Bayesian Personal Prep

Kalyango Jovan

This document contains five exercises of Bayesian statistics course, and detailed information about the differences between frequetists thinking and Bayesian thinking. In this document i will use JAGs and R coding styles, For JAGs all the supposed to be notepad codes will be used as comments but there is a number of data sets to be used as a way to apply the Bayesian thinking.

Tools /concepts

This section is describing all different tools used in both Frequentists and Bayesian statistics, but also highlighting the differences between the opposite or same tool in the two distant approaches.

Credible vs Confidence Interval

Frequentist: Cofidence interval, is interpreted in terms of repeated sampling(e.i., if we would repeat the experiment *infintly* many times then we expect 95% of the constructed CI to contain the true value of the parameter of interest).

Bayesian: Credible interval, is interpreted as summary of measure in terms of quantiles, we believe that the posterior distribution is indicating how parameters are distributed in the population.

Main Difference Frequentist define *CI* in terms of repeated sampling, while Bayesian defines it in terms of belief of how the parameters are distributed in the population.

Parameters

Frequentist: Parameters are fixed in the population and use different types of point- estimation procedures to estimate the parameters.

Bayesian: Parameters are random variables and use a posterior summarizations to describe these distributions. (they are derived using Bayesian theory)

Main Difference The difference lies in how the parameters are defined which is a key aspect to other practical differences.

Information Creteria Frequents: use log-likelihood evaluated at a point estimate(max likelihood estimate) , to calculate the misfit.

Bayesian : use 2*log-likelihood evaluated for a parameter value on posterior distribution(e.g using posterior mean), in order to calculate misfit(which is reffered to as Dhat in the complexity part of DIC) **Main diference**

The main difference is in calculating the misfit part, but also the whole posterior distribution(using each sample theta) to get 2*pd(complexity part), pd= Dbar-Dhat, where Dbar is the average of likelihood ratios evaluated for each sampled theta.

When To Use Gibbs

The **Gibbs sampler** is used when we have a multivariate posterior density, this is very hard to integrate or even the form of the integral is unknown. Possible to derive the conditional posterior densities of each parameter given other parameters. Researchers desire to use non-conjugate priors which results in multivariate

posteriors that are unknown and/or hard to integrate. Gibbs sampler is a special case of MH sampler with a perfect proposal density.

When To Use MH

The **Meteropopis Hastings Sampler** is used when we can not find the normalizing constant(unable to integrate the denominator of the Bayes rule) and/or when we can not recognize the form of *conditional density*(assuming our primary sampler was Gibbs), there is no conditional density, use a density propotional to conditional, and sample from a proposal and then based partly on the proportional density to accept or reject the sampled values. We can not use semi-conjugate prior, hence a conditional posterior has unknown from.

Example of MH

Given a data set $df < -load(file = "sampleExamQ4.Rdata"); unique(df)$, use unique function to check if there is high rate of repeated values in the chain to evaluate the acceptance rate, If the is repeation of values it indicates low acceptance rate, and MH sampler, it means if the propossed value is rejected, then the previously sampled value is written as the current value, which results in repeating the same value.

Convergence

In MH

No rush strategy , Change the proposal function This is due to the fact that the original proposal function is too narrow or too wide compared to the distribution we sampled from hence high or low acceptance rate respectively. So, we change to the optimal acceptance rate that affects the correlation. Therefore, multivariate proposals that are adjusted to the correlation structure and this is may be computationally intense.

Quick solution, Centering The differnce between the observed samples and their means, this diference is used in the model as paramters, this speeds up the convergence so fast and it is less time consuming.

Methods of Convergence

1. Trace plot

This check if the plot resembles "a big fat caterpillar" for the parameter(s), if the caterpillar is displayed it indicates that the parameter(s) converged at the same joint posterior distribution of the model parameter(s). If the was no convergence, we could increase the number of iterations or centering of the parameter(s) around their means to make sure the convergence occurs. but here no need.

2. Autocorrelation plot

The plot converge to zero immediately at the first lag for the parameter(s), this indicate rapid mixing. Hence, the sampler did not only appears to have reached the joint posterior distribution, but also appears to move through it effectively, thus subsequent values shows independence.

3. Gelman and Rubin statistics Paramters close to 1, implies no issue with convergence , hence independence.

Practicals

Exercise 1

With the system step done to perform bayesian analysis we need a cronological follow of steps below;

Step 1 In note pad create a file that difines a model to answer the research question, as follow. The file is *Exercise1Model.txt*

Cognitive behavioural therapy for PTSD: Is PE more effective than the baseline PC?

model{

likelihood of the data

y.PE ~ dbin(theta.PE, n.PE)

y.PC ~ dbin(theta.PC, n.PC)

prior distributions, Uninformative Prior

theta.PE ~ dbeta(1,1)

theta.PC ~ dbeta(1,1)

contrast/ output

RR <- theta.PC/theta.PE

}

Step 2

Obtain initial values.

For this particular model, it is not necessary to provide any initial values manually. JAGS automatically generates initial values when the model is specified, and no initial vales are provided. These are chosen to be a typical value from the prior distribution.

Step 3

Obtain samples from the posterior distribution of the parameters.

