베이지안 방법에 의한 K개 지수분포 모수들의 기하평균 추정에 관한 연구

김대황[1], 김혜중[2]

요 약

본 연구는 k개 지수분포 모수들의 기하평균에 대한 베이지안추정 방법을 제시하였다. 이를 위해 Tibshirani가 제안한 직교변환법으로 비정보적 사전확률분포를 도출하여 모수들의 결합사후확률분포를 유도해 내었으며, 이 분포 하에서 가중 몬테칼로 방법을 사용하여 기하평균을 추정하는 절차를 제안하였다. 모의실험과 실제자료의 예를 통해 제안된 베이지안 추정의 유효성 및 효용성을 보였으며, 본 연구에서 제안한 사전확률분포가 전통적인 포함확률을 기준으로 볼 때, Jeffrey의 사전확률분포 보다 더 유효한 추정을 함을 보였다.

주요용어 : 베이지안추정, 지수분포, 기하평균, 비정보적 사전확률분포, 가중몬테칼로 방법.

1. 서론

$X_1, X_2, \cdots, X_k$은 서로 독립이고, 평균이 각각 $\theta_1, \theta_2, \cdots, \theta_k$ 인 지수분포를 따르는 확률변수들이라 하자. 여기서 추정하고자 하는 모수는 $\theta_1, \theta_2, \cdots, \theta_k$의 산술평균 $\overline{\theta}(=1/k\sum_{i=1}^{k}\theta_i)$의 하계(lower bound)인 기하평균(geometric mean) $\delta=\left(\prod_{i=1}^{k}\theta_i\right)^{1/k}$이다 . 기하평균은 경영이나 경제학에서 주로 사용하는 위치측도로서 산술평균과는 달리 경제성장률, 물가상승률과 같은 변화율의 평균을 계산하는데 사용된다(Kenneth, John 와 Kenneth, 1998참조). 베이지안 방법으로 기하평균을 추정함에 있어 , 모수들에 대한 사전정보가 적거나 없을 때 그들의 사전확률분포 설정에 어려움이 있다. 일반적으로 사용되는 균등사전확률분포는 일대일 재모수화에 대해 불변성을 만족하지 못하는 것으로 알려져, 이를 이용하여 모수들의 기하평균을 추론하는데는 문제점이 있다. 이와 같은 불변성의 부제문제를 해결하기 위해 여러 연구가 진행되었다(Jeffrey, 1961 ; Bernaldo, 1979 ; Berger와 Bernaldo, 1989 ; Tibshirani, 1989참조). 이들 중 Tibshirani(1989)는 Stein(1985)의 연구에 기초를 두고 모수의 직교화 방법을 이용하여 전통적 신뢰구간 성질을 점근적으로 만족하는 비정보적 사전확률분포를 유도하는 방법을 제안하였다.

본 논문에서는 Tibshirani(1989)의 방법에 의해 얻은 θ_i들의 사전확률분포를 이용하여 모수들의 결합사후확률분포를 유도하고, 이로부터 θ_i들의 기하평균을 베이지안 추정하는 방법을 제안하고자 한다. 이를 위하여 직교모수화를 통한 사전확률분포 유도과정을 설명하고, θ_i들의 사후확률분포에 가중몬테칼로방법을 적용하여 기하평균을 추정하는 절차를 제안하였다. 또한 모의실험을 통해 제안된 추정방법의 유효성 및 효용성을 보였으며, 전통적인 포함확률을 기준

1) 동국대학교 통계학과 박사과정, 서울시 중구 필동 3가 26번지
2) 동국대학교 통계학과 교수 서울시 중구 필동 3가 26번지

으로 하여 본 연구에서 기하평균 추정을 위해 제안한 사전확률분포의 적절성도 함께 보였다.

