

시뮬레이션 출력분석을 위한 임계값 부트스트랩의 성능개선*

Improving the Performance of Threshold Bootstrap for Simulation Output Analysis*

김윤배**

Yun Bae Kim**

Abstract

Analyzing autocorrelated data set is still an open problem. Developing an easy and efficient method for severe positive correlated data set, which is common in simulation output, is vital for the simulation society. Bootstrap is an easy and powerful tool for constructing non-parametric inferential procedures in modern statistical data analysis. Conventional bootstrap algorithm requires iid assumption in the original data set. Proper choice of resampling units for generating replicates has much to do with the structure of the original data set, iid data or autocorrelated. In this paper, a new bootstrap resampling scheme is proposed to analyze the autocorrelated data set : the Threshold Bootstrap. A thorough literature search of bootstrap method focusing on the case of autocorrelated data set is also provided. Theoretical foundations of Threshold Bootstrap is studied and compared with other leading bootstrap sampling techniques for autocorrelated data sets. The performance of TB is reported using M/M/1 queueing model, also the comparison of other resampling techniques of ARMA data set is also reported.

1. 서론

자기상관(autocorrelation)이 존재하는 자료군을 분석하는 문제는 아직 그 해법이 완벽하지 못한 실정이며, 특히 자기상관 정도가 심하여 의미있는 척도(performance measure)를 구하기 위하여 많은 관측치가 요구되면 시뮬레이션 수행에 오랜 시간이 소요되며, 이런 경우의 출력 분석을 위해서 효율적인 분석기법의 개발이 절실하게 요구되고 있다.

시뮬레이션 출력분석에서는 강대수법칙(Strong Law of Large Numbers)에 의해 표본평균을 구해 시뮬레이션 모델을 수행하는 시스템의 참 값을 추정하며, 출력자료에 양의 상관(positive correlation)이 심하게 존재하면 분산이 iid 가정 때보다 부풀어지므로 신뢰구간 추정시 그 구간의 길이가 커진다. 이 문제를 해결하기 위한 여러 가지 기법들이 개발되었으나 통계적인 훈련이 없는 초보사용자가 사용하기에는 어렵다.

본 연구는 앞에서 언급한 시뮬레이션 출력과 시계열자

* 본 연구는 '96 한국과학재단 핵심전문연구비를 지원 받았음(961-1012-088-2).

** 성균관대학교 공과대학 산업공학과

료의 분석에 쓰일 새로운 기법개발과 그 기법의 성능향상에 그 목적이 있다. 강대수법칙의 적용으로 평균값 형태의 성능척도를 구하는 데에는 별 어려움이 없으나 분산의 추정에는 많은 어려움이 존재하는 시뮬레이션 자료의 분석에 쓰일 쉽고 새로운 기법의 개발은 다른 분야의 자기상관 자료군의 해석에도 유용하게 쓰일 것이다(예를 들면, 시계열 분석).

의미 있는 통계적 추정치를 구하기 위하여 다수의 반복실험(Independent Replication)을 수행하나, 복잡한 시뮬레이션 모델의 수행에는 너무 많은 시뮬레이션 수행시간이 요구된다. 또한, 배치평균방법(Batch Means Method)이 단독실험 추론방법으로 사용되나, 배치평균방법은 적절한 배치의 크기와 개수를 정하는데 어려움이 따른다 [Law and Kelton, 1991].

자료해석을 위한 재추출 기법(resampling schemes)은 현대통계학에서 매우 유용하게 비모수형 추론을 유도해 낼 수 있다. 특히, 최근 값싸고 쉽게 접근할 수 있는 계산환경을 바탕으로 재추출 기법들에 관한 연구가 활발하게 진행되고 있다. 더욱이 최소한으로 요구되는 계반 가정들은 재추출 알고리즘의 사용을 가속화시켰다. 그러나, 재추출 기법은 원시자료군에 iid가정을 요구하는데, 이런 가정이 만족되지 않는 시계열 자료나 시뮬레이션 출력자료를 분석하기 위한 재추출 기법의 변형을 위한 연구가 활발하게 진행되고 있고, 본 논문에서는 자기상관된 자료를 재추출하기 위한 방법들을 조사하여 소개하였다.

본 논문은 2절에서 재추출 기법을 시뮬레이션 출력분석에 이용하는 연구를 소개하였고, 3절에서는 임계값 부트스트랩의 개발에 관하여 설명하였고, 임계값 부트스트랩의 문제점과 개선방안, 또한 성능을 여러 가지 시계열 모형을 사용하여 비교하여 4절에서 소개하였다. 5절에서는 임계값 부트스트랩의 이론적 배경을 고찰하였고 유사한 부트스트랩 기법과의 이론적 배경을 비교하였다. 6절에서는 결론과 향후 연구방향이 언급되었다.

2. 재추출 기법(Resampling Techniques)의 시뮬레이션 출력분석에의 이용

2.1 고전적 부트스트랩(Conventional Bootstrap:CB)

미지의 확률분포 함수 F 에서 독립성과 동일성(iid)이 보

장되는 통계량의 추정값은 일반적으로 아래와 같이 정의된다.

$$\hat{\theta} = \theta(\hat{F}) \quad (1)$$

여기서 \hat{F} 는 경험적 확률분포함수를 나타내고, \hat{F} 는 독립성과 동일성이 보장되는 n 개의 모든 관측치 X_i 에 추출될 확률을 $\frac{1}{n}$ 로 보장한다.

$\sigma(F)$ 를 미지의 샘플링분포 F 와 함수관계에 있는 표준오차라고 하면, 부트스트랩으로 추정한 표준오차는 $\sigma(\cdot)$ 를 F 대신 \hat{F} 으로 대체했을 때의 값이다. 즉

$$se_{boot}(\hat{\theta}) = \sigma(\hat{F}) \quad (2)$$

만약 표본평균이 우리가 추정하는 통계량이면

$$\sigma(F) = \left\{ \frac{\mu_2(F)}{n} \right\}^{\frac{1}{2}} \quad (3)$$

이 된다. 여기서 $\mu_2(F)$ 는 $\int (x - \mu)^2 dF$ 이다. 따라서 표본평균에 대한 부트스트랩 표준오차는 (2)와 (3)식에서 아래와 같이 계산할 수 있다.

$$se_{boot}(\bar{X}_n) = \sigma(\hat{F}) = \left\{ \frac{\mu_2(\hat{F})}{n} \right\}^{\frac{1}{2}} \quad (4)$$

여기서 $\mu_2(\hat{F})$ 는 $\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n}$, F 의 2차 중심적률이다.

\bar{X}_n 이외의 통계량들에 대한 $\sigma(F)$ 의 공식은 일반적으로 알려져 있지 않고 $se_{boot}(\hat{\theta})$ 은 Monte Carlo 시뮬레이션을 통하여 추정할 수 있고 그 시뮬레이션 알고리즘은 아래와 같다.

1) 크기가 n 인 부트스트랩 표본을 iid 관측치 (X_1, X_2, \dots, X_n) 의 집단에서 생성한다.

$$\hat{F} \rightarrow (X_1^*, X_2^*, \dots, X_n^*)$$

2) 부트스트랩으로 구한 표본으로 통계량을 B 회만큼 계산한다.

$$\hat{\theta}_i^* = \hat{\theta}(X_1^*, \dots, X_n^*), \quad i=1, \dots, B \quad (5)$$

3) 통계량 θ 의 부트스트랩 추정치를 아래와 같이 계산한다.

$$se^*(\hat{\theta}^*) = \left\{ \frac{1}{B-1} \sum_{i=1}^B (\hat{\theta}_i^* - \hat{\theta}^*)^2 \right\}^{\frac{1}{2}} \quad (6)$$

$$\hat{\theta}^* = \frac{1}{B} \sum_{i=1}^B \hat{\theta}_i^* . B가 무한대로 커지면 (6)의 $se^*(\hat{\theta}^*)$ 는$$

(2)의 $se_{boot}(\theta)$ 로 접근한다.

