MA 578 — Bayesian Statistics
Fall 2019

Final Exam

Assigned: Tuesday 12/10/19, **Due**: Monday 12/16/19 at *noon*

1. Suppose you observe a zero-mean unit-variance random n-dimensional vector $\mathbf{y} \,|\, \rho \sim N(0, R(\rho))$, with an equi-correlation structure parameterized by $\rho \in (-1, 1)$:

$$R(\rho) = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix} = (1-\rho)I_n + \rho \mathbb{1}_n \mathbb{1}_n^\top.$$

   (a) Using the Sherman-Morrison identity and the matrix determinant lemma, show that

$$R(\rho)^{-1} = \frac{1}{1-\rho}\left(I_n - \frac{\rho}{1+(n-1)\rho}\mathbb{1}_n\mathbb{1}_n^\top\right) \quad \text{and} \quad |R(\rho)| = (1-\rho)^{n-1}(1+(n-1)\rho),$$

   and so argue that we actually need $\rho \in (-1/(n-1), 1)$. In this case, it is better to parameterize on a simpler scale:

$$\theta = \frac{\rho + 1/(n-1)}{1 + 1/(n-1)} = \frac{1}{n}(1+(n-1)\rho) \in (0,1).$$

   (b) Show that under this new parameterization the likelihood is

$$\mathbb{P}(\mathbf{y} \,|\, \theta) \propto \Big[\theta(1-\theta)^{n-1}\Big]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{(n-1)s^2(\mathbf{y})}{1-\theta} + \frac{\bar{y}^2}{\theta}\right]\right\},$$

   where $\bar{y} = \sum_{i=1}^n y_i/n = \mathbb{1}_n^\top \mathbf{y}/n$ and $s^2(\mathbf{y}) = \sum_{i=1}^n (y_i - \bar{y})^2/n = (\mathbf{y} - \mathbb{1}_n\bar{y})^\top(\mathbf{y} - \mathbb{1}_n\bar{y})/n$. *Hint:* expand $\mathbf{y}^\top\mathbf{y} = (\mathbf{y} - \mathbb{1}_n\bar{y} + \mathbb{1}_n\bar{y})^\top(\mathbf{y} - \mathbb{1}_n\bar{y} + \mathbb{1}_n\bar{y})$.

   (c) Give a conjugate prior for θ and, using the fact that $\mathbb{E}[\bar{y}^2] = \theta$ and $\mathbb{E}[s^2(\mathbf{y})] = 1 - \theta$, show that Jeffreys prior for θ is

$$\mathbb{P}(\theta) \propto \left[\frac{n-1}{(1-\theta)^2} + \frac{1}{\theta^2}\right]^{\frac{1}{2}}.$$

2. You observe J subjects independently and want to linearly regress their data $\mathbf{y}_j$ using a set of predictors X_j, for $j = 1, \dots, J$. However, instead of the usual assumption of independence in the observations for each subject, entries in each $\mathbf{y}_j$ are correlated:

$$\mathbf{y}_j \mid \beta, \sigma^2, \rho \overset{\text{ind}}{\sim} N\big(X_j\beta, \sigma^2 R(\rho)\big), \qquad j = 1, \dots, J.$$

   (a) Assuming the same semi-conjugate priors we discussed in class, $\beta \sim N(\beta_0, \Sigma_0)$ and $\sigma^2 \sim \textsf{Inv-}\chi^2(\nu, \tau^2)$, and Jeffreys prior for ρ, design a Gibbs sampler to infer the joint posterior on these parameters. For the conditional posterior step on ρ, sample θ using numerical integration based on a grid.

   (b) Use your Gibbs sampler to analyze the `stroke` dataset[1] by regressing subject `scores` on `week` using non-informative priors for β, σ, and ρ and summarize your findings. Compare your results to estimates for β, σ, and ρ from a mixed-effects model with the same equi-correlation structure using package `lme4`.

   (c) Conduct posterior predictive checks and outlier analysis using *uncorrelated* residuals.

   (d) Now regress `scores` on `week` and `group`, including an interaction, and informally test for a differential `group` effect on `week` slope, that is, compare two models,

$$\begin{aligned} H_0 &: \texttt{score} \sim \texttt{week} + \texttt{group} \\ H_1 &: \texttt{score} \sim \texttt{week} + \texttt{group} + \texttt{group:week} \end{aligned}$$

   by checking if the coefficients for the `group:week` interaction are significant in H_1.

