Homework 7

Hanao Li hl3202

November 21, 2018

i.

```
poisLoglik <- function(lambda, x){
  return(sum(x*log(lambda) - lambda - log(factorial(x))))
}
data <- c(1, 0, 0, 1, 1)
lambda <- 1
poisLoglik(lambda, data)
```

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

ii.

```
data <- read.csv("moretti.csv", header = TRUE)
count_new_genres <- function(x){
  return(sum(data$Begin == x))
}

count_new_genres(1803)
```

```
## [1] 0
```

```
count_new_genres(1850)
```

```
## [1] 3
```

iii.

```
new_genres <- c()
n <- 0
for (i in 1740:1900){
  n <- n + 1
  new_genres[n] <- count_new_genres(i)
}
# They should be at poistion 1803 - 1740 + 1 = 64 and 1850 - 1740 + 1 = 111 and their values should be
new_genres[1803 - 1740 + 1]
```

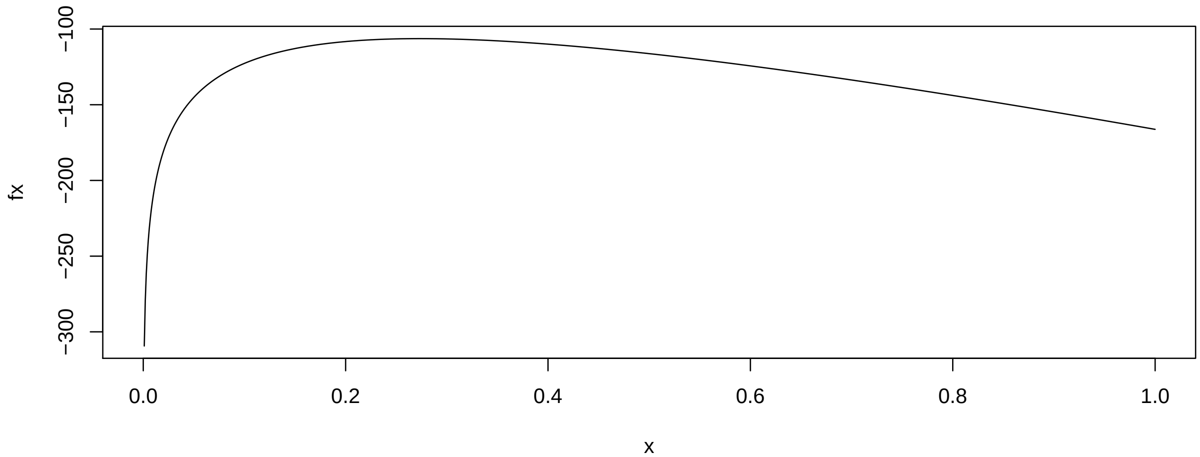
```
## [1] 0
```

```
new_genres[1850 - 1740 + 1]
```

```
## [1] 3
```

iv.

```
x <- seq(0, 1, 0.001)
fx <- c()
for (i in 1:length(x)){
  fx[i] <- poisLoglik(x[i], new_genres)
}
plot(x, fx, type = 'l')
```




```
# From the plot, we could see that the maximum is at around 0.273
```

v.

```
newpoisLoglik <- function(x){
  -(sum(new_genres*log(x)-x-log(factorial(new_genres))))
}
suppressWarnings(nlm(newpoisLoglik, 0.001))
```

```
## $minimum
## [1] 106.3349
##
## $estimate
## [1] 0.2732919
##
## $gradient
## [1] 4.334311e-09
##
## $code
## [1] 1
##
## $iterations
## [1] 12
```

vi.

```
intergenre_intervals <- diff(data$Begin)
mean(intergenre_intervals)
```

```
## [1] 3.44186
```

```
sd(intergenre_intervals)
```

```
## [1] 3.705224
```

```
Mcov <- sd(intergenre_intervals) / mean(intergenre_intervals)
Mcov
```

```
## [1] 1.076518
```

vii.

a.

```r
interval <- function(n){
  count <- 0
  intervals <- c()
  for (i in 2:length(n)){
    if (n[i] >= 2){
      intervals <- c(intervals, count + 1)
      for (j in 1:(n[i] - 1)){
        intervals <- c(intervals, 0)
      }
      count <- 0
      next
    }
    if (n[i] != 0 & n[i - 1] != 0){
      intervals <- c(intervals, 1)
      count <- 0
      next
    }
    else if (n[i] == 0){
      count = count + 1
    }
    else{
      intervals <- c(intervals, count + 1)
      count <- 0
    }
  }
  return(intervals)
}

interval(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  0  1
## [26]  1  1  1  0  0  1  6 11  3  1  0  1  3  8  1  0  3  0
```

```r
intergenre_intervals
```

```
##  [1]  8 11  7  2  2  3 16  1  1  9  4  4  6  8  3  1  2  2  0  2  6  1  7  0  1
## [26]  1  1  1  0  0  1  6 11  3  1  0  1  3  8  1  0  3  0
```

b.

```r
poisson.sim <- function(years, mean){
  r <- rpois(years, mean)
  i <- interval(r)
  cov <- sd(i) / mean(i)

  output <- list(i, cov)
  names(output) <- c("Inter-appearance Interval", "Coefficient of Variation")

  return(output)
}
poisson.sim(161, 0.273)
```

```
## $`Inter-appearance Interval`
```

```
##  [1]  5  0  4  1  2  1  3  0  3  1  0  0  1 10  1  1  2  1  0  7  9 10  1 11  0
## [26]  6  2  0  5  8  9  1  2  2 14  2  6  0  3  0  3  2  1  4  5  0  2  5
##
## $`Coefficient of Variation`
## [1] 1.068708
```

```
mean(poisson.sim(161, 0.273)[[1]])
```

```
## [1] 3.627907
```

viii.

```
covar <- c()
for (i in 1:10000){
  covar[i] <- poisson.sim(161, 0.273)[[2]]
}
summary(covar)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.5201  0.8849  0.9672  0.9812  1.0614  1.7616
```

```
p <- (sum(covar > Mcov) / 10000) * 100
cat("There are ", p,"% of simulation runs have higher coefficient of variation than Moretti's data")
```

```
## There are  22.25 % of simulation runs have higher coefficient of variation than Moretti's data
```

ix.

This result tells us that Moretti's data generally has a higher coefficient of variation than the simulation runs. The coefficient of variation is equal to the standard deviation divided by the mean. Fixing the mean of interval since we assume there will be the same amount of genres appear during the given period, we conclude that the Moretti's data have a higher variance. This leads to the conclusion that genres tend to appear together in burst.