STAT 597 Bayesian Statistics Homework #4

1. Use the following code to simulate data from a state space model (the same discussed in class)

set.seed(123)

u=74

b=.06

s2=3

z0=70

g1=40

g2=75

T=300

z=rep(z0,T+1)

for(t in 1:T){

z[t+1]=z[t]-b\*(z[t]-u)+rnorm(1,mean=0,sd=sqrt(s2))

if(t==100){

z[t+1]=z[t+1]+g1

}

if(t==200){

z[t+1]=z[t+1]+g2

}

}

y=rnorm(length(z),mean=z,sd=4)

plot(y)

## thin out y

y[-seq(10,300,by=10)]=NA

plot(y,col="red")

points(z,type="l")

## make “h”

h1=rep(0,300)

h1[100]=1

h2=rep(0,300)

h2[200]=1

Now construct a MCMC sampler BY HAND to fit this. You should work out full conditional distributions for all parameters in the following model:

After fitting this model with MCMC, compute the WAIC and DIC for this model. If you can’t get your own MCMC to work, compute the WAIC and DIC from nimble

1. Now, for the same data, fit the following model:

In the above model, is the k-th B-spline basis function in a semiparametric model for the mean of . Use 20 B-spline basis functions of order=4

B=create.bspline.basis(rangeval=c(mn,mx),nbasis=20,norder=4)

Phi=eval.basis(1:300,B)

The matrix **K** is **K=D’D** where **D** is a matrix that calculates the 2nd differences of **.** See BsplineEx.r for examples.

You may fit this model in nimble, or by hand. Either way, compute WAIC and DIC for this model, and then compare this model with the model in Q1. Which is better for this data?

Note that we have used the notation N(a,b) to denote a normal distribution with mean a and variance b, and IG(a,b) to denote an inverse gamma distribution with shape parameter a and scale parameter b.

The full conditional distributions used in the algorithm are derived as follows:

* z\_1 | y\_1, z\_2, ..., z\_300, u, beta, g\_1, g\_2, sigma^2, tau^2 ~ N((y\_1+u)/(1+beta), tau^2/(1+beta^2))
* z\_t | y\_1, ..., y\_t, z\_1, ..., z\_{t-1}, z\_{t+1}, ..., z\_300, u, beta, g\_1, g\_2, sigma^2, tau^2 ~ N(mu\_t, sigma^2), where mu\_t = (y\_t + beta*z\_{t-1} + g\_1*h\_1*t + g\_2*h\_2\*t)/(1 + beta^2)
* u | y\_1, ..., y\_{300}, z\_1, ..., z\_{300}, beta, tau^2 ~ N((beta\*z\_1)/(1+beta), tau^2/(1+beta^2))
* beta | y\_1, ..., y\_{300}, z\_1, ..., z\_{300}, u, g\_1, g\_2, sigma^2, tau^2 ~ N((z\_1-u)/(z\_2-z\_
* To construct a Markov Chain Monte Carlo (MCMC) sampler for this model, we need to derive the full conditional distributions for each parameter. The full conditional distribution for each parameter is the distribution of that parameter given all the other parameters. We will use these full conditional distributions to construct a Gibbs sampler, which alternates between updating each parameter by sampling from its full conditional distribution.
* Let's start with the likelihood function:
* p(y\_t | z\_t, τ^2) = N(y\_t | z\_t, τ^2)
* where N denotes the normal distribution. The likelihood function indicates that the data y\_t at time t are normally distributed with mean z\_t and variance τ^2.
* The next step is to derive the full conditional distribution for each parameter. We begin with u:
* p(u | z, β, g\_1, g\_2) = N(u | 0, 100^2)
* This is the prior distribution for u. Since u only appears in the model as a parameter of the prior distribution for z, its full conditional distribution does not depend on the data.
* The next parameter is β:
* p(β | z, u, g\_1, g\_2) = N(β | 0, 100^2)
* Similarly, β only appears in the model as a parameter of the prior distribution for z, so its full conditional distribution is also independent of the data.
* The next set of parameters are g\_1 and g\_2:
* p(g\_1 | z, u, β, g\_2) = N(g\_1 | 0, 100^2)
* p(g\_2 | z, u, β, g\_1) = N(g\_2 | 0, 100^2)
* These are also the prior distributions for g\_1 and g\_2, which are independent of the data.
* The next parameter is σ^2:
* p(σ^2 | z, y, τ^2) = IG(σ^2 | 10 + 150, 100 + 150τ^2 + ∑(y\_t-z\_t)^2)
* This is the full conditional distribution for σ^2. It is an inverse gamma distribution with shape parameter 10 + 150 and scale parameter 100 + 150τ^2 + ∑(y\_t-z\_t)^2. This distribution takes into account the likelihood function for the data y\_t and the normal distribution for the errors ε\_t = y\_t - z\_t.
* The final parameter is z:
* p(z\_t | z\_-t, y, τ^2, u, β, g\_1, g\_2) = N(z\_t | μ\_t, σ\_t^2)
* where μ\_t = (βu + (1-β)z\_t-1 + g\_1h\_1t + g\_2h\_2t)/(1+β) and σ\_t^2 = τ^2/(1+β)
* This is the full conditional distribution for z\_t. It is a normal distribution with mean μ\_t and variance σ\_t^2. The mean μ\_t depends on the previous value of z\_t, the current values of u, β, g\_1, and g\_2, and the time-varying covariates h\_1t and h\_2t. The variance σ\_t^2 only depends on τ^2 and β.

