CALVIN GUAN

MA 578

CLASS PAPER

NONPARAMETRIC KERNEL DENSITY ESTIMATION USING BAYESIAN SMOOTHING PARAMETER

Introduction

Bayesian Statistics is a powerful mathematically driven school of methods that incorporates both beliefs and data to make inferences. However, traditional Bayesian methods requires the likelihood of the data to belong to a parametric family. Oftentimes actual data are not distributed astraditional families of distributions, assuming so would lead to erroneous inferences. In such cases non-parametric methods are employed to create a data-driven density estimation. In this project, I will fit a kernel density estimator for the NBA data provided by the Professor. In addition, I will employ a Bayesian model to the smoothing parameter (a.k.a. bandwidth) to obtain the optimal bandwidth. This project uses the ideas outlined in the paper "Bayesian Approach to the Choice of Smoothing Parameter in Kernel Density Estimation" by Gangopadhyay and Cheung.

Kernel Density Estimation

Say we have samples from an unknown distribution $f(x)$, and we do not assume that $f(x)$belongs to any parametric family, we would derive an estimate $\hat{f}(x)$ from our data. One density estimation method is the Kernel Density Estimation, where we assume the data follows a distribution specified by the Kernel function in a small neighborhood. The length of the neighborhood is specified as a smoothing parameter (h). Kernel functions are non-negative symmetric functions around 0 that also integrates to 1 over its support.Let K(.) be the Kernel function, h the bandwidth, **X** the observed data, and **x** an evaluation point, the estimated Kernel Density is

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{X_i - x}{h})$$

Bayesian Approach to Bandwidth Selection

We can let h be a parameter in the estimated density, specify a prior distribution for h, and obtain the posterior of h. Because h is not a true parameter of the underlying density, we instead truncate the underlying density as a convolution of the density and the Kernel around a neighborhood of the evaluation point **x**. In the truncated density, h becomes the scale parameter. The advantage of bandwidth selection method is that the posterior distribution of h depends on the evaluation point **x**, make it more flexible to incorporate local features of the data. The posterior distribution of h is given by:

$$\pi(h|x) = \frac{f_h(x)\pi(h)}{\int f_h(x)\pi(h)dh}$$

Since $f_h(x)$ is unknown, we estimate it by $\hat{f}_h(x)$ and get the estimated posterior distribution:

$$\hat{\pi}(h|\boldsymbol{X}, x) = \frac{\hat{f}_h(x)\pi(h)}{\int \hat{f}_h(x)\pi(h)dh}$$

Where **X** = (X_1, …, X_n) is the observed data vector and $\hat{f}_h(x)$ is the estimated Kernel density. While usually the closed form for the posterior does not exist, it exists if we set the Kernel function to be the density function of a standard normal distribution. In this case, h is the standard deviation parameter of the data's local distribution. Thus, the conjugate prior follows an inverse gamma distribution. If $\pi(h) \sim IG(\alpha, \beta)$, the estimated posterior distribution of h can be computed:

$$\hat{\pi}(h|\boldsymbol{X}, x) = \frac{h^{-(2\alpha+2)} \sum_{i=1}^{n} \exp\{-\frac{(X_i-x)^2+2(\beta^{-1})}{2h^2}\}}{\frac{\Gamma(\alpha+\frac{1}{2})}{2} \sum_{i=1}^{n} \{\frac{(X_i-x)^2}{2} + \beta^{-1}\}^{-\alpha-\frac{1}{2}}}$$

The optimal local bandwidth (h^*) is then the estimated posterior mean:

$$h^*(x|\boldsymbol{X}, \alpha, \beta) = \frac{\Gamma(\alpha)}{\sqrt{2\beta}\Gamma(\alpha+\frac{1}{2})} \frac{\sum_{i=1}^{n} \{\beta(X_i - x)^2 + 2\}^{-\alpha}}{\sum_{i=1}^{n} \{\beta(X_i - x)^2 + 2\}^{-\alpha-\frac{1}{2}}}$$

<u>Applications</u>

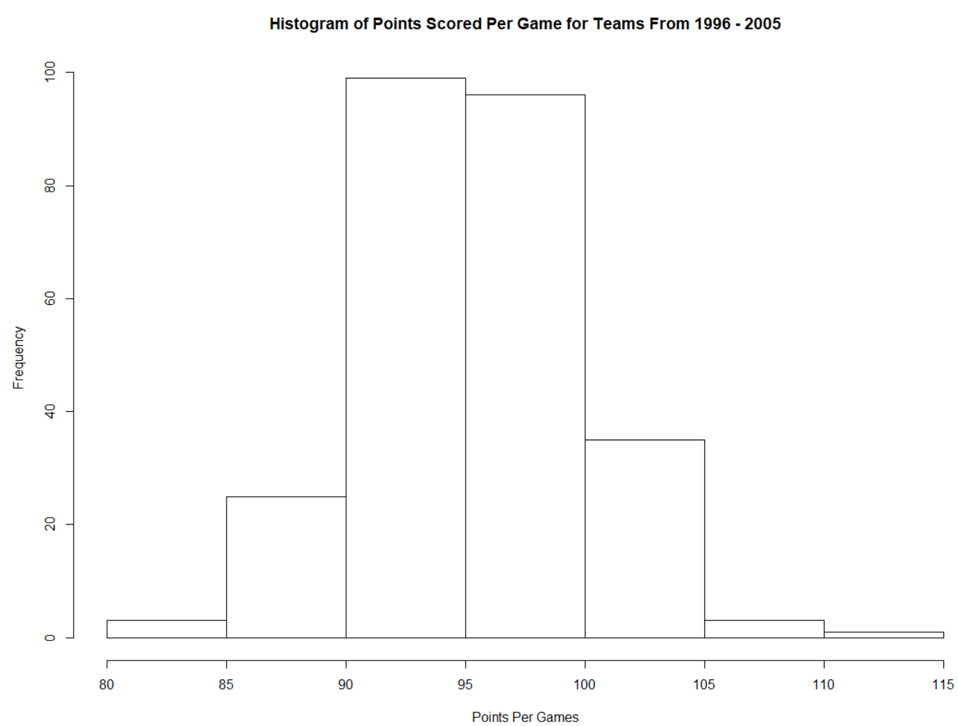
Bayesian bandwidth depends on the chosen prior parameters, the observed data, as well as the evaluation point. As such, it is a very flexible variable optimal bandwidth method. The prior parameters can be chosen to give us the exact amount of smoothing we need. The method will automatically look at the concentration of the observed data around the evaluation point. In a dense area of data, the bandwidth will be small to capture standout features. On the other hand, the bandwidth will be large in areas with sparse data to reduce variance and avoid over-fitting.

