

석유공급교란에 대한 변화점 분석 및 분포 추정: 베이지안 접근

박건*·이성수**

* 서울시정개발연구원

** 경기대학교 경제학부

A Change-Point Analysis of Oil Supply Disruption: Bayesian Approach

Chun Gun Park*·Sung-Su Lee**

* Seoul Development Institute

** Professor of School of Economics, Kyonggi University

Key Words : Oil Supply Disruption, Change-Point, Weibull Distribution, Markov Chain Monte Carlo(MCMC)

Abstract

Using statistical methods a change-point analysis of oil supply disruption is conducted. The statistical distribution of oil supply disruption is a weibull distribution. The detection of the change-point is applied to Bayesian method and weibull parameters are estimated through Markov chain monte carlo and parameter approach. The statistical approaches to the estimation for the change-point and weibull parameters is implemented with the sets of simulated and real data with small sizes of samples.

1. 서 론

Leiby(1997)가 석유비축모형의 핵심인 석유공급 교란의 형태를 파악하고자 와이블분포를 도입하였다. 일반적으로 와이블분포는 실패율 및 수명과 같은 자료에 사용되었으며, 모수추정에 관한 연구도 오래 전부터 수행되어왔다. 본 논문에서 제시된 와이블분포는 형태모수와 척도모수로 결정된 것이다.

Murthy(2004)가 와이블분포의 두 모수추정에 관한 방법론을 정리하였다. 두 모수를 가지는 와이블분포의 추정방법으로 (1) 그래픽적 방법과 (2) 최대우도방법 (3) 베이지안 방법을 들 수 있다. 특히, 최대우도추정량은 대표본에서 최적성질(Rao, 1973)을 가지며 반복수치적 기법으로 모수를 추정한다.

Lockhart와 Stephens(1994)는 세 개의 모수로 결정되는 와이블분포 추정에 대한 다른 반복수치적 방법을 제시하였다. 그래픽적 방법으로부터 얻어진 값들은 모수적 추정에서 이용되는 반복기법의 초기값으로 많이 쓰인다.

본 연구의 초점은 한 개의 변화점이 존재하는 데이터로부터 변화점의 추정뿐만 아니라 그 시점을 기준으로 분리된 데이터에 대한 와이블분포들을 베이지안 방법으로 추정하는데 있다. 일반적으로 베이지안 방법은 추정하고자 하는 모수의 사전정보와 데이터의 분포(우도)의 곱으로 표현되는 사후분포를 이용하는 기법이다. 여기서 관심이 집중되는 모수에 대한 주변조건부분포를 구하는 것이 관건인데 일반적으로 이끌어내기 쉬운 형태로 이루어진 경우가 드물다. 이러한 경우에는 MCMC(Markov Chain Monte Carlo) 기법이 적용되는데, 자세한 검토는 Gelman et al. 및

† 교신저자 chungunpark@daum.net

Gilks et al.의 저서(1996)와 논문(1992, 1995)에 제시되었다.

본 논문의 구성은 다음과 같다. 2절은 석유공급교란의 정의 및 적용분포에 관한 내용을 기술한다. 3절은 한 개의 변화점이 존재하는 공급교란의 모형설정 및 베이지안 추정 절차를 설명한다. 4절은 소표본에 대한 모의실험으로 변화점이 존재하는 와이블분포에 적용된 베이지안 추정과 제시된 모형을 공급교란에 적용한다. 마지막은 본 연구로부터 얻어진 결론 및 토의를 제시한다.

2. 자료 및 와이블 모수의 추정

석유공급교란은 일반적으로 특정한 국가 또는 국가들로부터 발생하는 석유의 손실을 의미하며¹⁾, 그 형태로는 통상금지(embargo)와 같이 산유국 및 국제 정세 불안, 산유국간의 전쟁과 분쟁, 자연재해로 인한 유전 및 정유시설의 손실 및 유전 및 유조선의 사고 등을 들 수 있다.

