

## 시계열자료에서 결측치 추정방법의 비교

이성덕<sup>1,a</sup>, 최재혁<sup>b</sup>, 김덕기<sup>a</sup>

<sup>a</sup>충북대학교 정보통계학과, <sup>b</sup>성균관대학교 통계학과

### 요약

시계열의 결측값은 미지의 모수로 취급될 수 있으며 최대우도방법 또는 확률변수방법에 의해 추정할 수 있으며 또한 주어진 자료 하에서 미지의 값에 대한 조건부기대치로 예측할 수 있다. 이 연구의 주된 목적은 불안정한 자료에 대해 ARMA 모형을 적용하여 두 가지 추정방법인 최대우도추정방법과 확률변수방법을 이용해 결측값을 대체하는 방법을 비교하는데 있다.

사례분석을 위해 한국질병관리본부에서 전산보고 하고 있는 전염병 자료 중에서 2001~2006년 동안의 월별 Mumps 자료를 이용하여 앞의 두 가지 추정방법을 예측오차제곱합(SSF)을 구하여 비교한다.

주요용어: 최대우도추정법, 확률변수법, ARMA 모형, Mumps 자료, 예측오차제곱합.

### 1. 서론

연구에 이용되는 자료들은 다양한 이유로 완전하지 않은 경우가 대부분이다. 따라서 실제의 불안정한 자료에서 새로운 결과를 만들어내야 한다. 실제로 완전한 관찰값은 결측값과 측정값의 합으로 생각해야 한다. 즉, 결측자료 또는 결측값이 존재하더라도 자료분석 방법은 달라지지 않고 전통적인 방법을 이용하여 분석한다.

결국 완전한 관찰값을 만들기 위해 결측값을 적절한 방법으로 대체시켜야 하는데 그 방법은 많은 이들로부터 연구 되어 왔다. 그 중 시계열 자료에서 존재하는 결측값을 대체시키는 방법은 크게 두 가지로 나뉘어져 있다. 첫째는 최우추정법을 이용하여 결측값을 대체하는 방법이고 둘째는 베이저안 방법을 통한 확률변수방법이다.

일반적으로 결측값을 추정하기 위한 최우추정법의 이용이나 확률변수의 결측값의 이용은 논쟁거리이다. 왜냐하면 연구자별로 추정량을 얻기 위해 각기 다른 우도함수를 사용하기 때문이다 (Bayarri 등, 1986).

시계열자료에서도 각 모형에 따라 우도함수가 다르기 때문에 이 문제로부터 자유롭지 못하다. 그러나 대부분의 연구자들은 ARMA모형을 이 문제에 적용시켜왔다. 시계열 모형중 가장 기본이 되는 모형이기 때문에 ARMA모형에 적용이 가능하다면 다른 어떠한 모형도 적용하기 쉽기 때문이다. Dunsmuir과 Robinson (1981)은 결측값이 존재할 때 결측값 대체방법으로 우도함수를 이용한 최우추정방법을 제안했고, Pena와 Tiao (1991)은 결측값 대체방법으로 확률변수방법을 제안하여 최우추정방법과 비교하였다. 두 연구 모두 ARMA모형을 통해 연구되었다.

본 논문은 실제 시계열자료를 사용하여 동일한 결측값을 최우추정방법과 확률변수방법을 이용하여 대체하고 ARMA모형에 적합시켜 두 방법에 따른 예측결과를 비교분석하였다. 실증분석 사례로 사용된 자료는 한국질병관리본부에서 제 2군 전염병으로 관리하고 있는 Mumps자료로써, 시간과 공간

이 논문은 2007학년도 충북대학교 학술연구지원 사업의 연구비 지원에 의하여 연구되었음.

