

정규확률변수 관측치열에 대한 베이지안 변화점 분석 : 서울지역 겨울철 평균기온 자료에의 적용 *

김경숙 ¹⁾ 손영숙 ²⁾

요약

본 논문에서는 일변량 정규분포를 따르는 확률변수의 관측치열에 대한 변화점 문제(change point problem)를 고찰한다. 변화점의 존재유무, 그리고 만일 변화점이 존재한다면 어떠한 유형으로 발생했는지 즉, 변화점 발생 이후로 평균만 변화, 분산만 변화, 또는 평균과 분산 모두가 변화했는지를 밝힌다. 가능한 여러 유형의 변화모형들 가운데 최적의 모형을 선택하기 위해 베이지안 모형선택 기법을 이용하고, 선택된 모형에 내재된 모수를 추정하기 위해 메트로폴리스-헤스팅스 알고리즘을 포함한 깃스샘플링을 이용한다. 이러한 방법론은 모의실험을 통해 검토되고, 또한 서울지역의 겨울철 평균기온 자료에 적용된다.

주요용어: 변화점 분석, 내재적 베이즈인자(intrinsic Bayes factor), 베이지안 모형선택, 모수추정, 깃스 샘플링, 메트로폴리스-헤스팅스 알고리즘, 일변량 정규분포

1. 서론

관측된 자료 열의 특성이 어느 시점에서 이전의 흐름과는 다르게 변화하는 변화점 문제(change point problem)는 여러 분야, 예를들어 품질관리, 공학, 환경과학, 생물학, 경제학, 수문학, 기상학, 신호처리 등 매우 많은 분야에서 발생한다. 변화점 문제는 일반적으로 최우도 방법 혹은 베이지안 방법과 같은 모수적 방법, 비모수적 방법, 또는 의사결정론적 방법 등에 의해 분석된다.

변화점 문제에 대한 논의는 Maquire 등 (1952)에 의해 상수 척도모수를 갖는 일모수 지수분포를 따르는 석탄광산의 폭발자료에 대해 척도모수의 변화점을 추정하는데서 비롯되었다. Page의 일련의 논문(1954, 1955, 1957)에 의해 변화점 문제는 하나의 연구분야로서 자리매김 되었는데, 특히 품질관리 분야에서 변화점을 탐지하기 위해 비모수적 방법인 CUSUM(cumulative sums) 기법을 이용하였다. 그 후, 1964년경부터는 대부분 모수적 통계 방법에 근거한 연구가 활발히 진행되었고, 특히 Chernoff와 Zacks (1964)는 정규분포를 따르는 독립적인 확률변수열에 있어서 평균수준에 대한 변화점 탐지를 위해 베이지안 검정

* 본 연구는 한국과학재단 목적기초연구 과제번호 R05-2000-000-00024-0 지원으로 수행되었음.

1) (500-757) 광주광역시 북구 용봉동 300 전남대학교 자연과학대학 통계학과 박사수료

E-mail: ksook620@stat.chonnam.ac.kr

2) (500-757) 광주광역시 북구 용봉동 300 전남대학교 자연과학대학 통계학과 교수

E-mail: ysson@chonnam.ac.kr

기법을 도출하였다. 정규확률변수열에 대한 변화점 문제를 베이지안 방식으로 접근한 다른 연구로는 Smith (1975), Lee와 Heghinian (1977), 그리고 Menzefricke (1981) 등이 있다.

최근에 Perreault 등 (2000a,b)은 일변량 정규분포를 따르는 관측치열에서 하나의 변화점이 존재하는 경우에 대해 변화점 분석을 위해 이전 연구보다 더 일반적인 형태의 베이지안 기법으로 접근하였다. 즉, 확률변수열 $\{X_t, t = 1, 2, \dots, n\}$ 에 대해 다음과 같은 확률모형을 가정하였다.

$$X_t \sim \begin{cases} N(\mu_1, \sigma_1^2), & t = 1, 2, \dots, r, \\ N(\mu_2, \sigma_2^2), & t = r + 1, r + 2, \dots, n, \end{cases}$$

여기서 r 은 가정된 변화점으로서 만일 이를 기준으로 $\mu_1 = \mu_2$ 이고 $\sigma_1^2 = \sigma_2^2$ 이면 관측치열은 변화가 발생하지 않은(즉, 변화점이 존재하지 않는) 형태이고, $\mu_1 \neq \mu_2$ 이고 $\sigma_1^2 = \sigma_2^2$ 이면 평균만 변화한 형태, $\mu_1 = \mu_2$ 이고 $\sigma_1^2 \neq \sigma_2^2$ 이면 분산만 변화한 형태, $\mu_1 \neq \mu_2$ 이고 $\sigma_1^2 \neq \sigma_2^2$ 이면 평균과 분산이 모두 변화한 형태로 분류하는 4가지 유형의 모형을 고려하였다. 여기에 수문기상 자료를 적용하여 변화점의 존재여부, 유형 및 변화의 정도 등을 검토하였다. 이때 모수들에 대해서는 공액 사전분포(conjugate priors)를 가정하고 보통의 베이지안자(Bayes factor)를 사용하여 모형선택을 하였고, 또한 선택된 모형의 모수추정을 위하여 깁스 샘플링(Gibbs sampling)을 수행하였다.

Perreault 등 (2000a,b) 이전의 정규확률변수열의 변화점 분석에 관한 논의는 일반적인 모형의 형태가 무변화(no change) 모형 대 평균변화(mean change) 모형, 무변화 모형 대 분산변화(variance change) 모형, 혹은 무변화 모형 대 평균 및 분산변화(mean and variance change) 모형으로 구성되었다. 그러나 이러한 모형실정보다는 4가지 모형을 동시에 비교하여 이 중에서 최적의 하나의 모형을 선택하는 문제가 훨씬 더 현실적일 것이다.

본 논문에서는 Perreault 등 (2000a,b)에 의해 연구되었던 일변량 정규분포를 따르는 4가지 유형의 모형에 대해 베이지안 기법을 적용하여 최적모형을 선택하고, 변화점 발생 위치를 탐지하며, 또한 선택된 모형에 내재된 모수를 추정함으로써 변화점 모형을 명확히 규명하고자 한다. 모수들에 대한 사전분포로서는 정규분포에 관련한 모수들에 대해서는 무정보 부적절 사전분포(noninformative improper priors)를 가정하고, 변화점 모수에 대해서는 이산적 균일분포를 가정함으로써 모수에 대한 최소한의 정보 하에서 베이지안 분석을 시도한다. 모형선택을 위한 베이지안자로는 Berger와 Pericchi (1996, 1998)가 제안한 내재적 베이지안자(intrinsic Bayes factor: IBF)의 산술평균(arithmetic IBF: AIBF), 중앙값(median IBF: MIBF), 그리고 기하평균(geometric IBF: GIBF)을 이용한다.

본 논문의 2절에서는 모형선택을 위한 통계적 도구인 AIBF, MIBF, GIBF와 이를 이용하는 각 모형의 사후확률과 변화점의 사후확률을 소개하고, 이를 산출하기 위해 필요한 주변확률밀도함수를 계산한다. 3절에서는 최적의 모형에 포함되어 있는 모수들을 추정하기 위한 방법으로서 깁스 샘플러(Gibbs sampler)와 메트로폴리스-헤스팅스(Metropolis-Hastings: M-H) 알고리즘 적용방법을 소개한다. 또한, 깁스 샘플러에 이용되는 완전조건부 사후분포(full conditional posterior distribution)를 계산한다. 4절과 5절에서는 모형선택 및 모수추정에 대한 모의실험과 실제 자료에 적용한 결과를 살펴봄으로서 변화점 분석을 위

한 베이지안 모형선택 방법 및 모수추정 방법에 대한 논리적 근거를 제시한다. 마지막으로 6절에서는 일반적인 토론과 결론을 언급한다.

2. 모형선택

2.1. 베이지안 모형선택 기법

정규분포를 따르는 서로 독립인 확률변수열 $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$ 내에 하나의 변화점이 존재하는지 검토하고, 만일 변화점이 존재한다면 그 점을 중심으로 어떠한 유형의 변화가 발생했는지를 논의한다. 본 논문에서 고려하는 4가지 유형의 모형들은 다음과 같다.

$$\begin{aligned}
 M_0 &: \mathbf{X}_t \sim N(\mu_1, \sigma_1^2), \quad t = 1, 2, \dots, n, \\
 M_{1,r_1} &: \mathbf{X}_t \sim \begin{cases} N(\mu_1, \sigma_1^2), & t = 1, 2, \dots, r_1, \\ N(\mu_2, \sigma_1^2), & t = r_1 + 1, r_1 + 2, \dots, n, \end{cases} \\
 M_{2,r_2} &: \mathbf{X}_t \sim \begin{cases} N(\mu_1, \sigma_1^2), & t = 1, 2, \dots, r_2, \\ N(\mu_1, \sigma_2^2), & t = r_2 + 1, r_2 + 2, \dots, n, \end{cases} \\
 M_{3,r_3} &: \mathbf{X}_t \sim \begin{cases} N(\mu_1, \sigma_1^2), & t = 1, 2, \dots, r_3, \\ N(\mu_2, \sigma_2^2), & t = r_3 + 1, r_3 + 2, \dots, n, \end{cases}
 \end{aligned}$$

여기서 $\mu_1 \neq \mu_2$, $\sigma_1^2 \neq \sigma_2^2$, 그리고 각 모형 M_{j,r_j} 에서 r_j ($j = 1, 2, 3$)는 변화점을 의미하며 $r_j \in I$ 이다. I 는 모든 변화 모형 M_{j,r_j} ($j = 1, 2, 3$)의 모수들이 추정될 수 있도록 구성되어야 하는 양의 정수들의 집합이다. 이 때 M_0 는 무변화 모형, M_{1,r_1} 은 평균변화 모형, M_{2,r_2} 는 분산변화 모형, M_{3,r_3} 는 평균 및 분산변화 모형에 해당한다.