Belgrave, Ebinger과 Okino(1988)는 “Energy Security to 2000”에서 석유공급교란을 초래할 수 있는 7가지 발생요인을 제시한 바 있다. 석유비축의 Risk Premium 산정과 적정수준 검토(박창원, 이영구, 2000)에서 석유공급교란의 측정으로 일평균공급부족량을 일평균소비량으로 나눈 값으로 표시했다. 세계석유소비량이 점차 증가추세이고 각 공급교란이 상호 독립이기 때문에 공급교란 비율이 다소 세계소비량에 영향을 받는 경향이 있다. 따라서 일평균공급부족율로 교란분포의 추정은 다소 석유소비량에 영향을 받을 수 있다. 이러한 문제를 해결하기 위해 공급교란의 측정을 일평균공급부족량으로 나타내는 것이 타당하다.

일평균공급부족량에 적용되는 통계적 분포는 와이블분포이다. 확률변수 X 가 형태모수(b)와 척도모수(a)를 가지는 와이블분포를 따른다면 확률밀도함수는 아래와 같이 표시된다.

$$f(x|a,b) = \frac{b}{a} \left(\frac{x}{a}\right)^{b-1} \exp\left[-\left(\frac{x}{a}\right)^b\right], \quad a,b > 0, x \geq 0. \quad (1)$$

주어진 자료 (x_1, x_2, \dots, x_n) 가 와이블분포를 따른다고 가정하면, 식 (1)로부터 로그우도함수는 다음과

같다.

$$\ln L(x_1, \dots, x_n | a, b) = n \ln b + n \ln a^* + (b-1) \sum_{i=1}^n \ln x_i - a^* \sum_{i=1}^n x_i^b \quad (2)$$

여기서 $a^* = 1/a$ 로 표시된다.

전통적인 방법으로 모수추정을 하려면, 식 (2)의 로그우도함수를 최대로 하는 모수값을 찾는 것이다. 식 (2)의 로그우도함수는 미지의 모수에 대해 비선형 연립방정식으로 구성되어 있어 수치적 접근인 Newton-Raphson법으로 모수의 추정값을 구해야한다(Murthy, 2004). 대표본의 경우에 최대우도추정량은 정칙조건(regularity conditions)하에서 일치성(consistency), 점근적 불편성(asymptotically unbiased), 효율성(efficiency), 그리고 정규분포(normally distributed)의 성질을 가지고 있다(Rao, 1973). 그러나 소표본에서 최대우도추정량은 위의 성질로 보존되지 않는다.

위의 문제를 극복하고자 제안된 방법은 베이지안 접근이다. 베이지안 방법은 자료의 정보인 우도함수와 관심 모수정보인 사전분포를 결합하여 사후분포를 얻은 후 각 모수에 대한 주변분포를 구하여 사후평균으로 모수추정값을 계산하는 것이다. 와이블분포를 결정하는 두 모수의 사전분포는 무정보(non-informative)를 바탕으로 설정한다.

$$p(a,b) \propto 1/a \quad (3)$$

그리고 식 (2)에서 지수를 취한 우도함수와 식 (3) 사전분포의 곱으로 표시된 결합사후분포는 다음과 같다.

$$p(a,b|X) \propto p(a,b) \prod_{i=1}^n f(x_i|a,b) \quad (4)$$

식 (4)에서 각 모수의 주변사후분포를 구하기 어렵기 때문에 베이즈이론(Bayes' theorem)을 적용하여 척도모수와 형태모수의 조건부사후분포를 구한다. 그 형태는 아래와 같다.

$$p(a|b,X) \propto a^{-nb-1} \exp\left[-a^{-b} \sum_{i=1}^n x_i^b\right]. \quad (5)$$

식 (5)로부터 척도모수의 조건부사후분포는 $Y = g(a) = a^b$ 로 변수변환을 하면 조건부확률변수 $Y|b,X$ 는 모수가 $\left(n, \sum_{i=1}^n x_i^b\right)$ 인 역감마함수(inverse gamma distribution)이다. 또한 조건부확률분포 $b|a,X$ 는 일반적으로 알려진 분포가 아니며, 변수변환이 어려운 형태이다. 그러나 형태모수 b 의 조건부확률분포는 로그오목(log-concave)의 조건을 만족한다.