<sup>1</sup> 교신저자: (361-763) 충북 청주시 흥덕구 충북대학교 정보통계학과, 교수. Email: sdlee@chungbuk.ac.kr

에 따라 전염성이 매우 강한 특징을 가지고 있다. 따라서 본 논문에서는 16개 시도 자료를 8개의 도별 자료로 재분류하여 공간에 대한 특성을 고정하여 분석하였다. 결측값이 제외된 Mumps 자료를 이용하여 모형식별과 모수를 추정하고 그에 따라 앞에서 추정된 모수를 사용하여 언급한 두 가지 추정방법을 통해서 결측값을 추정하였으며 또한 추정의 정도를 비교하였다. 또, 실제데이터를 이용하여 추정된 결측값을 대체한 후에 예측력을 비교하였다.

본 논문의 구성은 2절에서 최우추정방법과 확률변수방법으로 결측값을 대체하는 방법을 비교 설명하였고, 3절에서는 사례연구로써 Mumps 자료에 대해 결측값을 추정하여 추정의 정도를 비교하고, 4절에서는 각 방법에 대한 도별 Mumps 자료에 대한 월별 예측력을 비교하였고 5절에서는 결론 및 향후 연구 과제를 제시하였다.

## 2. 시계열 모형에서 결측값 대체

### 2.1. 최우추정법에 의한 결측값 대체

시계열 모형에서 결측값 대체방법을 살펴보기 위해 시계열  $z_t$ 는 정상 시계열이라고 가정하자.

$$x_t = \phi z_{t-1} + a_t, \quad |\phi| < 1, \quad a_t \stackrel{iid}{\sim} N(0, \sigma^2).$$

계산상의 편의를 위해  $\phi$ 와  $\sigma^2$ 는 알려져 있다고 가정하자.  $n$ 개의 관찰값  $z_t (t = 1, \dots, n)$  중  $z_T$ 를 결측값이라 가정하면  $(1 \leq T \leq n)$ 에 대하여,  $Z_n = (z_1, \dots, z_n)'$ 는  $n \times 1$  벡터로 정의되고  $Z_{(T)}$ 는  $Z_n$ 에서  $z_T$ 를 제외하고 얻어진  $(n-1) \times 1$  벡터로 정의된다.  $z_T$ 가 주어진 이용 가능한 자료  $Z_{(T)}$ 의 결합밀도함수는 다음과 같다.

$$f(Z_{(T)} | z_T) = \frac{f(Z_n)}{f(z_T)} \quad (2.1)$$

여기서,

$$f(Z_n) = (2\pi\sigma^2)^{-n/2} (1 - \phi^2)^{1/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[ (1 - \phi^2) z_1^2 + \sum_{i=2}^n (z_i - \phi z_{i-1})^2 \right] \right\} \quad (2.2)$$

$$f(z_T) = (2\pi\sigma^2)^{-1/2} (1 - \phi^2)^{1/2} \exp \left\{ -\frac{1}{\sigma^2} [(1 - \phi^2) z_T^2] \right\} \quad (2.3)$$

이다.

(2.1)에서  $Z_T$ 는  $Z_{(T)}$ 에 대한 미지의 모수로 취급할 수 있다. 그러면  $Z_T$ 의 우도함수는 다음과 같다.

$$l(Z_T | Z_{(T)}) = (2\pi\sigma^2)^{-(n-1)/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^{T-1} (z_i - \phi z_{i+1})^2 + \sum_{i=T+1}^n (z_i - \phi z_{i-1})^2 \right] \right\} \quad (2.4)$$

여기서 합의 범위가 양의 값이 아니라면 지수항은 보이지 않는다. 그러므로  $z_T$ 의 최우추정량은 다음과 같다.

$$\hat{z}_T = \begin{cases} \phi^{-1} z_{T+\delta}, & T = 1 \text{ 또는 } n, \\ (2\phi)^{-1} (z_{T+1} + z_{T-1}), & 1 < T < n. \end{cases} \quad (2.5)$$

이때  $T = 1$ 이면  $\delta = 1$ 이고  $T = n$ 이면  $\delta = -1$ 이다. 또한  $\hat{Z}_T$ 의 MSE는 (2.4)을 통해 다음과 같이 구할 수 있다.

