Computer Lab 4

Andrea Bruzzone, Thomas Zhang

2016 M05 18

Assignment 1

a,

We do a `glm()` fit and obtain the maximum liklihood estimator of β in the Poisson regression model for the eBay data.

```
## [1] "The maximum liklihood estimator of beta coefficients:"
```

```
## (Intercept) PowerSeller    VerifyID      Sealed     Minblem     MajBlem
##  1.07244206 -0.02054076 -0.39451647  0.44384257 -0.05219829 -0.22087119
##     LargNeg     LogBook MinBidShare
##  0.07067246 -0.12067761 -1.89409664
```

Through the `summary.glm()` function, we can say that it looks like VerifyID, Sealed, MajBlem, LogBook and MinBidShare are significant predictors in this model.

Next, we do a Bayesian analysis of the Poisson regression with prior distribution $\beta \sim \mathcal{N}(\mathbf{0}, 100 \cdot (X'X)^{-1})$. We know that we can use the `optim()` function to numerically fund the posterior mode $\tilde{\beta}$ and the Hessian $-J_{y,\tilde{\beta}}$ at that posterior mode. With these values we can approximate the posterior distribution as a multivariate normal distribution, $\beta|y \sim \mathcal{N}(\tilde{\beta}, J^{-1}_{y,\tilde{\beta}})$.

```
## [1] "The posterior mode beta coefficients: "
```

```
## [1]  1.06984118 -0.02051246 -0.39300599  0.44355549 -0.05246627 -0.22123840
## [7]  0.07069683 -0.12021767 -1.89198501
```

```
## [1] "The hessian at the posterior mode: "
```

```
##             [,1]        [,2]          [,3]          [,4]        [,5]
##  [1,] -3634.2841 -1574.88862 -1.284330e+02 -5.054825e+02 -308.957275
##  [2,] -1574.8886 -1574.88862 -6.049186e+01 -3.260764e+02 -104.461478
##  [3,]  -128.4330   -60.49186 -1.284330e+02 -6.148277e+01   -9.865298
##  [4,]  -505.4825  -326.07643 -6.148277e+01 -5.054825e+02    0.000000
##  [5,]  -308.9573  -104.46148 -9.865298e+00  0.000000e+00 -308.957275
##  [6,]  -126.9303   -68.96966  0.000000e+00  5.684342e-08    0.000000
##  [7,]  -385.7170   -53.05278  5.684342e-08  0.000000e+00  -33.905658
##  [8,]  -638.9730    71.79074 -6.887776e+01 -1.287304e+02  -22.297055
##  [9,]   729.8896   146.21556  2.380017e+01  8.975677e+01   55.182230
##                [,6]          [,7]        [,8]       [,9]
##  [1,] -1.269303e+02 -3.857170e+02  -638.97297  729.88955
##  [2,] -6.896966e+01 -5.305278e+01    71.79074  146.21556
##  [3,]  0.000000e+00  5.684342e-08   -68.87776   23.80017
##  [4,]  5.684342e-08  0.000000e+00  -128.73043   89.75677
##  [5,]  0.000000e+00 -3.390566e+01   -22.29705   55.18223
```

```
##  [6,] -1.269303e+02  0.000000e+00   -36.39914   34.16904
##  [7,]  0.000000e+00 -3.857170e+02  -220.23559  115.38523
##  [8,] -3.639914e+01 -2.202356e+02 -1930.07936  534.16381
##  [9,]  3.416904e+01  1.153852e+02   534.16381 -446.88731
```

```
## [1] "A posterior draw of beta: "
```

```
##          [,1]       [,2]       [,3]      [,4]       [,5]       [,6]
## [1,] 1.027218 0.01174206 -0.3210896 0.4104055 0.03598681 -0.1783029
##           [,7]        [,8]      [,9]
## [1,] 0.1131857 -0.08478953 -1.924996
```

