Homework #4

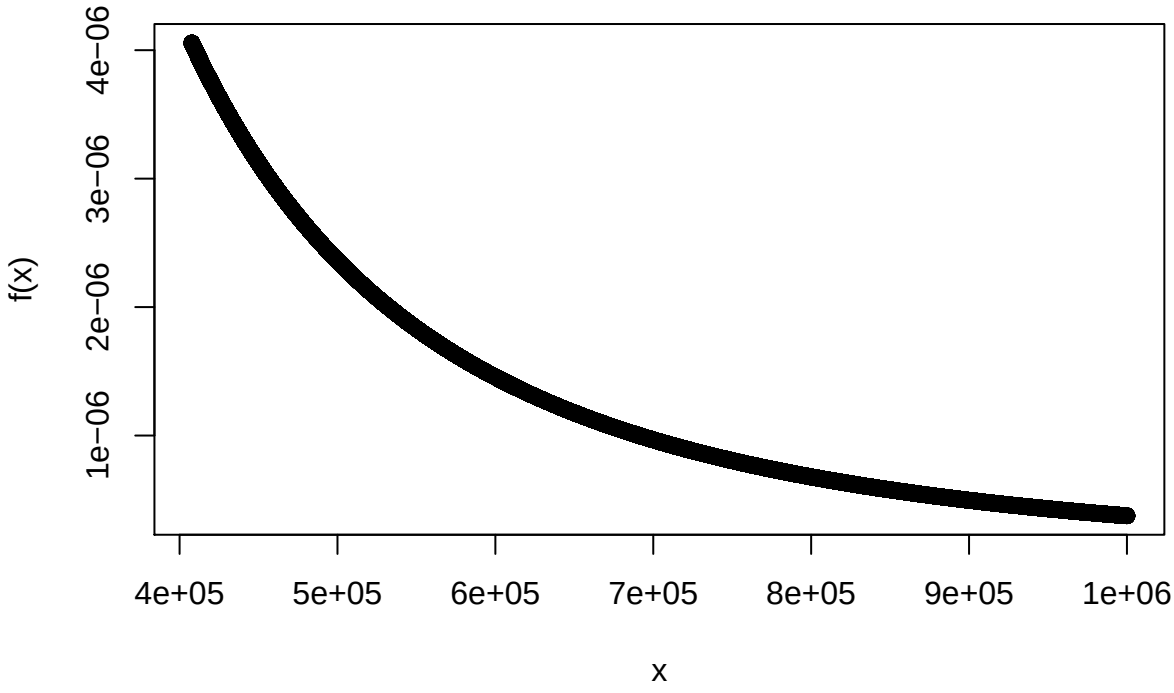
Fan Yang, fy2232

November 24, 2017

Part 1

i.

```
ahat = 2.654
xmin = 407760
f <- function(x, a = ahat, x_min = xmin){
  return ((a-1)/x_min*(x/x_min)^(-a))
}
xi <- 407760:1000000
plot(xi, f(xi),xlab = "x",ylab="f(x)")
```



ii. $F^{-1}(u) = x_{min}(1-u)^{\frac{1}{-a+1}}$

```
upper.income <- function(u, a = ahat,x_min=xmin ){
  return (x_min*(1-u)^(1/(1-a)))
}
upper.income(0.5)
```

```
## [1] 620020.2
```

