

# Bayesian Analysis and Mapping of Elderly Korean Suicide Rates

Jayoun Lee<sup>a</sup> · Dal Ho Kim<sup>a,1</sup>

<sup>a</sup>Department of Statistics, Kyungpook National University

(Received March 24, 2015; Revised March 31, 2015; Accepted March 31, 2015)

---

## Abstract

Elderly suicide rates tend to be high in Korea. Suicide by the elderly is no longer a personal problem; consequently, further research on risk and regional factors is necessary. Disease mapping in epidemiology estimates spatial patterns for disease risk over a geographical region. In this study, we use a simultaneous conditional autoregressive model for spatial correlations between neighboring areas to estimate standard mortality ratios and mapping. The method is illustrated with cause of death data from 2006 and 2010 to analyze regional patterns of elderly suicide in Korea. By considering spatial correlations, the Bayesian spatial models, mean educational attainment and percentage of the elderly who live alone was the significant regional characteristic for elderly suicide. Gibbs sampling and grid method are used for computation.

Keywords: Bayesian, conditional autoregressive, disease mapping, elderly suicide, Gibbs, small areas.

---

## 1. 서론

한국의 고령화는 매우 빠른 속도로 진행되고 있다. 통계청 조사에 따르면 2011년 국내의 전체인구 4,978만명 중에서 65세 이상의 노인인구가 차지하는 비중은 11.4%이며 2018년에는 노인인구비율이 14.3%로 고령사회에 진입할 예정이다. 따라서 고령화 인구에 대한 사회적 관심이 필요한 실정이다. 자살행위는 전 세계적으로 주요 보건문제가 되고 있다. 세계보건기구(WHO)에 따르면 매년 약 100만명이 자살로 사망하고 있고, 주요한 사망원인으로 자살을 꼽고 있으며, 최근에는 그 자살률이 증가하고 있는 경향을 보이고 있다. 따라서 자살행위는 전 세계적으로 주요 보건문제가 되고 있다. 국내에서는 1989년 십만명 당 7.4명이었던 자살률이 1990년대 초반부터 증가추세를 보이기 시작하여 2010년도에는 십만명당 31.2명에 이를 만큼 급증하였다. 특히 65세 이상 노인 인구는 전체인구의 10% 정도이지만 노인자살자는 전체 자살 인구의 25%를 차지한다. 노인자살은 노인의 주요 사망원인이며, 노인은 다른 연령층보다 자살의 고위험군으로 알려져 있다. 최근의 고령화시대에서 노인인구의 자살은 사회적인 문제로 대두되고 있으며, 잠재적으로 이를 예방하기 위한 사회적 노력이 필요하며, 사회문제로 인식하여 이에 대한 위협요인 파악과 지역적 차이를 확인하는 것이 중요하다. 특히, 노인의 자살은 다른 인구집단에 비해 원인이 다양한 것으로 알려져 있고, 그 중에서도 지역사회와의 통합결여 등이 큰 원인으로 고려되고 있다. Chang 등 (2011)은 대만에서 자살률에 영향을 미치는 사회경제학적 요인 및 지역에 따른 자

---

<sup>1</sup>Corresponding author: Department of Statistics, Kyungpook National University, 80 Daehakro, Bukgu, Daegu 702-701, Korea. E-mail: [dalkim@knu.ac.kr](mailto:dalkim@knu.ac.kr)

살률 패턴을 살펴보았다. 행정단위를 기본으로 강, 산과 같은 지형에 따라 나눈 지역을 기준으로 베이지안 계층적 모형을 이용하여 각 지역의 표준화 사망률을 산출한 결과, 산간지역이 대부분인 대만의 동쪽 지역에서 자살률이 높게 나타났고, 주요도시가 위치하고 있는 서쪽지역에서 상대적으로 자살률이 낮게 나타났다. 특히 공간상관관계를 고려한 베이지안 계층적 모형을 이용하여 SMR을 추정했을 때, 지역적 패턴이 더 뚜렷하게 나타났다. 유럽의 경우 국가별로 자살률이 10배 이상 차이가 난다는 보고가 있고 (Levi 등, 2003), 같은 국가 내에서도 자살률은 지역간에 차이가 나며 지리학적으로 각기 다른 형태로 분포한다 (Gunnell, 2005). 미국의 경우 농촌지역이 더 넓은 네바다 주의 자살률은 미국에서 가장 높은 십만명당 24.5명인데 반해 뉴욕주는 7.6명으로 나타났고 (American Association of Suicidology, 1999) 중국에서 도시지역과 시골지역 사이에서의 자살률을 비교하여 사회문화적 환경이 자살에 미치는 영향을 연구한 바 있다 (Ji 등, 2001). 우리나라에서도 농촌지역이 도시지역보다 자살률이 높다는 보고가 있다 (Seo, 2003).