베이지안 모형선택 방법의 기본원리는 각 모형에 대한 지지(support) 정도를 모형간의 비(ratio)로서 표현한 베이즈인자를 이용하여 각 모형에 대한 사후확률을 계산한 후, 이 중 가장 높은 사후확률을 갖는 모형을 선택하는 것이다. $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ 을 \mathbf{X} 의 관측자료라 하자. 이 때 무변화 모형 M_0 에 대해서 변화점 r_j 를 가지는 각 변화모형 M_j 의 베이즈인자는 다음과 같이 정의된다.

$$B_{j0}(\mathbf{x}|r_j) = \frac{m_j(r_j, \mathbf{x})}{m_0(\mathbf{x})}, \quad r_j \in I, \quad j = 1, 2, 3,$$

여기서 $m_0(\mathbf{x}) = \int_{\Theta_0} \pi_0(\theta_0) l_0(\theta_0|\mathbf{x}) d\theta_0$ 와 $m_j(r_j, \mathbf{x}) = \sum \int_{\Theta_j} \pi_j(\theta_j, r_j) l_j(\theta_j, r_j|\mathbf{x}) d\theta_j$ 는 무변화 모형 M_0 와 변화 모형 M_j, r_j 에 대한 각각의 주변확률밀도함수(marginal probability density) 또는 예측확률밀도함수를 말한다. 또한, Θ_0 와 Θ_j 는 각 해당 모형에 포함되어 있는 모수 θ_0, θ_j 의 영역을 의미하고, $\pi_0(\theta_0)$ 와 $\pi_j(\theta_j, r_j)$ 는 각 해당 모형에 대한 모수의 사전 분포(prior distribution), 그리고 $l_0(\theta_0|\mathbf{x})$ 와 $l_j(\theta_j, r_j|\mathbf{x})$ 는 각각의 우도함수를 나타낸다. 이때 $\theta_0 = (\mu_1, \sigma_1^2)$, $\theta_1 = (\mu_1, \mu_2, \sigma_1^2)$, $\theta_2 = (\mu_1, \sigma_1^2, \sigma_2^2)$, 그리고 $\theta_3 = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ 과 같다.

모수에 대한 사전분포로는 알려져 있는 정보가 있다면 그러한 정보 하에 주관적인 사전분포를 가정할 수 있으나, 알려진 사전정보가 없거나 객관적인 사전분포를 가정하고자 한다면 일반적으로 무정보(noninformative) 사전분포를 가정한다. 그런데, 무정보 사전분포는 부적절(improper) 확률분포인 경우가 많다. 부적절 무정보 사전분포를 사용하는 경우에는 알려지지 않은 상수로 인해 베이즈인자 계산 시에 문제가 발생한다. 이러한 문제는 몇 가지 해결방법 중에서 내재적 베이즈인자(IBF)가 비교적 간편하게 사용될 수 있다. 이는 Berger와 Pericchi(1996, 1998)에 의한 방법으로서, 디폴트(default) 베이즈인자 또는 자동적(automatic) 베이즈인자로 불린다.

IBF는 비교하는 두 모형에 포함된 모수들의 차원이 다른 경우, 각 모형의 주변확률밀도 함수에 내포된 정확히 정의되지 않은 임의의 상수가 상쇄되게 함으로서 임의의 상수를 밝혀내지 않더라도 베이즈인자를 정확히 계산해 낼 수 있는 장점을 가지고 있다. 변화점 r_j 에서의 무변화 모형 M_0 에 대한 변화모형 M_j 의 IBF인 $B_{j0}^I(\mathbf{x}|r_j)$ 는 다음과 같이 정의된다.

$$B_{j0}^I(\mathbf{x}|r_j) = B_{j0}^N(\mathbf{x}|r_j) \cdot B_{0j}^N(\mathbf{x}(\mathbf{l})|r_j) = \frac{m_j^N(r_j, \mathbf{x})}{m_0^N(\mathbf{x})} \cdot \frac{m_0^N(\mathbf{x}(\mathbf{l}))}{m_j^N(r_j, \mathbf{x}(\mathbf{l}))}, \quad j = 1, 2, 3,$$

여기서 윗첨자 N 은 무정보 부적절 사전분포를 사용함을 나타낸다. 윗식처럼 최소 트레이닝 표본(minimal training sample)이라고 하는 $\mathbf{x}(\mathbf{l})$ 을 이용하여 임의의 상수를 분자·분모에 대칭적으로 위치시킴으로서 상쇄시킬 수 있다. 여기서 최소 트레이닝 표본이란 가정된 모든 모형의 주변확률밀도함수가 유한해야 함을 만족하는 최소 크기의 표본을 말한다.

한편, 이러한 IBF는 하나의 최소 트레이닝 표본을 이용하므로 추출된 표본에 의존하는 결함을 가지고 있으므로, 이를 보완하기 위해 여러 개의 최소 트레이닝 표본을 사용하여 구한 IBF들의 산술평균(AIBF), 기하평균(GIBF) 또는 중앙값(MIBF) 등을 이용한다.

$$\begin{aligned} B_{j0}^{AIBF}(r_j) &= B_{j0}^N(\mathbf{x}|r_j) \cdot \frac{1}{L} \sum_{l=1}^L B_{0j}^N(\mathbf{x}(\mathbf{l})|r_j), \\ B_{j0}^{GIBF}(r_j) &= B_{j0}^N(\mathbf{x}|r_j) \cdot \left\{ \prod_{l=1}^L B_{0j}^N(\mathbf{x}(\mathbf{l})|r_j) \right\}^{1/L}, \\ B_{j0}^{MIBF}(r_j) &= B_{j0}^N(\mathbf{x}|r_j) \cdot \text{Median}_{1 < l \leq L} \left\{ B_{0j}^N(\mathbf{x}(\mathbf{l})|r_j) \right\}, \end{aligned}$$

여기서 l 은 추출된 최소 트레이닝 표본의 순번을 의미하고, L 은 가능한 모든 최소 트레이닝 표본의 수를 의미하며, $\mathbf{x}(\mathbf{l})$ 은 l 번째로 뽑힌 최소 트레이닝 표본을 말한다.

이제 위에서 정의된 베이즈인자들을 이용하여 변화점 r_j 를 가지는 변화모형 M_{j,r_j} 의 사후확률(posterior probability)을 다음과 같이 정의할 수 있으며, 이는 비교하고자 하는 모형을 여러 개 가정하는 경우에 유용하게 사용된다.

$$P(M_{j,r_j}|\mathbf{x}) = \left\{ \sum_{i=0}^3 \sum_{r_i \in I} \left(\frac{p_i}{p_j} \cdot \frac{B_{j0}^*(r_i)}{B_{j0}^*(r_j)} \right) \right\}^{-1}, \quad j = 1, 2, 3,$$

여기서 p_0 와 p_j ($j = 1, 2, 3$)는 각 모형에 대한 사전확률(prior probability)을 의미하는 것으로서 사전에 정보가 없는 경우에는 동일한 값이 부여되어 상쇄되는 경우가 많다. 또한 * 표시는 AIBF, GIBF, MIBF 등을 대신한다. 최종적으로 각 변화모형 M_j 의 사후확률은 다음의 식에 의해 계산된다.

$$P(M_j|\mathbf{x}) = \sum_{r_j \in I} P(M_{j,r_j}|\mathbf{x}), \quad j = 1, 2, 3.$$

한편 변화점 r_j 의 사후확률밀도함수는 다음과 같이 계산된다.

$$f_j(r_j|\mathbf{x}) = \frac{m_j^N(r_j, \mathbf{x})}{m_j^N(\mathbf{x})}, \quad j = 1, 2, 3,$$

여기서 $m_j^N(\mathbf{x}) = \sum_{r_j \in I} m_j^N(r_j, \mathbf{x})$ 를 의미한다. 변화점의 사후확률을 통해 변화점으로서 가능성이 높은 위치를 짐작할 수 있다. 이로부터 확률이 가장 높은 시점을 변화점인 것으로 잠정적인 결정을 내릴 수도 있겠으나, 더 정확한 추정을 위해 다음 장에서 다루는 모수 추정 단계에 필요한 사전 정보로서 유용하게 사용될 수 있다.

2.2. 모형에 적합된 계산식

각 모형에 대한 모수의 사전분포로서 정규분포의 평균과 분산 모수에 대해서는 다음과 같은 무정보 부적절 사전분포

$$\begin{aligned} \pi_0^N(\mu_1, \sigma_1^2) &= c_0 (\sigma_1^2)^{-1}, & \pi_1^N(\mu_1, \mu_2, \sigma_1^2) &= c_1 (\sigma_1^2)^{-1}, \\ \pi_2^N(\mu_1, \sigma_1^2, \sigma_2^2) &= c_2 (\sigma_1^2)^{-1} (\sigma_2^2)^{-1}, & \pi_3^N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) &= c_3 (\sigma_1^2)^{-1} (\sigma_2^2)^{-1} \end{aligned}$$

를 가정한다. 여기서 $c_j (j = 0, 1, 2, 3)$ 는 알려지지 않은 상수이다. 또한 변화점의 사전분포 $\pi_j(r_j) (j = 1, 2, 3)$ 는 균일분포를 가정한다.

각 모형은 정규분포를 따르므로 우도함수는 각각 다음과 같이 계산된다.

$$\begin{aligned} l_0(\mu_1, \sigma_1^2|\mathbf{x}) &= (2\pi\sigma_1^2)^{-n/2} \cdot \exp\left(-\frac{1}{2\sigma_1^2} \sum_{i=1}^n (x_i - \mu_1)^2\right), \\ l_1(\mu_1, \mu_2, \sigma_1^2, r_1|\mathbf{x}) &= (2\pi\sigma_1^2)^{-n/2} \\ &\quad \cdot \exp\left(-\frac{1}{2\sigma_1^2} \left\{ \sum_{i=1}^{r_1} (x_i - \mu_1)^2 + \sum_{i=r_1+1}^n (x_i - \mu_2)^2 \right\}\right), \\ l_2(\mu_1, \sigma_1^2, \sigma_2^2, r_2|\mathbf{x}) &= (2\pi)^{-n/2} (\sigma_1^2)^{-r_2/2} (\sigma_2^2)^{-(n-r_2)/2} \\ &\quad \cdot \exp\left(-\frac{1}{2\sigma_1^2} \sum_{i=1}^{r_2} (x_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=r_2+1}^n (x_i - \mu_1)^2\right), \\ l_3(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, r_3|\mathbf{x}) &= (2\pi)^{-n/2} (\sigma_1^2)^{-r_3/2} (\sigma_2^2)^{-(n-r_3)/2} \\ &\quad \cdot \exp\left(-\frac{1}{2\sigma_1^2} \sum_{i=1}^{r_3} (x_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=r_3+1}^n (x_i - \mu_2)^2\right). \end{aligned}$$