iii.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.2
```

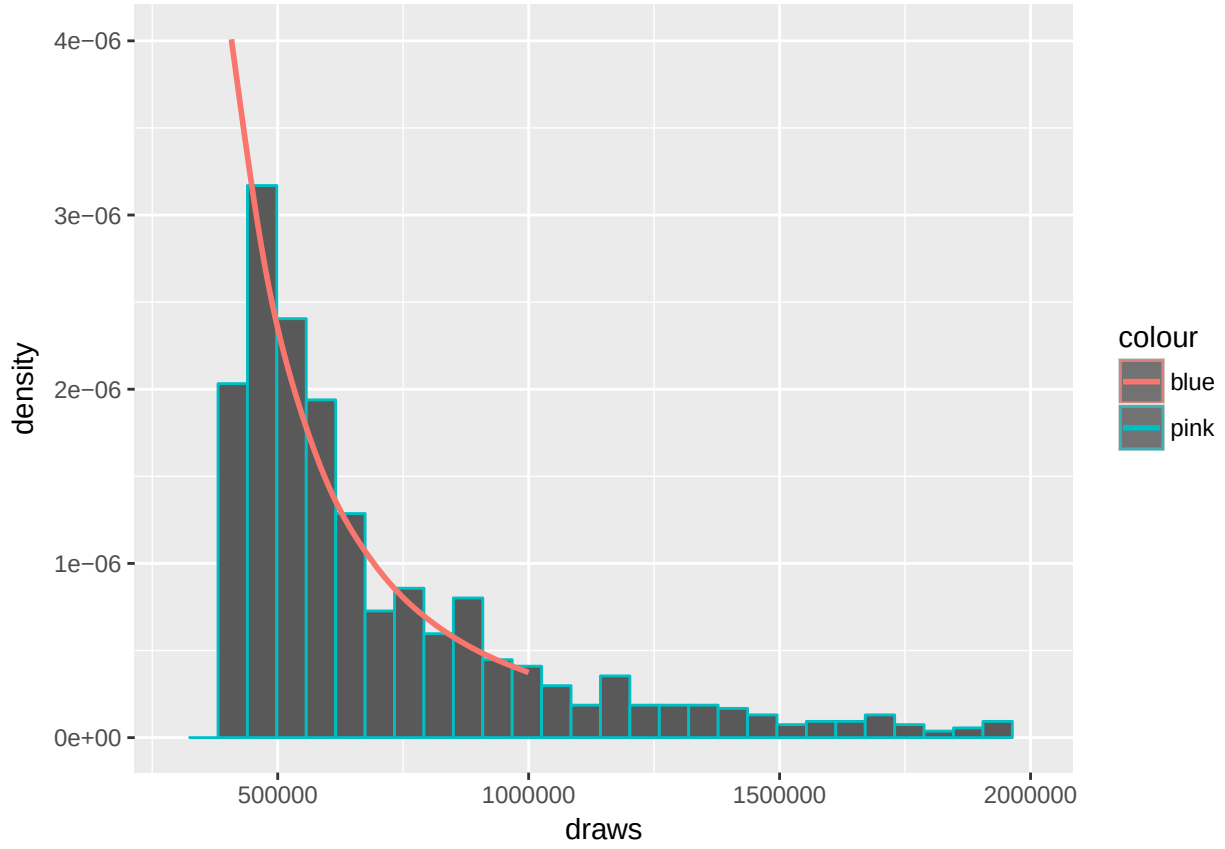
```
samp_p <- runif(1000)
draws <- upper.income(samp_p)
ggplot() +
  geom_histogram(aes(x = draws, y = ..density..,col="pink")) +
  #geom_density(aes(x = draws)) +
  geom_smooth(aes(x = xi, y = f(xi),col="blue")) +
  xlim(3*10^5,2*10^6)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 85 rows containing non-finite values (stat_bin).
```

```
## `geom_smooth()` using method = 'gam'
```

```
## Warning: Removed 1 rows containing missing values (geom_bar).
```



iv.

```
median(draws)
```

```
## [1] 607798.4
```

```
upper.income(0.5)
```

```
## [1] 620020.2
```

Part 2

i.

```
moretti <- read.csv("moretti.csv",header = T, as.is = T)
poisLoglik <- function(lambda, data){
   return (sum(log(lambda^data*exp(-lambda)/factorial(data))))
}
poisLoglik(1,c(1,0,0,1,1))
```

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

ii.

```
count_new_genres <- function(year){
   return (sum(moretti$Begin==year))
}
count_new_genres(1803)
```

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

```
count_new_genres(1850)
```

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

iii.

```
new_genres = sapply(1740:1900,count_new_genres)
1803-1740+1;1850-1740+1
```

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

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

```
new_genres[64]
```

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

```
new_genres[111]
```

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

```
new_genres[64]==count_new_genres(1803)
```

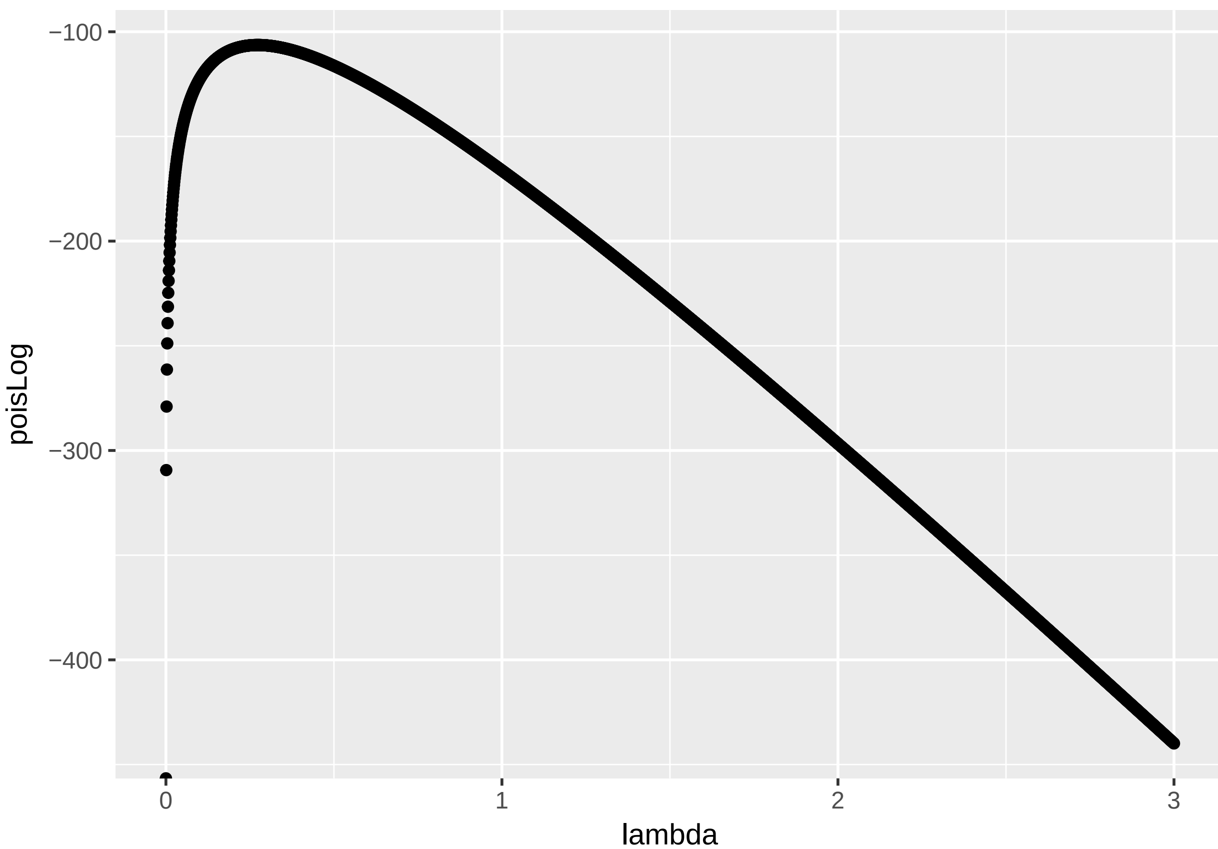
```
## [1] TRUE
```