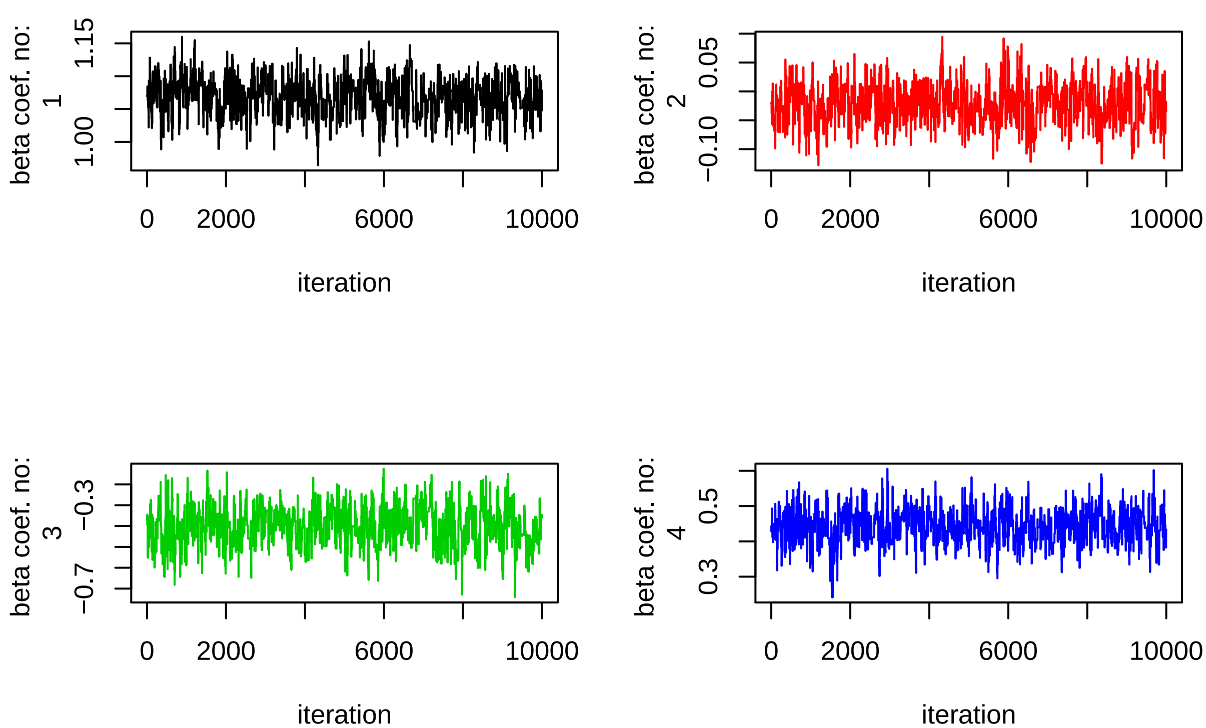
Now we simulate from the actual posterior of β using the random walk Metropolis-Hanstings algorithm. We are going to use a multivariate normal density, $\theta_p|\theta_c \sim \mathcal{N}(\theta_c, \tilde{c} \cdot \Sigma)$ as proposal density where $\Sigma = J^{-1}_{y,\tilde{\beta}}$ and $\tilde{c}$ is equal to 2.4 divided by the squre root of the number of parameters.

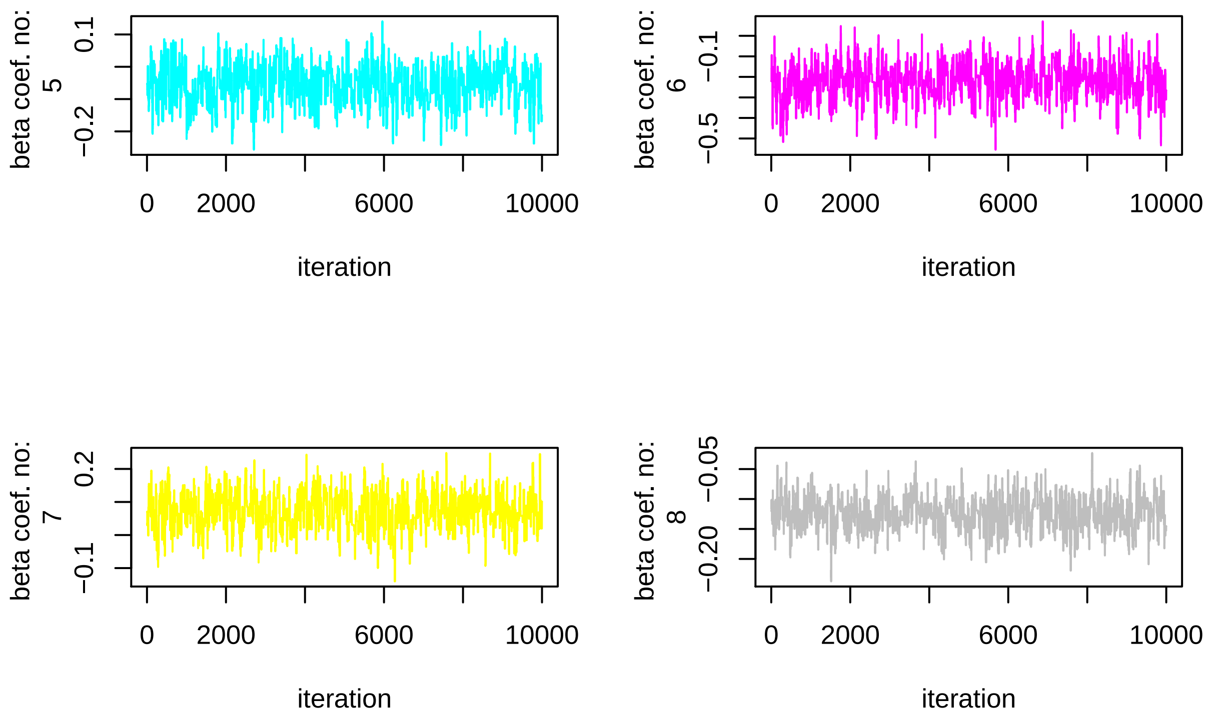
```
## [1] "last iteration of M-H algorithm: "
```