공간상관관계를 고려한 질병의 상대위험도에 대한 모형으로 Besag (1974)이 제안한 조건부 자기회귀모형(conditional autoregressive; CAR)은 상대위험도에 대한 모형에서 지역효과  $u_i$ 에 대해서 인접지역의 개수를  $n_i$ 라고 할 때 다음을 가정한다.

$$u_i \sim N\left(\bar{u}_i, \frac{\sigma_u^2}{n_i}\right), \quad (1.1)$$

여기에서  $\bar{u}_i = \sum_{j \in \text{neighbors}(i)} w_{ij} u_j / n_i$ 이고  $w_{ij}$ 는  $i$ 번째 지역과  $j$ 번째 지역이 인접지역이면 1, 아니면 0으로 정의되고,  $w_{ii} = 0$ 이다. 인접지역에 대한 정의는 자료에 따라 인접지역으로 정의할 수도 있고, 지역간 거리 등을 이용해서 정의할 수도 있다. 이 논문에서는 Besag (1974)이 제안한 CAR 모형을 확장한 simultaneous CAR model을 고려하고자 한다.

이 논문에서는 국내 노인인구의 지역별 자살의 위험요인 자료를 이용하여 공간상관관계를 고려하지 않은 소지역모형과 공간상관관계를 고려한 소지역모형을 적합해보았다. 2장에서 질병지도 작성을 위한 베이지안 추론 모형 및 계산방법에 대해 설명하였고, 3장에서 국내자료를 이용한 자료분석을 수행하였으며, 4장에서 결론을 제시하였다.

## 2. 질병지도 작성을 위한 베이지안 추론

질병지도에서 작성하는 지역에 따른 상대위험도에 대한 전통적 추정값은 최우추정법에 근거한 표준화 사망률(Standardized Mortality Ratio; SMR)이다. 질병지도를 작성하고자 하는 관심지역의 단위가 작게 나뉘어져 있어 지역별로 질병이 발생한 사람이 적은 경우에 지역별 질병이 발생한 환자수와 기대도수의 비로 산출되는 표준화사망률은 표준오차와 변동계수가 커지는 문제점이 발생할 수 있고, 지역간의 관계에 대해서 고려할 수 없다는 문제가 있다. 따라서 우리는 국내 시군구별 노인자살 자료를 이용하여 공간구조를 고려하지 않은 베이지안 소지역모형과 공간구조를 이용하여 이웃한 지역들의 정보를 빌려올 수 있는 베이지안 소지역모형을 고려하고자 한다.

### 2.1. 계층적 베이지안 소지역모형

지역별 자살자수를  $\mathbf{y} = (y_1, \dots, y_n)$ 라고 하고 모수를  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ 이라고 할 때, 표준화사망률을 추정하고자 한다. 추정된 표준화 사망률은 모수의 추정치를  $\hat{\boldsymbol{\mu}}$ , 지역별 기대도수를  $\mathbf{E}$ 라고 할 때,  $\hat{\boldsymbol{\mu}}/\mathbf{E}$ 로 산출되며, 지역별 기대도수  $\mathbf{E}$ 는 국내 자살률과 지역의 연앙인구를 이용하여 계산한다.  $i$ 번째 지

역의 자살자수를  $y_i$ 라 하고 공변량을  $\mathbf{x}_i$ 라고 할 때, 소지역모형은 식 (2.1)과 같다.

$$y_i | \mu_i \stackrel{ind}{\sim} \text{Poisson}(\mu_i), \quad i = 1, \dots, n,$$

$$\log(\mu_i) = \log(E_i) + \log(\gamma_i) + \mathbf{x}'_i \boldsymbol{\beta}, \quad (2.1)$$

여기에서  $\gamma_i$ 는 소지역 효과이고  $E_i$ 는 각 지역의 기대자살빈도로 주어진 자료에서 산출할 수 있다.  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ 는  $\mathbf{x}_i$ 의 회귀계수이며,  $\mathbf{y} = (y_1, \dots, y_n)$ 는  $n$ 개의 관측치로  $n$ 은 지역의 개수를 의미한다. 베이زي안 추론을 위해서 주어진 모수에 고려되는 사전분포는 다음과 같다.

$$\gamma_i \stackrel{ind}{\sim} \text{Gamma}(\eta, \eta),$$

$$\boldsymbol{\beta} \sim N_p(\hat{\boldsymbol{\beta}}, \sigma^2 \hat{\boldsymbol{\Sigma}}). \quad (2.2)$$

여기에서  $\text{Gamma}(\alpha, \beta)$ 는 감마분포를 나타내며(즉,  $f(x) \propto x^{(\alpha-1)} \exp(-\beta x)$ ),  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)$ 이라 두자. 공변량에 대한 회귀계수인  $\boldsymbol{\beta}$ 에 대한 사전분포의 모수인  $\hat{\boldsymbol{\beta}}$ 는  $\log(\mu_i) = \log(E_i) + \mathbf{x}'_i \boldsymbol{\beta}$ 로 가정했을 때  $\mathbf{y}$ 의 우도함수로 산출된  $\boldsymbol{\beta}$ 의 최대우도추정치(MLE)이고,  $\hat{\boldsymbol{\Sigma}}$ 는 H-행렬(Hessian matrix)의 절댓값으로 다음과 같다.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \text{diag}(E_1, \dots, E_n) \mathbf{X})^{-1} (\mathbf{1}_p \mathbf{X}' (\mathbf{y} - \mathbf{E})), \quad (2.3)$$

$$\hat{\boldsymbol{\Sigma}} = \mathbf{X}' \text{diag}(E_1 e^{\mathbf{x}'_1 \hat{\boldsymbol{\beta}}}, \dots, E_n e^{\mathbf{x}'_n \hat{\boldsymbol{\beta}}}) \mathbf{X}. \quad (2.4)$$

