체계가용도의 붓스트랩 로버스트 추정

홍 연 웅
동양대학교 산업공학과

Abstract

The bootstrap procedure is suggested as a useful method for point and interval estimation of system availability. Its validity and robustness has been shown in special, but representative case, by various sampling experiments. Alternative to the bootstrap suggest themselves(e.g. a variation of the 'F' technique, but remain to be evaluated, as do variations on the bootstrap itself.

I. 서 론

컴퓨터, 자동차, 통신시스템, 원자력발전소의 안전장치 등과 같이 수리나 교환이 가능한 부품이나 체계의 성능을 나타내기 위해서는 신뢰도 뿐만 아니라 가용도(availability)등의 척도가 사용된다. 내구연한내에 체계에 고장이 발생하여 수리할 경우 체계는 가동하는 기간과 고장난 기간을 반복적으로 가지게되는데 가용도는 특정시점에서 체계가 정상적으로 가동할 확률로 정의된다. 즉 가용도는 특정시점까지의 총가동기간과 고장으로 인한 비가동기간의 합에 대한 총가동기간의 비율이다. 고장을 수리하는데 필요한 행정·관리시간을 제외하고 순수한 고장기간만을 비가동기간으로 고려할 때를 고유가용도(intrinsic availability)라한다. 수리가 허용되지 않는 경우 가용도는 신뢰도와 동일하다.

가용도에 영향을 주는 주요 요소들은 (a) 고장기간 및 가동기간분포의 특성 : 독립적 고장 또는 공통요인고장(common-cause failure) (b) 수리정책, 수리부품의 성능, 대기부품의 가용도, 진단장치 및 기술자의 능력, (c) 시스템의 구조, (d) 예방정비 및 검사정책 등이 있다. 위의 요인들을 고려한 많은 연구가 있었으나 대부분이 확률모형에 대한 연구가 중심이 되었다. 통계적인 관점에서는 가동기간과 고장기간에 대한 확률분포를 가정하고 최우추정법이나 베이즈 추정방법이 많이 이용되나 이들은 모형의존적인 방법으로 실제 데이타와 가정한 확률모형 사이의 괴리가 클 경우에 적합성이 결여되어 통계적 신뢰성이 저하된다.

Gray와 Lewis(1967)는 가동기간을 지수분포, 고장기간을 대수정규분포라 가정하고 가용도에 대한 신뢰구간을 구하였다. Graver와 Chu(1979)는 재표본방법(resampling)의 한 가지인 잭나이프(jackknife) 방법을 이용하여 단일부품체계와 병렬체계의 가용도에 대하여 추정하였다.

붓스트랩 방법의 개념은 이미 오래전부터 있었으나 계산능력의 부재로 실용적인 측면에서는 발전을 보지 못하였으나 Efron(1979)의 통찰력에 의해서 동 방법의 여러가지 이점이 밝혀지고 컴퓨터의 계산력 증대에 힘입어 최근에는 많이 응용되고있다. 이 방법은 주어진 표본에 근거하여 재표본(resample)을 취하므로 모형설정이 필요없다(model-free).

기본적인 생각은 원래의 표본 $X=(X_1,\cdots,X_n)$에 근거한 추축통계량 $R_n(X,F)$와 붓스트

랩표본 $X^* = (X_1^*, \cdots, X_n^*)$에 의한 추축통계량 $R_n(X^*, F_n)$가 일정한 조건하에서 유사할 것이라는 점이다. 따라서 관찰된 자료에 근거한 붓스트랩 분포를 이용하여 표본분포를 추정할 수 있다(대문자는 확률변수, 그에 대응하는 소문자는 확률변수의 실현치를 나타냄). 실제의 표본분포 $J_n(x, F)$의 붓스트랩추정값 $J_n(x, F_n)$는 미지의 분포 F를 표본 $X = (X_1, \cdots, X_n)$에 근거한 경험분포 F_n으로 대치한 함수값이다. 따라서 F를 반드시 F_n으로만 추정하여 대치할 필요는 없다. 평균이 θ인 Poisson모형을 예로들면, 표본분포 $J_n(x, \theta)$의 붓스트랩 추정치는 F를 비모수적으로 추정한 F_n보다는 θ를 그의 좋은 추정값인 θ_n 으로 대치한 $J_n(x, \theta_n)$이 더 타당해 보인다. 이와 같은 방법을 모수적 붓스트랩이라 한다.

본 연구에서는 체계의 가용도에 대하여 고장기간 및 가동기간의 분포가 (a) 지수분포를 따를 때 (b)지수분포와 감마분포를 따를 때 (c) 지수분포와 유사하나 꼬리가 긴 분포(h-분포)를 따를 경우에 대하여 포함확률(coverage probability), 평균신뢰구간 및 신뢰구간의 분산관점에서 비모수적 붓스트랩방법, 모수적 붓스트랩 방법, 잭나이프 방법 및 F 통계량을 이용한 방법을 비교한다. 2장에서는 비모수적 붓스트랩방법과 모수적 붓스트랩방법으로 가용도의 신뢰구간 등을 구하고, 3장에서는 전술한 세 가지 경우에 대하여 소표본모의 실험한 결과를 비교한다.