1) <http://www.eia.doe.gov/emeu/security/distable.html> 을 참조하십시오.

그래서 이 조건부사후분포로부터 난수발생을 하기 위해 적응기각추출법(adaptive rejection sampling)이 적용된다. 적응기각추출기법의 자세한 설명은 Gilks (1992)과 Gilks와 Wild(1992)의 논문에 기술되어 있다.

와이블 분포의 형태모수와 척도모수에 관한 베이 지안 추론의 절차는 각 모수의 조건부사후분포로부터 깃스샘플기법과 적응기각추출을 이용한 것으로 아래와 같은 순서로 표본을 추출한다.

초기값이 주어짐: $(a^{(0)}, b^{(0)})$

반복 ($i=1:N$) $\{a^{(i)} \sim p(a | b^{(i-1)}, X)$ 에서 표본 생성
 $b^{(i)} \sim p(b | a^{(i)}, X)$ 에서 표본 생성}

여기서 $a|b, X$ 는 역감마함수로 변수변환을 하여 표 본을 추출하고 $b|a, X$ 의 난수생성은 적응기각추출법 이 적용된다.

위의 절차로부터 형태모수(b)와 척도모수(a)의 추 출된 N 개의 표본 중에서 초기값에 영향을 받지 않는 표본들만으로 모수추정에 사용한다. 모수추정은 주 어진 표본의 평균으로 구하며 그 평균을 사후평균 (posterior mean)이라고 한다.

$$\hat{a} = \frac{1}{N-k} \sum_{j=k+1}^N a^{(j)}, \quad \hat{b} = \frac{1}{N-k} \sum_{j=k+1}^N b^{(j)}$$

3. 변화점 추정

일평균공급부족량을 확률변수 $x_i, 1 \leq i \leq n$ 이 라 정의하자. 알려지지 않은 시점 t 을 변화점이라 가 정하면, $x_i \sim f_1(x|a_1, b_1), i = 1, \dots, t$ 이고 $x_i \sim f_2(x|a_2, b_2), i = t+1, \dots, n$ 이다. 한 개의 변화점이 있는 와이블분포의 결합분포(joint distribution)는 다음과 같다.

$$f_n(X|t, \theta) = \begin{cases} \left[\prod_{i=1}^t f_1(x_i | a_1, b_1) \right] \left[\prod_{j=t+1}^n f_2(x_j | a_2, b_2) \right], & 2 \leq t \leq n-1, \\ \prod_{i=1}^n f(x_i | a, b), & t = n \end{cases}$$

(6)

여기서 $\theta = (a_1, b_1, a_2, b_2)$ 이고 만약 $t = n$ 이면 $\theta = (a, b)$ 이다.

변화점에 대한 베이 지안 추론을 적용하기 위해서 변화점의 사전분포는 모든 시점이 동일하게 선택될

수 있는 균등분포로 정의한다. 식 (6)과 변화점의 사 전분포로부터 사후분포는 변화점이 존재하는 우도함 수와 같은 형태로 구성된다. 변화점의 추정에 사용되 는 방법으로는 가설검정 모형과 최대확률의 개념이 있다. 가설검정모형은 귀무가설과 대립가설을 설정 하여 이 둘 중에 하나를 베이 지안 접근으로는 선택하 는 방법으로 베이즈인자(Bayes factor)가 있다.

본 논문에서 변화점 추정은 주어진 변화점에 관한 사후확률을 구하여 최대확률을 가지는 시점을 선택 하는 방법을 채택하였다. 또한 변화점 추정의 과정 에서 추정해야할 미지모수들에 대해서는 모수적 접 근으로 구했고 그것들을 바탕으로 변화점의 사후확 률을 얻었다. 이와 같은 접근을 한 이유는 모든 모수 들에 대한 MCMC의 적용으로 발생하는 계산상의 복 잡성을 줄이는데 있다.