각 모형에 대한 사전분포함수와 우도함수를 이용하여 다음과 같이 주변확률밀도함수를 계산할 수 있다.

$$\begin{aligned} m_0^N(\mathbf{x}) &= \int_0^\infty \int_{-\infty}^\infty \pi_0^N(\mu_1, \sigma_1^2) l_0(\mu_1, \sigma_1^2 | \mathbf{x}) d\mu_1 d\sigma_1^2 \\ &= c_0 n^{-1/2} \pi^{-(n-1)/2} \Gamma\left(\frac{n-1}{2}\right) \left\{ \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2 / n \right\}, \end{aligned}$$

$$\begin{aligned} m_1^N(r_1, \mathbf{x}) &= \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \pi_1^N(\mu_1, \mu_2, \sigma_1^2) \pi_1(r_1) l_1(\mu_1, \mu_2, \sigma_1^2, r_1 | \mathbf{x}) d\mu_1 d\mu_2 d\sigma_1^2 \\ &= c_1 \pi_1(r_1) \pi^{-(n-2)/2} r_1^{-1/2} (n-r_1)^{-1/2} \Gamma\left(\frac{n-1}{2}\right) \\ &\quad \cdot \left\{ \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^{r_1} x_i\right)^2 / r_1 - \left(\sum_{i=r_1+1}^n x_i\right)^2 / (n-r_1) \right\}, \end{aligned}$$

$$\begin{aligned} m_2^N(r_2, \mathbf{x}) &= \int_{-\infty}^\infty \int_0^\infty \int_0^\infty \pi_2^N(\mu_1, \sigma_1^2, \sigma_2^2) \pi_2(r_2) l_2(\mu_1, \sigma_1^2, \sigma_2^2, r_2 | \mathbf{x}) d\sigma_1^2 d\sigma_2^2 d\mu_1 \\ &= c_2 \pi_2(r_2) \pi^{-n/2} \Gamma\left(\frac{r_2}{2}\right) \Gamma\left(\frac{n-r_2}{2}\right) W(r_2, \mathbf{x}) \\ &\quad \cdot \{(r_2-1)s_1^2\}^{-r_2/2} \{(n-r_2-1)s_2^2\}^{-(n-r_2)/2}, \end{aligned}$$

$$\begin{aligned} m_3^N(r_3, \mathbf{x}) &= \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \pi_3^N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \pi_3(r_3) l_3(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, r_3 | \mathbf{x}) d\mu_1 d\mu_2 d\sigma_1^2 d\sigma_2^2 \\ &= c_3 \pi_3(r_3) \pi^{-(n-2)/2} r_3^{-1/2} (n-r_3)^{-1/2} \Gamma\left(\frac{r_3-1}{2}\right) \Gamma\left(\frac{n-r_3-1}{2}\right) \\ &\quad \cdot \{(r_3-1)s_3^2\}^{-r_3/2} \{(n-r_3-1)s_4^2\}^{-(n-r_3)/2}, \end{aligned}$$

여기서 $W(r_2, \mathbf{x}) = \int_{-\infty}^\infty \{1 + Q_1(\mu_1)\}^{-n_1/2} \{1 + Q_2(\mu_1)\}^{-n_2/2} d\mu_1$, $Q_k(\mu_1) = C(\mathbf{x}_k)(\mu_1 - \bar{x}_k)^2$ ($k = 1, 2$), $C(\mathbf{x}_k) = n_k / \{(n_k - 1)s_k^2\}$, $n_1 = r_2, n_2 = n - r_2$ 이다. 또한, (\bar{x}_1, s_1^2) 은 관측치열 $\mathbf{x}_1 = \{x_i, i = 1, 2, \dots, r_2\}$ 에서, (\bar{x}_2, s_2^2) 는 관측치열 $\mathbf{x}_2 = \{x_i, i = r_2 + 1, r_2 + 2, \dots, n\}$ 에서 구한 각각의 평균과 분산((뚝수-1)로 나눔)이고, s_3^2 와 s_4^2 는 관측치열 $\mathbf{x}_3 = \{x_i, i = 1, 2, \dots, r_3\}$ 와 $\mathbf{x}_4 = \{x_i, i = r_3 + 1, r_3 + 2, \dots, n\}$ 에서 구한 각각의 분산((뚝수-1)로 나눔)이다.

이제, IBF를 계산하기 위해 필요한 최소 트레이닝 표본 $\mathbf{x}(l)$ 을 추출하기 위해 그 크기와 추출 방법을 결정한다. 먼저 최소 트레이닝 표본의 크기를 찾는다. 정의상으로는 가정된 모든 모형에 대해 각각 주변밀도함수가 유한하게 되는 최소의 표본 크기를 찾아야 하나, 특별한 경우로서 고려되는 모형들 가운데 자신 이외의 다른 모든 모형을 표현할 수 있는 포괄적 모형(encompassing model)이 존재하는 경우에는 포괄적 모형 속에 고려되는 모든 모수가 포함되어 있으므로 이 모형에 대한 최소 트레이닝 표본의 크기를 찾으면 충분하다. 본 논문의 경우에도 모형 M_{3, r_3} 가 다른 모형들을 표현할 수 있는 포괄적 모형이므로 M_{3, r_3} 모형에 대해서만 고려한다. 전체 표본은 변화점(r_3)를 중심으로 두 개의 집단으로 분리되는데, 각 집단의 최소 표본크기가 $r_3 \geq 2$ 와 $n - r_3 \geq 2$ 를 만족해야 모든 모수들이 추정 가능하므로, 결국 필요한 최소 트레이닝 표본의 크기는 4라고 할 수 있다.

다음으로 크기 4인 최소 트레이닝 표본의 추출방법을 정한다. 본 논문에서는 변화점(r_j)까지의 시점들($1 \sim r_j$)과 이후 시점들($r_j + 1 \sim n$)에서 각각 두 시점씩을 임의로 선택하여 그 시점에 해당하는 자료를 추출하는 단순임의추출 방법을 적용한다.

표본 크기가 n 일 때 모형 M_{j, r_j} ($j = 1, 2, 3$)에 있어서 변화점(r_j)으로서 가능한 시점들은 위에서 살펴보았듯이 실제적으로 $r_j = 2, 3, \dots, n - 2$ 이므로, 변화점에 대한 사전분포는 다

음과 같이 정의할 수 있다.

$$\pi_j(r_j) = \frac{1}{n-3}, \quad r_j = 2, 3, \dots, n-2, \quad j = 1, 2, 3.$$

위와 같은 방법으로 추출된 최소 트레이닝 표본 $\mathbf{x}(\mathbf{l})$ 의 주변확률밀도함수 $m_0^N(\mathbf{x}(\mathbf{l}))$ 과 $m_j^N(r_j, \mathbf{x}(\mathbf{l}))$ 는 $m_0^N(\mathbf{x})$ 와 $m_j^N(r_j, \mathbf{x})$ 의 계산식에, \mathbf{x} 대신 $\mathbf{x}(\mathbf{l})$, n 대신 4, r_j 대신 2를 대입하여 계산한다.

3. 모수추정

이제 적합한 최적의 변화모형이 선택되었다면 그 모형 내의 모수인 평균, 분산 및 변화점 등을 추정함으로써 최적 모형의 모형식을 추정하고자 한다. 모수추정은 선택된 모형이 어떤 형태의 모형인가에 따라 추정해야 할 모수가 달라진다. 즉, 평균변화 모형 M_{1,r_1} 은 $\mu_1, \mu_2, \sigma_1^2, r_1$ 을, 분산변화 모형 M_{2,r_2} 는 $\mu_1, \sigma_1^2, \sigma_2^2, r_2$ 를, 평균 및 분산변화 모형 M_{3,r_3} 는 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, r_3$ 를 추정해야 한다.

본 논문에서는 모수추정을 위해 깃스 샘플러와 메트로폴리스-해스팅스(M-H) 알고리즘을 이용한다. 먼저 모수들의 결합사후확률분포로부터 완전조건부 사후확률분포를 구하고, 각 조건부 사후확률분포들을 이용하여 M-H 알고리즘을 포함하는 깃스 샘플러를 구성한다.

3.1. 완전조건부 사후분포

각 모형 $M_j, r_j (j = 1, 2, 3)$ 에 대한 결합사후확률분포는 사전분포와 우도함수를 결합시킴으로서 각각 다음과 같이 얻어진다.

$$P_1(\mu_1, \mu_2, \sigma_1^2, r_1 | \mathbf{x}) \propto (\sigma_1^2)^{-(n+2)/2} \cdot \exp\left(-\frac{1}{2\sigma_1^2} \left\{ \sum_{i=1}^{r_1} (x_i - \mu_1)^2 + \sum_{i=r_1+1}^n (x_i - \mu_2)^2 \right\}\right),$$

$$P_2(\mu_1, \sigma_1^2, \sigma_2^2, r_2 | \mathbf{x}) \propto (\sigma_1^2)^{-(r_2+2)/2} (\sigma_2^2)^{-(n-r_2+2)/2} \cdot \exp\left(-\frac{1}{2\sigma_1^2} \sum_{i=1}^{r_2} (x_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=r_2+1}^n (x_i - \mu_1)^2\right),$$

$$P_3(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, r_3 | \mathbf{x}) \propto (\sigma_1^2)^{-(r_3+2)/2} (\sigma_2^2)^{-(n-r_3+2)/2} \cdot \exp\left(-\frac{1}{2\sigma_1^2} \sum_{i=1}^{r_3} (x_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=r_3+1}^n (x_i - \mu_2)^2\right).$$

위와 같은 각 모형 M_j, r_j 의 결합사후확률분포로부터 유도되는 모형 내의 각 모수에 대한 완전조건부 사후확률분포는 다음과 같다.

