Advanced Bayesian Methods - Assignment 2

통계데이터사이언스학과
박주연 (2022311137)

Q1-(a)

Mixed effect model $y \sim N_N(X\beta, \sigma^2 I_N)$, $\beta \sim N_{md}(1_m \otimes \mu_\beta, I_m \otimes \Sigma_\beta)$

$y = [y_1', \cdots, y_m']' \in \mathbb{R}^N$, $y_j \in \mathbb{R}^{n_j}$

$X = diag(X_1, \cdots, X_m) \in \mathbb{R}^{N \times (md)}$, X_j : design matrix for subject j $(j = 1, \cdots, m)$

$\beta = [\beta_1', \cdots, \beta_m']' \in \mathbb{R}^{md}$, $\beta_j \in \mathbb{R}^d$

$$p(y \mid \sigma^2, \mu_\beta, \Sigma_\beta) = \int p(y, \beta \mid \sigma^2, \mu_\beta, \Sigma_\beta)\, d\beta$$
$$= \int \underbrace{p(y \mid \beta, \sigma^2)}_{\propto MVN_N(X\beta, \sigma^2 I_N)} \underbrace{p(\beta \mid \mu_\beta, \Sigma_\beta)}_{\propto MVN_{md}(1_m \otimes \mu_\beta, I_m \otimes \Sigma_\beta)} d\beta$$

Since both distributions are MVN, $p(y, \beta \mid \sigma^2, \mu_\beta, \Sigma_\beta)$ is MVN.

And the marginal distribution $p(y \mid \sigma^2, \mu_\beta, \Sigma_\beta)$ is also MVN. ($\because$ marginal of MVN is MVN)

The parameters of $p(y \mid \sigma^2, \mu_\beta, \Sigma_\beta)$ are

$$E(y \mid \sigma^2, \mu_\beta, \Sigma_\beta) = E(E(y \mid \beta, \sigma^2, \mu_\beta, \Sigma_\beta)) = E(X\beta) = X(1_m \otimes \mu_\beta)$$

$$Var(y \mid \sigma^2, \mu_\beta, \Sigma_\beta) = E(Var(y \mid \sigma^2, \mu_\beta, \Sigma_\beta)) + Var(E(y \mid \beta, \sigma^2, \mu_\beta, \Sigma_\beta)) = \sigma^2 I_N + Var(X\beta) = \sigma^2 I_N + X(I_m \otimes \Sigma_\beta)X'$$

Therefore, $p(y \mid \sigma^2, \mu_\beta, \Sigma_\beta) \propto N_N(X(1_m \otimes \mu_\beta), X(I_m \otimes \Sigma_\beta)X')$

Q1-(b)

The prior $\mu_\beta \sim N_d(\xi, \Omega)$ $\sigma^2 \sim Inv\text{-}\chi^2(\nu, \tau^2)$ $\Sigma_\beta \sim Inv\text{-}Wishart_p(\Psi^{-1})$

i) $p(\beta \mid \sigma^2, \mu_\beta, \Sigma_\beta, y) \propto p(y, \beta \mid \sigma^2, \mu_\beta, \Sigma_\beta)$

$$= p(y \mid \beta, \sigma^2, \mu_\beta, \Sigma_\beta)\, p(\beta \mid \mu_\beta, \Sigma_\beta)$$

$$= \prod_{j=1}^{m} \underbrace{p(y_j \mid \beta_j, \sigma^2, \mu_\beta, \Sigma_\beta)}_{N(X_j\beta_j,\, \sigma^2 I_{n_j})} \underbrace{p(\beta_j \mid \mu_\beta, \Sigma_\beta)}_{N(\mu_\beta, \Sigma_\beta)} \qquad (\because y_j \perp\!\!\!\perp y_i ,\ j \neq i)$$

$$\propto \prod_{j=1}^{m} \exp\left\{-\frac{1}{2}\left[(y_j - X_j\beta_j)'\frac{1}{\sigma^2}(y_j - X_j\beta_j) + (\beta_j - \mu_\beta)'\Sigma_\beta^{-1}(\beta_j - \mu_\beta)\right]\right\}$$

$$(y_j - X_j\beta_j)'\frac{1}{\sigma^2}(y_j - X_j\beta_j) = \cancel{\frac{1}{\sigma^2}y_j'y_j} - 2\frac{1}{\sigma^2}y_j'X_j\beta_j + \frac{1}{\sigma^2}\beta_j'X_j'X_j\beta_j$$

$$(\beta_j - \mu_\beta)'\Sigma_\beta^{-1}(\beta_j - \mu_\beta) = \beta_j'\Sigma_\beta^{-1}\beta_j - 2\mu_\beta'\Sigma_\beta^{-1}\beta_j + \cancel{\mu_\beta'\Sigma_\beta^{-1}\mu_\beta}$$

$$\propto \prod_{j=1}^{m} \exp\left\{-\frac{1}{2}\left[\beta_j'\left(\frac{1}{\sigma^2}X_j'X_j + \Sigma_\beta^{-1}\right)\beta_j - 2\left(\frac{1}{\sigma^2}y_j'X_j + \mu_\beta'\Sigma_\beta^{-1}\right)\beta_j\right]\right\}$$

$$\propto \prod_{j=1}^{m} \exp\left\{-\frac{1}{2}(\beta_j - \mu_{\beta_j}^*)'(\Sigma_{\beta_j}^*)^{-1}(\beta_j - \mu_{\beta_j}^*)\right\} \quad (\text{for some } \mu_{\beta_j}^* \text{ and } \Sigma_{\beta_j}^*)$$

⊛ $(\Sigma_{\beta_j}^*)^{-1} = \left(\frac{1}{\sigma^2}X_j'X_j + \Sigma_\beta^{-1}\right)$

$\Leftrightarrow \Sigma_{\beta_j}^* = \left(\frac{1}{\sigma^2}X_j'X_j + \Sigma_\beta^{-1}\right)^{-1} \quad \in \mathbb{R}^{d \times d}$

⊛ $\cancel{-2}(\mu_{\beta_j}^*)'(\Sigma_{\beta_j}^*)^{-1}\cancel{\beta_j} = \cancel{-2}\left(\frac{1}{\sigma^2}y_j'X_j + \mu_\beta'\Sigma_\beta^{-1}\right)\cancel{\beta_j}$

$\Leftrightarrow (\mu_{\beta_j}^*)' = \left(\frac{1}{\sigma^2}y_j'X_j + \mu_\beta'\Sigma_\beta^{-1}\right)(\Sigma_{\beta_j}^*)$