이 모형에서  $(\sigma^2, \eta)$ 에 대한 초사전분포로 서로 독립인  $1/\sigma^2 \sim \chi^2(1)$ 와  $\pi(\eta) = 1/(1 + \eta)^2$ 을 가정하였다. 이에 따른 모수들의 결합사후분포  $\pi(\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma^2, \eta | \mathbf{y})$ 는 식 (2.5)와 같다.

$$\pi(\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma^2, \eta | \mathbf{y}) = \prod_{i=1}^n \pi(y_i | \gamma_i, \boldsymbol{\beta}) \pi(\gamma_i | \eta) \pi(\eta) \pi(\boldsymbol{\beta} | \sigma^2) \pi(\sigma^2). \quad (2.5)$$

다음으로 지리학적 요소가 자살률에 영향을 미칠 수 있다고 판단하여, 지역간의 공간정보를 활용하고자 (2.1) 모형에 공간임의효과를 추가한 소지역공간모형을 고려하였다.

$$y_i | \mu_i \stackrel{ind}{\sim} \text{Poisson}(\mu_i), \quad i = 1, \dots, n,$$

$$\log(\mu_i) = \log(E_i) + u_i + \log(\gamma_i) + \mathbf{x}'_i \boldsymbol{\beta}, \quad (2.6)$$

여기에서  $u_i$ 는 공간임의효과로 simultaneous conditional autoregressive을 사용하였다 (Clayton과 Kaldor, 1987; Besag, 1974).

$$(u_i | u_k, k \neq i) \sim N \left( \phi \sum_{k \neq i} C_{ik} u_k, \delta^2 \right), \quad (2.7)$$

위 모형에서  $\mathbf{C}$ 는  $n \times n$ 의 대칭인 공간상관행렬이다.  $\mathbf{I}$ 를  $n \times n$  단위행렬이라고 하고  $\mathbf{B} = \mathbf{I} - \phi \mathbf{C}$ 라고 하자. 또한  $\lambda_1 \leq \dots \leq \lambda_n$ 을 행렬  $\mathbf{C}$ 의 고유값이라고 하면  $\phi$ 가  $\lambda_1^{-1} < \phi < \lambda_n^{-1} > 0$ 을 만족할 때  $\mathbf{B}$ 는 양정치(positive definite)만족하고  $\mathbf{u}$ 의 결합사후분포는  $\mathbf{u} \sim N_n(0, \delta^2 \mathbf{B}^{-1})$ 가 된다. 또한  $\delta^2$ 와  $\phi$ 에 대해서는 각각 역감마분포와 균일분포로 가정하였다. 특히  $\phi$ 는 균일분포로 공간상관행렬  $\mathbf{C}$ 의 고유값 중 가장 작은 고유값( $\lambda_1$ )과 가장 큰 고유값( $\lambda_n$ )을 이용하여  $\phi \sim U(\lambda_1^{-1}, \lambda_n^{-1})$ 을 가정하였다 (He와 Sun, 2000). 이에 따른 모수들의 결합사후분포  $\pi(\mathbf{u}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \eta, \sigma^2, \delta^2, \phi | \mathbf{y})$ 는 식 (2.8)과 같다.

$$\pi(\mathbf{u}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \eta, \sigma^2, \delta^2, \phi | \mathbf{y}) = \prod_{i=1}^n \pi(y_i | u_i, \gamma_i, \boldsymbol{\beta}) \pi(\mathbf{u} | \delta^2, \phi) \pi(\gamma_i | \eta) \pi(\boldsymbol{\beta} | \sigma^2) \pi(\sigma^2) \pi(\delta^2) \pi(\phi) \pi(\eta). \quad (2.8)$$

## 2.2. 베이지안 계산

두 모형에 대해서 베이지안 추론을 하기 위해서 깃스 표집기(Gibbs sampler)를 이용하였다. 각 모수에 대한 조건부 사후분포(full conditional distribution)는 부록에 제시하였다. 소지역모형의 모수 중  $(\gamma, \sigma^2)$ 과 소지역 공간모형의 모수 중  $(\gamma, \mathbf{u}, \sigma^2, \delta^2)$ 는 알려져 있는 분포를 따르므로 깃스 표집기를 통해 난수를 생성할 수 있지만, 나머지 모수들의 경우 분포가 닫힌 형태로 나타나지 않아 다른 방법을 이용하여 난수를 생성하였다.  $\beta$ 는 조건부 사후분포에서 테일러(Taylor) 전개( $e^{\mathbf{x}'_i \beta} \approx (1 + \mathbf{x}'_i \beta + (\mathbf{x}'_i \beta)^2)$ )를 이용하여 계산된 정규분포를 제안분포(proposal distribution)로 한 메트로폴리스 헤스팅스(Metropolis-Hastings) 알고리즘을 이용하였다. 각 모형에서 계산된 제안분포는 다음과 같다.

$$N_p \left( \left( \mathbf{X}' \mathbf{A} \mathbf{X} + \frac{1}{\sigma^2} \hat{\Sigma}^{-1} \right)^{-1} \left( \mathbf{X}' (\mathbf{y} - \mathbf{a}) + \frac{1}{\sigma^2} \hat{\Sigma}^{-1} \hat{\beta} \right), \left( \mathbf{X}' \mathbf{A} \mathbf{X} + \frac{1}{\sigma^2} \hat{\Sigma}^{-1} \right)^{-1} \right). \quad (2.9)$$