   [*] Perform a formal comparison using a Bayes factor. You should be able to obtain $\mathbb{P}(\mathbf{y} \mid \rho, H_0)$ and $\mathbb{P}(\mathbf{y} \mid \rho, H_1)$ in closed form, but then you need to use numerical integration to marginalize ρ out.

Instructions

There are two questions in this exam. Please complete the two questions, being sure to show all your work. Read each question carefully and be sure to answer all components of each question. You may use the course textbook and your class notes, including R codes. You may NOT use any other sources. You may NOT discuss the material on this exam with anybody else before submitting it.

At the end of the exam, please copy down the following statement and sign your name:

> *I confirm that I have followed the instructions for this exam and have not discussed the problems on this exam or their solutions with anyone.*

[1] Check `r-session9-hierlinear`.

1.(a) Sherman-Morrison formula:

$$(A+UV^T)^{-1} = A^{-1} - \frac{A^{-1}UV^TA^{-1}}{1+V^TA^{-1}U}$$

$$\therefore \quad (1-\rho)I_n + \rho 1^{n\times1}\cdot 1^{1\times n} = (1-\rho)^{-1}I_n - \frac{(1-\rho)^{-1}I_n\cdot\rho\cdot 1^{n\times1}\cdot 1^{1\times n}(1-\rho)^{-1}I_n}{1+1^{1\times n}(1-\rho)^{-1}I_n\cdot\rho\cdot 1^{n\times1}}$$

$$= (1-\rho)^{-1}\left[I_n - \frac{\cdot\rho\cdot(1-\rho)^{-1}}{1+1^{1\times n}(1-\rho)^{-1}\cdot\rho\cdot 1^{n\times1}}\cdot 1^{n\times1}\cdot 1^{1\times n}\right]$$

$$= (1-\rho)^{-1}\left[I_n - \frac{\rho}{1-\rho+1^{1\times n}\cdot\rho\cdot 1^{n\times1}}\cdot 1^{n\times1}1^{1\times n}\right]$$

$$= (1-\rho)^{-1}\left[I_n - \frac{\rho}{1-\rho+n\rho}\cdot 1^{n\times1}\cdot 1^{1\times n}\right]$$

$$= (1-\rho)^{-1}\left[I_n - \frac{\rho}{1+(n-1)\rho}\cdot 1^{n\times1}\cdot 1^{1\times n}\right]$$

Matrix determinant lemma:

$$|A+UV^T| = (1+V^TA^{-1}U)|A|$$

$$\therefore \quad |R(\rho)| = [1+1^{1\times n}(1-\rho)^{-1}I_n\cdot\rho\cdot 1^{n\times1}](1-\rho)^n$$

$$= (1-\rho)^{-1}[1-\rho+1^{1\times n}\cdot I_n\cdot\rho^{n\times1}](1-\rho)^n$$

$$= [1-\rho+n\rho](1-\rho)^{n-1}$$

$$= [1+(n-1)\rho](1-\rho)^{n-1}$$

Because $R(\rho)$ is varance matrix.

$$\therefore \quad \begin{cases} 1-\rho>0 \\ 1+(n+1)\rho>0 \end{cases} \Rightarrow -\frac{1}{n-1}<\rho<1 \Rightarrow \rho\in\left(\frac{-1}{n-1},1\right).$$

(b) because $\frac{\partial\theta}{\partial\rho}$ is a constant.

$$\therefore \quad p(y|\theta)\propto p(y|\rho)$$

$$p(y|\rho)\propto \exp\left\{-\frac{1}{2}(y-0)^T\cdot R(\rho)^{-1}(y-0)\right\}$$

plug in $R(\rho)^{-1}$

$$= |R(\rho)|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}y^T\cdot\left(I_n-\frac{\rho}{1+(n-1)\rho}\cdot 1^{n\times1}1^{1\times n}\right)\cdot y\right\}$$

$$= [\theta(1-\theta)^{n-1}]^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[\frac{y^Ty-y^T\cdot\frac{\rho}{1+(n-1)\rho}\cdot 1^{n\times1}\cdot 1^{1\times n}\cdot y}{1-\rho}\right]\right\}$$