$\Leftrightarrow \mu_{\beta_j}^* = \underbrace{\left(\frac{1}{\sigma^2}X_j'X_j + \Sigma_\beta^{-1}\right)^{-1}}_{d \times d}\left(\frac{1}{\sigma^2}y_j'X_j + \mu_\beta'\Sigma_\beta^{-1}\right)' \in \mathbb{R}^d \qquad (\Sigma_{\beta_j}^* \text{ is symmetric}, (\Sigma_{\beta_j}^*)' = \Sigma_{\beta_j}^*)$

$\Rightarrow \beta_j \mid \sigma^2, \mu_\beta, \Sigma_\beta, y_j \sim N_d(\mu_{\beta_j}^*, \Sigma_{\beta_j}^*)$ where $\mu_{\beta_j}^* = \left(\frac{1}{\sigma^2}X_j'X_j + \Sigma_\beta^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}y_j'X_j + \mu_\beta'\Sigma_\beta^{-1}\right)'$ and $\Sigma_{\beta_j}^* = \left(\frac{1}{\sigma^2}X_j'X_j + \Sigma_\beta^{-1}\right)^{-1}$

$$\therefore \beta \mid \sigma^2, \mu_\beta, \Sigma_\beta, y_j \sim N_N(\mu, \Sigma) \text{ where } \mu = \begin{pmatrix} \mu_{\beta_1} \\ \mu_{\beta_2} \\ \vdots \\ \mu_{\beta_m} \end{pmatrix}, \ \Sigma = diag(\Sigma_{\beta_1}, \cdots, \Sigma_{\beta_m})$$

ii) $p(\sigma^2|\beta,\mu_\beta,\Sigma_\beta,y) \propto p(y,\sigma^2|\beta,\mu_\beta,\Sigma_\beta)$

$$= p(y|\beta,\sigma^2,\mu_\beta,\Sigma_\beta)\,p(\sigma^2|\beta,\mu_\beta,\Sigma_\beta)$$

$$= \underbrace{p(y|\beta,\sigma^2,\mu_\beta,\Sigma_\beta)}_{N(X\beta,\,\sigma^2 I_N)}\ \underbrace{p(\sigma^2|\nu,\tau^2)}_{Inv-\chi^2(\nu,\tau^2)}$$

$$\propto \det(\sigma^2 I_N)^{-1/2}\exp\left(\frac{-1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right)\cdot(\sigma^2)^{-\frac{\nu}{2}-1}\exp\left(-\frac{\nu\tau^2}{2\sigma^2}\right)$$

$$= (\sigma^2)^{-\frac{N+\nu}{2}-1}\exp\left(\frac{-1}{2\sigma^2}[(y-X\beta)'(y-X\beta)+\nu\tau^2]\right)$$

$\therefore\ \sigma^2|\beta,\mu_\beta,\Sigma_\beta,y \sim Inv-\chi^2\left(N+\nu,\ \frac{1}{N+\nu}[(y-X\beta)'(y-X\beta)+\nu\tau^2]\right)$

iii) $p(\mu_\beta|\beta,\sigma^2,\Sigma_\beta,y) \propto p(y,\mu_\beta|\beta,\sigma^2,\Sigma_\beta)$

$$\propto p(y|\beta,\sigma^2,\mu_\beta,\Sigma_\beta)\,p(\mu_\beta|\beta,\sigma^2,\Sigma_\beta)$$

$$\propto \cancel{p(y|\beta,\sigma^2,\mu_\beta,\Sigma_\beta)}\ p(\beta|\mu_\beta,\Sigma_\beta)\,p(\mu_\beta)$$

does not contain any term with μ_β

$$\propto \prod_{j=1}^m \underbrace{p(\beta_j|\mu_\beta,\Sigma_\beta)}_{N(\mu_\beta,\Sigma_\beta)}\ \underbrace{p(\mu_\beta|\xi,\Omega)}_{N(\xi,\Omega)}$$

$$\propto \prod_{j=1}^m\left\{\exp\left(-\tfrac{1}{2}(\beta_j-\mu_\beta)'\Sigma_\beta^{-1}(\beta_j-\mu_\beta)\right)\right\}\times\exp\left\{-\tfrac{1}{2}(\mu_\beta-\xi)'\Omega^{-1}(\mu_\beta-\xi)\right\}$$

$$= \exp\left[-\tfrac{1}{2}\sum_{j=1}^m\left\{(\beta_j-\mu_\beta)'\Sigma_\beta^{-1}(\beta_j-\mu_\beta)+\tfrac{1}{m}(\mu_\beta-\xi)'\Omega^{-1}(\mu_\beta-\xi)\right\}\right]$$

$$\circledast\ (\beta_j-\mu_\beta)'\Sigma_\beta^{-1}(\beta_j-\mu_\beta) = \mu_\beta'\Sigma_\beta^{-1}\mu_\beta - 2\beta_j'\Sigma_\beta^{-1}\mu_\beta + \cancel{\beta_j'\Sigma_\beta^{-1}\beta_j}$$

$$\tfrac{1}{m}(\mu_\beta-\xi)'\Omega^{-1}(\mu_\beta-\xi) = \tfrac{1}{m}\left[\mu_\beta'\Omega^{-1}\mu_\beta - 2\xi'\Omega^{-1}\mu_\beta + \cancel{\xi'\Omega^{-1}\xi}\right]$$

$$= \exp\left(-\tfrac{1}{2}\sum_{j=1}^m\left\{\mu_\beta'(\Sigma_\beta^{-1}+\tfrac{1}{m}\Omega^{-1})\mu_\beta - 2(\beta_j'\Sigma_\beta^{-1}+\tfrac{1}{m}\xi'\Omega^{-1})\mu_\beta\right\}\right)$$

$$= \exp\left(-\tfrac{1}{2}\left\{\mu_\beta'(m\Sigma_\beta^{-1}+\Omega^{-1})\mu_\beta - 2\left(\sum_{j=1}^m(\beta_j'\Sigma_\beta^{-1})+\xi'\Omega^{-1}\right)\mu_\beta\right\}\right)$$

$$\propto \exp\left(-\tfrac{1}{2}(\mu_\beta-\mu_*)'\Sigma_*^{-1}(\mu_\beta-\mu_*)\right)\quad \text{for some } \mu_* \text{ and } \Sigma_*$$

where $\Sigma_*^{-1} = (m\Sigma_\beta^{-1}+\Omega^{-1}) \Leftrightarrow \Sigma_* = (m\Sigma_\beta^{-1}+\Omega^{-1})^{-1}$

