Cassandra Ferring
ST 590 Mid-term Exam
April 10, 2017

This experiment was based on a previous x-ray crystallography experiment. Data consisted of n = 1,000 observations of angles (ɵ) and a count (Y). Figure 1 shows the plotted data with ɵ on the x-axis and Y on the y-axis. The objective of this study was to estimate the number of peaks and their locations in order to identify the material that was sampled.

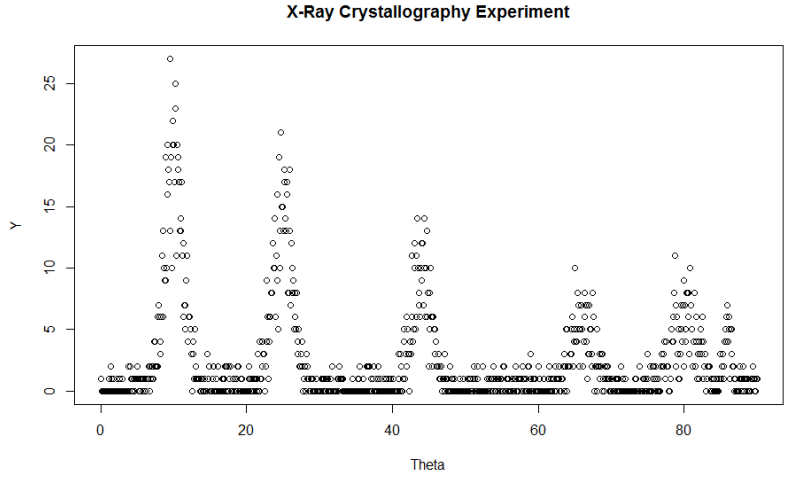

Figure 1: Data plotted of n = 1,000 observations with angles (ɵ) on the x axis and counts(Y) on the y axis

The following model was used to fit the data:

$$Y \sim \text{Poisson}[\lambda(\theta)] \text{ where } \lambda(\theta) = \alpha_0 + \sum_{j=1}^{J} \phi(\theta; \gamma j, \tau j)\ \beta j$$

Where $\alpha > 0$ is the background intensity or the amount of error of the x-ray crystallography, $\phi(\theta; \gamma, \tau) = \exp\left[-\frac{(\theta-\gamma)^2}{2\tau^2}\right]$ is the Gaussian peak function which details the shape of the peak, and the j^{th} peak is described by the location of the center of the peak γ_j which is between 0 and 90, $\tau_j > 0$ is the width or variance of the peak, and $\beta_j > 0$ is the intensity or height of the peaks.

First, three models with priors were created to potentially model the data. The three models were to model J= 5, 6, and 7 peaks, respectively. For all models, γ was fit with a uniform prior (0, 90) which was the range of the ɵ values and τ, β, and α were fit with gamma priors (0.1, 0.1). These priors were set to be uninformative because there was no prior information available. Initial values were also set for all the models. For the five peak model, the initial values of γ were set to 10, 23, 43, 65, and 80; τ were all set to 1; and β were set to 25, 19, 13, 9, and 10. For the six peak model, the initial values of γ were set to 10, 23, 43, 65, 80, and 83; τ were all set to 1; and β were set to 25, 19, 13, 9, 10, and 6. For the seven peak model, the initial values of γ were set to 10, 23, 43, 65, 80, 81, and 83; τ were all set to 1; and β were set to 25, 19, 13, 9, 10, 7, and 6. These values were based on the data as it was observed in figure 1.

Next, the models were created in JAGS using the jag.model function in R with 2 chains used in the creation of each model. After the models are created in JAGS, they are updated with a burn in. For each of the three models, a burn in of 5,000 iterations was done. After the 5,000 burn-in iterations, 10,000 more iterations were run to extract the samples from each model for γ, τ, and β. The penalized deviance

was extracted from the model to determine which model had the best fit. The penalized deviance for the models were 2901, 2788, and 2788 for the five, six, and seven peak models respectively. The lower the penalized deviance when comparing models, the better fit the model is for the data. Based on these penalized deviances, the six and seven peak models fit equally well to this data with a penalized deviance of 2788.

