R Notebook

Question 1 As part of their analysis of the Federalist papers, Mosteller and Wallace (1964) recorded the frequency of use of various words in selected articles by Alexander Hamilton and James Madison. The articles were divided into blocks (247 for Hamilton, 262 for Madison) of about 200 words each, and the number of instances of various words in each block were recorded. The table below displays the results for the word "may." # occurrences in a block 0 1 2 3 4 5 6 7+ Total blocks # blocks (Hamilton) 128 67 32 14 4 1 1 0 247 # blocks (Madison) 156 63 29 8 4 1 1 0 262

These data are also provided in the hamilton and madison vectors. a) Fit a Poisson model to these data, with different parameters for each author and a noninformative/weakly informative prior distirbution. Plot the posterior density of the Poisson mean parameter for each author. Interpret. Fit the model in Stan:

```r
library(rstan)
```

```
## Warning: package 'rstan' was built under R version 3.2.5
## Loading required package: ggplot2
## Warning: package 'ggplot2' was built under R version 3.2.5
## Loading required package: StanHeaders
## Warning: package 'StanHeaders' was built under R version 3.2.5
## rstan (Version 2.12.1, packaged: 2016-09-11 13:07:50 UTC, GitRev: 85f7a56811da)
## For execution on a local, multicore CPU with excess RAM we recommend calling
## rstan_options(auto_write = TRUE)
## options(mc.cores = parallel::detectCores())
```

```r
load("FinalExam.RData")
hamilton
```

```
##   [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [36] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [71] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [106] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1
## [141] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [176] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [211] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4
## [246] 5 6
```

```r
madison
```

```
##   [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [36] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [71] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [106] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [141] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [176] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [211] 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [246] 2 2 2 3 3 3 3 3 3 3 3 4 4 4 4 5 6
```

```r
mean(hamilton)
```

```
## [1] 0.8097166
```

```r
mean(madison)
```

```
## [1] 0.6564885
```

```
set.seed(500)
mod1code <- '
data{
  int<lower=0> y[247];
}
parameters{
  real <lower=0> lambda1;
}
model{
  lambda1 ~gamma(0.01,0.01);
  for(i in 1:247){
    y[i] ~ poisson(lambda1);
  }
}
'
mod1dat <- list(y=hamilton)
mod1 <- stan(model_code=mod1code, data=mod1dat)
```

```
## C:/Users/Palma/Documents/R/win-library/3.2/StanHeaders/include/stan/math/rev/core/set_zero_all_adjoi
##
## SAMPLING FOR MODEL '2dc97e7f3770ef45527d206bfcf757a0' NOW (CHAIN 1).
##
## Chain 1, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 1, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 1, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 1, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 1, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 1, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 1, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 1, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 1, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 1, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 1, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 1, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.084 seconds (Warm-up)
##                0.073 seconds (Sampling)
##                0.157 seconds (Total)
##
##
## SAMPLING FOR MODEL '2dc97e7f3770ef45527d206bfcf757a0' NOW (CHAIN 2).
##
## Chain 2, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 2, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 2, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 2, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 2, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 2, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 2, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 2, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 2, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 2, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 2, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 2, Iteration: 2000 / 2000 [100%]  (Sampling)
```

```
##  Elapsed Time: 0.076 seconds (Warm-up)
##                0.067 seconds (Sampling)
##                0.143 seconds (Total)
##
##
## SAMPLING FOR MODEL '2dc97e7f3770ef45527d206bfcf757a0' NOW (CHAIN 3).
##
## Chain 3, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 3, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 3, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 3, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 3, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 3, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 3, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 3, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 3, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 3, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 3, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 3, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.069 seconds (Warm-up)
##                0.085 seconds (Sampling)
##                0.154 seconds (Total)
##
##
## SAMPLING FOR MODEL '2dc97e7f3770ef45527d206bfcf757a0' NOW (CHAIN 4).
##
## Chain 4, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 4, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 4, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 4, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 4, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 4, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 4, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 4, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 4, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 4, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 4, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 4, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.083 seconds (Warm-up)
##                0.08 seconds (Sampling)
##                0.163 seconds (Total)
```

```r
print(mod1,pars="lambda1")
```

```
## Inference for Stan model: 2dc97e7f3770ef45527d206bfcf757a0.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##         mean se_mean   sd 2.5%  25%  50%  75% 97.5% n_eff Rhat
## lambda1 0.81       0 0.06  0.7 0.77 0.81 0.85  0.93  3142    1
##
## Samples were drawn using NUTS(diag_e) at Mon Dec 12 16:17:31 2016.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