제안분포는 첫 번째 모형에서는  $\mathbf{a} = (E_1 \gamma_1, \dots, E_n \gamma_1)'$ ,  $\mathbf{A} = \text{diag}(E_1 \gamma_1, \dots, E_n \gamma_1)$ 이고, 두 번째 모형에서는  $\mathbf{a} = (E_1 \gamma_1 e^{u_1}, \dots, E_n \gamma_1 e^{u_n})'$ ,  $\mathbf{A} = \text{diag}(E_1 \gamma_1 e^{u_1}, \dots, E_n \gamma_1 e^{u_n})$ 이다. 모형 1의  $\eta$ 와 모형 2의  $(\eta, \phi)$ 는 범위가 정해져있기 때문에 격자망 방법(grid method)을 고려한다.  $\eta$ 의 범위는  $\eta > 0$ 이지만  $w = 1/(1 + \eta)$ 로 치환하게 되면  $0 < w < 1$ 을 만족하므로,  $w$ 에 대해서도 격자망 방법을 고려하였다.

우선,  $\pi(\eta|\gamma)$ 를 치환한  $\pi(w|\gamma)$ 로 부터  $w(0 < w < 1)$ 을 생성하기 위하여 다음 절차를 따르는 격자망 방법을 고려하도록 한다.

1. 0에서 1사이 구간을 100개의 구간으로 분할하고, 각 구간의 경계점을  $I_k, k = 1, \dots, 101$ 이라 하자.
2. 각 구간의 중간값을 계산한다( $M_k, k = 1, \dots, 100$ ).
3. 각 중간값( $M_k$ )에 대한 조건부분포 함수값을 구한다( $a_k = \pi(M_k|\gamma), k = 1, \dots, 100$ ).
4.  $A = \sum_{k=1}^{100} a_k$ 라고 할 때,  $b_k = \sum_{i=1}^k a_i/A$ 를 계산한다.
5.  $U(0, 1)$ 로 부터 난수  $u_1$ 을 생성한다.
6. 1.에서 구한 구간 중에서  $b_k \leq u_1 < b_{k+1}$ 을 만족하는  $k$ 번째 구간을 선택한다.
7.  $U(I_k, I_{k+1})$ 로 부터  $u_2$ 를 생성하여  $w$ 라 한다.

비슷한 방법으로 소지역 공간모형에서  $\pi(\phi|\mathbf{u}, \delta^2)$ 로부터  $\phi$ 를 생성할 수 있다. 단,  $\phi$ 는  $\lambda_1^{-1}$ 에서  $\lambda_n^{-1}$ 사이 구간을 적용한다.

소지역 공간모형의  $\mathbf{u}$ 는 조건부사후분포가 정규분포이지만, 깃스 표집기에서 반복할 때마다  $\mathbf{B}$ 에 대한 역행렬을 계산해야하기 때문에 자료의 수가 많아질 경우 계산이 많아진다는 단점이 있다. 따라서 He와 Sun (2000)이 제안한 방법에 따라  $\mathbf{u}$ 와  $\delta^2$ 를 계산하였다. 깃스 표집기 내에서  $\mathbf{B}$ 에 대한 역행렬을 계산하는 대신, 공간상관행렬  $\mathbf{C}$ 을 행렬분해하여 구성되는 고유값( $\lambda$ )과 고유벡터( $\mathbf{G}$ )를 이용한다.  $\mathbf{u}$ 와  $\delta^2$ 를 계산하는 알고리즘은 다음과 같다.

1.  $\rho_i = \delta^2/(1 - \phi \lambda_i)$ 라고 할 때,  $\Psi = \text{diag}(\rho_1, \dots, \rho_n)$ 를 계산한다.
2.  $N_n(\mathbf{0}, \mathbf{I})$ 으로부터 난수  $\mathbf{K}$ 를 생성하여  $\mathbf{s} = (s_1, \dots, s_n)^T = \Psi \mathbf{G}^T \mathbf{K}$ 를 계산한다.
3.  $\mathbf{u} = \mathbf{G} \mathbf{s}$ 로 정의한다.
4.  $\text{IG}(n/2 + c, 1/2 \sum (1 - \phi \lambda_i) s_i^2 + d)$ 로부터  $\delta^2$ 를 생성한다.

