ECS 132 Term Project

Steven Alvarado, Russell Chien, and Ruth Hailu

University of California, Davis

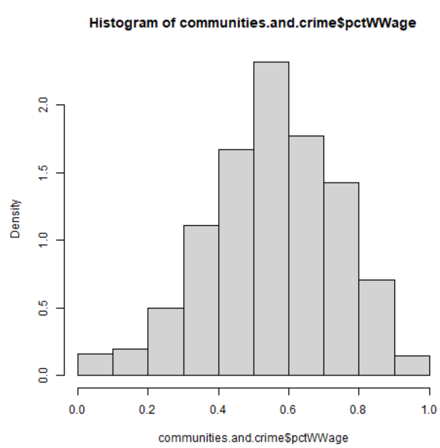
June 2023

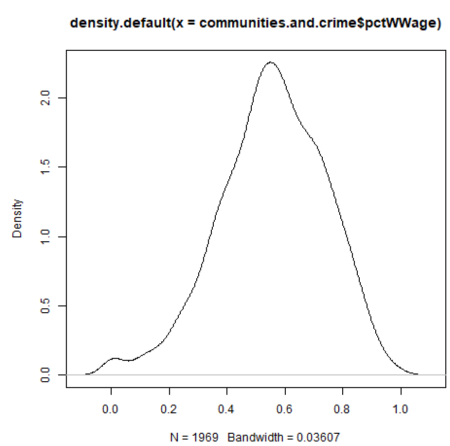
Chapter 1

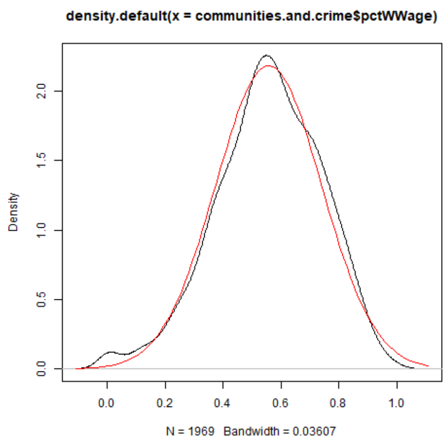
The Normal Family

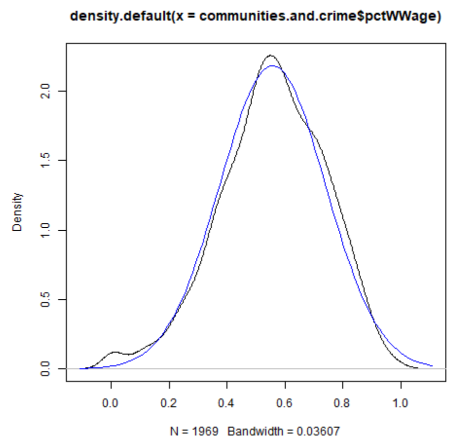
1.1 Communities and Crime: pctWWage

Our group observed that the variable **pctWWage** of the Communities and Crime dataset seemed well-approximated by the normal family of continuous distributions. According to the UCI Machine Learning Repository, **pctWWage** is described as the percentage of households within the United States with wage or salary income in 1989.






density.default(x = communities.and.crime$pctWWage)
Density
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
N = 1969 Bandwidth = 0.03607

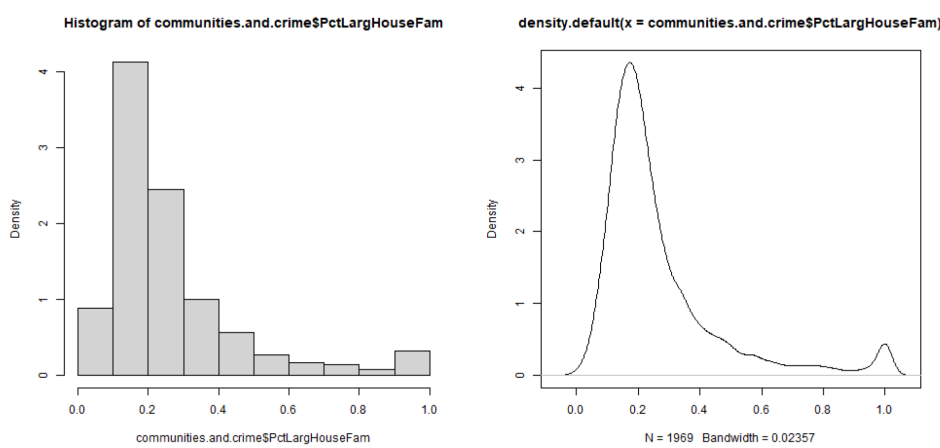

density.default(x = communities.and.crime$pctWWage)
Density
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
N = 1969 Bandwidth = 0.03607