모형 M_{1,r_1} 에 대해,

$$\begin{aligned} P_1(\mu_1|\sigma_1^2, r_1, \mathbf{x}) &\sim N\left(\frac{1}{r_1} \sum_{i=1}^{r_1} x_i, \frac{\sigma_1^2}{r_1}\right), \\ P_1(\mu_2|\sigma_1^2, r_1, \mathbf{x}) &\sim N\left(\frac{1}{n-r_1} \sum_{i=r_1+1}^n x_i, \frac{\sigma_1^2}{n-r_1}\right), \\ P_1(\sigma_1^2|\mu_1, \mu_2, r_1, \mathbf{x}) &\sim IG\left(\frac{n}{2}, 2\left\{\sum_{i=1}^{r_1} (x_i - \mu_1)^2 + \sum_{i=r_1+1}^n (x_i - \mu_2)^2\right\}^{-1}\right), \\ P_1(r_1|\mu_1, \mu_2, \sigma_1^2, \mathbf{x}) &\propto \exp\left(-\frac{1}{2\sigma_1^2} \left\{\sum_{i=1}^{r_1} (x_i - \mu_1)^2 + \sum_{i=r_1+1}^n (x_i - \mu_2)^2\right\}^{-1}\right). \end{aligned}$$

모형 M_{2,r_2} 에 대해,

$$\begin{aligned} P_2(\mu_1|\sigma_1^2, \sigma_2^2, r_2, \mathbf{x}) &\sim N\left(\left(\frac{r_2}{\sigma_1^2} + \frac{n-r_2}{\sigma_2^2}\right)^{-1} \left(\frac{1}{\sigma_1^2} \sum_{i=1}^{r_2} x_i + \frac{1}{\sigma_2^2} \sum_{i=r_2+1}^n x_i\right), \right. \\ &\quad \left. \cdot \left(\frac{r_2}{\sigma_1^2} + \frac{n-r_2}{\sigma_2^2}\right)^{-1}\right), \\ P_2(\sigma_1^2|\mu_1, r_2, \mathbf{x}) &\sim IG\left(\frac{r_2}{2}, 2\left\{\sum_{i=1}^{r_2} (x_i - \mu_1)^2\right\}^{-1}\right), \\ P_2(\sigma_2^2|\mu_1, r_2, \mathbf{x}) &\sim IG\left(\frac{n-r_2}{2}, 2\left\{\sum_{i=r_2+1}^n (x_i - \mu_1)^2\right\}^{-1}\right), \\ P_2(r_2|\mu_1, \sigma_1^2, \sigma_2^2, \mathbf{x}) &\propto \exp\left(-\frac{1}{2\sigma_1^2} \sum_{i=1}^{r_2} (x_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=r_2+1}^n (x_i - \mu_1)^2\right). \end{aligned}$$

모형 M_{3,r_3} 에 대해,

$$\begin{aligned} P_3(\mu_1|\sigma_1^2, r_3, \mathbf{x}) &\sim N\left(\frac{1}{r_3} \sum_{i=1}^{r_3} x_i, \frac{\sigma_1^2}{r_3}\right), \\ P_3(\mu_2|\sigma_2^2, r_3, \mathbf{x}) &\sim N\left(\frac{1}{n-r_3} \sum_{i=r_3+1}^n x_i, \frac{\sigma_2^2}{n-r_3}\right), \\ P_3(\sigma_1^2|\mu_1, r_3, \mathbf{x}) &\sim IG\left(\frac{r_3}{2}, 2\left\{\sum_{i=1}^{r_3} (x_i - \mu_1)^2\right\}^{-1}\right), \\ P_3(\sigma_2^2|\mu_2, r_3, \mathbf{x}) &\sim IG\left(\frac{n-r_3}{2}, 2\left\{\sum_{i=r_3+1}^n (x_i - \mu_2)^2\right\}^{-1}\right), \\ P_3(r_3|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \mathbf{x}) &\propto (\sigma_1^2)^{-r_3/2} (\sigma_2^2)^{-(n-r_3)/2} \\ &\quad \cdot \exp\left(-\frac{1}{2\sigma_1^2} \sum_{i=1}^{r_3} (x_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=r_3+1}^n (x_i - \mu_2)^2\right). \end{aligned}$$

3.2. 깃스 샘플러의 구성

3.1절에서 구한 완전조건부 사후분포를 이용하여 각 모형에 대해 M-H 알고리즘을 포함하는 깃스 샘플러를 구성하고자 한다. 이는 가장 많은 모수를 포함하고 있는 모형 M_{3,r_3} 에 대해서만 적용 방법을 제시하면 다음과 같다.

[단계 1 : 초기화단계] 평균과 변화점 모수에 대한 초기값 $\mu_1^{(0)}, \mu_2^{(0)}, r_3^{(0)}$ 를 정한다.

[단계 2 : 반복단계] 다음 단계를 $R(i=1, 2, \dots, R)$ 회 반복수행한다.

[단계 2-1] 각 조건부 사후확률분포로부터 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ 의 난수값을 발생시킨다.

$$\begin{aligned}\sigma_1^{2(i)} &\sim P_3(\sigma_1^2 | \mu_1^{(i-1)}, r_3^{(i-1)}, \mathbf{x}), \\ \sigma_2^{2(i)} &\sim P_3(\sigma_2^2 | \mu_2^{(i-1)}, r_3^{(i-1)}, \mathbf{x}), \\ \mu_1^{(i)} &\sim P_3(\mu_1 | \sigma_1^{2(i)}, r_3^{(i-1)}, \mathbf{x}), \\ \mu_2^{(i)} &\sim P_3(\mu_2 | \sigma_2^{2(i)}, r_3^{(i-1)}, \mathbf{x}).\end{aligned}$$

[단계 2-2] M-H 알고리즘을 이용하여 r_3 의 난수값을 발생시킨다.

$$h(r_3) = P_3(r_3 | \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \mathbf{x}) \text{라고 할 때,}$$

[단계 2-2-1 : 초기화단계] 초기값 $r_{3(0)}$ 또는 $r_3^{(i-1)}$ 값을 사용하고, r_3 의 적절한 조건부 전이확률함수 $g(\cdot)$ 를 설정한다.

[단계 2-2-2 : 반복단계] 다음의 과정을 $M(j = 1, 2, \dots, M)$ 번 반복수행한다.

1. $g(r_3 | r_{3(j-1)})$ 로부터 r_3^* 를 발생시킨다.
2. 전이확률 α 를 다음과 같이 구한다.

$$\alpha = \min \left\{ 1, \frac{h(r_3^*)g(r_{3(j-1)} | r_3^*)}{h(r_{3(j-1)})g(r_3^* | r_{3(j-1)})} \right\}.$$

3. $U(0, 1)$ 분포로부터 난수 u 를 발생시킨다.
4. 다음과 같이 $r_{3(j)}$ 를 결정한다.

$$r_{3(j)} = \begin{cases} r_3^*, & u \leq \alpha \text{인 경우,} \\ r_{3(j-1)}, & u > \alpha \text{인 경우.} \end{cases}$$

[단계 2-2-3] 최대 M 번 반복하여 얻은 $r_{3(M)}$ 값을 $r_3^{(i)}$ 값으로 대체시킨다.

4. 모의실험

지금까지 논의된 모형선택 및 모수추정을 위한 베이지안 절차를 이용한 모의실험을 통해 이론의 적합성을 검토하고자 한다. 모든 계산은 MATLAB(The MathWorks. Inc., 2002)를 사용하였다.

4.1. 모형선택 단계

변화점을 기준으로하여 두 집단으로 분리된 자료의 평균과 분산으로서 먼저, 변화점까지의 자료에는 $(\mu_1, \sigma_1^2) = (0, 1)$ 로 설정하고, 변화점 이후의 자료에는 $\mu_2 = 0, 1, 1.5, 2, 2.5$ 와 $\sigma_2^2 = 1, 2, 4, 6, 8$ 의 값으로 구성되는 모든 조합의 (μ_2, σ_2^2) 로 설정한다. 이 때 $(\mu_2, \sigma_2^2) = (0, 1)$ 인 경우는 무변화 모형 M_0 가, $\mu_2 = 1, 1.5, 2, 2.5$ 이고 $\sigma_2^2 = 1$ 인 경우는 평균변화 모형 M_{1,r_1} 이, $\mu_2 = 0$ 이고 $\sigma_2^2 = 2, 4, 6, 8$ 인 경우는 분산변화 모형 M_{2,r_2} 가, $\mu_2 = 1, 1.5, 2, 2.5$ 와 $\sigma_2^2 = 2, 4, 6, 8$ 인 경우에는 평균 및 분산변화 모형 M_{3,r_3} 가 참인 모형이다.

전체표본 크기(n)는 50, 각 변화모형 M_{j,r_j} 의 변화점 r_j 는 25로 하여 위에 제시한 총 25가지 모형에 대해 각각 100번씩 반복 수행한다. 최소 트레이닝 표본은 각각의 전체표본에 대해 변화점을 중심으로 양측에서 2개씩을 무작위 추출하며, 이를 30회 반복하여 얻는다.

모형 M_{2,r_2} 에 대한 주변확률밀도함수인 $m_2^N(r_2, \mathbf{x})$ 와 $m_2^N(r_2, \mathbf{x}(l))$ 의 계산은 평균 모수 μ_1 에 대한 해석학적 적분이 어려우므로 중요표본추출(importance sampling) 기법을 적용한 몬테카를로 적분법을 이용하였다. $m_2^N(r_2, \mathbf{x})$ 의 계산식에서 $W(r_2, \mathbf{x})$ 를 다음과 같이 재정의할 수 있다.

$$W(r_2, \mathbf{x}) = \int_{-\infty}^{\infty} g(\mu_1) d\mu_1 = \int_{-\infty}^{\infty} \frac{g(\mu_1)}{f(\mu_1)} f(\mu_1) d\mu_1 = E_{f(\mu_1)} \left[\frac{g(\mu_1)}{f(\mu_1)} \right],$$

여기서 $f(\mu_1)$ 은 μ_1 의 표본추출을 위한 확률밀도함수로서 중요함수(importance function)라고 한다. $f(\cdot) \propto |g(\cdot)|$ 이므로 $|g(\cdot)|$ 와 유사한 확률밀도함수를 $f(\mu_1)$ 으로 선택하는 것이 추정치의 정확도가 높다. 따라서 $g(\mu_1)$ 을 다음과 같이 다시 표현한다.

$$g(\mu_1) = \prod_{k=1}^2 (1 + Q_k(\mu_1))^{-n_k/2} = \prod_{k=1}^2 \exp \left\{ -\frac{n_k}{2} \ln(1 + Q_k(\mu_1)) \right\} \\ \propto \exp \left\{ -\frac{1}{2} (n_1 Q_1(\mu_1) + n_2 Q_2(\mu_1)) \right\},$$

여기서 둘째 줄의 비례식은 $\ln(1 + Q_k(\mu_1))$ 의 매크로린 급수(Maclaurin's series)에서 첫 번째 항만을 이용한 결과이다. 이로부터 중요함수로는 다음의 정규분포를 따르는 정규확률밀도함수를 사용한다.

$$N \left(\frac{u(\mathbf{x}_1)\bar{x}_1 + u(\mathbf{x}_2)\bar{x}_2}{u(\mathbf{x}_1) + u(\mathbf{x}_2)}, \frac{1}{u(\mathbf{x}_1) + u(\mathbf{x}_2)} \right),$$

여기서 $u(\mathbf{x}_1) = r_2 C(\mathbf{x}_1)$, $u(\mathbf{x}_2) = (n - r_2) \cdot C(\mathbf{x}_2)$ 이다. 따라서, $W(r_2, \mathbf{x})$ 에 대한 몬테카를로 추정치는 중요함수 $f(\mu_1)$ 으로부터 추출된 중요표본(importance sample)인 $\{\mu_{11}, \mu_{12}, \dots, \mu_{1N}\}$ 을 이용하여 $\hat{W}(r_2, \mathbf{x}) = \{ \sum_{j=1}^N g(\mu_{1j}) / f(\mu_{1j}) \} / N$ 와 같이 구하고, 이 때 사용된 중요표본수 N 은 안정성을 검토한 후 50개로 하였다.