II. 붓스트랩 신뢰구간

단일부품체계에 대하여 U 를 가용기간, D를 고장 기간을 나타내는 확률변수라하고, E(U)와 E(D)가 존재한다고 가정하면, 가용도는

$$
\begin{aligned}
A &= \lim_{x \to \infty} P\{t \text{ 시간에 체계가 가동할 사건}\} \\
&= E(U)/\{E(U) + E(D)\} \qquad (1)
\end{aligned}
$$

으로 정의된다. 본 절에서는 비모수적 붓스트랩 방법 및 모수적 붓스트랩 방법으로 다음 세가지 경우에 대하여 모의실험을 통하여 잭나이프 방법 및 F 통계량을 이용한 결과를 비교한다. U와 D에 대한 자료가 충분하지 않는 경우 이들의 기저 분포를 설정하기가 용이하지 않다. 본 연구에서는 실제 상황과 유사하거나 현실적으로 발생가능성이 큰 범위 내에서 대표적인 세 가지를 다음과 같이 정하였다.

(a) U와 D는 서로 독립인 지수분포를 따르며 $E(U) = 1/\lambda$, $E(D) = 1/\mu$인 경우

(b) U는 지수분포, D는 감마분포를 따르며 서로 독립인 경우

단, $E(U) = 1/\lambda$, $E(D) = 1/ku$, $Var(D) = 1/k\lambda^2$, 단 k>1,

(c) X가 평균 1인 지수분포를 따를 때 $U = (1-h)^2/\lambda X e^{hX}$, $0 < h < 1$의 경우

여기서 (a)와 (c)의 경우 E(U)와 E(D)는 같지만 (c)의 꼬리가 두껍다. 이상의 세 가지 경우는 실제로 가동기간과 및 고장기간이 지수분포 또는 이와 유사한 분포를 따를 경우에 모형화하는데 이용될 수 있다.

2.1. 비모수적 붓스트랩 방법

가용기간과 고장기간의 관측치를 $u_i, i=1,2,\cdots$와 $d_i, i=1,2,\cdots$ 라고하면 확률분포를 가정하지 않았을 경우에 가용도의 점추정치는 $\hat{A}=\bar{u}/(\bar{u}+\bar{d})$이다. 이러한 추정치는 표본크기가 유한할 때 편의되며 신뢰한계나 오차 등을 구할 수 없으므로 통계적인 안정성의 확보에 문제가 있다. 본 절에서는 소표본에 대하여 다음과 같이 비모수적 붓스트랩 방법을 적용한다.

단계 1. 표본의 확률분포를 정한다: $(F_U(\cdot), F_D(\cdot))$.

단계 2. $(F_U(\cdot), F_D(\cdot))$에서 랜덤추출한 n쌍의 관찰값 $(u_1, d_1),\cdots,(u_n, d_n)$으로부터 $(F_{U,n}(\cdot), F_{D,n}(\cdot))$를 만든다.

단계 3. $(F_{U,n}(\cdot), F_{D,n}(\cdot))$으로부터 붓스트랩 표본 $(u_1^*, d_1^*),\cdots,(u_n^*, d_n^*)$를 얻고 상응하는 $\hat{A}^*=\sum u_i^*/\sum(u_i^*+d_i^*)$ 를 계산한다.

단계 4. 단계 3을 독립적으로 B회 반복하여 $\hat{A}_1^*,\cdots, \hat{A}_B^*$를 얻는다.

가용도에 대한 붓스트랩 편의와 분산의 추정치는

$$bias_B=\hat{A}^*-A, \qquad var_B=\sum_{b=1}^{B}(\hat{A}_b^*-A)^2/(B-1)$$

이 된다. 여기서 $\hat{A}^*=\sum_{b=1}^{B}\hat{A}_b^*/B$이다. 또한 $B=\infty$일때, 즉 이상적인 붓스트랩 분포는

$$G(s)=\#\{\hat{A}_b^*<s\}/B$$

로 추정되어지며 편의보정치(bias correction)는 $z_o=\Phi^{-1}[\hat{G}(\hat{A}(s))]$ 으로 얻어진다. 여기서 $\Phi(\cdot)$는 표준정규분포의 누적분포함수이다. 따라서 가용도 A에 대한 $100(1-2\alpha)\%$ 백분위 신뢰구간은

$$[\hat{G}^{-1}(\alpha),\ \hat{G}^{-1}(1-\alpha)] \tag{2}$$

으로 주어지며, 편의보정된 $100(1-2\alpha)\%$ 백분위 신뢰구간은

$$[\hat{G}^{-1}\Phi(2z_o-z_\alpha),\ \hat{G}^{-1}\Phi(2z_o+z_\alpha)] \tag{3}$$

이다. 여기서 $z_\alpha=\Phi^{-1}(\alpha)$이다.

