STA 365 Homework #2 Ruike Xu.

Problem 1. (10 points)

The following matrix represents the transition matrix for a random walk on the integers $\{1, 2, 3, 4, 5\}$:

$$P = \begin{bmatrix} .2 & .8 & 0 & 0 & 0 \\ .2 & .2 & .6 & 0 & 0 \\ 0 & .2 & .6 & .2 & 0 \\ 0 & 0 & .6 & .2 & .2 \\ 0 & 0 & 0 & .8 & .2 \end{bmatrix}$$

(a) Suppose one starts walking at the state value 4. Find the probability of landing at each location after a single step.

(b) Starting at state value 4, find the probability of landing at each location after three steps.

(c) Explain what it means for this Markov Chain to be irreducible and aperiodic.

(a) Since the Person starts walking at the state value 4,

this Person's Current Location Probability is $p = (0, 0, 0, 1, 0)$

$$p^{(1)} = p^{(0)} \cdot P = [0, 0, 0, 1, 0] \begin{bmatrix} 0.2 & 0.8 & 0 & 0 & 0 \\ 0.2 & 0.2 & 0.6 & 0 & 0 \\ 0 & 0.2 & 0.6 & 0.2 & 0 \\ 0 & 0 & 0.6 & 0.2 & 0.2 \\ 0 & 0 & 0 & 0.8 & 0.2 \end{bmatrix}$$

$$= [0, 0, 0.6, 0.2, 0.2]$$

(b) $$p^{(3)} = p^{(0)} \cdot P^3 = [0, 0, 0, 1, 0] \begin{bmatrix} 0.2 & 0.8 & 0 & 0 & 0 \\ 0.2 & 0.2 & 0.6 & 0 & 0 \\ 0 & 0.2 & 0.6 & 0.2 & 0 \\ 0 & 0 & 0.6 & 0.2 & 0.2 \\ 0 & 0 & 0 & 0.8 & 0.2 \end{bmatrix}^3$$

$$= [0, 0, 0.6, 0.2, 0.2] \cdot P^2$$

$$= [0, 0.12, 0.48, 0.32, 0.08] \cdot P$$

$$= [0.024, 0.12, 0.552, 0.224, 0.08]$$

(c) This Markov Chain is irreducible because it's possible to go from any state to any other state. We can observe from the transition matrix that any state can go to other state given enough steps. We can see that $P(X_{n_{ij}} = j \mid X_0 = i) = P_{ij}^{(n_{ij})} > 0$ for all states i and j. States are communicating. This Markov chain is aperiodic because any state can move back to its own state in one step with probability higher than 0. We can see that from the diagonal of the transition matrix, it's possible for any states to go back to its own state in one step because the probability is higher than 0.

STA365_homework2_code

Ruike Xu

3/29/2022

Question 2

```r
# install.packages("rjags")
# install.packages("coda")

set.seed(1006562550)
library(coda)
```

```
## Warning: package 'coda' was built under R version 4.1.3
```

```r
library(R2jags)
```

```
## Warning: package 'R2jags' was built under R version 4.1.3
```

```
## Loading required package: rjags
```

```
## Warning: package 'rjags' was built under R version 4.1.3
```

```
## Linked to JAGS 4.3.0
```

```
## Loaded modules: basemod,bugs
```

```
##
## Attaching package: 'R2jags'
```

```
## The following object is masked from 'package:coda':
##
##     traceplot
```

```r
library(lattice)
```

$$Warren: \ \ \tau(1-\theta)+\theta(1-\lambda)$$

$$Trump(\psi): \ \ \lambda\theta+(1-\tau)(1-\theta)$$

```
# Model construction with likelihood function and prior
model.JAGS = function(){
    y ~ dbinom(psi, n)
    theta ~ dbeta(3, 7)
    tau ~ dbeta(6, 4)
    lambda ~ dbeta(7, 3)
    psi <- theta * lambda + (1 - theta) * (1 - tau)
  }
```

The model we are constructing is intended to predict which candidate is likely to win the 2024 US presidential election (Donald Trump or Elizabeth Warren), which is based on binomial likelihood and beta prior. θ is the probability that the voter claims to vote for trump (might lying), thus, $1-\theta$ is the probability that the person claims to vote for Warren. λ is the percentage of the population that is actually willing to vote for Trump while τ is the percentage of the population that is actually willing to vote for Warren. We can detect the 'Quasi Liberal' and 'Quasi Republican' as $1-\lambda$ and $1-\tau$. To be more specific, the percentage of the population that claimed they would vote for Trump/Warren but ended voting the opposite.

