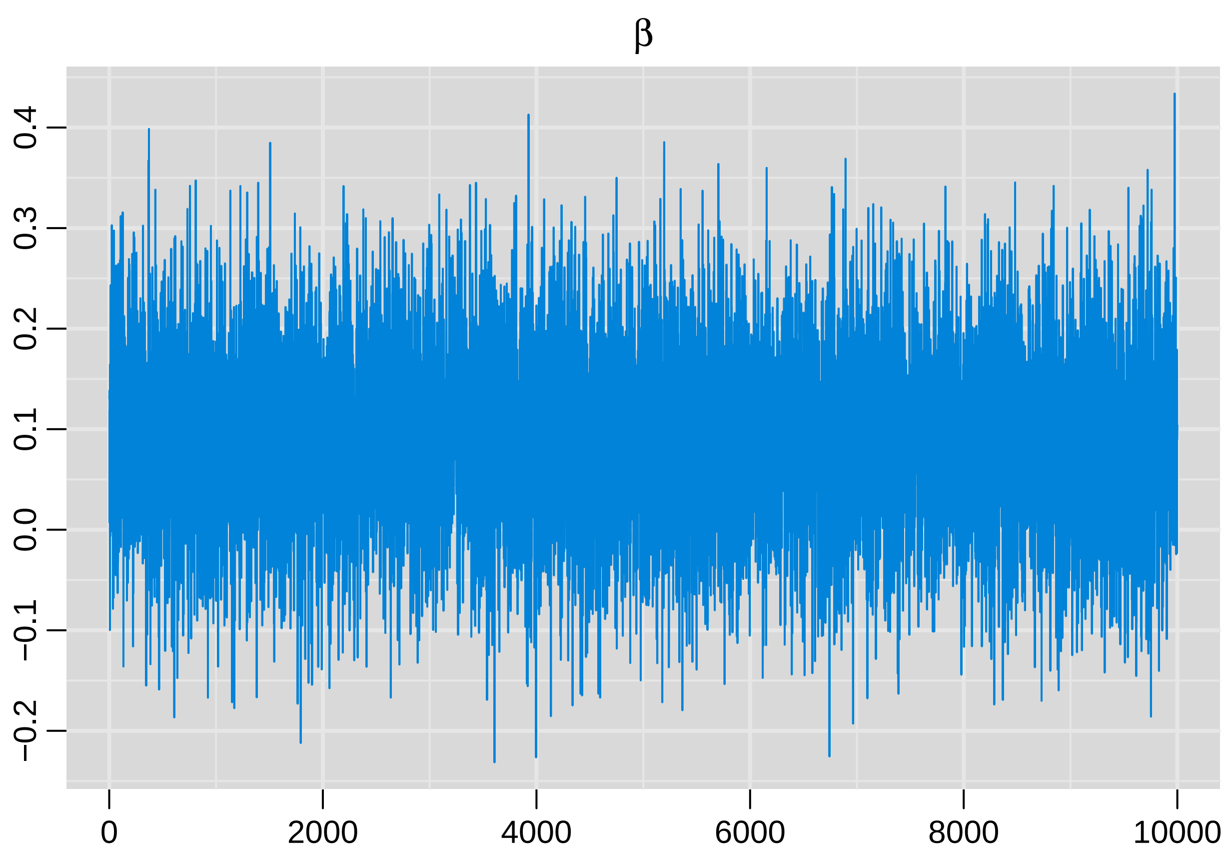
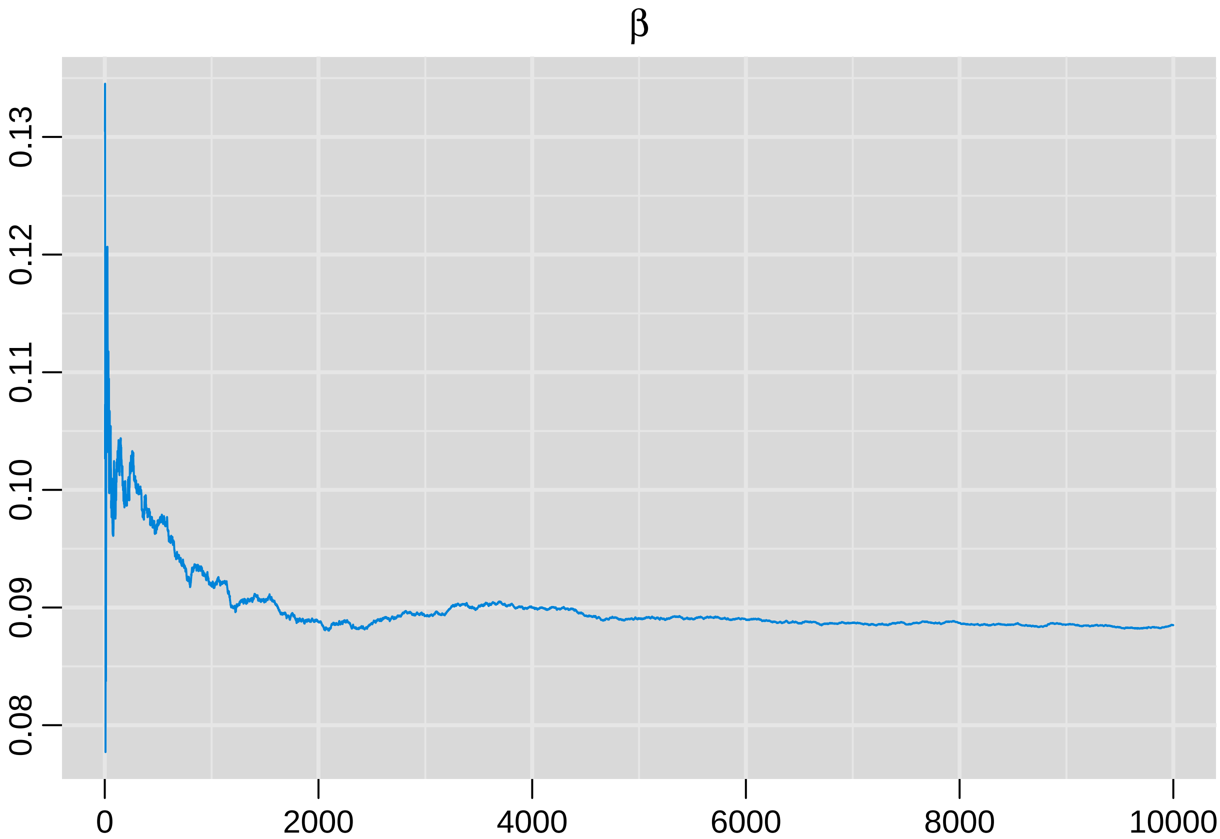
```
## ##
## ## Markov Chain Monte Carlo Package (MCMCpack)
## ## Copyright (C) 2003-2019 Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park
## ##
## ## Support provided by the U.S. National Science Foundation
## ## (Grants SES-0350646 and SES-0350613)
## ##
```