대부분의 부트스트랩 실험에서 B의 크기는 50에서 200 정도가 $se_{boot}(\hat{\theta})$ 를 적합하다고 권장되고 있다.[Efron and Tibshirani(1986)]

부트스트랩의 편의(bias)는 아래와 같이 정의되고 Monte Carlo 시뮬레이션으로 (2)의 $se_{boot}(\hat{\theta})$ 를 (6)의 $se_{boot}(\hat{\theta}^*)$ 로 추정하듯이 추정할 수 있다.

$$bias_{boot}(\theta) = E_{\hat{F}}(\hat{\theta}^*) - \theta(\hat{F}) \quad (7)$$

$$bias^*(\hat{\theta}^*) = \frac{1}{B} \sum_{i=1}^B \hat{\theta}_i^* - \theta(\hat{F}) \quad (8)$$

표본평균을 추정한다면 (8)은 아래와 같이 편이가 존재하지 않는다는 것을 보일 수 있다.

$$bias_{boot}(\bar{X}_n) = E_{\hat{F}}(\bar{X}_n^*) - \bar{X}_n = 0 \quad (9)$$

유사한 방법으로 (8)의 B가 무한대로 커지면 (8)의 $bias^*(\hat{\theta}^*)$ 가 (7)의 $bias_{boot}(\hat{\theta})$ 로 접근한다. 고전적인 부트스트랩에서는 원시자료군에 iid가 매우 중요한 가정이다. Liu(1989)가 처음으로 iid 가정을 완화시킬 때 고전적 부트스트랩의 강건성(robustness)에 대하여 언급하였다.

2.2 자기상관의 자료

자기상관된 자료를 재추출하려면 iid 경우의 재추출 방법의 재추출 단위를 조정하여 자기상관 구조를 보존할 수 있도록 하여야 하고, Carlstein(1993)이 대표적인 자기상관관계를 갖는 자료군인 시계열에 적용할 수 있는 재추출방법들을 조사하였다.

이번 절에서 간략하게 소개할 재추출 방법들은 대부분 비모수형 방법이나, 종속관계에 대하여 사전지식을 갖고 있을 때는 모수형 재추출 방법을 사용할 수 있다.

2.2.1 ARMA 모형을 위한 잔차 재추출 방법

원시자료군의 종속구조를 안다면 이 사전지식을 재추출 방법에 사용할 수 있다. 즉 iid가 보장되는 잔차에 대하여 고전적인 재추출 방법을 적용시키는 것이다. Efron과 Tibshirani(1986)에 의하여 고안된 이 방법은 1차 자기회귀 과정(first-order autoregressive process)의 표준편차 $\hat{\phi}$ 을 추정하는데 사용되었다. 자세한 알고리즘은 Efron과 Tibshirani(1986)의 논문을 참조하기 바란다. Thoms와 Schucany(1990)는 자기회귀 시계열 모형의 예측구간을 구하기 위하여 부트스트랩 방법을 사용했다. 실제로 종속관계에 대한 완전한 이해가 규명되기 어려우므로 이 모수형 방법을 사용하기에는 무리가 있다.

2.2.2 이동블록 부트스트랩(Moving Blocks Bootstrap: MBB)

Künsch(1989)와 Liu와 Singh(1992)에 의해 시계열을 비모수적으로 재추출하기 위한 이동블록 부트스트랩이 독립적으로 개발되었다. 이동블록 부트스트랩은 원시자료가 m -dependent 정상과정(m -dependent stationary process)에서 생성된다고 가정한다. 재추출의 기본 단위인 이동블록 B(크기가 b인 i번째 이동블록)는 아래와 같이 정의한다.

$$B_i = \{X_i, X_{i+1}, \dots, X_{i+b-1}\}, i = 1, 2, \dots, n-b+1$$

7개의 관측치로 이루어진 자료군을 갖고 재추출하는 이동블록의 크기는 4로 하는 간단한 예제를 그림 1에 소개하였다.

이렇게 정의한 이동블록 $\{B_1, B_2, \dots, B_{n-b+1}\}$ 의 집합에서 복원을 허락하며 재추출하면, 이동블록 부트스트랩의 확률적 구조 P^* 가 정의된다.

이동블록 부트스트랩으로 추정된 i 번째 \bar{X}_n^* 의 분산은 i 번째 \bar{X}_n^* 가 된다. 또한 $P\{\sqrt{n}(\bar{X}_n - \mu) \leq x\}$ 의 추정치는 $P^*\{\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \leq x\}$ 가 된다.

이동블록의 크기 b를 적절하게 정하는 것이 가장 큰 문제인데, 약한 종속구조(weak dependence)의 가정으로 이동블록의 크기 b를 적절하게 정하여 시차(lag) p의 자기상관을 무시할 수 있다. 또한, b만큼 떨어져 있는 관측치들간의 자기상관 구조는 보존될 수 있도록 이동블록의 크기를 정해야 한다.

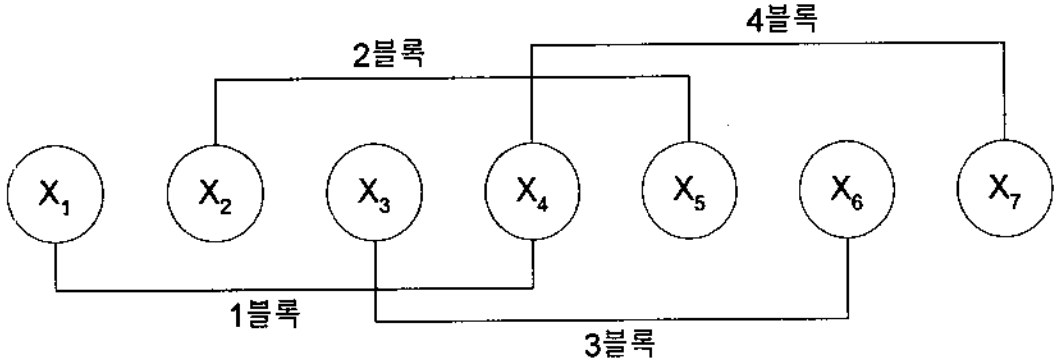


그림 1. 이동블록 부트스트랩의 블록구조도

Hall(1995)은 경험적으로 이동블록의 크기를 정하는 방법을 소개하였고, Lahiri(1991)와 Davison과 Hall(1992)은 이동블록의 크기를 정하는 문제를 분포추정의 관점에서 해석했다. Politis와 Romano(1992)는 이동블록 부트스트랩의 변형인 고리형 이동블록 부트스트랩을 소개하였다.