```r
set.seed(501)
mod2code <- '
data{
  int<lower=0> y[262];
}
parameters{
  real <lower=0> lambda2;
}
model{
  lambda2 ~gamma(0.01,0.01);
  for(i in 1:262){
    y[i] ~ poisson(lambda2);
  }
}
'

mod2dat <- list(y=madison)
mod2 <- stan(model_code=mod2code, data=mod2dat)
```

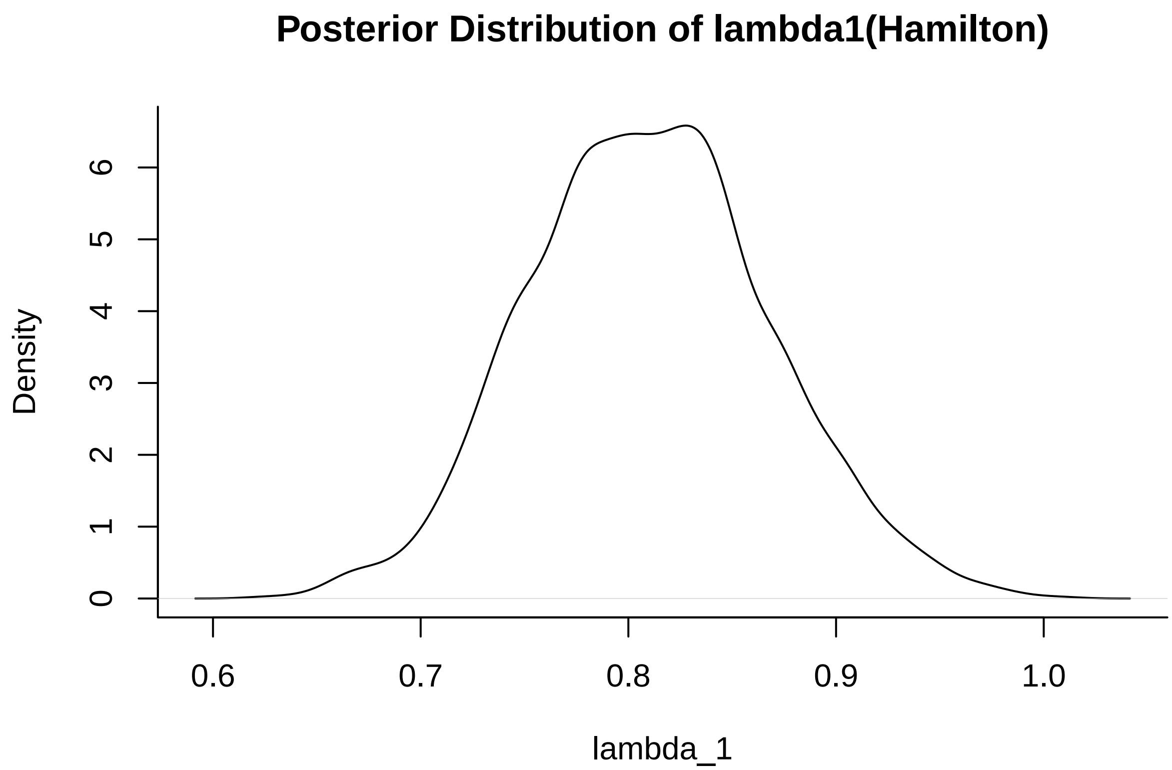
```
## C:/Users/Palma/Documents/R/win-library/3.2/StanHeaders/include/stan/math/rev/core/set_zero_all_adjoi
##
## SAMPLING FOR MODEL '6af5ecfd9319f3c6e80861de77ba94ac' NOW (CHAIN 1).
##
## Chain 1, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 1, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 1, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 1, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 1, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 1, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 1, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 1, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 1, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 1, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 1, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 1, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.069 seconds (Warm-up)
##                0.085 seconds (Sampling)
##                0.154 seconds (Total)
##
##
## SAMPLING FOR MODEL '6af5ecfd9319f3c6e80861de77ba94ac' NOW (CHAIN 2).
##
## Chain 2, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 2, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 2, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 2, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 2, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 2, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 2, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 2, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 2, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 2, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 2, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 2, Iteration: 2000 / 2000 [100%]  (Sampling)
```