```
new_genres[111]==count_new_genres(1850)
```

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

iv.

```
lambda <- c(0:3000)/1000
poisLog <- sapply(lambda, poisLoglik, data=new_genres)
ggplot() +
  geom_point(aes(x =lambda , y = poisLog))
```



v.

```
lambda <- c(0:3000)/1000
poisLoglik2 <- function(lambda, data=new_genres){
  return (-sum(log(lambda^data*exp(-lambda)/factorial(data))))
}

nlm(poisLoglik2,c(1))$estimate
```

```
## Warning in log(lambda^data * exp(-lambda)/factorial(data)): NaNs produced

## Warning in nlm(poisLoglik2, c(1)): NA/Inf replaced by maximum positive
## value

## Warning in log(lambda^data * exp(-lambda)/factorial(data)): NaNs produced

## Warning in nlm(poisLoglik2, c(1)): NA/Inf replaced by maximum positive
## value

## Warning in log(lambda^data * exp(-lambda)/factorial(data)): NaNs produced

## Warning in nlm(poisLoglik2, c(1)): NA/Inf replaced by maximum positive
## value

## Warning in log(lambda^data * exp(-lambda)/factorial(data)): NaNs produced

## Warning in nlm(poisLoglik2, c(1)): NA/Inf replaced by maximum positive
## value

## Warning in log(lambda^data * exp(-lambda)/factorial(data)): NaNs produced
```

```
## Warning in nlm(poisLoglik2, c(1)): NA/Inf replaced by maximum positive
## value
```

```
## Warning in log(lambda^data * exp(-lambda)/factorial(data)): NaNs produced
```

```
## Warning in nlm(poisLoglik2, c(1)): NA/Inf replaced by maximum positive
## value
```

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

vi.

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

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

```
sd(intergenre_intervals)
```

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

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

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

vii.

a)

```
pois_draws <- rpois(161,0.273)
```

b)

```
pois_f <- function(num_new_genres){
  yearinter = c()
  for (i in c(1:length(num_new_genres))){
    yearinter <- c(yearinter,rep(i,num_new_genres[i]))
  }
  return (diff(yearinter))
}
all(intergenre_intervals == pois_f(new_genres))
```

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

c)

```
simu_pois <- function(num.years,mean.genres){
  pois_draws <- rpois(num.years,mean.genres)
  interv <- pois_f(pois_draws)
  coeffov <- sd(interv) / mean(interv)
  return (list(intervals = interv,coefficient_of_variation=coeffov))
}
simu_pois(161,0.273)
```

```
## $intervals
##  [1]  2  4  1  6  7  6  1  4  1  6  3  2  8  1  0  7  5  0  7  9  1  1  1
## [24]  5  1  5  9  5  1  4  2  5  7  1 11  0  1 11
##
## $coefficient_of_variation
## [1] 0.8032696
```

viii.

```r
coeffofv <- c()
for (i in 1:10000){
  coeffofv[i] <- simu_pois(161,0.273)$coefficient_of_variation
}
coeffofv_Moretti <- sd(intergenre_intervals) / mean(intergenre_intervals)
sum(coeffofv>coeffofv_Moretti) / length(coeffofv)
```

```
## [1] 0.2208
```