Therefore, we can derive that the actual probability of a person voting for Trump as $\psi : \lambda\theta + (1-\tau)(1-\theta)$ and the actual probability of a person voting for Warren as $\tau(1-\theta) + \theta(1-\lambda)$

To define τ and λ, we would assume more Republican people are less willing to claim their feelings in Liberal states, so τ ~ Beta(6, 4) and λ ~ Beta(7,3)

```
# Simulating data
n = 100000
y = 30000
data.JAGS = list(y = y, n = n)
```

```
# Randomly select the initial values
inits.JAGS = function(){
  return(list(theta=rbeta(1, 3, 7),tau=rbeta(1, 6, 4),lambda=rbeta(1, 7, 3)))
}
```

```
# Select parameters that will be simulated with MCMC model
para.JAGS = c("theta", "tau", "lambda", "psi")
```

```
# Fit the MCMC model, number of iterations = 90000, burn in = 10000
fit.JAGS = jags(data=data.JAGS,inits=inits.JAGS,
                parameters.to.save = para.JAGS,
                n.chains=1,
                n.iter=90000,
                n.burnin=10000,
                model.file=model.JAGS)
```

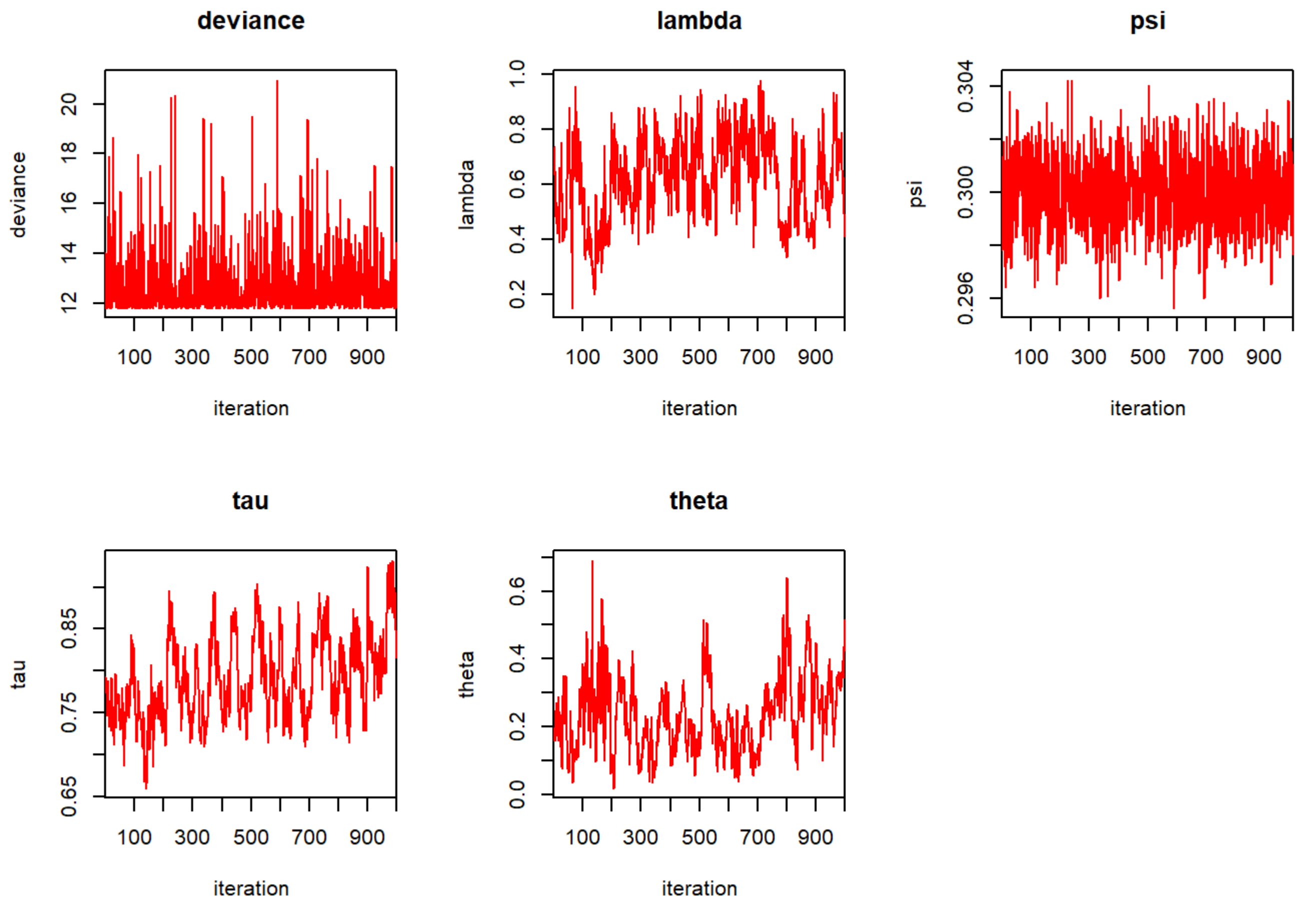
```
## module glm loaded
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 1
##    Unobserved stochastic nodes: 3
##    Total graph size: 15
##
## Initializing model
```