```
##  Elapsed Time: 0.083 seconds (Warm-up)
##                0.079 seconds (Sampling)
##                0.162 seconds (Total)
##
##
## SAMPLING FOR MODEL '6af5ecfd9319f3c6e80861de77ba94ac' NOW (CHAIN 3).
##
## Chain 3, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 3, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 3, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 3, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 3, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 3, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 3, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 3, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 3, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 3, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 3, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 3, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.069 seconds (Warm-up)
##                0.084 seconds (Sampling)
##                0.153 seconds (Total)
##
##
## SAMPLING FOR MODEL '6af5ecfd9319f3c6e80861de77ba94ac' NOW (CHAIN 4).
##
## Chain 4, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 4, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 4, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 4, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 4, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 4, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 4, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 4, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 4, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 4, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 4, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 4, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.085 seconds (Warm-up)
##                0.075 seconds (Sampling)
##                0.16 seconds (Total)
```

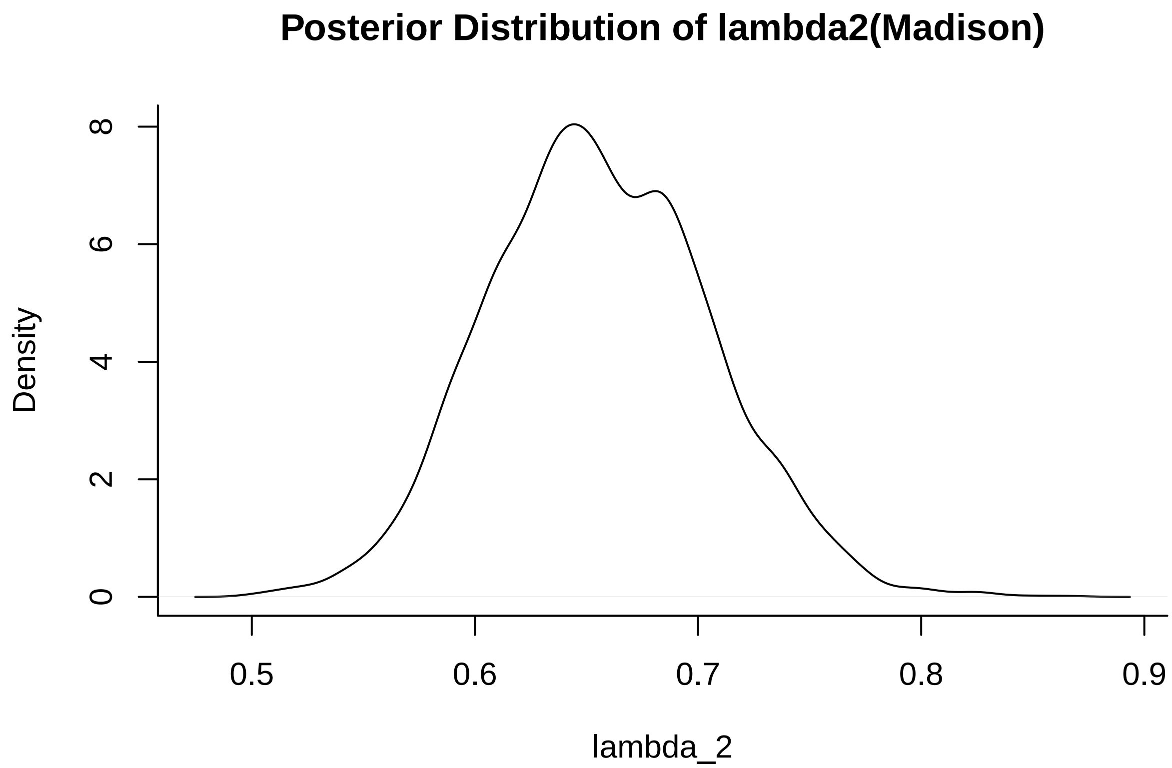
```r
print(mod2,pars="lambda2")
```

```
## Inference for Stan model: 6af5ecfd9319f3c6e80861de77ba94ac.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##         mean se_mean   sd 2.5%  25%  50%  75% 97.5% n_eff Rhat
## lambda2 0.66       0 0.05 0.56 0.62 0.65 0.69  0.76  3051    1
##
## Samples were drawn using NUTS(diag_e) at Mon Dec 12 16:18:13 2016.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

```
# Plot of posterior distribution of lambda1
lambda1<-extract(mod1)$lambda1
plot(density(lambda1), bty="l",xlab=expression(lambda_1), main="Posterior Distribution of lambda1(Hamil
```




```
# Plot of posterior distribution of lambda2
lambda2<-extract(mod2)$lambda2
plot(density(lambda2), bty="l",xlab=expression(lambda_2),main="Posterior Distribution of lambda2(Madiso
```



Thus using noninformative/weakly informative prior distirbution we chose a gamma distribution since we expect λ to be positive values greater than 0.We can see for the Hamilton data the posterior distribution of λ settles to a mean of 0.81 while the posterior distribution of λ settles to a mean of 0.66 for the Madison data. Given the distribution of the data , these values seem reasonable.