2.2.3 정상 부트스트랩(Stationary Bootstrap : SB)

정상 부트스트랩은 Politis와 Romano(1994)에 의해 개발되었는데, 이 방법 역시 관측치내에 약한 종속구조 (weakly dependent structure)를 가정한다. $\{X_1, \dots, X_n\}$ 을 약한 종속구조가 있는 정상(stationary) 관측치라고 하자. 표

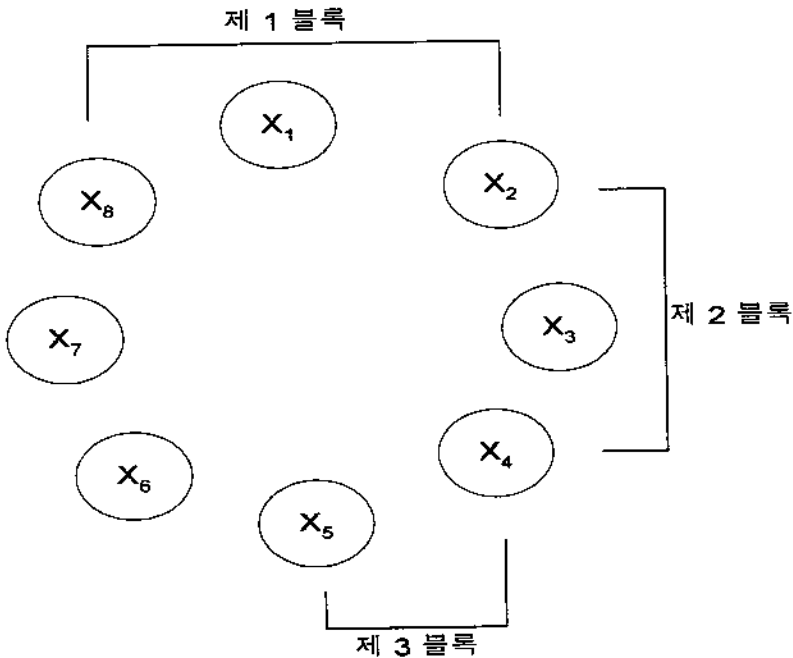


그림 2. 8개의 원시자료로 이루어지는 정상 부트스트랩의 블록 결정

본평균을 추정하기 위한 정상 부트스트랩은 이동블록 부트스트랩 방법과 마찬가지로 블록을 재추출의 기본단위로 사용하는데, X_i 에서 시작하여 b 개의 관측치를 갖는 블록을 아래와 같이 정의하고, 간단한 예를 그림 2에 소개하였다.

$$B_{i,b} = \{X_i, X_{i+1}, \dots, X_{i+b-1}\}$$

이렇게 정의된 블록을 사용하여 랜덤하게 재추출된 $B_{I_1, L_1}, B_{I_2, L_2}$ 로 의사시계열(pseudo time series)을 생성한다. 여기서 I_1, I_2 는 $\{1, \dots, n\}$ 까지 정의된 이산균등 분포를 따르고 L_1, L_2, \dots 는 모수를 p 로 하는 기하분포를 따른다. 만약 마지막으로 랜덤하게 고른 블록이 재생성한 의사시계열을 원시자료군의 크기인 n 을 넘어가면, 마지막으로 고른 블록내의 관측치를 $X_{K(K=j(\text{mod } n))}$ 로 제한한다. 정상 부트스트랩으로 추정할 \bar{X}_n 의 분산 추정치는 P^* 를 따르는 $(\bar{X}_n^* - \bar{X}_n)$ 의 분산값이 되고 $P\{|\bar{X}_n - \mu| \leq x\}$ 의 추정 값은 $P^*\{|\bar{X}_n^* - \bar{X}_n| \leq x\}$ 가 된다. 여기서 P^* 는 부트스트랩 샘플(즉 의사시계열)의 생성으로 정의된 확률측도(probability measure)이다. 이동블록 부트스트랩의 b 를 정하는 것과 유사한 문제인 p 를 모르는 것이 중요한 문제인데, p 는 기하분포를 따르는 블록의 평균크기 $\frac{1}{p}$ 을 결정한다. 이동블록 부트스트랩과 마찬가지로 적절하게 p 를 정하는 것이 정상 부트스트랩으로 추정하는 부트스트랩 추정량의 편차와 분산에 아주 민감하게 반응한다. 그림 2에 8개의 자료로 이루어진 정상 부트스트랩의 과정을 소개하였다.

2.3 이진 부트스트랩(Binary Bootstrap : BB)

이진 부트스트랩은 Kim(1993)에 의해 시뮬레이션 출력 분석을 위하여 개발되었다. 모든 정상 시계열은 임의의 임계값을 중심으로 이진형태로 변환될 수 있다. 임계값보다 큰 값은 1-run으로 적으면 0-run으로 변환시켜 이 run들을 번갈아 가며 재추출하여 부트스트랩 샘플을 생성해 낸다. 이 방법은 고정된 크기의 블록을 재추출의 기본단위로 사용하는 이동블록 부트스트랩이나 블록의 크기가 기하분포를 따르는 정상 부트스트랩보다 간편하다. Kim의 실험결과에 따르면 이진 부트스트랩은 단일반복 시뮬레이션 출력분석 기법인 배치평균 방법보다 M/M/1과

D/M/10에 우수한 성능을 보였다.

2.4 연결 블록 부트스트랩 (Linked Blockwise Bootstrap : LBB)

Künsch와 Carlstein(1990)은 이동블록 부트스트랩을 변형하여 연결블록 부트스트랩을 개발하였다. 이동블록 부트스트랩에서 블록을 랜덤하게 골라서 의사자료군을 생성하기 때문에 블록의 끝부분의 종속구조에 왜곡이 생긴다. 이 문제를 해결하기 위하여 연결되는 블록을 이웃하는 블록들에서만 추출할 것을 제안하여, 블록들의 순조로운 연결로 종속구조의 왜곡을 방지할 수 있다고 하였다. 이 주장은 임계값 부트스트랩에서도 자연스러운 연결을 위하여 다른 형태로 고려되었다.

2.5 최단 이웃 부트스트랩 (Nearest Neighbor Bootstrap : NNB)

Lall과 Sharma(1996)는 육수학(hydrology) 시계열을 부트스트랩을 사용하여 재추출하기 위하여 최단 이웃 부트스트랩을 연구하였다. 이 방법은 종속구조에 대한 정보를 알고 있다고 가정을 한다. 즉, 미래의 자료흐름이 어느 정도 시차까지 영향을 받는지를 미리 정하는 것이다. 이 가정은 모양벡터를 예측하거나 시뮬레이션 하여 우선 계승하는 값으로 조율되어진다. 현재 모양벡터의 최단 이웃 값들이 결정된 후 부트스트랩 샘플은 이 최단 이웃 값들에서 K 개를 추출한다. 균등하게 K 개의 후속 값들을 추출하기보다는 이산형 커널(Kernel)을 사용하여 재추출한다. 즉, 재추출이 필요하면 현재 모양벡터의 이웃 값들을 먼저 고려한다.

연결블록 부트스트랩과 최단 이웃 부트스트랩은 실제적으로 적용하기에 다음의 두 가지 이유로 어려움이 있다. 첫 번째 이유는 최적의 최단 이웃 값들을 결정하기 어렵고 두 번째 이유로는 후속 값이 그 전 값들에 강하게 의존되어 있다면 시차(lag)를 정하는데 어려움이 따르기 때문이다.