$$y^T \cdot y = (y-\bar{y}+\bar{y})^T (y-\bar{y}+\bar{y})$$
$$= (y-\bar{y})^T (y-\bar{y}) + (y-\bar{y})^T \bar{y} + \bar{y}^T (y-\bar{y}) + \bar{y}^T \bar{y} \qquad y^T \cdot 1^{n\times 1} \cdot 1^{1\times n} \cdot y = (\Sigma y_i)^2$$
$$= \Sigma (y_i-\bar{y})^2 + \Sigma \bar{y}^2$$

$$\therefore \left[\theta(1-\theta)^{n-1}\right]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{y^T y - y^T \cdot \frac{\rho}{1+(n-1)\rho} \cdot 1^{n\times 1} \cdot 1^{1\times n} \cdot y}{1-\rho}\right]\right\}$$

$$= \left[\theta(1-\theta)^{n-1}\right]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{\Sigma(y_i-\bar{y})^2 + \Sigma\bar{y}^2 - \frac{\rho}{1+(n-1)\rho}(\Sigma y_i)^2}{1-\rho}\right]\right\}$$

☆: $1+(n-1)\rho = n\theta$

$$= \left[\theta(1-\theta)^{n-1}\right]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{\frac{\Sigma(y_i-\bar{y})^2}{n} + \frac{\Sigma\bar{y}^2}{n} - \frac{\rho}{n^2\theta}(\Sigma y_i)^2}{\frac{1-\theta}{n-1}}\right]\right\}$$

☆ $1-\rho = \frac{1-\theta}{n-1}$

$$= \left[\theta(1-\theta)^{n-1}\right]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{(n-1)S^2 + (n-1)(\theta-\rho)\frac{\bar{y}^2}{\theta}}{1-\theta}\right]\right\}$$

☆: $\frac{\Sigma\bar{y}^2}{n} = \bar{y}^2$.

$$= \left[\theta(1-\theta)^{n-1}\right]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{(n-1)S^2}{1-\theta} + \frac{(n-1)(\theta-\rho)}{1-\theta}\frac{\bar{y}^2}{\theta}\right]\right\}.$$

as for $\frac{(n-1)(\theta-\rho)}{1-\theta}$, $\quad \because \theta = \frac{1}{n}[1+(n-1)\rho]$

$$n\theta = 1+(n-1)\rho$$
$$n\theta = 1+n\rho-\rho$$
$$n\theta - n\rho + \rho = 1$$
$$n\theta - \theta - n\rho + \rho = 1-\theta$$
$$\theta(n-1) - \rho(n-1) = 1-\theta$$
$$(n-1)(\theta-\rho) = 1-\theta$$
$$\frac{(n-1)(\theta-\rho)}{1-\theta} = 1.$$

$$\therefore \left[\theta(1-\theta)^{n-1}\right]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{(n-1)S^2}{1-\theta} + \frac{(n-1)(\theta-\rho)}{1-\theta}\frac{\bar{y}^2}{\theta}\right]\right\}.$$

$$= \left[\theta(1-\theta)^{n-1}\right]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{(n-1)S^2}{1-\theta} + \frac{\bar{y}^2}{\theta}\right]\right\} \quad \checkmark.$$

(c). $P(y|\theta) \propto [\theta(1-\theta)^{n-1}]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{(n-1)s^2}{1-\theta} + \frac{\bar{y}^2}{\theta}\right]\right\}$

let $P(\theta) \propto [\theta(1-\theta)^{u}]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{u\tau^2}{1-\theta} + \frac{\mu^2}{\theta}\right]\right\}$

$$P(\theta|y) \propto [\theta^2(1-\theta)^{n+u-1}]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{(n-1)s^2+u\tau^2}{1-\theta} + \frac{\bar{y}^2+\mu^2}{\theta}\right]\right\}$$