$$\text{MSE}(\hat{z}_T) = \begin{cases} \phi^{-2} \sigma^2, & T = 1 \text{ 또는 } n, \\ (2\phi^2)^{-1} \sigma^2, & 1 < T < n. \end{cases} \quad (2.6)$$

이를 ARMA( $p, q$ )에 적용하면 다음과 같이 최우추정법에 의해 결측값을 대체할 수 있다.

$$\hat{z}_T = \sum_{i=1}^p \frac{(z_{T+i} + z_{T-i})}{2\phi_i} + \sum_{j=1}^q \frac{(z_{T+j} + z_{T-j} - \hat{z}_{T+j} + \hat{z}_{T-j})}{2\theta_j}. \quad (2.7)$$

### 2.2. 확률변수에 의한 결측값 대체

그 동안 알려진 자료에 대한 결측값에 관하여 식 (2.2)의 결합밀도함수  $f(Z_n)$ 을 최대로 하는 최소제곱추정량이나 최우추정량을 이용하여 대체하는 연구들이 활발히 진행되어왔다 (Brubacher와 Wilson, 1976). 이 경우 추정량은 다음과 같이 쉽게 얻어진다.

$$\tilde{z}_T = \begin{cases} \phi z_{T+\delta}, & T = 1 \text{ 또는 } n, \\ (1 + \phi^2)^{-1}(z_{T+1} + z_{T-1}), & 1 < T < n. \end{cases} \quad (2.8)$$

그러나 이 추정량  $\tilde{z}_T$ 는 미지의 모수로 고려되는  $Z_T$ 에 대한 함수  $f(Z_n)$ 가 결합밀도함수가 아니기 때문에 최우추정량이 아니다. 그러므로 결측값  $z_T$ 와 측정값  $Z_{(T)}$ 에 대한  $f(Z_n)$ 는 우도함수가 아니다.

식 (2.8)의 의미를 설명하기 위해 식 (2.2)의 확률구조를 따르는 확률변수  $z_T$ 를 고려하자. 그러면  $Z_{(T)}$ 가 주어진  $z_T$ 의 분포는 다음과 같다.

$$f(z_T | Z_{(T)}) = \frac{f(Z_n)}{f(Z_{(T)})} \quad (2.9)$$

여기서  $f(Z_{(T)})$ 는  $f(Z_n)$ 으로 부터  $z_T$ 를 제거한 적분값으로 얻어진다. 잘 알려진 대로 식 (2.9)의 분포는 다음과 같은 정규분포이다.

$$E(z_T | Z_{(T)}) = \tilde{z}_T, \quad (2.10)$$

$$\text{VAR}(z_T | Z_{(T)}) = \begin{cases} \sigma^2 & T = 1 \text{ 또는 } n, \\ (1 + \phi^2)^{-1} \sigma^2, & 1 < T < n. \end{cases} \quad (2.11)$$

따라서 이를 ARMA( $p, q$ )에 적용하면 다음과 같이 확률변수방법에 의해 결측값을 대체할 수 있다.

$$\hat{z}_T = \sum_{i=1}^p \frac{(z_{T+i} + z_{T-i})}{(1 + \phi_i^2)} + \sum_{j=1}^q \frac{(z_{T+j} + z_{T-j} - \hat{z}_{T+j} + \hat{z}_{T-j})}{(1 + \theta_j^2)} \quad (2.12)$$

여기서,  $\phi_i$ 은  $i$ -번째 자기회귀계수,  $\theta_j$ 은  $j$ -번째 이동평균계수를 나타낸다.

### 2.3. 최우추정법과 확률변수방법의 비교

$f(Z_n)$ 은  $f(z_T | Z_{(T)})$ 와 비례하기 때문에 조건부 기대값  $E(z_T | Z_{(T)})$ 은 (2.8)의  $\tilde{z}_T$ 와 같다.  $z_T$ 의 최소 MSE 추정량인  $E(z_T | Z_{(T)})$ 가 (2.5)의 최우추정량  $\hat{z}_T$ 과 다르다는 점은 매우 중요하다. (2.6)의  $\hat{z}_T$ 의 MSE는 항상 (2.11)의  $\text{VAR}(z_T | Z_{(T)})$ 보다 크다. 물론 추정량  $E(z_T | Z_{(T)})$ 의 MSE 역시 마찬가지다.