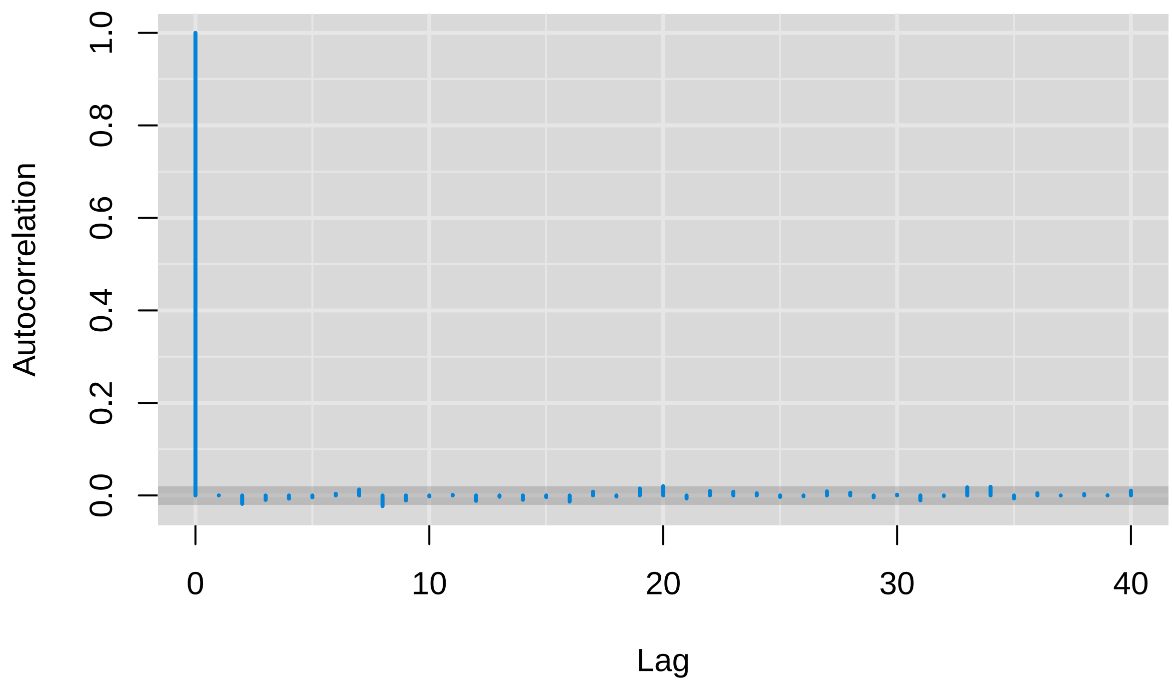
```
B <- 10000
ff <- read.table("forestfire.txt", header=TRUE)

Beta <- log( mean(ff[,1]) / (1-mean(ff[,1])) )
info <- 1/ ( mean(ff[,1])*( 1 - mean(ff[,1]) )*nrow(ff) )

set.seed(821); r.beta <- rnorm(B, mean=Beta, sd= sqrt(info) )

rmuList <- mcmc.list(list(mcmc(r.beta)))
mat <- matrix(r.beta , ncol = 1); colnames( mat ) <- "beta"
traplot( mat , greek = T )
```




```
rmeanplot(mat , greek = T)
```



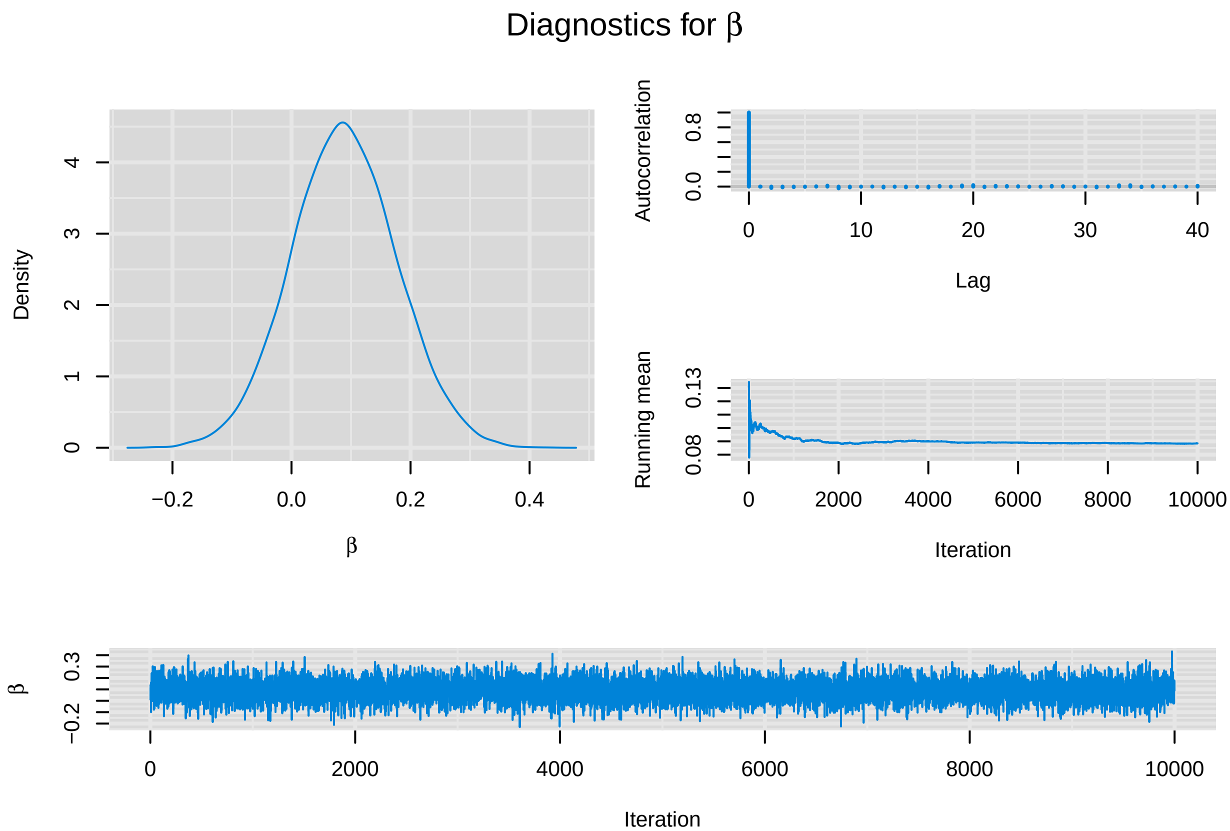

```r
autplot1( rmuList )
```




```r
geweke.diag( r.beta )
```

```
##
## Fraction in 1st window = 0.1
## Fraction in 2nd window = 0.5
##
## var1
## 1.25
```

```r
mcmcplot1(mat, greek = T)
```

