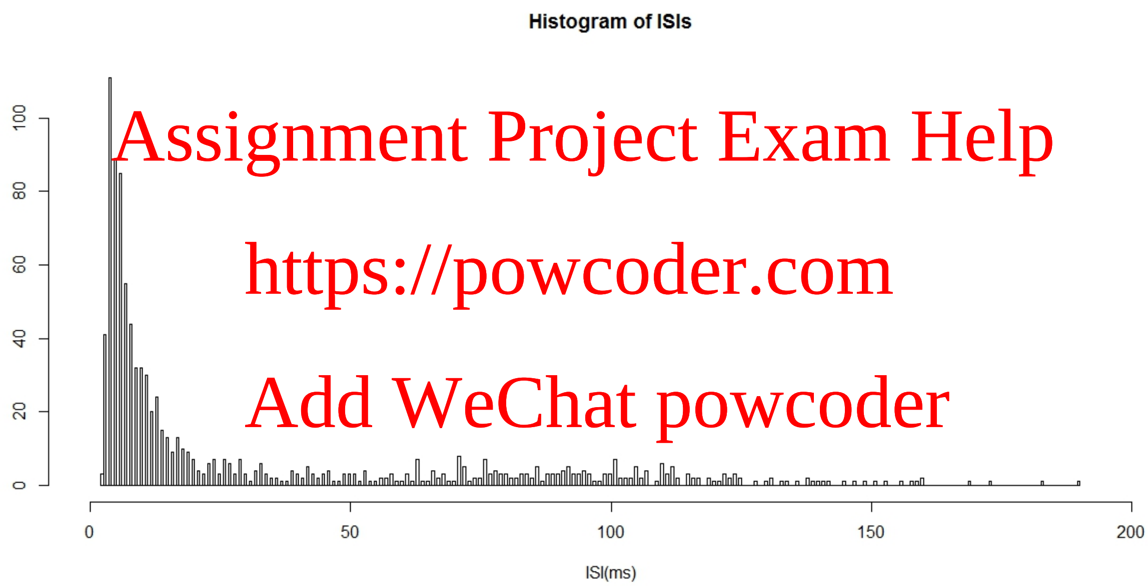
MA 568 Statistical Analysis of Point Process Data
Problem set 1

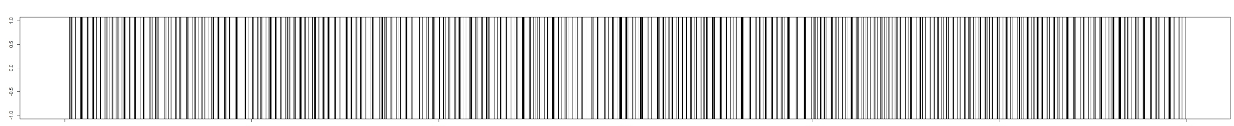
Long Tao and Uri Eden
Department of Mathematics and Statistics, Boston university

Question 1
Download the **Retinal_ISIs.mat** file from the course website. This file contains the waiting times, or inter-spike intervals (ISIs), in milliseconds, of a single retinal ganglion cell over 30 seconds.

```
library(R.matlab)
data <- readMat("C:\\Users\\Long_Tao\\Downloads\\Retinal_ISIs(1).mat")
ISIs <- data$ISIs
hist (ISIs , breaks = 300, xlab='ISI(ms)',xlim=c(0,200), ylab='')

spiketimes <- cumsum (ISIs)
T = max(spiketimes)
plot(1:T, rep(0,T),type='n',xlab='',ylab='')
abline(v=spiketimes)
```

First we plot the spiking activity as a histogram of the distribution of the times between spiking events and a spike train time series.

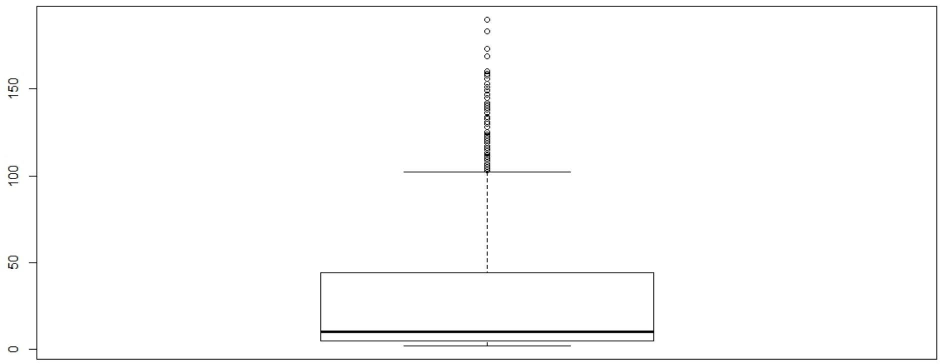




The histogram and spike train time series suggest that this neuron fires most of its spikes with an ISI between 5-40 ms. Features to note include an initial refractory period under 5 ms, a large number of spikes with ISIs around 4-8 ms, a large tail that includes ISIs out to 190 ms, and a small second model around 90-100 ms. This suggest that a simple Poisson model will not be able to fully capture the structure observed in this data.

