

Hierarchical Bayes Analysis of Smoking and Lung Cancer Data

Man-Suk Oh¹⁾, Hyun-Jin Park²⁾

Abstract

Hierarchical models are widely used for inference on correlated parameters as a compromise between underfitting and overfitting problems. In this paper, we take a Bayesian approach to analyzing hierarchical models and suggest a Markov chain Monte Carlo methods to get around computational difficulties in Bayesian analysis of the hierarchical models. We apply the method to a real data on smoking and lung cancer which are collected from cities in China.

Keywords : Hierarchical model, Correlated parameters, Markov chain Monte Carlo, Meta analysis.

1. 서론

통계추론의 기본요소는 변수 X 의 관측치 x 와 X 의 모형을 결정해주는 모수 θ 이다. 우리가 추정해야 할 모수 θ 가 한 개가 아니라 여러 개 $\theta_j, j=1, \dots, J$, 일 때 이 다중 모수들 사이가 독립이라면 그 모수들을 가지고 모형을 만드는 데는 별 어려움을 느끼지는 않을 것이다. 그러나 이런 경우는 매우 드물고 대부분의 경우에 다중 모수 간에 의존성이 존재하는데 이러한 의존성을 반영하는 결합확률모형을 구성하는 데는 어려움이 따른다. 이런 문제를 쉽게 해결하기 위한 방편으로 초모수(hyperparameter)라 정의되는 특정한 모수에 기반을 두고 계층적으로 모형화 해나가는 계층모형(hierarchical model)이 유용하게 사용되고 있다 (Good, 1965; Mosteller and Wallace, 1964; Hill, 1965; Lindley, 1971).

일반적으로 비계층 모형의 경우 모수의 수가 작은 간단한 모형은 좋은 적합을 얻기 어려우며 반대로 모수가 많은 복잡한 모형은 기존의 자료에는 잘 적합되지만 새로운 자료에 대해서는 큰 오차를 발생시킬 수 있는 이른바 과적합(overfitting)이 일어날 수 있는 문제가 생긴다. 계층모형은 이 둘을 보완하는 방법으로 과적합 문제를 피하면서도 모수들 사이에서의 의존성을 구조화한

1) Associate Professor, Department of Statistics, Ewha Womens University, Sodaemoon Gu, Seoul 120-750, KOREA, This work was supported by BK21 Korea.

E-mail: msoh@mm.ewha.ac.kr.

2) Consulting Division, SAS Korea, Youngdeungpo Gu, Seoul 150-010, KOREA,

E-mail: korhjp@kor.sas.com

모집단분포를 사용하여 자료를 적합시키기에 충분한 모수를 가지고 있다는 장점이 있어 실제 자료분석에서 널리 쓰이고 있다 (Fay and Herriot, 1979; Dempster and Raghunathan, 1987; Mollie and Richardson, 1991; Richardson and Gilks, 1993; Meng and Dempster, 1987, Gelman et. al., 1995).

본 논문에서는 중국 여러 도시에서 수집된 흡연과 폐암 발생률 자료를 계층 모형을 사용한 베이저안 기법을 사용하여 분석하고자 한다. 이는 자료의 형태가 같은 주제 아래서 수행되었으나 서로 다른 시간과 공간상에서 수집되었으므로 베이저안 메타분석(meta analysis)의 일종이라 할 수 있다 (Carlin, 1992).

2장에서는 계층모형에 대한 베이저안 분석에 대하여 설명하며 3장에서는 흡연과 폐암 발생률 자료에 대하여 계층적 베이저안 기법을 적용하여 분석한다.

2장. 정규 계층모형

각 실험단위별로 또는 그룹별로 각각 다른 평균을 가지며, 분산이 알려진 관측치가 다음과 같이 독립적으로 계층적 정규분포를 따른다고 하자:

$$y_j | \theta_j \sim N(\theta_j, \sigma_j^2),$$

$$\theta_j | \mu, \tau^2 \sim N(\mu, \tau^2), \quad i=1, \dots, n_j, \quad j=1, \dots, J.$$

이 때 σ_j 는 알려진 상수이며 θ_j 의 모수 μ 와 τ 를 초모수 (hyper-parameter)라 부른다. 베이저안 추론을 위하여 초모수에 대한 사전분포로 먼저 τ 의 사전분포를 $p(\tau)$ 라 하고 τ 가 주어졌을 때 μ 의 사전분포를 $p(\mu|\tau)$ 라 한다. $p(\mu|\tau)$ 로는 보통 상수 무정보사전 밀도함수, 즉, $p(\mu|\tau)=1$ 을 사용한다. 따라서 μ 와 τ 의 결합사전밀도함수는 $p(\mu, \tau) = p(\mu|\tau) \cdot p(\tau) \propto p(\tau)$ 와 같이 된다.