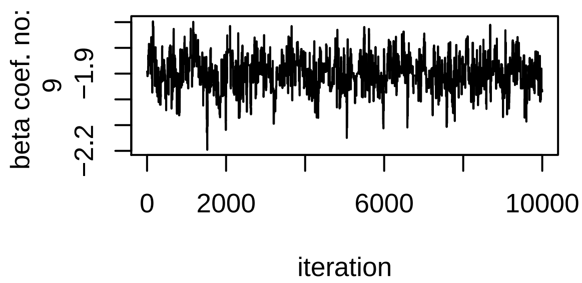
```
## [1]  1.04774337 -0.03437716 -0.35998463  0.43334543 -0.14985697 -0.26539786
## [7]  0.10138422 -0.14528573 -1.96953026
```

As it can be seen the draw from the M-H algorithm is very similar to the one found before.

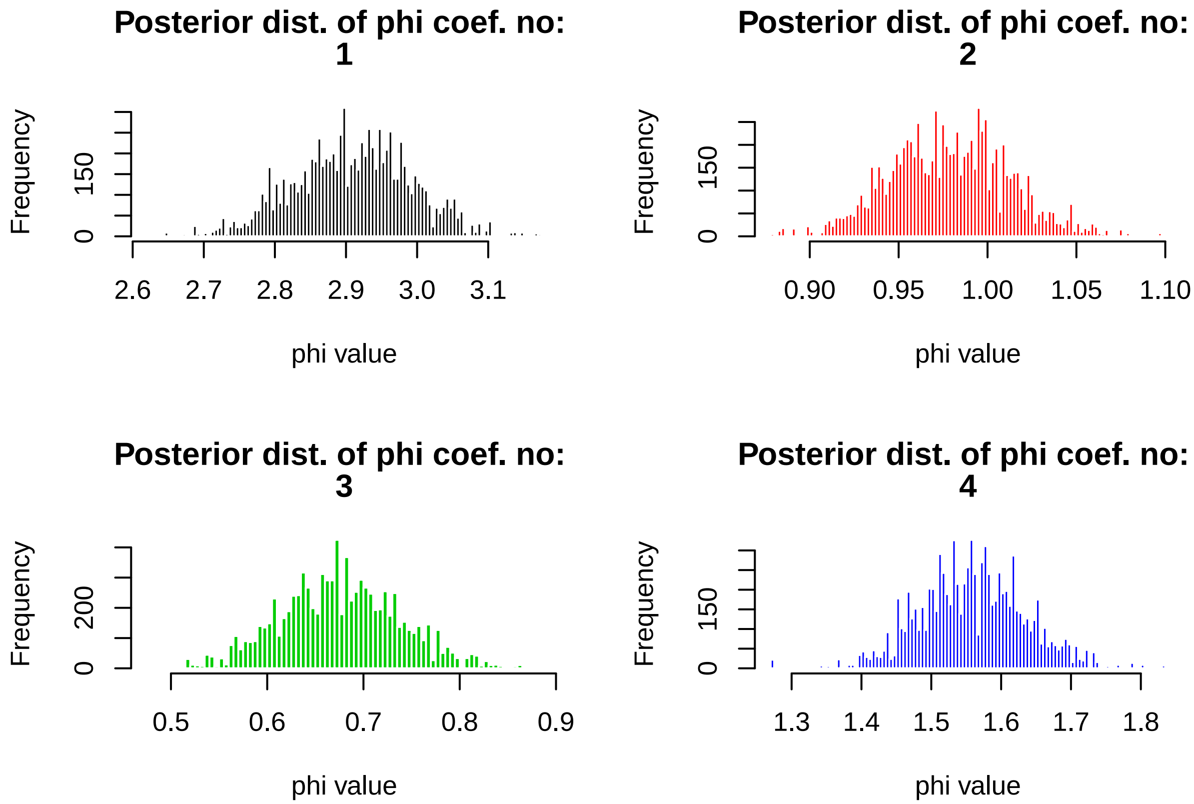
We plot the traceplots for each beta coefficient (where beta coefficient number one is the intercept) and then we plot the histograms for the posterior distributions of $\phi_j = \exp \beta_j$.