2. 사전분포와 사후분포의 유도

Tibshirani(1989)는 모수의 직교화 방법을 이용하여 전통적 신뢰구간 성질을 점근적으로 만족하도록 하는 비정보적 사전확률분포를 유도하는 방법을 제안하였다. 이 방법은 관심모수와 장애모수를 지닌 Fisher의 기대 정보행렬에서 이들의 비대각원소들이 모두 0이 되도록 하는 직교모수화의 방법(Cox and Reid, 1987참조)을 통해 관심모수의 사전확률분포를 유도하는 것이다. 이 방법을 지수분포 모수들의 사전확률분포 도출에 적용하면 다음과 같다.

x_{ij}, $i=1,2,\cdots,k$, $j=1,2,\cdots,n$,는 평균이 θ_i인 지수분포로부터의 표본이라고 하자. 이 때 우도함수는

$$L(\boldsymbol{\theta} \mid D) = \prod_{j=1}^{n}\left(\frac{1}{\left(\prod_{i=1}^{k}\theta_i\right)} \exp\left\{-\sum_{i=1}^{k}\frac{x_i}{\theta_i}\right\}\right) \tag{2.1}$$

이다. 여기서, $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)'$ 이고, $D = (x_1, x_2, \cdots, x_n)'$ 이다.

관심모수 δ와 장애모수 ζ를 각각 다음과 정의하자.

$$\delta = \left(\prod_{i=1}^{k}\theta_i\right)^{1/k}, \quad \zeta_i = \zeta_i(\boldsymbol{\theta}), \quad i = 2,3,\cdots,k$$

그리고, $\zeta_i^j = \partial\zeta_i(\boldsymbol{\theta})/\partial\theta_i$, $\eta_{(i)} = 1/k\left(\prod_{j=1}^{n}\theta\right)^{1/k}\theta_i^{-1}$라고 가정하면 Jacobian행렬은 다음과 같다.

$$\frac{\partial(\delta, \boldsymbol{\zeta})}{\partial(\boldsymbol{\theta})} = \begin{pmatrix} \eta_{(1)} & \eta_{(2)} & \cdots & \eta_{(k)} \\ \zeta_2^1 & \zeta_2^2 & \cdots & \zeta_2^k \\ \vdots & \vdots & \cdots & \vdots \\ \zeta_k^1 & \zeta_k^2 & \cdots & \zeta_k^n \end{pmatrix}. \tag{2.2}$$

따라서, Fisher의 기대정보행렬의 역행렬은

$$\begin{aligned} \Gamma^{-1}(\delta, \boldsymbol{\zeta}) &= \left(\frac{\partial(\delta, \boldsymbol{\zeta})}{\partial(\boldsymbol{\theta})}\right)\left(\frac{\partial(\delta, \boldsymbol{\zeta})}{\partial(\boldsymbol{\theta})}\right)^T \\ &= \begin{pmatrix} \sum_{i=1}^{n}\eta_{(i)}^2 & \phi^T \\ \phi & A \end{pmatrix} \end{aligned} \tag{2.3}$$

으로 표현된다. 여기서, $\phi = \left(\sum_{j=1}^{n}\eta_{(j)}\zeta_2^j, \cdots, \sum_{j=1}^{n}\eta_{(j)}\zeta_k^j\right)^T$ 이고, A는 $(k-1)\times(k-1)$ 인 정칙행렬(nonsingular matrix)이다.

만약, $\phi = 0$ 이라면 δ와 ζ는 서로 직교한다. 이러한 가정으로부터 $n-1$개의 동질적인 선형인 편미분방정식을 유도할 수 있다. $\phi(\theta_i^2 - \theta_j^2,\ i<j)$의 형식을 가진 어떤 함수가 방정식의 해가 될 수 있다. 예를 들어, $\zeta_i(\boldsymbol{\theta}) = \nu_i = (\theta_1^2 - \theta_i^2)/2$, $i = 2,3,\cdots,k$를 변환방법으로 취할 수 있다. 그러면 δ와 ζ들은 직교하고, 식 (2.2)는 다음과 같이 표현되어 진다.

$$\frac{\partial(\delta, \boldsymbol{\zeta})}{\partial(\boldsymbol{\theta})} = \begin{pmatrix} \eta_{(1)} & \eta_{(2)} & \eta_{(3)} & \cdots & \eta_{(k)} \\ \theta_1 & -\theta_2 & 0 & \cdots & 0 \\ \theta_1 & 0 & -\theta_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \theta_1 & 0 & 0 & \cdots & -\theta_k \end{pmatrix}.$$