$-2\mu_*'\Sigma_*^{-1} = -2\left(\sum_{j=1}^m(\beta_j'\Sigma_\beta^{-1})+\xi'\Omega^{-1}\right) \Leftrightarrow \mu_* = (m\Sigma_\beta^{-1}+\Omega^{-1})^{-1}\left(\sum_{j=1}^m(\beta_j'\Sigma_\beta^{-1})+\xi'\Omega^{-1}\right)' \in \mathbb{R}^d$

$\therefore\ \mu_\beta|\beta,\sigma^2,\Sigma_\beta,y \sim N_d\left((m\Sigma_\beta^{-1}+\Omega^{-1})^{-1}\left(\sum_{j=1}^m(\beta_j'\Sigma_\beta^{-1})+\xi'\Omega^{-1}\right)',\ (m\Sigma_\beta^{-1}+\Omega^{-1})^{-1}\right)$

iv) $p(\Sigma_\beta|\beta,\sigma^2,\mu_\beta,y) \propto p(y,\Sigma_\beta|\beta,\sigma^2,\mu_\beta)$

$$\propto \cancel{p(y|\beta,\sigma^2,\mu_\beta,\Sigma_\beta)}\ p(\beta|\mu_\beta,\Sigma_\beta)\,p(\Sigma_\beta)$$

doesn't contain any term with Σ_β

$$\propto \prod_{j=1}^m \underbrace{p(\beta_j|\mu_\beta,\Sigma_\beta)}_{N(\mu_\beta,\Sigma_\beta)} \times \underbrace{p(\Sigma_\beta|\Psi^{-1},\nu)}_{Inv-wishart_\nu(\Psi^{-1})}$$

$$\propto \prod_{j=1}^m\left\{|\Sigma_\beta|^{-1/2}\exp\left(-\tfrac{1}{2}(\beta_j-\mu_\beta)'\Sigma_\beta^{-1}(\beta_j-\mu_\beta)\right)\right\}\times|\Sigma_\beta|^{-(p+d+1)/2}\exp\left(-\tfrac{1}{2}tr(\Psi\Sigma_\beta^{-1})\right)$$

$$= |\Sigma_\beta|^{-m/2}\exp\left(-\tfrac{1}{2}\sum_{j=1}^m(\beta_j-\mu_\beta)'\Sigma_\beta^{-1}(\beta_j-\mu_\beta)\right)\times|\Sigma_\beta|^{-(p+d+1)/2}\exp\left(-\tfrac{1}{2}tr(\Psi\Sigma_\beta^{-1})\right)$$

$$(\beta_j-\mu_\beta)'\Sigma_\beta^{-1}(\beta_j-\mu_\beta) = tr\left((\beta_j-\mu_\beta)'\Sigma_\beta^{-1}(\beta_j-\mu_\beta)\right)\quad (\because \text{scalar})$$

$$= tr\left((\beta_j-\mu_\beta)(\beta_j-\mu_\beta)'\Sigma_\beta^{-1}\right)\quad (\because tr(ABC) = tr(CAB))$$

$$= |\Sigma_\beta|^{-(m+p+d+1)/2}\exp\left(-\tfrac{1}{2}\sum_{j=1}^m tr\left((\beta_j-\mu_\beta)(\beta_j-\mu_\beta)'\Sigma_\beta^{-1}\right) - \tfrac{1}{2}tr(\Psi\Sigma_\beta^{-1})\right)$$

$$= |\Sigma_\beta|^{-(m+p+d+1)/2}\exp\left(-\tfrac{1}{2}tr\left\{\left(\sum_{j=1}^m(\beta_j-\mu_\beta)(\beta_j-\mu_\beta)'+\Psi\right)\Sigma_\beta^{-1}\right\}\right)$$

$\therefore\ \Sigma_\beta|\beta,\sigma^2,\mu_\beta,y \sim Inv-wishart\left(\sum_{j=1}^m\{(\beta_j-\mu_\beta)(\beta_j-\mu_\beta)'\}+\Psi,\ m+p\right)$

Q1-(c)

When we change the prior, we now update each component of Σ_β, $\sigma_{\beta,k}^2$ one by one, instead of updating Σ_β all at once.

Suppose $\beta_j = [\beta_{j,1}, \cdots, \beta_{j,d}]' \in \mathbb{R}^d$

$\mu_\beta = [\mu_{\beta,1}, \cdots, \mu_{\beta,d}]' \in \mathbb{R}^d$

$\Sigma_\beta^{(-k)} = \{ \sigma_{\beta,1}^2, \cdots, \sigma_{\beta,k-1}^2, \sigma_{\beta,k+1}^2, \cdots, \sigma_{\beta,d}^2 \}$

Conditional posterior $p(\sigma_{\beta,k}^2 \mid \beta, \sigma^2, \mu_\beta, y, \Sigma_\beta^{(-k)}) \propto p(y, \sigma_{\beta,k}^2 \mid \beta, \sigma^2, \mu_\beta, \Sigma_\beta^{(-k)})$

$$= p(y \mid \beta, \sigma^2, \mu_\beta, \Sigma_\beta)\, p(\sigma_{\beta,k}^2 \mid \beta, \sigma^2, \mu_\beta, \Sigma_\beta^{(-k)})$$

$$\propto p(\beta, \sigma_{\beta,k}^2 \mid \mu_\beta, \Sigma_\beta^{(-k)})$$

$$= p(\beta \mid \mu_\beta, \Sigma_\beta)\, p(\sigma_{\beta,k}^2)$$

$$= \prod_{j=1}^m \underbrace{p(\beta_j \mid \mu_\beta, \Sigma_\beta)}_{\propto N(\mu_\beta, \Sigma_\beta)} \underbrace{p(\sigma_{\beta,k}^2)}_{\propto \text{Inv-}\chi^2(\rho, \psi^2)}$$

$$= \prod_{j=1}^m |\Sigma_\beta|^{-1/2} \exp\left(-\tfrac{1}{2}(\beta_j - \mu_\beta)' \Sigma_\beta^{-1} (\beta_j - \mu_\beta)\right) \times (\sigma_{\beta,k}^2)^{-\rho/2-1} \exp\left(-\frac{1}{\sigma_{\beta,k}^2}\left(\frac{\rho\psi^2}{2}\right)\right)$$

$$|\Sigma_\beta|^{-1/2} = |\text{diag}(\sigma_{\beta,1}^2, \cdots, \sigma_{\beta,d}^2)|^{-1/2} = \left(\prod_{k=1}^d \sigma_{\beta,k}^2\right)^{-1/2} \propto (\sigma_{\beta,k}^2)^{-1/2}$$

$$\propto (\sigma_{\beta,k}^2)^{-(m+\rho)/2} \exp\left(-\tfrac{1}{2}\left[\sum_{j=1}^m \{ \underbrace{(\beta_j - \mu_\beta)'}_{1\times m} \underbrace{\Sigma_\beta^{-1}}_{m\times m} \underbrace{(\beta_j - \mu_\beta)}_{m\times 1} \} + \frac{1}{\sigma_{\beta,k}^2}\rho\psi^2\right]\right)$$

