Determining the Direction of the Path Using a Bayesian Semiparametric Model

Kei Miyazaki
Nagoya University
JSPS Research Fellow

Takahiro Hoshino
Nagoya University

Kazuo Shigemasu
Teikyo University

Outline

1. Introduction
    1-1 parameter estimation methods when the error variables are nonnormal
    1-2 semiparametric Bayesian estimation methods
2. Research purpose

Determining the direction of the path using a Bayesian semiparametric model
ー an application of Dirichlet process priors ー

3. Simulation studies
    3-1 Simulation 1 ーwhen the explanatory variables are normalー
    3-2 Simulation 2 ーwhen the explanatory variables are nonnormalー
    3-3 Simulation 3 ーwhen the nonlinear regression model is assumedー
4. Conclusion

Introduction

- In behavioral sciences, we face a situation where the observed variables do not follow normal distributions

→when the data are obtained from heterogeneous populations

- the methods that do not require normality for the error variables are required

- Asymptotic distribution free method (Browne, 1984) → used only when the sample size is large

- Estimation method when the error variables follow the elliptical distributions (Kano, Berkane & Bentler, 1993)→ it is necessary to assume symmetric distributions

Introduction

- An estimation method that uses a higher-order moment structure sparked interest among researchers in this field (Bentler, 1983; Shimizu & Kano, 2008)

→This method can be applied to data that are generated from very skewed distributions

→this method makes it possible to determine the direction of path among the models that have the same values of goodness of fit (that is, equivalent models)

Bayesian estimation

- Several features are pointed out

1. Not based on asymptotic theory
2. Derivation of theoretically correct predictive distributions
3. We can assume any complex models
4. we can compare several models accurately by using marginal likelihoods

Marginal likelihood

- Chib (1995) suggested a calculation method for the marginal likelihood from the Gibbs output
- The following identity is used

$$m(\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{\pi(\theta|\mathbf{y})}$$

Substituting the posterior means for the parameter, and taking logarithms,

$$\log \widehat{m}(\mathbf{y}) = \log f(\mathbf{y}|\theta^*) + \log \pi(\theta^*) - \log \widehat{\pi}(\theta^*|\mathbf{y})$$

The terms in the right side of the above equation are to be calculated

The problem is to calculate $\widehat{\pi}(\theta^*|\mathbf{y})$

Marginal likelihood

- The posterior distributions is expressed as follows

$$\pi(\theta^*|\mathbf{y}) = \pi(\theta_1^*|\mathbf{y})\times\pi(\theta_2^*|\mathbf{y},\theta_1^*)\times\cdots\times\pi(\theta_B^*|\mathbf{y},\theta_1^*,\cdots,\theta_{B-1}^*)$$

Each term in the right side of the above equation is calculated as follows

$$\begin{aligned}&\widehat{\pi}(\theta_r^* \mid \mathbf{y},\theta_s^*(s<r)) \\ &\quad = \; G^{-1}\sum_{j=1}^{G}\pi(\theta_r^*|\mathbf{y},\theta_1^*,\theta_2^*,\cdots,\theta_{r-1}^*,\theta_l^{(j)}(l>r))\end{aligned}$$

Marginal likelihood is calculated using the following equation

$$\begin{aligned}\log\widehat{m}(\mathbf{y}) = \log f(\mathbf{y}|\theta^*) &+ \log\pi(\theta^*) \\ &- \sum_{r=1}^{B}\log\widehat{\pi}(\theta_r^*|\mathbf{y},\theta_s^*(s<r))\end{aligned}$$

Bayesian estimation

- Several features are pointed out

1. Not based on asymptotic theory
2. Derivation of theoretically correct predictive distributions
3. We can assume any complex models
4. we can compare several models accurately by using marginal likelihoods

- it is necessary to make a distributional assumption for random variables in existing general Bayesian estimation method

→ is it impossible to perform estimation for any assumed shape of distributions for the parameters

Solution approach: Dirichlet process priors

- Ferguson(1973, 1974) proposed
- Assuming Dirichlet prior distribution $DP(\gamma, G_0)$

for θ that is the parameter of the distribution of random variable *U,*

$$\boldsymbol{U} \sim \sum_{l=1}^{\infty} p_l f(\cdot|\theta_l)$$

where $p_l = \prod_{m=1}^{l-1}(1 - V_m)V_l, \ V_l \sim \mathrm{Be}(1, \gamma)$

(definition by Sethuraman(1994))

