Kernel Density Estimation Using Earthquake Magnitude Data

The first portion of this program generates the vector of earthquake magnitudes and displays descriptive statistics for these data. Necessary library packages are also identified.

Earthquake magnitude descriptive statistics

```
mag <- c(0.70, 0.74, 0.64, 0.39, 0.70, 2.20, 1.98, 0.64, 1.22, 0.20, 1.64, 1.02,
         2.95, 0.90, 1.76, 1.00, 1.05, 0.10, 3.45, 1.56, 1.62, 1.83, 0.99, 1.56,
         0.40, 1.28, 0.83, 1.24, 0.54, 1.44, 0.92, 1.00, 0.79, 0.79, 1.54, 1.00,
         2.24, 2.50, 1.79, 1.25, 1.49, 0.84, 1.42, 1.00, 1.25, 1.42, 1.15, 0.93,
         0.40, 1.39)

my_stats <- function(x) {
  cat("\n     mean:", mean(x, na.rm = TRUE))
  cat("\n   median:", median(x, na.rm = TRUE))
  cat("\n    range:", range(x))
  cat("\n       sd:", sd(x, na.rm = TRUE))
  cat("\n variance:", var(x, na.rm = TRUE))
  cat("\n       Q1:", quantile(x, probs = c(0.25), na.rm = TRUE))
  cat("\n       Q3:", quantile(x, probs = c(0.75), na.rm = TRUE))
  cat("\n      P10:", quantile(x, probs = c(0.10), na.rm = TRUE),"\n")
}
my_stats(mag)
```

```
##
##      mean: 1.2336
##    median: 1.1
##     range: 0.1 3.45
##        sd: 0.6634165
##  variance: 0.4401215
##        Q1: 0.8
##        Q3: 1.555
##       P10: 0.526
```

```
library(sROC)
```

```
## sROC 0.1-2 loaded
```

Kernel Density Estimation

Kernel density estimation is a non-parametric way to estimate the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing procedure involving inferences about the population based on a data sample. It is useful for enhancing displays, showing differences between distributions and in some cases estimating quantiles.

Use With Histograms

A kernel density estimate can be superimposed upon a histogram in base R. To do this, it is necessary to scale the density estimate to reflect the class frequencies in the histogram. A multiplier must be calculated and applied to the kernel density estimate. The object generated by hist() contains both counts and density values which can be used. Checking str() reveals this.

```
str(hist(mag))
```

```
## List of 6
##  $ breaks  : num [1:8] 0 0.5 1 1.5 2 2.5 3 3.5
##  $ counts  : int [1:7] 5 18 13 9 3 1 1
##  $ density : num [1:7] 0.2 0.72 0.52 0.36 0.12 0.04 0.04
##  $ mids    : num [1:7] 0.25 0.75 1.25 1.75 2.25 2.75 3.25
##  $ xname   : chr "mag"
##  $ equidist: logi TRUE
##  - attr(*, "class")= chr "histogram"
```

```
multiplier <- mean(hist(mag, cells <- seq(from = 0.0, to = 3.5, by = 0.5),
                        right = FALSE)$counts/hist(mag)$density)
```

test <- hist(mag,cells <- seq(from = 0.0, to = 3.5, by = 0.5),right = FALSE) There are different options for kernel density estimation that can be found using the help function. Checking the structure of the object generated shows the variable y which must be adjusted by the multiplier for scaling purposes.

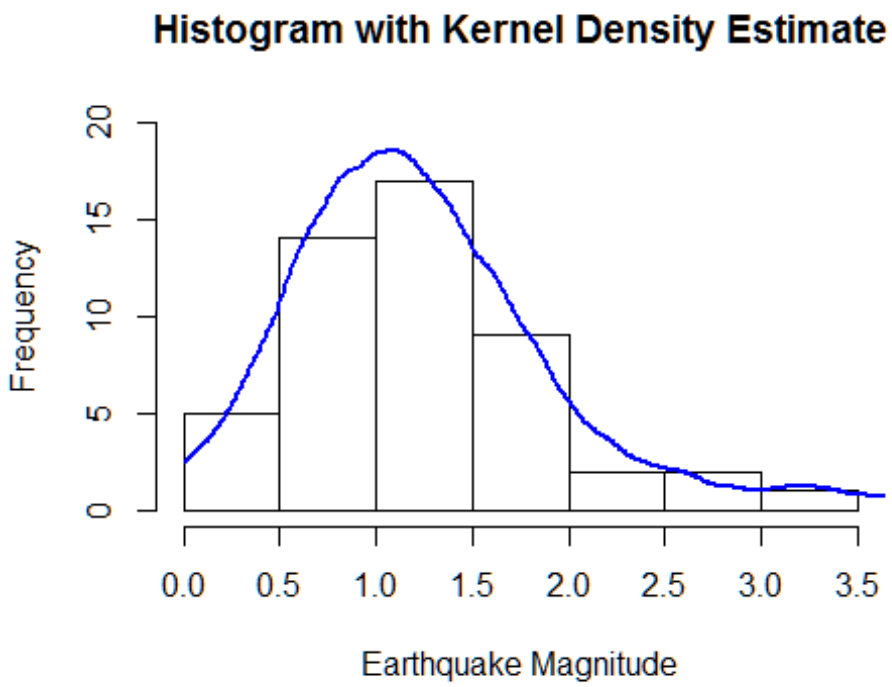
```
object <- density(mag,bw="nrd0",adjust=1,kernel=c("epanechnikov"),from=0,to=4)
str(object)
```

```
## List of 7
##  $ x        : num [1:512] 0 0.00783 0.01566 0.02348 0.03131 ...
##  $ y        : num [1:512] 0.0886 0.091 0.0935 0.0961 0.0992 ...
##  $ bw       : num 0.232
##  $ n        : int 50
##  $ call     : language density.default(x = mag, bw = "nrd0", adjust = 1, kernel
= c("epanechnikov"),      from = 0, to = 4)
##  $ data.name: chr "mag"
##  $ has.na   : logi FALSE
##  - attr(*, "class")= chr "density"
```

```
object$y <- object$y*multiplier
```

