Homework 7 Solutions Help

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
```

```
genres$Begin
```

```
##  [1] 1740 1748 1759 1766 1768 1770 1773 1789 1790 1791 1800 1804 1808 1814
## [15] 1822 1825 1826 1828 1830 1830 1832 1838 1839 1846 1846 1847 1848 1849
## [29] 1850 1850 1850 1851 1857 1868 1871 1872 1872 1873 1876 1884 1885 1885
## [43] 1888 1888
```

```
which(genres$Begin==1846)
```

```
## [1] 24 25
```

```
#genres
```

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
```

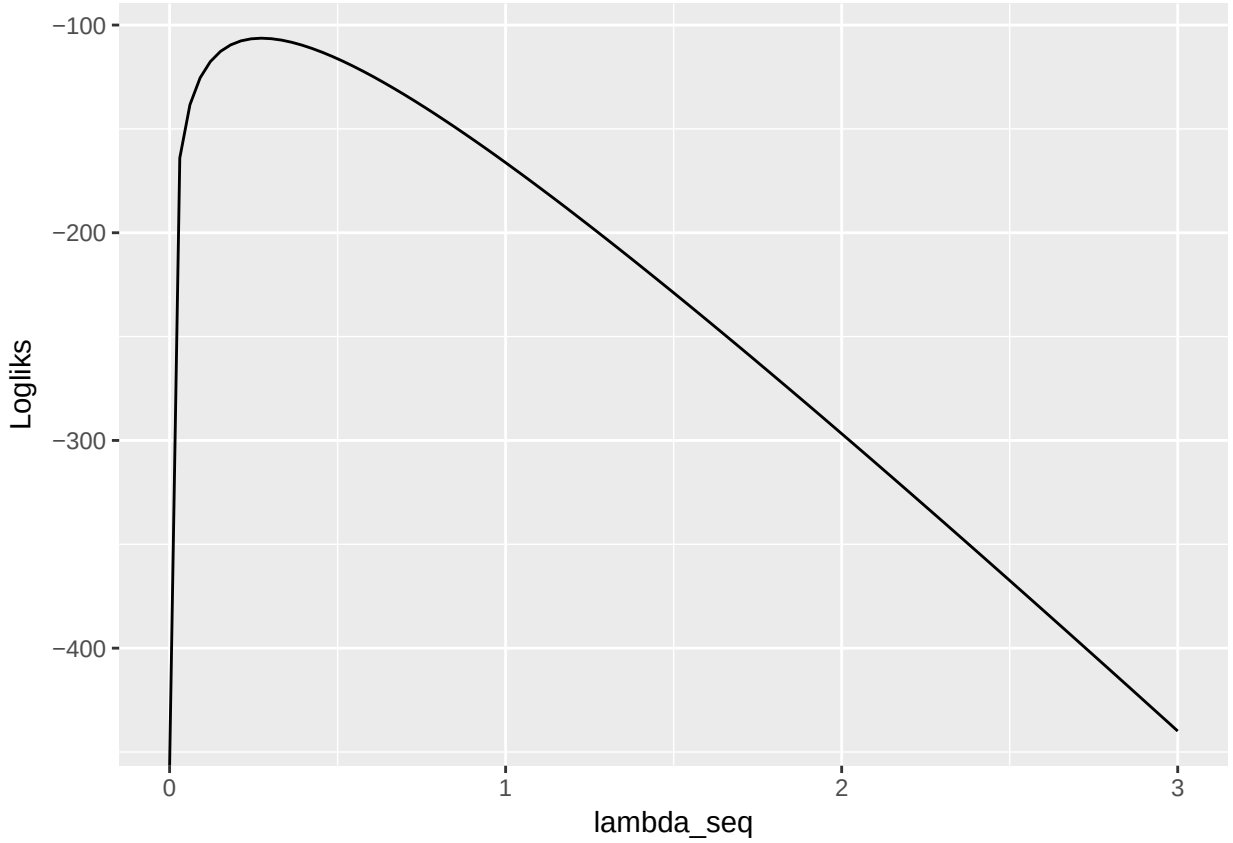
```
new_genres
```

```
## 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754
##    1    0    0    0    0    0    0    0    1    0    0    0    0    0    0
## 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769
##    0    0    0    0    1    0    0    0    0    0    0    1    0    1    0
## 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784
##    1    0    0    1    0    0    0    0    0    0    0    0    0    0    0
## 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799
##    0    0    0    0    1    1    1    0    0    0    0    0    0    0    0
## 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814
##    1    0    0    0    1    0    0    0    1    0    0    0    0    0    1
## 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829
##    0    0    0    0    0    0    0    1    0    0    1    1    0    1    0
## 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844
##    2    0    1    0    0    0    0    0    1    1    0    0    0    0    0
## 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859
##    0    2    1    1    1    3    1    0    0    0    0    0    1    0    0
## 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874
##    0    0    0    0    0    0    0    0    1    0    0    1    2    1    0
## 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889
##    0    1    0    0    0    0    0    0    0    1    2    0    0    2    0
## 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900
##    0    0    0    0    0    0    0    0    0    0    0
```