Since models six and seven were equally fit based on the penalized deviance, the mean estimates and convergence were assessed to determine if one of the models fit the data better than the other. The mean estimates and standard deviations for γ, τ, and β for the six and seven peak models is found in table 1. The values for γ_6 and γ_7 for the seven peak model had greatly increase standard deviations compare to the standard deviations of the six peak model, which could be indicative of a problem with model convergence at that location. Since there is a potential problem at γ_6 and γ_7, the convergence plots were considered for these locations. Figure 2 compares the convergence between γ_6 in the six peak model and γ_6 and γ_7 in the seven peak model. This figure clearly shows that the two chains run in the six peak model converged together, while the chains in the seven peak model did not show convergence i.e. the lines did not come together at the same point at the same time. With this information, it appears as if the six peak model was the best fit for this data.

Variable	Six Peak Model		Seven Peak Model	
	Mean	SD	Mean	SD
γ_1	10.00	0.05	10.00	0.05
γ_2	24.92	0.06	24.92	0.06
γ_3	43.96	0.08	43.96	0.08
γ_4	66.01	0.13	66.00	0.13
γ_5	80.13	0.13	80.13	0.13
γ_6	85.97	0.08	65.66	27.41
γ_7	.	.	65.83	27.45
τ_1	1.24	0.04	1.24	0.04
τ_2	1.31	0.05	1.31	0.05
τ_3	1.24	0.06	1.24	0.06
τ_4	1.54	0.11	1.54	0.11
τ_5	1.72	0.11	1.72	0.11
τ_6	0.47	0.06	0.28	0.22
τ_7	.	.	0.25	0.23
β_1	18.52	0.93	18.50	0.95
β_2	14.37	0.80	14.34	0.81
β_3	9.53	0.69	9.52	0.68
β_4	4.95	0.46	4.95	0.46
β_5	5.44	0.46	5.43	0.47
β_6	4.75	0.81	2.48	2.38
β_7	.	.	2.86	3.15

Table 1: Column one shows the variables γ, τ, and β for 1-7. Columns 2 and 4 are the means of the sampled variables for the six and seven peak models, respectively and columns 3 and 5 are the standard deviations for the sampled variables for the models as well.

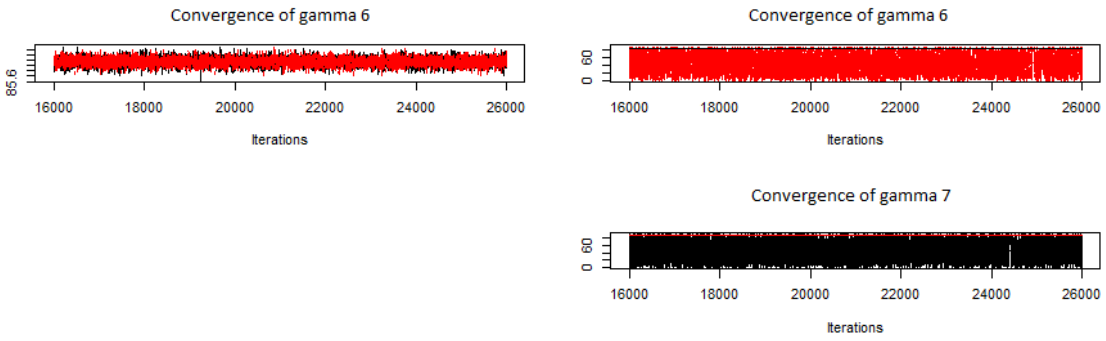

Figure 2: The plot on the left shows the convergence of γ_6 for the six peak model. The two plots on the right show the convergence of γ_6 and γ_7 for the seven peak model.

Finally, it can be concluded that the best model to fit this data is a six peak model (J=6) and the locations of these peaks are based on the γ from this model which were at 10.00, 24.92, 43.96, 66.01, 80.13, and 85.97. The 95% intervals quantify the uncertainty for these γ estimates and can be found in table 2. Figure 3 shows how the mean estimates fit the data with the 95% intervals. The plots in figure 4 show that the two chains that were run for the six peak model converged for all γ.

Variable	2.5% Quantile	97.5% Quantile
γ_1	9.90	10.10
γ_2	24.80	25.05
γ_3	43.81	44.12
γ_4	65.75	66.26
γ_5	79.88	80.38
γ_6	85.82	86.12

Table 2: The γ variables are found in the first column. Column two contains the lower bound of the 95% interval and column three contains the upper bound of the 95% interval.