베이저안 분석은 모수의 사후분포에 근거한다. 그런데 계층 모형의 경우 사후분포가 쉬운 형태로 주어지지 않는 경우가 대부분이다. 이를 해결하기 위한 방법으로 몬테칼로 기법을 사용하여 모수의 사후표본을 생성하고 이들 사후표본을 기반으로 한 베이저안 기법이 주로 사용되고 있다. 따라서 모수의 사후표본 생성이 본 논문에서 다루는 분석기법의 기반이라 할 수 있다 (Gelfand and Smith, 1990).

이제 계층 모형 하에서 각 모수의 사후표본 생성을 위한 사후분포를 유도해 보자. 먼저 자료가 주어졌을 때 모수 $\theta=(\theta_1, \dots, \theta_j), \mu, \tau$ 의 결합사후밀도함수는

$$p(\theta, \mu, \tau | y) \propto p(\mu, \tau) p(\theta | y, \mu, \tau)$$

$$\propto p(\mu, \tau) p(y | \theta) p(\theta | \mu, \tau)$$

$$\propto p(\tau) p(y | \theta) p(\theta | \mu, \tau)$$

$$\propto p(\tau) \prod_{j=1}^I \phi((y_j - \theta_j)/\sigma_j) \prod_{j=1}^I \phi((\theta_j - \mu)/\tau) \quad (1)$$

이다. 여기에서 $\phi(\cdot)$ 은 표준 정규 밀도함수를 나타낸다. 위 식으로부터 θ_j 의 조건부 사후분포를 구하면 잘 알려진 바와 같이

$$\theta_j \sim N(\hat{\theta}_j, V_j),$$

$$\hat{\theta}_j = \frac{y_j/\sigma_j^2 + \mu/\tau^2}{1/\sigma_j^2 + 1/\tau^2}, \quad V_j = \frac{1}{1/\sigma_j^2 + 1/\tau^2} \quad (2)$$

이다. 다음, μ, τ 의 주변사후분포를 유도하면 $p(\mu, \tau) \propto p(\mu, \tau) p(y|\mu, \tau)$ 인데,

$y_j | \theta_j \sim N(\theta_j, \sigma_j^2)$, $\theta_j \sim N(\mu, \tau^2)$ 로부터 $y_j | \mu, \tau \sim N(\mu, \sigma_j^2 + \tau^2)$ 라는 것을 쉽게 구할 수 있다. 그러므로,

$$p(\mu, \tau) \propto p(\mu, \tau) \prod_{j=1}^I \phi((\mu - y_j)/\sqrt{\sigma_j^2 + \tau^2})$$

가 된다. 위의 $p(\mu, \tau)$ 로부터 τ 가 주어졌을 때 μ 의 사후분포 $p(\mu|\tau, y)$ 는

$$\begin{aligned} p(\mu|\tau, y) &\propto p(\mu, \tau) \propto p(\mu, \tau) \prod_{j=1}^I \phi((y_j - \mu)/\sqrt{\sigma_j^2 + \tau^2}) \\ &\propto \prod_{j=1}^I \phi((y_j - \mu)/\sqrt{\sigma_j^2 + \tau^2}) \propto \exp\left(-\sum_{j=1}^I \frac{(y_j - \mu)^2}{2(\tau^2 + \sigma_j^2)}\right) \\ &\propto \exp\left(-\frac{1}{2}\left(\sum_{j=1}^I \frac{1}{\tau^2 + \sigma_j^2}\right)\mu^2 - 2\left(\sum_{j=1}^I \frac{y_j}{\tau^2 + \sigma_j^2}\right)\mu\right) \\ &= \exp\left(-\frac{1}{2}\sum_{j=1}^I \frac{1}{\tau^2 + \sigma_j^2}(\mu^2 - 2\mu \sum_{j=1}^I \frac{y_j}{\tau^2 + \sigma_j^2})\right) \\ &\propto \exp\left(-\frac{1}{2}\sum_{j=1}^I \frac{1}{\tau^2 + \sigma_j^2}\left(\mu - \frac{\sum_{j=1}^I \frac{y_j}{\tau^2 + \sigma_j^2}}{\sum_{j=1}^I \frac{1}{\tau^2 + \sigma_j^2}}\right)^2\right) \end{aligned}$$

$$\propto \phi((\mu - \hat{\mu})/\sqrt{V_\mu})$$

인데 이 때

$$\hat{\mu} = \frac{\sum_{j=1}^I \frac{y_j}{\tau^2 + \sigma_j^2}}{\sum_{j=1}^I \frac{1}{\tau^2 + \sigma_j^2}}, \quad V_\mu = \frac{1}{\sum_{j=1}^I \frac{1}{\tau^2 + \sigma_j^2}} \tag{3}$$

이다.