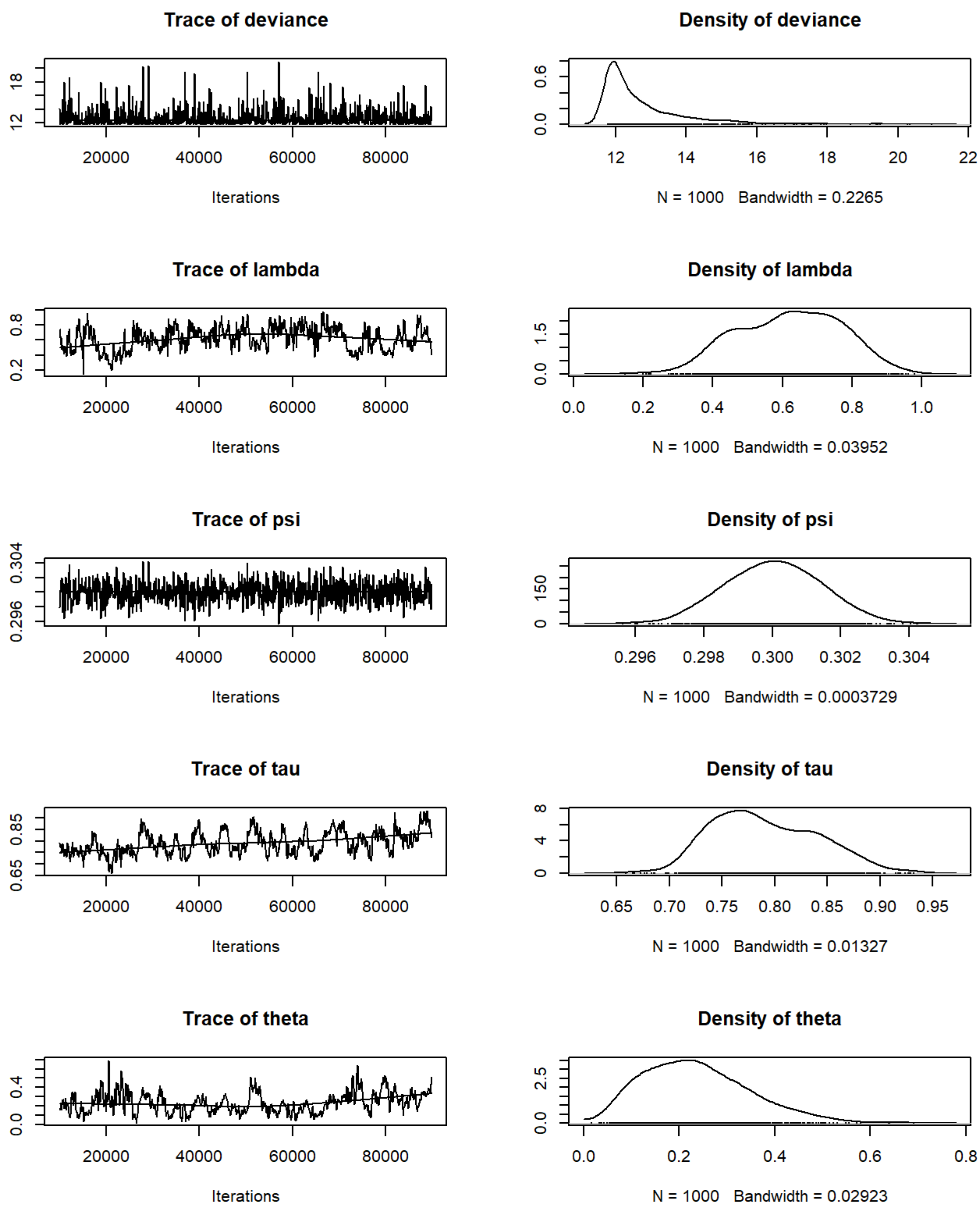
```
# Print model fit summary
print(fit.JAGS)
```

```
## Inference for Bugs model at "C:/Users/Ruike Xu/AppData/Local/Temp/Rtmpgtofxy/model862071d15d2
a.txt", fit using jags,
##  1 chains, each with 90000 iterations (first 10000 discarded), n.thin = 80
##  n.sims = 1000 iterations saved
##          mu.vect sd.vect   2.5%    25%    50%    75%  97.5%
## lambda     0.620   0.148  0.348  0.507  0.625  0.732  0.880
## psi        0.300   0.001  0.297  0.299  0.300  0.301  0.303
## tau        0.791   0.050  0.713  0.752  0.784  0.830  0.891
## theta      0.236   0.110  0.067  0.152  0.226  0.308  0.478
## deviance  12.724   1.272 11.791 11.890 12.245 13.030 16.160
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 0.8 and DIC = 13.5
## DIC is an estimate of expected predictive error (lower deviance is better).
```

```
# Generate traceplots
traceplot(fit.JAGS,mfrow=c(2,3),ask=FALSE)
```



