Bayesian Poisson Regression of Fatal Accidents

Greg Johnson

We have a city's data on the number of fatal traffic accidents between 1976 and 1985.

Simple Poisson Model

Let's assume the number of fatal accidents independently follow a Poisson distribution. Because of lack of background information, we wish to use a noninformative prior, specifically we want to derive Jeffry's prior and derive the posterior distribution.

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta) \propto L(\theta)J(\theta)^{1/2} = \exp(-n\theta)\theta^{\sum y_i} J(\theta)^{1/2}$$

$$J(\theta) = -E\left(\frac{d^2}{d\theta^2}\log p(\mathbf{y}|\theta)\middle|\theta\right) = -E\left(\frac{d^2}{d\theta^2}\left[-n\theta + \sum y_i \log\theta - \sum \log y_i!\right]\middle|\theta\right)$$

$$= -E\left(\frac{d}{d\theta}\left[-n + \frac{1}{\theta}\sum y_i\right]\middle|\theta\right) = E\left(\frac{1}{\theta^2}\sum y_i\middle|\theta\right) = \frac{1}{\theta^2}\sum E(y_i|\theta) = \frac{n\theta}{\theta^2} = \frac{n}{\theta}$$

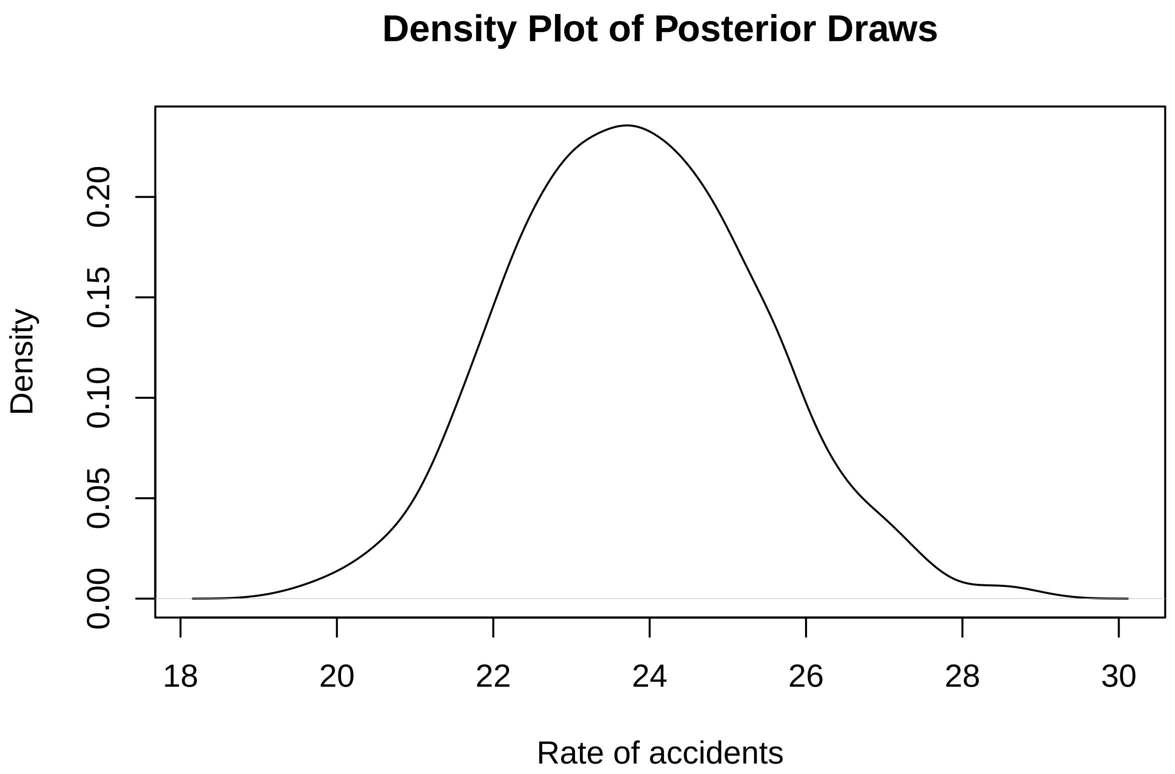
$$p(\theta|\mathbf{y}) \propto \exp(-n\theta)\theta^{\sum y_i}\left(\frac{n}{\theta}\right)^{1/2} \propto \exp(-n\theta)\theta^{(\sum y_i - 1/2)}$$

We recognize the parameteric form of the posterior: $\theta|\mathbf{y} \sim \text{Gamma}(\sum y_i + 1/2, -n)$

Sampling from Posterior

```r
dat1 = data.frame(year=seq(1976,1985),accidents=c(24,25,31,31,22,21,26,20,16,22))

#posterior draws using Monte Carlo
post_draws = rgamma(2000,shape=sum(dat1$accidents)+.5,rate=nrow(dat1))
#density plot
dpost = density(post_draws,adjust=1)
plot(dpost,xlab="Rate of accidents",main="Density Plot of Posterior Draws")
```



```
#95% credible intervals
CredInt95 = quantile(post_draws,c(.025,.975))
CredInt95
```