끝으로 τ 의 주변 사후분포 $p(\tau|y)$ 는

$$\begin{aligned} p(\tau|y) &= p(\mu, \tau|y) / p(\mu|\tau, y) = p(\tau)p(y|\mu, \tau) / p(\mu|\tau, y) \\ &\propto \frac{p(\tau) \prod_{j=1}^I \phi((y_j - \mu) / \sqrt{\sigma_j^2 + \tau^2})}{\phi((\mu - \hat{\mu}) / \sqrt{V_\mu})} \\ &\propto \frac{p(\tau) \prod_{j=1}^I (\sigma_j^2 + \tau^2)^{-1/2} \exp(-\frac{(y_j - \mu)^2}{2(\sigma_j^2 + \tau^2)})}{V_\mu^{-1/2} \exp(-\frac{(\mu - \hat{\mu})^2}{2V_\mu})} \end{aligned} \tag{4}$$

이다 (Oh, 1999). 위 식은 임의의 μ 에 대하여 성립하므로 $\mu = \hat{\mu}$ 로 놓으면,

$$p(\tau|y) \propto p(\tau) V_\mu^{1/2} \prod_{j=1}^I (\sigma_j^2 + \tau^2)^{-1/2} \exp(-\frac{(y_j - \hat{\mu})^2}{2(\sigma_j^2 + \tau^2)})$$

인데 보다시피 τ 의 함수는 매우 복잡한 형태를 가지고 있다. 이제 위에서 유도한 사후분포들을 이용하여, 우리는 모수 θ, μ, τ 의 사후분포를 쉽게 얻을 수 있다. 제일 먼저 $p(\tau|y)$ 로부터 τ 를 생성하는데 $p(\tau|y)$ 의 식이 편리한 형태가 아니므로 채택-기각 방법(Accept-Reject Method)을 이용한다 (Rubinstein, 1981). 생성된 τ 가 주어진 조건 하에서 $p(\mu|\tau, y)$ 로부터 μ 를 생성하고 다음 $p(\theta|\mu, \tau, y)$ 로부터 θ 를 생성하면 된다. 이렇게 생성된 μ, τ, θ 의 사후표본을 이용하여 각 모수에 대한 사후추론을 수행하면 된다.

3장. 흡연과 폐암 발생률 자료의 분석

이 장에서는 중국의 8개 도시에서 조사된 흡연자와 비흡연자 사이의 폐암 발생유무에 대한 범주형 자료를 2장의 계층모형을 이용하여 분석하고자 한다. 여러 곳에서 모아진 자료를 하나의 주제아래 결합시켜 분석하려는 것을 메타분석(meta analysis)이라고 하는데, 각 도시별로 차이점은 인정하지만, 예측할 수 있는 영향을 가지는 사전정보(priori)를 기대하지 않고, 각 도시는 모두 동일하다고 가정한다.

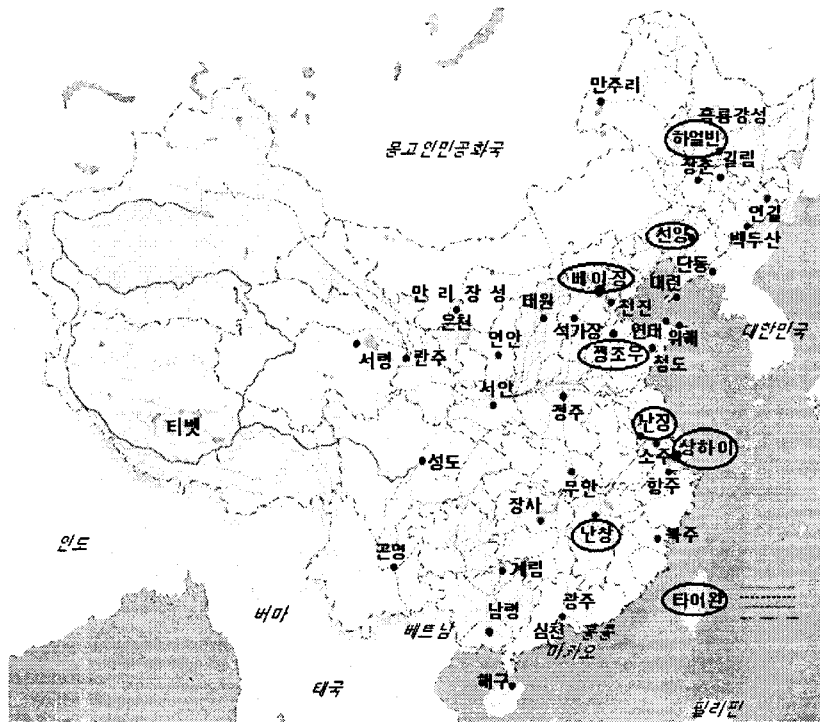


그림 1. 중국지도

도시	흡연	폐암	
		예	아니오
베이징	예	126	100
	아니오	35	61
상하이	예	908	688
	아니오	497	807
선양	예	913	747
	아니오	336	598
난징	예	235	172
	아니오	58	121
하얼빈	예	402	308
	아니오	121	215
쑹조우	예	182	156
	아니오	72	98
타이완	예	60	99
	아니오	11	43
난창	예	104	89
	아니오	21	36

표 1. 중국인의 흡연과 폐암자료 (Liu, 1992)

<그림 1>에 표시되어 있는 곳이 조사가 행해진 도시 8곳이다. 데이터는 <표 1>에 주어져 있다.