Jefery's Prior

$$P(y|\theta) \propto [\theta(1-\theta)^{n-1}]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{(n-1)s^2}{1-\theta} + \frac{\bar{y}^2}{\theta}\right]\right\}$$

$$l = \log[P(y|\theta)] = -\frac{1}{2}\left[\log\theta + (n-1)\log(1-\theta) + \frac{(n-1)s^2}{1-\theta} + \frac{\bar{y}^2}{\theta}\right]$$

$$\frac{\partial l}{\partial\theta} = -\frac{1}{2}\left[\frac{1}{\theta} - \frac{n-1}{1-\theta} + \frac{(n-1)s^2}{(1-\theta)^2} - \frac{\bar{y}^2}{\theta^2}\right]$$

$$\frac{\partial l}{\partial\theta\partial\theta} = -\frac{1}{2}\left[-\frac{1}{\theta^2} - \frac{n-1}{(1-\theta)^2} + \frac{2(n-1)s^2}{(1-\theta)^3} + \frac{2\bar{y}^2}{\theta^3}\right]$$

$$I(\theta) = \mathbb{E}\left(-\frac{\partial l}{\partial\theta^2}\right) = \frac{1}{2}\left[-\frac{1}{\theta^2} - \frac{n-1}{(1-\theta)^2} + \frac{2(n-1)(1-\theta)}{(1-\theta)^3} + \frac{2\theta}{\theta^3}\right]$$

$$= \frac{1}{2}\left[\frac{n-1}{(1-\theta)^2} + \frac{1}{\theta^2}\right]$$

$$\therefore P(\theta) \propto \left[\frac{n-1}{(1-\theta)^2} + \frac{1}{\theta^2}\right]^{\frac{1}{2}}$$

2. (a). Compare to normal linear OLS regression, which:

$$Y \sim N(X\beta, \sigma^2 I_n).$$

here we have $Y \sim N(X\beta, \sigma^2 R(\rho))$.

so if we can transform this into a normal situation, we can use a normal Gibbs Sampler:

$$Y \sim N(X\beta, \sigma^2 R(\rho))$$

$$P(\rho, Y, \beta, \sigma^2) \propto |\sigma^2 R(\rho)|^{-\frac{1}{2}} \cdot \exp\{-\frac{1}{2\sigma^2} \cdot (Y-X\beta)^T R(\rho)^{-1} (Y-X\beta)\}$$
$$\cdot \exp\{-\frac{1}{2}(\beta-\beta_0)^T \Sigma_0^{-1} (\beta-\beta_0)\}$$
$$\cdot (\sigma^2)^{-(\frac{\nu}{2}+1)} \exp\{-\frac{\nu \tau^2}{2\sigma^2}\}$$
$$\cdot \left[\frac{n-1}{1-\rho^2} + \frac{1}{\rho^2}\right]$$

$$\therefore P(\rho | Y, \beta, \sigma^2) \propto |\sigma^2 R(\rho)|^{-\frac{1}{2}} \cdot \exp\{-\frac{1}{2\sigma^2} \cdot (Y-X\beta)^T R(\rho)^{-1} (Y-X\beta)\}$$
$$\cdot \left[\frac{n-1}{1-\rho^2} + \frac{1}{\rho^2}\right]$$

$\therefore$ Our Gibbs Sample is:

① Sample ρ from $P(\rho | Y, \beta, \sigma^2)$, and calculate $R(\rho)$;

② Do Choleski Decomposition on $R(\rho)$ and $R(\rho) = C^T C$.

③ $\tilde{Y} = (C^T)^{-1} \cdot Y$, $\tilde{X} = (C^T)^{-1} X$. $(C^T)^{-1} Y \sim N[(C^T)^{-1} X\beta, (C^T)^{-1} R(\rho) \cdot C^{-1}]$

$\because R(\rho) = C^T C$. $\therefore (C^T)^{-1} \cdot R(\rho) C^{-1} = (C^T)^{-1} \cdot C^T \cdot C \cdot C^{-1} = I_n$.

$\therefore \tilde{Y} \sim N(\tilde{X} \cdot \beta, I_n)$.