이 행렬의 행렬식(Determinant)은 $\frac{1}{n}\left(\prod_{i=1}^{k}\theta_i\right)^{1+\frac{1}{k}}\left(\sum_{i=1}^{k}\theta_i^{-2}\right)$이 된다. 위의 행렬을 이용하면 식 (2.2)의 Fisher의 정보행렬은 다음의 식으로 다시 표현할 수 있다.

$$I(\delta,\zeta)^{-1}=\begin{pmatrix}\sum_{i=1}^{n}\eta_{(i)}^2 & 0 & \cdots & 0\\ 0 & \theta_1^2+\theta_2^2 & \cdots & \theta_1^2\\ \vdots & \vdots & \cdots & \theta_1^2\\ 0 & \theta_1^2 & \cdots & \theta_1^2+\theta_2^2\end{pmatrix}$$

Tibshirani의 방법(Berger; 1992 참조)을 이용하면 다음과 같은 사전분포를 구할 수 있다.

$$\pi(\delta,\zeta)\propto g(\zeta)\left(\sum_{i=1}^{n}\eta_{(i)}^2\right)^{-1/2}$$

여기서, $g(\zeta)$는 임의의 양의 함수이다.

위의 사전확률분포를 변환시켜 원래의 모수 $\boldsymbol{\theta}$의 사전확률분포를 구하면

$$\begin{aligned}\pi^*(\boldsymbol{\theta}) &\propto g(\zeta(\boldsymbol{\theta}))\left(\sum_{i=1}^{k}\eta_{(i)}^2\right)^{-1/2}\left|\frac{\partial(\theta,\zeta)}{\partial(\boldsymbol{\theta})}\right|\\ &\propto g(\zeta(\boldsymbol{\theta}))\frac{\frac{1}{k}\left(\prod_{i=1}^{k}\theta_i\right)^{1+1/k}\sum_{i=1}^{k}\theta_i^{-2}}{\frac{1}{k}\sqrt{\left(\prod_{i=1}^{k}\theta_i\right)^{2/k}(\theta_1^{-2}+\theta_2^{-2}+\cdots\theta_k^{-2})}}\\ &\propto g(\zeta(\boldsymbol{\theta}))\left(\prod_{i=1}^{k}\theta_i\right)\sqrt{\sum_{i=1}^{k}\theta_i^{-2}}\end{aligned}\tag{2.4}$$

이다. 따라서 식(2.4)를 사용하여 얻은 결합사후분포는 다음과 같아진다.

$$\begin{aligned}\pi(\boldsymbol{\theta}\mid D) &\propto g(\zeta(\boldsymbol{\theta}))\left(\prod_{i=1}^{k}\theta_i\right)\sqrt{\sum_{i=1}^{k}\theta_i^{-2}}\prod_{j=1}^{n}\left(\frac{1}{\left(\prod_{i=1}^{k}\theta_i\right)}\exp\left\{-\sum_{i=1}^{k}\frac{x_{ij}}{\theta_i}\right\}\right)\\ &\propto g(\zeta(\boldsymbol{\theta}))\sqrt{\sum_{i=1}^{k}\theta_i^{-2}}\frac{1}{\left(\prod_{i=1}^{k}\theta_i\right)^{m-1}}\exp\left\{-\sum_{i=1}^{k}\frac{\sum_{j=1}^{n}x_{ij}}{\theta_i}\right\}.\end{aligned}\tag{2.5}$$

여기서, $g(\zeta(\boldsymbol{\theta}))=1$이라 하면 식(2.5)는 다음의 식으로 축소되어 진다.

$$\pi(\boldsymbol{\theta}\mid D)\propto\sqrt{\sum_{i=1}^{k}\theta_i^{-2}}\frac{1}{\left(\prod_{i=1}^{k}\theta_i\right)^{n-1}}\exp\left\{-\sum_{i=1}^{k}\frac{\sum_{j=1}^{n}x_{ij}}{\theta_i}\right\}.\tag{2.6}$$