두 개의 추정량  $\tilde{z}_T$ 과  $\hat{z}_T$ 의 차이점은 베이저안 관점에서 보면 매우 간단하다. 추정량  $\tilde{z}_T$ 은 (2.9)의 사후분포  $f(z_T | Z_{(T)})$ 의 평균 또는 최빈값이다. 이것은 (2.4)의 우도함수  $l(z_T | Z_{(T)})$ 와 (2.3)의 사전분포  $f(Z_T)$ 의 곱에 비례한다. 반면, 추정량  $\hat{z}_T$ 은  $l(z_T | Z_{(T)})\rho_0(z_T)$ 과 비례하는  $z_T$ 의 사후분포의 평균 또는 최빈값이다. 여기서  $\rho_0(z_T)$ 는 국소균일(locally uniform)이고 비정보적 사전분포(noninformative prior)이

다 (Box와 Tiao, 1973). 즉, 정상성 계열의 경우( $|\phi| < 1$ ) 매우 다른 사전분포를 사용하기 때문에 두 평균은 매우 다르다. 이것은  $\phi$ 가 1에 가까워지면(비정상성 모형) 사전분포  $f(z_T)$ 역시 국소균일이 되기 때문에 이 두 추정량의 차이가 없다는 것을 의미한다.

이것은 두 추정량은 매우 다른 가정 하에 얻어지기 때문에 (2.6), (2.11)의  $\tilde{z}_T$ 과  $\hat{z}_T$ 의 최적 특성을 실제로 비교할 수 없다는 빈도학파(frequentist) 관점에서는 논쟁의 여지가 있다. 최우추정량  $\hat{z}_T$ 은 결측값  $z_T$ 를 고정된 모수로 간주하는 가정 하에 (2.6)에 의해  $MSE(\hat{z}_T)$ 를 얻는다. 반면 추정량  $\tilde{z}_T$ 은  $\hat{z}_T$ 을 (2.2)의 구조를 따르는 확률변수로 간주하고 (2.11)에 의해  $MSE(\tilde{z}_T) = \text{VAR}(z_T|Z_{(T)})$ 가 얻어진다.  $\tilde{z}_T$ 의 MSE는 다음과 같다.

$$MSE^*(\tilde{z}_T) = \begin{cases} \phi^2\sigma^2 + (1 + \phi^2)^2 z_T^2, & T = 1 \text{ 또는 } n, \\ (1 + \phi^2)^{-2} [2\phi^2\sigma^2 + (1 - \phi^2)^2 z_T^2], & 1 < T < n. \end{cases} \quad (2.13)$$

따라서 어떠한 값  $z_T$ 에 대해서도  $MSE^*(\tilde{z}_T)$ 는  $MSE(\hat{z}_T)$ 보다 작다.

앞서 살펴보았듯이 시계열 자료에서의 결측값의 대체에서 최우추정법은 정상성 가정 하에서 최적의 예측값을 얻는 것과 매우 다른 결과를 가져온다. 결측값을 고정된 모수로 간주하는 것은 모순적인 것으로 생각될 수 있다. 시계열 분석에서 정상성을 만족한다면 결측값을 확률변수로 간주하는 것이 적용하기에 더 일반적일 것이다. 그러나 최적추정량처럼 조건부 기대값인지 사후분포의 평균인지는 분석가의 선택에 달려있다. 본 논문에서는 이 두 방법들을 사례연구를 통해 비교해 보았다.