본 모의실험에서 전체표본 자료의 생성과 몬테카를로 적분을 위한 중요표본의 추출은 *MATLAB*에서 제공되는 *NORMRND* 함수를 이용하고, 최소 트레이닝 표본 추출은 *UNIDRND* 함수를 이용한다. 표 4.1은 AIBF를 이용한 모형선택의 결과를 제시한다. 첫째, 평균 혹은 분산의 변화 정도가 커질수록, 즉 두 모형이 달라질수록 참 모형을 선택하는 능력이 좋아지는 경향을 보인다. 한편, 참 모형의 선택율이 낮은 모형들은 변화량이 비교적 적은 경우로서, M_3 모형에서 평균 변화량 $\mu_2 = 1$ 일 때와 분산 변화량 $\sigma_2^2 = 2$ 일 때이다. 둘째, 모형에 대한 사후확률은 분산변화 정도에 비해 평균변화 정도에 대해서 훨씬 더 민감하게 작용하는 특성을 보인다. 즉, 평균만 변화하는 경우에 변화량이 많을수록 평균변화 모형 M_1 의 확률이 거의 1에 가까워지고 분산변화 모형 M_2 의 확률은 거의 0에 가까워진다. 또한, 평균과 분산이 모두 변화하는 경우에는 이와 비교할 때 분산변화 모형의 확률이 더 빠르게 0에 가까워지는 것을 볼 수 있다. 반면에 분산만 변화하는 경우에는 평균이 변화한 동일 수준과 비교했을 때, 그 변화량이 많을수록 평균변화 모형에 비해 분산변화 모형의 사후확률이 현저히 낮아진다.

지면 관계상 제시하지 못했으나 MIBF와 GIBF를 이용하여 계산된 사후확률의 결과도 AIBF와 유사한 경향을 보인다. 그림 4.1 ~ 그림 4.6은 세 가지 IBF의 참 모형 선택율을 비교한 것으로서 이 결과에 의하면, M_0 모형에서는 참 모형의 선택율이 GIBF, MIBF, AIBF 순으로 좋은 결과를 보였고, M_1 모형에서는 IBF들간에 비슷한 수준을 보였으며, M_2 모형에서는 AIBF, MIBF, GIBF 순으로 좋은 결과를 보였다. 또한, M_3 모형에서는 μ_2 가 작은 경우에는 σ_2^2 이 커질수록 GIBF, MIBF, AIBF 순으로 좋은 결과를 보였으나, μ_2 가 큰 경우에는 IBF들간에 비슷한 수준을 보였다. 그림 4.7 ~ 그림 4.9는 변화점의 최대 사후확률의 100회 반복치들을 히스토그램한 것이다. 평균 또는 분산의 변화량이 커질수록 변화량의 참 값인 25시점의 비율이 커지는 경향을 보인다. 한편, M_2 모형에서 $(\mu_2, \sigma_2^2) = (0, 2)$ 인 경우는 변화점 참값 25보다는 양 끝 극단점에서의 확률이 더 높은 것으로 나타났는데, 이는 표 4.1에 제시되었듯이 M_0 모형의 사후확률이 비슷하게 크기 때문인 것으로 생각할 수 있다.

4.2. 모수추정 단계

변화점을 기준으로 두 집단으로 분리된 자료의 평균과 분산은 이전의 모형선택 단계에서 설정한 방식과 동일하다. 무변화 모형을 제외한 24가지 가능한 모든 변화 모형에 대해 전체표본 크기(n)는 50이고, 각 모형 M_{j,r_j} ($j = 1, 2, 3$)의 변화점 r_j 는 25로 설정하였다. 깃스 샘플러는 두 세트를 병렬로 구성하고, 이에 필요한 모수의 초기값은 두 경우 모두 동일값을 주었다. 변화점에 대한 초기값 $r_{j(0)}$ 는 모형선택 단계에서 구한 최대의 변화점 사후확률을 갖는 시점으로 하고, 평균에 대한 초기값으로는 M_{1,r_1} 과 M_{3,r_3} 모형은 $\mu_1^{(0)} = (\sum_{i=1}^{r_{j(0)}} x_i)/r_{j(0)}$, $\mu_2^{(0)} = (\sum_{i=r_{j(0)+1}}^n x_i)/(n - r_{j(0)})$ 로 하고, M_{2,r_2} 모형은 $\mu_1^{(0)} = (\sum_{i=1}^n x_i)/n$ 로 하였다. 깃스 샘플러를 1,000번 반복하여 얻은 총 2,000 쌍의 결과 중 초기 절반을 제외한 1,000 쌍만을 사용하였다. 또한, 깃스 샘플러 내의 M-H 알고리즘은 50번 반복수행하고, 이때 사용된 전이확률함수로는 변화점의 사후확률을 고려하여 $r_{j(0)}$ 를 중심으로 ± 5 이내의 범위를 갖는 균일분포를 사용하였다. 표 4.2에 제시된 결과를 살펴보면, 전반적으로 각 모형의 모수들이 잘 추정된 것으로 볼 수 있다.

5. 서울 지역의 일일평균기온 자료에 적용

지구의 온난화 현상으로 기상자료에 변화점이 발생했을 것이라는 예상 하에 우리나라 서울 지역의 일일평균기온 자료를 계절별로 나누어 본 논문의 방법론을 적용하기에 적합한 한지를 살펴본 결과 4계절 자료 중에서 겨울철 자료가 적합하였다. 한편, 겨울철 자료는 원 자료인 12월, 1월, 2월의 자료를 합하여 구성한 변인이므로, '겨울철'의 종합적인 특성 외에 각 달(12월, 1월, 2월)의 구체적인 특징을 살펴보기로 하였다. 따라서 우리나라 서울 지역의 '61.1. ~ 2002.12.까지 42년 간의 일일평균기온 자료 가운데, 12월과 1월, 2월의 자료를 추출하여 겨울철의 일일평균기온에 대해 변화점이 존재하는지, 있다면 어떠한 유형으로 발생되었는지를 살펴보았다.

겨울철인 12월 ~ 2월의 일일평균기온 자료를 평균하여 41개 자료의 새로운 '겨울철' 변수를 생성하였다. 또한, 42년 간의 일일평균기온 자료 중 12월, 1월, 2월을 각각 월평균 내

어 42개 자료로 이루어진 3개의 변수를 생성하였다. 아래의 그림 5.1~ 그림 5.4는 관찰 자료를 그래프로 표현한 것이다.

분석에 사용되는 서울 지역의 겨울철 및 12월, 1월, 2월의 기온자료가 본 연구에 적용되기 위해서는 기본적 가정인 정규성 및 독립성을 만족해야 한다. 정규성에 대해서는 Shapiro-Wilks의 검정방법을 통해 '자료는 정규분포를 따른다'는 가설에 대해 검정하였고, 독립성에 대해서는 Durbin-Watson의 검정방법을 통해 '1차 자기상관관계가 존재하지 않는다'는 가설에 대해 검정하였다. 그 결과, 표 5.1에 제시되었듯이 4가지 자료 모두 기본 가정에 적합하다고 볼 수 있다. 여기서 사용된 변화점은 표 5.2에 제시한 추정된 변화점을 사용하였다. 또한, 각 자료별 변화점을 중심으로 나뉜 두 집단(전·후)의 분산에 대한 동일성을 검정($H_0 : \sigma_{before}^2 = \sigma_{after}^2$, $H_1 : \sigma_{before}^2 \neq \sigma_{after}^2$) 한 후, 그 결과에 따라 등분산 또는 이분산 하에서 평균에 대한 동일성 검정($H_0 : \mu_{before} = \mu_{after}$, $H_1 : \mu_{before} \neq \mu_{after}$)을 한 결과 표 5.1의 마지막 두개 열에 제시되어 있다. 12월은 확실히 이분산(p-value=0.0049)을 갖으며, 1월(p-value=0.0318) 및 겨울철(p-value= 0.0741)도 이분산의 징후가 큰 편이다. 이분산의 가정 하에 두 집단 간의 평균 차이 검정은 모두 매우 유의한 결과를 보였다. 이것은 평균과 분산이 같이 변화하는 것을 예상해 볼 수 있다. 2월 자료의 경우는 두 집단 간에 등분산을 갖지만 평균은 차이를 보이는 것으로 해석할 수 있다.

5.1. 모형선택 단계

표 5.2에 제시된 모형선택의 결과를 살펴보면, 겨울철과 2월은 M_1 모형에 대한 사후확률이 매우 높기 때문에 평균변화 모형이 적합한 것으로 볼 수 있고, 12월은 평균 및 분산변화 모형이 적합한 것으로 볼 수 있다. 반면, 1월은 평균변화 모형의 확률이 다른 모형에 비해 상대적으로 가장 높긴 하나, 겨울철과 2월 자료에 비해서는 낮은 편이라고 할 수 있다. 또한 1월의 무변화 모형과 평균 및 분산변화 모형의 결과와 비교하더라도 큰 차이는 보이고 있지 않다. 이러한 결과는 그림 5.1 ~ 그림 5.4 및 표 5.1의 결과와도 상통하고 있다.