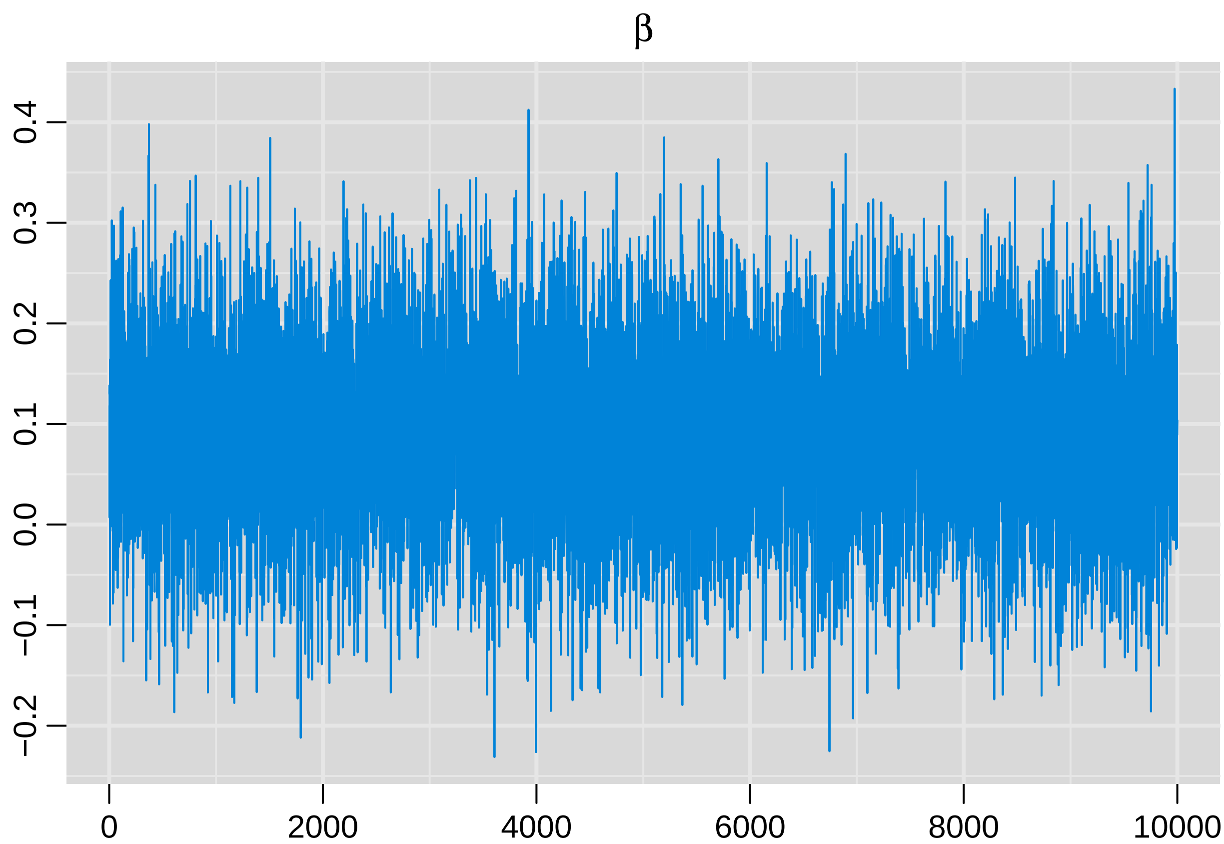


For the trace plot we see no trend and random scatter. The running means plot shows the running mean flattening out as the number of samples increases. And the ACF plot shows independence of our samples, as the vertical lines drop between the blue dashed lines almost immediately. The Geweke Convergence Diagnostic is also showing convergence with a z-score that is within two standard devations from zero.

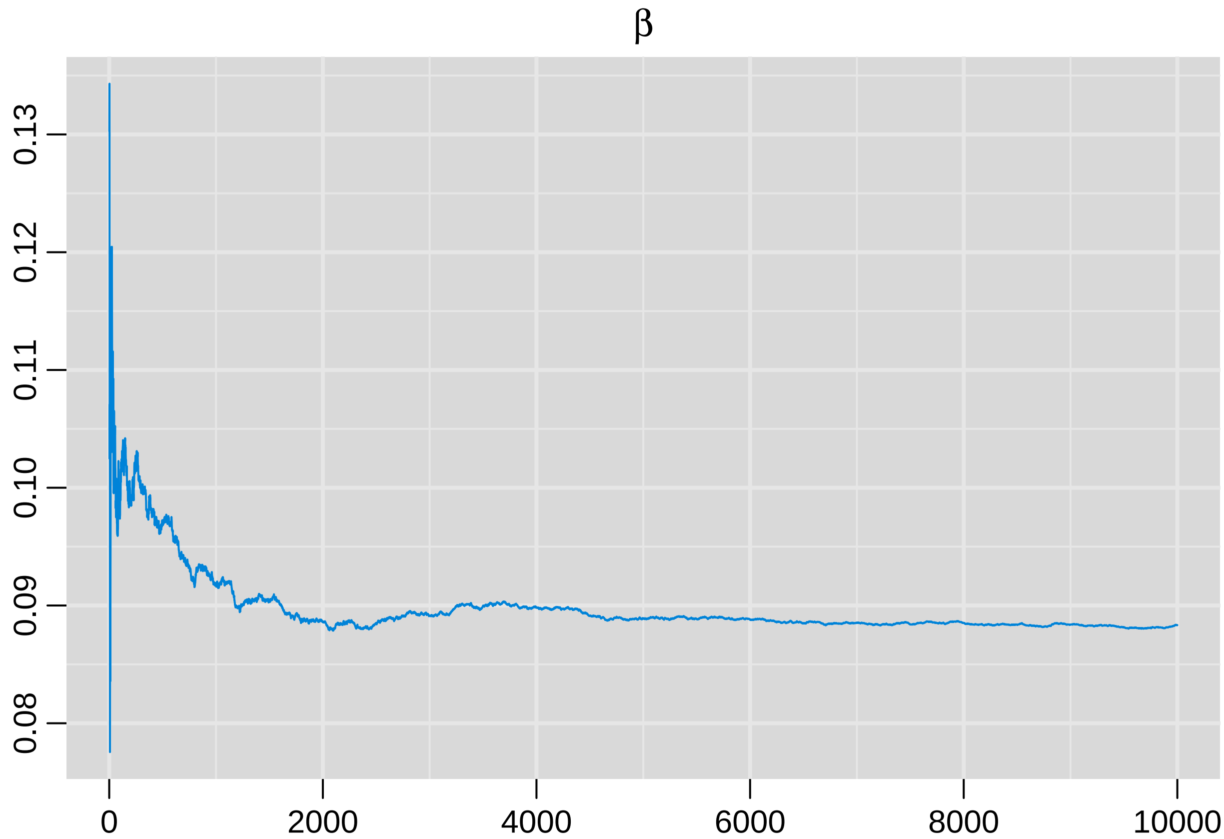
```
n <- nrow(ff)
ybar <- mean(ff[,1])

BJeff <- log( ( n*ybar + .5 )  /  ( n + .5 - n*ybar )  )
IJeff <-  (n+1) * ( ( exp(BJeff) ) /  ( 1 + exp(BJeff) )^2 )
1/IJeff
```

```
## [1] 0.007737262
```