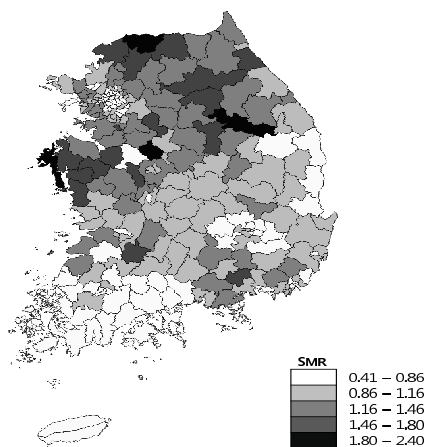


Figure 3.1. Standardized Mortality Ratio of the elderly suicide

### 3. 자료분석

국내의 시군구별 노인 자살률을 산출하기 위해서 소지역모형과 소지역 공간모형을 적합해보았다. 2006~2010년(5개년도) 통계청 사망자료를 이용하여 질병사인분류가 고의적 자해(자살)인 만 65세 이상의 노인을 대상으로 시군구별 자살자수를 산출하였으며, 2008년 통계청 주민등록 연앙인구자료를 이용하여 기대자살자수를 산출하였다. 공변량으로 시군구별 교육수준과 독거노인비율을 고려하였다. 시군구별 교육수준은 2008년 지역사회건강조사 자료를 이용하였으며, 시군구별 독거노인 비율은 인구주택총조사 5% 표본자료를 활용하였다. 2006년부터 2010년까지의 국내 행정구역이 변화함에 따라 2007년을 기준으로 총 248개의 시군구 행정구역으로 정리하였다. 전체 248개 지역 중 도서지역 등 인접지역이 없는 6개 지역을 제외한 242개 시군구를 분석대상으로 하였다. 2006년부터 2010년까지 5년 동안 평균 표준화 사망률(SMR)을 지도에 나타내면 Figure 3.1과 같다.

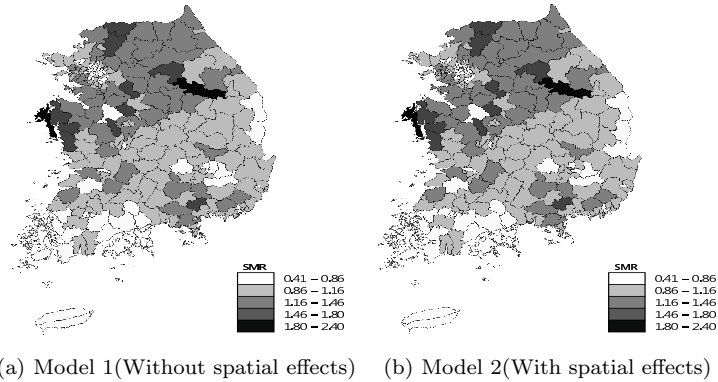
그림에서 살펴보면, 강원도 지역의 표준화사망률이 전반적으로 높고, 전라도 지역의 표준화 사망률이 전반적으로 낮게 나타난 것을 확인할 수 있다. 또한, 서울 등 대도시 주변의 경우 표준화사망률이 낮게 나타났다. 시군구별 교육수준과 독거노인 비율을 공변량으로 고려했을 때,  $\beta$ 에 대한 사전분포의 평균과 분산은 다음과 같다. 즉, 시군구별 교육수준과 독거노인비율에 대한 회귀계수 최대우도추정치(MLE)는 각각  $-0.0213, 0.4668$ 로 교육수준이 낮을수록, 독거노인의 비율이 높을수록 노인자살률이 높은 것으로 나타났다.

$$\hat{\beta} = (-0.0213, 0.4668), \tag{3.1}$$

$$\hat{\Sigma} = \begin{pmatrix} 0.000005916016 & -0.000119828 \\ -0.000119828 & 0.003323877 \end{pmatrix}.$$

공간구조를 활용한 두번째 모형에서 시군구의 공간구조를 나타내는 행렬  $C$ 의 고유값 중 최소값( $\lambda_1$ )과 최대값( $\lambda_n$ )은 각각  $-3.192, 6.142$ 였다. 따라서  $\phi$ 의 사전분포는  $U(-0.313, 0.162)$ 가 된다.

마코브 연쇄 몬테칼로 계산(MCMC)에서 부록에 제시한 조건부 사후분포 및 3절에서 소개한 알고리즘에 따라 난수를 추출하였다. 이 때 2개의 다중체인을 고려하였으며, 각 체인에서 10,000개의 난수를 추출하며 초기값의 영향을 제거하기 위해서 초기 5,000개의 난수를 제거한 후 나머지 5,000개의 난수를



**Figure 3.2.** Estimated SMR of the elderly suicide

**Table 3.1.** Bayesian estimates and HPD credible intervals

Parameter	Model 1				Model 2			
	Mean	SD	95% HPD		Mean	SD	95% HPD	
$\beta_1$	-0.0204	0.0029	-0.0258	-0.0139	-0.0203	0.003	-0.026	-0.014
$\beta_2$	0.4624	0.0652	0.3139	0.5912	0.4618	0.068	0.310	0.597
$\sigma^2$	3.9871	19.5177	0.0675	12.4631	4.4416	31.559	0.069	13.276
$\eta$	19.5934	3.3526	13.1137	26.0877	19.5513	3.358	13.114	26.000
$\delta^2$					0.0000833	0.0000076	0.0000690	0.0000985
$\phi$					-0.0077	0.038	-0.084	0.064

이용하여 모수를 추정하였다. 여기에서 공간정보를 활용하지 않은 모형(소지역모형)과 공간정보를 활용한 모형(소지역 공간모형)을 비교하고자 한다. 소지역모형과 소지역 공간모형에서 추정된 회귀계수는 교육수준이 각각  $-0.0204$ ,  $-0.0203$ 으로 교육수준이 높을수록 지역의 자살율이 낮은 것으로 나타났고, 독거노인 비율은 각각  $0.4624$ ,  $0.4618$ 로 독거노인 비율이 높을수록 지역의 자살률이 높은 관계를 가지는 것으로 나타났다. 또한 각 모형에서 추정된 표준화사망률을 이용하여 질병지도를 작성한 결과, 평활화된 표준화사망률을 확인할 수 있고, 공간상관관계를 고려했을 때 주변 지역의 정보를 빌려옴으로써, 포항시 북구, 대구 동구 등의 지역에서 표준화사망률이 달라진 것을 확인할 수 있었다 (Figure 3.2).

