677 Project: Discerning Wet and Dry Years

Will Zen

5/11/2022

packages; read data

```
library(fitdistrplus)
```

```
## Warning: package 'fitdistrplus' was built under R version 4.1.2
## Loading required package: MASS
## Loading required package: survival
```

```
library(readxl)
data <- read_excel("Illinois_rain_1960-1964.xlsx")
```

Put data in vector form for fitting

```
data_vec <- c(na.omit(data$`1960`),na.omit(data$`1961`),na.omit(data$`1962`),
              na.omit(data$`1963`),na.omit(data$`1964`))

sixty <- sort(na.omit(data$`1960`), decreasing = TRUE)
sixty1 <- sort(na.omit(data$`1961`), decreasing = TRUE)
sixty2 <- sort(na.omit(data$`1962`), decreasing = TRUE)
sixty3 <- sort(na.omit(data$`1963`), decreasing = TRUE)
sixty4 <- sort(na.omit(data$`1964`), decreasing = TRUE)
```

Fit gamma to each year. Specifying MLE is unnecessary.

```
fit <- fitdist(data_vec, 'gamma')
fit0 <- fitdist(sixty, distr = 'gamma', method = 'mle')
fit1 <- fitdist(sixty1, distr = 'gamma', method = 'mle')
fit2 <- fitdist(sixty2, distr = 'gamma', method = 'mle')
fit3 <- fitdist(sixty3, distr = 'gamma', method = 'mle')
fit4<- fitdist(sixty4, distr = 'gamma', method = 'mle')
```

This table helps a little maybe.

```
tab <- matrix(c(fit$estimate, fit0$estimate, fit1$estimate,
                fit2$estimate, fit3$estimate, fit4$estimate),
                ncol=2, byrow=TRUE)
colnames(tab) <- c('shape','rate')
rownames(tab) <- c('total','60','61','62','63','64')
tab <- as.table(tab)
tab
```

```
##           shape      rate
## total 0.4408386 1.9648409
## 60    0.3542986 1.6081421
## 61    0.5783901 2.1037195
```

```
## 62    0.4130575 2.2357824
## 63    0.5283565 2.0131998
## 64    0.4454876 2.3806945
```

Skewness = 2/sqrt(shape). Some of the shapes are slightly smaller, which indicates more skew.

I did not have to calculate the joint likelihood by multiplying f(x=x1) by f(x=x2) etc. I did not have to find the MLE of the rate in terms of the shape by setting the derivative to 0. I did not have to plug the MLE into the joint likelihood. Wikipedia just gives the formula of the profile log-likelihood. I applied that formula to the data from every year.

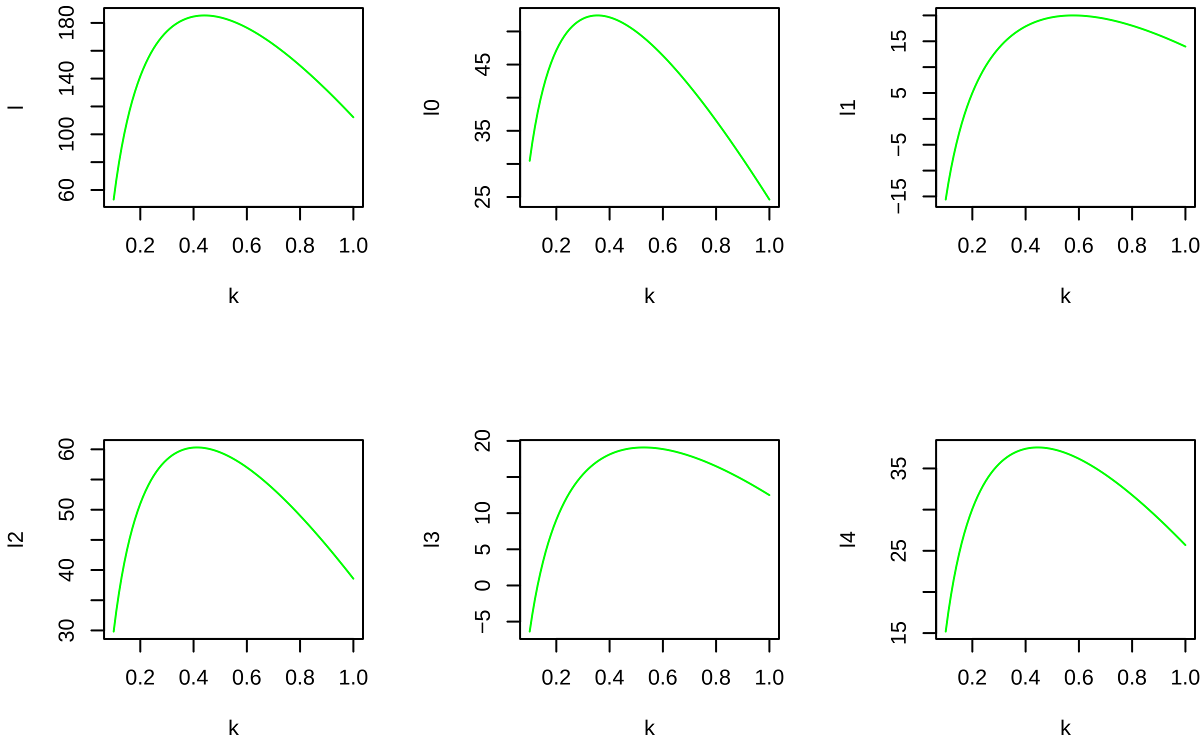
These log-likelihood vectors can be verified by the models above, each of which have the maximum log-likelihood as an argument.

```
k <- seq(.1, 1, .01) # k is shape parameter

# https://en.wikipedia.org/wiki/Gamma_distribution#Maximum_likelihood_estimation
# equation 4
# L <- prod(data_vec)^(k-1) / exp(N*k) / (sum(data_vec)/k/N)^(N*k) / gamma(k)^N

N <- length(data_vec)
l <- (k-1)*sum(log(data_vec)) - N*k - N*k*log(sum(data_vec)/k/N) - N*log(gamma(k))
N <- length(sixty)
l0 <- (k-1)*sum(log(sixty)) - N*k - N*k*log(sum(sixty)/k/N) - N*log(gamma(k))
N <- length(sixty1)
l1 <- (k-1)*sum(log(sixty1)) - N*k - N*k*log(sum(sixty1)/k/N) - N*log(gamma(k))
N <- length(sixty2)
l2 <- (k-1)*sum(log(sixty2)) - N*k - N*k*log(sum(sixty2)/k/N) - N*log(gamma(k))
N <- length(sixty3)
l3 <- (k-1)*sum(log(sixty3)) - N*k - N*k*log(sum(sixty3)/k/N) - N*log(gamma(k))
N <- length(sixty4)
l4 <- (k-1)*sum(log(sixty4)) - N*k - N*k*log(sum(sixty4)/k/N) - N*log(gamma(k))

{
par(mfrow = c(2,3))
plot(k,l,type="l",col="green")
plot(k,l0,type="l",col="green")
plot(k,l1,type='l',col="green")
plot(k,l2,type='l',col="green")
plot(k,l3,type='l',col="green")
plot(k,l4,type='l',col="green")
}
```


I
I0
I1
I2
I3
I4
k
0.2
0.4
0.6
0.8
1.0
60
100
140
180
25
35
45
−15
−5
5
15
30
40
50
60
−5
0
5
10
20
15
25
35