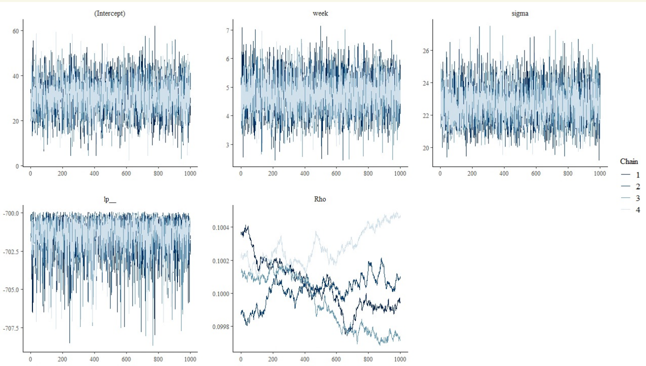
④ Use Gibbs Sample designed before to sample $\hat{\beta}$ and $\hat{\sigma}^2$.

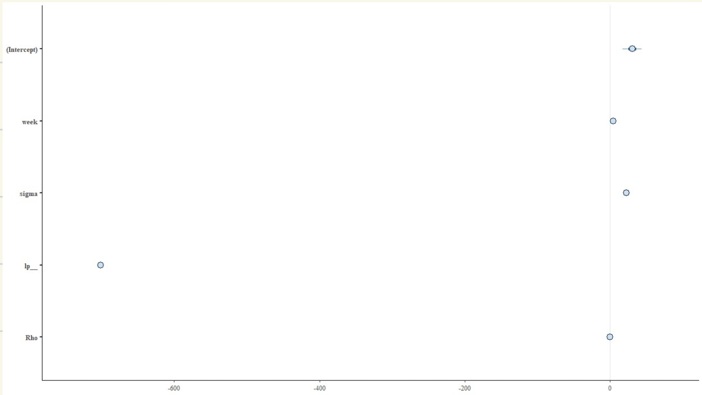
⑤ Sample new ρ from $P(\rho | Y, \hat{\beta}, \hat{\sigma}^2)$ and calculate new $R(\rho)$.

Repeat ②-⑤.

(b). Because the prior for β is $N(\beta_0, \Sigma_0)$, and prior for σ^2 is $Inv.\chi^2(\nu, \tau^2)$. so I just set variance of both prior to a relatively large number to approximate non-informative flat prior, below is the result:

	Q5 <dbl>	Q50 <dbl>	Q95 <dbl>	Mean <dbl>	SD <dbl>	Rhat <dbl>	Bulk_ESS <dbl>	Tail_ESS <dbl>
(Intercept)	17.6	30.9	44.0	30.8	8.1	1.00	1904	1955
week	3.6	4.8	5.9	4.8	0.7	1.00	1921	1852
sigma	20.8	22.7	24.7	22.7	1.2	1.00	1789	1851
lp__	-703.8	-701.2	-700.1	-701.5	1.2	1.00	1692	1788
Rho	0.1	0.1	0.1	0.1	0.0	3.04	5	11





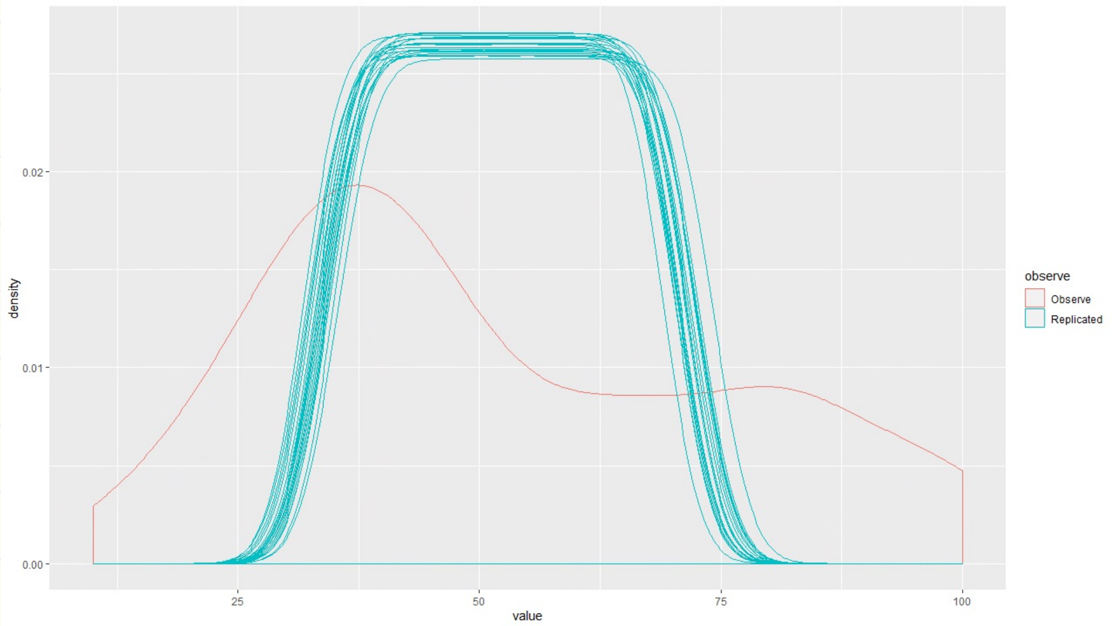
we can see that β, σ^2 and log-likelihood converge well, but ρ does not converge at all