A histogram with the kernel density estimate superimposed may now be presented. The kernel density estimate has been scaled to correspond to the frequencies and superimposed on the histogram using base R. Comparable displays may be generated using ggplot2.

```
plot(hist(mag, cells <- seq(from = 0.0, to = 3.5, by = 0.5),right = FALSE),
     ylim = c(0,20),xlab = c("Earthquake Magnitude"), ylab = c("Frequency"),
     main = "Histogram with Kernel Density Estimate")
lines(object, col = "blue", lwd = 2)
```



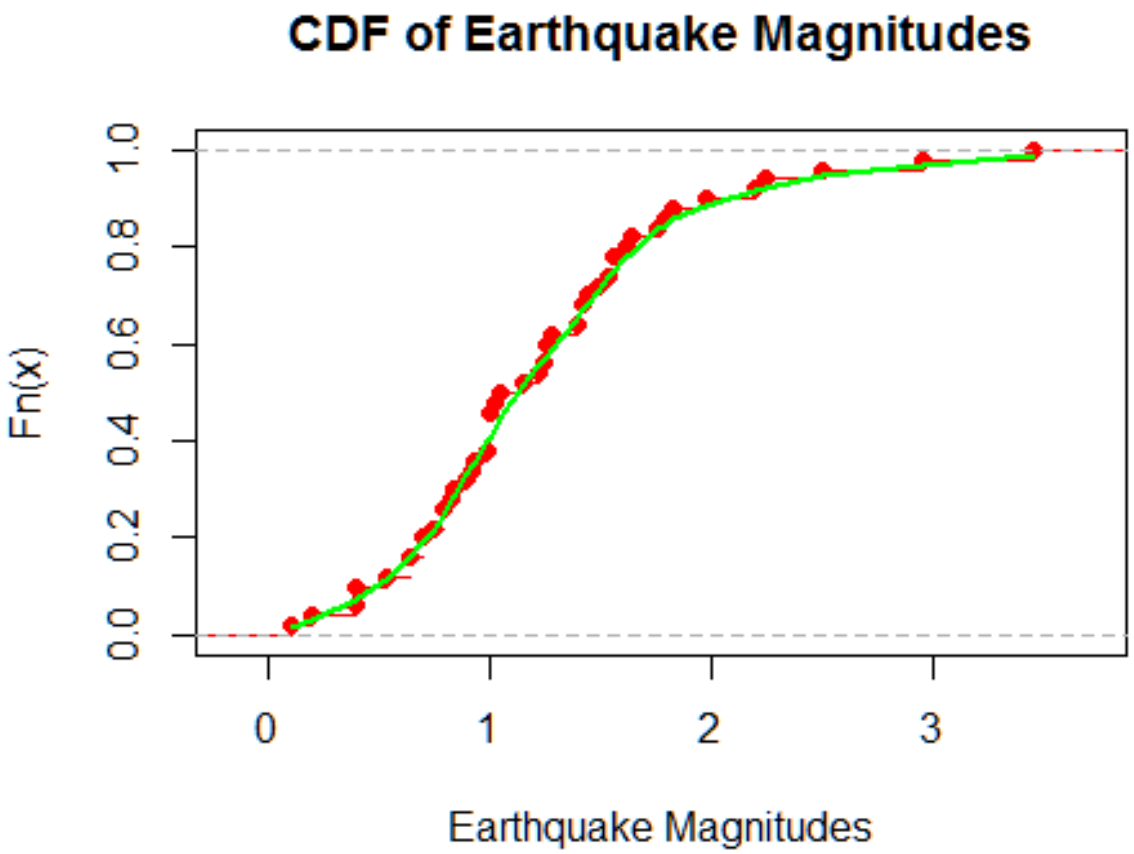
Use with Cumulative Distribution Function

Another application is to superimpose an estimate of the cumulative distribution function on a plot of the empirical distribution function. There are various options which must be specified in the kCDF() function. Checking the structure of the resulting object helps to clarify the variables used for display.

```
result <- kCDF(mag,from = 0,bw = "npdf",kernel=c("epanechnikov"),xgrid=mag)
str(result)

## List of 8
##  $ x        : num [1:50] 0.1 0.2 0.39 0.4 0.4 0.54 0.64 0.64 0.7 0.7 ...
##  $ Fhat     : num [1:50] 0.0147 0.0284 0.0704 0.0729 0.0729 ...
##  $ bw       : num 0.271
##  $ n        : int 50
##  $ call     : language kCDF(x = mag, bw = "npdf", kernel = c("epanechnikov"), xg
rid = mag,      from = 0)
##  $ data.name: chr "x"
##  $ data     : num [1:50] 0.7 0.74 0.64 0.39 0.7 2.2 1.98 0.64 1.22 0.2 ...
##  $ has.na   : logi FALSE
##  - attr(*, "class")= chr "CDF"

plot.ecdf(mag, col = "red", pch =16, xlab = "Earthquake Magnitudes",
          main = "CDF of Earthquake Magnitudes"          )
lines(result$x,result$Fhat, col = "green", lwd = 2)
```


CDF of Earthquake Magnitudes
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
Earthquake Magnitudes