Lab 4 Solutions

Shahab Geravesh, UC Riverside

Part I

1. Load the file [http://faculty.ucr.edu/~jflegal/206/mlb-obp.csv] into a variable of your choice in R. How many players have been included? What is the minimum number of plate appearances required to appear on this list? Who had the most plate appearances? What are the minimum, maximum and mean OBP?

```
a <- read.csv("http://faculty.ucr.edu/~jflegal/206/mlb-obp.csv", as.is = TRUE)
head(a)
```

```
##   X              Name  PA   OBP
## 1 1        Mike Trout 705 0.377
## 2 2 Andrew McCutchen 648 0.410
## 3 3 Michael Brantley 676 0.385
## 4 4   Anthony Rendon 683 0.351
## 5 5       Alex Gordon 643 0.351
## 6 6   Josh Donaldson 695 0.342
```

```
num.players <- nrow(a)
num.players
```

```
## [1] 441
```

```
min.PA <- min(a$PA)
min.PA
```

```
## [1] 103
```

```
a$Name[which.max(a$PA)]
```

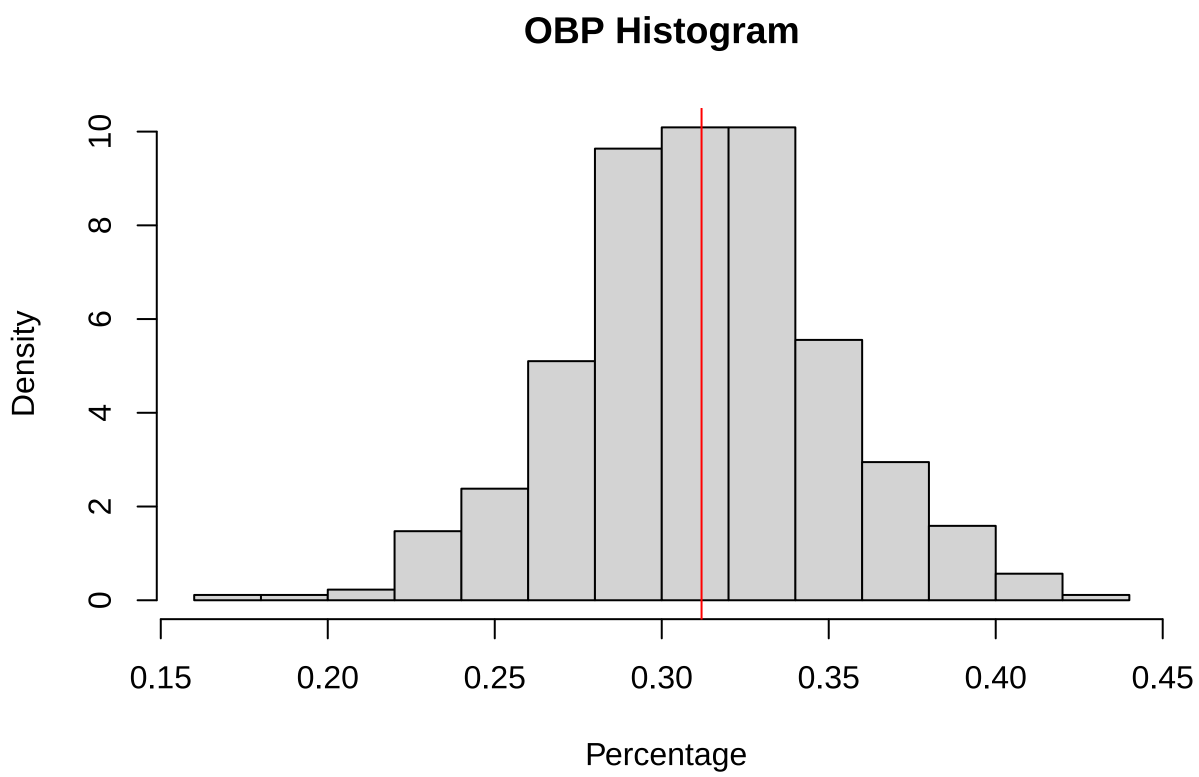
```
## [1] "Ian Kinsler"
```

```
summary(a$OBP)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1680  0.2870  0.3130  0.3119  0.3370  0.4320
```