이와 같은 범주형 자료의 경우 각 도시별로 로그오즈비(log odds ratio)에 대하여 정규분포 근사를 사용할 수 있는데 우리는 이 로그오즈비에 대하여 정규 계층 모형을 가정하고 베이지안 분석을 행하고자 한다.

먼저 각 도시별로 로그오즈비를 살펴보자. n_{11j} 를 j번째 도시에서 흡연자중 폐암에 걸린 사람 수, n_{12j} 를 j번째 도시에서 흡연자중 폐암에 안 걸린 사람 수, n_{21j} 를 j번째 도시에서 비흡연자 중 폐암에 걸린 사람수, n_{22j} 를 j번째 도시에서 비흡연자중 폐암에 안 걸린 사람수라 하면 각 도시별 로그오즈비의 추정값은

$$y_j \equiv \hat{\theta}_j = \log\left(\frac{n_{11j}}{n_{12j}}\right) - \log\left(\frac{n_{21j}}{n_{22j}}\right)$$

이고 이에 대한 분산은 대략적으로

$$\sigma_j^2 = \frac{1}{n_{11j}} + \frac{1}{n_{12j}} + \frac{1}{n_{21j}} + \frac{1}{n_{22j}}$$

이며, 표본크기가 큰 경우 근사적으로 $y_j \sim N(\theta_j, \sigma_j^2)$ 이다 (박태성, 이승연, 1998). 따라서 2장에서와 같이 계층 모형, 즉, $\theta_j \sim N(\mu, \tau^2)$ 를 가정할 수 있다.

계층 모형의 베이저안 분석에 들어가기 전에 로그오즈비의 추정치를 살펴보면 <표 2>과 같이 된다. <표 2>의 오즈비로부터 우리가 알 수 있는 사실은 우선 흡연이 비흡연 보다 폐암과의 연관이 대체로 2배정도 높다는 것이다. 특히, 난징의 오즈비가 2.850으로 제일 높고 쑹조우의 오즈비가 1.588로 가장 낮다.

중국의 경우 서양에서의 짧은 담배보다 담뱃대를 이용한 흡연자들이 많다. 필터가 긴 담배의 경우가 짧은 담배보다 폐암에 걸릴 확률이 더 낮다는 연구결과를 보고 유추하건데, 짧은 담배를 많이 피는 서양이나 우리나라의 경우 오즈비는 더 커질 것으로 예상할 수 있겠다.

이제 로그오즈비에 대한 계층모형에 2장에서 기술한 베이저안 분석을 수행하여 보자. 베이저안 분석을 위하여 τ 의 사전분포가 요구되는데, 이 예에서 우리는 무정보 사전분포 $p(\tau) = 1$ 을 사용하기로 하겠다. 모수들의 사후표본을 얻기 위하여 τ, μ, θ 의 순으로 사후표본을 생성하기로 한다. 언급한 바와 같이 τ 의 주변사후분포 $p(d\tau)$ 는 편리한 형태가 아니므로 채택-기각기법을 사용하여 생성하고자 하는데 채택-기각기법에서는 봉투함수의 선택이 가장 중요하다 (성내경, 오만숙, 1995). 적절한 봉투함수 선택을 위하여 $p(d\tau)$ 를 그려보니 <그림 2>와 같았다. 이 $p(d\tau)$ 와 모양이 비슷하면서 난수생성이 쉬운 봉투함수로 우리는 여러 시행착오를 거쳐 $N(0, 0.2^2)$ 를 선택하였다. 이 봉투함수를 사용한 기각-채택 기법으로부터 3000개의 τ 난수를 발생시켜 히스토그램을 그려본 결과 $p(d\tau)$ 에 매우 근접함을 알 수 있었다.

이제 생성된 각 τ 가 주어진 조건 하에서 $p(\mu|\tau, y)$ 인 $N(\hat{\mu}, V_\mu)$ 로부터 μ 의 난수를 생성하고 다음 μ, τ 가 주어진 조건 하에서 $p(\theta_j|\mu, \tau, y)$ 인 $N(\hat{\theta}_j, V_j)$ 로부터 θ_j 를 생성하였다.

서로 다른 도시의 로그오즈비의 평균인 θ_j 의 사후분포를 한눈에 알기 쉽게 비교하기 위해 위에서 생성된 θ_j 의 난수를 가지고 상자그림을 그리면 <그림 3>과 같다. 위의 상자그림을 보면 전체적으로 θ_j 는 대칭이 아닌 약간 기울어진 분포를 하고 있고, 범위가 가장 큰 도시는 7번 도시(타이완)이며 가장 작은 도시는 2번 도시(상하이)이다. 그리고, 2번 도시(상하이)와 3번 도시(선양)의 분포가 상당히 집중되어 있는 편이다. 대체로 중앙값은 0.8 내외에 분포한다. 결론적으로 말하자면, 각 도시별로 추정된 θ_j 의 분포에 어느 정도 차이가 있음을 알 수 있다.

