5206hw7_br2498

Bo Rong br2498

November 12, 2016

```
#i.
poisLoglik<-function(lambda,data) {
  result<-c()
  for(i in 1:length(data))
    result[i]<-log((lambda^data[i])*exp(-lambda)/factorial(data[i]))
  return(sum(result))
  }
poisLoglik(lambda=1,data=c(1,0,0,1,1))
```

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

```
#ii.
filename <- "~/Downloads/moretti.csv"
moretti<- read.csv(file=filename, header=T)
count_new_genres<-function(x){
  number<-length(which(moretti$Begin == x))
return(number)
}
count_new_genres(1803)
```

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

```
count_new_genres(1850)
```

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

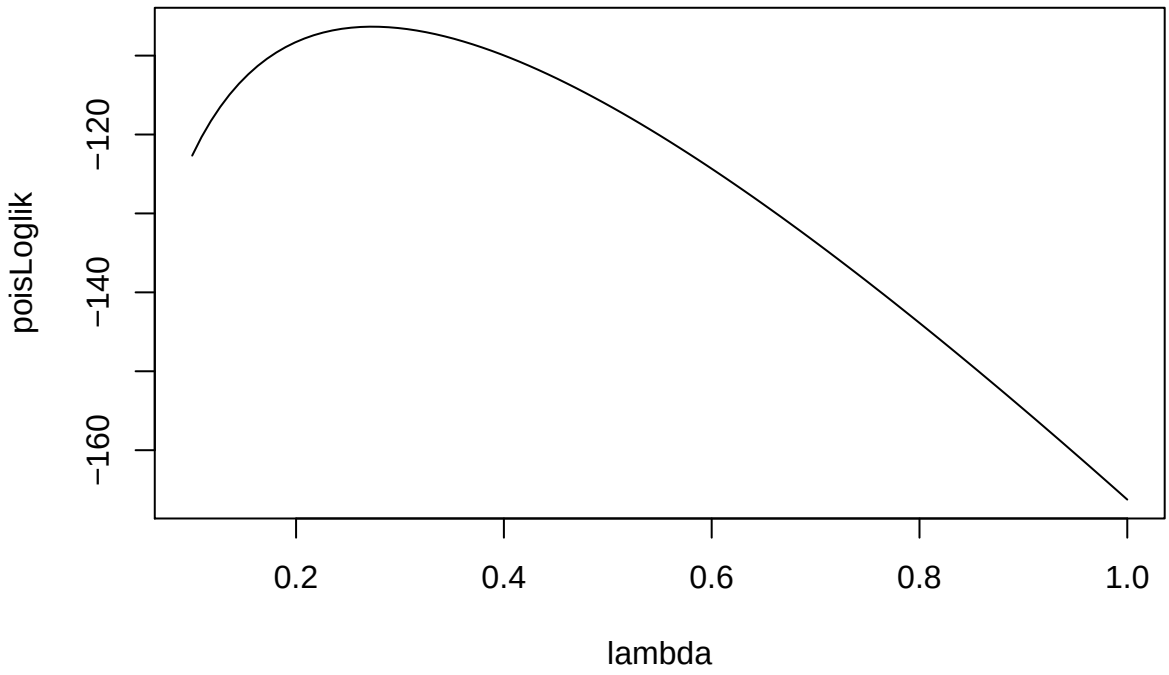
```
#iii.
new_genres<-sapply(unique(1740:1900),count_new_genres)
newdata<-data.frame("Year"=1740:1900,"countnum"=new_genres)
which(newdata$Year==1803)
```

```
## [1] 64
```

```
which(newdata$Year==1850)
```

```
## [1] 111
```

```
#iv.
poisLoglik<-function(lambda,data=new_genres) {
  result<-c()
  for(i in 1:length(data))
    result[i]<-log((lambda^data[i])*exp(-lambda)/factorial(data[i]))
  return(sum(result))
  }
h <- Vectorize(poisLoglik); curve(h, 0.1, 1,xlab = "lambda",ylab = "poisLoglik")
```




```
optimize(poisLoglik, interval=c(0.1, 1), maximum=TRUE)
```

```
## $maximum
## [1] 0.2733037
##
## $objective
## [1] -106.3349
```

```
#v.
poisLoglik_max<-function(lambda,data=new_genres) {
  result<-c()
  for(i in 1:length(data))
    result[i]<--log((lambda^data[i])*exp(-lambda)/factorial(data[i]))
  return(sum(result))
  }
nlm(poisLoglik_max,c(1))$estimate
```

```
## [1] 0.2732914
```

```
#vi.
intergenre_intervals<-NULL
for(i in 1:length(moretti$Begin)-1)
  intergenre_intervals[i]<-moretti$Begin[i+1]-moretti$Begin[i]
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
## [24]  0  1  1  1  1  0  0  1  6 11  3  1  0  1  3  8  1  0  3  0
```

```r
mean(intergenre_intervals)
```

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

```r
sd(intergenre_intervals)
```

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

```r
sd(intergenre_intervals)/mean(intergenre_intervals)
```

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

```r
#vii.
#(a)
check_interval <- function(x){
 n <- length(x)
 data<- cbind(c(1:n), x)
 trans <- NULL
 for(i in 1:n){
   t <- rep(data[i,1], data[i,2])
   trans <- c(trans,t)
 }
 difference<- NULL
 for (i in 1:(length(trans)-1)){
   interdiff <- trans[i+1] - trans[i]
   difference <- c(difference,interdiff)
 }
  return(as.integer(difference))
}

check_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
## [24]  0  1  1  1  1  0  0  1  6 11  3  1  0  1  3  8  1  0  3  0
```

```r
#(b)

Poisson_process<-function(num_years,mean_genres){
  x<-rpois(num_years,mean_genres)
  coef_of_var<-sd(check_interval(x))/mean(check_interval(x))
  return(list(check_interval(x),coef_of_var))
}
Poisson_process(161,0.273)
```

```
## [[1]]
##  [1]  1  1  0  2  1  4  1  3  6  0  3  3  1  3  1  8  3  1 13  3  1 12  6
## [24]  2  2  0  3  0  1  6  2  3  0  3  4  0  0  5  0 12  1  2  0  9  2  1
## [47]  0  6  9
##
## [[2]]
## [1] 1.096501
```

```r
#viii.
simu<-100000
sample<-rep(NA, simu)
for(i in 1:simu){
  sample[i] <- as.numeric(Poisson_process(161, 0.273)[2])
}
fraction <- length(sample[sample>sd(intergenre_intervals)/mean(intergenre_intervals)])/simu
fraction
```

```
## [1] 0.2293
```

```r
#ix.
#The fraction is closed to the p-value, and it is greater than 0.05, thus we need to accept our assumpt
#So genres don't tend to appear together in bursts.
```