4. 모의실험 및 적용

4.1 모의실험

변화점 및 와이블모수의 추정에 대한 정확도를 설 명하기 위해 와이블분포를 따르며 한 개의 변화점이 있는 데이터에 대한 모의실험을 실행한다. 각 모의실 험에 적용될 표본의 크기는 25개이며 변화점($t=14$)을 기점으로 14개의 표본은 형태모수와 척도모수가 (2.0, 1.5)이고 나머지 11개는 (1.0, 1.3)으로 100개 의 반복 데이터를 생성한다. <그림 1>은 설정된 모 수에 따른 분포형태를 나타낸 것이며, 모수 (2.0, 1.5)은 (1.0, 1.3)에 비해 최빈수가 우측으로 이동한 반면 우측 꼬리부분은 얇은(light tailed) 경향을 보 이다.

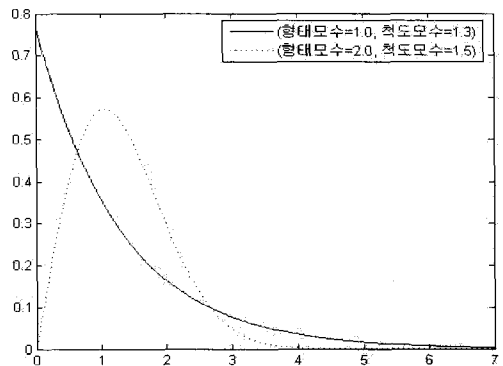


그림 1] 와이블분포의 형태

모수추정의 정확성을 모의실험을 통하여 다음과 같은 기준으로 평가한다.

$$\text{Mean Squared Error} : \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \theta)^2,$$

여기서 θ 는 참모수이며 $\hat{\theta}_i$ 는 각 모의실험의 추정값이다.

변화점이 존재하는 데이터에서 변화점을 고려하지 않고 와이블모수를 추정하면 그 접근 방법이 무엇이든 추정값은 변화점으로 인해 오염될 것이다. 위의 모의실험 설정에 대해 앞에서 언급한 접근들로 모수를 추정한 결과가 <표 1>에 제공되며, 형태모수와 척도모수의 표본평균은 데이터가 많은 모수값들에 치우치는 경향을 보인다.

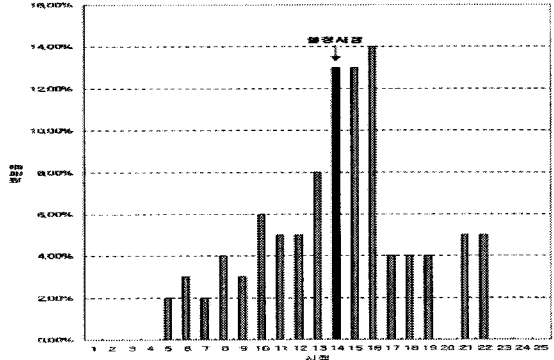
<표 1> 변화점이 있는 경우에 변화점이 없다는 가정하에 와이블 분포의 추정(표본의 크기=25)

모 수	통계량	Bayesian	MLE
형태모수 (= 2.0, 1.0)	Max	1.9919	2.1398
	Min	0.7604	0.8732
	Mean	1.2214	1.3397
	Variance	0.0682	0.0704
척도모수 (= 1.5, 1.3)	Max	2.3381	2.3450
	Min	0.9454	0.9472
	Mean	1.3719	1.3689
	Variance	0.0644	0.0651

변화점이 존재하는 데이터에서 올바른 추론방법은 식 (6)의 모형을 기반으로 변화점과 모수를 모두 고려하는 것이다. 본 논문에서 제시한 변화점의 추정방법은 베이지안 접근으로 변화점에 대한 조건부사후 분포인 식 (6)을 최대확률로 하는 변화점을 구하고 각 와이블분포에 대한 모수를 추정하는 것이다.

변화점의 추정결과가 <그림 2>에 제시되었다. 100개의 데이터집합으로부터 정확한 시점을 찾아낸 것은 13개고 주변 시점, 즉 13, 15 및 16을 포함하면 48개 정도이다.