3. 임계값 부트스트랩(Threshold Bootstrap: TB)

임계값 부트스트랩(TB)은 시뮬레이션의 단일실행의 결

과를 해석하기 위하여 고전적인 부트스트랩의 재추출 단위인 단독 관측치를 관측치내에 존재하는 종속관계를 유지하도록 '주기'로 바꾸었다. Kim(1992)은 재추출 단위를 주기로 바꾼 상태에서 재추출된 부트스트랩 모의자료군이 원시자료에 존재하는 종속관계를 유지하는 것을 보였다. TB는 Kim(1992)의 이진 부트스트랩을 이진형태가 아닌 정상의 자료에 적용하도록 고안하였고 재추출단위로 이진 부트스트랩의 'run'(1-run 또는 0-run)에서 주기(low run과 high run의 결합)로 바꾸었다. 임계값을 시계열을 관통하는 수준(표본평균이나 표본의 중앙값 등)으로 정하면 시계열은 임계값보다 높은 자료의 연속인 high런과 낮은 자료의 연속인 low런으로 구분되어진다. 재추출의 기본단위로 중복되지 않고 연속되는 두 개의 런(high런과 low런의 결합)으로 정하고, 이를 '주기'라고 하였다. 임계값을 정하면 관측치 자체가 주기를 결정하게 되어, 재추출의 단위가 자동으로 정해지는 것이 TB의 장점일 수 있다. 임계값 부트스트랩은 일반적인 시뮬레이션 출력분석에 사용되는 다수의 반복 시뮬레이션 실행을 부트스트랩 샘플로 대체할 수 있어서 시뮬레이션 실행시간을 단축할 수 있고, 동류의 단일실행 분석기법인 배치평균방법보다

사용하기 쉽다는 장점이 있다. 정확도는 실험을 통하여 3.2에 소개하였다.

3.1 임계값 부트스트랩의 구조와 재추출 알고리즘

확률체계 P에 의해 생성된 정상 시계열이 약한 종속구조를 갖고, 임의의 임계값이 R개의 주기를 만든다고 가정하자. $C_i = \{X_{i,1}, X_{i,2}, \dots, X_{i,n_i}\}$ 는 크기를 n_i 로 하는 i번째 주기라면, $i=1, 2, \dots, R$ 이고 $\sum_{i=1}^R n_i = n$ 이 된다. 주기의 크기 n_i 와 주기의 수 R은 둘 다 확률변수이다. 이 두 개의 확률변수들은 관측자료군의 자기상관과 임계값에 매우 강한 종속관계를 갖는다. TB는 집합 $\{C_1, C_2, \dots, C_R\}$ 에서 랜덤하게 주기를 추출하여 부트스트랩 샘플을 만든다. 미지의 P는 재추출 구조에 근거하여 P^* 로 추정할 수 있다. 관측치 $\{X_1, X_2, \dots, X_n\}$ 의 표본평균 \bar{X}_n 의 편차와 표준오차를 추정하려 한다. $\{X_1^*, X_2^*, \dots, X_n^*\}$ 가 P^* 에서 얻은 부트스트랩 샘플이라면, 즉 $P^* \rightarrow \{X_1^*, X_2^*, \dots, X_n^*\}$, \bar{X}_n 에 관한 편차와 표준오차에 대한 부트스트랩 추정치는 재추출 확률구조 P^* 에 의거한 \bar{X}_n^* 의 편차와 표준오차와 일치하게 된다. 임계값 부트스트랩의 구조에 대한 이해를 돕기 위하여 20개의 관

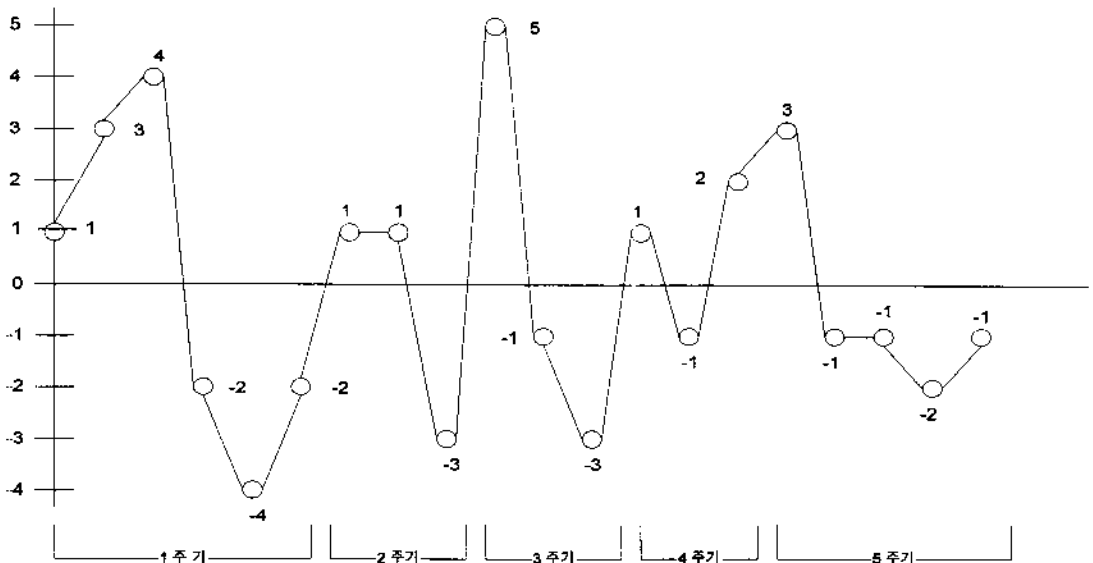


그림 3. 임계값 부트스트랩의 주기결정

측치로 이루어진 표본과 TB를 3회 반복한 예제를 소개 하였다.

1주기={1,3,4,-2,-4,-2}

2주기={1,1,-3}

3주기={5,-1,-3}

4주기={1,-1}

5주기={2,3,-1,-1,-2,-1}

임계값= 표본평균=0

반복 1: 1,-1 1,1,-3 1,1,-3 1,3,4,-2,-4,-2 5,-1,-3 1,1,-3 → $\bar{X}_1^* = -0.1$

반복 2: 2,3,-1,-1,-2,-1 5,-1,-3 1,-1 1,1,-3 5,-1,-3 5,-1,-3 → $\bar{X}_2^* = 0.1$

반복 3: 1,3,4,-2,-4,-2 2,3,-1,-1,-2,-1 1,-1 5,-1,-3 1,1,-3 → $\bar{X}_3^* = 0.0$

편차=0, 표준오차=0.1

TB의 재추출 알고리즘은 다음과 같다.

step 0: N개로 이루어진 자기상관이 존재하는 자료군을 획득한다.

step 1: 임계값을 설정한다(예를 들면, 표본평균이나 표본의 중앙값)

step 2: 자료군을 임계값을 중심으로 두 개의 자료 군으로 분류한다. 자료군은 임계값보다 높은 자료의 연속인 "high-run"과 낮은 자료의 연속인 "low-run"으로 구분되어진다.

step 3: 중복되지 않고 연속되는 두 개의 런(high런과 low런의 결합)을 복원을 허락하며 재추출된 주기들을 연결하여 부트스트랩 샘플을 원시자료의 크기만큼 재생성한다. R개의 주기에 추출될 확률을 각각 $\frac{1}{R}$ 로 할당하고, 균등분포(0,1)를 따르는 난수를 생성시켜 재추출할 주기를 선택한다.

* 만약 마지막 랜덤하게 재추출된 주기가 재생성한 의사 시계열을 원시자료군의 크기인 n을 넘어가면, 마지막으로 고른 주기내의 관측치를 X_n 으로 제한한다.

step 4: 재 생성된 부트스트랩 샘플로 관심 있는 통계량(예를 들면, 표본평균)을 계산한다.

step 5: step 3과 step 4를 총 B번 반복한다.

step 6: B번의 반복으로 계산한 통계량의 표본 형태를 계산한다(예를 들면, 신뢰구간).