좀더 세밀한 분석을 위하여 각 도시별 θ_j 의 사후분위수와 표준편차 (<표 3>) 그리고 비대칭도(skewness)와 첨도(kurtosis) (<표 4>)를 구하였다. <표 3>을 보면 타이완이 제일 많이 퍼져있고, 상하이가 제일 적게 퍼져있음을 알 수 있다. 비대칭도(skewness)를 <표 4>에서 살펴보면 쑹조우의 경우만 빼고 대체로 0에 가까운 값을 가지므로, 대칭모양을 하고는 있지만 약간씩 왼쪽, 오른쪽으로 기울어져있다고 판단할 수 있다. 첨도(kurtosis)를 보면 대체로 정규분포 보다 약간 평평한 모양을 가진다고 판단 할 수 있겠다.

다음으로 초모수 μ 와 τ 에 대한 사후추론을 살펴 보기로 한다. τ 에 대한 사후분포를 보면 <그림 4>에서와 같이 0주위로 값이 몰려있는 절단된 형태를 가지고 있으며 τ 의 사후 중앙값은 0.07116 이다. 비록 τ 의 최빈값이 0이긴 하나 이와 같은 사후분포를 기반으로 하는 베이저안 추

론은 0 이외에 다른 값들도 고려하기 때문에 τ 를 0으로 고정시키는 통합된 (pooled) 방법과는 다른 결과를 준다.

	도시	y_j	σ_j^2	$odds_j$
1	베이징	0.78663	0.0629	2.196
2	상하이	0.76219	0.0058	2.143
3	선양	0.77715	0.0071	2.175
4	난징	1.04744	0.0356	2.850
5	하얼빈	0.8412	0.0186	2.319
6	징조우	0.46245	0.0360	1.588
7	타이완	0.86253	0.1409	2.369
8	난창	0.69475	0.0962	2.003

표 2. 각 도시별 로그오즈비와 분산

종합적인 평균값 μ 의 사후분포를 보면 <표 5>에서 보는 바와 같이 거의 대칭인 모양을 가지고 있다. 이는 각 도시별 θ_j 가 약간 기울어진 형태를 가진 것과 대조적이다. 메타분석에서는 각 도시별 평균 추정치 $\hat{\theta}_j$ 보다 μ 를 추정하는데 더 관심이 있는데 이는 메타분석의 목적이 전체를 분석해보고자 하는 것이기 때문이다.

식 (3)에 있는 $\text{var}(\mu | \tau, y) = V_\mu = \frac{1}{\sum_{j=1}^I \frac{1}{\tau^2 + \sigma_j^2}}$ 에 제곱근을 씌운 $\text{sd}(\mu | \tau, y)$ 을 τ 에 대해

평균을 낸 값 $\text{sd}(\mu | y)$ 는 0.07195 이다.

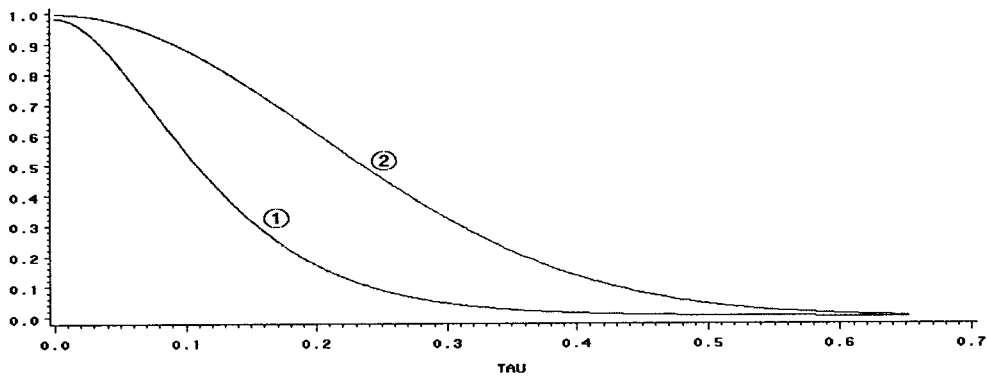


그림 2 : ① $p(\tau|y)$, ② $N(0, 0.04)$

	도시	2.5%	25%	50%	75%	97.5%	표준편차
1	베이징	0.56293	0.721155	0.777998	0.83806	1.02998	0.107694
2	상하이	0.65214	0.732905	0.771519	0.811845	0.88976	0.06118
3	선양	0.65383	0.736181	0.776943	0.820317	0.90278	0.065498
4	난징	0.66522	0.75978	0.815626	0.888345	1.09789	0.109813
5	하얼빈	0.6381	0.743136	0.792564	0.844697	0.98911	0.084841
6	징조우	0.41767	0.659378	0.739464	0.791619	0.89357	0.118558
7	타이완	0.54828	0.726362	0.780388	0.838732	1.04772	0.11952
8	난창	0.51464	0.714904	0.772396	0.827438	1.01141	0.114529

