hw7

Autumn Li UNI:ql2280

11/12/2016

```
moretti <- read.csv("~/Desktop/moretti.csv")
head(moretti)
```

```
##          Name Begin  End
## 1   Courtship  1740 1820
## 2 Picareseque  1748 1790
## 3    Oriental  1759 1787
## 4  Epistolary  1766 1795
## 5 Sentimental  1768 1790
## 6         Spy  1770 1800
```

Part 1

i. Write a function poisLoglik

```
poisLoglik =  function(lamda = 1 , data = c(1,0,0,1,1)){
  result = sum(log(lamda^(data)*exp(-lamda)/factorial(data)))
  return(result)
}
poisLoglik(lamda = 1 , data = c(1,0,0,1,1))
```

```
## [1] -5
```

ii. Write a function count new genres which takes in a year, and returns the number of new genres which appeared in that year

```r
moretti <- read.csv("~/Desktop/moretti.csv")
begin = moretti$Begin
end = moretti$End
begin
```

```
##  [1] 1740 1748 1759 1766 1768 1770 1773 1789 1790 1791 1800 1804 180
8 1814
## [15] 1822 1825 1826 1828 1830 1830 1832 1838 1839 1846 1846 1847 184
8 1849
## [29] 1850 1850 1850 1851 1857 1868 1871 1872 1872 1873 1876 1884 188
5 1885
## [43] 1888 1888
```

```r
length(begin[begin==1830])
```

```
## [1] 2
```

```r
count_new_genres = function(year = 1850){
  if(length(begin[begin==year]) == 0){
      result = 0
  }

  if(length(begin[begin == year]) == 1){
    result = 1
  }
  if(length(begin[begin == year]) == 2){
    result = 2
  }
  if(length(begin[begin == year]) == 3){
    result = 3
  }
  if(length(begin[begin == year]) == 4){
    result = 4
```

```
  }
  if(length(begin[begin == year]) == 5){
    result = 5
  }
  return (result)
}
count_new_genres (year = 1850)

## [1] 3

count_new_genres (year = 1803)

## [1] 0
```

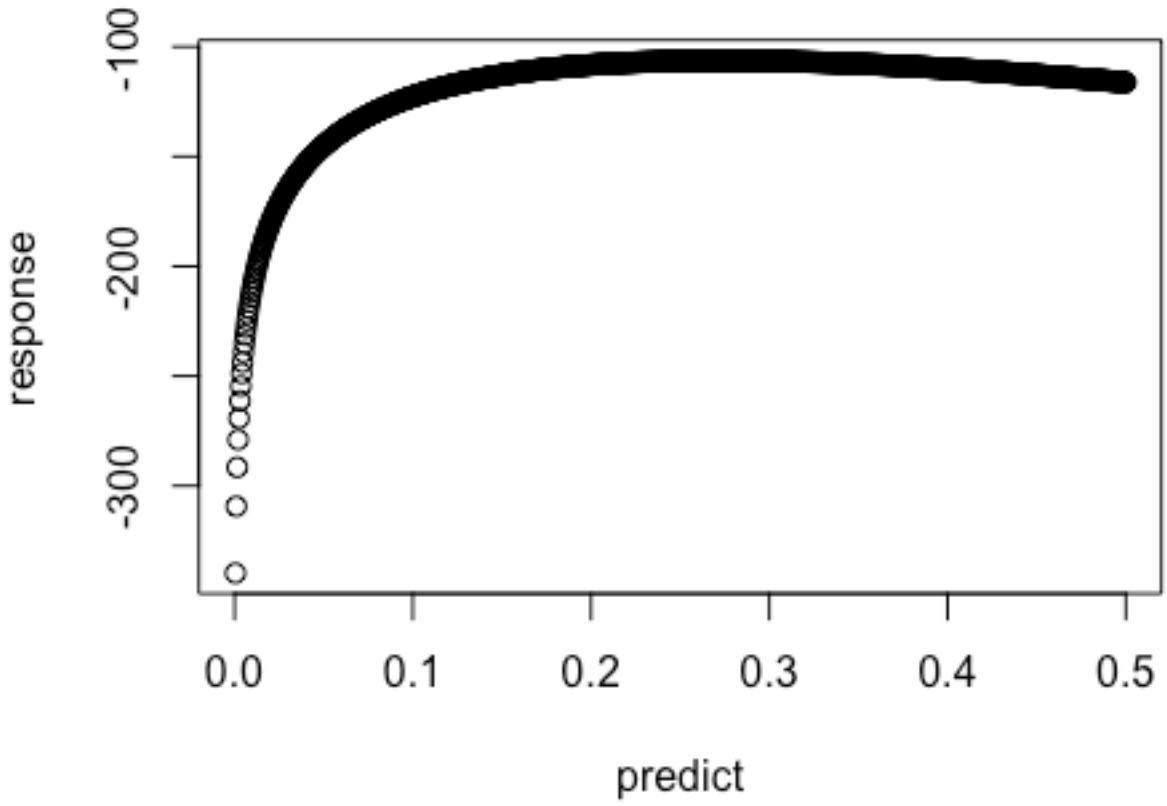
iii. Create a vector, new genres, which counts the number of new genres which appeared in each year of the data, from 1740 to 1900.

```
new_genres = NULL
 i = 1740
while(i < 1901) {
   new_genres[i-1739] = count_new_genres (year = i)
   i = i+ 1
}
new_genres

##   [1] 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0 1
0 0 1 0
##  [36] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 1 0 0 0 1 0
0 0 1 0
##  [71] 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 1 1 0 1 0 2 0 1 0 0 0 0 0 1 1 0
0 0 0 0
## [106] 0 2 1 1 1 3 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 1 2 1 0 0
1 0 0 0
## [141] 0 0 0 0 1 2 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0
```