The τ^2 full conditional refers to the conditional distribution of τ^2 given the observed data and all other parameters in the model. To derive the full conditional distribution of τ^2, we need to apply Bayes' theorem and obtain the posterior distribution of τ^2, which is proportional to the product of the prior distribution of τ^2 and the likelihood function of the data, given τ^2. The constant of proportionality is determined by integrating over the entire range of τ^2.

Using the given model specification, the prior distribution of τ^2 is an inverse gamma distribution with parameters (a=10, b=100), denoted as IG(a,b). The likelihood function of the data, given τ^2, is a normal distribution with mean z\_t and variance τ^2, denoted as N(z\_t,τ^2).

Let D denote the observed data, and let θ denote the vector of all other parameters in the model, i.e., θ = {u, β, g\_1, g\_2, z\_1, σ^2}. Then the full conditional distribution of τ^2, denoted as f(τ^2|D,θ), is given by:

f(τ^2|D,θ) ∝ f(D|τ^2,θ) \* f(τ^2),

where f(D|τ^2,θ) is the likelihood function of the data, given τ^2 and θ, and f(τ^2) is the prior distribution of τ^2.

Using the given model specification, we have:

f(D|τ^2,θ) = ∏\_{t=1}^300 N(z\_t|z\_{t-1}-β(z\_{t-1}-u)+g\_1 h\_1t+g\_2 h\_2t, τ^2) = N(z|μ, Σ),

where z = (z\_1, z\_2, ..., z\_300), μ is the mean vector, and Σ is the covariance matrix. The mean vector μ and covariance matrix Σ can be computed using the Kalman filter or other numerical methods.

The prior distribution of τ^2 is an inverse gamma distribution with parameters a=10 and b=100, denoted as IG(10,100):

f(τ^2) = IG(τ^2|a=10, b=100) = (b^a / Gamma(a)) \* (τ^2)^{-a-1} \* exp(-b/τ^2),

where Gamma(a) is the gamma function evaluated at a.

Therefore, the full conditional distribution of τ^2 is:

f(τ^2|D,θ) ∝ N(z|μ, Σ) \* IG(τ^2|a=10, b=100),

which is proportional to the product of a normal density and an inverse gamma density. This is a known distribution called the inverse gamma-normal distribution, which has a closed-form expression that can be easily computed using numerical methods.

Based on the given model and data, the full conditional for σ^2 can be derived as follows:

First, we can write the likelihood function for the data as:

L(σ^2|y,z) = ∏\_(t=1)^n [1/(√(2πσ^2)) \* exp(-0.5\*(y\_t-z\_t)^2/σ^2)]

where y is the observed data and z is the latent variable.