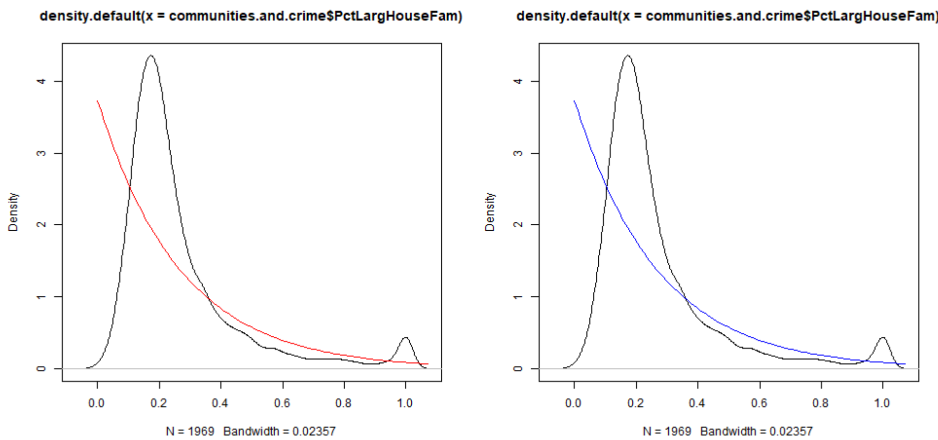
Chapter 2

The Exponential Family

2.1 Communities and Crime: PctLargHouse-Fam

For the exponential family of continuous distributions, we observed that the variable **PctLargHouseFam** was a suitable approximation. According to the UCI Machine Learning Repository, **PctLargHouseFam** is described as the percentage of family households with six or more family members.



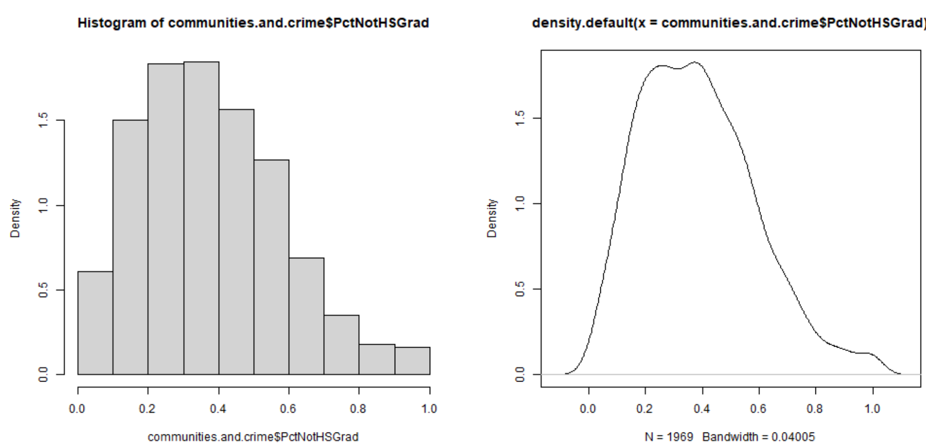

density.default(x = communities.and.crime$PctLargHouseFam)
Density
N = 1969 Bandwidth = 0.02357
density.default(x = communities.and.crime$PctLargHouseFam)
Density
N = 1969 Bandwidth = 0.02357

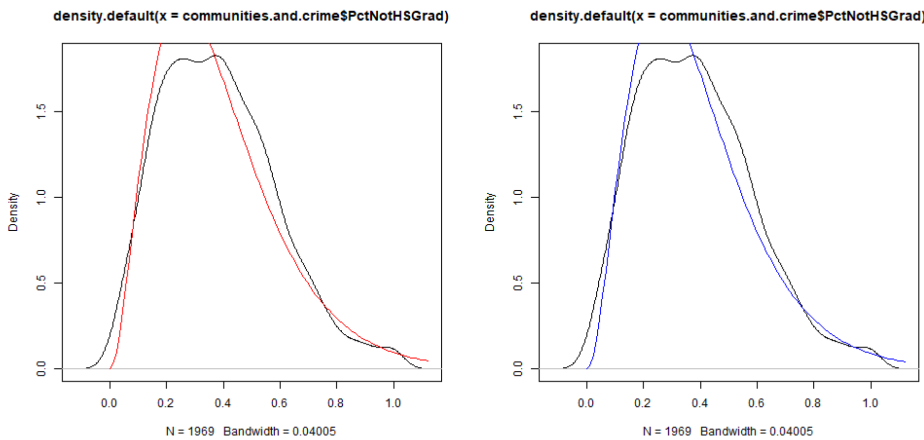
Chapter 3

The Gamma Family

3.1 Communities and Crime: PctNotHsGrad

We observed that the variable **PctNotHsGrad** of the Communities and Crime dataset seemed well-approximated by the gamma family of continuous distributions. According to the UCI Machine Learning Repository, **PctNotHsGrad** is described as the percentage of people 25 and over that are not high school graduates.