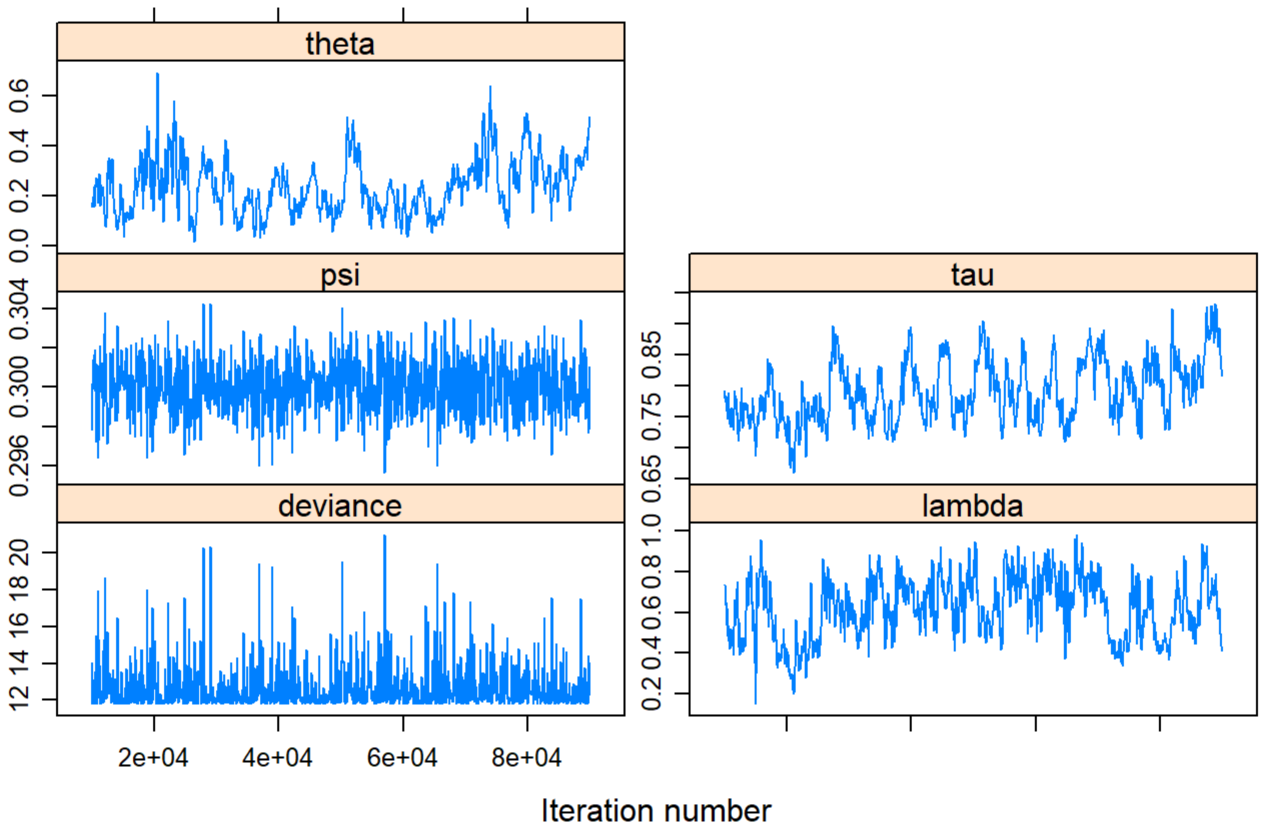

```
# Plot MCMC object to show all parameter densities along with traceplots
fit.JAGS.mcmc = as.mcmc(fit.JAGS)
plot(fit.JAGS.mcmc,ask=FALSE)
```