Question 2
Compute a 5-number summary (min, .25 quantile, median, .75 quantile and max) and a box plot for the ISI distribution.

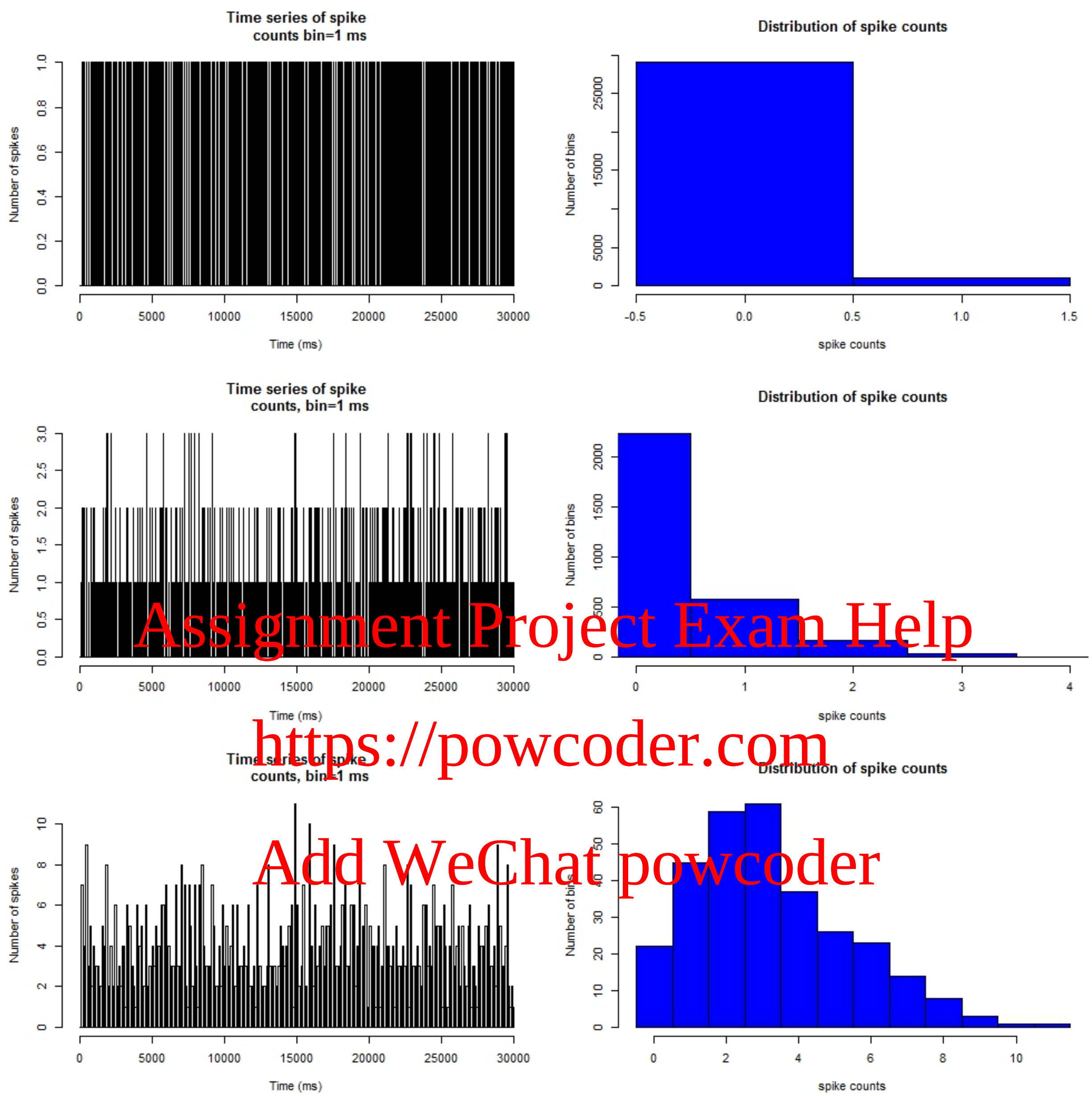
```r
quantile(ISIs, probs=seq(0,1,by=.25))
0%   25%   50%   75%  100%
2      5     10     44    190
boxplot(ISIs)
```



The 5-number summary and boxplot confirm our earlier observation about the distribution of ISIs. Most spikes occur at small ISIs between 5-40 ms, but there are a large number of long ISIs over 100 ms.