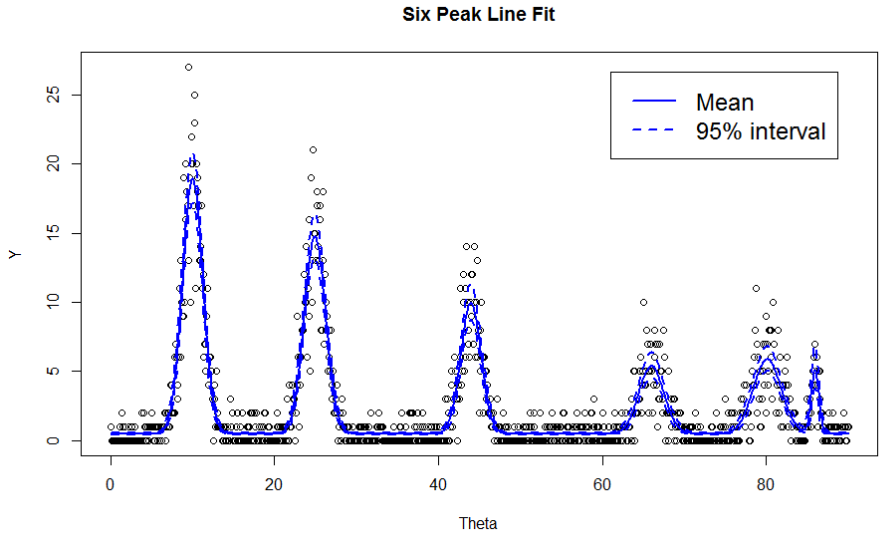

Figure 3: Mean of λ estimates fit to the data (solid line) and the 95% interval also fit to the data (dashed line)