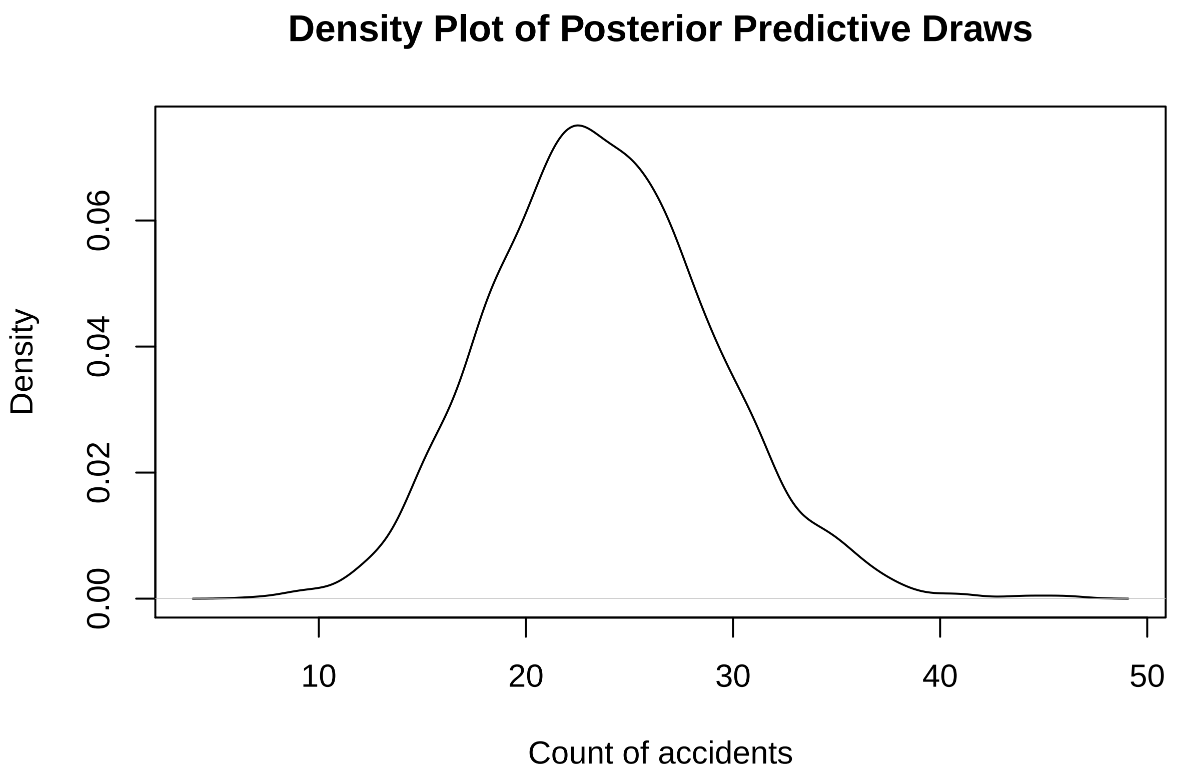
```
##     2.5%    97.5%
## 20.81071 27.02759
```

```
#MAP estimate
i = which.max(dpost$y)
post_mode = dpost$x[i]
post_mode
```

```
## [1] 23.7022
```

Sampling from Posterior Predictive

```
#Monte Carlo simulation of posterior predictive
postpred_draws = rpois(length(post_draws),post_draws)
dpred = density(postpred_draws,adjust=1)
plot(dpred,xlab="Count of accidents",main="Density Plot of Posterior Predictive Draws")
```



```
#95% credible intervals
CredInt95 = quantile(postpred_draws,c(.025,.975))
CredInt95
```

```
##  2.5% 97.5%
##    14    35
```

```
#MAP estimate
i = which.max(dpred$y)
pred_mode = dpred$x[i]
pred_mode
```

```
## [1] 22.47815
```

So under the model that the number of accidents every year is iid Poisson with the same rate, our MAP estimate for the number of fatal accidents in 1986 is just the most likely value in our posterior predictive distribution, 22.4781454.

Bayesian Poisson Regression

Let the number of fatal accidents in each year t follow a Poisson where the rate parameter is linked to the covariate year by the logarithmic link function.

Derive Posterior

The default noninformative prior for α and β is the product of independent uniform distributions which is appropriate given that either parameter can (potentially) take on any values in the Reals.

$$p(\alpha, \beta) = p(\alpha)p(\beta) \propto 1$$

Our likelihood takes a simple form due to the fact that the Poisson distribution has no variance parameter to deal with.

$$L(\alpha,\beta) = \prod_{i=1}^{n} p(y_i|\alpha,\beta,t_i) = \prod_{i=1}^{n} \text{Poisson}\big(\theta_i\big) = \prod_{i=1}^{n} \text{Poisson}\big(\exp(\alpha+\beta t_i)\big)$$

$$\propto \prod_{i=1}^{n} \exp\Big\{-\exp(\alpha+\beta t_i)\Big\}\Big(\exp(\alpha+\beta t_i)\Big)^{y_i} = \prod_{i=1}^{n} \exp\Big\{-\exp(\alpha+\beta t_i)+\alpha y_i+\beta t_i y_i\Big\}$$

$$L(\alpha,\beta) \propto \exp\Big\{-\sum_{i=1}^{n}\exp(\alpha+\beta t_i)+\alpha\sum_{i=1}^{n} y_i+\beta\sum_{i=1}^{n} t_i y_i\Big\}$$

Thus our unnormalized posterior density takes the form:

$$p(\alpha,\beta|\mathbf{y}) \propto \exp\Big\{-\sum_{i=1}^{n}\exp(\alpha+\beta t_i)+\alpha\sum_{i=1}^{n} y_i+\beta\sum_{i=1}^{n} t_i y_i\Big\}$$

Formulating the Metropolis-Hastings Algorithm

One of the most important decisions for implementing MH is the jumping distribution. Given the fact that both parameters have no boundaries on their respective spaces, the bivariate normal is a good, default choice. As a quick check (and because we have the luxury of having only two parameters), let's plot the unnormalized density on a grid centered on estimates of α and β found by IRLS.

```r
y = dat1[["accidents"]]
t = dat1[["year"]]-1976 #number of years since 1976 to avoid overflow
n = nrow(dat1)


glm1 = glm(accidents~I(year-1976),family="poisson",data=dat1)
coef(glm1)
```