There are multiple parameterizations of the negative binomial distribution, but the most convenient for our purpose in Stan is neg_binomial_2; The mean and variance of this distribution are given by EX $= \mu$ and VarX $= \mu + \mu\ \mu\ /\phi$. Here, is called the "dispersion parameter," and $\delta = 1/\phi$

b) Suggest reasonable noninformative/weakly informative priors for ?? and ϕ (or, if you prefer, for ?? and δ).

The negative binomial model can be used to address potential overdispersion relative to the Poisson. The negative binomial distribution converges to a possion distribution as ϕ to becomes very large . Small values of ϕ drag the mode of the distribution towards 0.A reasonable noninformative/weakly informative priors for hamilton are : μ must be a positive real number from say a gamma distribution. Since the distribution is skewed towards 0, we choose a Gamma(0.01,0.01) and also ϕ must also be a positive real number hence we chose phi from a gamma distribution.Let ϕ be a Gamma(0.01,0.02). A reasonable noninformative/weakly informative priors for madison are for μ is Gamma(0.01,0.01) and ϕ be a Gamma(0.01,0.01) . ALso note that there is not much dispersion expected since the count of number of occurrences in a block cannot necessarily be heavily overdispersed.

c) Using your priors from part b, fit a negative binomial model to these data, with different parameters for each author. Plot the posterior densities of each parameter. Interpret.

```
set.seed(503)
mod3code <- '
data{
  int<lower=0> y[247];
```

```
}
parameters{
  real <lower=0> mu3;
  real<lower=1>  phi3;
}
model{
  mu3 ~gamma(0.01,0.01);
  phi3 ~ gamma(0.01,0.02);
  for(i in 1:247){
    y[i] ~ neg_binomial_2(mu3,phi3);
  }
}
'
mod3dat <- list(y=hamilton)
mod3 <- stan(model_code=mod3code, data=mod3dat)
```

```
## C:/Users/Palma/Documents/R/win-library/3.2/StanHeaders/include/stan/math/rev/core/set_zero_all_adjoi
##
## SAMPLING FOR MODEL 'a3a1cd10d84fda2e4442aee730ae801a' NOW (CHAIN 1).
##
## Chain 1, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 1, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 1, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 1, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 1, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 1, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 1, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 1, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 1, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 1, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 1, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 1, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.594 seconds (Warm-up)
##                0.508 seconds (Sampling)
##                1.102 seconds (Total)
##
## [1] "The following numerical problems occured the indicated number of times after warmup on chain 1"
##                                                                                                     c
## Exception thrown at line 14: neg_binomial_2_log: Location parameter is 1.#INF, but must be finite!
## [1] "When a numerical problem occurs, the Hamiltonian proposal gets rejected."
## [1] "See http://mc-stan.org/misc/warnings.html#exception-hamiltonian-proposal-rejected"
## [1] "If the number in the 'count' column is small,  do not ask about this message on stan-users."
##
## SAMPLING FOR MODEL 'a3a1cd10d84fda2e4442aee730ae801a' NOW (CHAIN 2).
##
## Chain 2, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 2, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 2, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 2, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 2, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 2, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 2, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 2, Iteration: 1200 / 2000 [ 60%]  (Sampling)
```

```
## Chain 2, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 2, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 2, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 2, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.585 seconds (Warm-up)
##                0.501 seconds (Sampling)
##                1.086 seconds (Total)
##
## [1] "The following numerical problems occured the indicated number of times after warmup on chain 2"
##                                                                                                    c
## Exception thrown at line 14: neg_binomial_2_log: Location parameter is 1.#INF, but must be finite!
## [1] "When a numerical problem occurs, the Hamiltonian proposal gets rejected."
## [1] "See http://mc-stan.org/misc/warnings.html#exception-hamiltonian-proposal-rejected"
## [1] "If the number in the 'count' column is small,  do not ask about this message on stan-users."
##
## SAMPLING FOR MODEL 'a3a1cd10d84fda2e4442aee730ae801a' NOW (CHAIN 3).
##
## Chain 3, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 3, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 3, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 3, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 3, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 3, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 3, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 3, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 3, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 3, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 3, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 3, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.617 seconds (Warm-up)
##                0.517 seconds (Sampling)
##                1.134 seconds (Total)
##
## [1] "The following numerical problems occured the indicated number of times after warmup on chain 3"
##                                                                                                    c
## Exception thrown at line 14: neg_binomial_2_log: Location parameter is 1.#INF, but must be finite!
## [1] "When a numerical problem occurs, the Hamiltonian proposal gets rejected."
## [1] "See http://mc-stan.org/misc/warnings.html#exception-hamiltonian-proposal-rejected"
## [1] "If the number in the 'count' column is small,  do not ask about this message on stan-users."
##
## SAMPLING FOR MODEL 'a3a1cd10d84fda2e4442aee730ae801a' NOW (CHAIN 4).
##
## Chain 4, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 4, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 4, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 4, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 4, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 4, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 4, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 4, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 4, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 4, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 4, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 4, Iteration: 2000 / 2000 [100%]  (Sampling)
```

```
##  Elapsed Time: 0.669 seconds (Warm-up)
##                0.554 seconds (Sampling)
##                1.223 seconds (Total)
##
## [1] "The following numerical problems occured the indicated number of times after warmup on chain 4"
##                                                                                                 count
## Exception thrown at line 14: neg_binomial_2_log: Location parameter is 0, but must be > 0!     1
## [1] "When a numerical problem occurs, the Hamiltonian proposal gets rejected."
## [1] "See http://mc-stan.org/misc/warnings.html#exception-hamiltonian-proposal-rejected"
## [1] "If the number in the 'count' column is small,  do not ask about this message on stan-users."
```

```
print(mod3,pars=c("mu3","phi3 "))
```

```
## Inference for Stan model: a3a1cd10d84fda2e4442aee730ae801a.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##      mean se_mean   sd 2.5%  25%  50%  75% 97.5% n_eff Rhat
## mu3  0.81    0.00 0.07 0.69 0.76 0.81 0.85  0.95  2568 1.00
## phi3 2.13    0.02 1.11 1.11 1.51 1.86 2.40  4.75  2411 1.01
##
## Samples were drawn using NUTS(diag_e) at Mon Dec 12 16:18:59 2016.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