각 자료에서 변화점의 사후확률이 가장 높은 시점만을 고려하면, 전반적으로 '85년부터 '88년 사이에 변화점이 발생한 것으로 추정할 수 있다. 변화점의 사후확률에 대한 더 자세한 정보는 그림 5.5 ~ 그림 5.8에 제시되었다. 여기에서도 확률이 상대적으로 높은 시점들은 최고 확률이 발생한 시점과 동일한 기간대를 형성하고 있다. 한편, 1월 자료는 그림 5.7을 살펴보면, 시작 시점 부근에서 높은 사후확률 값을 갖는 것으로 나타나 무변화 모형 M_0 에 해당될 가능성을 보여주며, 이는 표 5.2에서 M_0 의 사후확률 값이 비교적 높은 값을 갖는 것에도 상통한다. 변화점의 사후확률에 대한 이러한 정보는 모수추정 단계에서 변화점 추정을 위한 사전정보로서 유용하게 사용된다.

한편, 표 5.2에서 $P(M_2|x)$ 의 확률이 모두 0.0000의 값을 갖는데, 이와 비슷하게 설정된 모의실험에서의 결과 또한 유사한 경향을 띄고 있었다. 이러한 원인은 본 연구에서 제시된 모형선택의 기준이 분산 차에 대해서 보다는 평균 차에 대해 훨씬 더 민감하게 작용하는 특성으로부터 비롯된 결과이다. 즉 평균 및 분산 변화가 있는 경우에는 평균변화 모형(M_1)과 평균 및 분산 변화 모형(M_3)의 사후확률이 모두 크게 나타나는 것으로 고려된다.

5.2. 모수추정 단계

깁스 샘플러는 2,000번 반복하고 M-H 알고리즘은 각각 50번씩 반복하여 생성된 모수들 가운데, 전반 1,000 쌍을 제외하고 후반 1,000 쌍의 결과만 분석에 사용하였다. 이때, 변화점 모수에 대한 초기값 $r_{j(0)}$ 은 각 시점에 대한 변화점 사후확률을 계산하여 최대 확률을 갖는 시점으로 설정하고, 평균 모수에 대한 초기값은 $r_{j(0)}$ 시점을 중심으로 두 집단으로 나뉜 각 자료의 평균을 사용하였다.

추정 결과로서 평균, 표준편차, 중앙값, 그리고 95% HPD(highest probability density)를 표 5.3에 제시하였다. 95% HPD는 Chen-Shao HPD 추정 알고리즘을 적용하여 산출하였다(Ming-Hui Chen 등, 2000). 겨울철 자료에서는 26번째('86.12~'87.2) 자료에서 변화가 발생하여 변화 이전의 평균 수준 -1.85에서 변화 이후 -0.04 수준으로 상승한 것으로 나타났다. 12월 자료에서는 26번째('86.12)에서 변화가 발생하였고, 변화 이전에는 평균과 분산이 -0.62와 4.27인데 비해 변화 이후에는 0.95와 1.14로서, 변화 시점 이후에 기온의 평균 수준은 상승하고 분산은 감소한 형태를 보였다. 또한, 1월 자료에서는 변화점이 28번째('88.1)에서 발생하여 기온의 평균 수준이 -3.58에서 -1.70으로 상승한 것으로 나타났으며, 2월 자료에서는 변화점이 27번째('87.2)에서 발생하였고, 기온의 평균 수준이 -1.35에서 0.77 수준으로 상승한 것으로 나타났다.

6. 결론

본 논문에서는 정규확률 관측치열에 있어서 변화점 문제에 대해 다루었다. 관측치열 내에 하나의 변화점이 존재하는지, 존재한다면 변화점 발생 이후 자료는 어떠한 형태로 변화되었는지를 밝히기 위해 4가지 유형의 모형을 가정하였다. 즉, 변화가 없는 형태, 평균만 변화한 형태, 분산만 변화한 형태, 그리고 평균 및 분산이 모두 변화한 형태 등이다. 4가지의 가능한 유형 가운데 어느 것에 속하는지를 파악하기 위해 모수에 대한 사전분포는 객관성을 갖기 위해서 무정보 부적절 사전분포를 가정하였고, 디폴트 베이즈인자인 AIBF, MIBF, 그리고 GIBF를 이용하여 최적의 적합모형을 선택하였다. 최적의 모형을 선택한 후에는 선택된 모형에 내재되어 있는 모수를 추정하여 모형을 더 명확히 규명하고자 했으며, 이를 위해서 메트로폴리스-헤스팅스 알고리즘을 포함한 깁스 샘플러를 이용하였다.

이론에 대한 검토로서 모의실험한 결과, 모형선택 단계에서 변화량이 비교적 작은 경우를 제외하고는 참 모형의 선택율이 높게 나타났으며, 모수추정 단계에서도 대부분 적절한 추정치를 제시하였다. 또한 우리나라 서울 지역의 '61.1.~'2002.12.까지 42년간의 겨울철 일일평균기온 자료에 적용한 사례에서도 자료의 특성에 부합되는 결과를 얻었다.

이론에 대한 검토로서 모의실험 및 우리나라 서울 지역의 42년 간('61.1~'2002.12)의 겨울철(12월, 1월, 2월 포함) 일일평균기온 자료에 적용하였다. 그 결과, 본 연구에서 제시된 모형선택의 기준이 분산 차에 대해서 보다는 평균 차에 대해 더 민감하게 작용하는 특성이 있음을 알 수 있었다. 모의실험 결과로는 모형선택 단계에서 변화량이 비교적 작은 경우를 제외하고는 참 모형의 선택율이 높게 나타났으며, 모수추정 단계에서도 대부분 적절한 추정치를 제시하였다. 실제 자료의 결과에서도 자료의 특성에 부합되는 결과를 얻었다.

참고문헌

- Berger, J. O. and Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction, *Journal of the American Statistical Association*, **91**(433), 109-122.
- Berger, J. O. and Pericchi, L. R. (1998). Accurate and stable Bayesian model selection: the median intrinsic Bayes factor, *Sankhya, B*, **60**, 1-18.
- Chernoff H. and Zacks S. (1964). Estimating the current mean of a normal distribution which is subject to changes over-time, *Annals of Mathematical Statistics*, **35**, 999-1018.
- Lee A. S. F. and Heghinian S. M. (1977). A shift of the mean level in a sequence of independent normal random variables - A Bayesian approach, *Technometrics*, **19**, 503-506.
- Maguire B. A., Pearson, E. S. and Wynn A. H. A. (1952). The time intervals between industrial accidents, *Biometrika*, **38**, 168-180.
- Menzefricke, U. (1981). A Bayesian analysis of a change in the precision of a sequence of independent normal random variables at an unknown time point, *Applied Statistics*, **30**, 141-146.
- Ming-Hui Chen, Qi-Man Shao and Joseph G. Ibrahim (2000), *Monte Carlo methods in Bayesian computation*, Springer-Verlag New York, Inc.
- Page E. S. (1954). Continuous inspections schemes, *Biometrika*, **41**, 100-114.
- Page E. S. (1955). A test for a change in parametric occurring at an unknown point, *Biometrika*, **42**, 523-527.
- Page E. S. (1957). On problems on which a change in a parameter occurs at an unknown point, *Biometrika*, **44**, 248-252.
- Perreault, L., Bernier, J., Bobe, B. and Parent, E. (2000,a). Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited, *Journal of Hydrology*, **235**, 221-241.
- Perreault, L., Bernier, J., Bobe, B. and Parent, E. (2000,b). Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting, *Journal of Hydrology*, **235**, 242-263.
- Smith, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables, *Biometrika*, **62**, 407-416.
- The MATH WORKS Inc. (1998). *MATLAB/Statistics Toolbox*, Version 5.3, Natick, MA.

[2003년 8월 접수, 2004년 2월 채택]

부록

표 4.1: AIBF를 이용한 모형선택의 모의실험 결과 (Mean(S.D.), 반복 100회)

참 모형	모수값		사후 확률			
	μ_2	σ_2^2	$P(M_0 \mathbf{x})$	$P(M_1 \mathbf{x})$	$P(M_2 \mathbf{x})$	$P(M_3 \mathbf{x})$
$P(M_0)$	0.0	1	0.5210 (0.1124)	0.2287 (0.0937)	0.1589 (0.0984)	0.0913 (0.0651)
$P(M_1)$	1.0	1	0.1382 (0.1552)	0.6363 (0.2245)	0.0465 (0.0566)	0.1790 (0.1608)
	1.5		0.0180 (0.0523)	0.7893 (0.1575)	0.0069 (0.0188)	0.1858 (0.1497)
	2.0		0.0002 (0.0006)	0.8316 (0.1106)	0.0001 (0.0003)	0.1681 (0.1107)
	2.5		0.0000 (0.0000)	0.8565 (0.0652)	0.0000 (0.0000)	0.1875 (0.1530)
$P(M_2)$	0.0	2	0.3459 (0.1823)	0.1707 (0.1192)	0.3527 (0.2104)	0.1308 (0.1278)
		4	0.1149 (0.1253)	0.0683 (0.0987)	0.6883 (0.1892)	0.1285 (0.0737)
		6	0.0285 (0.0538)	0.0275 (0.0892)	0.8136 (0.1543)	0.1303 (0.1028)
		8	0.0122 (0.0386)	0.0083 (0.0260)	0.8428 (0.1607)	0.1367 (0.1470)
$P(M_3)$	1.0	2	0.1468 (0.1542)	0.4275 (0.2443)*	0.1358 (0.1484)	0.2899 (0.2275)
		4	0.0779 (0.0911)	0.1435 (0.1914)	0.4384 (0.2699)*	0.3401 (0.2528)
		6	0.0197 (0.0439)	0.0627 (0.1483)	0.5762 (0.2950)*	0.3414 (0.2469)
		8	0.0034 (0.0100)	0.0072 (0.0234)	0.6393 (0.2942)*	0.3501 (0.2880)
	1.5	2	0.0445 (0.0891)	0.5277 (0.3011)*	0.0572 (0.1261)	0.3706 (0.2831)
		4	0.0403 (0.0742)	0.1809 (0.2295)	0.2263 (0.2575)	0.5525 (0.2996)
		6	0.0080 (0.0245)	0.0459 (0.1132)	0.3044 (0.2841)	0.5768 (0.2921)
		8	0.0026 (0.0114)	0.0148 (0.0841)	0.4272 (0.3201)	0.5888 (0.3110)
	2.0	2	0.0082 (0.0277)	0.6554 (0.2554)*	0.0000 (0.0000)	0.3364 (0.2581)
		4	0.0152 (0.0624)	0.2087 (0.2547)	0.0000 (0.0000)	0.7761 (0.2739)
		6	0.0063 (0.0173)	0.0552 (0.1118)	0.0000 (0.0000)	0.9384 (0.1162)
		8	0.0037 (0.0180)	0.0145 (0.0673)	0.0000 (0.0000)	0.9818 (0.0736)
	2.5	2	0.0001 (0.0003)	0.6054 (0.2723)*	0.0000 (0.0000)	0.3945 (0.2723)
		4	0.0012 (0.0053)	0.1532 (0.2059)	0.0000 (0.0000)	0.8456 (0.2064)
		6	0.0034 (0.0227)	0.0697 (0.1570)	0.0000 (0.0000)	0.9269 (0.1623)
		8	0.0024 (0.0128)	0.0242 (0.0731)	0.0000 (0.0000)	0.9733 (0.0778)

* : 참 모형의 사후확률이 최대가 아닌 경우.