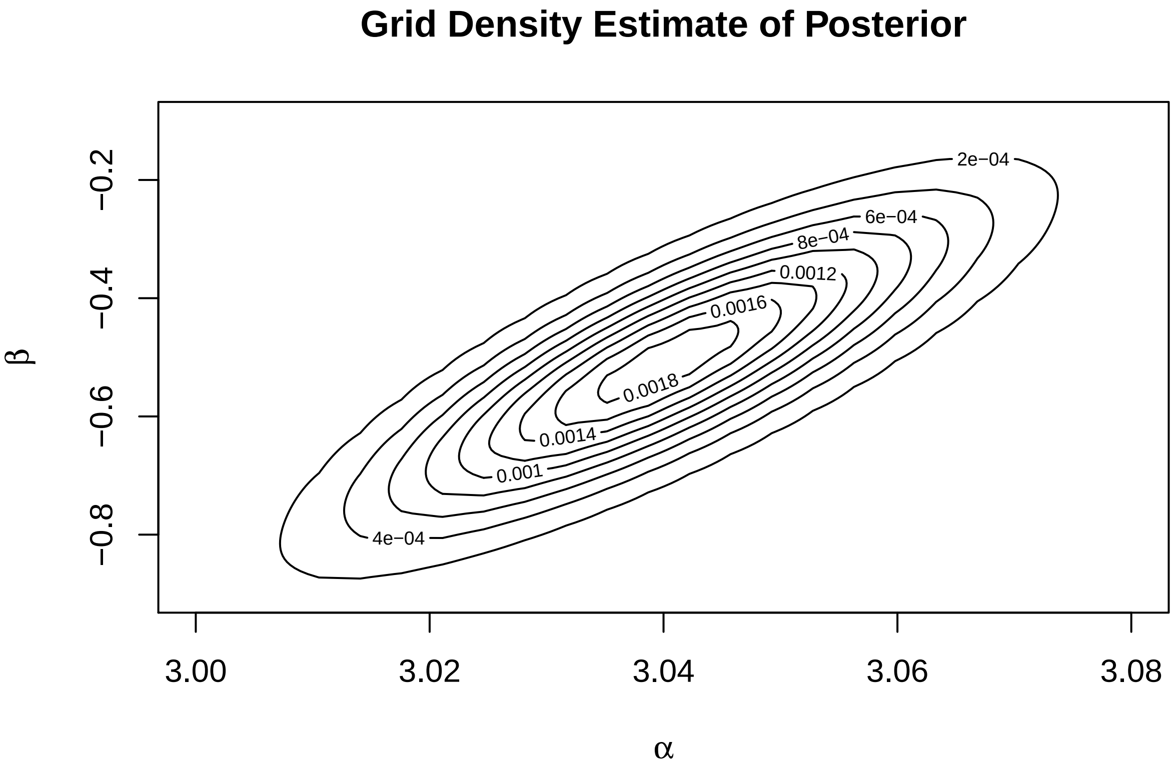
```
##    (Intercept) I(year - 1976)
##      3.3381019     -0.0388044
```

```r
alpha_grid = seq(3,3.7,length.out=200)
beta_grid = seq(.02,-1,length.out=200)
post_grid = matrix(NA,200,200)

#unnormalized density
post_density = function(a,b){
  exp( -sum(exp(a+b*t)) + a*sum(y) + b*sum(y*t) )
}

for(i in 1:200){
  for(j in 1:200){
    a = alpha_grid[j]
    b = beta_grid[i]

    post_grid[i,j] = post_density(a,b)
  }
}
plot(alpha_grid,beta_grid,type="n",xlim=c(3,3.08),ylim=c(-.9,-.1),main="Grid Density Estimate of Posteri
contour(alpha_grid,rev(beta_grid),post_grid/sum(post_grid),add=TRUE)
```



The distribution looks elliptical, suggesting that a bivariate Gaussian would be a good proposal density for Metropolis Hastings (which would actually just be Metropolis given the symmetry of our proposal density). Our algorithm will proceed as follows:

1. Draw a starting point $\boldsymbol{\theta}^0 = (\alpha^0, \beta^0)$ from a starting distribution. In our case we will just start with the estimates obtained from fitting the data with IRLS: $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}) = (3.34, -0.04)$.
2. For $t = 1, 2, ...$ repeat steps 3-5:
3. Sample proposal $\boldsymbol{\theta}^*$ from the proposal distribution $N_2\Big(\boldsymbol{\mu} = (\alpha^{t-1}, \beta^{t-1}), \Sigma\Big)$. Σ will be tuned according to acceptance ratio and autocorrelation.
4. Calculate:

$$r = \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}^{t-1}|\mathbf{y})}$$

5. Update $\boldsymbol{\theta}^t \leftarrow \boldsymbol{\theta}^*$ with probability $\min(r, 1)$. Otherwise retain $\boldsymbol{\theta}^{t-1}$.

Sampling Posterior

Implement Metropolis

```
mh_poisson = function(start=c(3.34,-0.04),ndraws,jump_cov){

  #to collect the output from the MH loop
  aratio = rep(1,ndraws) #by default, all iterations rejected
  theta_draws = matrix(NA,ndraws,2)

  #Provided starting value
  theta = start
  for(i in 1:ndraws){
    #1: Proposal
    prop = rmvnorm(1,mean=theta,sigma=jump_cov)
    #2: Ratio
    r = post_density(prop[1],prop[2])/post_density(theta[1],theta[2])
    #3: Stochastic update
    if(runif(1)<min(r,1)){
      theta = prop
      aratio[i] = 0 #indicate NOT rejected
    }
    #Save
    theta_draws[i,] = theta
  }

  return(list(
    "Acceptance Ratio" = 1-mean(aratio),
    "Posterior Draws" = theta_draws
  ))
}
```

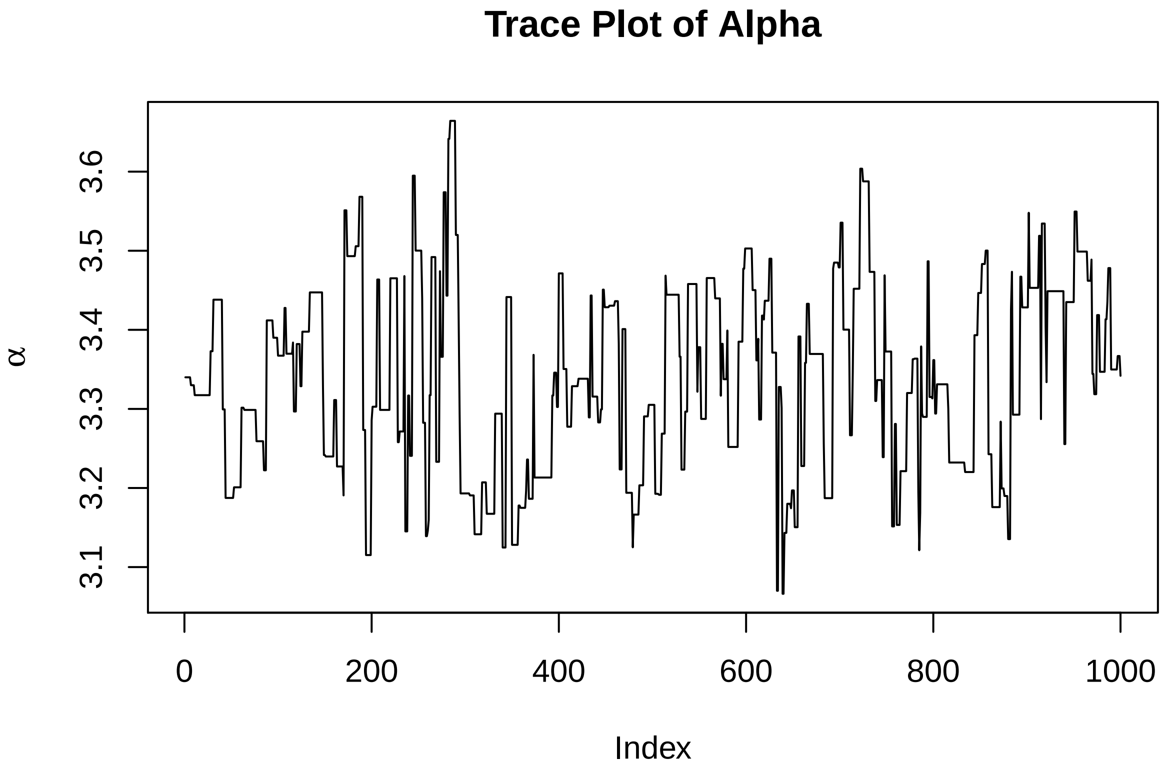
After some fiddling with the tuning parameter, I found that $\Sigma = \text{diag}(.05, .001)$ produces the best acceptance ratio, approximately 23%.

```
set.seed(1)
post_draws = mh_poisson(ndraws=1000,jump_cov=diag(c(.05,.001)))
post_draws$`Acceptance Ratio`
```