2. Plot the OBP data as a histogram with the option `probability=TRUE`. Add a vertical line for the mean of the distribution. Does the mean coincide with the mode of the distribution?

```
hist(a$OBP, probability = TRUE, xlab = " Percentage", main = "OBP Histogram")
abline(v = mean(a$OBP), col = "red")
```



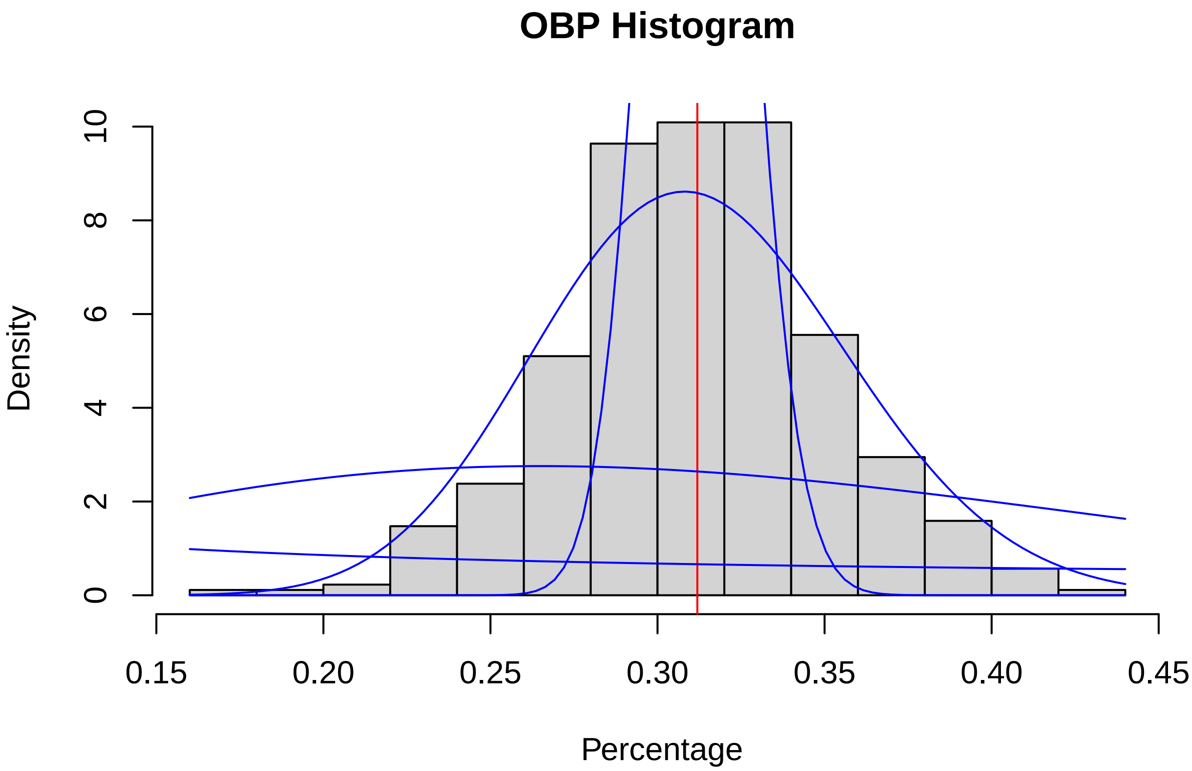
3. Eyeball fit. Add a `curve()` to the plot using the density function `dbeta()`. Pick parameters α and β that match the mean of the distribution but where their sum equals 1. Add three more `curve()`s to this plot where the sum of these parameters equals 10, 100 and 1000 respectively. Which of these is closest to the observed distribution?

```
mean <- mean(a$OBP)

Parameters  <- function(mean, sum) {
  a <- mean*sum
  return(c(alpha = a, beta = sum - a))
}

hist(a$OBP, probability = TRUE, xlab = " Percentage", main = "OBP Histogram")
abline(v = mean(a$OBP), col = "red")

curve(dbeta(x, shape1 = Parameters (mean, sum = 1)[1], shape2 = Parameters (mean, sum = 1)[2]), add = TR
curve(dbeta(x, shape1 = Parameters (mean, sum = 10)[1], shape2 = Parameters (mean, sum = 10)[2]), add =
curve(dbeta(x, shape1 = Parameters (mean, sum = 100)[1], shape2 = Parameters (mean, sum = 100)[2]), add
curve(dbeta(x, shape1 = Parameters (mean, sum = 1000)[1], shape2 = Parameters (mean, sum = 1000)[2]), ad
```