3. 가중 몬테칼로방법

베이지안 추론에서 사후 기대값 $E(h(\boldsymbol{\theta})\mid D)$의 계산에서 복잡하거나 불가능한 고차원 적분이 존재할 수 있다. 이와 같은 계산 문제는 몬테칼로 방법으로 해결할 수 있다. 특히, 주표본 방법과 조건부 몬테칼로 방법과 같은 대부분의 기본적인 몬테칼로 방법들(Hammersley와 Handscomb; 1964, Trotter와 Tukey; 1956 참조)은 현재도 유용하게 사용되어지고 있다.

사후분포 $\pi(\boldsymbol{\theta}\mid D)$가 식(2.6)과 같이 복잡하여 이 분포로부터 표본추출이 용이하지 않으면 마코브체인 몬테칼로 방법을 적용하는 대신 주표본 함수를 이용한 가중 몬테칼로 방법이 사용

된다.

식(2.6)의 주표본 함수 $g(\boldsymbol{\theta})$를 아래와 같이 가정하여 보자.

$$g(\boldsymbol{\theta}) = \frac{1}{\left(\prod_{i=1}^{k}\theta_i\right)^{n-1}} \exp\left\{-\sum_{i=1}^{k}\frac{\sum_{j=1}^{n}x_{ij}}{\theta_i}\right\} \tag{3.1}$$

식(3.1)에서 $\theta_{-i}=(\theta_1,\cdots,\theta_{i-1},\theta_{i+1},\cdots,\theta_k)'$와 D가 주어졌을 때 θ_i의 조건부 분포를 구하면 아래와 같아진다.

$$g(\theta_i \mid \theta_{-i}, D) \propto \frac{1}{\theta_i^{n-1}} \exp\left\{-\sum_{j=1}^{n}\frac{x_{ij}}{\theta_i}\right\},\ i=1,2,\cdots,k \tag{3.2}$$

그러므로, θ_i^{-1}의 조건부 분포는 모수가 $\gamma=n-2$, $\beta=1/(\sum_{j=1}^{n}x_{ij})$인 감마분포를 따른다.

이 조건부 사후분포들을 깁스표집법에 적용하여 얻은 표본을 $\{(\boldsymbol{\theta}_i,\ w_i),\ i=1,2,\cdots,n\}$라고 하자. 이 때, $h(\boldsymbol{\theta})=(\prod_{i=1}^{k}\theta_i)^{1/k}$의 사후 추정값은 다음과 같이 구해진다.

$$\hat{E}_w(h) = \frac{\sum_{i=1}^{m} w_i h(\boldsymbol{\theta}_i)}{\sum_{l=1}^{m} w_l} \tag{3.3}$$

여기서, w_j는 $\{(\boldsymbol{\theta}^{(i)}=(\theta_1,\cdots,\theta_k),\ w^{(i)}),\ i=1,2,\cdots,m\}$, j 단계에서의 가중치로 다음과 같이 구해진다.

$$w_j = \sqrt{\sum_{j=1}^{k}\theta_j^{(i)-2}} \tag{3.4}$$

주표본 방법의 유효성과 정확성은 주표본 함수 $g(\boldsymbol{\theta})$의 적절한 선택에 달려 있다. Huzurbazar과 Butler(1998 참조)가 제안한 방법과 같이, 모의실험으로 식(3.2)에서 정의되는 γ값의 변화에 따른 유효성을 비교해 본 결과 $\gamma=n-2$인 식(3.2)를 $\gamma=n$으로 변형하면 최적의 주표본 함수가 됨을 보였다. [표 1]은 $(\theta_1,\theta_2,\theta_3)$과 $n=64$에 의해 얻어진 모의실험 결과를 정리한 것이다.