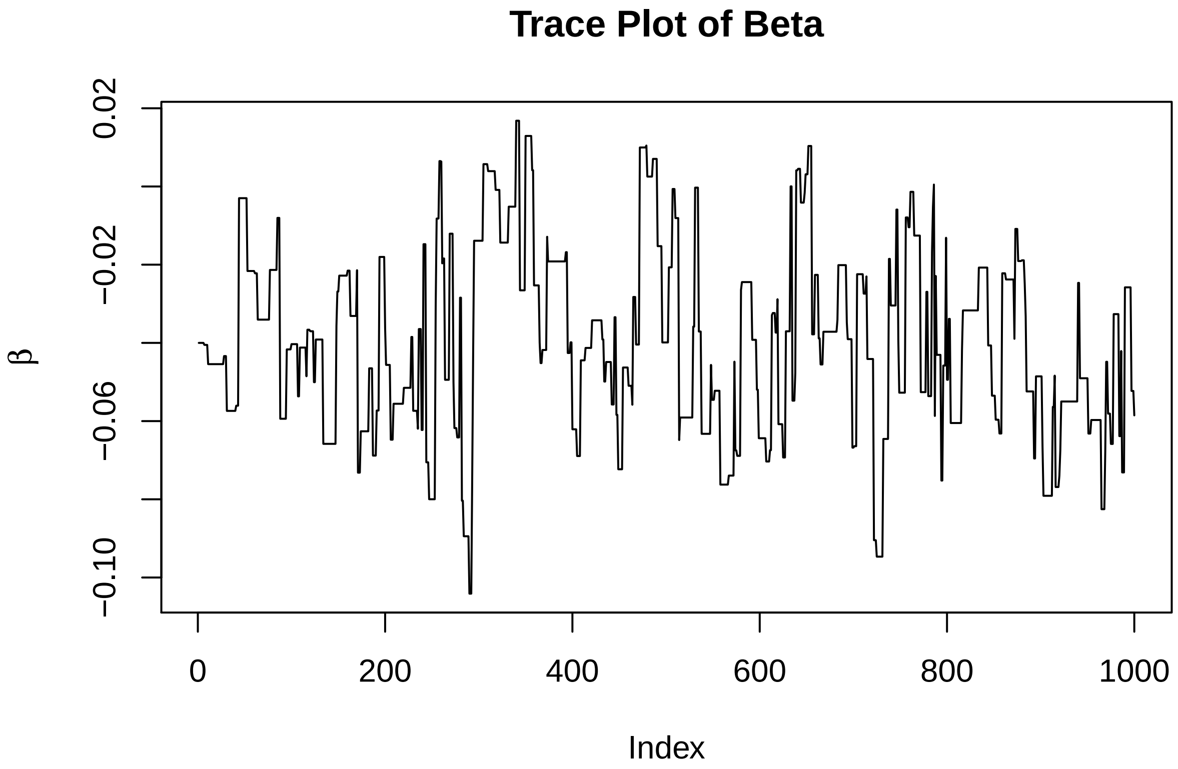
```
## [1] 0.23
```

Convergence: Trace Plots

```
plot(post_draws$`Posterior Draws`[,1],type="l",xlab="Index",ylab=expression(alpha),main="Trace Plot of
```



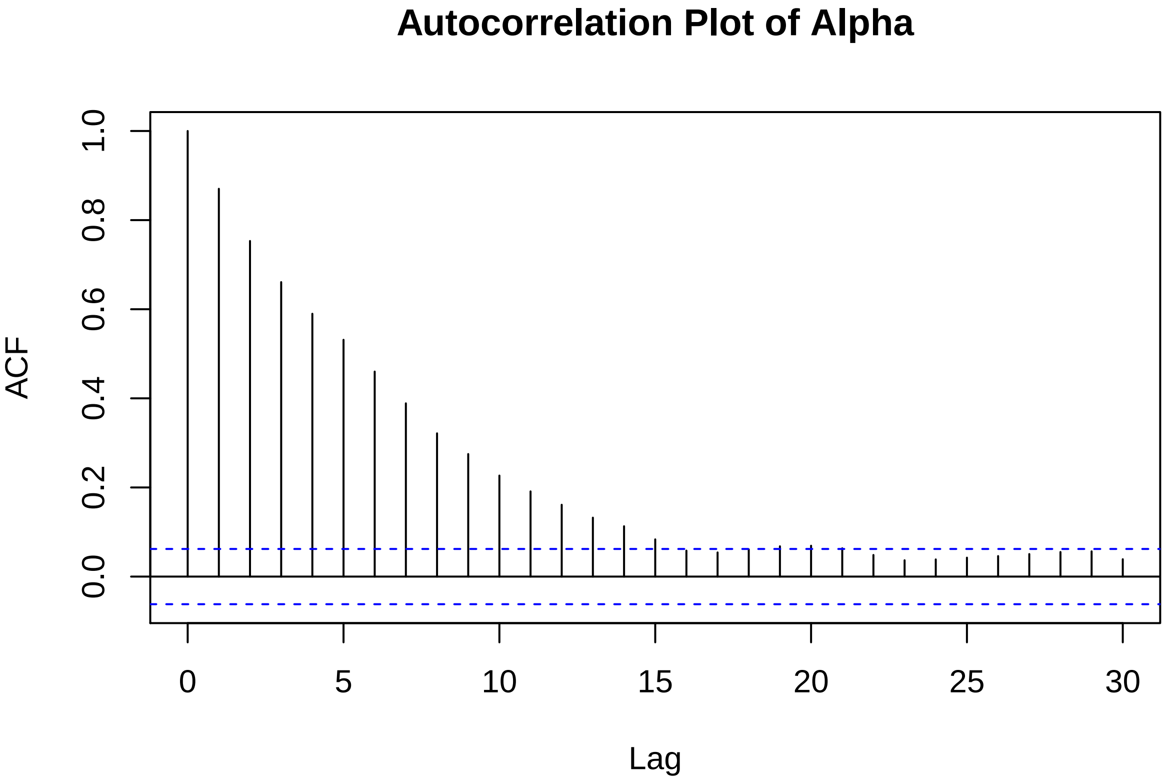
```
plot(post_draws$`Posterior Draws`[,2],type="l",xlab="Index",ylab=expression(beta),main="Trace Plot of Be
```