NBA Data

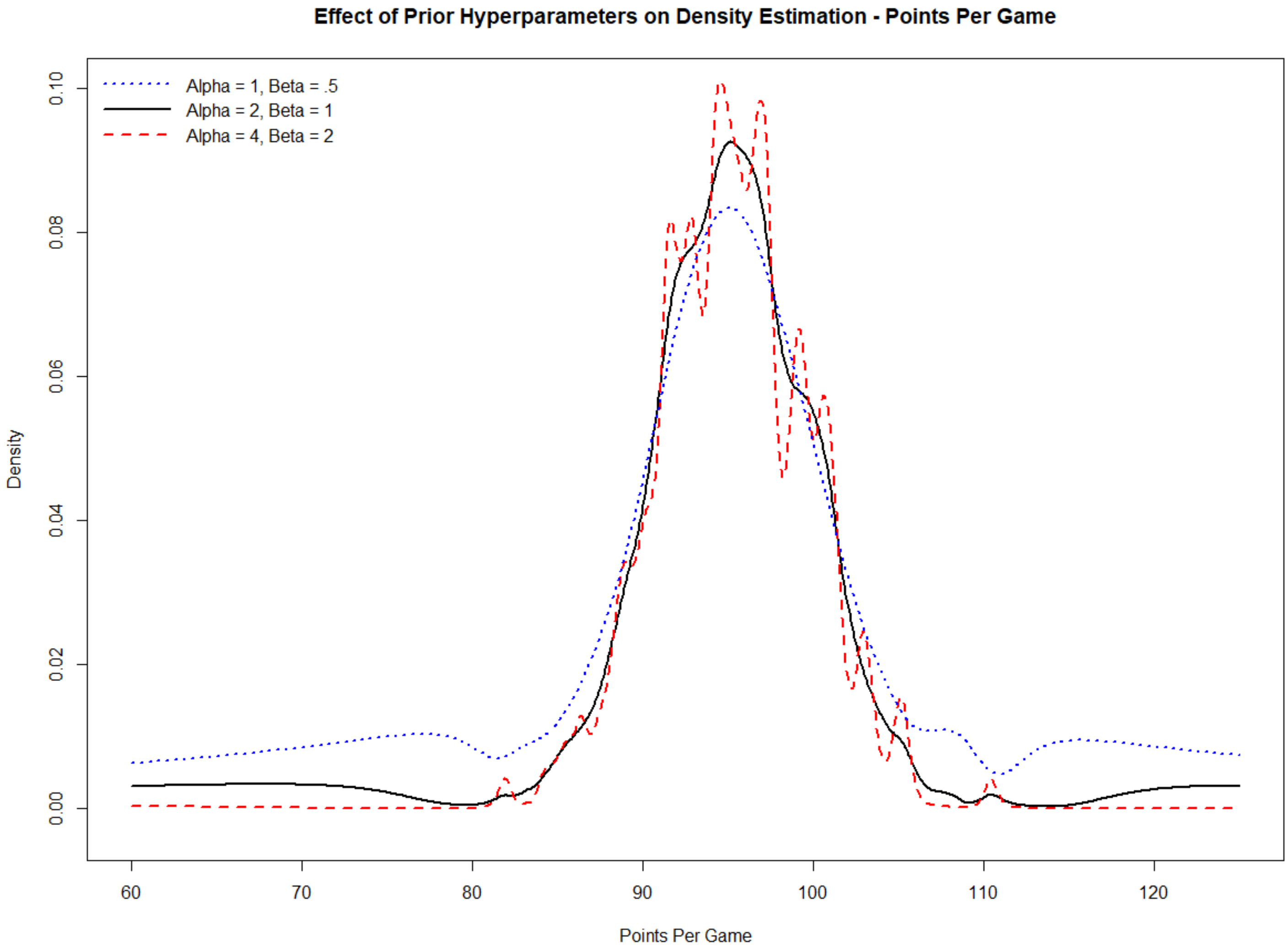
The NBA data kindly provided by the professor contains information on NBA teams' performance from year 1946 to 2005. We are interested in how some team statistics change throughout the years. Specifically we want to estimate the distribution of three stats: Average points scored per game, proportion of field goals made, and proportion of free throws made. The observation level is by teams, and we group the data to six 10-year periods per group, the first group is from 1946 to 1955, and the last group from 1996 to 2005.

Points Per Game

The first variable I looked at is Points Per Game. It would be interesting to see how the distribution of the points scored evolved from NBA's inception to the most recent season in the data, year 2005. I first used the most recent 10 years' data to come up with prior parameters that results in a good fit, defined as neither over-smoothed nor under-smoothed. Perhaps the most used density estimation is the histogram, thus I first start with a histogram to take a preliminary look at the data:

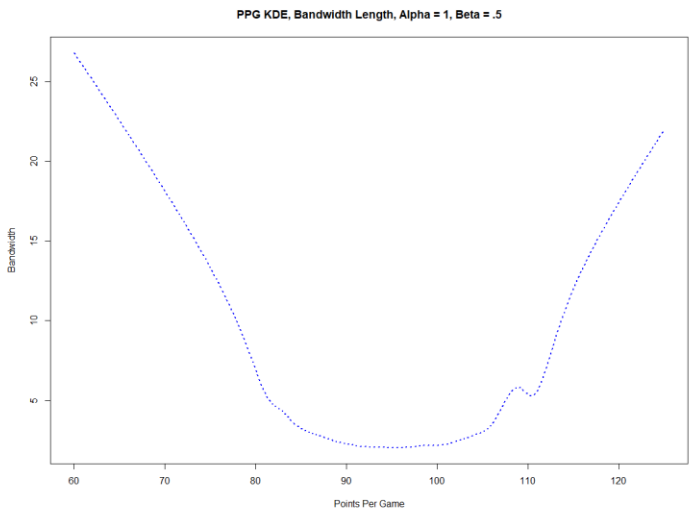


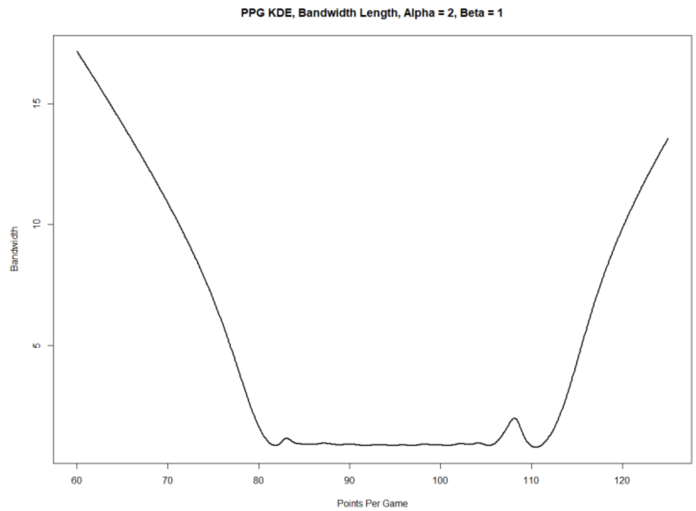
While the histogram can visualize the center and general spread of the data, it is far from able to describe the shape of the data. Instead, I fitted a Kernel Density Estimation, using a Gaussian Kernel and the Inverse-Gamma prior to the bandwidth. Three different hyper-parameters for the prior were used and compared, for simplicity, I fixed the ratio between α and β:

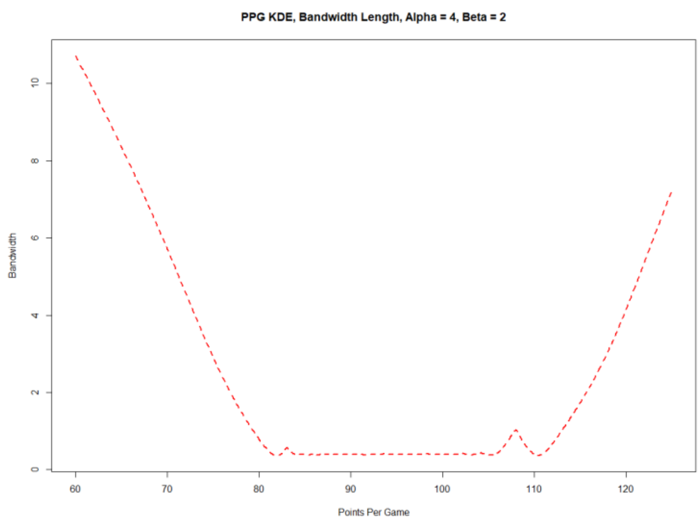


Clearly, the parameters (α = 4, β = 2) under-smoothed the data creating a sharp and rough density. On the other hand, the parameters (α = 1, β = .5) over-smoothed the data, resulting in a density very similar to a Gaussian density. I chose (α = 2, β = 1) as the prior parameters to analyze the data, for consistency, the same prior is used for all the 10 year periods.