Question 3 Discrete time bins of fixed width and count the number of events that occur in each time bin. Set bin size to be 1 ms, 10 ms and 100 ms. Plot the time series of spike counts and the distribution of spike counts as a histogram for each bin width. They look similar to the model distribution of increments of Poisson processes.

```r
par(mfrow=c(3,2))
spike1 <- hist(spiketimes, breaks=c(0:30000), main = 'Time series of spike
counts bin=1 ms', xlab='Time (ms)', ylab = 'Number of spikes')
hist(spike1$counts, breaks=c(0:2)-.5, col= 'blue', xlab='spike counts',
main = 'Distribution of spike counts', ylab='Number of bins')
spike10 <- hist(spiketimes, breaks = seq(0,30000,by=10),, main = 'Time series of spike
counts, bin=1 ms', xlab='Time (ms)', ylab = 'Number of spikes')
hist(spike10$counts, breaks=c(0:5)-.5, col= 'blue', xlim=c(0,4),, xlab='spike counts',
main = 'Distribution of spike counts',ylab='Number of bins')
spike100 <- hist(spiketimes, breaks = seq(0,30000,by=100), main = 'Time series of spike
counts, bin=1 ms', xlab='Time (ms)', ylab = 'Number of spikes')
hist(spike100$counts, breaks=c(0:12)-.5, col= 'blue',, xlab='spike counts',
main = 'Distribution of spike counts',ylab='Number of bins')
```



Question 4

The log-likelihood for a homogeneous Poisson process with rate parameter λ is given by

$$\log L(\lambda) = N(T)\log(\lambda\Delta t) - \lambda T$$

thus the MLE of λ is

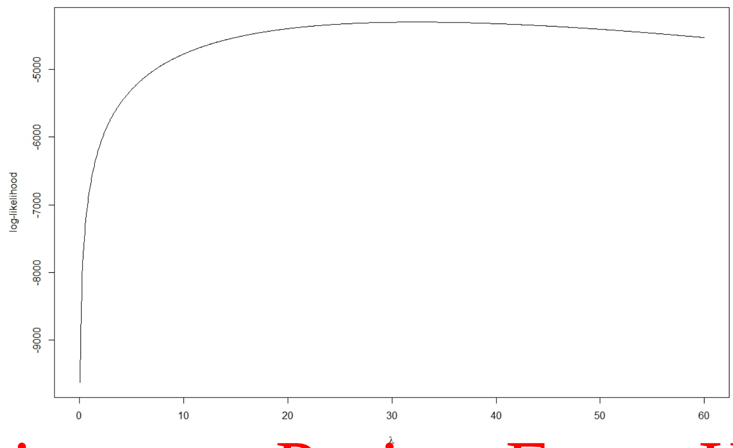
$$\hat{\lambda} = \frac{N(T)}{T} = \frac{972}{30} = 32.4000$$