iv. Plot poisLoglik as a function of λ on the new genres data.

```
par(mfrow=c(1,1))
predict = seq(0,0.5,length = 1000)
response = NULL
for (i in 1:1000){
  response[i] = poisLoglik(lamda = predict[i] , data = new_genres)
}
plot(predict,response)
```




```
max_response = max(response)
max_lamda = predict[which(response==max_response)];max_lamda
```

```
## [1] 0.2732733
```

v. Use nlm() to maximize the log likelihood to check the λ = 0.273 value suggested in the previous question.

```
poisLoglik =  function(lamda  , data ){
  result = sum(log(lamda^(data)*exp(-lamda)/factorial(data)))
  return(-result)
}
nlm(poisLoglik,p = 0.27,data = new_genres)$estimate

## [1] 0.2732919
```

vi. To investigate whether genres appear in bunches or randomly, we look at the spacing between genre births

```
begin1 = begin
begin= begin[-44]
begin2 = begin1[-1]
intergenre_intervals = begin2-begin
mean(intergenre_intervals)

## [1] 3.44186

sd(intergenre_intervals)

## [1] 3.705224

coefficient_of_variation = sd(intergenre_intervals)/mean(intergenre_int
ervals);coefficient_of_variation

## [1] 1.076518

diff(intergenre_intervals)

##  [1]   3  -4  -5   0   1  13 -15   0   8  -5   0   2   2  -5  -2   1
   0
```

```
## [18]  -2   2   4  -5   6  -7   1   0   0   0  -1   0   1   5   5  -8
  -2
## [35]  -1   1   2   5  -7  -1   3  -3
```

vii. For a Poisson process, the coefficient of variation is expected to be around 1

```
years = NULL
a = 1
b = 1
length_genres = length(new_genres)
genres_appear = function(new_genres){
  for (i in 2:length_genres){
    if (new_genres[i] == 0){
      a = a+1
    }
    if (new_genres[i] == 1){
      years[b] = a
      a = 1
      b = b + 1
    }
    if (new_genres[i] == 2){
      years[b] = a
      b = b+1
      years[b] = 0
      a = 1
      b = b+1
    }
    if (new_genres[i] == 3){
      years[b] = a
      b = b + 1
      years[b] = 0
```

```
      b = b+1
      years[b] = 0
      a = 1
      b = b+1
      }
  }

  return(years)
}
genres_appear(new_genres)
```

```
##  [1]  8 11  7  2  2  3 16  1  1  9  4  4  6  8  3  1  2  2  0  2  6
 1  7
## [24]  0  1  1  1  1  0  0  1  6 11  3  1  0  1  3  8  1  0  3  0
```

(b)

```
poisson = function (year_num, genres_mean){
  data = rpois(year_num, genres_mean)
  ia_interval = genres_appear(data)
  coef = sd(ia_interval)/mean(ia_interval)
  result = list(inter_appearance_interval = ia_interval,coefficients =
coef,mean_value = mean(ia_interval))
  return(result)
}
poisson(161,0.273)
```

```
## $inter_appearance_interval
##  [1]  4  6  5  8  2  2 19  6  4  0  5  6  4  0  1  2  2  0  6  2  1
 9  5
## [24]  5  6  0  1  0  9  2  0  2  6  2  4  1  4  1  7  0  1  2  1  3
##
## $coefficients
```

```
## [1] 0.9910691
##
## $mean_value
## [1] 3.545455
```

viii. Run your simulation 100,000 times, taking the coefficient of variation (only) from each.

```
repeat_1000 = matrix(NA,ncol= 1000,nrow = 1)
high_coef = 0
for (i in 1:1000){
  repeat_1000[i] = poisson(161,0.273)$coefficient
  if(repeat_1000[i] > coefficient_of_variation){
    high_coef = high_coef + 1
  }
}
fraction = high_coef/1000;fraction

## [1] 0.216
```

ix. Explain what this does and does not tell you about the conjecture that genres tend to appear together in burst?

```
#Nearly 22 percent of variation than morett's data, which suggests that
 the conjecture that genres tend to appear together in burst
```

```
```