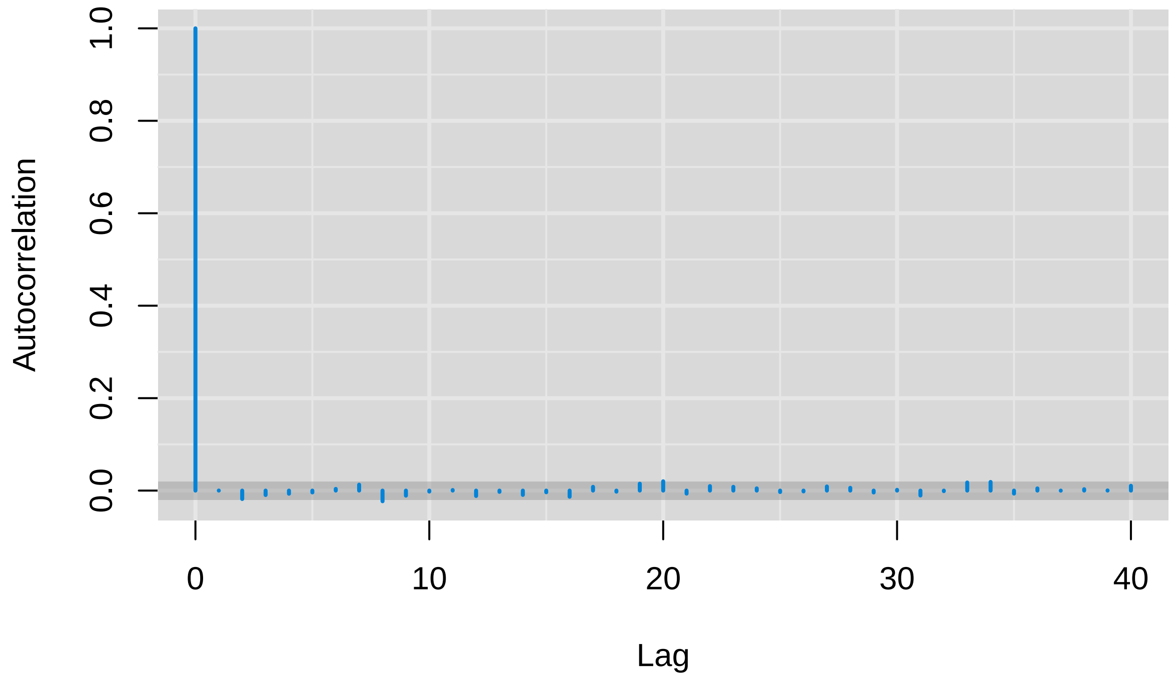
```
set.seed(821); rj.beta <- rnorm(B , mean=BJeff , sd= 1/sqrt(IJeff ) )

rmuListj <- mcmc.list(list(mcmc(rj.beta)))
matj <- matrix(rj.beta , ncol = 1); colnames( matj ) <- "beta"
traplot( matj , greek = T )
```




```
rmeanplot(matj , greek = T)
```



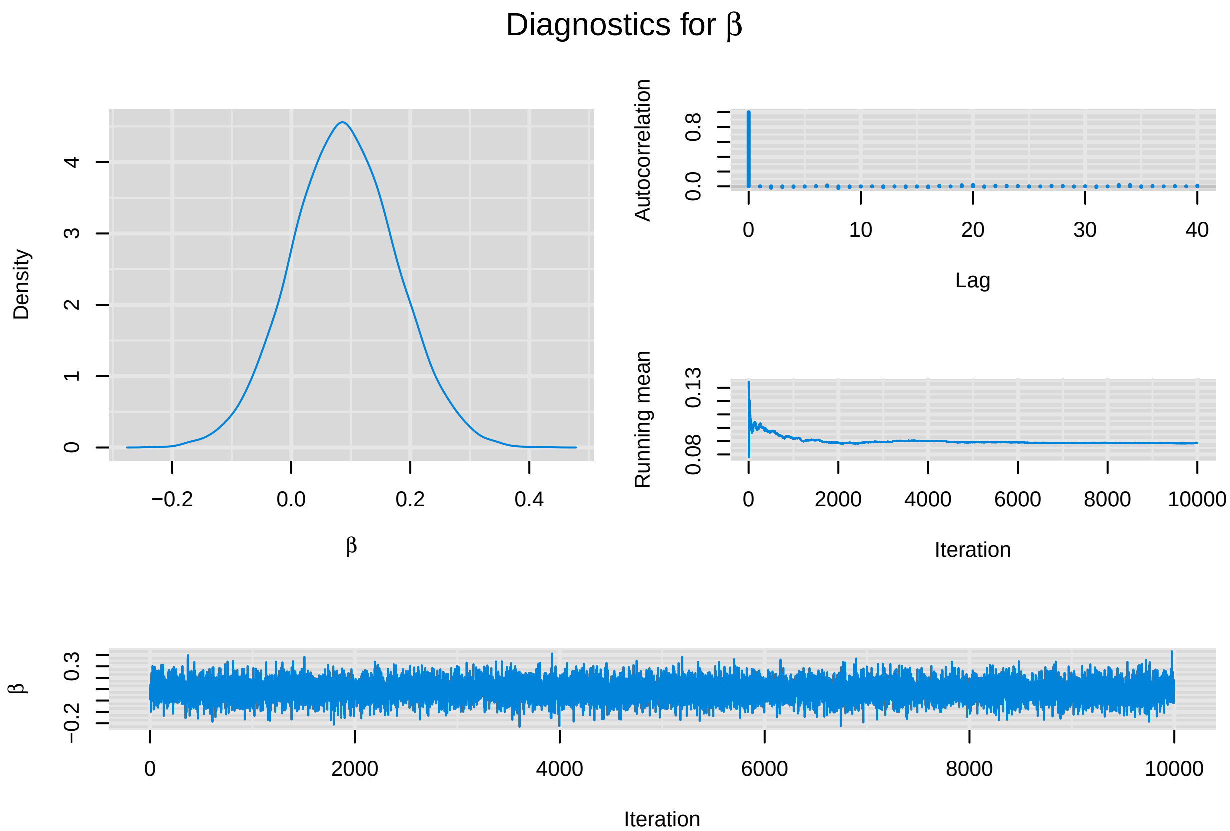

```
autplot1( rmuListj )
```




```r
geweke.diag( rj.beta )
```

```
##
## Fraction in 1st window = 0.1
## Fraction in 2nd window = 0.5
##
## var1
## 1.25
```

```r
mcmcplot1(mat, greek = T)
```



For the trace plot we see no trend and random scatter. The running means plot shows the running mean flattening out as the number of samples increases. And the ACF plot shows independence of our samples, as the vertical lines drop between the blue dashed lines almost immediately. The Geweke Convergence Diagnostic is also showing convergence with a z-score that is within two standard devations from zero.

