STAT40810 — Stochastic Models

Brendan Murphy

Week 9

Inhomogeneous Poisson Process

Inhomogeneous Poisson process: definition

In some cases, the assumption that the expected arrival rate is constant is not realistic.

Definition

Let $\lambda(t) > 0$ be a function of the time t. An inhomogeneous Poisson process with intensity $\lambda(t)$ is a counting process $\{X_t\}$ with

1. independent increments;
2. $\mathbb{P}(X_{t+h} - X_t = 1) = \lambda(t)h + o(h)$ when $h \to 0$;
3. $\mathbb{P}(X_{t+h} - X_t > 1) = o(h)$ when $h \to 0$.

Inhomogeneous Poisson process: properties

Let $\{X_t\}$ be an inhomogeneous Poisson process with intensity $\lambda(t)$. Then we can prove that

$$N_{(s,t]} = X_t - X_s \sim \mathcal{P}\left(\int_s^t \lambda(x)\mathrm{d}x\right).$$

In particular, when $\lambda(t) = \lambda$ is constant, $\{X_t\}$ is a Poisson process and we have:

$$N_{(s,t]} = X_t - X_s \sim \mathcal{P}\left((t-s)\lambda\right).$$

Example Code: Homogeneous

- We can fit a Poisson process to the Ireland vs New Zealand data using the `NHPoisson` package.
- Let's start with a homogeneous Poisson process.

```
# Load the NHPoisson package
library(NHPoisson)

# Read in the scoring time data
x <- c(3,4,10,17,21,23,34,49,52,57,59,64,76)

# Fit a homogeneous Poisson process
fit0 <- fitPP.fun(n=80,posE=x,start=list(b0=0))
summary(fit0)
```

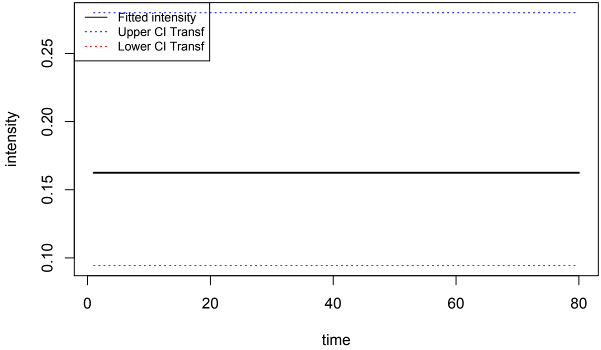
Fit

- The output is:

```
Maximum likelihood estimation

Call:
fitPP.fun(start = list(b0 = 0), posE = x, nobs = 80)

Coefficients:
    Estimate Std. Error
b0 -1.817077  0.2773501

-2 log L: 73.24401
```

Estimated Intensity

- The estimated intensity (plotted) is:



Example Code: Linear

- The `NHPoisson` package allows the intensity to depend on covariates (in this case we use time).

```
# Allow for a linear intensity function
# Set up the time covariate (I have made this 1 to 80 in 1 minute intervals)
timerange<-seq(1,80,by=1)
covariates<-timerange

#Fit the inhomogeneous Poisson process
fit1 <- fitPP.fun(covariates=covariates,posE=x,start=list(b0=0,b1=0))
summary(fit1)
```

Fit

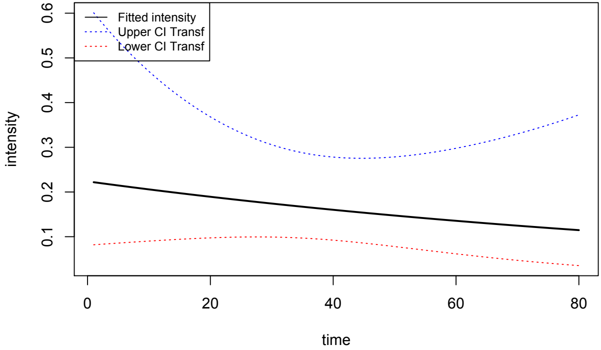
- The output is:

```
Maximum likelihood estimation

Call:
fitPP.fun(covariates = covariates, start = list(b0 = 0, b1 = 0),
    posE = x)

Coefficients:
       Estimate Std. Error
b0 -1.497201255 0.51854630
b1 -0.008356157 0.01214462

-2 log L: 72.76531
```

- We haven't gained much by allowing for inhomogeneity using a linear intensity.