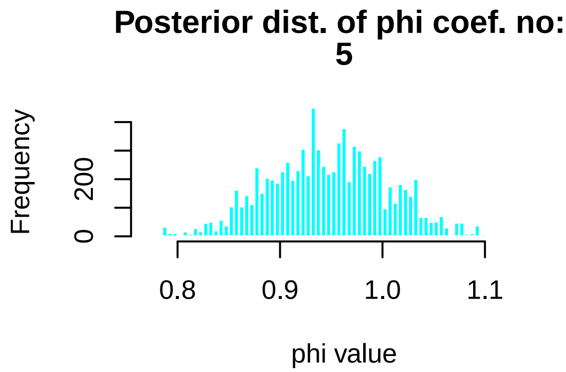


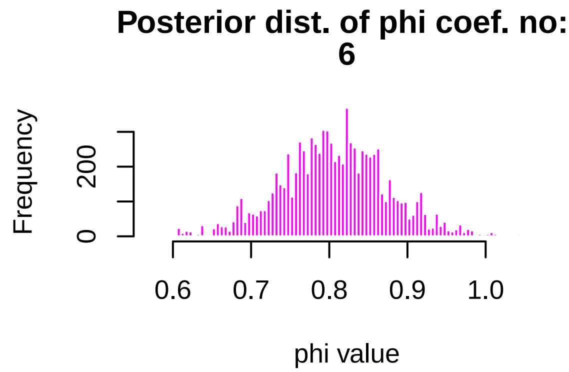

beta coef. no: 5
0.1
−0.2
0
2000
6000
10000
iteration
beta coef. no: 6
−0.1
−0.5
0
2000
6000
10000
iteration
beta coef. no: 7
0.2
−0.1
0
2000
6000
10000
iteration
beta coef. no: 8
−0.05
−0.20
0
2000
6000
10000
iteration

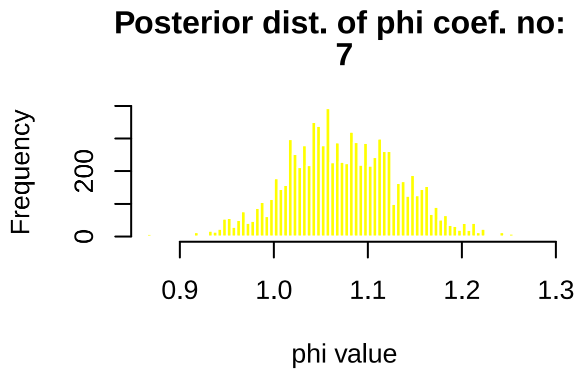


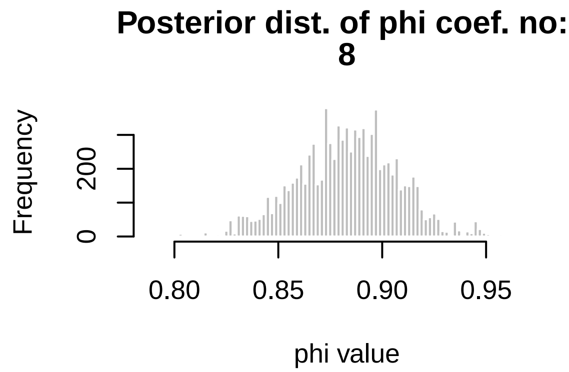

```
## [1] "Effective sizes of the MCMC chains, 10000 iterations: "

##      var1     var2     var3     var4     var5     var6     var7     var8
## 359.1155 348.7120 292.5276 328.3734 323.2014 344.4381 299.0249 342.5584
##      var9
## 343.3265
```