Trace of deviance
Iterations
Density of deviance
N = 1000 Bandwidth = 0.2265
Trace of lambda
Iterations
Density of lambda
N = 1000 Bandwidth = 0.03952
Trace of psi
Iterations
Density of psi
N = 1000 Bandwidth = 0.0003729
Trace of tau
Iterations
Density of tau
N = 1000 Bandwidth = 0.01327
Trace of theta
Iterations
Density of theta
N = 1000 Bandwidth = 0.02923

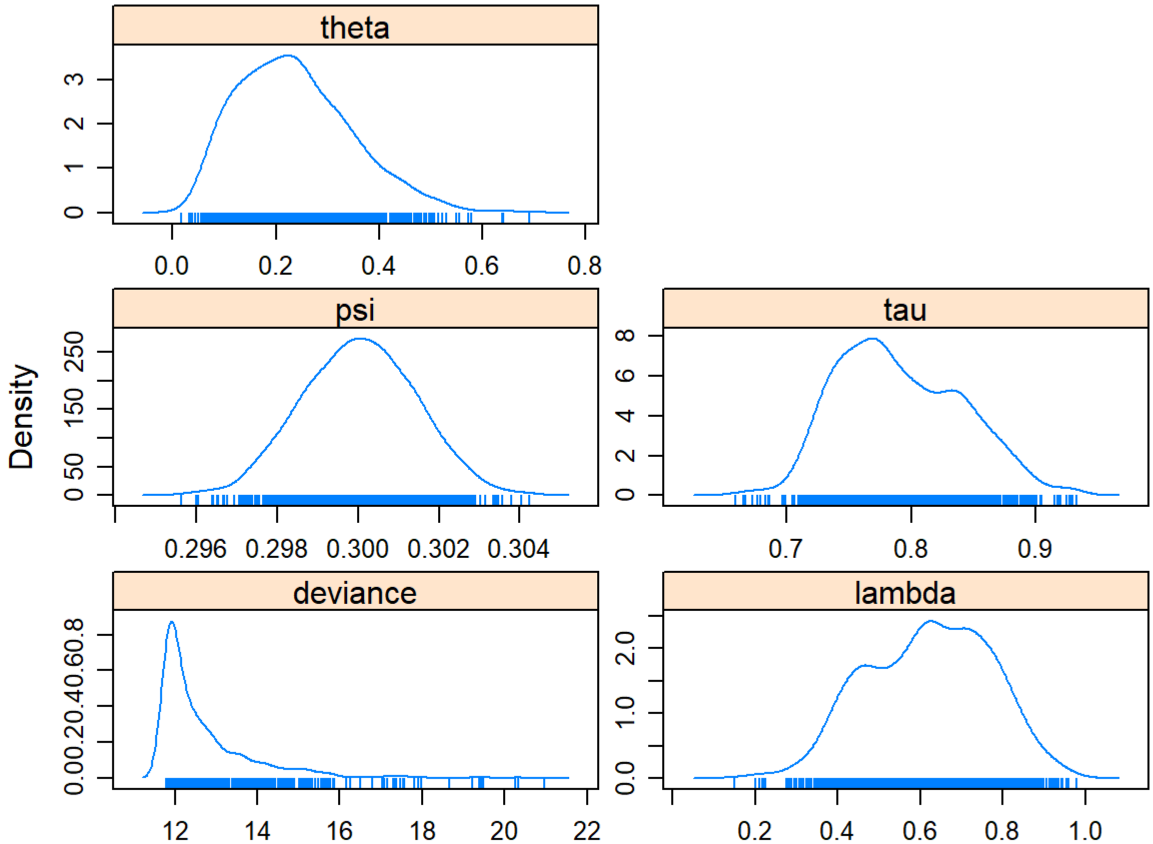
```r
# Summary as a MCMC object
summary(fit.JAGS.mcmc)
```

```
##
## Iterations = 10001:89921
## Thinning interval = 80
## Number of chains = 1
## Sample size per chain = 1000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##             Mean      SD  Naive SE Time-series SE
## deviance 12.7236 1.27180 4.022e-02      4.022e-02
## lambda    0.6201 0.14844 4.694e-03      2.407e-02
## psi       0.3000 0.00140 4.429e-05      4.429e-05
## tau       0.7912 0.04983 1.576e-03      8.174e-03
## theta     0.2364 0.10978 3.472e-03      1.788e-02
##
## 2. Quantiles for each variable:
##
##              2.5%     25%     50%     75%   97.5%
## deviance 11.79091 11.8904 12.2452 13.0302 16.1600
## lambda    0.34765  0.5065  0.6245  0.7321  0.8804
## psi       0.29735  0.2990  0.3000  0.3010  0.3027
## tau       0.71344  0.7523  0.7839  0.8302  0.8906
## theta     0.06721  0.1524  0.2263  0.3076  0.4782
```

```r
# Traceplots as a MCMC object
xyplot(fit.JAGS.mcmc,layout=c(2,3))
```




```
# Density plot as a MCMC object
densityplot(fit.JAGS.mcmc)
```



Part(d)