우리가 고려한 모형을 비교하기 위해 세 가지 척도를 비교해보았다. 첫 번째는 Waller 등 (1997)이 제안한 EPD(posterior expected predictive deviance)이다.  $\mathbf{y}^{rep}$ 는 MCMC 계산에서 반복적으로 계산된 값을 의미하며, 다음 식에 따라 임의 추출된  $\mathbf{y}^{rep}$ 를 이용하여 계산되며,  $k$ 는 반복횟수를 의미한다 ( $k = 1, \dots, \#of\ replication$ ).

$$P(\mathbf{y}^{rep}|\mathbf{y}^{obs}) = \int g(\mathbf{y}^{rep}|\boldsymbol{\mu})h(\boldsymbol{\mu}|\mathbf{y}^{obs})d\boldsymbol{\mu}. \quad (3.2)$$

모형선택문제에서 차이를 나타내는 함수  $d(\mathbf{y}^{rep}, \mathbf{y}^{obs})$ 를 정의하여 다음을 계산하여 값이 가장 작은 모형을 선택하게 된다.

$$E[d(\mathbf{y}^{rep}, \mathbf{y}^{obs})|\mathbf{y}^{obs}], \quad (3.3)$$

여기에서는 EPD를 계산하기 위한 함수  $d$ 를 카이제곱과 포아송 기준으로 두 가지를 고려하였다. 카이

제공의 경우

$$d_1(\mathbf{y}^{obs}, \mathbf{y}^{rep}) = \sum_k \frac{(y_i^{obs} - y_{i,k}^{rep})^2}{(y_{i,k}^{rep} + 0.5)} \tag{3.4}$$

이고, 포아송분포를 가정하는 경우는  $d_2(\mathbf{y}^{obs}, \mathbf{y}^{rep}) = 2 \sum_k \{y_i^{obs} \log(y_i^{obs}/y_{i,k}^{rep}) - (y_i^{obs} - y_{i,k}^{rep})\}$ 이다. 하지만, 이 식에서  $y_i^{obs}$ 가 0인 경우는 계산할 수 없기 때문에, 실제로 고려한 식은 다음과 같다.

$$d_2(\mathbf{y}^{obs}, \mathbf{y}^{rep}) = 2 \sum_k \left\{ (y_i^{obs} + 0.5) \log \left( \frac{y_i^{obs} + 0.5}{y_{i,k}^{rep} + 0.5} \right) - (y_i^{obs} - y_{i,k}^{rep}) \right\}. \tag{3.5}$$

두 번째로 Gelfand와 Ghosh (1998)가 제안한 PPL(minimum posterior predictive loss approach)을 확인해보았다.  $PPL = G_1 + P_1$ 으로 구성되며,  $G_1$ 과  $P_1$ 은 다음과 같이 정의된다.

$$G_1 = \sum_k \left\{ E(y_{i,k}^{rep} | \mathbf{y}^{obs}) - y_i^{obs} \right\}^2, \quad P_1 = \sum_k \text{var}(y_{i,k}^{rep} | \mathbf{y}^{obs}), \tag{3.6}$$

여기에서  $G_1$ 과  $P_1$ 은 식 (3.2)를 이용해 계산되며,  $G_1$ 은 모형적합도를 의미하며,  $P_1$ 은 벌점항(penalty term)으로, 만약 모형이 과소추정 또는 과대추정하게 되면 위 값은 커지게 되므로 값이 작을수록 좋은 모형임을 의미한다.

Gelfand 등 (1992)은 cross-validatory(또는 “leave-one-out”) 방법인 CPO(conditional predictive ordinate)를 제안하였다. 베이시안 질병지도 모형에서 leave-one-out cross-validatory density는 다음과 같다.

$$P(Y = y | \mathbf{y}_{(i)}) = \int f(Y = y | \mathbf{y}_{(i)}, \Omega) \pi(\Omega | \mathbf{y}_{(i)}) d\Omega, \tag{3.7}$$

여기에서  $\mathbf{y}_{(i)}$ 는  $i$ 번째 관측치를 제외한 것을 의미하며,  $\Omega = (\boldsymbol{\mu}, \sigma^2, \eta)$ 이다. MCMC 계산결과를 이용하여 CPO를 계산하는 식은 다음과 같다.

$$\widehat{CPO}_i = \sum_{h=1}^H w_h P(Y = y | \Omega^{(h)}), \quad i = 1, \dots, n, \tag{3.8}$$

$$w_h = \frac{1}{f(Y = y | \Omega^{(h)})} / \sum_{h=1}^H \frac{1}{f(Y = y | \Omega^{(h)})}, \quad h = 1, \dots, H.$$

따라서, CPO 값이 클수록 모형이 더 좋음을 의미한다. “pseudo marginal likelihood”로 CPO 값의 곱을 살펴보기도 하는데, 우리는 계산의 편의를 위해 로그함수를 이용하여 로그(CPO)의 평균을 하였다. 따라서 LCPO는 값이 작을수록 좋은 모형임을 의미한다.

$$\overline{LCPO} = -\frac{1}{n} \sum_{i=1}^n \log(\widehat{CPO}_i). \tag{3.9}$$

우리가 고려한 자료를 이용하여 소지역모형과 소지역 공간모형을 비교했을 때 위 세 가지 척도에 대한 결과는 Table 3.2와 같다. 결과적으로 공간상관관계를 고려한 소지역 공간모형이 잘 적합하는 것으로 나타났다.