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)
```

```
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.

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

```
mean(intergenre_intervals)
```

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

```
sd(intergenre_intervals)
```

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

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

vii.

a.

```
# Look at body of the function
names(new_genres) <- 1:length(new_genres)
new_genres
```

```
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
##   1   0   0   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0
##  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
##   0   1   0   0   0   0   0   0   1   0   1   0   1   0   0   1   0   0
##  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54
##   0   0   0   0   0   0   0   0   0   0   0   0   0   1   1   1   0   0
```

```
##  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72
##   0   0   0   0   0   0   1   0   0   0   1   0   0   0   1   0   0   0
##  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
##   0   0   1   0   0   0   0   0   0   0   1   0   0   1   1   0   1   0
##  91  92  93  94  95  96  97  98  99 100 101 102 103 104 105 106 107 108
##   2   0   1   0   0   0   0   0   1   1   0   0   0   0   0   0   2   1
## 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126
##   1   1   3   1   0   0   0   0   0   1   0   0   0   0   0   0   0   0
## 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144
##   0   0   1   0   0   1   2   1   0   0   1   0   0   0   0   0   0   0
## 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161
##   1   2   0   0   2   0   0   0   0   0   0   0   0   0   0   0   0
```

```
new_genres        <- new_genres[new_genres !=0]
new_genres
```

```
##   1   9  20  27  29  31  34  50  51  52  61  65  69  75  83  86  87  89
##   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  91  93  99 100 107 108 109 110 111 112 118 129 132 133 134 137 145 146
##   2   1   1   1   2   1   1   1   3   1   1   1   1   2   1   1   1   2
## 149
##   2
```

```
 years             <- as.numeric(rep(names(new_genres), new_genres))
years
```

```
##  [1]   1   9  20  27  29  31  34  50  51  52  61  65  69  75  83  86  87
## [18]  89  91  91  93  99 100 107 107 108 109 110 111 111 111 112 118 129
## [35] 132 133 133 134 137 145 146 146 149 149
```

```
# define function
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] FALSE
```

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(161, 0.273)
  print(mean(res$intergenre_intervals))
}
```

```
## [1] 3.137255
## [1] 3.5
## [1] 4.25
## [1] 3.590909
## [1] 3.361702
## [1] 2.823529
## [1] 3.55814
## [1] 3.418605
## [1] 3.145833
## [1] 3.4
```

viii.

```
n <- 10000
coef.of.var <- rep(NA, n)

for (i in 1:n) {
  res <- Pois.sim(161, 0.273)
  coef.of.var[i] <- res$coef.of.var
}

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

```
## [1] 0.2276
```

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 percent of the time we would see results as or more clustered than Moretti's data (when we consider the coefficient of variation as a measure of the amount of cluster).