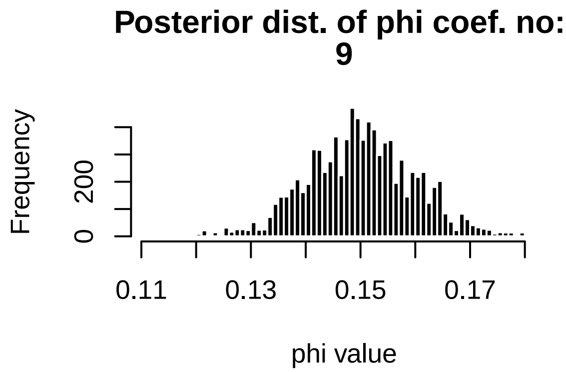

Posterior dist. of phi coef. no: 1
Frequency
0
150
2.6
2.7
2.8
2.9
3.0
3.1
phi value
Posterior dist. of phi coef. no: 2
Frequency
0
150
0.90
0.95
1.00
1.05
1.10
phi value
Posterior dist. of phi coef. no: 3
Frequency
0
200
0.5
0.6
0.7
0.8
0.9
phi value
Posterior dist. of phi coef. no: 4
Frequency
0
150
1.3
1.4
1.5
1.6
1.7
1.8
phi value


Posterior dist. of phi coef. no: 5
Frequency
0
200
0.8
0.9
1.0
1.1
phi value


Posterior dist. of phi coef. no: 6
Frequency
0
200
0.6
0.7
0.8
0.9
1.0
phi value

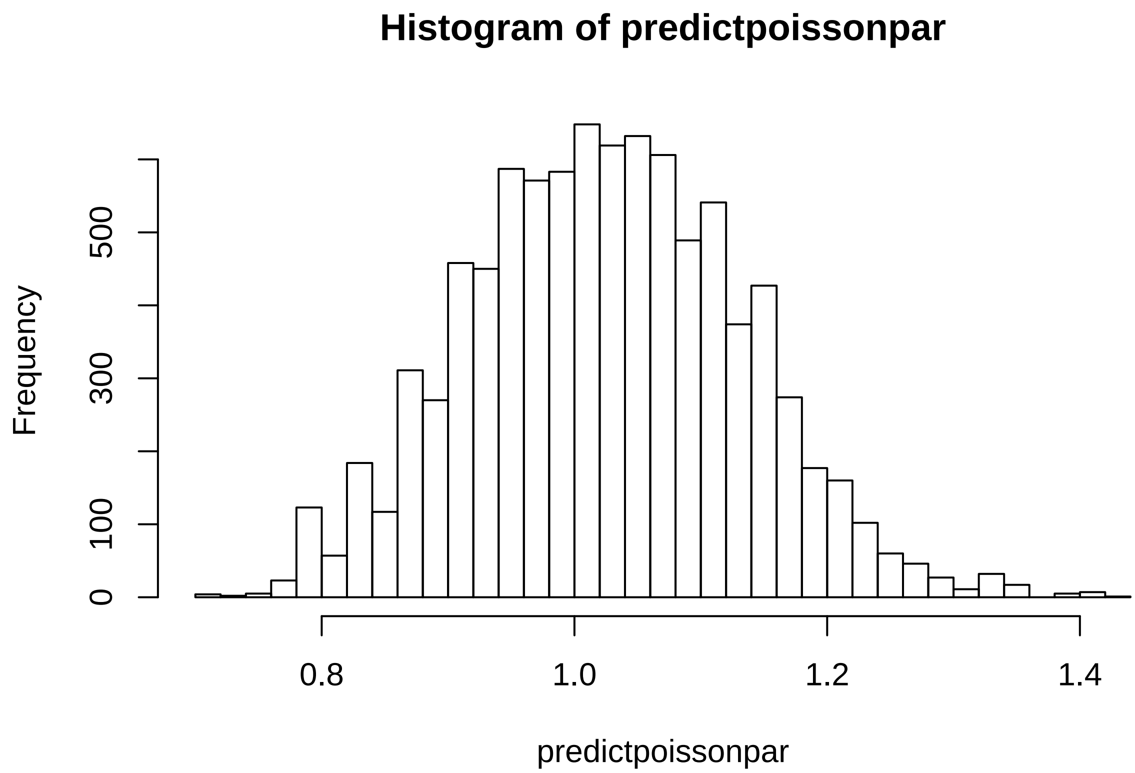

Posterior dist. of phi coef. no: 7
Frequency
0
200
0.9
1.0
1.1
1.2
1.3
phi value


Posterior dist. of phi coef. no: 8
Frequency
0
200
0.80
0.85
0.90
0.95
phi value



From the traceplots we can see that all the β values seem to converge. Anyways the effective samples size is only about 3.5%.

Finally we plot a histogram over the predictive distribution of poisson parameter λ for the auction given in the lab instructions and calculate the probability that that auction will have zero bids.




```
## [1] "Mean probability for no bids on the new auction: 0.361"
```

We can tell that the coin object and the seller is of good quality, but that the MinBidShare and LogBook values were rather higher than average. These factors balance each other out and we believe that the result is reasonable.