표 3. 각 도시별 θ_j 의 사후분위수와 표준편차

	도시	skewness	kurtosis
1	베이징	0.142624	2.046571
2	상하이	-0.07315	0.278323
3	선양	-0.03108	0.297455
4	난징	0.901144	1.466277
5	하얼빈	0.390352	1.338345
6	징조우	-1.09965	2.008515
7	타이완	0.24988	3.048353
8	난창	-0.20848	2.700613

표 4. 각 도시별 비대칭도(Skewness)와 첨도(Kurtosis)

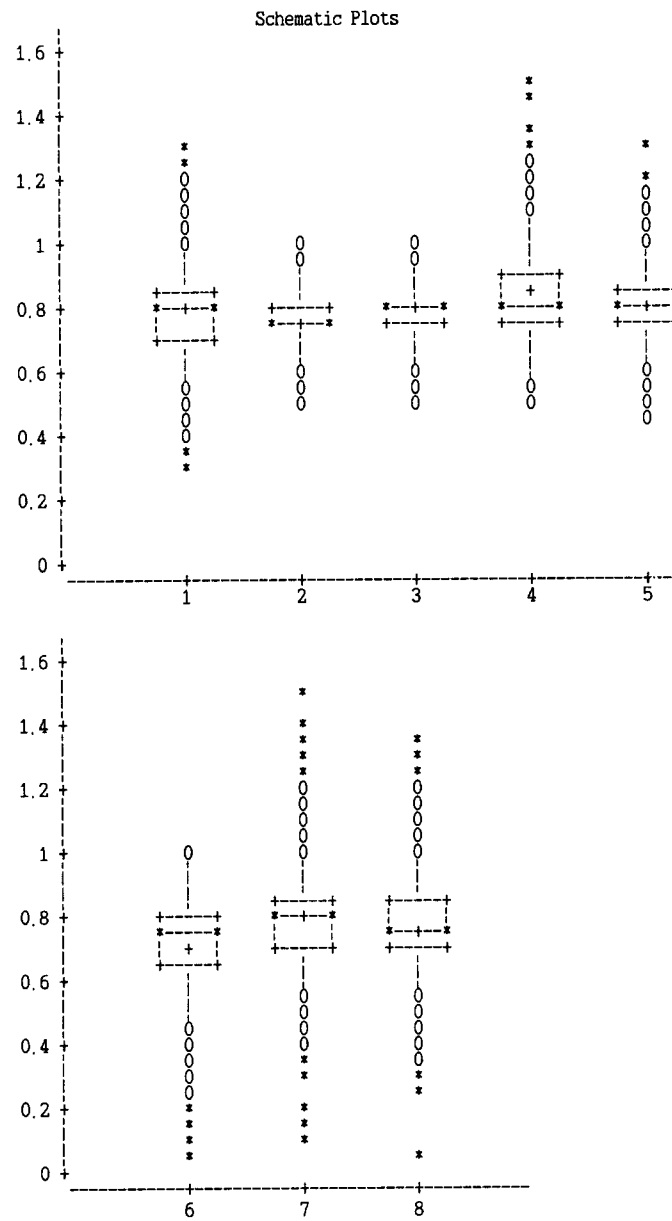


그림 3. 각 도시별 θ_j 의 상자그림

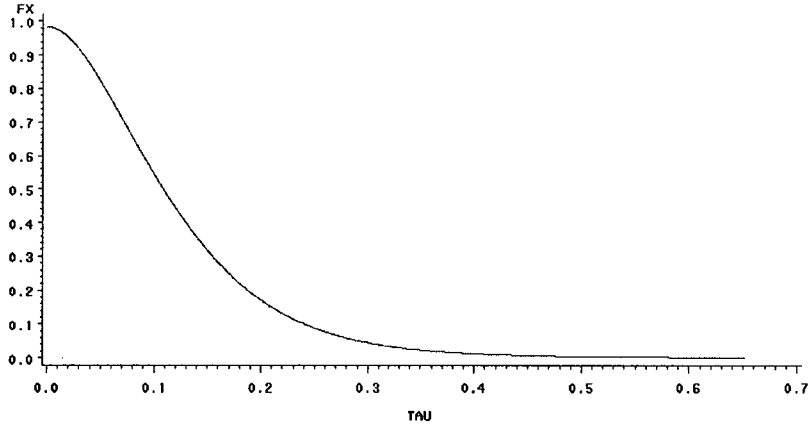


그림 4 : $p(\tau|y)$ 의 그림

<표 5>는 초모수 μ , τ 의 사후분위수값이다. μ 의 95% 사후밀도 구간은 [0.64276, 0.92234]인 반면, $\tau = 0$ 이라는 가정하에서는 식 (5)에서 보듯 완전히 풀링된 추정치가 구해지게 되는데,

$$y_{\cdot} = \frac{\sum_{j=1}^I \frac{y_j}{\sigma_j^2}}{\sum_{j=1}^I \frac{1}{\sigma_j^2}}, \quad \sigma^2 = \frac{1}{\sum_{j=1}^I \frac{1}{\sigma_j^2}} \quad (5)$$