For the next steps in the analysis you will run JAGS from R using the rjags package. First load data then use jags model

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 2
##    Unobserved stochastic nodes: 2
##    Total graph size: 8
##
## Initializing model
```

Burn-in Period

Subsequently, use the *update()* function to run a large number of burn-in iterations (for example 1000 iterations) for your model:

```
#burn-in period :
update(object = model.def, n.iter =1000 )
```

WHY? caters for quick convergence(later)

Coda

Then, use the coda.samples() function to set monitors on the parameters of interest and draw a large number of samples from the posterior distribution, (for example 10000): obtain samples from the posterior distribution of the parameters and monitor these:

```
parameters <- c('theta.PE', 'theta.PC', 'RR')
res <- coda.samples(model = model.def, variable.names = parameters, n.iter = 10000)
summary(res)
```

```
##
## Iterations = 2001:12000
## Thinning interval = 1
## Number of chains = 2
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##            Mean      SD  Naive SE Time-series SE
## RR       0.6920 0.11510 0.0008139      0.0010320
## theta.PC 0.2830 0.03741 0.0002645      0.0003370
## theta.PE 0.4131 0.04101 0.0002900      0.0003741
##
## 2. Quantiles for each variable:
##
##            2.5%    25%    50%    75%  97.5%
## RR       0.4903 0.6116 0.6841 0.7641 0.9395
## theta.PC 0.2116 0.2573 0.2819 0.3078 0.3584
## theta.PE 0.3348 0.3850 0.4126 0.4408 0.4953
```

covergence check

Interpretation

The posterior mean is **0.69**, CI is $[0.49, 0, 94]$, lie below 1, thus we are certain or believe that PE therapy gives a higher change of recovery than PC therapy. Since, posterior mean of PC is **0.28** with Central Credible Interval between $[0.21, 0, 36]$, laying below 1, and posterior mean of PE is **0.41** with Central Credible interval is [0.33, 0.49].

Historical data comparison

In the model, inside the notepad we make changes in the priors, then we re-run the steps above, the interpretation is given as.

The posterior mean is **0.79**, CI is [0.62, 0, 99], lie below 1, thus we are certain or believe that PE therapy gives a higher change of recovery than PC therapy. Since, posterior mean of PC is 0.31 with Central Credible Interval is $[0.25, 0, 37]$, laying below 1, and posterior mean of PE is **0.34** with Central Credible interval is $[0.34, 0.46]$.The 95% CCIs of the RR are smaller, and still do not include 1. Therefore, the RR is still in favor of PE, but less strongly than with uninformative priors.

ExerciseExam-1

This question asks for the cohen's d and we have a data set with a number of predictors but in this example we are asked to use sex as a factor or predictor only, and the dependent variable is postnumb.

The very first step is to think of a model and write in the note pad file, it must contain the standand objects, likelihood, prior then output.

Load data

```
sesamedata<-read.table("sesame2.txt",header=TRUE)
View(sesamedata)
```

create a dummy for sex

```
## [1] 240  21
```

Creating a sampler to answer questions below.

Model creation

model{

likelihood

for(i in 1 : 240){

postnumb[i]~ dnorm(mu[i], tau)

mu[i] <- alpha + beta* sexDummy[i] }

Priors

alpha ~ dnorm(0, 0.0001)

beta ~ dnorm(0, 0.0001)

tau ~ dgamma(0.001, 0.001)

Output

sigma <- 1/tau

cohensd <- beta/ sigma

}

In R

```
library(rjags)
model.def <- jags.model(file = "modelssesa1.txt", data = sesamedata, n.chains = 2)
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 240
##    Unobserved stochastic nodes: 3
##    Total graph size: 494
##
## Initializing model
```

```
## burn-in period
update(object = model.def, n.iter =1000)

## define the parameters
parameters <- c("alpha", "beta", "sigma", "cohensd")

## sample with coda

results <- coda.samples(model = model.def, variable.names = parameters, n.iter = 10000)
summary(results)
```

```
##
## Iterations = 1001:11000
## Thinning interval = 1
## Number of chains = 2
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
```

```
##              Mean     SD  Naive SE Time-series SE
## alpha   3.004e+01 1.2046 0.0085181       0.015063
## beta    1.236e-02 1.6731 0.0118304       0.020935
## cohensd 9.574e-04 0.1294 0.0009149       0.001618
## sigma   1.291e+01 0.5905 0.0041755       0.004176
##
## 2. Quantiles for each variable:
##
##             2.5%     25%       50%      75%  97.5%
## alpha   27.6692 29.2316 30.045759 30.85403 32.408
## beta    -3.2545 -1.1044  0.014099  1.14137  3.297
## cohensd -0.2506 -0.0854  0.001089  0.08872  0.254
## sigma   11.8137 12.5041 12.896757 13.30031 14.117
```

convergence

Interpretation

Looking at the beta value= 0.04, using uninformative priors the means of boys(0) and girls(1) are equal, indicates that girls coded as 1, on average have a 0.04 higher mean than that of boys (coded as 0). However the 95% credible interval for beta includes zero which is (-3.2, 3.3), so, we can conclude that there is no difference between the means of boys and girls on the dependent variable(postnumb).

For cohen's d, the value 0.00298 , wit 95% credible interval between (-0.25, 0.26), this means that it is our belief that the value of cohen's d is zero with a probability of 95% .