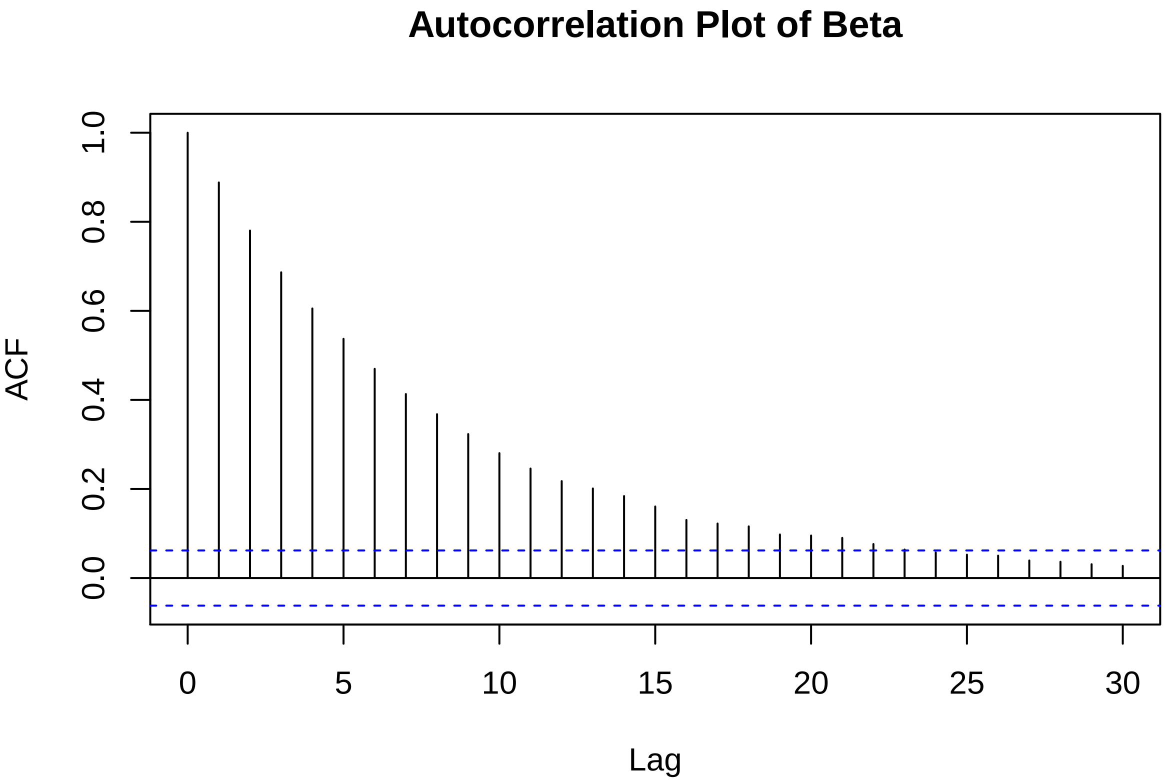
Looks like no burn-in is necessary - the Markov Chain has converged from the start. This is probably because we started with the IRLS estimates.

Effective Sample Size: Autocorrelation & Thinning

```r
acf(post_draws$`Posterior Draws`[,1],main="Autocorrelation Plot of Alpha")
```



```r
acf(post_draws$`Posterior Draws`[,2],main="Autocorrelation Plot of Beta")
```

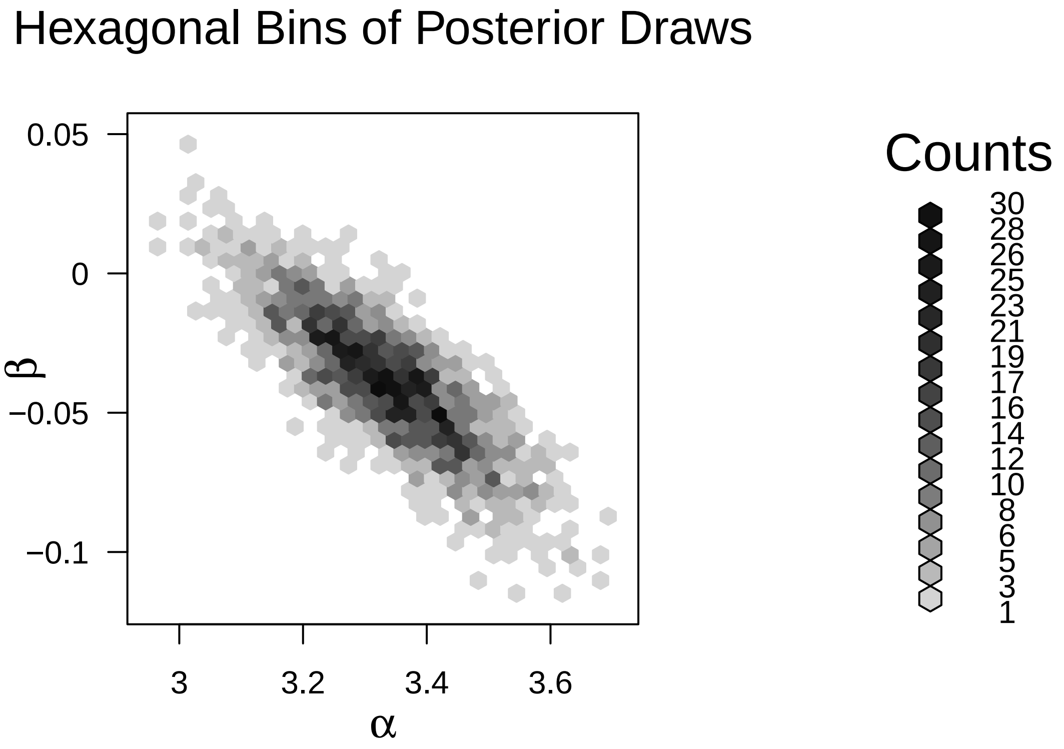


Looks like β has the greater lag of about 20. So if we thin down to every 20th draw then we can remove any autocorrelation from the MCMC. This means that to get an effective sample size of 2000 we need 40000 draws.