변화점의 정확한 추정도 중요하지만 그 시점으로 전후한 분포의 형태를 올바르게 추정하는 것도 중요하다. 그 이유는 변화점을 기준으로 표본의 수가 상당히 작아지고 모수의 추정도 표본의 수와 비례해서 신뢰도가 떨어지기 때문이다.



[그림 2] 변화점의 추정 결과

100개의 데이터집합으로부터 변화점을 모두 정확하게 추정했다는 가정하에 소표본에서의 모수추정을 베이지안 방법과 모수적 방법을 비교한 결과가 <표 2>와 <표 3>에 제시되어 있다. 그 결과로부터 베이지안 방법의 추정값이 모수적 방법보다 대체적으로 참모수에 가까운 결과를 보였다.

<표 2> 와이블 분포의 모수추정(표본의 크기=14, 100개의 집합)

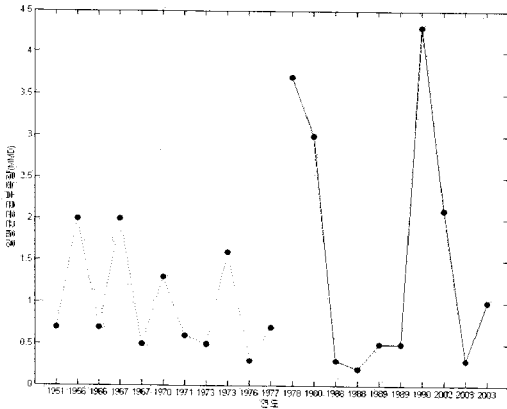
모 수	통 계 량	Bayesian	MLE
형태모수 (= 2.0)	MSE	1.5842	1.9396
	Max	4.3206	4.5543
	Min	1.2719	1.3789
	Mean	2.1319	2.2747
	Variance	0.3059	0.3179
척도모수 (= 1.5)	MSE	0.1203	0.1486
	Max	2.4010	2.3714
	Min	0.9171	0.8829
	Mean	1.5239	1.4893
	Variance	0.0709	0.0697

4.2 실제자료

[그림 3]은 1951년부터 2003년까지 일평균공급 부족량을 표시한 것이다. 이 데이터의 분포 형태를 알아보기 위해 앞 장에서 제시한 방법들이 적용된다.

<표 3> 와이블 분포의 모수추정(표본의 크기=11, 100개의 집합)

모수	통계량	Bayesian	MLE
형태모수 (= 1.0)	MSE	0.0622	0.0752
	Max	1.7440	1.8893
	Min	0.5301	0.6450
	Mean	0.9824	1.1082
	Variance	0.0625	0.0641
척도모수 (= 1.3)	MSE	0.1612	0.1552
	Max	2.5792	2.5283
	Min	0.5020	0.4763
	Mean	1.3033	1.2539
	Variance	0.1628	0.1547



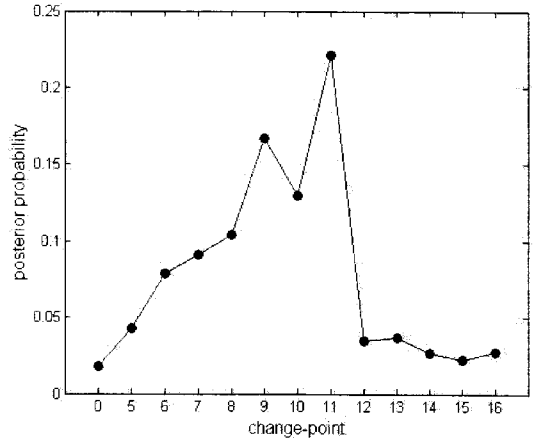
<그림 3> 일평균공급부족량

<표 4>는 일평균공급부족량에 변화점이 없다는 가정으로 한 와이블모수 추정으로 모수적 방법과 베이저안 방법이 적용된 결과이다. 두 방법에서 척도모수의 추정은 차이가 없지만 형태모수에서 다소 모수적 방법이 큰 값으로 추정되었다. 앞에서 실행한 모의실험 결과에 의하면 변화점이 존재하면 추정된 모수는 적용방법에 관계없이 변화점에 의해 왜곡된 값으로 변질되었다.