[표 1] 최적의 주표본 함수

$(\theta_1,\theta_2,\theta_3)$	True value	original($\gamma=n-2$) Estimated value	original($\gamma=n-2$) MSE	best($\gamma=n$) Estimated value	best($\gamma=n$) MSE
(1,1,1)	1.000000	1.0387489	0.0128259	1.0017394	0.0105368
(1,2,3)	1.817121	1.8984333	0.0444161	1.8219971	0.0359836
(1,5,10)	3.684032	3.8230321	0.1729619	3.6876061	0.1386465
(2,2,2)	2.000000	2.0738037	0.0496321	2.0017574	0.0434291
(3,3,3)	3.000000	3.1094335	0.1118389	3.0142650	0.0942327
(5,5,5)	5.000000	5.1909142	0.3187658	5.0227563	0.2412505
(10,10,10)	10.00000	10.3669150	1.3265981	10.0359229	1.0215633

4. 사후 추정량에 대한 전통적 신뢰확률

추론에 사용된 비정보적 사전확률분포가 적절한 지를 평가하는 기준으로 베이지안 추론결과가 전통적 성질(frequentist property)을 만족하는 것인 지를 평가하는 방법이 널리 사용되고 있다(Berger 와 Bernardo, 1989 ; Efron, 1987 ; Ghosh 와 Mukerjee, 1992 ; Stein, 1985 ; Welch

와 Peers, 1963 ; Ye, 1993 ; Ye 와 Berger, 1991참조). 여기에 사용되는 평가방법으로는 베이지안추정치들의 $1-\alpha$번째 사후분위수(posterior quantile)의 포함확률(frequentist coverage probability)이 $1-\alpha$ 값에 근사하는 지를 평가하는 방법이다. [표 2]는 제안된 사전확률분포와 Jeffreys의 사전확률분포(균등사전확률분포)를 사용하여 모의실험으로 포함확률을 계산하여 서로 대비시킨 표이다. 모의실험에서는 여러 가지 $\boldsymbol{\theta}=(\theta_{10},\theta_{20},\cdots,\theta_{k0})$의 참값과 α = 0.05, 0.95, 그리고 k=3으로 두었다. 여기서 사용된 포함확률의 계산알고리즘은 다음과 같은 단계로 이루어졌다.

(단계1). 평균이 $\boldsymbol{\theta}$인 지수분포로부터 $\boldsymbol{x}$를 발생시킨다.

(단계2). 3장에 소개된 주표본방법으로 사후표본 $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_k)$를 m_1번 반복 발생시킨 후, 이들을 이용하여 m_1개 표본 중에서 $(\prod_{i=1}^{k}\theta_i)^{1/k}\leq\delta_0$를 성립시키는 표본의 비율 ρ를 구한다. 여기서 $\delta_0=(\prod_{i=1}^{k}\theta_{i0})^{1/k}$이며, 이것은 주어진 참값에서 계산된다. 그리고, ρ는 δ가 구간 $(0,\delta_0)$에 포함될 주변 사후확률의 추정값이 된다.

(단계3). 단계1과 단계2을 m_2번 반복시행해서 $\rho\leq\alpha$를 만족하는 시행의 비율을 ϕ로 구하면, 이것이 α번째 사후분위수의 포함확률 추정값이 된다.

[표 2]는 위의 알고리즘에서 $n=64$, m_1= 10000, 그리고 m_2= 20000을 사용하여 얻은 값들이다. 이 표에 의하면 제안된 사전확률분포를 사용한 사후확률분포로 부터 추정된 모수값들의 포함확률이 Jeffreys의 사전확률분포를 이용하였을 때 보다 더 α값에 정확히 근사함을 알 수 있다. 이는 제안된 사전확률분포가 지수분포 모수들의 기하평균을 추정하는데 Jeffreys의 사전확률분포보다 더 적당한 것임을 나타낸다.