It appears as if the model is not sensitive to the choice of non-informative priors. The estimates for β and the variance are roughly the same regardless of the prior we choose.

2 Using the daily Capital Bike Share data, `bikeshare.txt`, fit a Poisson regression model first with the casual users count as the outcome and then with the registered users count as the outcome. Using Bayesian inference, determine the best set of predictors for each outcome. For each final model, check the convergence of all model parameters using visual inspection as well as Geweke's diagnostic. Also provide a brief interpretation of each parameter. Do the models differ substantially from each other? Data variable names are provided alongside this assignment, `bikeshareDataDictionary.rtf`. For each model, set the seed to 1959.

```r
bike <- read.table("bikeshare.txt", header=T)
fit <- glm( casual ~ . -registered  , data = bike , family = poisson )
bhat    <- coef(fit)
vbeta   <- vcov(fit)
B <- 10000

set.seed(1959)
Beta <- rmvnorm(B, mean = bhat, sigma = vbeta)

t(apply(Beta, 2, quantile, probs = c(0.5, 0.025, 0.975)))
```

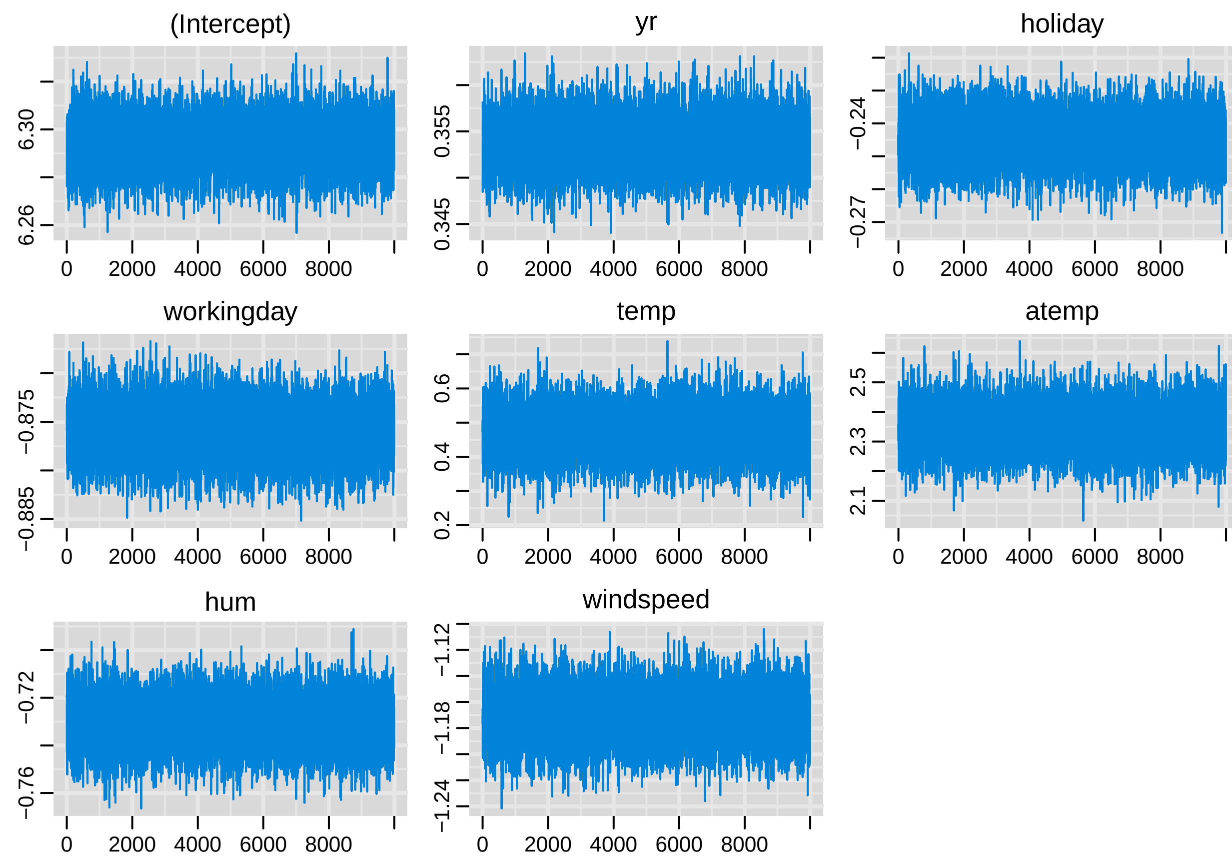
```
##                    50%       2.5%      97.5%
## (Intercept)  6.2936371  6.2744323  6.3130403
## yr           0.3537921  0.3486128  0.3588844
## holiday     -0.2450467 -0.2588736 -0.2311165
## workingday  -0.8760760 -0.8812143 -0.8709988
## temp         0.4704731  0.3453926  0.5987697
## atemp        2.3472809  2.2011027  2.4898816
## hum         -0.7309822 -0.7503327 -0.7113461
## windspeed   -1.1710228 -1.2083146 -1.1336541
```

```r
exp(t(apply(Beta, 2, quantile, probs = c(0.5, 0.025, 0.975))))
```