즉, 완전히 풀링된 추정치의 95% 사후구간은 좀 더 좁은 [0.68271, 0.86989]이다. 이런 결과가 나오게 된 이유는 모든 도시가 동일하다는 가정하에서 부적합한 모형을 사용했기 때문이다. 수학적 인 의미에서 완전한 풀링이란 τ 가 정확히 0이라는 가정을 전제하고 있다. 그렇지만, τ 는 불확실한 성질을 가진 모수일 뿐, 어떤 특정한 값이라고 판단하기엔 우리가 가진 자료를 가지고 설명하기엔 역부족이다. τ 가 0에 매우 가까울 것이란 판단은 가능할지 모르나 말이다.

이번엔 μ 에 대해 averaging한 조건부 사후평균 $E(\theta_j | \tau, y)$ 를 그려보았고 <그림 5>와 같다. 다음으로 조건부 사후표준편차 $SD(\theta_j | \tau, y)$ 의 그림을 그리면 <그림 6>과 같다. $\tau = 0$ 일 때는 각 도시별로 $E(\theta_j | \tau, y)$, $SD(\theta_j | \tau, y)$ 는 같은 값을 가지다가 τ 가 1로 커져갈수록 차이가 남을 알 수 있다. 수렴해 가는 값은 앞에서 제시한 y_j , σ_j^2 와 거의 비슷하다. 우리는 여기서 $\tau=0$ 일 때는 모든 도시가 비슷한 양상을 가지다가 τ 가 커짐에 따라 더욱더 확연히 달라지고 있음을 확인할 수 있다. 즉, $\tau \approx 0$ 일 경우에는 $\hat{\theta}_j = y_{\cdot} = 0.7763$, $\hat{\sigma}_j^2 = \sigma^2 = 0.00468$ 로 $\tau \approx 1$ 일수록 $\hat{\theta}_j = y_j$, $\hat{\sigma}_j^2 = \sigma_j^2$ 로 결정할 수 있겠다.

추정치	2.5 %	25 %	median	75%	97.5%
평균, μ	0.64276	0.73572	0.77553	0.81787	0.92234
표준편차, τ	0.00269	0.03257	0.07116	0.13396	0.31555