density.default(x = communities.and.crime$PctNotHSGrad)
Density
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
N = 1969 Bandwidth = 0.04005
density.default(x = communities.and.crime$PctNotHSGrad)
Density
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
N = 1969 Bandwidth = 0.04005

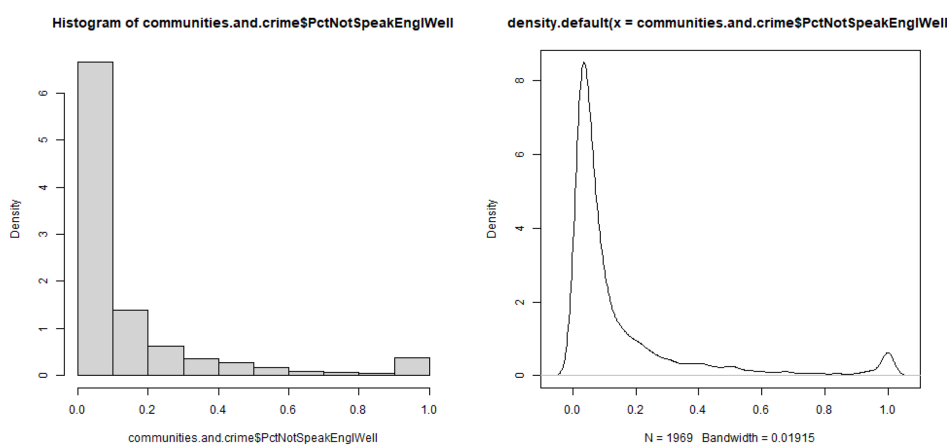
Chapter 4

The Beta Family

4.1 Communities and Crime: PctNotSpeakEnglWell

For the beta family of continuous distributions, we observed that the variable **PctNotSpeakEnglWell** was a suitable approximation. According to the UCI Machine Learning Repository, **PctNotSpeakEnglWell** is described as the percentage of people who do not speak English well.

4.2 Histogram and Density



4.3 MLE and MM

To find the MLE of the beta family, we first had to scale our data so it was within the support of $(0, 1)$:

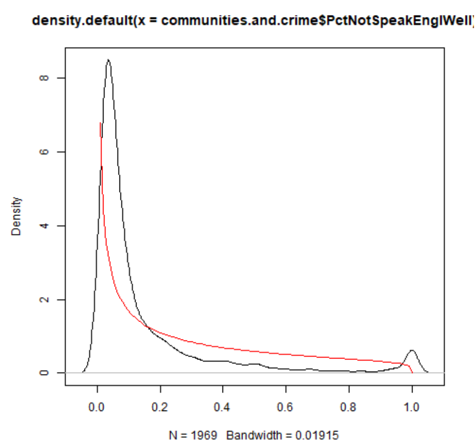
```
x[which(x == 0)] <- 0.0001
x[which(x == 1)] <- 0.9999
```

We then found the log likelihood function:

$$L(\alpha, \beta) = n \log\left(\Gamma(\alpha + \beta)\right) - n \log\left(\Gamma(\alpha)\right) - n \log\left(\Gamma(\beta)\right) + \\ (\alpha - 1) \sum (\log(x)) + (\beta - 1) \sum (\log(1 - x)) \quad (4.1)$$

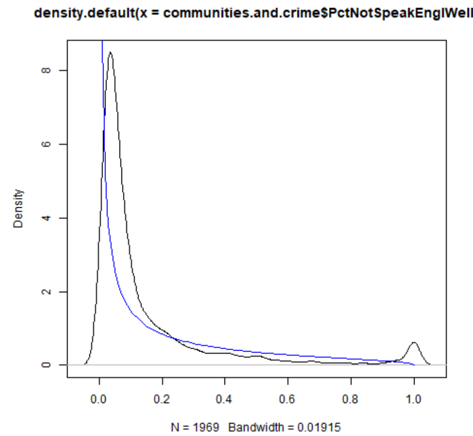
To find the MLE, we use R's built in mle() function with the negative value of our 4.1 log likelihood function.

```
z <- mle(minuslogl = ll, start = c(list(alpha = 1), list(beta
    = 1)))
```



To find the MM of the beta family, we used the following function to estimate the α and β values:

```
mm <- function(x) {
  mu <- mean(x)
  var <- var(x)
  alpha <- mu * (mu * (1 - mu) / var - 1)
  beta <- (1 - mu) * (mu * (1 - mu) / var - 1)
  return(c(alpha, beta))
}
```



4.4 Conclusion