3.2 TB의 M/M/1 적용결과

TB를 사용하여 M/M/1($\rho=0.9$) 대기모형의 특징을 분석

표 1. M/M/1 delay에 대한 BM과 TB의 Monte Carlo Comparison

| Server Utilization | $\rho=0.9$ | $\rho=0.9$ |
|-----------------------------------|--------------|--------------|
| # Data Points | 20,000 | 100,000 |
| Average of Sample Mean | | |
| Theory | 8.100 | 8.100 |
| BM ± 1.65SE | 8.110 ± .186 | 8.089 ± .087 |
| TB ± 1.65SE | 8.159 ± .183 | 8.119 ± .084 |
| Standard Deviation of Sample Mean | | |
| Theory ¹⁾ | 1.340 | .599 |
| BM ± 1.65SE | 1.101 ± .156 | .277 ± .028 |
| TB ± 1.65SE | 1.028 ± .069 | .520 ± .023 |
| Confidence Interval Coverage | | |
| Nominal | .900 | .900 |
| BM ± 1.65SE | .780 ± .068 | .830 ± .062 |
| TB ± 1.65SE | .850 ± .059 | .880 ± .054 |

¹⁾ Asymptotic approximation(Whitt 1989)

BM은 100번의 반복시행을 기초로한 결과 치이다.

TB는 각 독립반복시행을 B=500인 Bootstrap 반복시행으로한 100번의 결과 치이다.

한 결과 배치평균방법(BM: Batch Means method)보다 TB의 성능이 우수함을 발견했다. 표본평균의 평균의 신뢰구간 측면에서 볼 때 N이 20,000과 100,000의 경우, 두 방법 모두 참값을 포함한다. 하지만 표본평균의 표준편차의 신뢰구간에서는 N이 20,000일 경우 두 방법 모두 참값을 포함하지 못하는데, 이는 자료의 수가 자기상관 정도에 비해 너무 적기 때문일 것이다. 그러나 N이 100,000일 경우 배치평균보다는 TB의 신뢰구간이 참값에 근접하였다. Coverage test에서도 TB의 우수성이 입증되었는데, 20,000의 자료군을 갖고도 TB만이 원하는 nominal coverage를 달성했다. Nominal coverage란 90% 신뢰구간을 사용했을 때 100번의 다른 모집단의 샘플을 사용한 결과 이론 값 을 cover하는 비율이 90% 달성함을 목표로 한다.

4. TB의 문제점과 개선 방안 및 시계열에의 적용

TB는 원시자료가 임계값 수준을 교차하는 횟수와 매우 밀접한 관계를 갖고 있다. 또한 강한 양의 상관관계를 가진 원시자료군에서는 주기의 숫자가 적고 주기의 길이가 긴 반면 음의 상관관계의 자료군에서는 주기의 숫자가 많고 그 길이가 짧다. 경험적으로 AR(1) 시계열의 평균주기의 길이와 1차 자기회귀계수와 관계를 -0.9 에서 $+0.9$ 로 변화시켜 보면 평균주기의 길이가 증가한다. 이런 현상이 그림 4에 나타나 있다.

따라서 일반적으로 교차횟수가 많을수록 재추출 단위가 짧아지므로 재추출 횟수는 많아지는데 이런 현상은 대표적으로 강한 음의 상관관계가 존재할 때 나타난다. 반면에 시계열이 반드럽고 변화의 속도가 느리면 재추출해야 하는 횟수는 적어진다. 문제는 강한 음의 상관관계가 존재하는 자료군에서 발생하는데, 많은 횟수의 재추출을 시행하면 주기간의 접합점(주기의 양 끝점)이 많아져서 접합점 주변의 자료에 상관관계가 무너지는 경우가 발생한다. 따라서 음의 상관관계가 존재하는 자료군의 주기의 길이에 대한 의문점이 발생한다. 즉 상관관계는 증가하지만 평균주기의 길이는 줄어든다. 그런데 TB의 주목적은

원시자료에 존재하는 상관관계를 재추출한 의사 시계열에도 보존하는 것이므로 임계값을 교차하는 횟수가 많아지면 재추출하는 단위에 대한 조정이 필요하다.

Politis와 Romano(1992)는 적절한 블록크기인 b 를 선택하는데 어려움이 있다고 지적하였는데 같은 문제가 TB에서도 발생하는 것을 알 수 있다. 새로이 고안한 재추출 단위는 토막(chunk)이라고 명명하고, 그 크기는 하나 혹은 그 이상의 주기로 이루어진다. 새롭게 고안한 TB의 재추출 구조를 검증하기 위해서 ARMA 시계열 자료를 사용하여 표본평균 표준편차의 평균제곱오차를 최소화 하는 블록/토막의 크기를 결정하였고, 이동블록 부트스트랩(MBB)과 정상 부트스트랩(SB)등과 비교하였다.

재추출의 기본단위인 블록의 크기는 관측자료군의 크기와 자기상관 정도에 연관이 되어있어 단순히 결정할 수 없다. 최적의 블록크기를 구하기 위한 해석적인 방법은 아직까지 소개되어 있지 않아, 경험적인 수치발견 기법을 적용하였다. TB에서는 최초의 토막 크기를 1주기로 하였고, 한 주기씩 늘려가며 평균제곱오차를 관측하였다. MBB와 SB에서는 초기블록의 크기를 5개의 관측 치로 정했고, 매번 2개의 관측치를 늘려가며 평균제곱 오차를 관측하였다.

ARMA 시계열을 다음과 같이 고려해 보자.

(평균런의 길이)

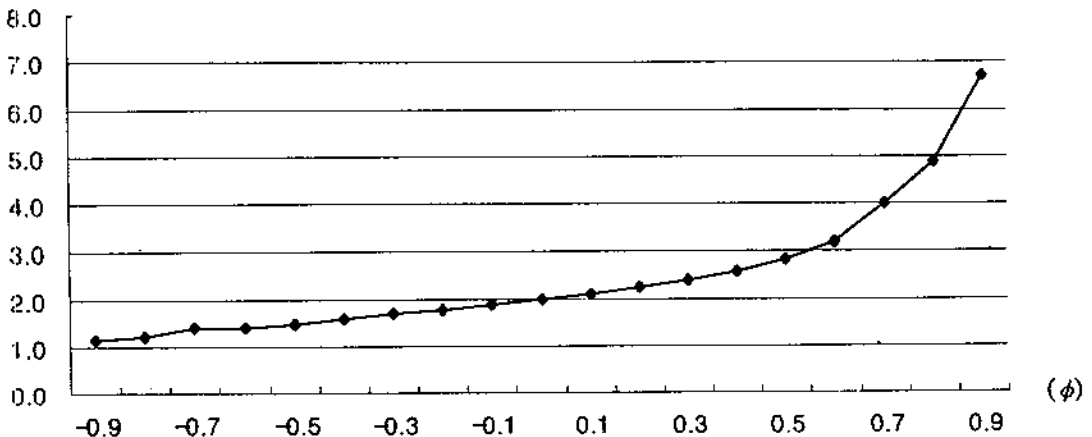


그림 4. 종속구조와 run의 크기의 상관관계

$$(1 - \phi B)X_t = \mu' + (1 - \theta B)\varepsilon_t, \quad \infty < t < \infty \quad (10)$$

ε_t 는 iid 표준정규분포를 따르고, 시계열의 평균 $E(X_t) = \frac{\mu'}{(1-\phi)}$ μ 는 0이라고 가정하며, μ 는 포본평균 \bar{X}_n 로 추정한다. 세 가지 재추출 방법으로 추정하는 표본평균의 표준오차, $se(\bar{X}_n)$ 는 아래의 식으로 구했다.

$$se^*(\bar{X}_n) = \sqrt{\frac{\sum_{i=1}^B (\bar{X}_i^* - \bar{X})^2}{B-1}} \quad (11)$$

$\bar{X}^* = \frac{\sum_{i=1}^B \bar{X}_i^*}{B}$, $\bar{X}_i^* = \frac{\sum_{j=1}^n \bar{X}_{ij}^*}{n}$ 이고, X_{ij}^* 는 i 번째 부트스트랩 반복의 j 번째 부트스트랩 관측치이다. 시계열자료의 크기는 $n=1,000$ 으로 하였고, 세 방법 모두 부트스트랩 반복횟수인 B 를 300으로 하였다. 실험에 사용된 시계열의 모형과 특징은 다음과 같다.