표 5. μ , τ 의 사후분위수값

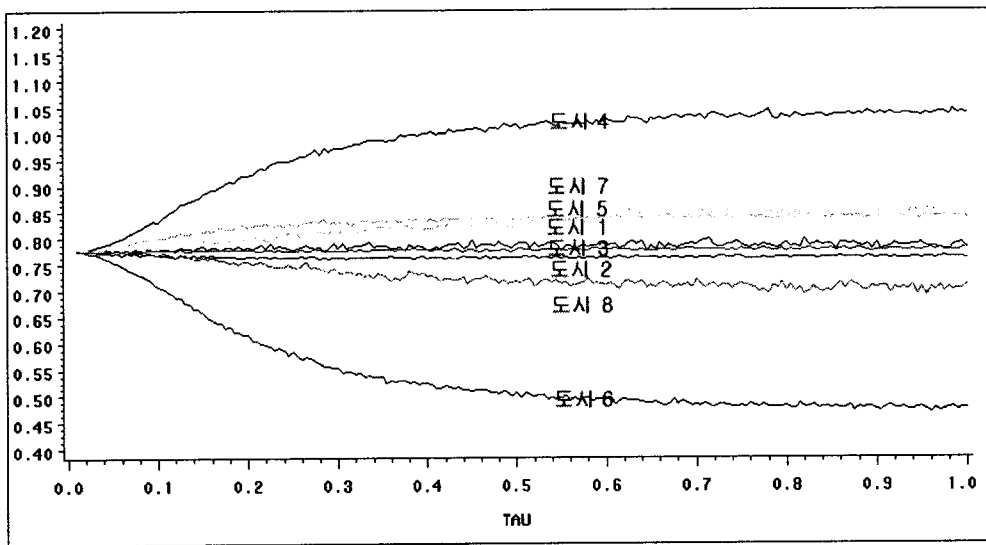


그림 5: $E(\theta_j | \tau, y)$ 의 그림

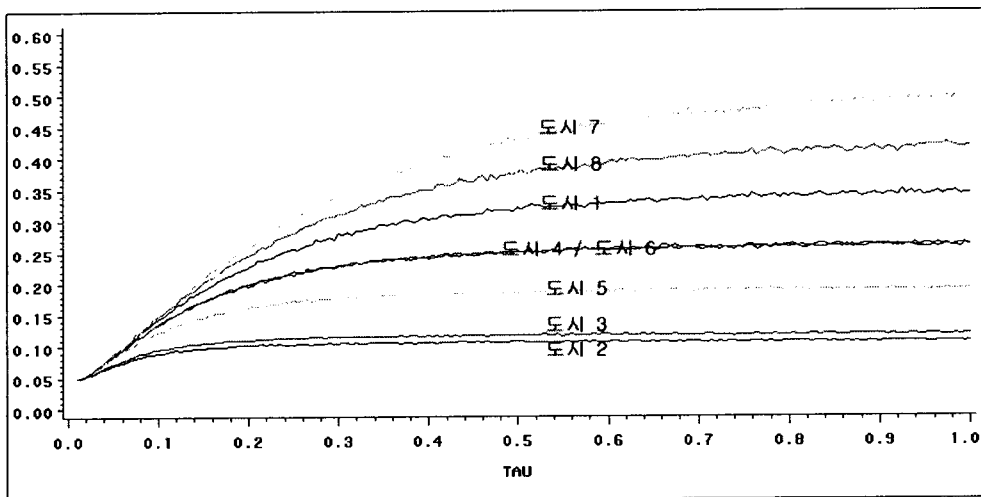


그림 6. $SD(\theta_j | \tau, y)$ 의 그림

이렇게 베이저안 분석은 흥미있는 많은 모수들의 분포에 대해 세밀한 추론을 할 수 있을 뿐 아니라 계층모형을 자료에 적용시키는데 충분히 융통성을 가지고 있으며 초모수에 대한 불확실성을 설명하는데 좀더 사실적인 사후추론을 제공한다. 그렇지만, 이 방법은 초모수의 분포나 모수의 사전정보에 따라 결과가 완전히 달라질 수 있다는 이유로 실제 사용을 꺼리게 만드는 면도 있다. 그러나 불확실한 상황에서 하나의 답을 제시하는 것 보다는 불확실성을 반영하여 융통성을 지니는 해답을 제시하는 것이 더 바람직한 일이라 할 수 있겠다.

참고문헌

- [1] 박태성, 이승연(1998) 「범주형자료분석」. 서울 : 자유아카데미.
- [2] 성내경, 오만숙(1995) 「몬테칼로 적분」. 서울 : 자유아카데미.
- [3] Carlin, J.B. (1992) Meta-analysis for 2x2 tables: a Bayesian approach, *Statistics in Medicine*, Vol. 11, 141-158.
- [4] Dempster, A.P. and Raghunathan, T.E. (1987), Using a covariate for small area estimation: a common sense Bayesian approach. *Small Area Statistics: An International Symposium*, Platek, R., Rao, J.N.K., et. al. editors.
- [5] Fay, R.E. and Harriot, R.A. (1979), Estimates of income for small places: An application of James-Stein procedures to census data, *Journal of the American Statistical Association*, Vol. 74, 269-277.
- [6] Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, Vol. 85, 398-409.
- [7] Gelman, Carlin, Stern, Rubin. (1995) *Bayesian Data Analysis*. London : Chapman & Hall, p.119 - 156.
- [8] Good, I.J. (1965) *The estimation of probabilities: An essay on modern Bayesian methods*, Cambridge, Massachusettes, MIT press.
- [9] Hill, S.E. (1965), Inference about variance components in the one-way model, *Journal of the American Statistical Association*, Vol. 60, 806-825.
- [10] Lindley, D.V. (1971), The estimation of many parameters, *Foundations of Statistical Science*, Godambe, V.P., Sprott, D.A. et. al. editors.
- [11] Liu, Z (1992), Study on smoking and lung cancer, *International Journal of Epidemiology*, Vol. 21, 197-201.
- [12] Meng, C.Y.K. and Dempster, A.P. (1987), A Bayesian approach to multiplicity problem for significance testing with binomial data, *Biometrics*, Vol. 43, 301-311.
- [13] Mollie, A. and Richardson, S. (1991), Empirical Bayes estimates for cancer mortality rates using spatial models, *Statistics in Medicine*, Vol. 10, 95-112.
- [14] Mosteller, F. and Wallace, D.L. (1964), *Applied Bayesian and classical inferences: The case of the federalist papers*, New York, Springer-Verlag.

- [15] Oh, M-S. (1999). Estimation of Posterior Density Functions from a Posterior Sample, *Computational Statistics and Data Analysis*, Vol. 29, 411-427.
- [16] Richardson, S. and Gilks, W.R. (1993), A Bayesian approach to measurement error problems in epidemiology using conditional independence model, *American Journal of Epidemiology*, Vol. 138, 430-442.
- [17] Rubinstein, R.Y. (1981), *Simulation and the Monte Carlo Method*, Wiley, New York.

[2001년 11월 접수, 2002년 2월 채택]