→Any distributions can be expressed as mixture distributions of conventional distributions such as normal distributions

Finite Dirichlet process prior distribution

- Using *L*-dimensional finite Dirichlet process prior distribution $DP_L(\gamma, G_0)$,

$$\boldsymbol{U} \sim \sum_{l=1}^{L} p_l f(\cdot|\theta_l)$$

which approximates Dirichlet process prior distributions with satisfactory accuracy when *L* is large (Ishwaran & Zarepour, 2000)

- Ishwaran & James (2001) proposed Blocked Gibbs Sampler as a parameter estimation algorithm when finite Dirichlet process prior distribution is assumed

Research purposes

- We show Bayesian solution method for analyzing nonnormal data using Dirichlet process priors and calculation method of marginal likelihood
- We set two Dirichlet process mixture models wherein the explanatory and dependent variables are alternated with each other, calculate the marginal likelihoods and compare the resulting values
- We evaluate the performance of marginal likelihoods for determining the direction of the path

Model Assumption

- We consider two simple single regression models wherein the explanatory and dependent variables are alternated with each other

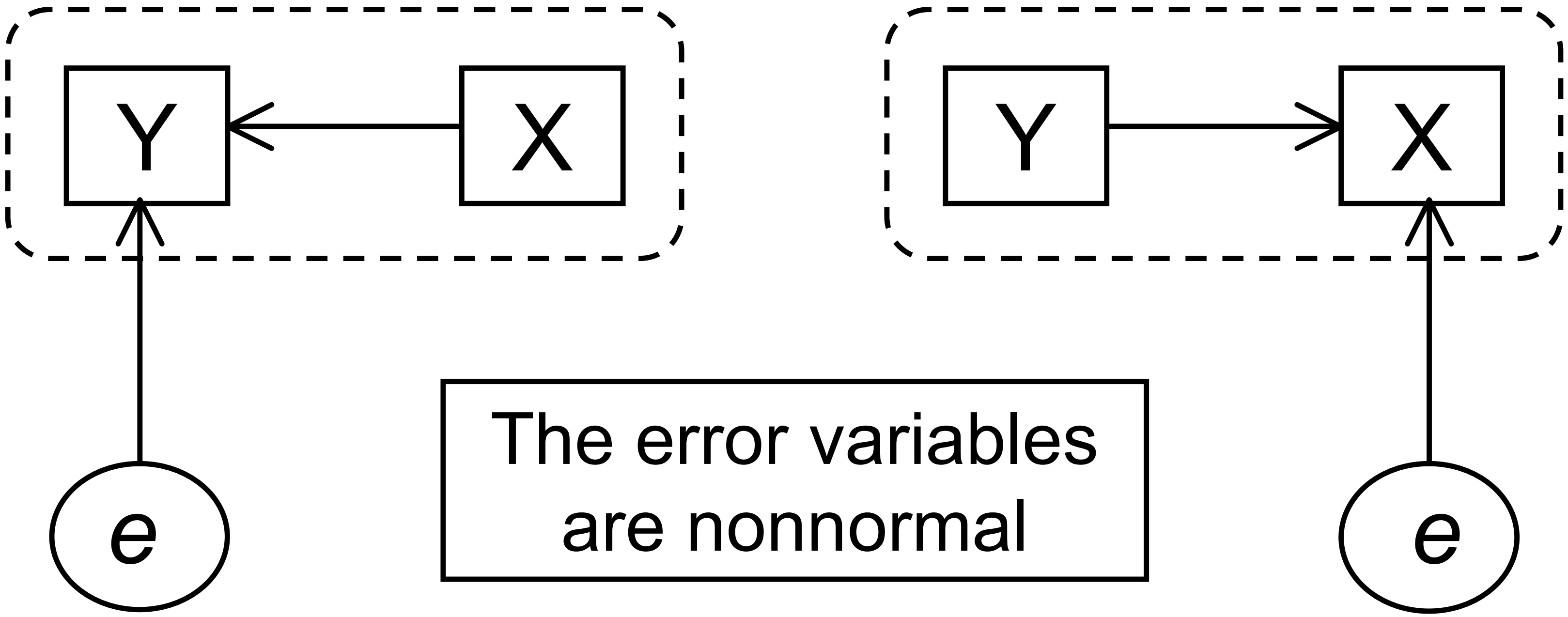
y_i : dependent variable of *i*-th observation