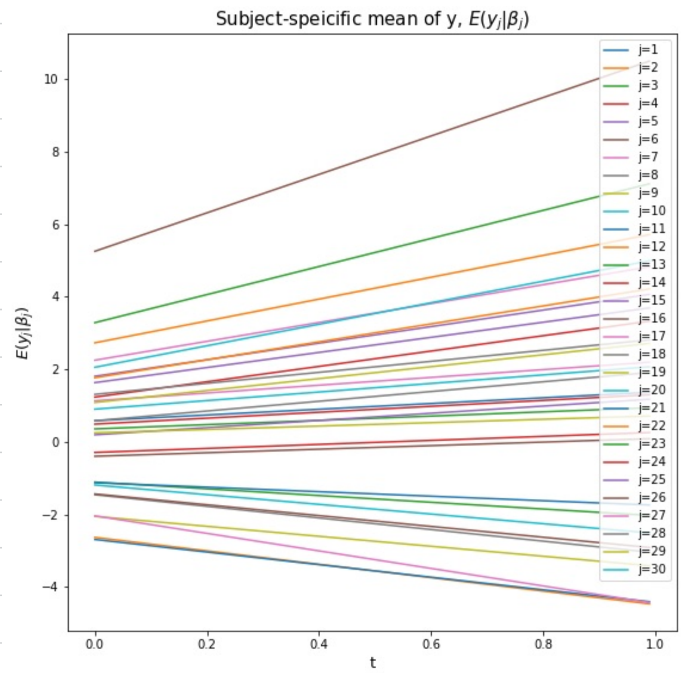
$$\sum_{\ell}^d (\sigma_{\beta,\ell}^2)^{-1} (\beta_{j,\ell} - \mu_{\beta,\ell})^2 \propto \frac{1}{\sigma_{\beta,k}^2}(\beta_{j,k} - \mu_{\beta,k})^2$$

$$\propto (\sigma_{\beta,k}^2)^{-(m+\rho)/2} \exp\left(-\frac{1}{2\sigma_{\beta,k}^2} \sum_{j=1}^m (\beta_{j,k} - \mu_{\beta,k})^2 + \rho\psi^2\right)$$

$$\propto \text{Inv-}\chi^2\left(m+\rho, \frac{1}{m+\rho}\left[\sum_{j=1}^m (\beta_{j,k} - \mu_{\beta,k})^2 + \rho\psi^2\right]\right)$$

Q1-(d)

i) Draw the subject-specific mean of y as a function of t



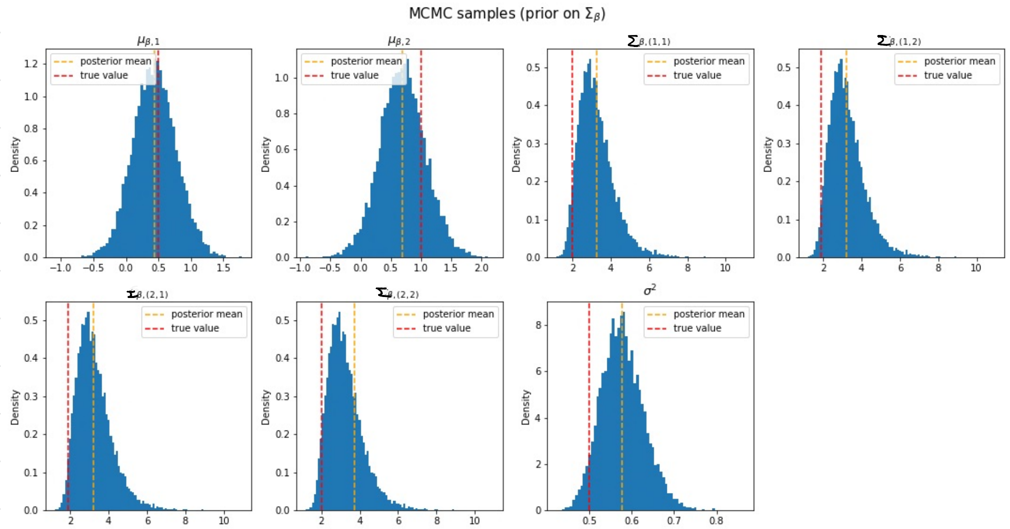
ii) Draw the marginal posterior distribution for μ_β, Σ_β and σ^2

① Case 1) prior on Σ_β as (b)

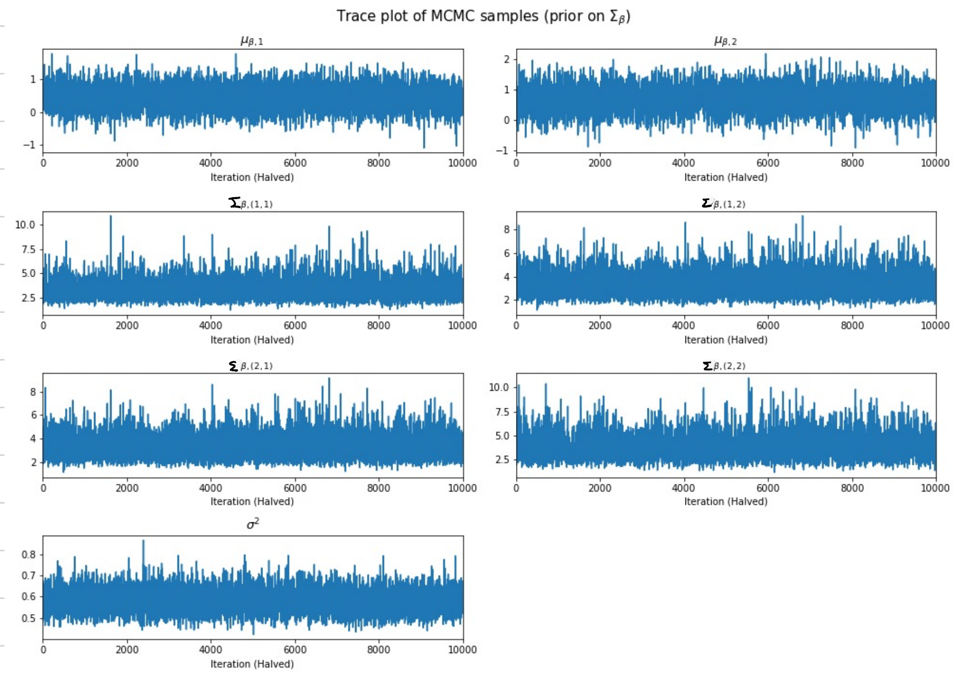
when looking at the traceplots of MCMC samples, it seems like the chain has converged well,

but comparing the actual value with the posterior mean, it cannot be said that the gibbs sampler has been performed well.