To further look at how the prior parameters affect the posterior optimal bandwidth $h^{*}(x|\boldsymbol{X},\alpha,\beta)$, I plotted the bandwidth against **x** for each set of parameters:

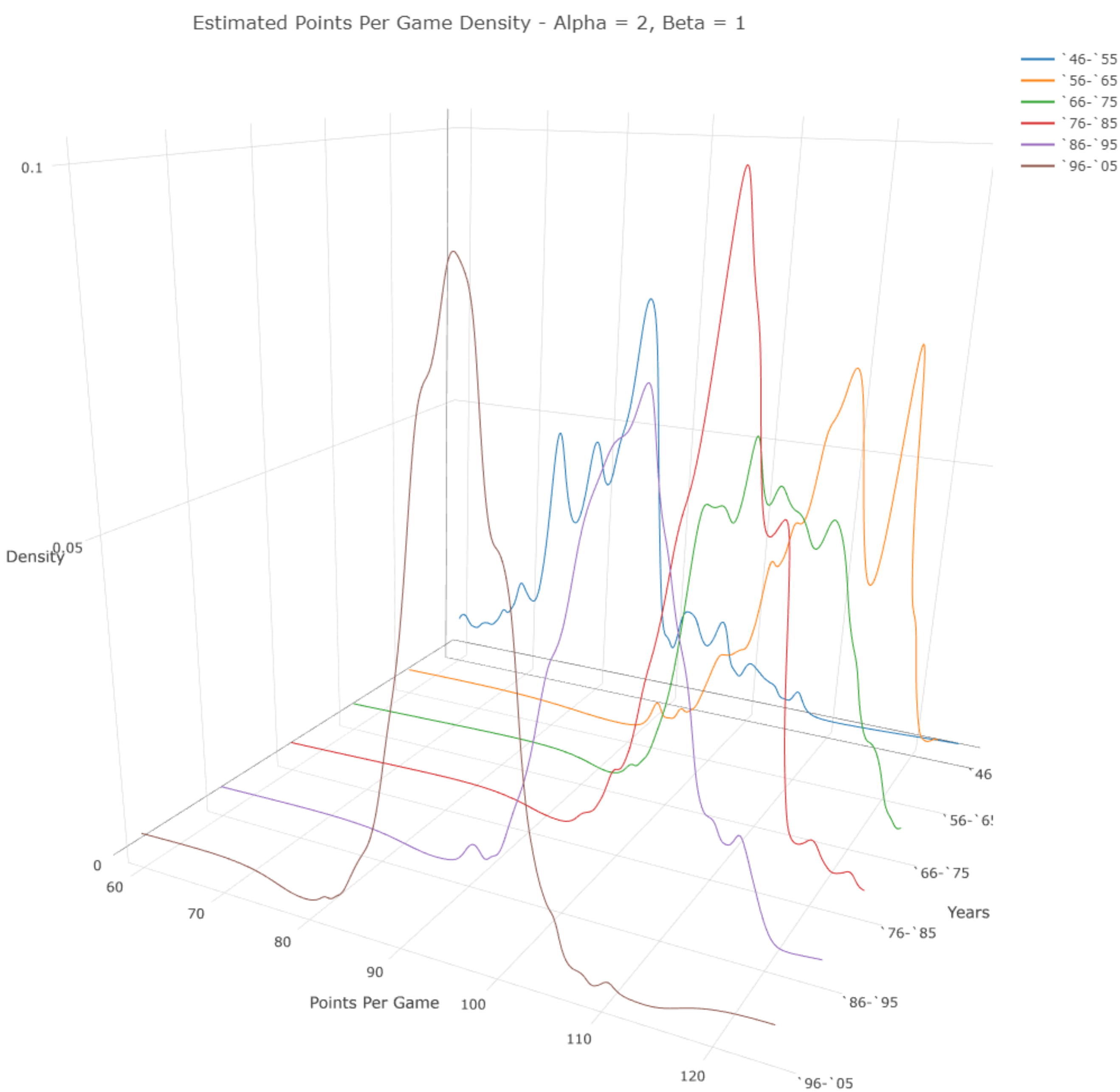






Clearly we notice a trend where the area with more data points has a small bandwidth, while those with fewer data points has a large bandwidth. In general, the bandwidth increases as α and β decreases. But what is perhaps more interesting, is that as α and β increases, the bandwidth becomes more sensitive to the data. Thus very small values of α and β will lead to a density that is highly influenced by the prior, and will most likely not be able to express the intricate features of the data.

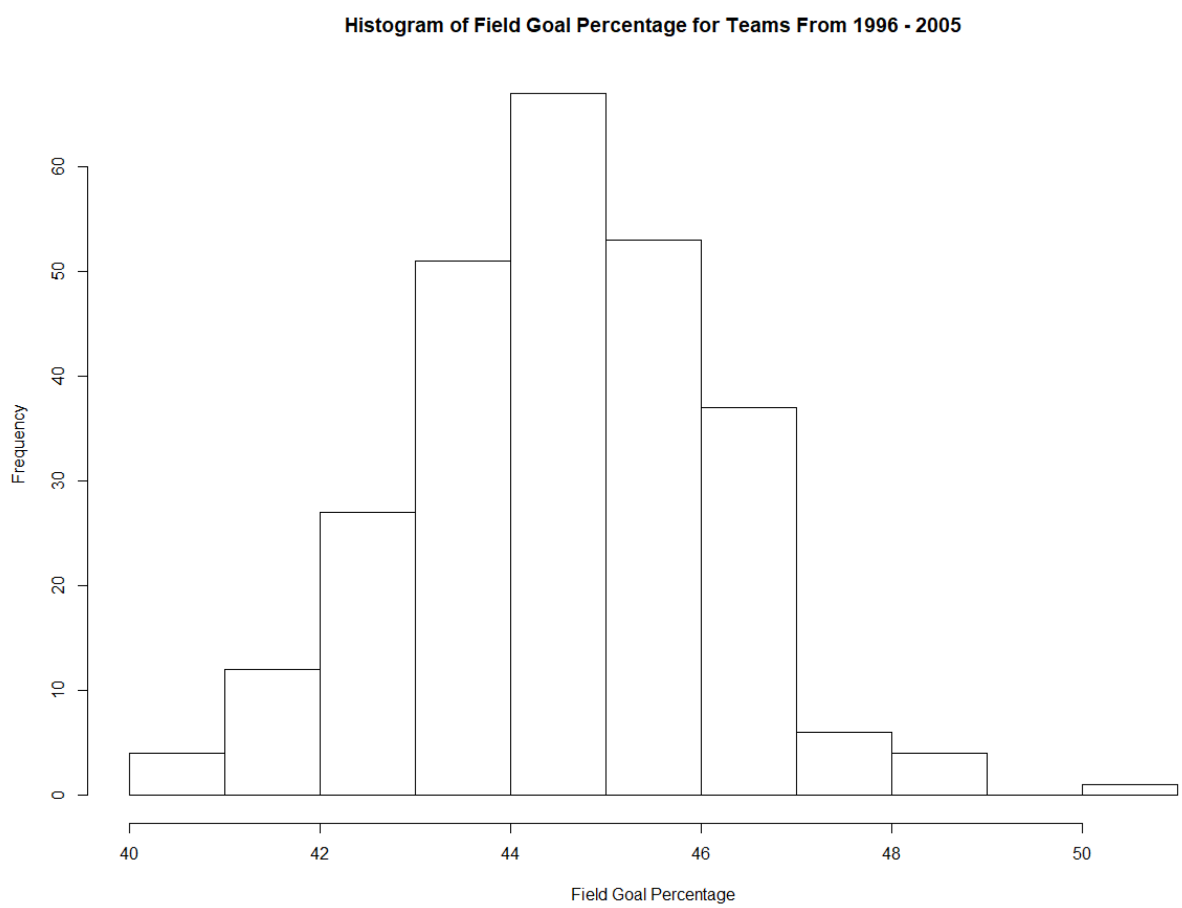
Having decided on the prior parameters, I ran the estimation for the six 10-year periods and plotted a 3D-plot:



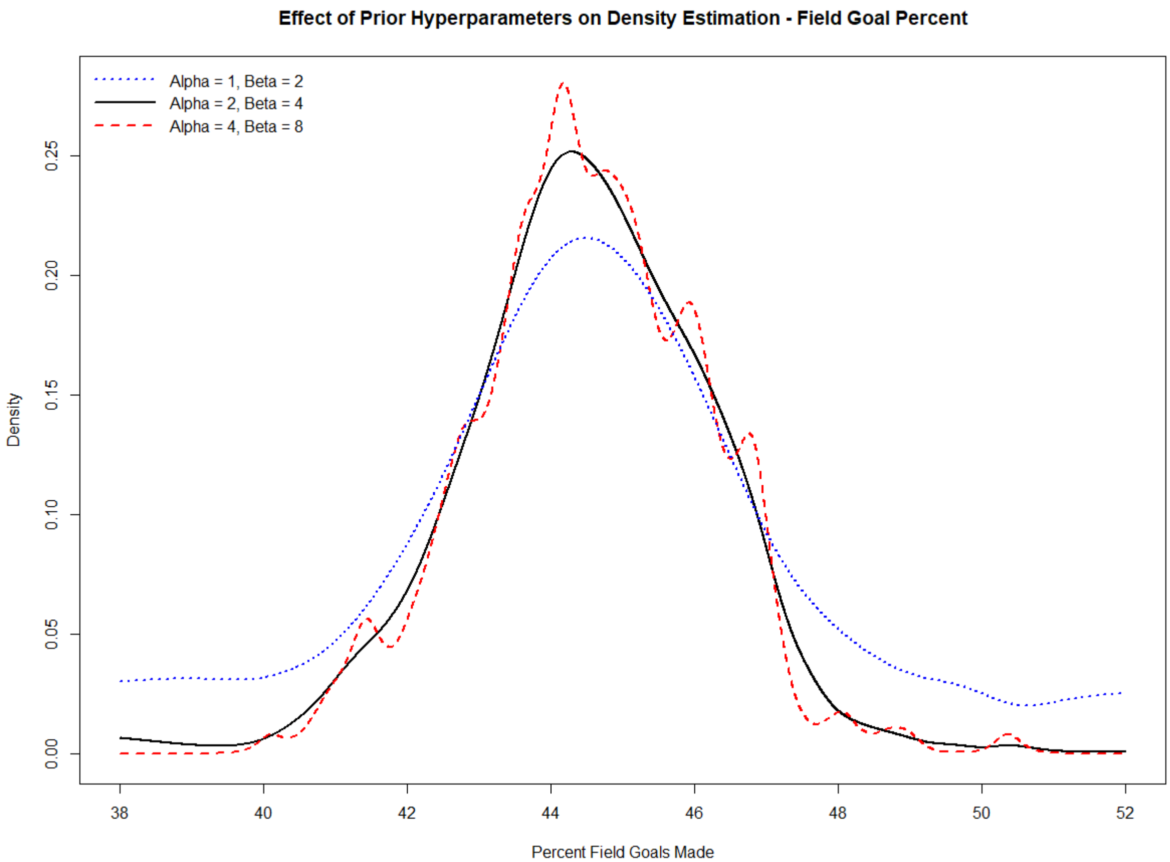
The results are very interesting: In the first 10 years, the distribution of points per page are to the left of the other years. The next 10 years shown a big increase in the points scored, and has a bi-modal distribution. Perhaps at that time period (1956-1965) there are two sets of teams: strong dominant teams and weak teams. The next 10 years have roughly the same range, but has a flat plateau-like mode. Suggesting that strong and teams still exist, but average teams are becoming more prominent. The next three decades (1976-2005) of data are unimodal, but we see a steady shift to the left in points scored. In these periods, average teams are the norm, with fewer strong and weak teams. We also see that teams are scoring many teams from the 70's to the 80's, indicating that offensive strategies are ahead of defensive strategies. But as time progresses, defensive strategies catches up, and the points scored per game steadily decreases. Note that the earliest 10 years (1946 – 1955) have fewer data points than the other periods, thus the bandwidth we chose under-smoothed the data, as larger bandwidths are needed to account for fewer data points. But for consistency I chose to leave the parameters as-is.

Field Goal Percentage

The next variable that I was interested in is the proportion of field goals were made, defined as number of field goals made divided by the number of field goals attempted and multiplied by 100. The main reason is to see how the accuracy of the shots changed over the years. Each data point identifies a team's field goal percentage in a season.

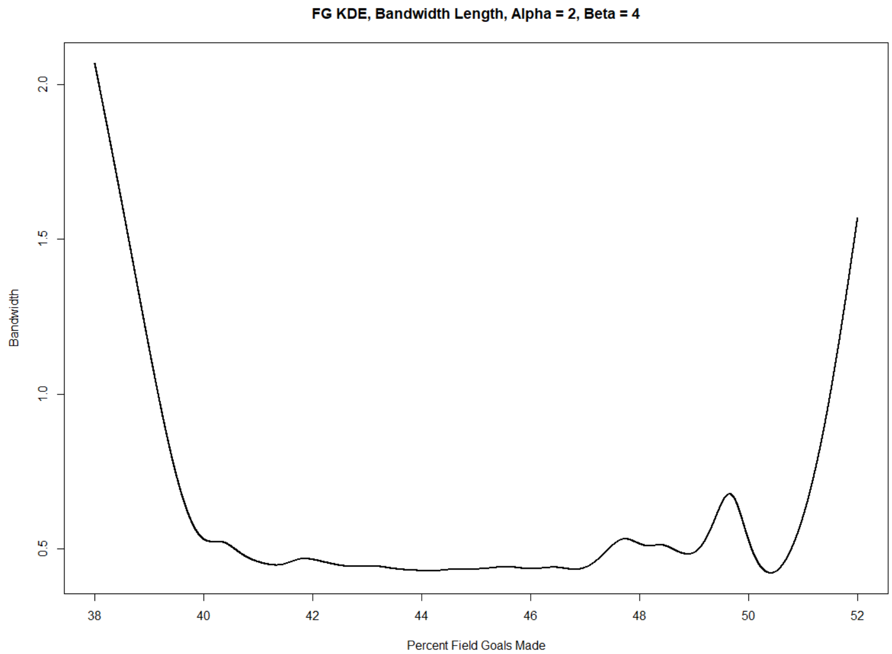


The preliminary histogram shows that in recent times the average percent of field goals made is around 44%, with an almost symmetric distribution. I use the same methods as points scored to estimate the density of percent field goals made:

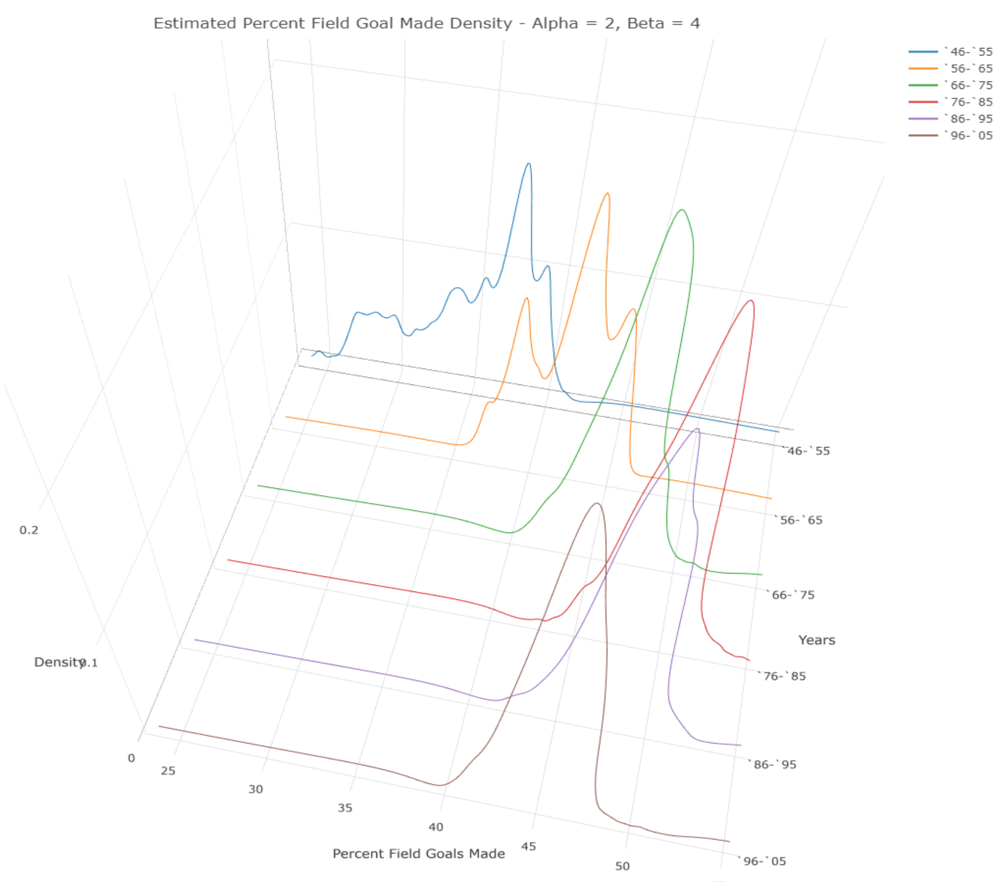


The above plot shows that ($\alpha = 4$, $\beta = 8$) under-smoothed the data, ($\alpha = 1$, $\beta = 2$) over-smoothed the data, and ($\alpha = 2$, $\beta = 4$) is just right.

The posterior bandwidth curve shows that smaller bandwidth are needed for the field goal data compared to the points scored data. This fact suggests that there is a much smaller variance for the percent of field goals made.



After running the process for the six time periods, we have the following plot:

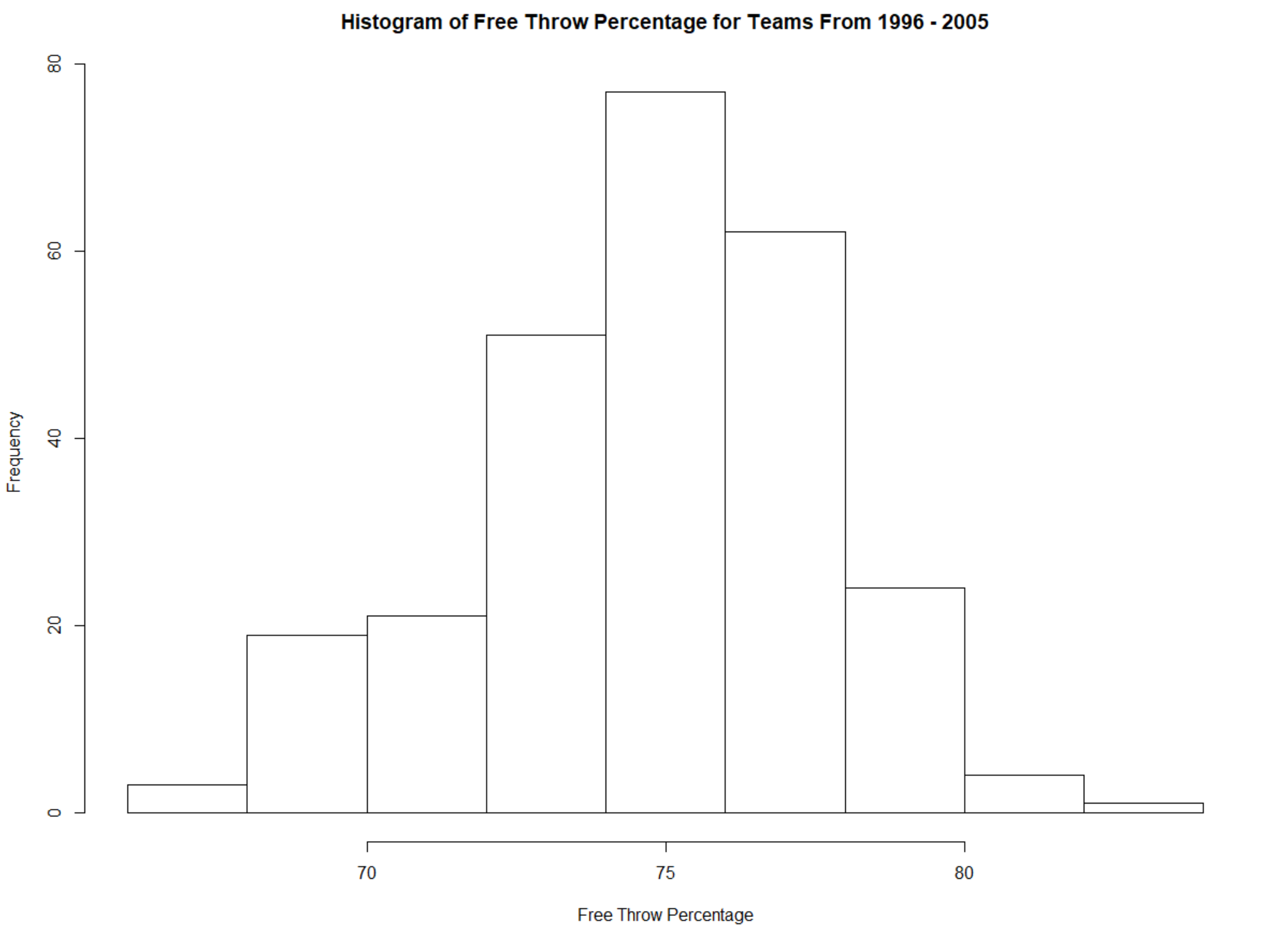


The distribution of the percent of field goals made looks similar as to that for the points scored. This intuitively makes sense: more points are scored as a higher percent of field goals are made. A notable difference is that percent field goals made has a very high peak for the last four decades, suggesting a very small variance for this distribution. The earliest two decades have a higher variance, this could be explained by the lack of data points for the first decade. The multiple modes in the distribution for the second decade matches with the bi-modal distribution in the points scored for the same time period.

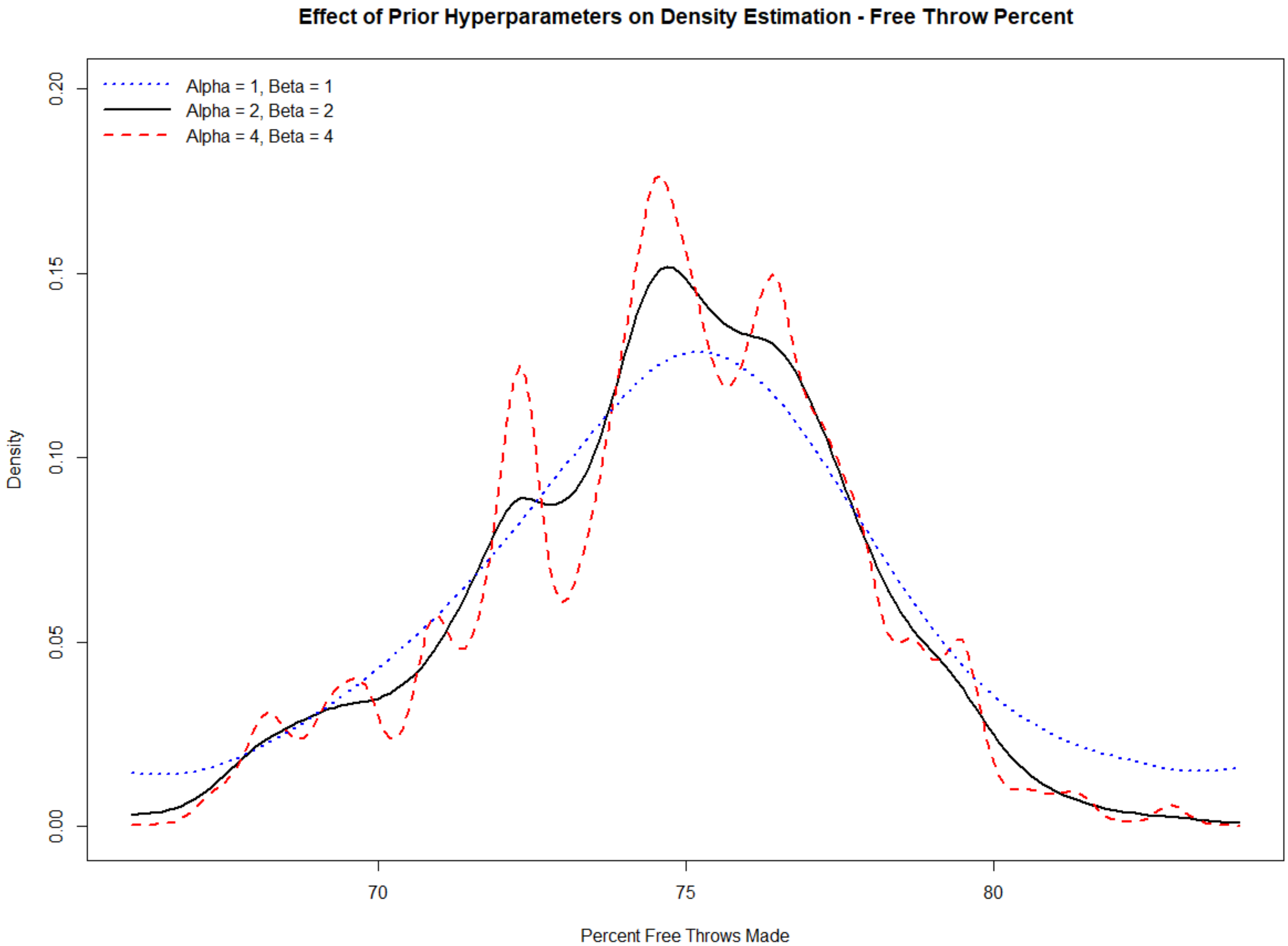
Free Throws Made Percent

The last variable I wanted to look at is the percent of free throws made. Perhaps as time passes, people will be more and more accurate with free throws. Free throws percent are defined same as field goals percent, by the free throws made over attempted, times 100. As field goals, each point indicates a team's overall free throws percent in a season.

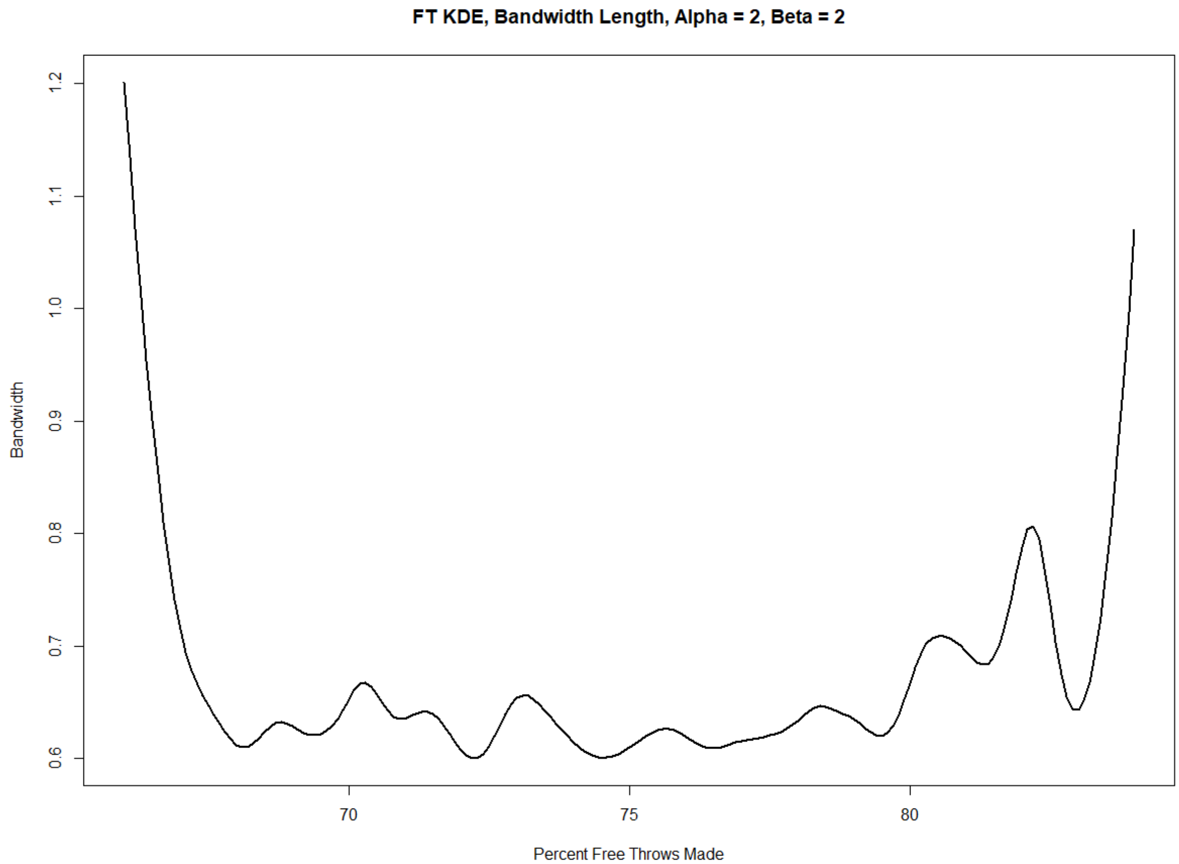
The histogram from the most recent decade is:



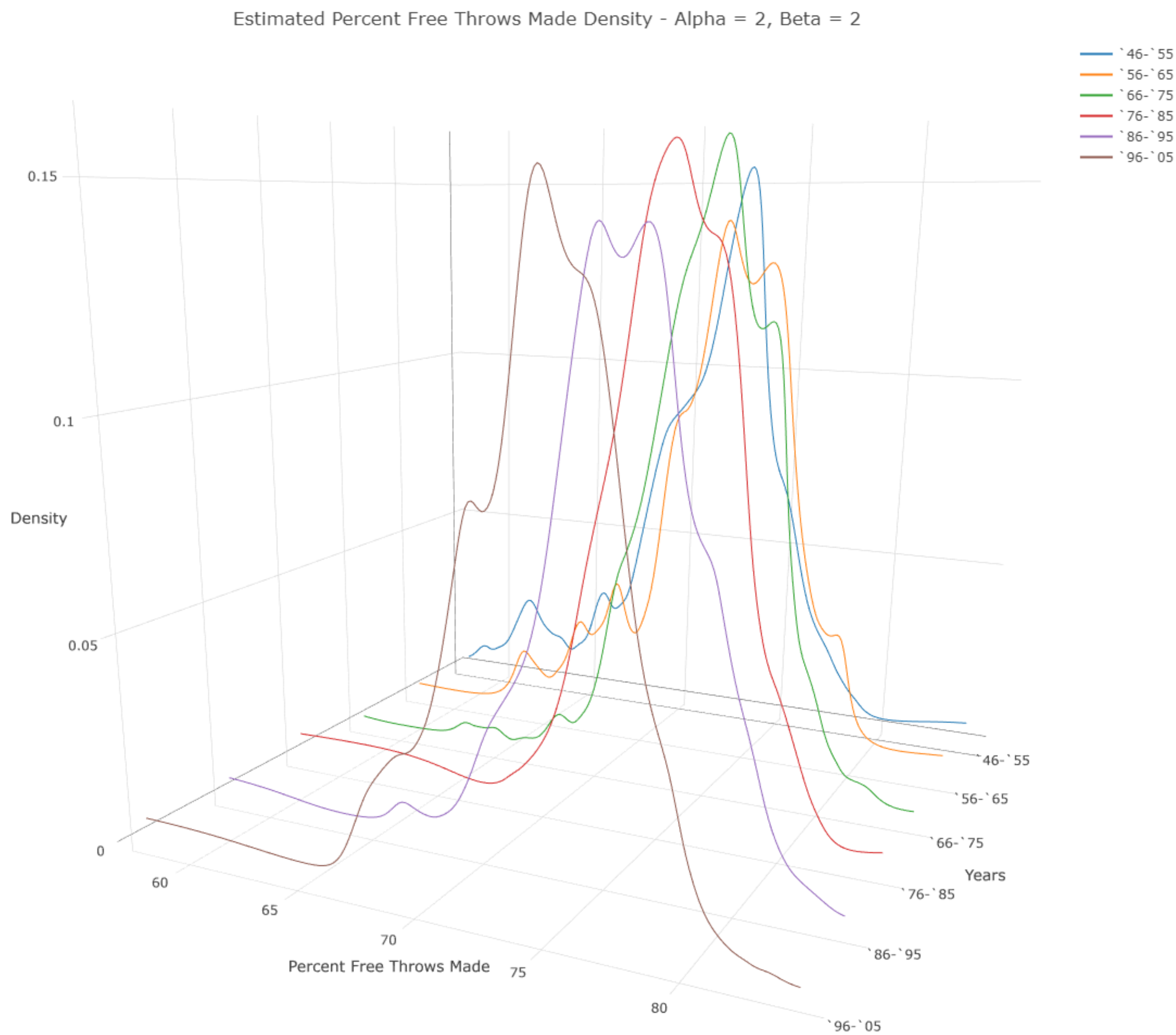
The distribution looks almost symmetric, with center at around 75%. I perform the Kernel Density estimation with three sets of hyper-parameters:



The above plot shows that ($\alpha = 4$, $\beta = 4$) under-smoothed the data, ($\alpha = 1$, $\beta = 1$) over-smoothed the data, and ($\alpha = 2$, $\beta = 2$) is about just right. Once again, I took a look at the posterior optimal bandwidth curve on **x**.



In this case, the optimal bandwidth appears to fluctuate heavily with the **x** value, suggesting existence of many clusters of data. Plotting the Kernel Density for each time period shows a surprising result:



Aside from minute details, there is no major change in free throw percent across the years. Therefore, from the beginning free throws have been at its optimal level. Over 50 years of practice have not had a major impact on the performance of free throw shooters. Intuitively, this does make some sense. Shooting a free throw only involves the player and the basket, there are no teammates, opponents, or strategies involved. There is very little to improve in shooting a free throw aside from being more accurate. Most NBA level basketball players have shot over tens of thousands of free throws, so they are likely shooting as accurate as they could.

Conclusion

Non-parametric methods can lead to powerful inference results, as illustrated in this project. While the majority of non-parametric methods do not involve Bayesian Statistics (as they do not involve parameters), in specific cases such as this one connections can be made. With a good choice in prior-parameters, Bayesian bandwidth can lead to the best density estimate results, out-performing fixed-bandwidth methods such as BCV and LSCV. This project showed just how dependent the density estimate is on the prior-parameters; just a tweak in the numbers can change a very smooth density estimate to a very rough one. A good topic to consider in the future is finding an algorithm to pick the optimal prior parameters that would lead to the most optimal bandwidth.

Reference

[1] A. Gangopadhyay and K. Cheung, Bayesian Approach to the Choice of Smoothing Parameter in Kernel Density Estimation, *Journal of Nonparametric Statistics*, vol. 14, 2002, pp.655-664.