<표 4> 변화점이 존재하지 않는 가정으로 일평균공급부족량의 와이블분포 추정

추정법	형태모수	척도모수
Bayesian	1.0797	1.3650
MLE	1.1897	1.3620

[그림 4]로부터 일평균공급부족량의 변화점의 최대사후확률을 가지는 연도는 1977년이다. 따라서 1951년부터 1977년까지의 공급부족량에 대한 분포와 1978년~2003년에 해당하는 분포는 다른 형태를 갖는다.



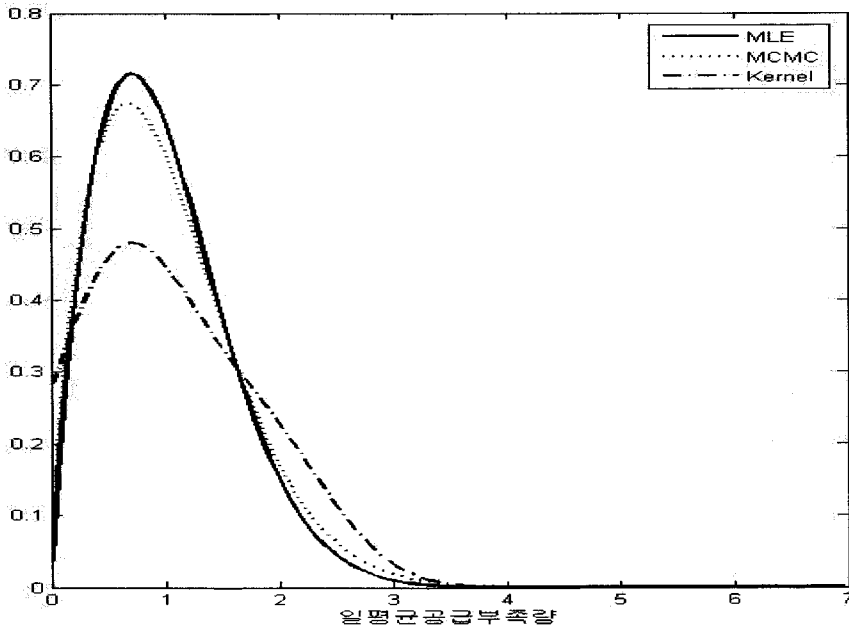
<그림 4> 일평균공급부족량의 변화점에 관한 사후확률

1951년부터 1977년까지의 공급부족량에 대한 분포 추정과 1978년~2003년에 해당하는 와이블모수 추정이 <표 5>에 제시되었다.

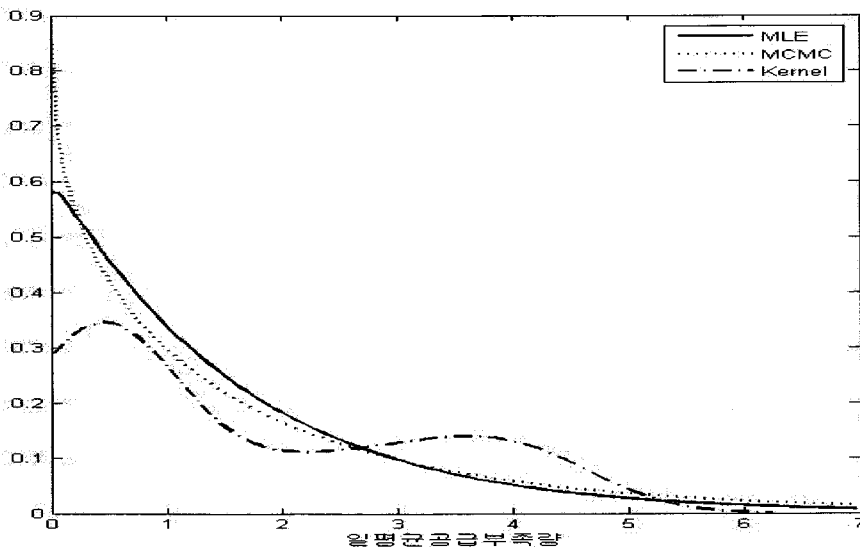
<표 5> 일평균공급부족량에서 변화점이 존재하는 와이블 분포로 모수추정

연도	추정법	형태모수	척도모수
1951~1977 (n=11)	Bayesian	1.6841	1.1532
	MLE	1.7949	1.1227
1978~2003 (n=10)	Bayesian	0.9000	1.7622
	MLE	1.0162	1.6011