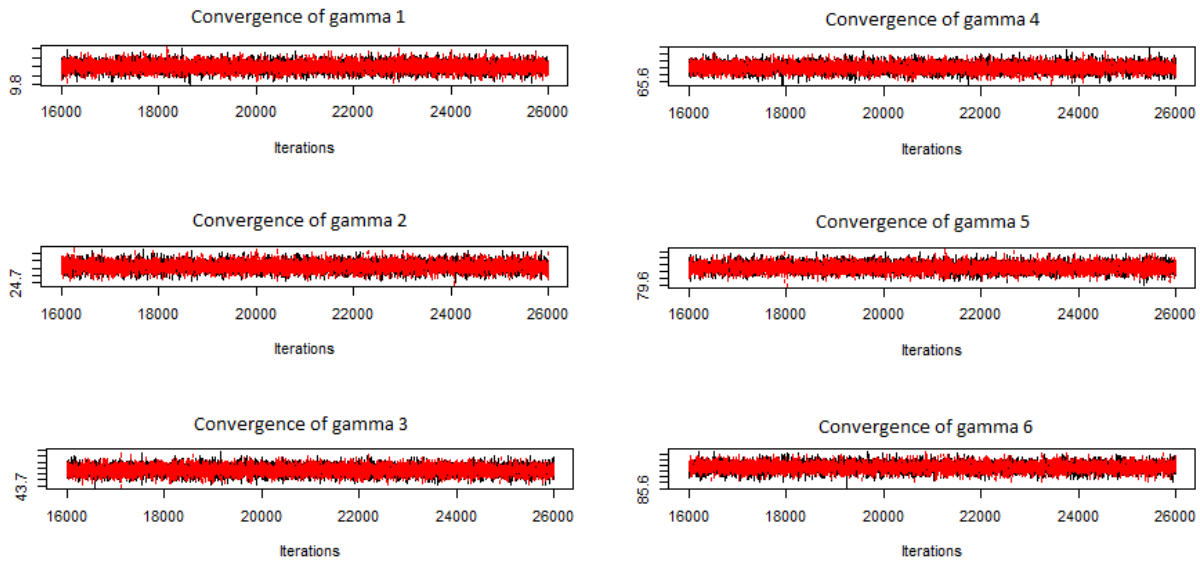

Figure 4: Convergence of all γ variables using the six peak model

Cassandra Ferring
ST 590 Mid-term Exam
April 10, 2017

R Code:

```
rm(list = ls())
dev.off()
data = read.csv("C:/Users/Cassie/Documents/Bayes/xraydata.csv", header=TRUE, na.strings = ".")
Y = data$Y
theta = data$theta
n = 1000
library(rjags)
#Initial Plot
plot(data$theta,data$Y, main = "X-Ray Crystallography Experiment",  xlab = "Theta", ylab = "Y")
### 5 Peaks ###
J = 5
model5peak<- "model{
#Model and Priors
for(i in 1:n){
Y[i] ~ dpois(lambda[i])
lambda[i] <- alpha + exp(-((theta[i]-gamma[1])^2)/(2*tau[1]^2))*beta[1] +
exp(-((theta[i]-gamma[2])^2)/(2*tau[2]^2))*beta[2]+
exp(-((theta[i]-gamma[3])^2)/(2*tau[3]^2))*beta[3]+
exp(-((theta[i]-gamma[4])^2)/(2*tau[4]^2))*beta[4]+
exp(-((theta[i]-gamma[5])^2)/(2*tau[5]^2))*beta[5]
}
for(i in 1:J){
tau[i] ~ dgamma(0.1,0.1)
beta[i] ~ dgamma(0.1,0.1)
gamma[i] ~ dunif(0,90)
}
alpha ~ dgamma(0.1,0.1)
}"
#Initial Values
initvals  <- function() {list(tau = c(1, 1, 1, 1, 1), beta = c(25, 19, 13, 9, 10), gamma = c(10, 23, 43, 65, 80))}
```

```
#Run JAGS

model5 <- jags.model(textConnection(model5peak), data = list(Y=Y, theta=theta, n=n, J=J),inits = initvals, n.chains =
2, quiet=TRUE)

#Burn in -> 5,000 iterations

update(model5, 5000, progress.bar="none")

#Extracting Samples from the Model -> 10,000 iterations

samp51 <- coda.samples(model5, variable.names=c("beta", "gamma", "tau"), n.iter=10000, progress.bar="none")

samp52 <- coda.samples(model5, variable.names=c("lambda"), n.iter=10000, progress.bar="none")

#Penalized Deviance from the Model

dic1 <- dic.samples(model5, variable.names=c("beta", "gamma", "tau"), n.iter=10000, progress.bar="none")

 #Conversion Plots

plot(samp51)

#Summary of Samples

sum51 <- summary(samp51)

print(sum51)

sum52 <- summary(samp52)

#Plotting Mean and 95% Quantiles for Samples

q1 <- sum52$quantiles

plot(data$theta,data$Y, main = "Five Peak Line Fit",  xlab = "Theta", ylab = "Y")

lines(theta[1:n],q1[,1],col=4,lty=2,lwd=2)

lines(theta[1:n],q1[,3],col=4,lty=1,lwd=2)

lines(theta[1:n],q1[,5],col=4,lty=2,lwd=2)

legend("topright",c("Mean","95% interval"),lty=1:2,col=4,lwd=2,bg=gray(1),inset=0.05,cex=1.5)

### 6 Peaks ###

J= 6

model6peak<- "model{

#Model and Priors

for(i in 1:n){

Y[i] ~ dpois(lambda[i])

lambda[i] <- alpha + exp(-((theta[i]-gamma[1])^2)/(2*tau[1]^2))*beta[1] +

exp(-((theta[i]-gamma[2])^2)/(2*tau[2]^2))*beta[2]+

exp(-((theta[i]-gamma[3])^2)/(2*tau[3]^2))*beta[3]+
```

```
exp(-((theta[i]-gamma[4])^2)/(2*tau[4]^2))*beta[4]+

exp(-((theta[i]-gamma[5])^2)/(2*tau[5]^2))*beta[5]+

exp(-((theta[i]-gamma[6])^2)/(2*tau[6]^2))*beta[6]

}

for(i in 1:J){

tau[i] ~ dgamma(0.1,0.1)

beta[i] ~ dgamma(0.1,0.1)

gamma[i] ~ dunif(0,90)

}

alpha ~ dgamma(0.1,0.1)

}"

init  <- function() {list(tau = c(1, 1, 1, 1, 1, 1), beta = c(25, 19, 13, 9, 10, 6), gamma = c(10, 23, 43, 65, 80, 83))}

#Run JAGS