**Table 3.2.** Model comparison for two models

	LCPO	EPD(Chi-square)	EPD(Poisson)	PPL
Model 1	164.1	523.7	508.2	23381.4
Model 2	163.5	523.2	507.9	23361.0

#### 4. 결론

본 논문에서는 국내 노인사살자료에 대해 추정된 표준화 사망률로 질병지도를 작성하였다. 표준화사망률을 추정하기 위해서 공간상관관계를 고려하지 않은 소지역 모형과 공간상관관계를 고려한 소지역 공간모형을 고려하였다. 특히, 소지역 공간모형에서는 이웃한 지역들의 정보를 빌려올 수 있는 simultaneous CAR model을 이용한 베이지안 추론을 고려하였다. 그 결과, 소지역 공간모형이 자료에 더 잘 적합하는 것으로 나타났으며 표준화사망률이 매우 높거나 낮은 지역에서 주변 지역의 영향을 받아 주변 지역 간에 표준화사망률의 변동이 줄어든 것을 확인할 수 있었다. 또한 소지역 공간모형에 대해서 He와 Sun (2000)이 제안한 알고리즘을 사용함으로써 효율적인 베이지안 추론을 할 수 있었다.

#### 부록 A: MLE and H-matrix of $\beta$

$\beta$ 의 사전분포는 우도함수를 다음과 같이 가정했을 때  $\beta$ 의 최대우도추정치와 H-matrix를 이용하였다.

$$L(\beta) \propto \prod_{i=1}^n \exp\{-E_i e^{\mathbf{x}'_i \beta}\} E_i^{y_i} \exp\{\mathbf{x}'_i \beta y_i\}, \quad (\text{A.1})$$

$$\log L(\beta) = -\sum_{i=1}^n E_i e^{\mathbf{x}'_i \beta} + \sum_{i=1}^n \mathbf{x}'_i \beta y_i, \quad (\text{A.2})$$

$$\frac{\partial}{\partial \beta} \log L(\beta) = -\sum_{i=1}^n E_i \mathbf{x}'_i e^{\mathbf{x}'_i \beta} + \sum_{i=1}^n \mathbf{x}'_i y_i = 0 \quad (\text{A.3})$$

$$\Leftrightarrow \sum_{i=1}^n E_i \mathbf{x}'_i (1 + \mathbf{x}'_i \beta) = \sum_{i=1}^n \mathbf{x}'_i y_i \quad (\text{by Taylor series}), \quad (\text{A.4})$$

$$\frac{\partial^2}{\partial \beta^2} \log L(\beta) = -\sum_{i=1}^n E_i \mathbf{x}_i \mathbf{x}'_i e^{\mathbf{x}'_i \beta} = -\mathbf{X}' \text{diag}(E_1 e^{\mathbf{x}'_1 \beta}, \dots, E_n e^{\mathbf{x}'_n \beta}) \mathbf{X}, \quad (\text{A.5})$$

$$\hat{\beta}^{\text{MLE}} = (\mathbf{X}' \text{diag}(E_1, \dots, E_n) \mathbf{X})^{-1} (\mathbf{1}_p \mathbf{X}' \mathbf{y} - \mathbf{X}' \mathbf{E}), \quad (\text{A.6})$$

$$\hat{\Sigma}^{-1} = \mathbf{X}' \text{diag}(E_1 e^{\mathbf{x}'_1 \beta}, \dots, E_n e^{\mathbf{x}'_n \beta}) \mathbf{X}. \quad (\text{A.7})$$

#### 부록 B: 조건부 사후분포

소지역 공간모형에 대해서 깃스 표집기를 사용하기 위한 조건부사후분포이다.

$$(a) [\mathbf{u} | \gamma, \beta, \delta^2, \phi] \sim N_n(\delta^2 \mathbf{B}^{-1}(\mathbf{y} - \mathbf{A}), \delta^2 \mathbf{B}^{-1}), \quad \text{where } \mathbf{A} = (E_1 \gamma_1 e^{\mathbf{x}'_1 \beta}, \dots, E_n \gamma_n e^{\mathbf{x}'_n \beta})^T. \quad (\text{B.1})$$

$$(b) [\gamma_i | y_i, \beta, \eta] \stackrel{\text{ind}}{\sim} \text{Gamma}(y_i + \eta, E_i e^{\mathbf{x}'_i \beta} + \eta). \quad (\text{B.2})$$

$$(c) [\delta^2 | \mathbf{u}, \phi] \stackrel{\text{ind}}{\sim} \text{IG}\left(\frac{n}{2} + c, \frac{1}{2} \mathbf{u}' \mathbf{B} \mathbf{u} + d\right). \quad (\text{B.3})$$