Below is result from build-in function:

```
Coefficients:
             Estimate Naive S.E.   Naive z Robust S.E. Robust z
(Intercept) 31.034226  4.2200929  7.35392   4.1024357 7.564829
week         4.729663  0.2867134 16.49613   0.6226925 7.595503
```

we can see that take mean as point estimator, the result of Gibbs Sampler is quite close to build-in function, but due the poor convergence of ρ, the standard-deviation is larger than build-in function.

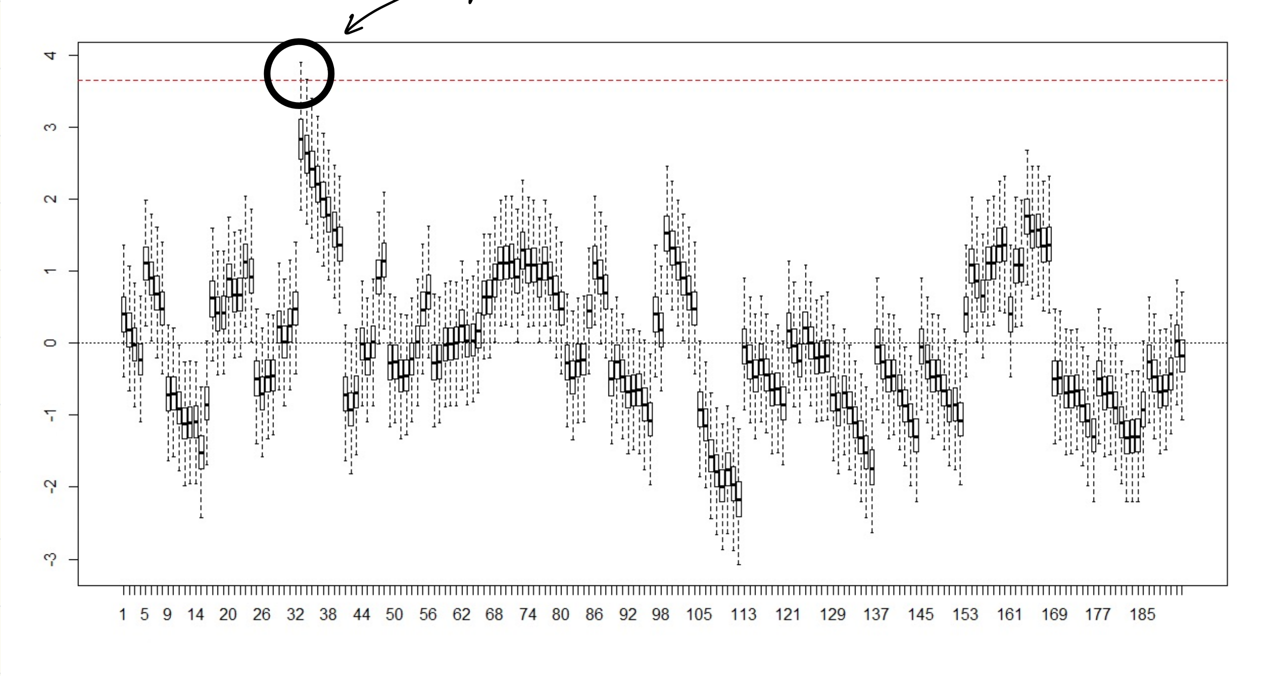
(c). Posterior Check.



Out model captured some trend of the data.
but generally speaking, it fits poor.

Outliers Detect:

potential outlier.



(d) model selection:

Bayes Factor.