model6 <- jags.model(textConnection(model6peak), data = list(Y=Y, theta=theta, n=n, J=J),inits = initvals, n.chains =
2, quiet=TRUE)

#Burn in -> 5,000 iterations

update(model6, 5000, progress.bar="none")

#Extracting Samples from the Model -> 10,000 iterations

samp61 <- coda.samples(model6, variable.names=c("beta", "gamma", "tau"), n.iter=10000, progress.bar="none")

samp62 <- coda.samples(model6, variable.names=c("lambda"), n.iter=10000, progress.bar="none")

#Penalized Deviance from the Model

dic6 <- dic.samples(model6, variable.names=c("beta", "gamma", "tau"), n.iter=10000, progress.bar="none")

#Conversion Plots

plot(samp61)

#Summary of Samples

sum61 <- summary(samp61)

print(sum61)

sum62 <- summary(samp62)

#Plotting Mean and 95% Quantiles for Samples

q2 <- sum62$quantiles

plot(data$theta,data$Y, main = "Six Peak Line Fit",  xlab = "Theta", ylab = "Y")

lines(theta[1:n],q2[,1],col=4,lty=2,lwd=2)
```

```
lines(theta[1:n],q2[,3],col=4,lty=1,lwd=2)

lines(theta[1:n],q2[,5],col=4,lty=2,lwd=2)

legend("topright",c("Mean","95% interval"),lty=1:2,col=4,lwd=2,bg=gray(1),inset=0.05,cex=1.5)

### 7 Peaks ###

J = 7

model7peak<- "model{

#Model and Priors

for(i in 1:n){

Y[i] ~ dpois(lambda[i])

lambda[i] <- alpha + exp(-((theta[i]-gamma[1])^2)/(2*tau[1]^2))*beta[1] +

exp(-((theta[i]-gamma[2])^2)/(2*tau[2]^2))*beta[2]+

exp(-((theta[i]-gamma[3])^2)/(2*tau[3]^2))*beta[3]+

exp(-((theta[i]-gamma[4])^2)/(2*tau[4]^2))*beta[4]+

exp(-((theta[i]-gamma[5])^2)/(2*tau[5]^2))*beta[5]+

exp(-((theta[i]-gamma[6])^2)/(2*tau[6]^2))*beta[6]+

exp(-((theta[i]-gamma[7])^2)/(2*tau[7]^2))*beta[7]

}

for(i in 1:J){

tau[i] ~ dgamma(0.1,0.1)

beta[i] ~ dgamma(0.1,0.1)

gamma[i] ~ dunif(0,90)

}

alpha ~ dgamma(0.1,0.1)

}"

initvals  <- function() {list(tau = c(1, 1, 1, 1, 1, 1, 1), beta = c(25, 19, 13, 9, 10, 7, 6), gamma = c(10, 23, 43, 65, 80, 81, 83))}

#Run JAGS

model7 <- jags.model(textConnection(model7peak), data = list(Y=Y, theta=theta, n=n, J=J),inits = initvals, n.chains = 2, quiet=TRUE)

#Burn in -> 5,000 iterations

update(model7, 5000, progress.bar="none")
```

```
#Extracting Samples from the Model -> 10,000 iterations
samp71 <- coda.samples(model7, variable.names=c("beta", "gamma", "tau"), n.iter=10000, progress.bar="none")
samp72 <- coda.samples(model7, variable.names=c("lambda"), n.iter=10000, progress.bar="none")
#Penalized Deviance from the Model
dic7 <- dic.samples(model7, variable.names=c("beta", "gamma", "tau"), n.iter=10000, progress.bar="none")
Print(dic7)
#Conversion Plots
plot(samp71)
#Summary of Samples
sum71 <- summary(samp71)
print(sum71)
sum72 <- summary(samp72)
#Plotting Mean and 95% Quantiles for Samples
q3 <- sum72$quantiles
plot(data$theta,data$Y, main = "Seven Peak Line Fit",  xlab = "Theta", ylab = "Y")
lines(theta[1:n],q3[,1],col=4,lty=2,lwd=2)
lines(theta[1:n],q3[,3],col=4,lty=1,lwd=2)
lines(theta[1:n],q3[,5],col=4,lty=2,lwd=2)
legend("topright",c("Mean","95% interval"),lty=1:2,col=4,lwd=2,bg=gray(1),inset=0.05,cex=1.5)
```