```
set.seed(504)
mod4code <- '
data{
  int<lower=0> y[262];

}
parameters{
  real <lower=0> mu4;
  real<lower=1>  phi4;
}
model{
  mu4 ~ gamma(0.01,0.01);
  phi4 ~ gamma(0.01,0.01);
  for(i in 1:262){
    y[i] ~ neg_binomial_2(mu4,phi4);
  }
}
'
mod4dat <- list(y=madison)
mod4 <- stan(model_code=mod4code, data=mod4dat)
```

```
## C:/Users/Palma/Documents/R/win-library/3.2/StanHeaders/include/stan/math/rev/core/set_zero_all_adjoi
##
## SAMPLING FOR MODEL 'ffe59bb1b110ba36a1513916665912b9' NOW (CHAIN 1).
##
## Chain 1, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 1, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 1, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 1, Iteration:  600 / 2000 [ 30%]  (Warmup)
```

```
## Chain 1, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 1, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 1, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 1, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 1, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 1, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 1, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 1, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.61 seconds (Warm-up)
##                0.5 seconds (Sampling)
##                1.11 seconds (Total)
##
## [1] "The following numerical problems occured the indicated number of times after warmup on chain 1"
##                                                                                              count
## Exception thrown at line 14: neg_binomial_2_log: Location parameter is 0, but must be > 0!     3
## [1] "When a numerical problem occurs, the Hamiltonian proposal gets rejected."
## [1] "See http://mc-stan.org/misc/warnings.html#exception-hamiltonian-proposal-rejected"
## [1] "If the number in the 'count' column is small,  do not ask about this message on stan-users."
##
## SAMPLING FOR MODEL 'ffe59bb1b110ba36a1513916665912b9' NOW (CHAIN 2).
##
## Chain 2, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 2, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 2, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 2, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 2, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 2, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 2, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 2, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 2, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 2, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 2, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 2, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.601 seconds (Warm-up)
##                0.504 seconds (Sampling)
##                1.105 seconds (Total)
##
## [1] "The following numerical problems occured the indicated number of times after warmup on chain 2"
##                                                                                                     co
## Exception thrown at line 14: neg_binomial_2_log: Location parameter is 1.#INF, but must be finite!
## [1] "When a numerical problem occurs, the Hamiltonian proposal gets rejected."
## [1] "See http://mc-stan.org/misc/warnings.html#exception-hamiltonian-proposal-rejected"
## [1] "If the number in the 'count' column is small,  do not ask about this message on stan-users."
##
## SAMPLING FOR MODEL 'ffe59bb1b110ba36a1513916665912b9' NOW (CHAIN 3).
##
## Chain 3, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 3, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 3, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 3, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 3, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 3, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 3, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 3, Iteration: 1200 / 2000 [ 60%]  (Sampling)
```