• histogram of μ_β, Σ_β and σ^2



. Traceplot for MCMC samples



• Actual value and posterior mean of μ_β, Σ_β and σ^2

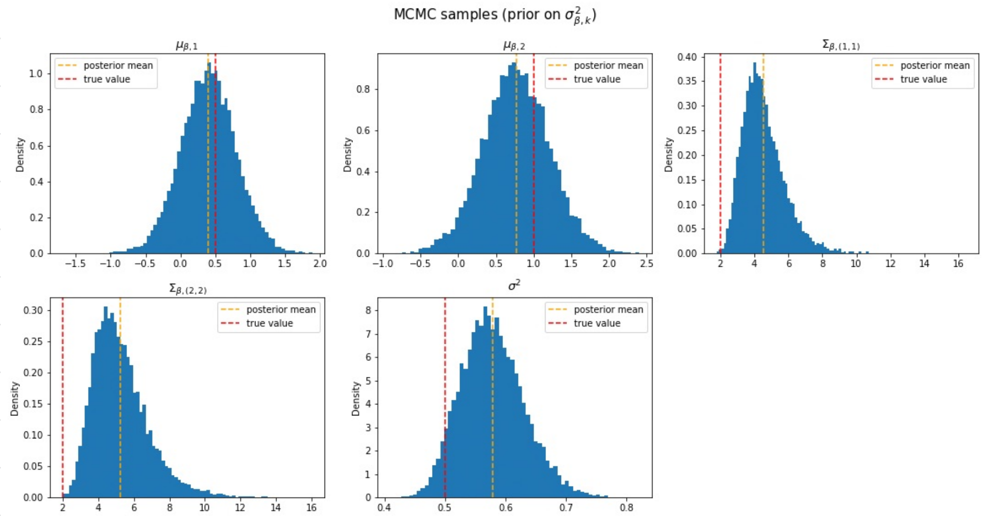
```
True values
mu_beta: [0.5 1. ]
cov_beta:
 [[2.  1.9]
 [1.9 2. ]]
sigma^2: 0.5

The means of MCMC samples
mu_beta:  [0.44243912 0.69901112]
cov_beta:
 [[3.25474284 3.21958513]
 [3.21958513 3.7233238 ]]
sigma^2: 0.5780527115203976
```

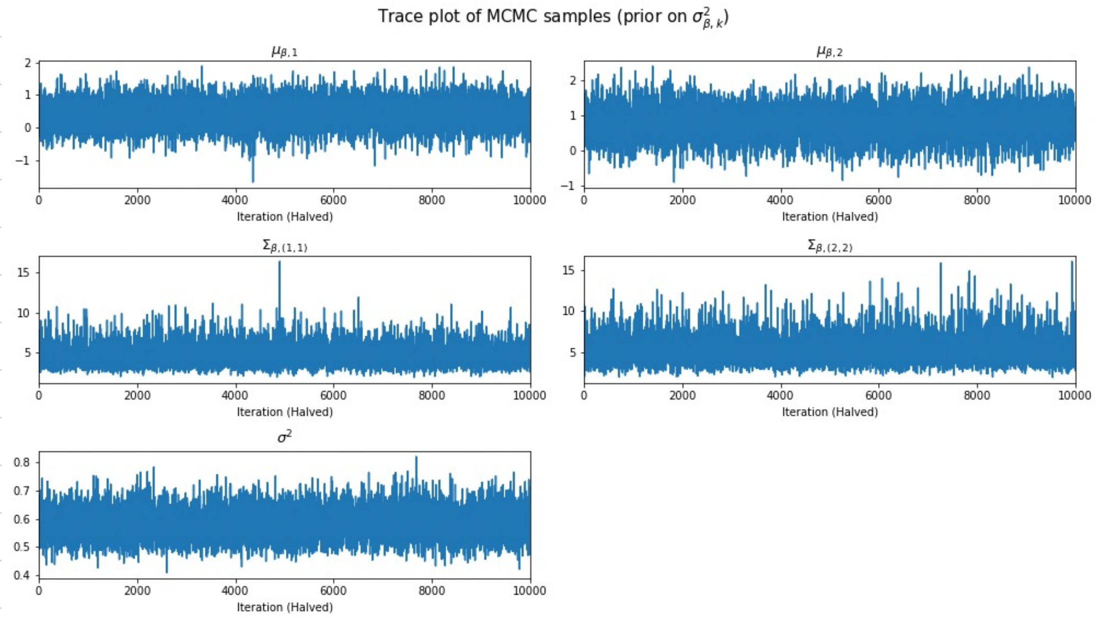
② Case 2) prior on $\sigma^2_{\beta,k}$ as (c)

Similar to the previous prior case, it seems converge well in traceplot, but there is difference between the posterior mean and actual value.