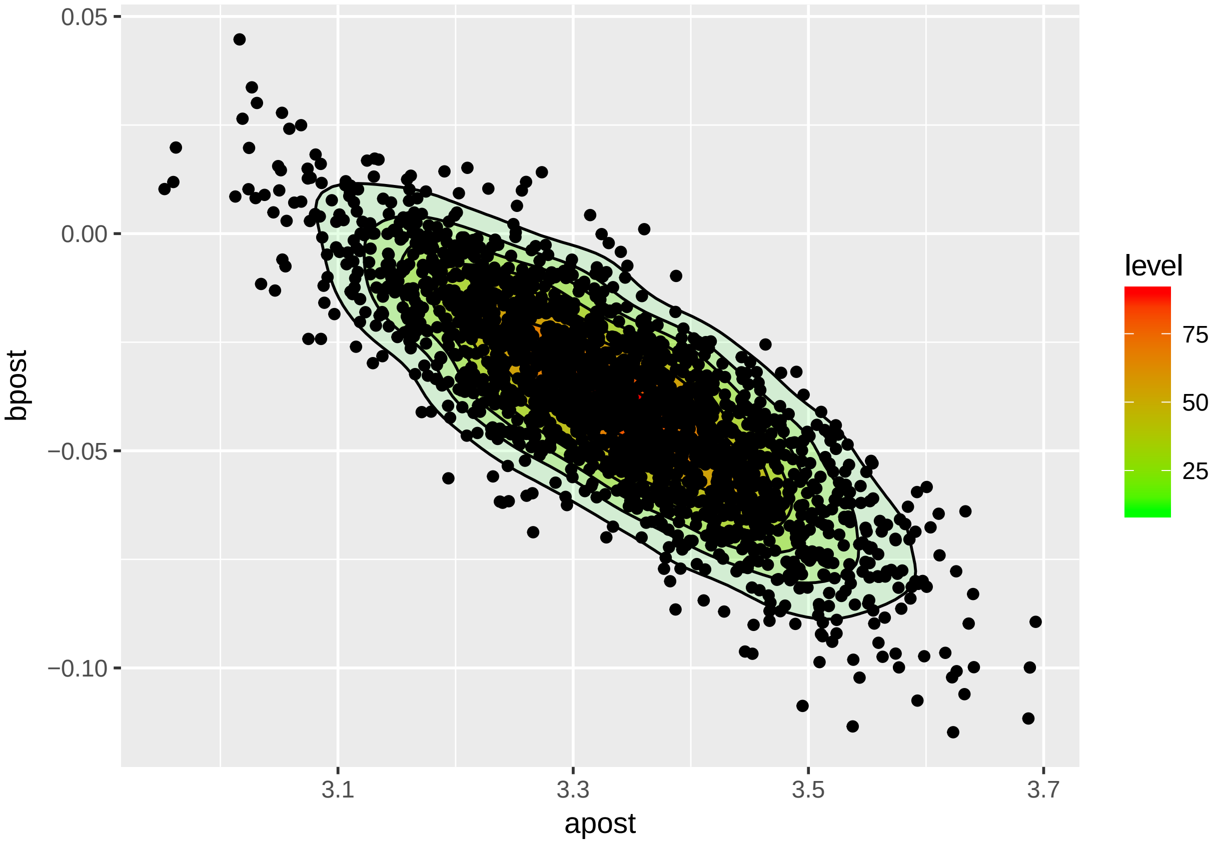
```
set.seed(1)
post_draws = mh_poisson(ndraws=40000,jump_cov=diag(c(.05,.001)))
alpha_post = post_draws$`Posterior Draws`[seq(1,40000,20),1]
beta_post = post_draws$`Posterior Draws`[seq(1,40000,20),2]
```

Joint and Marginal Posterior Densities

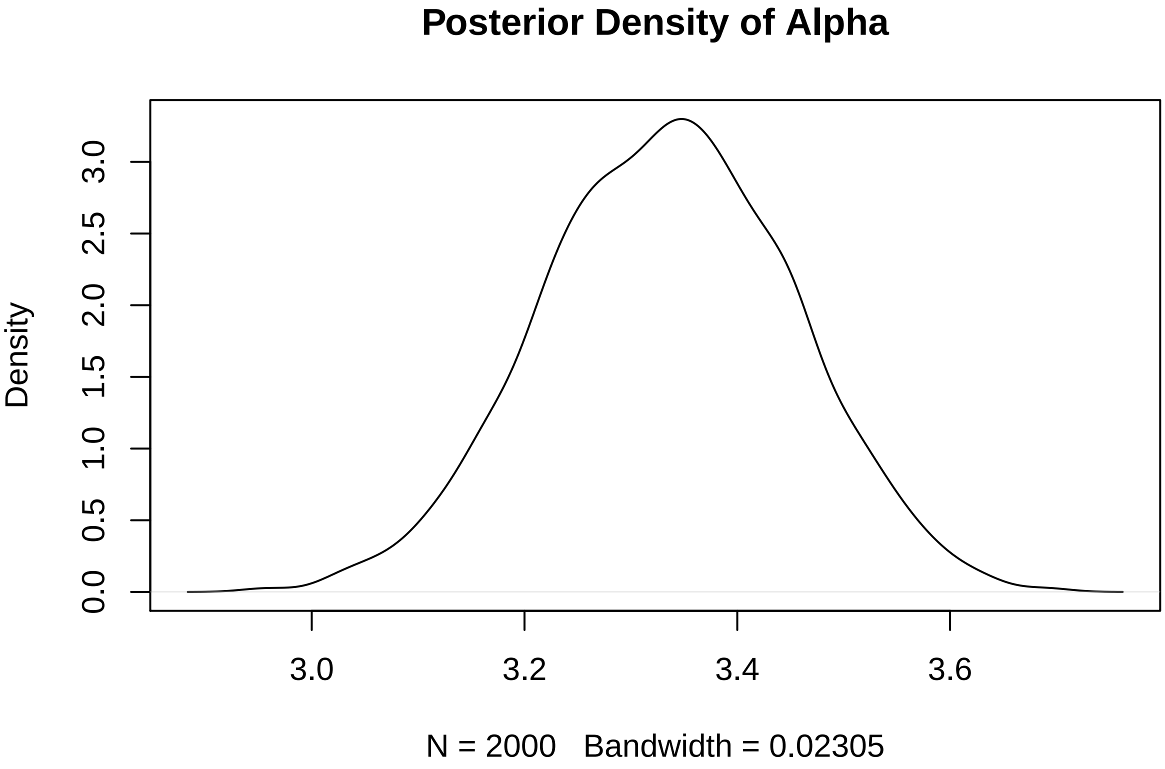
```
#bivariate histogram - hexagonal binning
plot(hexbin(alpha_post,beta_post),xlab=expression(alpha),ylab=expression(beta),main="Hexagonal Bins of P
```