```
## Chain 3, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 3, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 3, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 3, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.616 seconds (Warm-up)
##                0.522 seconds (Sampling)
##                1.138 seconds (Total)
##
## [1] "The following numerical problems occured the indicated number of times after warmup on chain 3"
##                                                                                                   count
## Exception thrown at line 14: neg_binomial_2_log: Location parameter is 0, but must be > 0!     3
## [1] "When a numerical problem occurs, the Hamiltonian proposal gets rejected."
## [1] "See http://mc-stan.org/misc/warnings.html#exception-hamiltonian-proposal-rejected"
## [1] "If the number in the 'count' column is small,  do not ask about this message on stan-users."
##
## SAMPLING FOR MODEL 'ffe59bb1b110ba36a1513916665912b9' NOW (CHAIN 4).
##
## Chain 4, Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 4, Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 4, Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 4, Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 4, Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 4, Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 4, Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 4, Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 4, Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 4, Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 4, Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 4, Iteration: 2000 / 2000 [100%]  (Sampling)
##  Elapsed Time: 0.6 seconds (Warm-up)
##                0.575 seconds (Sampling)
##                1.175 seconds (Total)
##
## [1] "The following numerical problems occured the indicated number of times after warmup on chain 4"
##                                                                                                        c
## Exception thrown at line 14: neg_binomial_2_log: Location parameter is 1.#INF, but must be finite!
## [1] "When a numerical problem occurs, the Hamiltonian proposal gets rejected."
## [1] "See http://mc-stan.org/misc/warnings.html#exception-hamiltonian-proposal-rejected"
## [1] "If the number in the 'count' column is small,  do not ask about this message on stan-users."
```

```r
print(mod4,pars=c("mu4","phi4"))
```

```
## Inference for Stan model: ffe59bb1b110ba36a1513916665912b9.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##      mean se_mean   sd 2.5%  25%  50%  75% 97.5% n_eff Rhat
## mu4  0.66    0.00 0.06 0.55 0.62 0.65 0.69  0.78  2387    1
## phi4 1.49    0.01 0.48 1.02 1.16 1.36 1.66  2.72  2647    1
##
## Samples were drawn using NUTS(diag_e) at Mon Dec 12 16:19:46 2016.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

So we can see that for hamilton, the posterior distribution for the dispersion parameter has mean of 2.13 and the the posterior distribution for the dispersion parameter for madison has mean of 1.49. Thus there is more overdispersion in the hamilton counts data than the madison data.

d) Use posetrior predictive checks to determine which model provides a better fit to our data. #####Poisson Model ##############

```
# Obtain posterior distribution of lambda for Hamilton
par(mfrow=c(1,2))
lambda1<- extract(mod1)$lambda1

# Construct and plot the posterior predicitve distribution, 95% PI
x.pred.a <- rpois(4000,lambda1)
hist(x.pred.a,breaks=seq(min(x.pred.a)-.5,max(x.pred.a)+.5,1),
col="lightgray",freq=FALSE,xlab=expression(tilde(x)),
main="Poisson Post. Predictive (Hamilton)")
x.pred.a.95 <- quantile(x.pred.a,c(.025,.975))
abline(v=x.pred.a.95+c(-.5,.5),lty=2,lwd=2,col="red")
mean(x.pred.a.95)
```

```
## [1] 1.5
```