- histogram of μ_β, Σ_β and σ^2



. Traceplot for MCMC samples



- Actual value and posterior mean of μ_β, Σ_β and σ^2

```
True values
mu_beta: [0.5 1. ]
cov_beta:
[[2.  1.9]
 [1.9 2. ]]
sigma^2: 0.5

The means of MCMC samples
mu_beta: [0.40207464 0.77142362]
cov_beta:
 [[4.55200841 0.        ]
 [0.         5.26628787]]
sigma^2: 0.5782189042899708
```

iii) Posterior predictive check

To conduct posterior predictive check,

- Generate y^{rep}
  - Since $p(y^{rep} \mid y) = \int p(y^{rep} \mid \theta) p(\theta \mid y) \, d\theta$, we can get $y^{rep(s)}$ in Gibbs sampler.
  - i.e. get posterior simulations of $(y^{rep(s)}, \theta^{(s)})$
- Graphical comparison
  - compare the histogram of y and some y^{rep}
- Calculate Posterior Predictive P-value using residual ~~test~~ quantity. $T(y \mid \theta)$
  - $P_B = \iint \mathbb{1}\{T(y^{rep} \mid \theta) \geq T(y \mid \theta)\} \, p(y^{rep} \mid \theta) p(\theta \mid y) \, dy^{rep} \, d\theta$

$$= \mathbb{E}\left(\mathbb{1}\{T(y^{rep} \mid \theta) \geq T(y \mid \theta)\}\right)$$

$$\approx \frac{1}{N} \sum_{s=1}^{S} \mathbb{1}\{T(y^{rep(s)} \mid \theta^{(s)}) \geq T(y \mid \theta^{(s)})\}$$

① case 1) prior on Σ_β as (b)

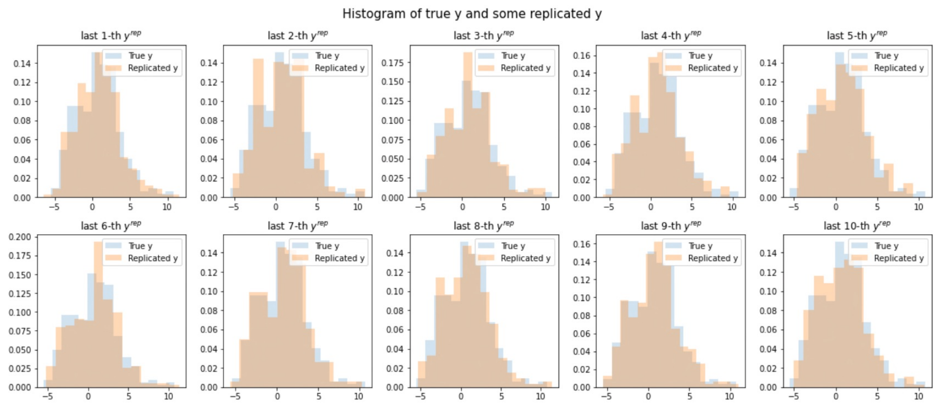
- Posterior predictive p-value

It seems that Gibbs sampler did not work well and resulted in an extreme value.

```
Posterior predictive p-value: 1.000
```

- Graphical comparison

Although the predictive posterior p-value seems strange, the distribution of y^{rep} seems to resemble the dist of actual y.



② case 2) prior on $\sigma^2_{\beta,k}$ as (c)

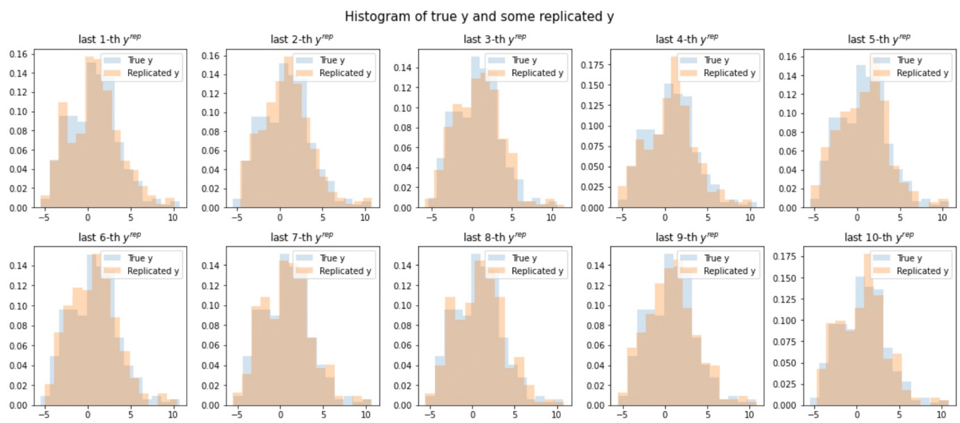
- Posterior predictive p-value

It seems that Gibbs sampler did not work well and resulted in an extreme value like previous case 1.

```
Posterior predictive p-value: 1.000
```

- Graphical comparison

Although the predictive posterior p-value seems strange, the distribution of y^{rep} seems to resemble with actual value y.



Q2-(a)

$y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, $\beta \in \mathbb{R}^p$, X_i : row vector $\in \mathbb{R}^p$, $\eta_i = X_i\beta$

$$\text{loglikelihood } L(y_i|X_i,\beta,\phi) \approx L(y_i|X_i,\beta,\phi)\big|_{\beta=\hat\beta} + \cancel{(\beta-\hat\beta)\frac{\partial}{\partial\beta}L(y_i|X_i,\beta,\phi)\big|_{\beta=\hat\beta}} \to 0 \quad (\because \hat\beta : \text{MLE})$$

$$+ \frac{1}{2}(\beta-\hat\beta)'\frac{\partial^2}{\partial\beta\partial\beta'}L(y_i|X_i,\beta,\phi)\big|_{\beta=\hat\beta}(\beta-\hat\beta) + \cancel{\text{Reminder}} \to \text{ignorable}$$

$$\approx L(y_i|X_i,\hat\beta,\phi) + \frac{1}{2}(\beta-\hat\beta)'\frac{\partial^2}{\partial\beta\partial\beta'}L(y_i|X_i,\beta,\phi)\big|_{\beta=\hat\beta}(\beta-\hat\beta)$$

$$p(y|X,\beta,\phi) = \prod_{i=1}^n \exp\left(L(y_i|X_i,\beta,\phi)\right)$$

$$= \prod_{i=1}^n \exp\Big(\underbrace{L(y_i|X_i,\hat\beta,\phi)}_{\text{ignorable } (\because \beta \text{ is not included})} + \frac{1}{2}(\beta-\hat\beta)'\frac{\partial^2}{\partial\beta\partial\beta'}L(y_i|X_i,\beta,\phi)\big|_{\beta=\hat\beta}(\beta-\hat\beta)\Big)$$

$$\propto N(\beta|\hat\beta, V)$$

$$V_i^{-1} = \frac{\partial^2}{\partial\beta\partial\beta'}L(y_i|X_i,\beta,\phi)\big|_{\beta=\hat\beta}$$

$$= -\frac{\partial\eta_i}{\partial\beta}\left[\frac{\partial^2}{\partial\eta_i^2}L(y_i|\eta_i,\phi)\big|_{\eta_i=\hat\eta_i}\right]\left(\frac{\partial\eta_i}{\partial\beta}\right)' \qquad \left(\because \frac{\partial L_i}{\partial\beta} = \frac{\partial L_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta},\ \hat\eta_i = X_i\hat\beta\right)$$