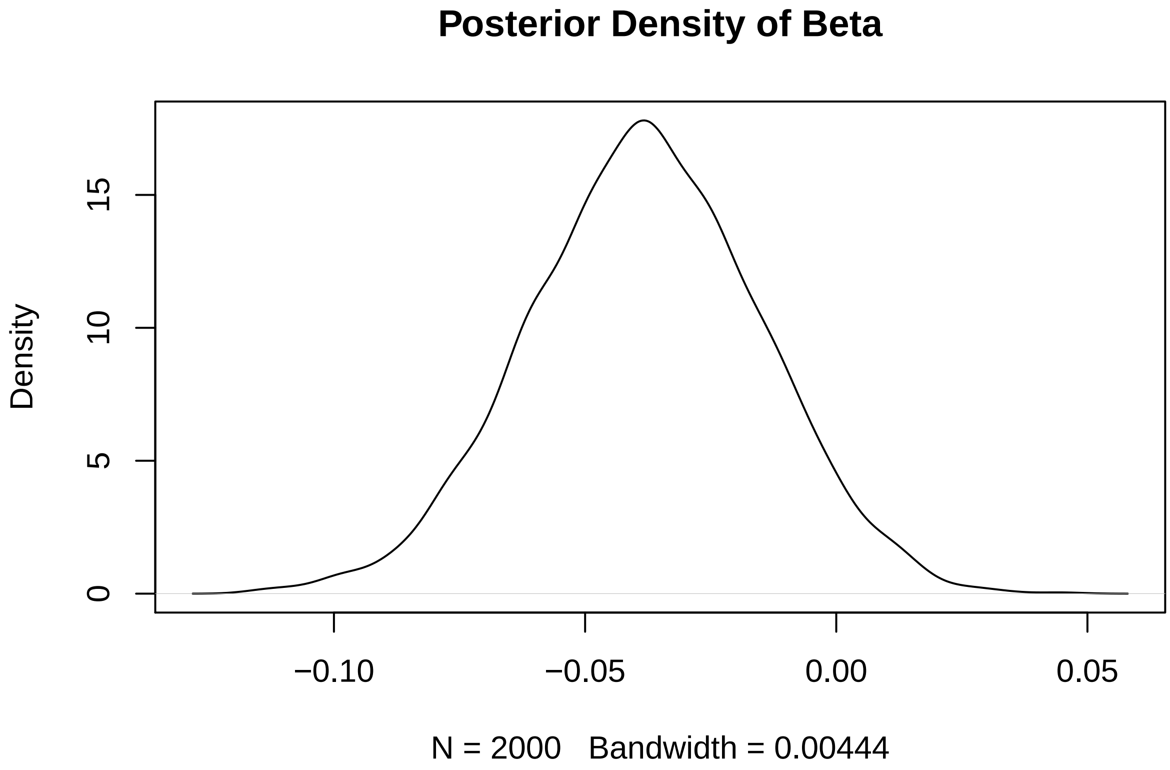
```
ggplot(data=data.frame(apost=alpha_post,bpost=beta_post),aes(apost,bpost)) + stat_density2d(aes(fill=..l
```



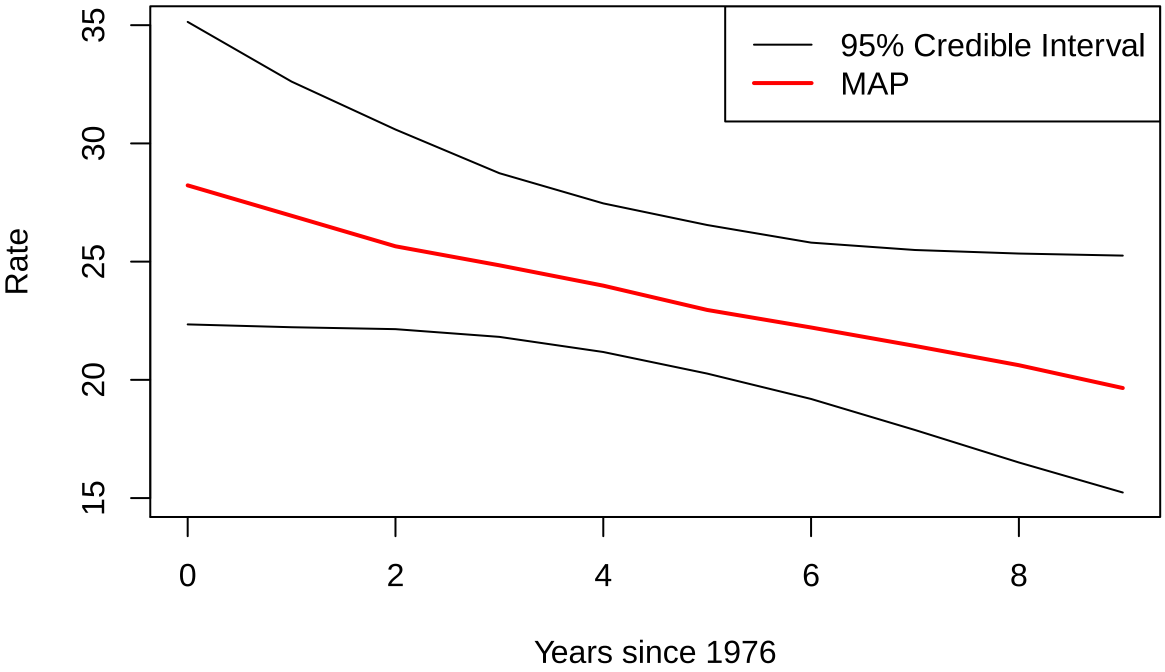
```r
plot(density(alpha_post,adjust=1),main="Posterior Density of Alpha")
```



```r
plot(density(beta_post,adjust=1),main="Posterior Density of Beta")
```



```r
Results = matrix(NA,10,3)
colnames(Results)=c("Lower 95CI","Upper 95CI","MAP")
rownames(Results)=seq(1976,1985)
for(i in 1:10){
  temp = exp(alpha_post+beta_post*t[i])
  Results[i,1:2] = quantile(temp,c(.025,.975))
  dpost = density(temp,adjust=1)
  j = which.max(dpost$y)
  Results[i,3] = dpost$x[j]
}
plot(0:9,Results[,3],type="l",lwd=2,col="red",ylim=c(15,35),xlab="Years since 1976",ylab="Rate")
lines(x=0:9,Results[,1])
lines(x=0:9,Results[,2])
legend("topright",c("95% Credible Interval","MAP"),col=c("black","red"),lwd=c(1,2))
```