### 3. 사례연구

#### 3.1. 자료설명

유행성이하선염(이하 Mumps)은 홍역, B형 간염 등과 더불어 우리나라에서 제 2군 전염병으로 분류하고 있으며, 직접적인 비말(droplet) 또는 다른 오염된 물질이 코나 입으로 들어가서 감염된다. Mumps는 1998년 4,461명의 환자가 발생한 이후 점차 감소하는 추이를 보이지만 여전히 전국에 걸쳐 높은 발생률을 보이는 전염병으로 알려져 있다 Mumps는 뇌수막염, 고환염, 부고환염, 난소염, 췌장염 등의 합병증을 동반하고 일단 감염되면 특이요법이 없어 예방을 최우선으로 한다. 이에 예방을 위해 우리나라에서는 생후 12 ~ 15개월과 4 ~ 6세에 MMR 백신 (홍역-유행성이하선염(볼거리) -풍진 혼합백신)접종을 추천하고 있다.

본 논문에서 사용된 Mumps는 한국질병관리본부(Korea Center for Disease Control and Prevention)에서 2001년 1월부터 2006년 9월까지 전염성 감시체계로 전산 보고된 16개 시도의 월별 Mumps 자료이다. 여기서 한국의 Mumps 자료는 8개 도로 재분류하였다. 사용된 Mumps 자료는 연속성 자료가 아닌 빈도 자료로 Poisson분포를 따르는 변수이므로, 원래의 자료를 분산안정화변환(여기서,  $\sqrt{Z}$ )과 12차 계절차분을 취한 후 정상시계열이 되었고, 이 자료를 정규분포로 근사시키기 위해 각 도 별로 평균과 표준편차를 구해 자료를 표준화 하였다. 여기서 2005년 5월 자료가 결측값이고 2001년 1월 ~ 2005년 12월 자료까지의 월별 자료만을 이용하여 모형을 추정, 결측값을 대체했고 나머지 자료(2006년 1월 ~ 2006년 9월)는 예측력을 비교하는데 사용했다.

#### 3.2. 모형식별 및 추정과 결측값 대체

모형식별은 자기상관함수(Autocorrelation Function: ACF)와 편차자기상관함수(Partial Autocorrelation Function: PACF) 등의 감소패턴을 보면서 차수를 식별하였고 AIC, BIC등을 통해 모형을 선택하였다. 이에 8개 도의 식별된 모형과 모수추정 결과는 표 1과 같다.

표 1: 모형식별 및 모수추정결과

지역	모형	모수추정	AIC
경기	ARMA(1,1)	$\phi = 0.91430, \theta = 0.55868$	113.73
강원	ARMA(0,1)	$\theta = -0.42854$	128.82
충남	ARMA(1,0)	$\phi = 0.36068$	128.28
충북	ARMA(0,1)	$\theta = -0.44161$	126.63
전북	ARMA(1,1)	$\phi = 0.80889, \theta = 0.42455$	120.52
전남	ARMA(1,0)	$\phi = 0.44825$	124.57
경북	ARMA(1,0)	$\phi = 0.71602$	102.85
경남	ARMA(0,1)	$\theta = -0.31073$	130.51

표 2: 결측값 추정결과

지역	관측값(2005년 5월자료)	최우추정방법(MLE)	확률변수방법(RV)
경기	129	187	142
강원	16	0	0
충남	5	66	42
충북	26	23	22
전북	4	2	5
전남	5	0	0
경북	16	2	3
경남	35	3	4

표 3: 결측값 대체 후 모수추정결과

지역	모형	최우추정방법	확률변수방법
경기	ARMA(1,1)	$\phi = 0.91648, \theta = 0.56120$	$\phi = 0.91352, \theta = 0.55245$
강원	ARMA(0,1)	$\theta = -0.41189$	$\theta = -0.41189$
충남	ARMA(1,0)	$\phi = 0.40654$	$\phi = 0.40489$
충북	ARMA(0,1)	$\theta = -0.45496$	$\theta = -0.45495$
전북	ARMA(1,1)	$\phi = 0.80251, \theta = 0.40790$	$\phi = 0.81504, \theta = 0.43090$
전남	ARMA(1,0)	$\phi = 0.45562,$	$\phi = 0.45562$
경북	ARMA(1,0)	$\phi = 0.72235$	$\phi = 0.72247$
경남	ARMA(0,1)	$\theta = -0.30128$	$\theta = -0.30496$

결측값인 2005년 5월 자료에 대해 표 1과 같이 선택된 모형에 따라 최우추정방법과 확률변수방법으로 결측값을 추정한 결과가 표 2와 같다.