Then, we can write the joint posterior distribution for σ^2 as:

p(σ^2|y,z) ∝ L(σ^2|y,z) \* p(σ^2)

where p(σ^2) is the prior distribution for σ^2.

Assuming an inverse gamma prior for σ^2, we have:

p(σ^2) = IG(σ^2|a,b) = (b^a / Γ(a)) \* (1/σ^2)^(a+1) \* exp(-b/σ^2)

where a = 10 and b = 100.

Using the fact that the full conditional for σ^2 is proportional to the joint posterior distribution, we can write:

p(σ^2|y,z) ∝ L(σ^2|y,z) \* IG(σ^2|a,b)

Taking the logarithm of both sides, we have:

log p(σ^2|y,z) ∝ log L(σ^2|y,z) + log IG(σ^2|a,b)

Simplifying the first term using the likelihood function, we have:

log L(σ^2|y,z) = -0.5 \* ∑\_(t=1)^n log(2πσ^2) - 0.5 \* ∑\_(t=1)^n (y\_t - z\_t)^2 / σ^2

Simplifying the second term using the prior distribution, we have:

log IG(σ^2|a,b) = a \* log(b) - log Γ(a) - (a+1) \* log(σ^2) - b/σ^2

Combining the two terms and dropping the constant terms that do not depend on σ^2, we have:

log p(σ^2|y,z) ∝ -0.5 \* ∑\_(t=1)^n log(2πσ^2) - 0.5 \* ∑\_(t=1)^n (y\_t - z\_t)^2 / σ^2 - (a+1) \* log(σ^2) - b/σ^2

Taking the exponential of both sides, we have:

p(σ^2|y,z) ∝ 1/σ^(n+a+1) \* exp[-0.5 \* ∑\_(t=1)^n (y\_t - z\_t)^2 / σ^2 - b/σ^2]

which is the full conditional for σ^2 in this model.

The full conditional for **z[1]** is:

**z[1] | z[2], y, u, beta, g1, g2, tau2, sigma2 ~ N(mu\_1, sigma\_1^2)**

where **mu\_1** and **sigma\_1^2** are:

**mu\_1 = (beta \* u + g1 \* h1[1] + g2 \* h2[1]) / (1 + beta)**

**sigma\_1^2 = sigma2 / (1 + beta)**

The full conditional distribution of $z\_{t+1}$ given the data and parameters can be derived from the joint distribution of $z\_{t+1}$, $z\_t$ and the data $y$.

Using Bayes theorem, the joint posterior distribution of $z\_{t+1}$, $z\_t$ and the parameters is proportional to the product of the likelihood and the prior distribution:

$p(z\_{t+1}, z\_t, \theta | y) \propto p(y | z\_{t+1}, \theta) \times p(z\_{t+1} | z\_t, \theta) \times p(z\_t | z\_{t-1}, \theta) \times p(\theta)$

where $\theta$ represents the set of parameters ${u, b, s^2, g\_1, g\_2, \sigma^2, \tau^2}$.

To obtain the full conditional distribution of $z\_{t+1}$, we can integrate out $z\_t$ and $\theta$ from the joint posterior distribution:

$p(z\_{t+1} | y, z\_{-t-1}) \propto \int \int p(y | z\_{t+1}, \theta) \times p(z\_{t+1} | z\_t, \theta) \times p(z\_t | z\_{t-1}, \theta) \times p(\theta) dz\_t d\theta$

where $z\_{-t-1}$ denotes all the $z$ values except for $z\_{t+1}$ and $z\_t$.

Since the joint distribution of $z\_t$ and $\theta$ is not available in closed-form, we can instead use Gibbs sampling to sample from the joint posterior distribution. In the Gibbs sampling algorithm, we iteratively sample from the full conditional distributions of each variable, while holding the other variables fixed at their current values.

Therefore, to obtain the full conditional distribution of $z\_{t+1}$, we need to derive its conditional distribution given all the other variables in the model.