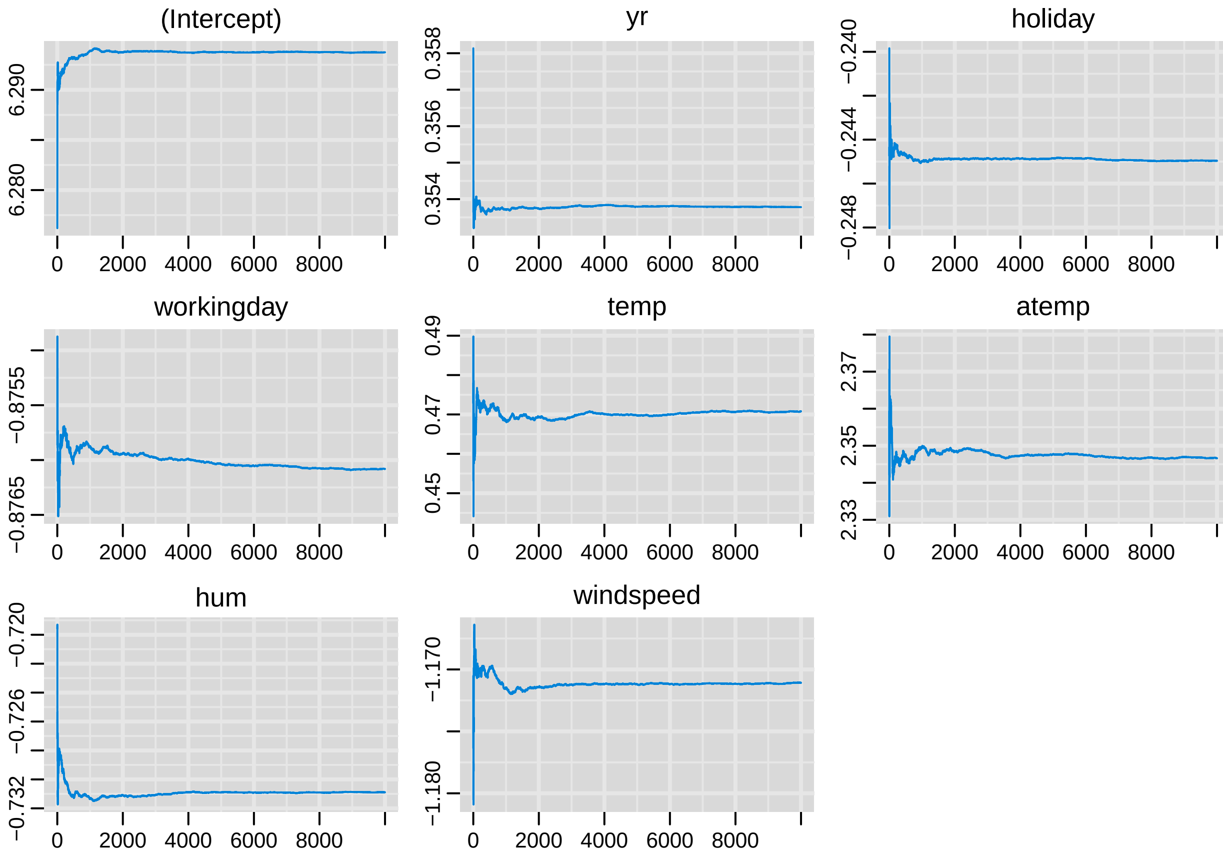
```
##                     50%        2.5%       97.5%
## (Intercept) 541.1178326 530.8249664 551.7198161
## yr            1.4244589   1.4171005   1.4317313
## holiday       0.7826680   0.7719206   0.7936470
## workingday    0.4164137   0.4142795   0.4185333
## temp          1.6007513   1.4125444   1.8198784
## atemp        10.4570969   9.0349712  12.0598483
## hum           0.4814359   0.4722094   0.4909828
## windspeed     0.3100496   0.2987003   0.3218550
```

None of the exponentiated credible intervals contain 0, so we use each variable in the regression.

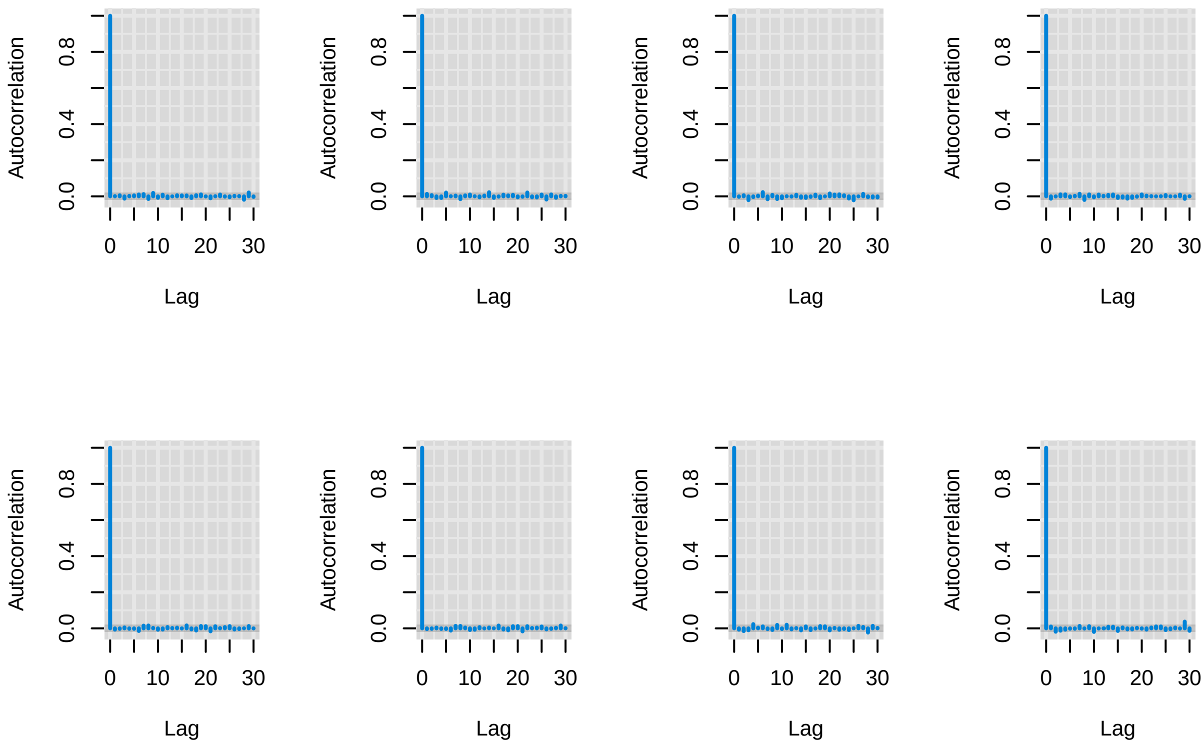
```r
mat <- as.matrix(Beta   )
colnames( Beta ) <-  names( coef(fit) )
rmuList <- mcmc.list(list(mcmc(mat )))
traplot( mat , greek = T )
```




```r
rmeanplot(mat , greek = T)
```




```
par(mfrow = c(2,4)); autplot1( rmuList )
```




```r
geweke.diag( Beta  )
```

```
##
## Fraction in 1st window = 0.1
## Fraction in 2nd window = 0.5
##
## (Intercept)          yr     holiday  workingday        temp       atemp
##      0.4189     -0.4988      0.4859      2.8449     -1.5535      1.5916
##         hum   windspeed
##     -1.3125     -0.7739
```

All the parameters looks as if they have converged, but there is some concern over the convergence of the `workingday` parameter. Its Geweke diagnoistic is greater than 2, and it may not have converged from inspection of the running mean plot. However, it is not showing any drastic signs of not converging.

Interpretation of the coefficients:

`yr`: Holding all others constant, for an additional one unit in years, the expected rides for casual users increases by about 42%.

`holiday`: Holding all others constant, for holidays, the expected rides for casual users decreases about 22%.

`workingday`: Holding all others constant, for working days, the expected rides for casual users decreases about 58%.