표 4.2: 깃스 샘플링에 의한 모수추정의 모의실험 결과 (Mean(S.D.), 반복 1,000회)

모형	참 값				추정 값				
	μ_1	μ_2	σ_1^2	σ_2^2	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	\hat{r}_j
M_1	0	1	1	-	0.01 (0.21)	0.92 (0.23)	1.05 (0.23)	-	26.38 (2.02)
		1.5			-0.18 (0.21)	1.52 (0.23)	1.12 (0.25)	-	25.35 (0.84)
		2			0.09 (0.21)	2.09 (0.30)	1.07 (0.25)	-	25.00 (1.51)
		2.5			0.08 (0.22)	2.40 (0.30)	1.08 (0.25)	-	24.93 (1.56)
M_2	0	-	1	2	0.14 (0.17)	-	1.12 (0.37)	2.01 (0.62)	24.20 (2.70)
				4	-0.03 (0.18)	-	0.92 (0.32)	4.28 (1.31)	23.19 (2.27)
				6	0.03 (0.20)	-	1.08 (0.37)	6.06 (1.88)	23.16 (1.74)
				8	-0.24 (0.23)	-	1.51 (0.49)	8.39 (2.55)	23.86 (1.41)
M_3	0	1	1	2	-0.28 (0.21)	1.06 (0.28)	0.92 (0.31)	2.01 (0.60)	23.70 (1.88)
				4	0.15 (0.24)	1.13 (0.42)	1.24 (0.42)	4.32 (1.39)	25.06 (2.68)
				6	0.20 (0.23)	0.83 (0.51)	1.28 (0.41)	6.60 (2.24)	24.96 (1.83)
				8	-0.09 (0.18)	1.01 (0.59)	0.84 (0.26)	8.58 (2.65)	24.44 (0.98)
	0	1.5	1	2	0.04 (0.21)	1.72 (0.32)	1.00 (0.32)	2.35 (0.76)	25.28 (1.35)
				4	-0.27 (0.18)	1.72 (0.42)	0.89 (0.29)	4.37 (1.36)	25.03 (1.15)
				6	-0.12 (0.20)	1.44 (0.50)	0.92 (0.31)	6.21 (1.88)	23.94 (1.58)
				8	-0.24 (0.21)	1.94 (0.60)	1.01 (0.34)	8.48 (2.77)	25.43 (0.89)
	0	2	1	2	-0.13 (0.22)	2.13 (0.32)	1.09 (0.36)	2.34 (0.78)	25.03 (1.49)
				4	0.19 (0.20)	1.92 (0.39)	0.92 (0.30)	4.30 (1.44)	24.18 (1.05)
				6	-0.10 (0.21)	2.13 (0.54)	0.98 (0.33)	6.99 (2.18)	24.34 (1.23)
				8	-0.06 (0.21)	1.76 (0.60)	1.01 (0.33)	8.79 (2.89)	24.51 (1.28)
	0	2.5	1	2	-0.18 (0.20)	2.62 (0.32)	1.08 (0.34)	2.39 (0.76)	25.35 (0.82)
				4	0.03 (0.26)	2.62 (0.45)	1.41 (0.45)	4.58 (1.48)	25.31 (1.75)
				6	0.02 (0.25)	2.51 (0.50)	1.45 (0.51)	6.28 (1.94)	24.97 (1.15)
				8	0.12 (0.23)	2.29 (0.59)	1.09 (0.38)	8.61 (2.75)	23.90 (1.25)

표 5.1: 적용된 자료의 기본 가정에 대한 적합성 검토

계절 / 월별	n	Shapiro-Wilks	Durbin-Watson	평균 및 분산		평균차 검정	등분산 검정	
		p-value	검정통계량			p-value	p-value	
겨울	변화점 까지	25	.670	1.779	-1.91	1.998	.0000	.0741
	변화점 이후	16	.576	2.030	-.07	.811		
12월	변화점 까지	25	.990	1.982	-.70	3.913	.0007	.0049
	변화점 이후	17	.423	2.321	.99	.942		
1월	변화점 까지	27	.611	1.827	-3.64	5.757	.0020	.0318
	변화점 이후	15	.880	2.524	-1.71	1.876		
2월	변화점 까지	28	.744	2.191	-1.36	2.595	.0001	1.000
	변화점 이후	14	.476	1.808	.86	2.587		

표 5.2: 베이즈 인자를 이용한 모형선택의 결과 (기간 : '61.1~'2002.12)

계절/월별	베이즈 인자	사 후 확 률				변화점 사후확률 최고 시점(기간)
		$P(M_0 \mathbf{x})$	$P(M_1 \mathbf{x})$	$P(M_2 \mathbf{x})$	$P(M_3 \mathbf{x})$	
겨울철	AIBF	0.0129	0.6405*	0.0000	0.3465	$r_1 = 25$ ('85.12~'86.2)
	MIBF	0.0190	0.6519*	0.0000	0.3292	
	GIBF	0.0237	0.6896*	0.0000	0.2867	
12월	AIBF	0.0853	0.1979	0.0000	0.7168*	$r_3 = 25$ ('85.12)
	MIBF	0.1195	0.2326	0.0000	0.6479*	
	GIBF	0.1492	0.2289	0.0000	0.6219*	
1월	AIBF	0.2356	0.4048*	0.0000	0.3596	$r_1 = 27$ ('87.1)
	MIBF	0.2919	0.4166*	0.0000	0.2915	
	GIBF	0.4037	0.4832*	0.0000	0.1131	
2월	AIBF	0.0654	0.8247*	0.0000	0.1100	$r_1 = 28$ ('88.2)
	MIBF	0.0805	0.8365*	0.0000	0.0830	
	GIBF	0.1169	0.8777*	0.0000	0.0054	

표 5.3: 깃스 샘플링에 의한 모수추정의 결과 (반복 1,000회)

계절/ 월별	선택 모형	모수	초기값	사 후 분 포				
				Mean	S.D.	Median	95 % HPD	
							Lower	Upper
겨울철	M_1	μ_1	-1.91	-1.85	0.25	-1.84	-2.37	-1.40
		μ_2	-0.07	-0.04	0.32	-0.05	-0.67	0.58
		σ_1^2	1.40	1.68	0.39	1.62	1.02	2.46
		r_1	25.00	25.88	0.46	26.00	25.00	27.00
12월	M_3	μ_1	-0.70	-0.62	0.41	-0.62	-1.45	0.20
		μ_2	0.99	0.95	0.26	0.96	0.44	1.49
		σ_1^2	3.91	4.27	1.21	4.07	2.28	6.83
		σ_2^2	0.94	1.14	0.52	1.02	0.50	2.13
		r_3	25.00	25.72	0.57	26.00	25.00	27.00
1월	M_1	μ_1	-3.64	-3.58	0.42	-3.58	-4.34	-2.71
		μ_2	-1.71	-1.70	0.56	-1.72	-2.91	-0.65
		σ_1^2	3.82	4.72	1.11	4.52	2.76	7.00
		r_1	27.00	27.60	0.79	28.00	27.00	30.00
2월	M_1	μ_1	-1.36	-1.35	0.32	-1.35	-1.95	-0.69
		μ_2	0.86	0.77	0.46	0.76	-0.11	1.63
		σ_1^2	2.59	2.81	0.65	2.75	1.74	4.11
		r_1	28.00	27.38	0.64	27.00	26.00	28.00

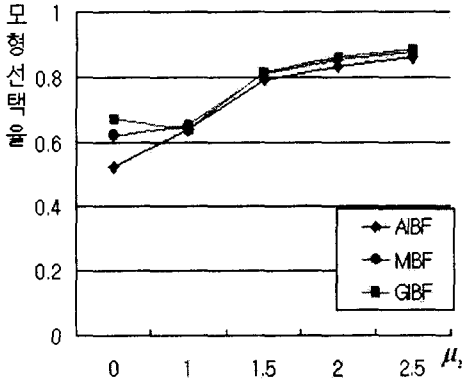


그림 4.1: M_0 또는 $M_1(\sigma_2^2 = 1)$ 모형

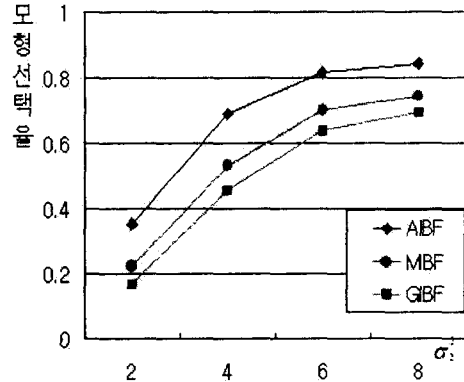


그림 4.2: M_2 모형 ($\mu_2 = 0$ 일때)

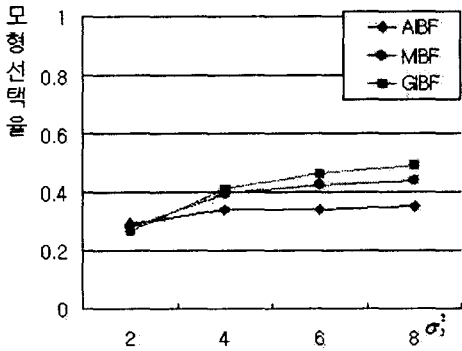


그림 4.3: M_3 모형 ($\mu_2 = 1$ 일때)