- (1) $X_t = 0.9X_{t-1} + \varepsilon_t$: 자기상관함수(autocorrelation function)가 천천히 감소한다
- (2) $X_t = -0.9X_{t-1} + \varepsilon_t$: 자기상관함수가 자주 변동한다
- (3) $X_t = 0.4X_{t-1} + 0.2X_{t-2} + 0.1X_{t-3} + 0.1X_{t-4} + 0.1X_{t-5} + \varepsilon_t$: 복잡한 자기상관함수를 가지며 t 시점의 값은 과거 5개의 값과 자기회귀 계수에 의존한다

(4) $X_t = \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3} + \varepsilon_{t-4}$: 자기상관함수가 4번째 lag이후에는 없다

(5) $X_t = 0.9X_{t-1} + \varepsilon_t + 0.9\varepsilon_{t-1}$: 자기회귀 계수와 이동평균 계수에 종속적인 자기상관함수가 초기값 ρ_1 에 서부터 지수 적으로 감소한다

표 3에 최소 표준평균오차와 최적의 블록/토막의 크기가 요약되어 있는데, TB와 토막 크기가 실험한 모든 시계열 모형에서 작음을 관측할 수 있었다. Carlstein(1986)이 AR(1) 모형($\phi = \pm 0.9$)이 제시한 이론적인 최적의 크기인 45($\phi = \pm 0.9$ 두 경우 모두)와 경험적으로 찾은 크기(MBB와 SB)인 38-54와 비교할 때 매우 근접하나 변동이 있음을 관측하였는데, 이는 Carlstein의 결과는 자기상관에 의존하지 않으나, 수치발견기법으로 찾은 결과는 자기상관 정도에 의존하고 있기 때문이다. 시계열 모형을 실험한 수치발견 기법과 TB의 결과를 아래의 표 2에 요약하였다.

5. TB의 이론적 고찰

부트스트랩을 사용하여 추정하는 θ^* 의 평균과 분산은 관측치 $\{X_1, \dots, X_n\}$ 에 제한하여 재추출 알고리즘을 사용하여 구한 부트스트랩 표본 $\{X_1^*, \dots, X_n^*\}$ 를 통하여 계산하는 것이 일반적인 방법이다.

표 2. 자기상관 자료군을 위한 재추출 기법들의 성능비교

| 시계열 모형 | 최소 표준평균 오차*10 ³ | | | 최적 블록크기 | | |
|--|----------------------------|------------------|------------------|---------|-----|----|
| | TB | MBB | SB | TB | MBB | SB |
| AR(1), $\phi=0.9$ | 3.1267 ± .5831 | 3.2463 ± .3075 | 4.0782 ± .3686 | 14.11 | 38 | 38 |
| AR(1), $\phi=-0.9$ | 0.0038 ± .0006 | 0.0046 ± .0006 | 0.0051 ± .0009 | 16.16 | 54 | 46 |
| AR(5), $\phi=0.4, 0.2, 0.1, 0.1, 0.1$ | 6.1982 ± .6780 | 7.1595 ± .7363 | 8.8871 ± .8831 | 35.22 | 84 | 62 |
| MA(4), $\theta=-1, -1, -1, -1(N=1000)$ | 0.2976 ± .0382 | 0.2110 ± .03000 | 0.2639 ± .0343 | 9.79 | 18 | 20 |
| MA(4), $\theta=-1, -1, -1, -1(N=5000)$ | 0.0183 ± .0022 | 0.0204 ± .0028 | 0.0219 ± .0027 | 9.75 | 32 | 38 |
| ARMA(1,1), $\phi=0.9, \theta=-0.9$ | 11.4694 ± 1.6144 | 11.5661 ± 1.5739 | 15.2453 ± 1.7919 | 20.18 | 50 | 38 |

- 1. MSE * 10³ ± se ; se는 고전적인 부트스트랩 방법으로 구함.
- 2. 부트스트랩 반복 횟수 ; B=300
- 3. 가장 작은 평균제곱오차와 최적의 블록크기는 굵게 표시하였다.
- 4. MBB:Moving Block Bootstrap, SB: Stationary Bootstrap

특히 추정하는 파라미터가 $\hat{\theta}^* = \bar{X}_n^*$ 이면, 표본평균 $E^*(\bar{X}_n^*)$ 나 $Var^*(\sqrt{n} \cdot \bar{X}_n^*)$ 는 재추출하는 과정 없이도 계산이 가능하다. 이 절에서는 원시 관측치 $\{X_1, \dots, X_n\}$ 이 주어질 때 조건부 부트스트랩 표본평균 $E^*(\bar{X}_n^*)$ 과 조건부 부트스트랩 표본평균의 분산 $Var^*(\sqrt{n} \cdot \bar{X}_n^*)$ 을 구하는 이론적인 배경을 고찰하였다.

5.1 임계값부트스트랩 표본평균의 조건부 평균: $E^*(\bar{X}_n^*)$

Politis와 Romano(1994)는 SB를 통하여 표본평균 \bar{X}_n 의 불편추정치인 \bar{X}_n^* 를 구할 수 있다고 하였는데, 즉 $E(\bar{X}_n^*/X_1, \dots, X_n) = \bar{X}_n$ 임을 보였다. Liu와 Singh(1992)은 MBB를 통한 조건부 부트스트랩 표본평균이 $E(\bar{X}_n^*/X_1, \dots, X_n) = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} \frac{1}{b} \sum_{j=1}^{i+b-1} X_j \bar{X}_n + O_p(b/n)$ 이라고 하였고, n 이 커지면 (b/n) 은 0으로 접근한다고 주장했다.

TB의 조건부 부트스트랩 표본평균을 구하기 위하여, 먼저 $\{X_1, \dots, X_n\}$ 이 정상의 약한 종속구조를 갖는 시계열자료이고 각각 n_i 의 크기를 갖는 R개의 토막이 있다고 가정하면 $\sum_{i=1}^R n_i = n$ 이 성립한다.

크기가 n 인 부트스트랩 표본을 생성하기 위하여 k 개의 토막을 복원을 허락하며 재추출하면 $\sum_{i=1}^k n_i^* = n$ 이 되며, n_i^* 는 $\{n_1, n_2, \dots, n_k\}$ 에서 랜덤하게 추출한 i 번째 토막이다. k 는 토막의 크기 n_i 의 함수인 확률변수이며, n_i 가 클수록 k 는 작게 되며 그 역도 성립한다.