After we observe the results from part(c), the traceplots for the parameters seem fairly stationary, however, λ, τ, θ seems a bit less appropriate than ψ, but I think the variations are acceptable. When we identify the density plots, all parameters are in appropriate shapes as the priors we set up. Therefore, I think the model we constructed seems reasonable without any adjustments.

Part(e)

The estimated probability of Trump winning the 2024 US Election is 0.236, which is lower than 0.3. Since we set the prior that people that claimed to vote for Warren but vote for Trump are slightly higher than people who claimed to vote for Trump but vote for Warren, we would expect more Quasi Liberal to vote for Trump. However, the prior setting does not influence much due to the huge likelihood probability difference.

Question 3

Part(a)

$$\theta_1 = 1 + 0.065X_1 = 1 + 0.065 * 95 = 7.175$$
$$H_1 \sim Poisson(7.175)$$

```
# Probability of Poisson distribution for at least 2 infections
P_atleast2 <- 1 - ppois(1, 7.175)
P_atleast2
```

```
## [1] 0.9937422
```

Therefore, the probability that Indonesia observes at least 2 infections in the given time period is 0.994

Part(b)

$$\theta_2 = 1 + 0.065X_2 = 1 + 0.065 * 150 = 10.75$$
$$H_2 \sim Poisson(10.75)$$

```
# Probability of Poisson distribution for at least 28 infections
P_atleast28 <- 1 - ppois(27, 10.75)
P_atleast28
```

```
## [1] 8.378949e-06
```

Therefore, the probability that Singapore observes 28 or more infections in the given time period is approximately 0.