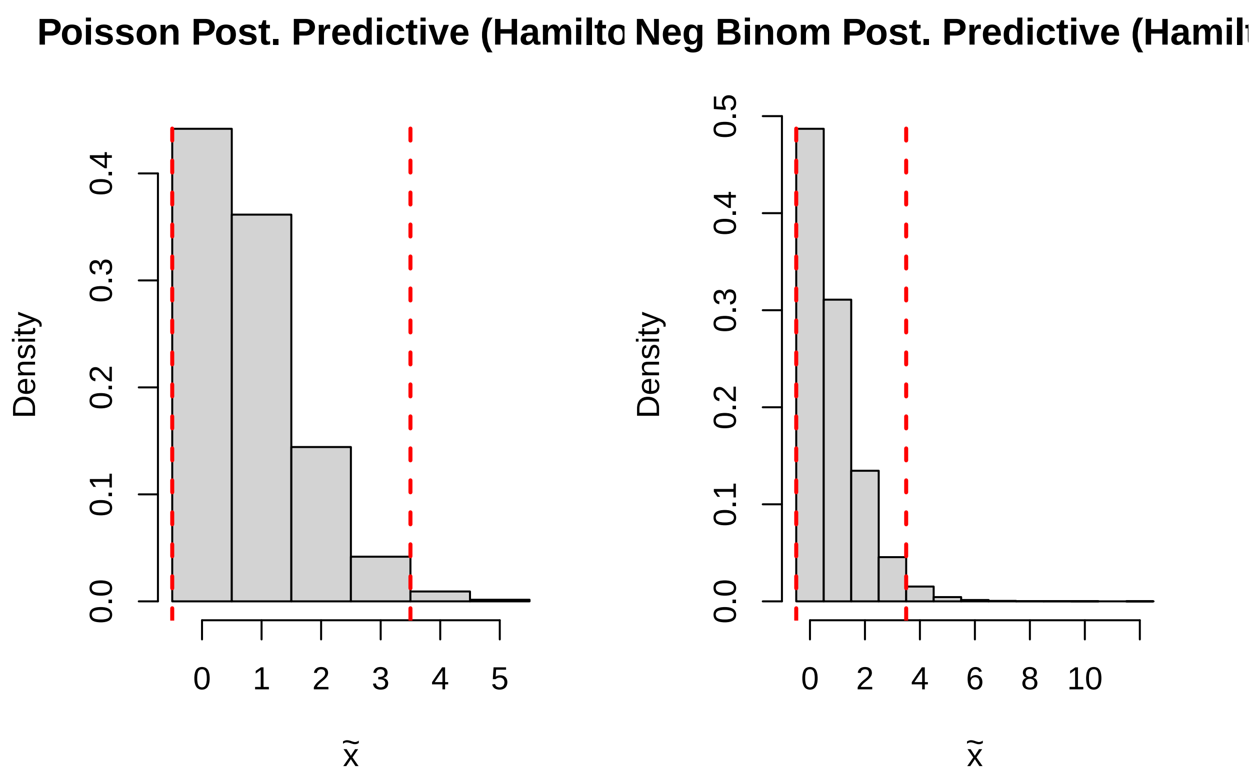
```
x.pred.a.95
```

```
##  2.5% 97.5%
##     0     3
```

```
# Obtain posterior distribution of lambda for Hamilton
mu3<- extract(mod3)$mu3
phi3<-extract(mod3)$phi3

# Construct and plot the posterior predicitve distribution, 95% PI
x.pred.a<-matrix(NA,4000,8)
for(i in 1:8){
x.pred.a[,i] <- rnbinom(4000,size=i,mu=mu3)
}

hist(x.pred.a,breaks=seq(min(x.pred.a)-.5,max(x.pred.a)+.5,1),
col="lightgray",freq=FALSE,xlab=expression(tilde(x)),
main="Neg Binom Post. Predictive (Hamilton)")
x.pred.a.95 <- quantile(x.pred.a,c(.025,.975))
x.pred.a.95
```

```
##  2.5% 97.5%
##     0     3
```

```
abline(v=x.pred.a.95+c(-.5,.5),lty=2,lwd=2,col="red")
```




```r
mean(x.pred.a.95)
```

```
## [1] 1.5
```

```r
# Obtain posterior distribution of lambda for Madison
par(mfrow=c(1,2))
lambda2<- extract(mod2)$lambda2
# Construct and plot the posterior predicitve distribution, 95% PI
x.pred.a <- rpois(4000,lambda2)
hist(x.pred.a,breaks=seq(min(x.pred.a)-.5,max(x.pred.a)+.5,1),
col="lightgray",freq=FALSE,xlab=expression(tilde(x)),
main=" Poisson Posterior Predictive (Madison)")
x.pred.a.95 <- quantile(x.pred.a,c(.025,.975))
abline(v=x.pred.a.95+c(-.5,.5),lty=2,lwd=2,col="red")
mean(x.pred.a.95)
```

```
## [1] 1.5
```

```r
x.pred.a.95
```