with

$$s.e.(\hat{\lambda}) = \sqrt{N(T)}/T = 1.039$$

and 95% CI to be

$$[\lambda - 1.96 * s.e., \lambda + 1.96 * s.e.] = [30.3631, 34.4369]$$

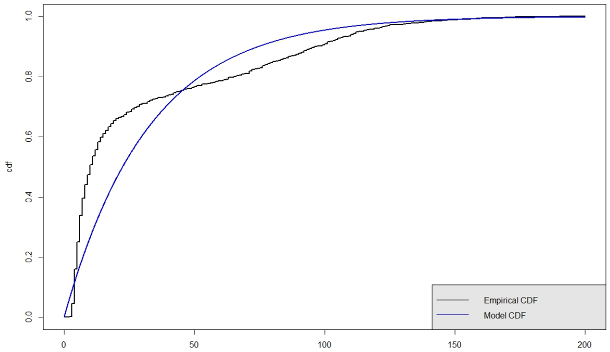


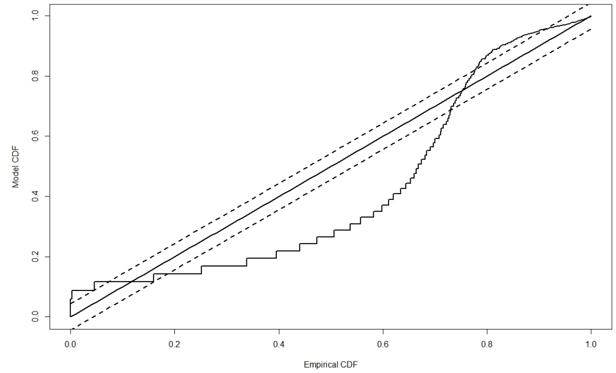
Question 5


```
lam = N/30
w <- seq(0, 200, by=.01)
Femp <- numeric(length(w))
Qs <- ISIs[order(ISIs)]
for (i in 1:length(w)){Femp[i] <- sum(Qs<=w[i])/length(Qs)}
plot(w, Femp, type='l', xlab ='', ylab = 'cdf', lwd=2)
lines(w, 1-exp(-w/lam),col='blue', lwd=2)
legend('bottomright', c('Empirical_CDF', "Model_CDF"), col = c('black','blue'),
lty = c(1, 1),merge = TRUE, bg = "gray90")

KSstat <- max(abs(1-exp(-w/lam) - Femp/N))
# KSstat = 0.2518

plot(Femp, 1-exp(-w/lam), type='l', ylab='Model_CDF', xlab='Empirical_CDF', lwd=2)
x=seq(0,1,by=.01);  y=x
lines(x,y,lty=1, lwd=2)
x=seq(0,1,by=.01);  y=x+1.36/sqrt(N)
lines(x,y,lty=2, lwd=2)
x=seq(0,1,by=.01);  y=x-1.36/sqrt(N)
lines(x,y,lty=2, lwd=2)
```

$$KS - statistic = \max |\hat{F}_{Emp} - F_{dis}| = 0.2518$$

Note that the KS-statistic value may be different if you choose different precision of the Empirical cdf here.

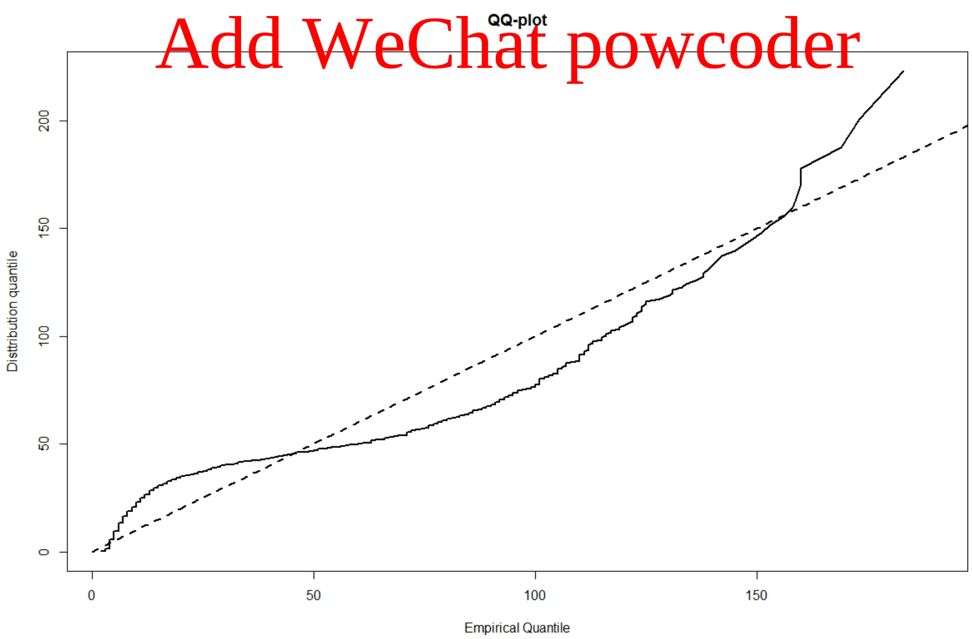




Question 6

```
x <- seq(1/N, 1, by= 1/N)
Qmodel <- qexp(x,rate = 1/lam)
plot (Qs, Qmodel, type='l',lwd=2, xlab='Empirical Quantile',
main='QQ-plot',ylab='Distribution quantile')
x<-c(0:200); y <- x
lines (x,y, lty=2, lwd=2)
```


From QQ-plot, we see that for large ISI values (170-182 ms), the QQ plot most deviates the expect.



