Homework 7 Solutions

Cynthia Rush (cgr2130)

November 14, 2016

Part 1

```
genres <- read.csv("moretti.csv", as.is = TRUE)
head(genres)
```

```
##          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
```

i.

```
poisLoglik <- function(lambda, data) {
  return(sum(dpois(data, lambda = lambda, log = TRUE)))
}

data1 <- c(1, 0, 0, 1, 1)
lambda1 <- 1

poisLoglik(lambda1, data1)
```

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

```
poisLoglik(0, data1)
```

```
## [1] -Inf
```

ii.

```
count_new_genres <- function(year) {
  return(sum(genres$Begin == year))
}

count_new_genres(1803)
```

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

```
count_new_genres(1850)
```

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

iii.

```
years <- seq(1740, 1900)
num.years <- length(years)
new_genres <- rep(NA, num.years)
names(new_genres) <- years

for (i in 1:num.years) {
  new_genres[i] <- count_new_genres(years[i])
}

twoyears <- which(years == 1803 | years == 1850)
twoyears
```

```
## [1]  64 111
```

```
new_genres[twoyears]
```

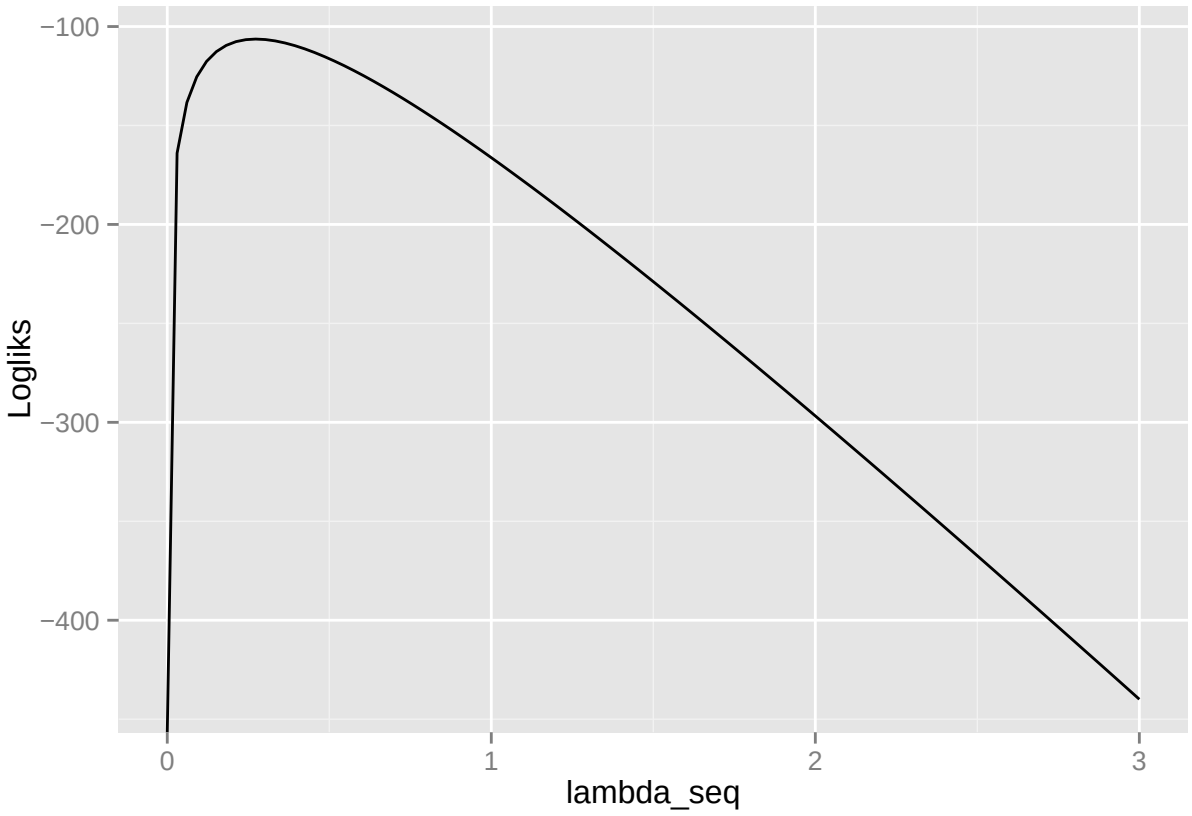
```
## 1803 1850
##    0    3
```

iv.

```
lambda_seq  <- seq(0, 3, length.out = 100)
num.lambdas <- length(lambda_seq)
Logliks     <- rep(NA, num.lambdas)
for (i in 1:num.lambdas) {
  Logliks[i] <- poisLoglik(lambda_seq[i], new_genres)
}
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
ggplot() +
  geom_line(mapping = aes(x = lambda_seq, y = Logliks)) +
  labs(main = "Plot of the Log Likelihood Function", ylab = "Log Likelihood", xlab = "Lambda")
```




```
lambda_seq[which.max(Logliks)]
```

```
## [1] 0.2727273
```

v.

```
NegpoisLoglik <- function(lambda, data) {
  return(-sum(dpois(data, lambda = lambda, log = TRUE)))
}
nlm(NegpoisLoglik, 1, new_genres)
```

```
## Warning in dpois(data, lambda = lambda, log = TRUE): NaNs produced

## Warning in nlm(NegpoisLoglik, 1, new_genres): NA/Inf replaced by maximum
## positive value

## Warning in dpois(data, lambda = lambda, log = TRUE): NaNs produced

## Warning in nlm(NegpoisLoglik, 1, new_genres): NA/Inf replaced by maximum
## positive value

## Warning in dpois(data, lambda = lambda, log = TRUE): NaNs produced

## Warning in nlm(NegpoisLoglik, 1, new_genres): NA/Inf replaced by maximum
## positive value
```

```