S_i 를 i 번째($i=1, \dots, R$) 토막내의 관측치의 합이라고 하고, S_i^* 는 i 번째($i=1, \dots, k$) 재추출한 토막내의 관측치의 합으로 정의하자. 그러면,

$$\bar{X}_n^* = \sum_{i=1}^k \sum_{j=1}^{n_i^*} X_{ij}^* / n = \sum_{i=1}^k S_i^* / n \tag{12}$$

(12)식에 의하면 $E^*(\bar{X}_n^*) = E[\bar{X}_n^*/X_1, \dots, X_n] = \frac{1}{n} E[\sum_{i=1}^k S_i^*/X_1, \dots, X_n]$ 이 되고, S_i^* 가 iid이고 k 가 확률변수이기 때문에, $E^*(\bar{X}_n^*) = \frac{1}{n} E[k/X_1, \dots, X_n]$.

$E[S_i^*/X_1, \dots, X_n]$ 가 된다.

각각의 토막에 재추출될 확률을 균등하게 $\frac{1}{R}$ 씩 할당했으므로,

$$E^*(\bar{X}_n^*) = \frac{1}{n} E[k/X_1, \dots, X_n] \cdot \frac{1}{R} (S_1 + S_2 + \dots + S_R) = \frac{E[k/X_1, \dots, X_n]}{R} \cdot \bar{X}_n \tag{13}$$

이 된다.

$E[k/X_1, \dots, X_n]$ 을 구하기 위하여 $\sum_{i=1}^k n_i^* = n$ 의 양변을 $\{X_1, \dots, X_n\}$ 에 조건을 취해 조건부 기대값을 구하면,

$$E[n_i^*/X_1, \dots, X_n] = \frac{1}{R} \sum_{r=1}^R n_r \frac{n}{R} \text{이므로 } E[k/X_1, \dots, X_n] \cdot E[n_i^*/X_1, \dots, X_n] = n \text{이 되고}$$

$$E[kX_1, \dots, X_n] = R \tag{14}$$

이 된다.

(14)식을 (13)식에 대입하면, $E^*(\bar{X}_n^*) = \bar{X}_n$ 이 된다. SB의 결과처럼 TB도 조건부 불편 추정치를 제공함을 보여준다.

5.2 임계값부트스트랩 표본 평균의 분산: $Var^*(\sqrt{n} \cdot \bar{X}_n^*)$

Liu와 Singh(1992)는 MBB의 $Var^*(\sqrt{n} \cdot \bar{X}_n^*)$ 가

$U_i = (\sum_{j=1}^{i+b-1} X_j / b)$, $i=1, 2, \dots, n-b+1$ 의 표본분산인 $Var^*(\sqrt{n} \cdot \bar{X}_n^*) = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} (U_i - \bar{U})^2$ 로 간략해질 수 있는 것을 소개했다.

위식의 $\bar{U} = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} U_i = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} \sum_{j=1}^{i+b-1} X_j / b$ 이다.

따라서,

$$Var^*(\sqrt{n} \cdot \bar{X}_n^*) = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} \left[\left(\sum_{j=1}^{i+b-1} X_j / b \right) - \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} \sum_{j=1}^{i+b-1} X_j / b \right]^2$$

$$= \frac{b}{n-b+1} \sum_{i=1}^{n-b+1} \left[\frac{1}{b} \sum_{j=1}^{i+b-1} X_j - E^*(\bar{X}_n^*) \right]^2 \quad (15)$$

가 된다.

SB의 경우에 Politis와 Romano(1994)는 $Var^*(\bar{Y}_n \cdot \bar{X}_n^*)$ 의 공식을 고리형(circular) 자기공분산이 포함된 형태로 아래처럼 유도하였다.

$$Var^* \left[\bar{Y}_n \cdot \bar{X}_n^* \right] = \hat{C}_n(0) + 2 \sum_{i=1}^{n-1} \left(1 - \frac{i}{n} \right) (1-p)^i \hat{C}_n(i) \quad (16)$$

여기서 $\hat{C}_n(i) = \frac{1}{n} \sum_{j=1}^n [(X_j - \bar{X}_n)(X_{j+i} - \bar{X}_n)]$ 이다.

TB의 경우에는 (12)번 식부터 출발할 수 있다.

$$\begin{aligned} Var^* (\bar{Y}_n \cdot \bar{X}_n^*) &= \frac{1}{n} Var \left[\sum_{i=1}^k S_i^* / X_1, \dots, X_n \right] \\ &= \frac{1}{n} E(k/X_1, \dots, X_n) Var(S_i^* / X_1, \dots, X_n) \\ &\quad + \frac{1}{n} [E(S_i^* / X_1, \dots, X_n)]^2 Var(k/X_1, \dots, X_n) \end{aligned} \quad (17)$$

(17)번 식의 각 항을 계산하면 (14)번 식에서 $E(k/X_1, \dots, X_n) = R$ 이고,

$$E(S_i^* / X_1, \dots, X_n) = \frac{(S_1 + \dots + S_R) n \bar{X}_n}{R} = \bar{X}_n \quad (18)$$

$$\begin{aligned} Var(S_i^* / X_1, \dots, X_n) &= E(S_i^{*2} / X_1, \dots, X_n) - [E(S_i^* / X_1, \dots, X_n)]^2 \\ &= \sum_{i=1}^R \frac{S_i^2}{R} - \frac{n^2 \bar{X}_n^2}{R^2} \end{aligned} \quad (19)$$

이다.

$Var(kX_1, \dots, X_n)$ 을 계산하기 위해서 $\sum_{i=1}^k n_i^* = n$ 의 양변을 $\{X_1, \dots, X_n\}$ 에 조건을 취해 조건부 분산을 구하면

$$Var \left[\sum_{i=1}^k n_i^* / X_1, \dots, X_n \right] = 0 \text{이 되어,}$$

결국

$$E(k/X_1, \dots, X_n) Var(n_i^* / X_1, \dots, X_n) + [E(n_i^* / X_1, \dots, X_n)]^2 Var(k/X_1, \dots, X_n) = 0 \quad (20)$$

이 된다.

두 번째 항의

$$\begin{aligned} Var(n_i^* / X_1, \dots, X_n) &= E(n_i^{*2} / X_1, \dots, X_n) - [E(n_i^* / X_1, \dots, X_n)]^2 \\ &= \sum_{i=1}^R \frac{n_i^2}{R} - \left[\sum_{i=1}^R \frac{n_i}{R} \right]^2 \\ &= \frac{(n_1 + n_2 + \dots + n_R)^2 - 2 \sum_{i < j} n_i n_j}{R} - \frac{n^2}{R^2} \\ &= \frac{n^2 - 2 \sum_{i < j} n_i n_j}{R} - \frac{n^2}{R^2} \end{aligned} \quad (21)$$

이 된다.

세 번째 항에서 $E(n_i^* / X_1, \dots, X_n) = n/R$ 이다.

(14)식, (21)식, (22)식을 (20)식에 대입하면 $n^2 - 2 \sum_{i < j} n_i n_j - \frac{n^2}{R} + \frac{n^2}{R^2}$

$Var(k/X_1, \dots, X_n) = 0$ 이다.

따라서

$$Var(kX_1, \dots, X_n) = R - R^2 + 2 \frac{R^2}{n} \sum_{i < j} n_i n_j \quad (23)$$

이 된다.