결과를 살펴보면, 8개 도의 결측값 추정결과 강원, 전남의 경우는 같았고, 충북을 제외한 나머지 5개 지역의 경우 확률변수방법이 관측 값에 좀더 가깝게 추정되었음을 알 수 있다. 추정된 결측값을 대체하여 완비자료에 대한 모수추정을 한 결과는 표 3과 같다.

표 2의 8개 도의 결측값에 대한 최대우도추정방법과 확률변수방법의 추정값에 차이가 비교적 나는 경우에만 표 3의 모수추정치에 다소 차이가 남을 알 수 있다. 이는 각 도의 결측치의 수가 1개인 경우의 추정된 결측값을 대체하여 모수를 추정한 것으로 결측치의 수가 여러 개로 늘어나면 모수추정 및 AIC의 값에 차이가 다소 확연할 것이다.

#### 4. 예측결과

본 논문에 사용된 Mumps 자료는 추정에 2001년 1월 ~ 2005년 12월 자료까지의 월별자료를 활용하였고 나머지 자료(2006년 1월 ~ 2006년 9월)는 예측력을 파악하는데 사용하였다. 최우추정법과 확률변수방법에 의해 결측값을 대체한 후의 모형의 예측력 비교는 아래의 예측오차 제곱합(Sum of

표 4: 8개 도의 월별 Mumps 자료에 대한 예측오차제곱합의 비교

지역	최우추정방법(SSF)	확률변수방법(SSF)
경기	13632	8167
강원	496	496
충남	5715	2936
충북	2083	2046
전북	197	173
전남	222	222
경북	1410	1377
경남	8272	8102

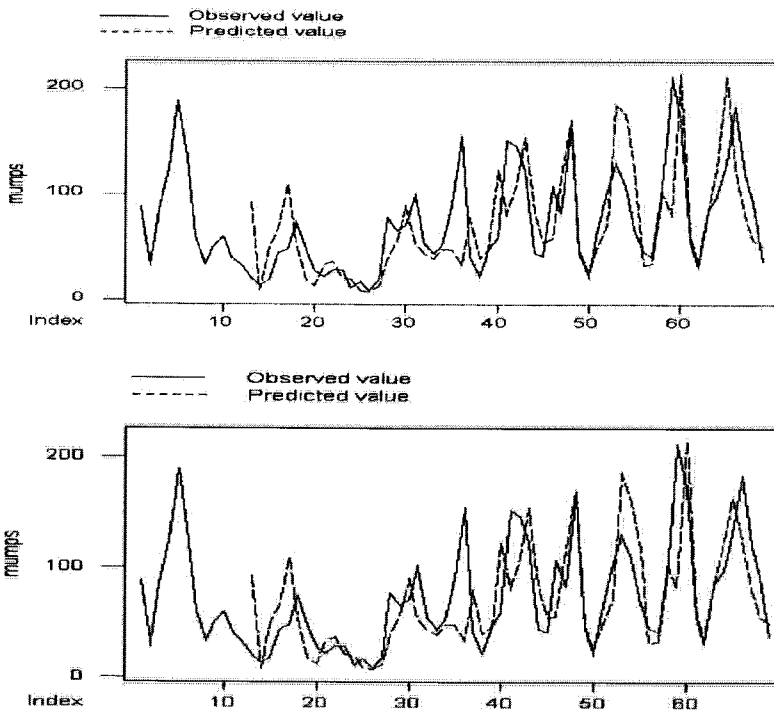


그림 1: 최우추정방법(상단)과 확률변수방법(하단)에 의한 대체 후의 예측결과

square for forecasting error: SSF)을 사용하였고 8개 도의 Mumps 월별자료에 대한 관측값과 예측값의 비교결과는 표 4와 같다.

$$SSF = \sum_{i=1}^9 [\text{observed value} - \text{predicted value}]^2$$

예측오차 제곱합(SSF)을 비교해 보면 강원, 전남은 동일한 예측값을 보였고, 그 이외의 지역에 대한 예측력의 결과를 보면 확률변수방법이 최우추정방법에 비해 더 좋은 예측력을 보여주고 있음을 알 수 있다.