`temp`: Holding all others constant, for one additional unit of temperature, the expected rides for casual users decreases about 60%.

`atemp`: Holding all others constant, for one additional unit of "feeling" temperature, the expected rides for casual users increases about 945%.

`hum`: Holding all others constant, for one additional unit of humidity, the expected rides for casual users decreases about 52%.

`windspeed`: Holding all others constant, for one additional unit of windspeed, the expected rides for casual users decreases about 69%.

```r
bike <- read.table("bikeshare.txt", header=T)
fit <- glm( registered ~ . -casual  , data = bike , family = poisson )
bhat    <- coef(fit)
vbeta   <- vcov(fit)
B <- 10000

set.seed(1959)
Beta <- rmvnorm(B, mean = bhat, sigma = vbeta)

t(apply(Beta, 2, quantile, probs = c(0.5, 0.025, 0.975)))
```

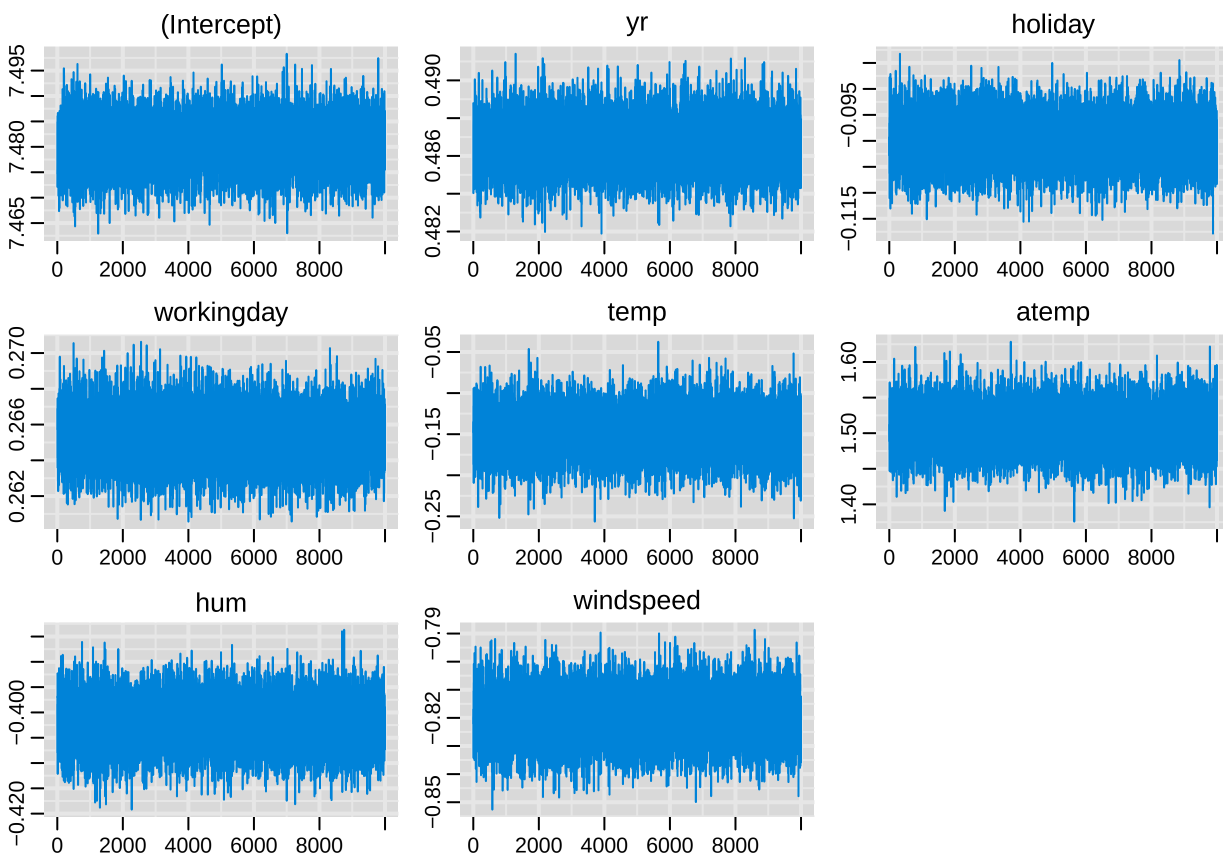
```
##                    50%       2.5%       97.5%
## (Intercept)  7.4802873  7.4711965  7.48951004
## yr           0.4866650  0.4841393  0.48916502
## holiday     -0.1000658 -0.1088724 -0.09122383
## workingday   0.2652855  0.2624238  0.26810864
## temp        -0.1493753 -0.2013942 -0.09608735
## atemp        1.5069657  1.4463507  1.56629925
## hum         -0.4023300 -0.4115116 -0.39303909
## windspeed   -0.8198421 -0.8371358 -0.80255936
```

```r
exp(t(apply(Beta, 2, quantile, probs = c(0.5, 0.025, 0.975))))
```

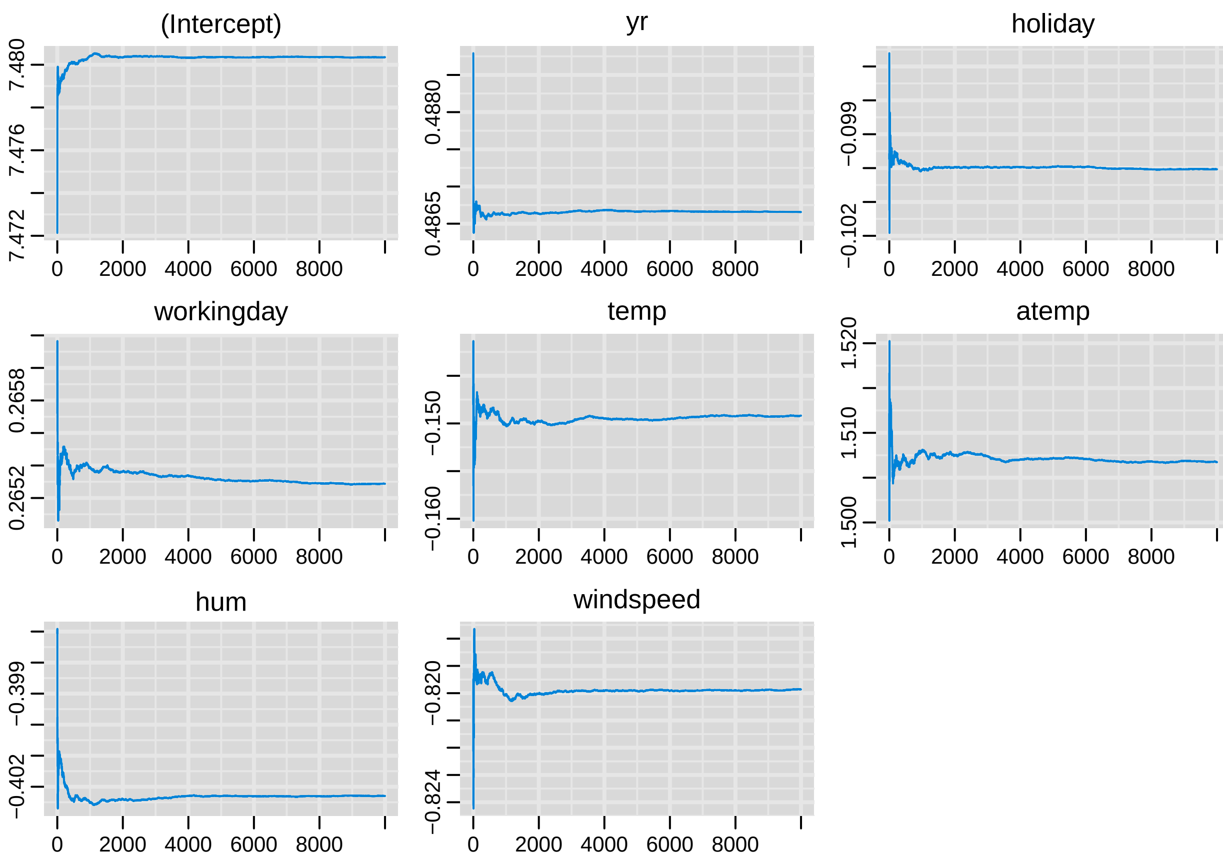
```
##                      50%         2.5%        97.5%
## (Intercept) 1772.7499452 1756.7072499 1789.1752442
## yr             1.6268816    1.6227777    1.6309538
## holiday        0.9047779    0.8968449    0.9128134
## workingday     1.3038032    1.3000774    1.3074892
## temp           0.8612459    0.8175901    0.9083847
## atemp          4.5130162    4.2475854    4.7888929
## hum            0.6687600    0.6626478    0.6750024
## windspeed      0.4405012    0.4329488    0.4481804
```