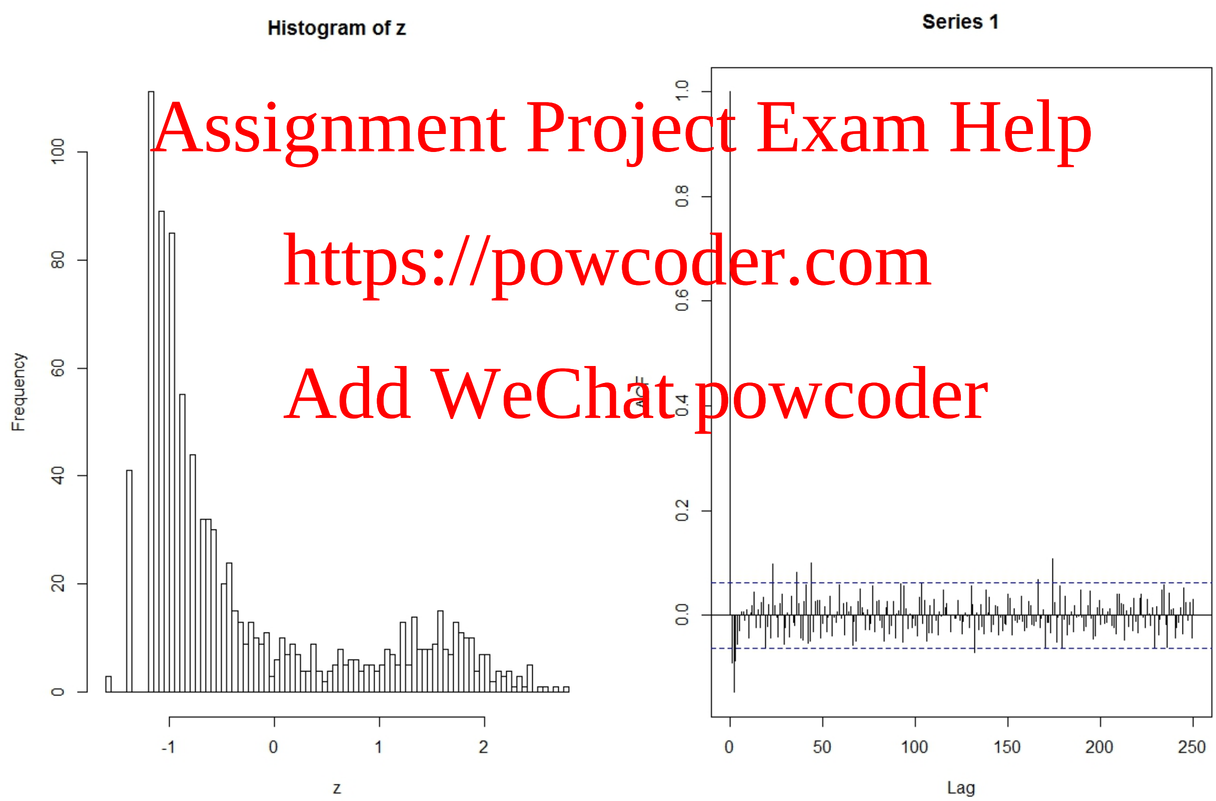
Question 7

```r
FF1 <- var(spike1$counts)/ mean(spike1$counts); FF1 # 0.9676
qgamma(c(.025,.975), length(spike1$counts)/2, scale=2/length(spike1$counts)) #0.9841 1.0161
FF10 <- var(spike10$counts)/ mean(spike10$counts); FF10  # 1.1661
qgamma(c(.025,.975), length(spike10$counts)/2, scale=2/length(spike10$counts))# 0.9500 1.0512
FF100 <- var(spike100$counts)/mean(spike100$counts); FF100   # 1.4583
qgamma(c(.025,.975), length(spike100$counts)/2, scale=2/length(spike100$counts)) # 0.8464 1.1662
```

Question 8

Plot the rescaled ISIs and the autocorrelation function of the observed ISIs with 95% confidence bounds. Note the 95% confidence interval for the correlation coefficient of independent Gaussian rv is $1.96/\sqrt{n}$.

```r
par(mfrow=c(1,2))
z = qnorm(pexp(ISIs,1/lam)); hist(z, breaks=100)
acf(ISIs, lag = 250)
```



The rescaled ISIs are not normally distributed and at small lags, the autocorrelation function falls outside of the 95% confidence bounds more often than expected by chance alone for small lags, suggesting that nearby spikes have ISIs that are not independent.

Question 9

From the above analysis, we can conclude that under a simple Poisson model the firing rate of this neuron is likely somewhere between 30-34 Hz. However, this Poisson model fails to account for many aspects of the spiking activity including its refractoriness, bursting, and long ISIs. Our analysis of the Fano Factor suggests that there is less variability in the 1 ms increments and more in the 2o ms and 100 ms increments than what would be expected from a Poisson process. The autocorrelation analysis suggests a dependence structure on past spiking activity. Therefore, describing the structure of this data will require history dependent point process models.