그림 1은 두 가지 방법에 의한 결측치 대체 후의 8개 도의 예측결과 자료중 경기도에 대한 예측적합을 보여주는 그림이다.

## 5. 결론 및 향후과제

본 논문에서는 시계열자료에서 존재하는 결측값을 대체하는 방법으로 최우추정방법과 베이저안 방법의 확률변수방법을 이용해 8개도의 Mumps 자료에 대해 각각 1개의 결측값을 만들고 두 가지 추정 방법에 따른 결측값을 추정하여 그 추정의 정도를 비교한 결과 대부분의 지역에서 확률변수방법이 추정의 정도가 더 좋았으며, 두 가지 방법에 의해 추정된 결측값을 각각 대체한 후 추가적으로 미래 9개월에 대한 예측력을 예측오차제곱합(SSF)를 구하여 비교한 결과 역시 확률변수방법이 좋은 예측력을 보임을 알 수 있었다.

본 논문에서는 공간을 고정해 놓고 시간만을 고려한 선형시계열모형인 ARMA모형에 대해서 결측치 대체방법을 살펴보고, 추후 연구계획으로 ARMA모형을 포함하는 보다 폭넓은 비선형시계열모형인 중선형모형(Bilinear Model)과 시간과 공간을 동시에 고려한 공간시계열 중선형모형(Space-Time Bilinear Model)으로 확장하여 시간과 공간에 종속되어 있는 결측치의 대체방법에 대해 연구할 계획이다.

## 참고 문헌

- Bayarri, M. J., DeGroot, M. H. and Kadane, J. B. (1986). What is the Likelihood Function? In: *Statistical Decision Theory and Related Topics IV, Volume 1.*, (S. S. Gupta and J. O. Berger eds), New York: Springer-Verlag.
- Box, G. E. P. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis*, Reading, M. A, Addison-Wesley.
- Brubacher, S. R. and Wilson, T. (1976). Interpolating time series with application to the estimation of holiday effects on electricity demand, *Applied statistics*, **25**, 107–116.
- Dunsmuir, W. and Robinson, P. M. (1981). Estimation of time series models in the presence of missing data, *Journal of the American Statistical Association*, **76**, 560–68.
- Pena, D. and Tiao, G. C. (1991) A note on likelihood estimation of missing values in time series, *The American Statistician*, **45**, 212–213.

# The Comparison of Imputation Methods in Time Series Data with Missing Values

SungDuck Lee<sup>1,a</sup>, JaeHyuk Choi<sup>b</sup>, DuckKi Kim<sup>a</sup>

<sup>a</sup>Department of Information and Statistics, Chungbuk National University

<sup>b</sup>Department of Statistics, Sung Kyun Kwan University

---

## Abstract

Missing values in time series can be treated as unknown parameters and estimated by maximum likelihood or as random variables and predicted by the expectation of the unknown values given the data. The purpose of this study is to impute missing values which are regarded as the maximum likelihood estimator and random variable in incomplete data and to compare with two methods using ARMA model. For illustration, the Mumps data reported from the national capital region monthly over the years 2001 ~ 2006 are used, and results from two methods are compared with using SSF(Sum of square for forecasting error).

**Keywords:** Maximum likelihood estimation, random variables, ARMA model, Mumps data, sum of square for forecasting error.

---

---

This Paper was supported by the Chungbuk National University in 2007.

<sup>1</sup> Corresponding author: Professor, Department of Information and Statistics, Chungbuk National University, Chungbuk 361-763, Korea. E-mail: sdlee@chungbuk.ac.kr