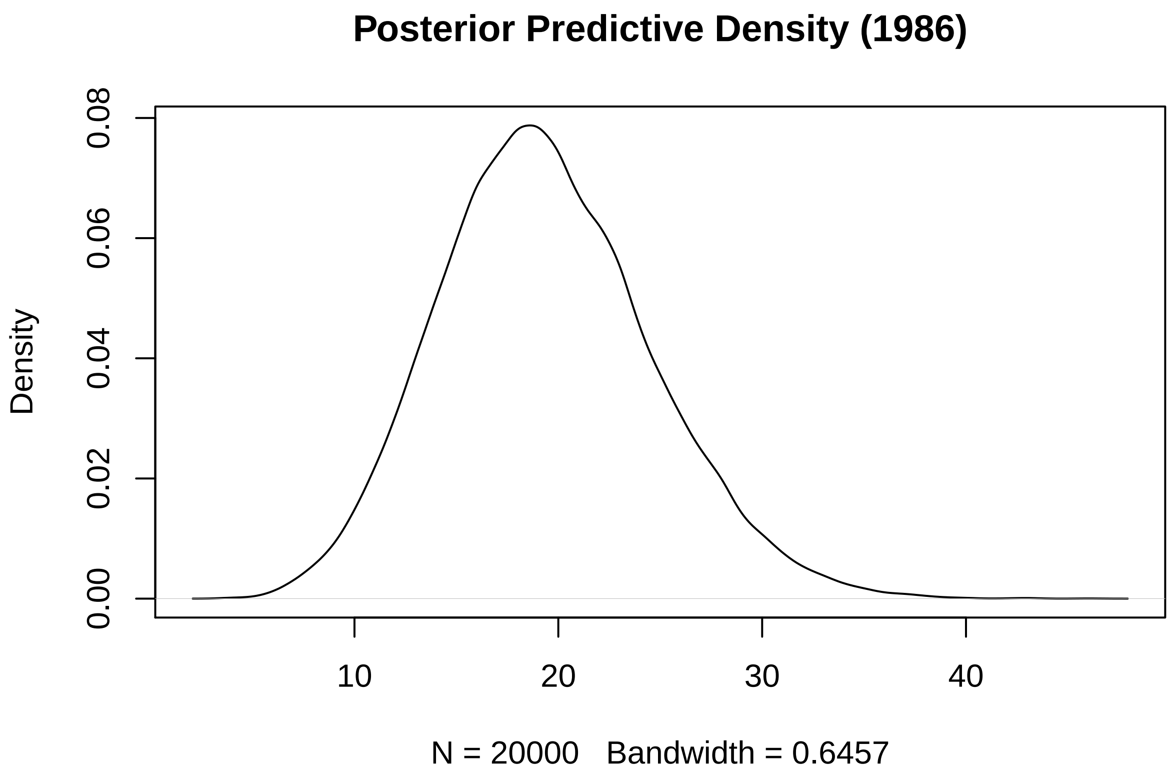


It looks like over time, the average rate of fatal accidents has decreased, almost linearly.

Sampling from Posterior Predictive

Let's use posterior samples of α and β to predict the number of fatal accidents in 1986.

```
postpred = rpois(20000,exp(alpha_post+beta_post*10))
plot(density(postpred,adjust=1),main="Posterior Predictive Density (1986)")
```



```
CredInt95 = quantile(postpred,c(.025,.975))
CredInt95
```

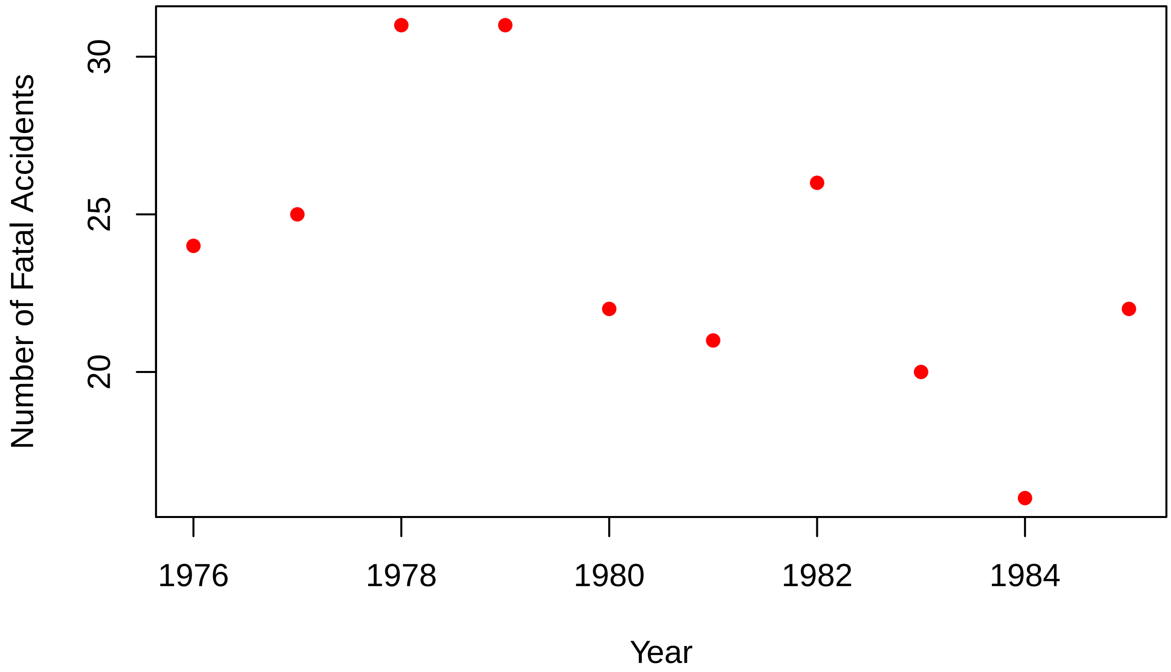
```
##  2.5% 97.5%
##    10    30
```

```
MAP = names(which.max(table(postpred)))
MAP
```

```
## [1] "19"
```

The point estimate (MAP) for the Poisson model is less than that for the iid model. Why? Because the Poisson model is incorporating information from the time covariate and a simple glance at a plot of time versus number of fatal accidents reveals that a relationship exists between time and accidents. In fact, it appears to be a linear downward trend - something that the iid model completely misses.

```
plot(dat1[,1],dat1[,2],xlab="Year",ylab="Number of Fatal Accidents",col="red",pch=16)
```


Number of Fatal Accidents
20
25
30
1976
1978
1980
1982
1984
Year