Estimated Intensity

- The estimated intensity (plotted) is:



Example Code: Quadratic

- We also investigate if a quadratic intensity is worthwhile. That is,

$$\log \lambda(t) = b_0 + b_1 t + b_2 t^2.$$

```
# Allow for a quadratic intensity function
timerange<-seq(1,80,by=1)
covariates<-cbind(timerange,timerange^2)

#Fit the inhomogeneous Poisson process
fit2 <- fitPP.fun(covariates=covariates,posE=x,start=list(b0=0,b1=0,b2=0))
summary(fit2)
```

Fit

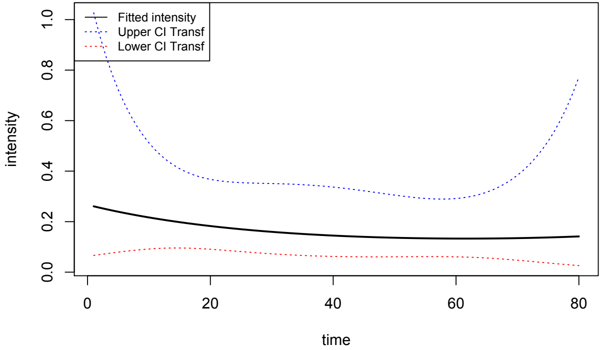
- The output is:

```
Maximum likelihood estimation

Call:
fitPP.fun(covariates = covariates, start = list(b0 = 0, b1 = 0,
    b2 = 0), posE = x)

Coefficients:
        Estimate   Std. Error
b0 -1.3217711615 0.7377627733
b1 -0.0225982871 0.0459999286
b2  0.0001835105 0.0005719938

-2 log L: 72.66407
```

- We haven't gained much by allowing for inhomogeneity using a quadratic intensity either.

Estimated Intensity

- The estimated intensity (plotted) is:



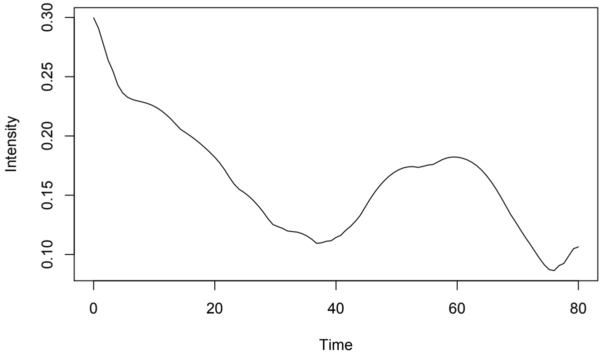
Non parametric Intensity

- The `lpint` package in R allows for the intensity to be estimated non-parametrically.
  It estimates the intensity in an analogous manner to the kernel smoothing regression methods we have already seen.

```
#Load the lpint package
library(lpint)

# Read in the scoring time data
x <- c(3,4,10,17,21,23,34,49,52,57,59,64,76)

# Set up the number of scores at each time point
#Set to be 1 in this case
y<-rep(1,length(x))

#Fit the inhomogeneous Poisson process
fit <- lpint(jmptimes=x,jmpsizes=y,Tau=80)
plot(fit,type="l",xlab="Time",ylab="Intensity")
```

Estimated Intensity

- The estimated intensity (plotted) is:



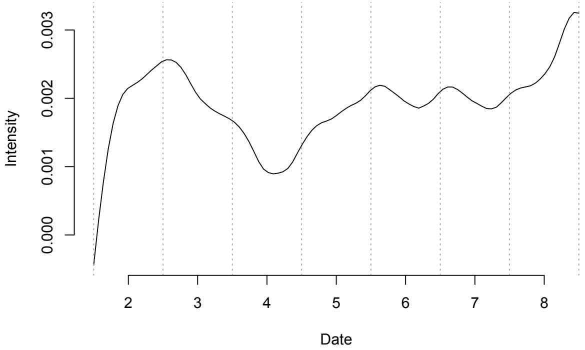
Example: Washington DC Bikeshare Scheme

- Data were collected from the Capital Bikeshare scheme in Washington DC.

- The times at which bikes were collected from the “Massachusetts Ave & Dupont Circle NW” bike station were recorded from July 2nd to July 8th, 2016.
- A total of 1166 collection events were observed in this seven day period.

Estimated Intensity: Bikeshare Scheme

- The intensity of the process was estimated using the `lpint` package:



- Notice the dip in intensity on the 4th of July!

Code

```
# Read in data (file downloaded from the Capital bikeshare scheme website)
dat0 <- read.csv("~/Downloads/2016-Q3-cabi-trips-history-data/2016-Q3-Trips-History-Data-1.csv")

# Extract data on collections from station 31200
dat <- dat0[dat0$Start.station.number==31200,]

# Extract the relevant events (found manually)
dat<-dat[10510:9345,]

# Extract the event times
x<- as.character(dat$Start.date)

# Put into numeric format (seconds from start at midnight 7/2/2016)
library(lubridate)
x <- mdy_hm(x)
x <- as.numeric(x)
x <- x-as.numeric(mdy_hm("7/2/2016 0:00"))

# Set up the number of scores at each time point
# Set to be 1 in this case
y<-rep(1,length(x))

#Fit the inhomogeneous Poisson process
fit <- lpint(jmptimes=x,jmpsizes=y,Tau=7*24*60*60)

# Plot the fit
plot(fit,type="l",xlab="Date",ylab="Intensity",axes=FALSE)
abline(v=(0:7)*(24*60*60),col="darkgray",lty=3)
axis(1,at=((0:6)+0.5)*(24*60*60),labels=c(2:8))
axis(2)
```