x_i : explanatory variable of *i*-th observation

$$y_i = \beta x_i + e_i \qquad x_i = \beta' y_i + e_i'$$

Graphical explanation

the same data



$$e_i \sim \sum_j^J p_j N(\mu_j, \sigma_j^2)$$

Priors

$$\boldsymbol{e} \sim DP_L(\gamma, N(\boldsymbol{\mu}, \boldsymbol{\sigma}))$$

$$\beta \sim N(\beta_0, \sigma_\beta^2)$$

$$\boldsymbol{p} = (p_1, \cdots, p_L) \sim \mathrm{Dirichlet}(\boldsymbol{a}, \boldsymbol{b})$$

- The values of hyperparameters are fixed
- γ : the parameter that indicates ease of transition to other components（fixed）

Blocked Gibbs Sampler

- Blocked Gibbs Sampler is the parameter estimation method for the models in which finite Dirichlet process prior distributions are assumed
- '･･･' represents that the other parameters are given

(a) Generate β

$$p(\beta|\cdots) \propto \prod_i^N p(y_i|k_i = l, \beta, \mu_l, \sigma_l^2)p(\beta)$$

※For the components which no subjects belong to, parameters are generated only from prior distribution: $p(\beta)$

(b) Generate μ (the mean of the error variables)

$$p(\mu_l|\cdots) \propto \prod_i^N p(y_i|k_i = l, \mu_l, \sigma_l^2, \beta)p(\mu_l)$$

(c) Generate σ (variance for the error variables)

$$p(\sigma_l|\cdots) \propto \prod_i^N p(y_i|k_i = l, \mu_l, \sigma_l^2, \beta)p(\sigma_l^2)$$

(d) Generate *k* (component to which each observation belongs) $p(k_i|\cdots) \sim \sum_{l=1}^L \pi_{li}\delta_l(\cdot)$

$$\pi_{li} \propto \frac{p_l\sigma_l^{-1}\,\text{exp}\left[-\frac{1}{2\sigma_l^2}(e_i - \mu_l)^2\right]}{\sum_l^L p_l\sigma_l^{-1}\,\text{exp}\left[-\frac{1}{2\sigma_l^2}(e_i - \mu_l)^2\right]}$$

where $e_i = y_i - \beta x_i$

(e) generate *p* (estimation of the probability of each component membership)

$$p_l = \prod_{m=1}^{l-1}(1 - V_m)V_l$$

$$V_l \sim \text{Be}(a_l + M_l, b_l + \sum_{m=l+1}^{L} M_m)$$

M_m : the number of subjects assigned to *m*-th component

a,b : hyperparameters

<u>This part is the core of Dirichlet process priors and largely different from normal finite mixture models</u>

Marginal likelihood when Dirichlet process priors are used

- Basu & Chib (2003) proposed the calculation method of marginal likelihood when Dirichlet process priors are used

$$\log p(\boldsymbol{y}) = \underline{\log p(\boldsymbol{y}|\beta^*)} + \log p(\beta^*) - \underline{\log p(\beta^*|\boldsymbol{y})}$$

This term is rather difficult to calculate

Chib (1995)

- We used Sequential Importance Sampling (Kong, Liu & Wong, 1994). Let $\boldsymbol{\theta} = (\boldsymbol{\mu}^t, \boldsymbol{\sigma}^t)^t$