공급부족량의 분포들을 추정하는데 베이저안 방법과 모수적 방법이 이용되었으며 1951년~1977년에서 추정된 와이블분포의 모수와 1978년~2003년의 일평균석유공급부족량에 해당하는 와이블분포의 모수들은 다소 차이가 있다. <그림 5>와 <그림 6>은 변화점을 기점으로 한 추정된 와이블분포의 형태이다. 여기서 제시된 핵밀도추정은 적은 표본을 바탕으로 추정된 분포이기 때문에 모수적 방법이나 베이저안 방법으로 추정된 와이블분포와 다소 차이는 있을 수 있다.



<그림 5> 1951년~1977년 일평균공급부족량의 와이블 분포의 형태



<그림 6> 1978년~2003년 일평균공급부족량에 대한 와이블 분포의 형태

5. 결론 및 토의

변화점이 존재하는 데이터의 분포추정에서 먼저

다루어야 할 부분은 변화점의 탐색이다. 본 논문은 변화점이 존재하는 와이블모형을 정의하고 베이지안 방법인 사후확률이 최대인 변화점을 선택하는 기법을 기술하였다. 특히, 소표본의 모의실험을 통하여

변화점의 추정결과와 정확도가 대략 50%정도로 나타났으며 와이블모수 추정에 있어 베이지안 방법이 모수적 방법보다 정확하게 나왔다. 따라서 소표본 실제 데이터를 분석하는데 적용되는 기법은 앞의 모의 실험의 결과와 모수적 방법이 소표본에 대해 여러 가지 성질을 보존하기 어렵기 때문에 베이지안 접근이 타당하다.

변화점 와이블모형이 21개 일평균공급부족량에 적용되었고 베이지안 방법으로 추정된 변화점은 석유교란 및 유가에 상당한 의미를 제공하였다. 또한 변화점이 존재하는 와이블모형의 일반화 및 다양한 모수값들에 대한 추정과 신뢰성에 관한 연구가 필요하다.

참 고 문 헌

[1] 박창원, 이영구 (2002), 석유비축의 Risk Premium 산정과 적정수준 검토, 에너지경제연구원.
 [2] Belgrave, R., Ebinger, C. K. and Okino, H. (1988), "Energy Security to 2000" , *Energy Policy* Vol. 16.
 [3] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, Chapman & Hall.

[4] Gilks, W. R. (1992), "Derivative Free Adaptive Rejection Sampling for Gibbs Sampling", *Bayesian Statistics*, Oxford University Press, Vol.4, pp. 859.
 [5] Gilks, W. R. and Wild, P. (1992) "Adaptive rejection sampling for Gibbs sampling" , *Applied Statistics*. 41, pp. 337-348.
 [6] Gilks, W. R., Best, N. G. and Tan, K. K. C. (1995), "Adaptive rejection Metropolis sampling" , *Applied Statistics*, 44, pp. 455-472.
 [7] Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, pp. 486.
 [8] Leiby, P. et al. (1997), "Oil Imports: An Assessment of Benefits and Costs." , Oak Ridge National Laboratory.
 [9] Lockhart, R. A. and Stephens, M. A. (1994), "Estimation and tests of fit for the three parameter Weibull distributions" , *Journal Royal Statistics Society*, B56, pp. 491-500.
 [10] Prabhaker, D. N., Murthy, Xie, M. and Jiang, R. (2004), *Weibull Models*, John Wiley and Sons.
 [11] Rao, C. R. (1973), "Linear Statistical Inference and its Applications" , John Wiley and Sons, pp. 625.