2.2. 모수적 붓스트랩 방법

비모수적 방법의 단계2에서 $(F_{U,n}(\cdot), F_{D,n}(\cdot))$대신에 모수의 좋은 추정값을 사용하며 $(F_U(\cdot), F_D(\cdot))$를 만들고 단계 3과 단계 4를 시행한다. 모수적 붓스트랩방법을 이용

할 경우의 가용도 A에 대한 $100(1-2\alpha)\%$ 백분위 신뢰구간 및 편의보정백분위 신뢰구간은 식(2) 및 (3)과 동일한 형태로 주어진다.

Ⅲ. 모의 실험에 의한 비교

가용기간 U의 확률분포함수를 $F_U(\cdot)$, 고장기간 D의 확률분포함수를 $F_D(\cdot)$라고하면 전술한 방법에 의한 신뢰구간의 타당성을 포함확률의 관점에서 보이고자 한다. 공칭(nominal)포함확률을 0.95라하고 표본의 크기가 15및 25인 두 가지 경우에 대하여 붓스트랩 표본을 400회 씩 반복하는 실험을 1000번 시행하여 <표 Ⅲ-1> 및 <표 Ⅲ-2>를 얻었다. <표 Ⅲ-1>과 <표 Ⅲ-2>에서 볼 수 있듯이 붓스트랩 방법은 F분포를 이용한 근사방법이나 잭나이프 방법보다 우월함을 알 수 있다. 또한 표본의 크기가 크면 더욱 유효한 신뢰구간을 얻을 수 있었다.

<표 Ⅲ-1> 모의실험 결과 (양측 95% 신뢰구간)
n=15, 1000번 반복

기저분포	포함확률(%)	평균신뢰 구간길이	신뢰구간의 분산
가동기간과 고장기간이 지수분포를 따를 경우	JK : 95.0	1.94	10
	F : 94.4	1.63	4
	NP : 91.0	1.79	7
	NBC : 91.3	1.81	7
	P : 95.2	1.78	7
	BC : 95.3	1.80	7
가동기간은 지수분포, 고장기간은 감마분포를 따를 경우	JK : 94.1	1.39	3
	F : 97.7	1.64	3
	NP : 87.1	3.91	23
	NBC : 88.1	3.87	23
	P : 94.3	2.57	19
	BC : 94.5	2.60	19
가동기간은 꼬리긴 h분포, 고장기간은 지수분포를 따를 경우	JK : 92.4	2.42	18
	F : 88.0	1.75	7
	NP : 91.5	1.39	4.2
	NBC : 91.5	1.42	4.5
	P : 94.3	1.38	4.1
	BC : 94.3	1.39	4.2

JK : 잭나이프방법
NP : 비모수적 붓스트랩
NBC : 편의수정된 비모수적 붓스트랩
P : 백분위 방법
BC : 편의수정된 백분위 방법

<표 Ⅲ-2> 모의실험 결과 (양측 95% 신뢰구간)
n=25, 1000번 반복

기저분포	포함확률(%)	신뢰구간평균길이	신뢰구간의 분산
가동기간과 고장기간이 지수분포를 따를 경우	JK : 96.2	1.27	2
	F : 95.9	1.21	1
	NP : 92.1	1.09	1.4
	NBC : 92.6	1.11	1.5
	P : 96.3	1.10	1.4
	BC : 96.3	1.11	1.5
가동기간은 지수분포, 고장기간은 감마분포를 따를 경우	JK : 94.2	0.99	1
	F : 98.8	1.22	1
	NP : 91.7	2.88	7
	NBC : 91.9	2.88	7
	P : 94.7	2.72	7
	BC : 94.7	2.72	7
가동기간은 꼬리긴 h분포, 고장기간은 지수분포를 따를 경우	JK : 94.1	1.64	4
	F : 88.7	1.27	2
	NP : 92.2	1.10	1.4
	NBC : 92.8	1.19	1.4
	P : 94.2	1.32	1.1
	BC : 94.5	1.32	1.1

JK : 잭나이프방법
NP : 비모수적 붓스트랩
NBC : 편의수정된 비모수적 붓스트랩
P : 백분위 방법
BC : 편의수정된 백분위 방법

Ⅳ. 결 론

체계의 가용도에 대한 구간추정 방법으로 붓스트랩 방법을 제안하였다. 그 결과 모수적 붓스트랩 신뢰구간이 다른방법에 비하여 우수하게 나타남을 알 수 있었다. 특히 꼬리가 긴 분포일수록 붓스트랩 추론방법이 더욱 효율적있었다. 추후 연구과제로는 사용기간과 고장기간이 여러가지 분포를 가질때 가용도에 대한 추론과 보다 복잡한 체계의 가용도에 대한 추론을 고려할 수 있다.

참 고 문 헌

1. Efon, B(1979) "Bootstap methods ; another look at the jackknife", Annals of Statistics. Vol. 1, 1-16.
2. Efron, B and Tibshirani, R(1986) "Bootstrap methods for standard errors,

confidence intervals, and other measured of statistical accuracy" Statistical Science, Vol. 1, 54-77.

3. Gaver, D.P and Chu, B.B(1979) "Jackknife Estimates of Component and System Availability". Technometrics. Vol. 21, 443-450.
4. Gray, H.L and Lewis, T.O.(1967) "A Confidence Interval for the Availability rate" Technometrics. Vol. 9, 465-471.