statistic for randomly censored data, *Biometrika*, **63**, 465-474.

[11] Jie, M. (1990). Asymptotic results for the Koziol-Green model, *Communication in statistics - Theory and Methods*, **19**, 2767-2779.

[1997년 5월 접수, 1997년 11월 최종수정]

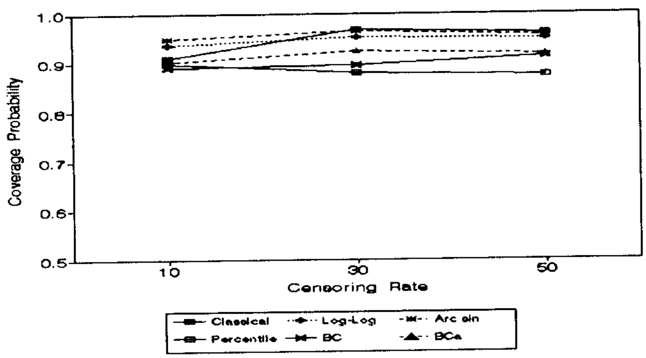

그림 1. 생존함수의 25백분위수에 대한 신뢰구간의 포함확률 (n=20)

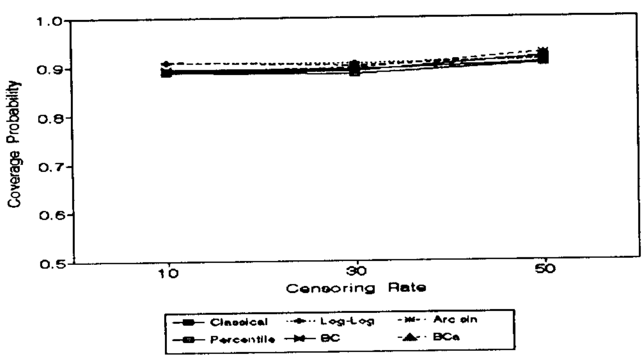

그림 2. 생존함수의 50백분위수에 대한 신뢰구간의 포함확률 (n=30)

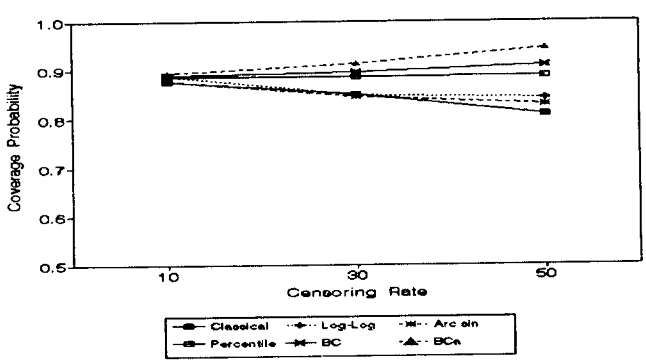

그림 3. 생존함수의 75백분위수에 대한 신뢰구간의 포함확률 (n=50)

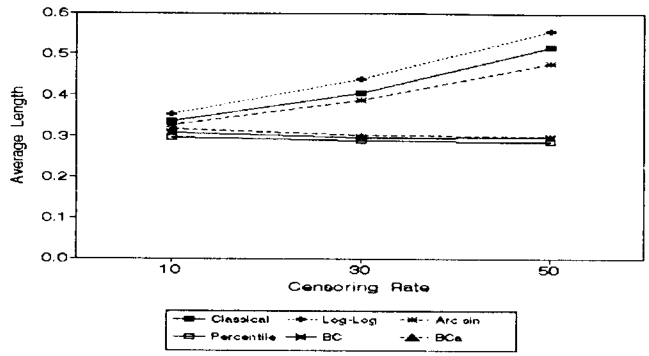

그림 4. 생존함수의 25백분위수에 대한 신뢰구간의 평균길이 (n=20)

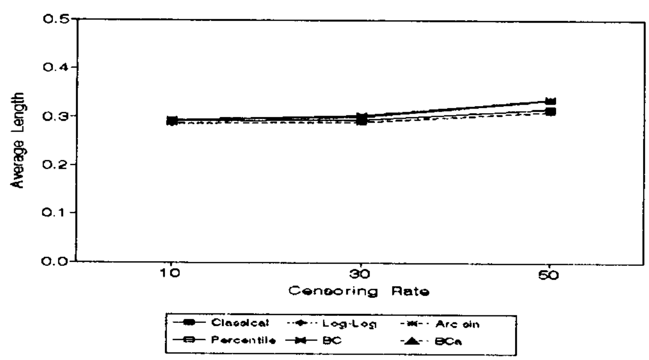

그림 5. 생존함수의 50백분위수에 대한 신뢰구간의 평균길이 (n=30)

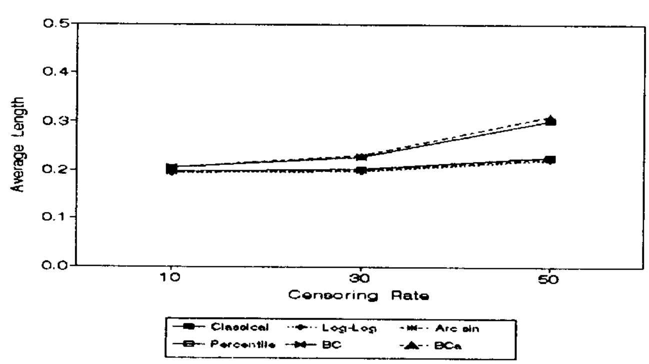

그림 6. 생존함수의 75백분위수에 대한 신뢰구간의 평균길이 (n=50)

Bootstrap Confidence Interval for Survival Function in the Koziol-Green Model *

Kil-Ho Cho †, Seong-Hwa Jeong ‡, Dal-Woo Choi §, Hyeon-Sook Chae ¶

ABSTRACT

We study the bootstrap interval estimation for survival function in the Koziol-Green model. We construct the approximate bootstrap confidence intervals for suvival function and prove the strong consistency for the bootstrap estimator of survival function. Finally we show that the approximate bootstrap confidence intervals are better in terms of coverage probability than confidence intervals based on asymptotic normal distribution and transformations of survival function via Monte Carlo simulation study.

*This paper was supported by Kyungpook National University Research Fund, 1997.
† Associate Professor, Department of Statistics, Kyungpook National University, Taegu 702-701, Korea.
‡ Ph. D. candidate, Department of Statistics, Kyungpook National University, Taegu 702-701, Korea.
§ Part-time Lecturer, Department of Statistics, Kyungpook National University, Taegu 702-701, Korea.
¶ Part-time Lecturer, Department of Statistics, Kyungpook National University, Taegu 702-701, Korea.