[표 2] k=3일 때 사후분위수 0.05(0.95)의 포함확률

$\boldsymbol{\theta}_0$	(1,1,1)	(1,2,3)	(1,3,5)	(2,2,2)	(3,3,3)	(5,5,5)	(10,10,10)
uniform prior	.080(.970)	.065(.961)	.082(.974)	.078(.966)	.072(.969)	.074(.971)	.065(.969)
suggested prior	.050(.948)	.041(.945)	.054(.943)	.046(.942)	.047(.954)	.057(.945)	.050(.941)

5. 결론

본 연구에서는 경영이나 경제학 등에서 변화율의 대표값으로 널리 사용하는 기하평균의 베이지안추정을 위해 Stein과 Tibshirani의 연구에 기초한 사전확률분포 유도하고, 이를 이용하여 가중 Monte Carlo방법으로 기하평균을 추정하는 방법 제안하였다. 또한 모의실험을 통해, 제안된 사전확률분포가 기하평균의 추정에 적절함을 Jeffreys의 사전확률분포를 사용한 경우와 대비하여 보였다.

본 연구에서는 지수분포모수들의 기하평균에 대한 베이지안 추정법에 한하여 논하였으나, 이를 조금만 변형하면 타 분포 모수들의 기하평균 추정에도 쉽게 적용할 수 있을 것이다.

참고문헌

[1] Berger, J. (1992), Discussion of "Non-Informative Priors" by J. K. Ghosh and R

Mukerjee, in *Bayesian Statistics 4*, edu. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, London: Oxford University Press, pp 205-206

[2] Berger, J., and Bernardo, J. (1989), "Estimation a Product of Means: Bayesian Analysis With Reference Priors", *Journal of the American Statistical Association*, 84, 200-207

[3] Berger, J., and Robert, C. (1992), "Estimation of Quadratic Function(Reference Priors for Noncentrality Parameters)", unpublished manuscript, Purdue University, Dept. of Statistics.

[4] Bernardo, J. (1979), "Reference Posterior Distributions for Bayesian Inference" (with discussion), *Journal of Royal Statistical Society*, Ser B, 41, 113-147.

[5] Cox, D. R. and Reid, N. (1987), "Parameter Othogonality and Approximate Conditional Inference"(with Discussion), *Journal of Royal Statistical Society*, Ser B, 49, 1-39.

[6] Efron, B. (1987), "Why Isn't Everyone a Bayesian?", *American Statistician*, 40, 1-11.

[7] Ghosh, J. K. and Mukerjee, R. (1992), "Non-Informative Priors", in *Bayesian Statistics 4*, edu. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, London: Oxford University Press, 195-210.

[8] Hammersley, J. M., and Handscomb, D. C. (1964), *Monte Carlo Methods*. London: Metheun.

[9] Huzurbazar, S and Butler, R. W. (1998), "Importance Sampling for p-value Computations in Multivariate Tests", *Journal of Computational and Graphical Statistics*, Vol. 7, 342-355.

[10] Jeffrey, H. (1961), *Theory of Probability*(3rd ed.). Oxford, U.K.: Clarendon Press.

[11] Kenneth V. D., John S. G., and Kenneth J. S. (1998), "Incorporating a geometric mean formula into the CPI", *Monthly Labor Review*, October, 2-7.

[12] Stein, C. (1985), "On the Coverage Probability of Confidence Sets Based on a Prior Distribution", in *Sequential Methods in Statistics*, ed. R. Zielinski, Warsaw: PWN-Polish Scientific Publishers, 485-514.

[13] Tibshirani, R. (1989), "Non-Informative Priors for One Parameter of Many", *Biometrika*, 76, 604-608.

[14] Trotter, H. F. and Tukey, J. W. (1956), “Conditional Monte Carlo for normal Samples”. In symposium on Monte Carlo Methods(Ed. H. A. Meyer). New York: Wiley, 64-79.

[15] Welch, B. L., and Peers H. W. (1963), "On Formulae for Confidence Points Based on Integrals of Weighted Likelihoods", *Journal of Royal Statistical Society*, Ser. B, 25. 318-329.

[16] Ye, K. (1993), "Sensitivity Study of the Reference Priors in an Random Effect Model", Technical Report 92-18, Virginia Polytechnic Institute and State University, Dept. of Statistics.

[17] Ye, K., and Berger, J. (1991), "Noninformative Priors for Inferences in Exponential Regression Models", *Biometrika*, 78, 645-656.