Problem 4. (20 points)

Suppose data $(y_1, ..., y_J)$ follow a multinomial distribution with parameters $(\theta_1, ..., \theta_J)$. Also, suppose that $\theta = (\theta_1, ..., \theta_J)$ has a Dirichlet prior distribution. Let $\alpha = \frac{\theta_1}{\theta_1+\theta_2}$.

(a) Write down the marginal posterior distribution for α.

(b) Show that this distribution is identical to the posterior distribution for α obtained by treating y_1 as an observation from the binomial distribution with probability α and sample size $y_1 + y_2$, ignoring the data $y_3, ..., y_n$.

(a) Suppose that the multinomial distribution with parameters $(\theta_1, \cdots\cdots \theta_J)$ is given by

$$P(Y|\theta) = \frac{n!}{y_1!\,y_2!\cdots\cdots y_J!}\,\theta_1^{y_1}\theta_2^{y_2}\cdots\cdots\theta_J^{y_J}$$

The Dirichlet prior distribution is

$$P(\theta|\alpha) = \frac{\Gamma(\alpha_1+\alpha_2+\cdots\cdots+\alpha_J)}{\prod_{i=1}^{J}\Gamma(\alpha_i)}\,\theta_1^{\alpha_1-1}\,\theta_2^{\alpha_2-1}\cdots\cdots\theta_J^{\alpha_J-1}$$

The posterior distribution is :

$$P(\theta|y) \propto P(\theta_1,\theta_2,\cdots\cdots,\theta_J|\alpha_1,\alpha_2,\cdots\alpha_J)\cdot P(Y|\theta)$$

$$\propto \prod_{i=1}^{J}\theta_i^{\alpha_i-1}\cdot\frac{(y_1+y_2+\cdots\cdots+y_J)!}{y_1!\,y_2!\cdots\cdots y_J!}\,\theta_1^{y_1}\theta_2^{y_2}\cdots\cdots\theta_J^{y_J}$$

$$\propto \theta_1^{y_1+\alpha_1-1}\,\theta_2^{y_2+\alpha_2-1}\cdots\cdots\theta_J^{y_J-\alpha_J-1}$$

We can observe that the posterior distribution based on Dirichlet prior and multinomial likelihood is Dirichlet $(y_1+\alpha_1, y_2+\alpha_2, \cdots\cdots, y_J+\alpha_J)$

In a model where a Dirichlet prior distribution is placed over a set of categorical-valued observations, the marginal joint distribution of the observations is also a Dirichlet-multinomial distribution. (Wikimedia, 2022)

Therefore, by the marginal joint Dirichlet distribution property,

$$P(\theta_1,\theta_2|Y) \propto \theta_1^{y_1+\alpha_1-1}\,\theta_2^{y_2+\alpha_2-1}(1-\theta_1-\theta_2)^{y_3+y_4+\cdots+y_J+\alpha_3+\alpha_4+\cdots\cdots+\alpha_J-1}$$

Let $\alpha = \frac{\theta_1}{\theta_1+\theta_2}$

if we define $\beta = \theta_1+\theta_2$

$$\theta_1 = h_1(\alpha,\beta) = \frac{\theta_1}{\theta_1+\theta_2}\cdot(\theta_1+\theta_2) = \alpha\beta$$

$$\theta_2 = h_2(\alpha,\beta) = \beta-\theta_1 = \beta-\alpha\beta$$

Here, we perform a Jacobian transformation for change of variable. (Evans & Rosenthal, 2022)

$$J(\alpha,\beta) = \det\begin{pmatrix}\frac{dh_1}{d\alpha} & \frac{dh_2}{d\alpha}\\ \frac{dh_1}{d\beta} & \frac{dh_2}{d\beta}\end{pmatrix}$$