## Warning in dpois(data, lambda = lambda, log = TRUE): NaNs produced

## Warning in nlm(NegpoisLoglik, 1, new_genres): NA/Inf replaced by maximum
## positive value

## Warning in dpois(data, lambda = lambda, log = TRUE): NaNs produced

## Warning in nlm(NegpoisLoglik, 1, new_genres): NA/Inf replaced by maximum
## positive value

## Warning in dpois(data, lambda = lambda, log = TRUE): NaNs produced

## Warning in nlm(NegpoisLoglik, 1, new_genres): NA/Inf replaced by maximum
## positive value

## $minimum
## [1] 106.3349
##
## $estimate
## [1] 0.2732914
##
## $gradient
## [1] 2.984279e-07
##
## $code
## [1] 1
##
## $iterations
## [1] 11
```

vi.

```
intergenre_intervals <- diff(sort(genres$Begin))
mean(intergenre_intervals)
```

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

```
sd(intergenre_intervals)
```

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

```
moretti.coef <- sd(intergenre_intervals)/mean(intergenre_intervals)
```

vii.

a.

```
intergenre_calc <- function(new.genres) {
  names(new.genres) <- 1:length(new.genres)
  new.genres        <- new.genres[new.genres !=0]
  years             <- as.numeric(rep(names(new.genres), new.genres))
  return(diff(sort(years)))
}

intergenre_intervals2 <- intergenre_calc(new_genres)
all(intergenre_intervals == intergenre_intervals2)
```

```
## [1] TRUE
```

b.

```
Pois.sim <- function(num.years, mean.genres) {
  samples <- rpois(num.years, lambda = mean.genres)
  intergenre_intervals <- intergenre_calc(samples)
  coef.of.var <- sd(intergenre_intervals)/mean(intergenre_intervals)
  return(list(intergenre_intervals = intergenre_intervals, coef.of.var = coef.of.var))
}

for (i in 1:10) {
  res <- Pois.sim(141, 0.273)
  print(mean(res$intergenre_intervals))
}
```

```
## [1] 2.708333
## [1] 3.4
## [1] 3.414634
## [1] 3.186047
## [1] 4.090909
## [1] 3.219512
## [1] 2.327586
## [1] 3.857143
## [1] 3.702703
## [1] 3.261905
```

viii.

```
n <- 100000
coef.of.var <- rep(NA, n)

for (i in 1:n) {
  res <- Pois.sim(141, 0.273)
  coef.of.var[i] <- res$coef.of.var
}

mean(coef.of.var > moretti.coef)
```

```
## [1] 0.23229
```

ix.

The above result tells us that there isn't really any evidence that new genres appear together in bursts. If the appearance of new genres were truly random (not clustered), meaning they appear according to a Poisson process, then around 23 precent of the time we would see results as or more clustered than Moretti's data (when we consider the cefficient of variation as a measure of the amoutn of cluster).