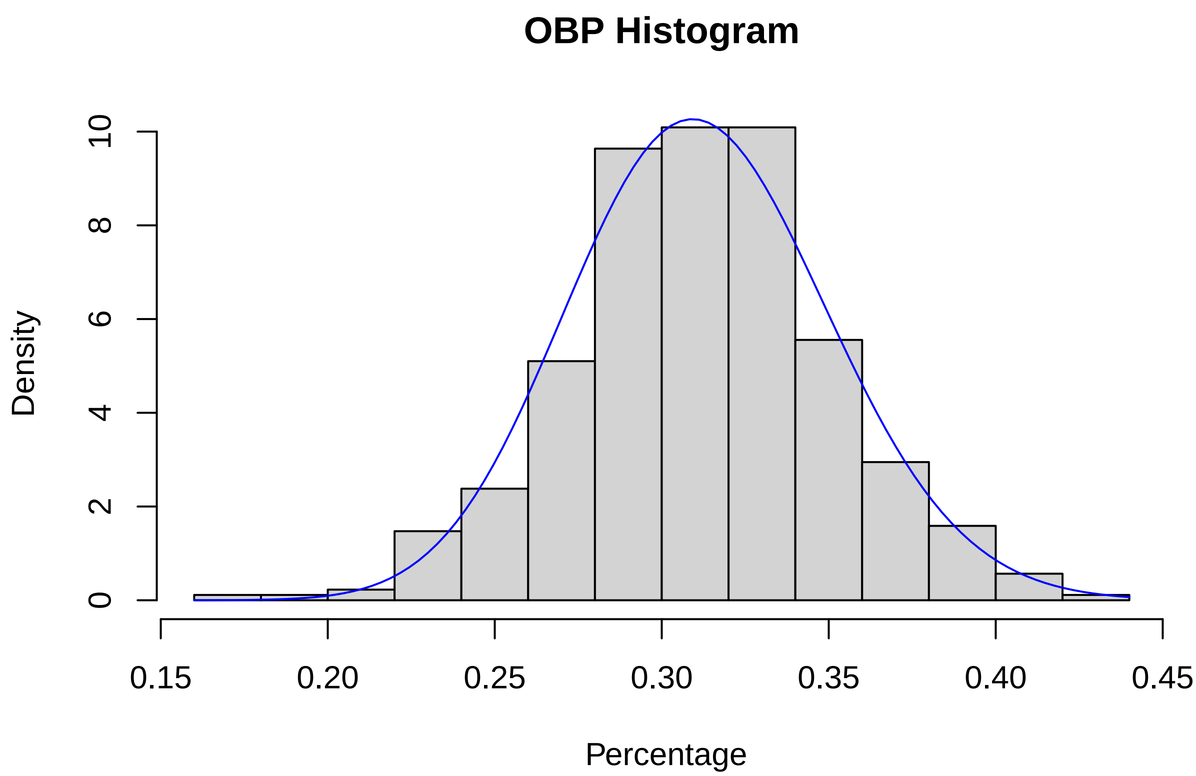


Part II

4. Method of moments fit. Find the calculation for the parameters from the mean and variance from [http://en.wikipedia.org/wiki/Beta_distribution] and solve forα and β. Create a new density histogram and add this 'curve()' to the plot. How does it agree with the data?