$$= \det\begin{pmatrix}\beta & -\beta\\ \alpha & 1-\alpha\end{pmatrix}$$

$$= \beta-\alpha\beta+\alpha\beta = \beta$$

$$1-\theta_1-\theta_2 = 1-\alpha\beta-\beta+\alpha\beta = 1-\beta$$

$$P_{\alpha,\beta}(\alpha,\beta|y) \propto \beta\cdot(\alpha\beta)^{\alpha_1+y_1-1}(\beta-\alpha\beta)^{\alpha_2+y_2-1}(1-\beta)^{\alpha_3+\alpha_4+\cdots+\alpha_J+y_3+y_4+\cdots y_J-1}$$

$$\propto \beta\,\alpha^{\alpha_1+y_1-1}\cdot\beta^{\alpha_1+y_1-1}[\beta(1-\alpha)]^{\alpha_2+y_2-1}(1-\beta)^{\alpha_3+\alpha_4+\cdots+\alpha_J+y_3+y_4+\cdots+y_J-1}$$

$$\propto \beta\,\alpha^{\alpha_1+y_1-1}\beta^{\alpha_1+y_1-1}\beta^{\alpha_2+y_2-1}(1-\alpha)^{\alpha_2+y_2-1}(1-\beta)^{\alpha_3+\alpha_4+\cdots+\alpha_J+y_3+y_4+\cdots+y_J-1}$$

$$\propto \alpha^{\alpha_1+y_1-1}(1-\alpha)^{\alpha_2+y_1-1}\beta^{\alpha_1+y_1+\alpha_2+y_1-2}\cdot\beta\,(1-\beta)^{\alpha_3+\alpha_4+\cdots+\alpha_J+y_3+y_4+\cdots+y_J-1}$$

$$\propto \mathrm{Beta}(y_1+\alpha_1, y_2+\alpha_2)\cdot\mathrm{Beta}(y_1+\alpha_1+y_2+\alpha_2-1, y_3+y_4+\cdots+y_J+\alpha_3+\alpha_4+\cdots+\alpha_J-1)$$

Therefore, the marginal posterior distribution for α is $\mathrm{Beta}(y_1+\alpha_1, y_2+\alpha_2)$

(b) Since we are treating y_1 as an observation from a binomial distribution with probability α and sample size y_1+y_2.

$$f(y_1|\alpha) = \binom{y_1+y_2}{y_1}\alpha^{y_1}(1-\alpha)^{y_2}$$

Since we know that α is $\mathrm{Beta}(y_1+\alpha_1, y_2+\alpha_2)$ distribution from the previous part, to avoid any confusion by notation, let $\alpha \sim \mathrm{Beta}(m,n)$

$$P(\alpha|y_1) \propto P(y_1|\alpha)\,P(\alpha)$$

$$\propto \alpha^{y_1}(1-\alpha)^{y_2}\cdot\alpha^{m-1}(1-\alpha)^{n-1}$$

$$= \alpha^{y_1+m-1}(1-\alpha)^{y_2+n-1}$$

We can observe that the posterior distribution is $\mathrm{Beta}(y_1+m, y_2+n)$, which is identical to the marginal posterior distribution of previous part if we assume $\alpha \sim \mathrm{Beta}(\alpha_1,\alpha_2)$

Therefore, the distribution in part (a) is identical to the posterior distribution for α obtained by treating y_1 as an observation from binomial with probability α and sample size y_1+y_2.

Reference

Wikimedia Foundation. (2022, March 26). *Dirichlet distribution*. Wikipedia. Retrieved March 29, 2022, from https://en.wikipedia.org/wiki/Dirichlet_distribution#:~:text=In%20a%20model%20where%20a,is%20a%20Dirichlet%2Dmultinomial%20distribution.

Evans, M., & Rosenthal, J. (n.d.). *Probability and statistics - the science of uncertainty, Second edition*. Probability and Statistics - The Science of Uncertainty. Retrieved March 29, 2022, from http://www.utstat.toronto.edu/mikevans/jeffrosenthal/

Mavavilj. (2015, December 7). *How do you see a Markov chain is irreducible?* Cross Validated. Retrieved March 29, 2022, from https://stats.stackexchange.com/questions/186033/how-do-you-see-a-markov-chain-is-irreducible