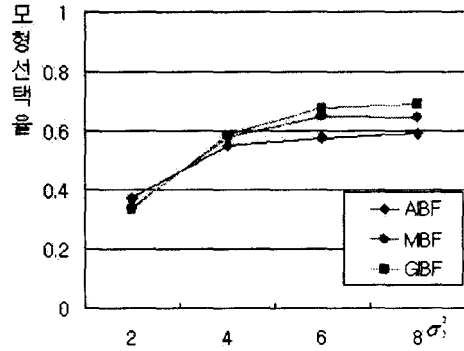


그림 4.4: M_3 모형 ($\mu_2 = 1.5$ 일때)

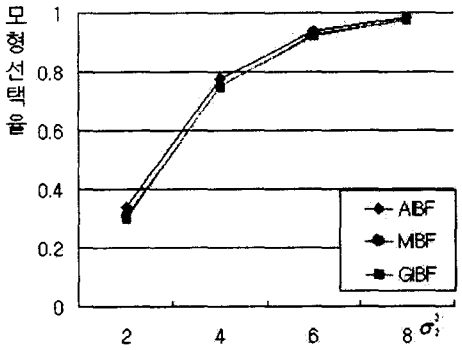


그림 4.5: M_3 모형 ($\mu_2 = 1$ 일때)

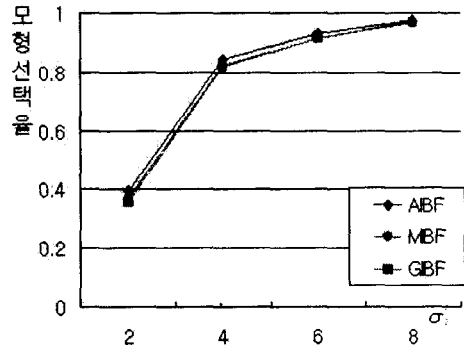


그림 4.6: M_3 모형 ($\mu_2 = 1.5$ 일때)

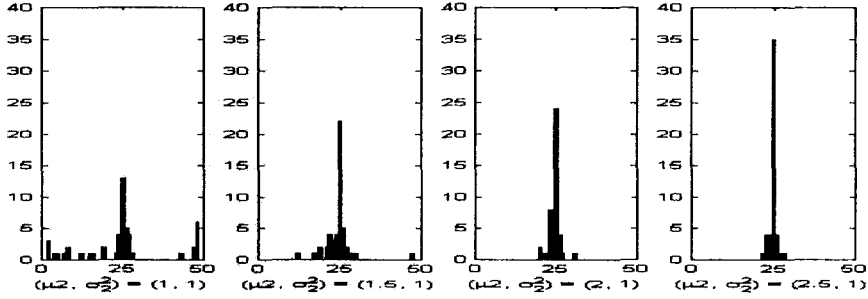


그림 4.7: M_1 모형의 변화점 최대 사후확률 (반복 100회)

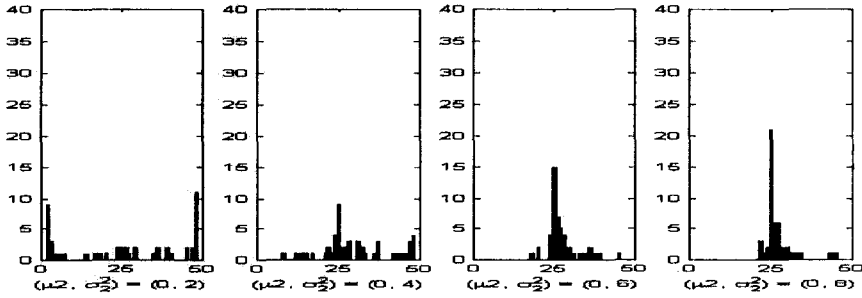


그림 4.8: M_2 모형의 변화점 최대 사후확률 (반복 100회)

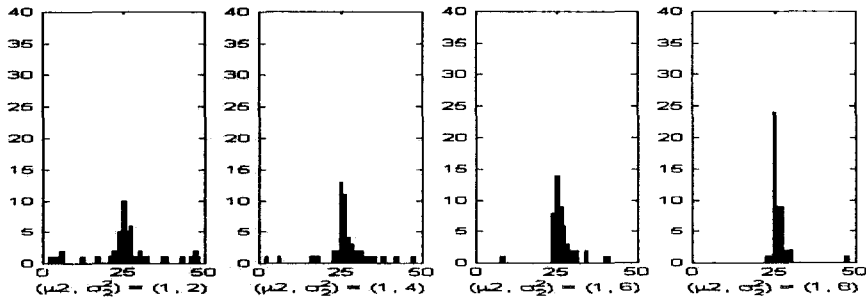


그림 4.9: M_3 모형의 변화점 최대 사후확률 (반복 100회)

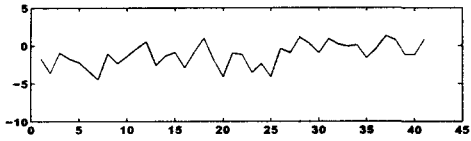


그림 5.1: 겨울철의 평균기온 자료

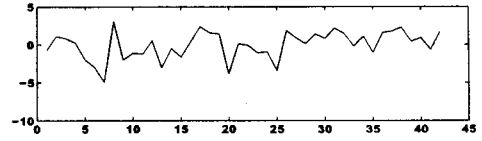
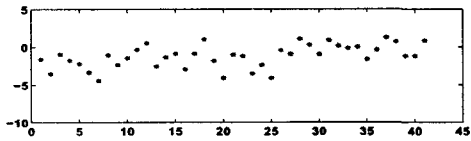


그림 5.2: 12월의 평균기온 자료

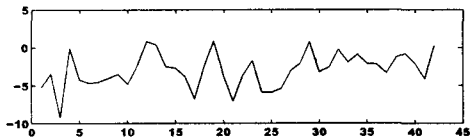
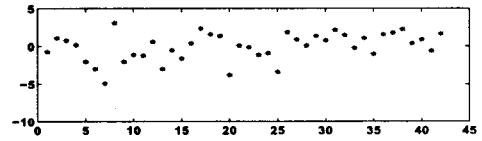


그림 5.3: 1월의 평균기온 자료

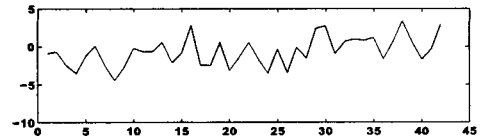
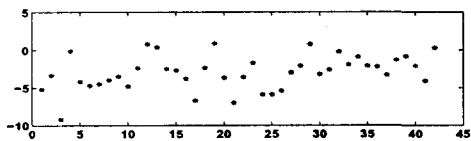


그림 5.4: 2월의 평균기온 자료

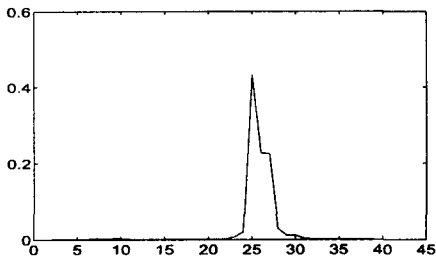
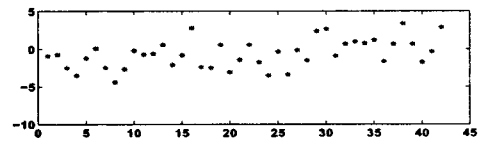


그림 5.5: 겨울철 자료의 변화점 사후확률

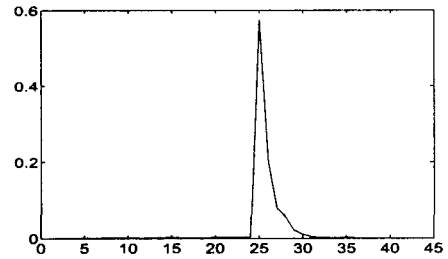


그림 5.6: 12월 자료의 변화점 사후확률

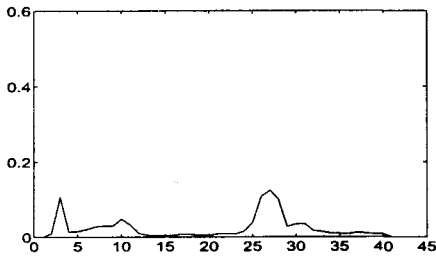


그림 5.7: 1월 자료의 변화점 사후확률

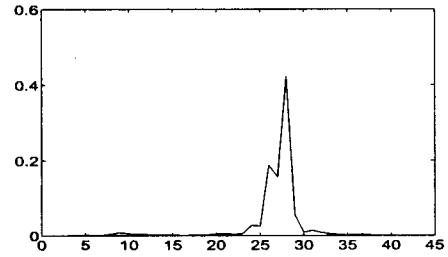


그림 5.8: 2월 자료의 변화점 사후확률

Bayesian Change Point Analysis for a Sequence of Normal Observations: Application to the Winter Average Temperature in Seoul *

Kyungsook Kim¹⁾ Young Sook Son²⁾

ABSTRACT

In this paper we consider the change point problem in a sequence of univariate normal observations. We want to know whether there is any change point or not. In case a change point exists, we will identify its change type. Namely, it can be a mean change, a variance change, or both the mean and variance change. The intrinsic Bayes factors of Berger and Pericchi (1996, 1998) are used to find the type of optimal change model. The Gibbs sampling including the Metropolis-Hastings algorithm is used to estimate all the parameters in the change model. These methods are checked via simulation and applied to the winter average temperature data in Seoul.

Keywords: Change point analysis, Intrinsic Bayes factor, Bayesian model selection, Parameter estimation, Gibbs sampling, Metropolis-Hastings algorithm, Univariate normal distribution.

* This work was supported by grant No. R05-2000-000-00024-0 from the Korea Science & Engineering Foundation.

1) Doctoral Course, Dept. of Statistics, Chonnam National University, 300, Yongbong-dong, Buk-gu, Gwangju, 500-757, Rep of Korea.

E-mail: ksook620@stat.chonnam.ac.kr

2) Professor, Dept. of Statistics, Chonnam National University, 300, Yongbong-dong, Buk-gu, Gwangju, 500-757, Rep of Korea.

E-mail: ysson@chonnam.ac.kr