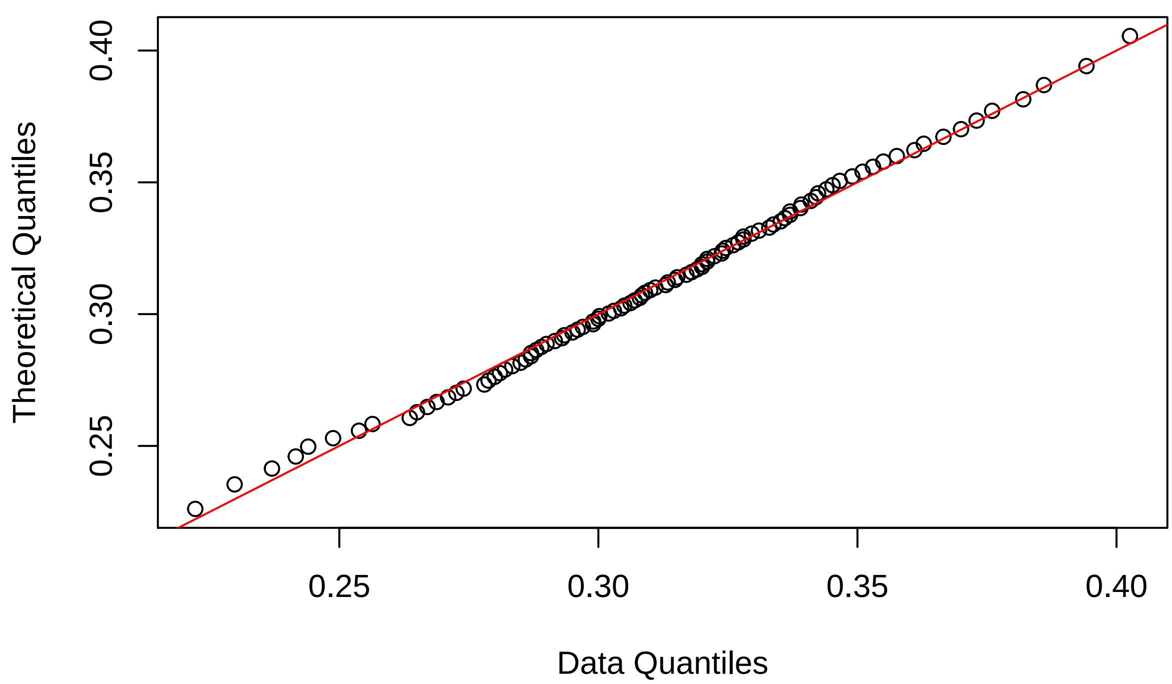
```
beta.MM <- function(data) {
  m <- mean(data)
  v <- var(data)
  return(c(alpha = (m/v)*(m*(1-m) - v), beta = ((1-m)/v)*(m*(1-m) - v)))
}
MM <- beta.MM(a$OBP)


hist(a$OBP, probability = TRUE, xlab = " Percentage", main = "OBP Histogram")
curve(dbeta(x, shape1 = MM[1], shape2 = MM[2]), add = TRUE, col = "blue")
```



5. Calibration. Find the 99 percentiles of the actual distribution of the data using the `quantile()` function using `quantile(a$OBP, probs = seq(1, 99)/100)` and plot them against the 99 percentiles of the beta distribution you just fit using `qbeta()`. How does the fit appear to you?

```
data.quant   <- quantile(a$OBP, probs = seq(1, 99)/100)
theory.quant <- qbeta(seq(1, 99)/100, shape1 = MM[1], shape2 = MM[2])
plot(data.quant, theory.quant, xlab = "Data Quantiles", ylab = "Theoretical Quantiles")
abline(a = 0, b = 1, col = "red")
```



6. Create a function for the log-likelihood of the distribution that calculates -sum(dbeta(your.data.here, your.alpha, your.beta, log=TRUE)) and has one argument p=c(your.alpha, your.beta). Use nlm() to find the minimum of the negative of the log-likelihood. Take the MOM fit for your starting position. How do these values compare?

```
logLikelihood <- function(Parameters ) {
  return(-sum(dbeta(a$OBP, Parameters [1], Parameters [2], log = TRUE)))
}

MLE <- nlm(logLikelihood, MM)$est
MLE
```

```
## [1] 43.73915 96.49892
```

```
MM
```

```
##    alpha     beta
## 44.31690 97.76163
```