$$= -X_i'\left[\frac{\partial^2}{\partial\eta_i^2}L(y_i|\eta_i,\phi)\big|_{\eta_i=\hat\eta_i}\right]X_i \qquad \left(\because \frac{\partial\eta_i}{\partial\beta} = X_i'\right)$$

$$= X_i'\left[-L''(y_i|\eta_i,\phi)\right]X_i$$

$$\Rightarrow V_i = \left[X_i'(-L''(y_i|\eta_i,\phi))X_i\right]^{-1}$$

$$\therefore V = \left(\left[X_1' \cdots X_n'\right]\operatorname{diag}(-L''(y_i|\eta_i,\phi))\begin{bmatrix}X_1\\ \vdots \\ X_n\end{bmatrix}\right)^{-1}$$

$$= \left[X'\operatorname{diag}(-L''(y_i|\eta_i,\phi))X\right]^{-1} \text{ where } L''(y_i|\eta_i,\phi) = \frac{d^2}{d\eta_i^2}L(y_i|\eta_i,\phi)\big|_{\eta_i=\hat\eta_i}$$

Q2-(b)

Note that

$$\frac{d}{d\eta_i}\Phi(\eta_i) = \phi(\eta_i) \quad [\text{pdf of standard normal dist}]$$

$$\frac{d^2}{d\eta_i^2}\Phi(\eta_i) = \frac{d}{d\eta_i}\phi(\eta_i) = -\eta_i\phi(\eta_i)$$

Likelihood $\quad p(y|\beta) = \prod_{i=1}^n \binom{n_i}{y_i}[\Phi(\eta_i)]^{y_i}[1-\Phi(\eta_i)]^{n_i-y_i}$

Log likelihood $\quad L(y|\beta) = \sum_{i=1}^n L(y_i|\beta)$

$$L(y_i|\beta) \propto y_i\log(\Phi(\eta_i)) + (n_i-y_i)\log(1-\Phi(\eta_i))$$

$$L'(y_i|\hat\eta_i) \propto y_i\Phi(\eta_i)^{-1}\phi(\eta_i) - (n_i-y_i)(1-\Phi(\eta_i))^{-1}\phi(\eta_i) \leftarrow \frac{d}{d\eta_i}\Phi(\eta_i)$$

$$= y_i\frac{\phi(\eta_i)}{\Phi(\eta_i)} - (n_i-y_i)\frac{\phi(\eta_i)}{1-\Phi(\eta_i)}$$

$$L''(y_i|\hat\eta_i) \propto y_i\frac{\phi'(\eta_i)\Phi(\eta_i) - \phi(\eta_i)\Phi'(\eta_i)}{\Phi(\eta_i)^2} - (n_i-y_i)\frac{\phi'(\eta_i)(1-\Phi(\eta_i)) - \phi(\eta_i)[1-\Phi(\eta_i)]'}{(1-\Phi(\eta_i))^2}$$

$$= y_i\frac{-\eta_i\phi(\eta_i)\Phi(\eta_i) - \phi(\eta_i)^2}{\Phi(\eta_i)^2} - (n_i-y_i)\frac{\phi(\eta_i)^2 + \eta_i\phi(\eta_i)(1-\Phi(\eta_i))}{(1-\Phi(\eta_i))^2}\Bigg|_{\eta_i=\hat\eta_i}$$

Express in terms of $\hat\beta$

$$L''(y_i|\hat\eta_i) = y_i\frac{-(X_i\hat\beta)\phi(X_i\hat\beta)\Phi(X_i\hat\beta) - \phi(X_i\hat\beta)^2}{\Phi(X_i\hat\beta)^2} + (n_i-y_i)\frac{\phi(X_i\hat\beta)^2 + (X_i\hat\beta)\phi(X_i\hat\beta)(1-\Phi(X_i\hat\beta))}{(1-\Phi(X_i\hat\beta))^2}$$

Then V can expressed as

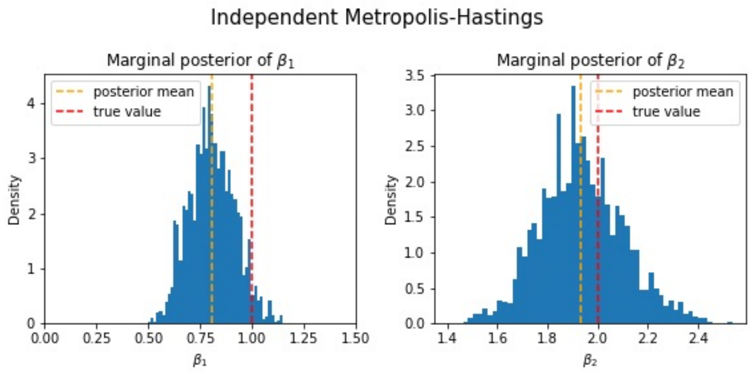
$$V = \left[X'\operatorname{diag}(-L''(y_i|\eta_i))X\right]^{-1}$$

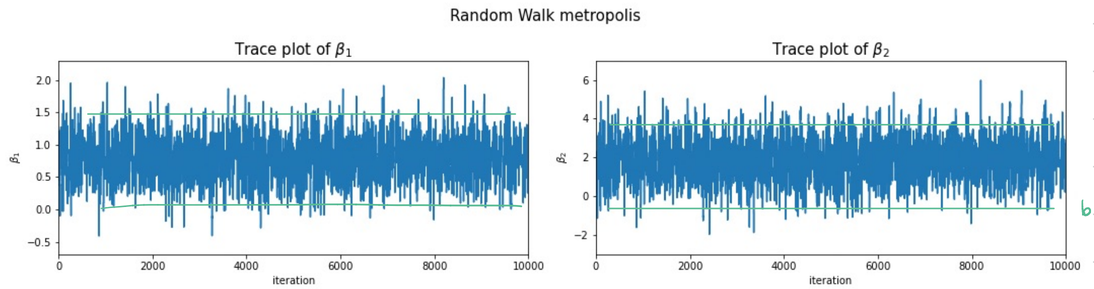
$$\text{where } L''(y_i|\hat\eta_i) = y_i\frac{-(X_i\hat\beta)\phi(X_i\hat\beta)\Phi(X_i\hat\beta) - \phi(X_i\hat\beta)^2}{\Phi(X_i\hat\beta)^2} + (n_i-y_i)\frac{\phi(X_i\hat\beta)^2 + (X_i\hat\beta)\phi(X_i\hat\beta)(1-\Phi(X_i\hat\beta))}{(1-\Phi(X_i\hat\beta))^2}$$