$$(d) [\beta|\gamma, \beta, \sigma^2] \propto \exp \left\{ - \sum_{i=1}^n E_i \gamma_i e^{\mathbf{x}'_i \beta + u_i} + \sum_{i=1}^n \mathbf{x}'_i \beta y_i - \frac{1}{2\sigma^2} (\beta - \hat{\beta})' \hat{\Sigma}^{-1} (\beta - \hat{\beta}) \right\}. \quad (B.4)$$

$$(e) [\sigma^2|\beta] \sim \text{IG} \left( \frac{p+1}{2}, \frac{1}{2} \left\{ (\beta - \hat{\beta})' \hat{\Sigma}^{-1} (\beta - \hat{\beta}) + 1 \right\} \right). \quad (B.5)$$

$$(f) [\phi|\mathbf{u}, \delta^2] \propto |(\mathbf{I} - \phi\mathbf{C})^{-1}|^{-\frac{1}{2}} \exp \left\{ - \frac{1}{2\delta^2} \mathbf{u}' (\mathbf{I} - \phi\mathbf{C}) \mathbf{u} \right\}, \quad \text{for } \phi \in (\lambda_1^{-1}, \lambda_n^{-1}). \quad (B.6)$$

$$(g) [\eta|\gamma] \propto \left[ g_1 \left(\frac{1}{w}\right)^{-2} \exp \left\{ \left(\frac{1}{w} - 1\right) g_0 \right\} \right]^n, \quad \text{where } g_0 = \frac{\sum_{i=1}^n \gamma_i}{n}; \quad g_1 = \sqrt[n]{\prod_{i=1}^n \gamma_i}. \quad (B.7)$$

## References

- American Association of Suicidology (1999). *Suicide Prevention Facilitator's Training Manual*, Final report to the Department of the Navy, American Association of Suicidology, Washington, DC.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion), *Journal of the Royal Statistical Society, Series B*, **36**, 192–236.
- Chang, S. S., Sterne, J. A., Wheeler, B. W., Lu, T. H., Lin, J. J. and Gunnell, D. (2011). Geography of suicide in Taiwan: Spatial patterning and socioeconomic correlates. *Health Place*, **17**, 641–650.
- Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping, *Biometrics*, **43**, 671–681.
- Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based method (with discussion), In *Bayesian Statistics 4*, edited by J.M. Bernardo, et al., Oxford University Press, Oxford, 147–167.
- Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach, *Biometrika*, **85**, 1–11.
- Gunnell, D. (2005). Time trends and geographic differences in suicide: Implications for prevention. In: Hawton, K. (Ed.). *Prevention and Treatment of Suicidal Behaviour: From Science to Practice*, Oxford University Press, 293–306.
- He, X. and Sun, D. (2000). Hierarchical Bay estimation of hunting success rates with spatial correlation, *Biometrics*, **56**, 360–367.
- Ji, J., Kleinman, A., and Becker, A. E. (2001). Suicide in contemporary China: A review of China's distinctive suicide demographics in their sociocultural context, *Harvard Review of Psychiatry*, **9**, 1–12.
- Levi, F., La Vecchia, C., Lucchini, F., Negri, E., Saxena, S., Maulik, P. K., and Saraceno, B. (2003). Trends in mortality from suicide, 1965–1999. *Acta Psychiatrica Scandinavica*, **108**, 341–349.
- Seo, D. W. (2003). Status of suicide rates and role of the Mental Health Center in Korea, *2nd Stakeholder's Workshop by 2003 Community Mental Health Services Project*.
- Waller, A. L., Carlin, B. P., Xia, H. and Gelfand, A. E. (1997). Hierarchical Spatio-Temporal Mapping of Disease Rates, *Journal of the American Statistical Association*, **92**, 607–617.

# 베이지안 모형을 활용한 국내 노인 자살률 질병지도

이자연<sup>a</sup> · 김달호<sup>a,1</sup>

<sup>a</sup>경북대학교 통계학과

(2015년 3월 24일 접수, 2015년 3월 31일 수정, 2015년 3월 31일 채택)

---

## 요약

한국의 고령화는 매우 빠른 속도로 진행되고 있고, 노인자살은 노인의 주요 사망원인이며 노인은 다른 연령층보다 자살의 고위험군으로 알려져있다. 고령화 시대에서 노인의 자살은 사회적인 문제로 대두되고 있으며 이를 예방하기 위해 노인자살에 대한 위험요인을 파악하고, 지역적 차이를 확인하는 것이 중요하다. 특히 노인의 자살문제에서는 지역사회의 통합결여 등이 큰 원인으로 고려되기 때문이다. 따라서, 본 논문에서는 공간적 상관관계를 고려하여 추정된 표준화사망률을 이용하여 질병지도를 작성하고자 하였다. 공간적 상관관계를 고려하기 위해서 simultaneous CAR model을 사용하였다. 2006년부터 2010년까지 통계청 사망자료를 이용하여 국내 시군구별 노인자살자수에 대해 두 모형을 적합시켜본 결과, 공간적 상관관계를 고려하지 않은 모형보다 공간적 상관관계를 고려한 모형이 더 좋은 모형임을 보였다. 또한 효율적인 베이지안 추론을 위해 격자망 방법 등을 고려하였다.

주요용어: 노인자살, 깃스, 베이지안, 소지역, 조건부 자기회귀, 질병지도.

---

<sup>1</sup>교신저자: (702-701) 대구광역시 북구 대학로 80, 경북대학교 통계학과. E-mail: dalkim@knu.ac.kr