Using the joint distribution of $z\_t$ and $z\_{t+1}$, we can write the conditional distribution of $z\_{t+1}$ as:

$p(z\_{t+1} | z\_t, y, z\_{-t-1}, \theta) \propto p(y | z\_{t+1}, \theta) \times p(z\_{t+1} | z\_t, \theta)$

Plugging in the expressions for the likelihood and the transition probability, we get:

$p(z\_{t+1} | z\_t, y, z\_{-t-1}, \theta) \propto \exp\left{-\frac{(y\_t - z\_{t+1})^2}{2 \tau^2}\right} \times \exp\left{-\frac{(z\_{t+1} - z\_t + \beta(z\_t - u) - g\_1 h\_{1,t} - g\_2 h\_{2,t})^2}{2 \sigma^2}\right}$

Taking the logarithm of the above expression and dropping the constant terms, we get:

$\log p(z\_{t+1} | z\_t, y, z\_{-t-1}, \theta) \propto -\frac{(y\_t - z\_{t+1})^2}{2 \tau^2} -\frac{(z\_{t+1} - z\_t + \beta(z\_t - u) - g\_1 h\_{1,t} - g\_2 h\_{2,t})^2}{2 \sigma^2}$

o implement the MCMC sampler, you need to define the log posterior distribution of the parameters and sample from their full conditional distributions.

For this model, the log posterior distribution is given by:

log(p(u, beta, g1, g2, z[2:T+1], tau2, sigma2 | y))

= log(p(y | z, tau2, sigma2)) + log(p(z[2:T+1] | z[1:T], u, beta, g1, g2, sigma2))

* log(p(u)) + log(p(beta)) + log(p(g1)) + log(p(g2)) + log(p(z[1])) + log(p(tau2)) + log(p(sigma2))

= log(Normal(y | z, tau2^0.5)) + log(Normal(z[2:T+1] | z[1:T], beta\*(z[1:T]-u)+g1*h1[2:T+1]+g2*h2[2:T+1], sigma2))

* log(Normal(u | 0, 100)) + log(Normal(beta | 0, 100)) + log(Normal(g1 | 0, 100)) + log(Normal(g2 | 0, 100))
* log(Normal(z[1] | 0, 100)) + log(IG(tau2 | 10, 100)) + log(IG(sigma2 | 10, 100))

To perform WAIC and DIC by hand, you need to calculate the log-likelihood of the model and then use it to compute the necessary statistics. Here are the steps to follow:

1. Calculate the log-likelihood of each observation in the dataset under the model. In this case, the likelihood function for the i-th observation can be written as:

**loglik\_i = dnorm(y[i], mean = z[t.obs[i]], sd = sqrt(tau2), log = TRUE)**

1. Compute the log-likelihood of the entire dataset by summing the log-likelihoods of each observation:

**loglik = sum(loglik\_i)**

1. Calculate the deviance of the model by subtracting the log-likelihood of the saturated model (i.e., a model with a separate parameter for each observation) from the log-likelihood of the current model. In this case, the deviance can be computed as:

**deviance = -2\*(loglik - sum(dnorm(y, mean = mean(y), sd = sd(y), log = TRUE)))**

1. Compute the effective number of parameters (p\_DIC) as the difference between the deviance and the log-likelihood of the current model, divided by 2:

**p\_DIC = (deviance - loglik)/2**

1. Calculate the WAIC by first computing the sum of the pointwise log-likelihoods (lppd), and then adding to it the effective number of parameters multiplied by the sum of the variances of the pointwise log-likelihoods (p\_WAIC):

**lppd = sum(loglik\_i)**

**p\_WAIC = sum((loglik\_i - loglik)^2)**

**WAIC = -2\*(lppd - p\_WAIC)**

Note that in practice, you would want to compute these statistics for multiple MCMC samples and then summarize the results using the mean or median values. Also, these statistics can be sensitive to the choice of reference distribution for the deviance, so you may want to consider using different reference distributions (e.g., the prior distribution or a null model) and compare the results.