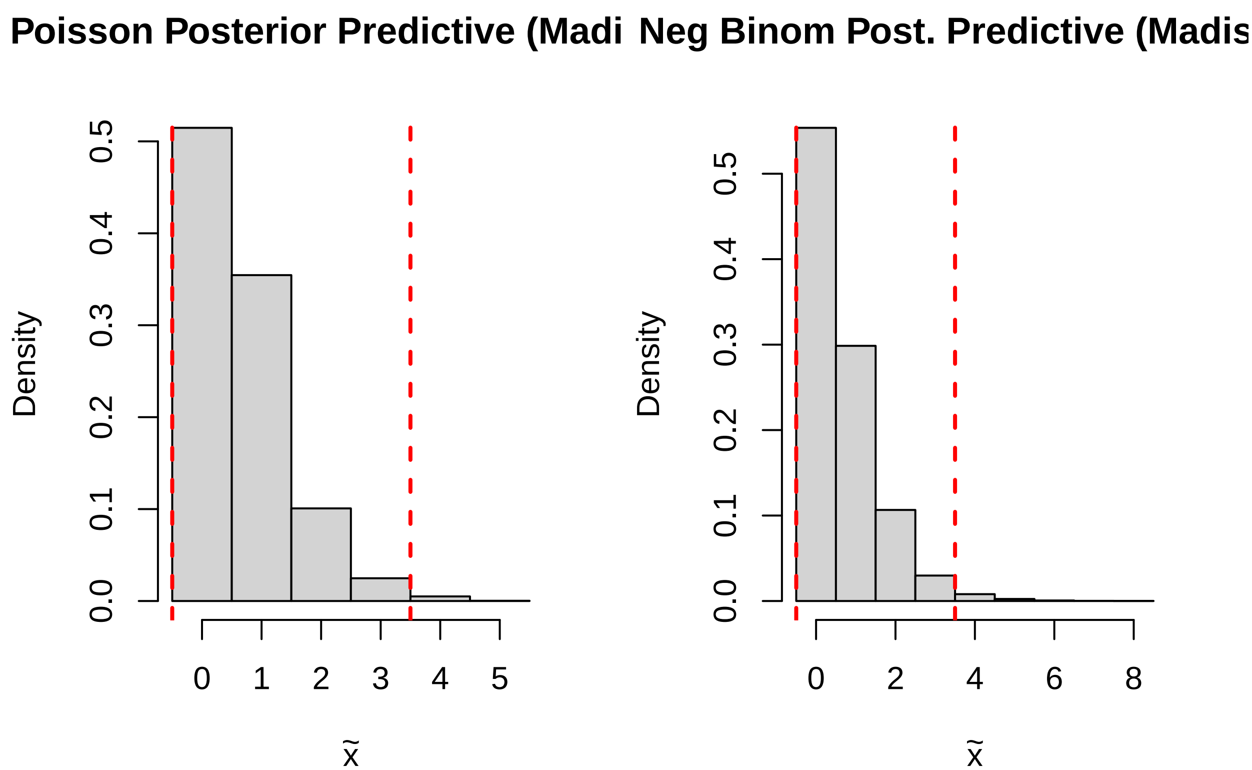
```
##  2.5% 97.5%
##     0     3
```

```r
# Obtain posterior distribution of lambda for Madison
mu4<- extract(mod4)$mu4
phi4<-extract(mod4)$phi4

# Construct and plot the posterior predicitve distribution, 95% PI
```

```
x.pred.a<-matrix(NA,4000,8)
for(i in 1:8){
x.pred.a[,i] <- rnbinom(4000,size=i,mu=mu4)
}

hist(x.pred.a,breaks=seq(min(x.pred.a)-.5,max(x.pred.a)+.5,1),
col="lightgray",freq=FALSE,xlab=expression(tilde(x)),
main="Neg Binom Post. Predictive (Madison)")
x.pred.a.95 <- quantile(x.pred.a,c(.025,.975))
x.pred.a.95
```

```
##  2.5% 97.5%
##     0     3
```

```
abline(v=x.pred.a.95+c(-.5,.5),lty=2,lwd=2,col="red")
```



Thus given our results from predictive distribution of both models we can see that the models perform well in predicting and telling us about the data . We can see the skewed distribution of the predictive model which closely aligns with our data. Both models seem to give indistinguishable results and this confirms the assertion that the Negative binomial converges to a poission under certain limiting conditions. The Negative binomial is better for 2 reasons : Unlike the poisson model ,It does not require the mean of the data to be equal to the variance for this model to work well

Secondly , it at least has a dispersion parameter which tell us some information about the dispersion in the data , which is very useful for future extrapolations and analysis .