(a) generate $\boldsymbol{\theta}_1^{(g)}, \cdots, \boldsymbol{\theta}_N^{(g)}$ from $p(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_N | \boldsymbol{y}, \beta^*)$

$$p(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_N | \boldsymbol{y}, \beta^*) = \prod_{i=1}^N \underline{p(\boldsymbol{\theta}_i | \boldsymbol{y}_{(i)}, \boldsymbol{\theta}_{(i-1)}, \beta^*)}$$

$$\underline{p(\boldsymbol{\theta}_i | \boldsymbol{y}_{(i)}, \boldsymbol{\theta}_{(i-1)}, \beta^*)} \propto \frac{\gamma}{\gamma + i - 1} p(y_i | \boldsymbol{\theta}_i, \beta^*) p(\boldsymbol{\theta}_i | \boldsymbol{\tau}) + \sum_{j=1}^{m_{i-1}} \frac{n_{j,i-1}}{\gamma + i - 1} p(y_i | \boldsymbol{\theta}_i, \beta^*) \delta_{\boldsymbol{\theta}_j}(\boldsymbol{\theta}_i)$$

where $\boldsymbol{y}_{(i)} = (y_1, \cdots, y_i)^t$ $\boldsymbol{\theta}_{(i)} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_i)^t$

m_{i-1} : The number of patterns of different parameter values within $(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{i-1})$

$n_{j,i-1}$: the frequency of observations that have *j*-th parameter value

(b) calculate w

$$w = w(\boldsymbol{\theta}_1^{(g)}, \cdots, \boldsymbol{\theta}_N^{(g)}) = \prod_{i=1}^{N} p(y_i | \boldsymbol{y}_{(i-1)}, \boldsymbol{\theta}_{i-1}^{(g)}, \beta^*)$$

$$p(y_i | \boldsymbol{y}_{(i-1)}, \boldsymbol{\theta}_{(i-1)}, \beta^*) = \frac{\gamma}{\gamma + i - 1} \int p(y_i | \boldsymbol{\theta}_i, \beta^*) p(\boldsymbol{\theta}_i | \boldsymbol{\tau}) d\boldsymbol{\theta}_i + \sum_{j=1}^{m_{i-1}} \frac{n_{j,i-1}}{\gamma + i - 1} p(y_i | \boldsymbol{\theta}_j, \beta^*)$$

The above integration is approximated by Monte Carlo method

(c) Calculate the mean of w

$$\bar{w} = \frac{1}{G} \sum_{g=1}^{G} w(\boldsymbol{\theta}_1^{(g)}, \cdots, \boldsymbol{\theta}_N^{(g)})$$

Simulation studies

- We considered a simple single regression model, generated 100 data set, analyzed each data set and calculated the marginal likelihood
- Sample size＝100, μ = (-2.0,2.0)’, σ = (0.5,2.0)’, *p*=(0.5,0.5)’, β = 0.5, γ =3.0, *L*=10
- We considered two cases about the explanatory variables (X).

$$x_i \sim N(0, 10.0) \quad \text{(normal)}$$

$$\begin{aligned} x_i &\sim 0.3 \cdot N(-2.0, 1.0) \\ &+ 0.7 \cdot N(2.0, 2.0) \end{aligned} \quad \text{(nonnormal)}$$

Model comparison via the joint distribution

- In determining the direction of the path, it is necessary to compare the joint distribution of x and y:

We compare

$$\log m(\boldsymbol{y}|\boldsymbol{x}) + \log m(\boldsymbol{x})$$

and

$$\log m(\boldsymbol{x}|\boldsymbol{y}) + \log m(\boldsymbol{y})$$

Simulation results

Data generating	$Y=\beta X+e$ X is normal	$Y=\beta X+e$ X is nonnormal	$Y=\beta X^3+e$
Data analysis	(1-a) $Y=\beta X+e$ (**87**) $Y=\beta X^3+e$ (**13**) (1-b) $Y=\beta X+e$ (**92**) $X=\beta' Y+e'$ (**8**)	(2-a) $Y=\beta X+e$ (**78**) $Y=\beta' X^3+e'$ (**22**) (2-b) $Y=\beta X+e$ (**72**) $X=\beta' Y+e'$ (**28**)	(3-a) $Y=\beta X^3+e$ (**85**) $Y=\beta' X+e'$ (**15**) (3-b) $Y=\beta X+e$ (**42**) $X=\beta' Y+e'$ (**58**)

※ parentheses: the number of times that the model is adopted via marginal likelihood

Discussion

- We illustrated Bayesian approach from parameter estimation to model comparison for nonnormal data
- true models are not always adopted in determining the direction of the path
- Naturally, the direction of the path does not relate to the direction of causation
- At least, it is meaningless to identify the causation by determining the direction of the path using nonnormal error variables

Thank you very much !

miyazaki.behaviormetrics@gmail.com