(14)식, (18)식, (19)식, (23)식을 (17)식에 대입하면

$$Var^* (\bar{Y}_n \cdot \bar{X}_n^*) = \frac{1}{n} \left[R \left(\sum_{i=1}^R S_i^2 / R - n \bar{X}_n^2 / R^2 \right) + (n \bar{X}_n^2 / R^2) \cdot \left(R - R^2 + 2 \frac{R^2}{n} \sum_{i < j} n_i n_j \right) \right]$$

$$\begin{aligned} &= \frac{1}{n} \left\{ \sum_{i=1}^R S_i^2 - n^2 \overline{X_n^2} / R + n^2 \overline{X_n^2} / R - n^2 \overline{X_n^2} + 2 \overline{X_n^2} \sum_{i,j} n_i n_j \right\} \\ &= \frac{1}{n} \left\{ \sum_{i=1}^R S_i^2 - \overline{X_n^2} (n^2 - 2 \sum_{i,j} n_i n_j) \right\} \\ &= \frac{1}{n} \left\{ \sum_{i=1}^R S_i^2 - \sum_{i=1}^R n_i^2 \overline{X_n^2} \right\} \end{aligned}$$

가 된다.

6. 결론

시계열과 시뮬레이션 출력 등의 자기상관된 자료의 해석을 위한 새로운 부트스트랩 기법(임계값 부트스트랩: Threshold Bootstrap)을 개발하였다. 이 기법을 기존의 시뮬레이션 단일실행 출력분석 기법인 배치평균 방법과 성능을 비교하였다. M/M/1 모델을 이용한 실험결과, TB는 배치평균방법보다 우수한 성능을 나타내었다. 이 기법은 시뮬레이션 실행 시간이 오래 소요되거나, 다수의 자료를 구하기 어려운 시계열 자료를 분석하는데 쓰일 수 있으며 실행방법이 쉽고 체계적 통계 배경이 요구되지 않는다는 장점이 있다.

TB의 약점을 발견하였는데, 그 약점은 TB의 성능이 원시자료군에 음의 상관관계(negative correlation)가 있거나 또는 원래의 시뮬레이션 모델에 주기성(periodicity)이 있으면 그 수행도가 떨어지는 것이다. 이를 극복하기 위하여 자료의 재추출 단위를 변화시켜 그 수행도를 향상시켰다. 재추출시 최적의 자료크기("run"의 크기: 토막(chunk))를 정하기 위하여 numerical search를 수행하였다.

TB의 이론적 배경을 소개하였고, 현재 자기상관된 자료군을 재추출하기 위해 사용되는 기법들의 이론적 배경과도 비교하였다.

자기상관자료를 위한 기타 부트스트랩 방법과 수행도를 비교하였다. 비교 대상으로는 기존의 상관관계하의 자료를 resampling하는 기법들인 이동블록부트스트랩(Moving Block Bootstrap)과 정상부트스트랩(Stationary Bootstrap)을 선정했다. TB는 여러 가지 형태의 시계열 모형에서 성능이 우수함을 보였다.

향후 연구과제로는 대표적인 양의 상관관계를 나타내는 M/M/1 모델과 음의 상관관계를 나타내는 재고관리 모형의 시뮬레이션 모델에 적용하여 TB의 성능을 이동 부

트스트랩과 정상 부트스트랩과 비교하는 것이다.

참고문헌

- [1] Carlstein, E., "The Use of Subseries values for Estimating the variance of a General Statistic from a Stationary Sequence", *The Annals of Statistics*, Vol. 14, NO. 3, pp. 1171-1179, 1986.
- [2] Carlstein, E., "Resampling Techniques for Stationary Time-Series: Some Recent Developments", in *New Directions in Time Series Analysis*, eds. Brillinger, D., Springer-Verlag, New York, pp. 75-85, 1993.
- [3] Davison, A., and Hall, P., "On Studentizing and Blocking Methods for implementing the Bootstrap with Dependent Data", *Australian Journal of Statistics*, Vol. 35, pp. 21-224, 1993.
- [4] Efron, B., "Bootstrap Methods: Another Look at the Jackknife", *The Annals of Statistics*, Vol. 7, pp. 1-26, 1979.
- [5] Efron, B. and Tibshirani R. *An Introduction to the Bootstrap*, Chapman and Hall, New York, 1982.
- [6] Efron, B., "The Jackknife, the Bootstrap and other Resampling Plans", *Society for Industrial and Applied Mathematics*, Philadelphia, 1982.
- [7] Efron, B. and Tibshirani, R., "Bootstrap Methods for Standard Errors, Confidence Intervals, and other Measures of Statistical Accuracy", *Statistical Science*, Vol. 1, No.1, pp. 54-77, 1986.
- [8] Hall, P., Horowitz, J., and Jirg, B., "On Blocking Rules for the Bootstrap with Dependent Data", *Biometrika*, Vol. 82, No.4, pp. 375-386, 1995.
- [9] Kim, Y., "Single Replication Methods for Simulation Experiments", *Ph. D Dissertation*, Decision Science and Engineering Systems, Rensselaer Polytechnic Institute, Troy, New York, 1992.
- [10] Kim Y., Haddock J. and Willemain T., "The Binary Bootstrap: Inference with Autocorrelated Binary Data", *Communications in Statistics: Simulation and Computation* Vol. 22, pp. 205-216, 1993.

- [11] Künch H., "The Jackknife and the Bootstrap for General Stationary Observations", *Annals of Statistics* Vol., 17, pp. 1217-1241, 1989.
- [12] Künsch, H., and Carlstein, E., "The Linked Blockwise Bootstrap for Serially Dependent Observation", Technical Report Dept. of Statistics, Univ. of North Carolina, Chapel Hill, 1990.
- [13] Lahiri, S., "Edgeworth Correction by 'Moving Block' Bootstrap for Stationary and Nonstationary Data", in *Exploring the Limits of the Bootstrap*, eds. R. LePage and L. Billard, John Wiley, NY, pp. 183-214, 1992.
- [14] Lall, U. and Sharma, A., "A Nearest Neighbor Bootstrap for Resampling Hydrology Time Series", *Water Research*, Vol. 32, pp. 679-693, 1996.
- [15] Law, A. and Kelton, W., *Simulation Modeling and Analysis*, 2nd ed. McGraw-Hill, New York, 1991.
- [16] Leger, C., Politis D. and Romano J., "Bootstrap Technology and Applications", *Technometrics* Vol. 34, pp. 378-398, 1992.
- [17] Liu, R., "Bootstrap Procedures under Some Non-I.I. D. Models", *Technometrics*, Vol. 16, No.4, pp. 1696-1708, 1988.
- [18] Liu, R. and Singh, K., "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence", in *Exploring the Limits of Bootstrap*, eds. R. LePage and L. Billard, John Wiley, New York, pp. 225-248, 1992.
- [19] Politis, D. and Romano J., "The Stationary Bootstrap", in *Exploring the Limits of Bootstrap*, eds. R. LePage and L. Billard, John Wiley, New York, 1992.
- [20] Politis, D. and Romano, J., "The Stationary Bootstrap", *Journal of the American Statistical Association*, vol.89, No.428, pp. 1303-1313, 1994.
- [21] Thoms, L. and Schucany, W., "Bootstrap prediction Intervals for Autoregression", *Journal of the American Statistical Association*, Vol. 85, pp. 486-492, 1990.
- [22] Singh, K., "On the Asymptotic Accuracy of Efron's Bootstrap", *The Annals of Statistics*, Vol. 9, NO. 6, pp. 1187-1195, 1981.
- [23] Whitt, W., "Planning Queueing Simulation", *Management Science*, 35: 11, pp. 1341-1366, 1989.