Q2-(c)

① Independent Metropolis - Hastings

- proposal : $J_t(\beta^* | \beta^{t-1}) = J_t(\beta^*) \propto N(\beta^* | \hat{\beta}_{pos}, V_{pos})$
- posterior : $p(\beta | \phi, y) \propto p(y | \beta, \phi)\, p(\beta | \phi) \propto \prod_{i=1}^{n} (\Phi(\eta_i))^{y_i} (1 - \Phi(\eta_i))^{n_i - y_i}$
- Acceptance prob : $\min\left(1, \dfrac{p(\beta^* | \phi, y)\, N(\beta^{t-1} | \hat{\beta}_{pos}, V_{pos})}{p(\beta^{t-1} | \phi, y)\, N(\beta^* | \hat{\beta}_{pos}, V_{pos})}\right)$
- posterior mean : [0.8 1.9] (cf. true value is [1 2])
- Histogram and trace plot

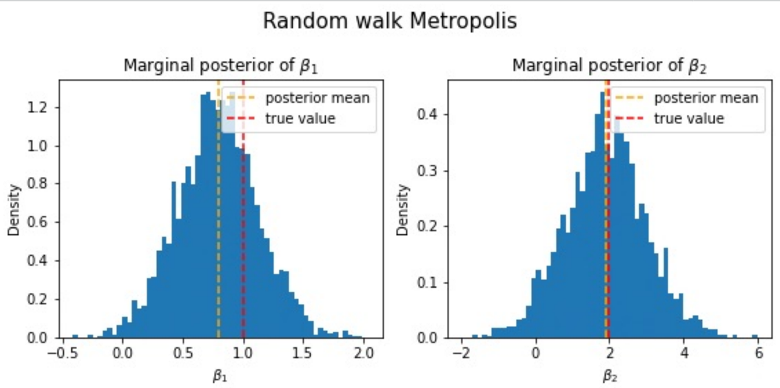


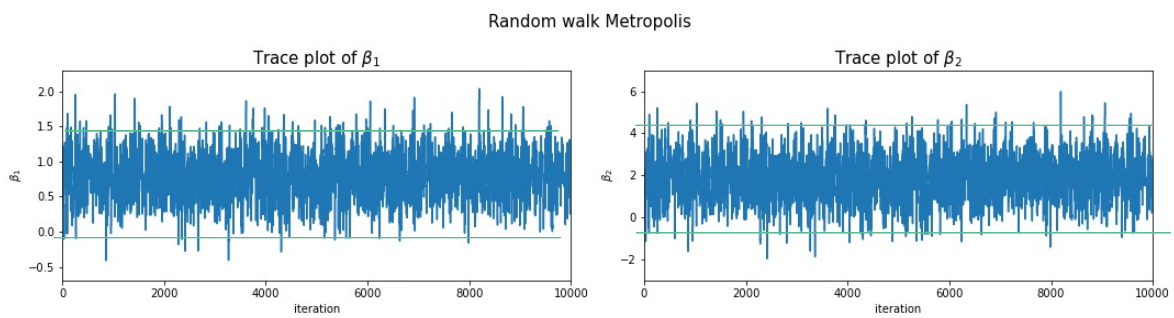


bounded

② Random walk Metropolis

- Proposal : $J_t(\beta^* | \beta^{t-1}) \propto N(\beta^* | \beta^{t-1}, (2.38^2/d) V_{pos})$
- posterial : $p(\beta | \phi, y) \propto N(\beta | \hat{\beta}_{pos}, V_{pos})$
- Acceptance prob : $\min\left(1, \dfrac{p(\beta^* | \phi, y)}{p(\beta^{t-1} | \phi, y)}\right)$
- posterior mean : [0.79 1.8] (cf. true value is [1 2])
- Histogram and trace plot





bounded

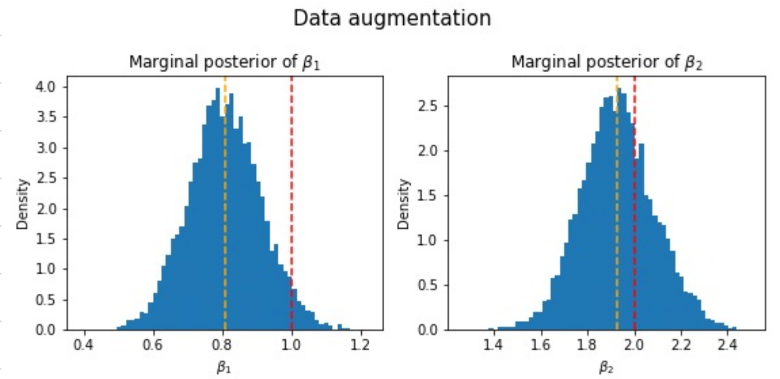
② Data Augmentation

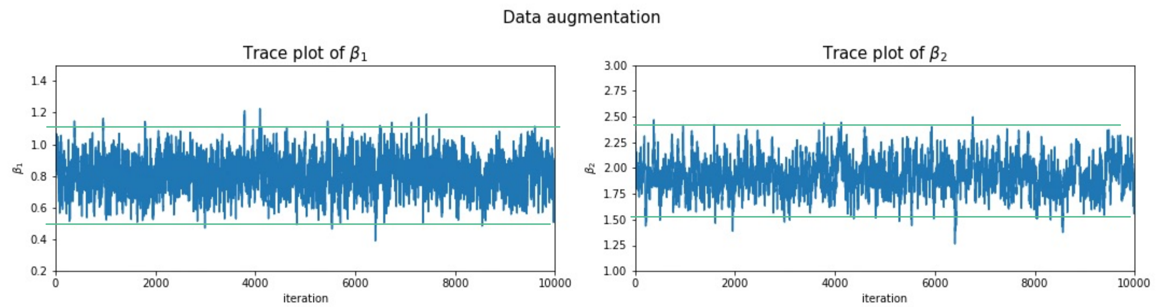
· Gibbs sampling i) $\beta \mid u, y \sim N\left((X'X)^{-1}X'u,\ (X'X)^{-1}\right)$

ii) $u_i \mid \beta, y \sim \begin{cases} TN_{(0,\infty)}(X_i\beta,\ 1) & y_i = 1 \\ TN_{(-\infty,0)}(X_i\beta,\ 1) & y_i = 0 \end{cases}$

· posterior mean : [0.8 1.9] (cf. true value is [1 2])

· Histogram and trace plot





bounded