None of the exponentiated credible intervals contain 0, so we use each variable in the regression.

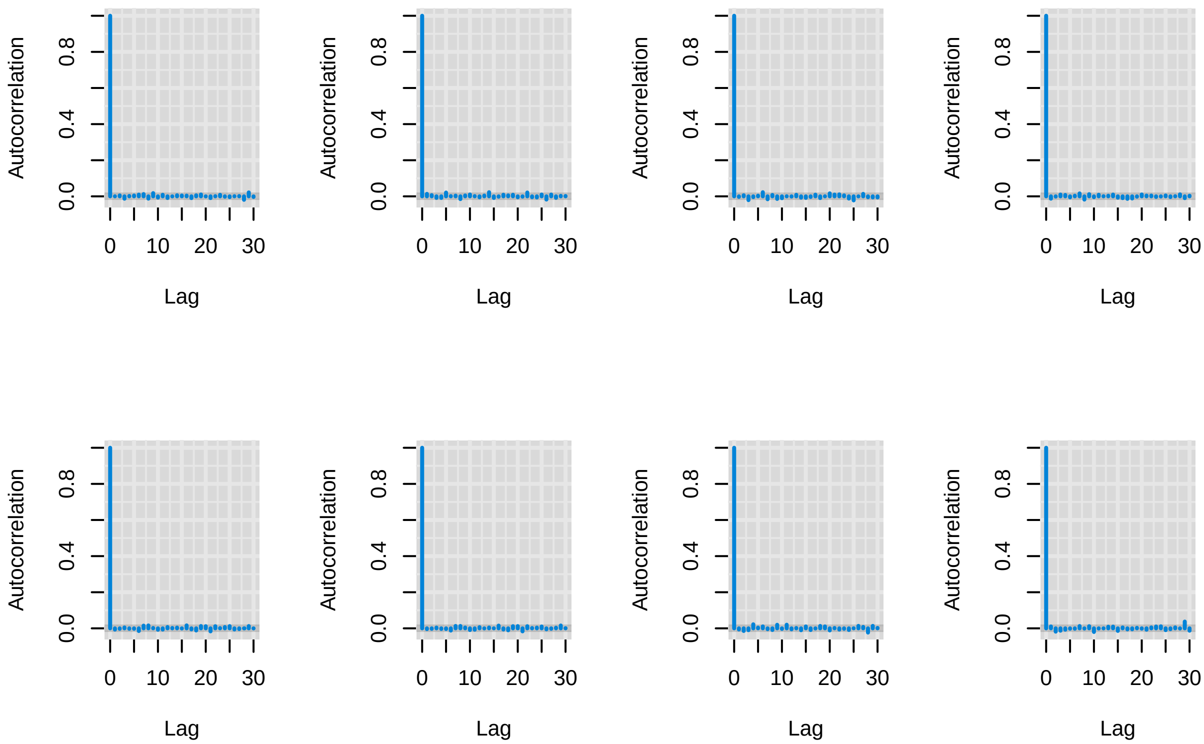
```r
mat <- as.matrix(Beta   )
colnames( Beta ) <-  names( coef(fit) )
rmuList <- mcmc.list(list(mcmc(mat )))
traplot( mat , greek = T )
```




```r
rmeanplot(mat , greek = T)
```




```
par(mfrow = c(2,4)); autplot1( rmuList )
```




```r
geweke.diag( Beta  )
```

```
##
## Fraction in 1st window = 0.1
## Fraction in 2nd window = 0.5
##
## (Intercept)          yr     holiday  workingday        temp       atemp
##      0.3536     -0.5846      0.5749      2.5350     -1.5389      1.5868
##         hum   windspeed
##     -1.3354     -0.7584
```

All the parameters looks as if they have converged, but there is some concern over the convergence of the `workingday` parameter. Its Geweke diagnoistic is greater than 2, and it may not have converged from inspection of the running mean plot. However, it is not showing any drastic signs of not converging.

Interpretation of the coefficients:

`yr`: Holding all others constant, for an additional one unit in years, the expected rides for registered users increases by about 62%.

`holiday`: Holding all others constant, for holidays, the expected rides for registered users decreases about 10%.

`workingday`: Holding all others constant, for working days, the expected rides for registered users increases about 30%.

`temp`: Holding all others constant, for one additional unit of temperature, the expected rides for registered users decreases about 14%.

`atemp`: Holding all others constant, for one additional unit of "feeling" temperature, the expected rides for registered users increases about 351%.

`hum`: Holding all others constant, for one additional unit of humidity, the expected rides for registered users decreases about 33%.

`windspeed`: Holding all others constant, for one additional unit of windspeed, the expected rides for registered users decreases about 56%.

One big difference between the models is `workingday` variable. For registered users it signals an increase in rides and for casual users it signals a decrease. This makes sense because many registered uses would probably use bikeshare for commuting to work. Also, the temperature variables do not have as big of an effect on the registered users. Again, this is because many of the registered users may use the bikes to commute to work.