R-Codes:

```
#Calvin Guan
#MA 578 Final Project
library(plotly)
library(plyr)
library(magrittr)

#Read in data, keep only NBA National leagues, and replace 0's with missing values
A0 <- read.csv("C:/Users/cxgua/OneDrive/Documents/MA578/Final Project/NBA Data/team_season.txt", header = 1)
A1 <- A0[A0$leag=='N',] ; A1[A1==0] <- NA
A2 <- A1[,c(2,4:18,34:36)]

A.sub <- function(var,yr){
  var[A2$year %in% yr]/(A2$won[A2$year %in% yr]+ A2$lost[A2$year %in% yr])
}

#Bayesian Posterior Smoothing Parameter
h.bayes <- function(x,X,a,b){
  l <- length(x); R <- numeric(l)
  for (i in 1:l){
    R[i] <- (gamma(a)/(gamma(a+.5)*sqrt(2*b)))*sum((1/(b*(X-x[i])^2+2))^a)/sum((1/(b*(X-x[i])^2+2))^(a+.5))
  } ; R
}

#Kernel Density Estimation using Bayesian Bandwidth and Gausian Kernel
KDE <- function(x,h,X){
  n <- length(X); l <- length(x); R <- numeric(l)
  for (i in 1:l){
    R[i] <- sum(dnorm((X-x[i])/h[i]))/(n*h[i])
  } ; R
}

########################
#Histogram of points
s1 <- A.sub(A2$o_pts,1996:2005)

hist(s1, main = 'Histogram of Points Scored Per Game for Teams From 1996 - 2005', xlab = "Points Per Games")

########################
### Density Estimation for Points Scored
#Try 3 different parameters for h
#1. a = 2, b = 1
#2. a = 4, b = 2
#3. a = 1, b = .5

x <- seq(60,125,.1)

a <- 2; b <- 1
plot(x, KDE(x,h.bayes(x,s1,a,b),s1), type = 'l', main = "Effect of Prior Hyperparameters on Density Estimation - 
Points Per Game",
     ylab = 'Density', xlab = 'Points Per Game', ylim = c(-0.003,0.1),lwd = 2)
a <- 4; b <- 2
lines(x, KDE(x,h.bayes(x,s1,a,b),s1), lty = 2, lwd = 2, col="red")
a <- 1; b <- .5
lines(x, KDE(x,h.bayes(x,s1,a,b),s1), lty = 3, lwd = 2, col = "blue")
legend('topleft',bty='n',lty=c(3,1,2), col = c("blue","black","red"), lwd = 2,
       legend=c('Alpha = 1, Beta = .5','Alpha = 2, Beta = 1','Alpha = 4, Beta = 2'), seg.len=4, cex=1)

### Bandwidth Values for Points Scored
#Try 3 different parameters for h
#1. a = 2, b = 1
#2. a = 4, b = 2
#3. a = 1, b = .5

a <- 2; b <- 1
plot(x, h.bayes(x,s1,a,b), type = 'l', main = "PPG KDE, Bandwidth Length, Alpha = 2, Beta = 1",
     ylab = 'Bandwidth', xlab = 'Points Per Game', lwd = 2)
a <- 4; b <- 2
plot(x, h.bayes(x,s1,a,b), main = "PPG KDE, Bandwidth Length, Alpha = 4, Beta = 2",
     ylab = 'Bandwidth', xlab = 'Points Per Game', lty = 2, lwd = 2, col="red", type = 'l')
a <- 1; b <- .5
plot(x, h.bayes(x,s1,a,b), main = "PPG KDE, Bandwidth Length, Alpha = 1, Beta = .5",
     ylab = 'Bandwidth', xlab = 'Points Per Game', lty = 3, lwd = 2, col = "blue", type = 'l')

###Influence of Prior on Bandwidth DECREASES as the parameter values increases
###########################################
### Group Data By Decades
```

```
g1 <- A.sub(A2$o_pts,1946:1955)
g2 <- A.sub(A2$o_pts,1956:1965)
g3 <- A.sub(A2$o_pts,1966:1975)
g4 <- A.sub(A2$o_pts,1976:1985)
g5 <- A.sub(A2$o_pts,1986:1995)
g6 <- A.sub(A2$o_pts,1996:2005)


### Get Estimated Density for each Decade
x <- seq(60,125,.1); l <- length(x)

a <- 2; b <- 1

k1 <- KDE(x,h.bayes(x,g1,a,b),g1)
k2 <- KDE(x,h.bayes(x,g2,a,b),g2)
k3 <- KDE(x,h.bayes(x,g3,a,b),g3)
k4 <- KDE(x,h.bayes(x,g4,a,b),g4)
k5 <- KDE(x,h.bayes(x,g5,a,b),g5)
k6 <- KDE(x,h.bayes(x,g6,a,b),g6)

k <- c(k1,k2,k3,k4,k5,k6)

###Decade labels
yr <- c(rep("`46-`55",l),rep("`56-`65",l),rep("`66-`75",l),rep("`76-`85",l),rep("`86-`95",l),rep("`96-`05",l))


plot_ly(y = rep(x,6), z = k, x = yr, type = "scatter3d", mode = "lines", split = yr) %>%
layout(
  title = "Estimated Points Per Game Density - Alpha = 2, Beta = 1",
  scene = list(
    yaxis = list(title = "Points Per Game"),
    xaxis = list(title = "Years"),
    zaxis = list(title = "Density")
  ))

#########################################
###Repeat for Field Goal Percentage
###Observation Level: Teams, Response: Overall FG Percentage in a season

########################
#Histogram of Field Goal Percentage
fg1 <- A.sub(A2$o_fgm,1996:2005)/A.sub(A2$o_fga,1996:2005)*100

hist(fg1, main = 'Histogram of Field Goal Percentage for Teams From 1996 - 2005', xlab = "Field Goal
Percentage")


########################
### Density Estimation for Points Scored
#Try 3 different parameters for h
#1. a = 2, b = 4
#2. a = 4, b = 8
#3. a = 1, b = 2
x <- seq(38,52,.1); l <- length(x)

a <- 2; b <- 4
plot(x, KDE(x,h.bayes(x,fg1,a,b),fg1), type = 'l', main = "Effect of Prior Hyperparameters on Density Estimation
- Field Goal Percent",
     ylab = 'Density', ylim = c(-0.001,0.28), xlab = 'Percent Field Goals Made', lwd = 2)
a <- 4; b <- 8
lines(x, KDE(x,h.bayes(x,fg1,a,b),fg1), lty = 2, lwd = 2, col="red")
a <- 1; b <- 2
lines(x, KDE(x,h.bayes(x,fg1,a,b),fg1), lty = 3, lwd = 2, col = "blue")
legend('topleft',bty='n',lty=c(3,1,2), col = c("blue","black","red"), lwd = 2,
       legend=c('Alpha = 1, Beta = 2','Alpha = 2, Beta = 4','Alpha = 4, Beta = 8'), seg.len=4, cex=1)

### Bandwidth Values for Points Scored
#Try 3 different parameters for h
#1. a = 2, b = 4
#2. a = 4, b = 8
#3. a = 1, b = 2

a <- 2; b <- 4
plot(x, h.bayes(x,fg1,a,b), type = 'l', main = "FG KDE, Bandwidth Length, Alpha = 2, Beta = 4",
     ylab = 'Bandwidth', xlab = 'Percent Field Goals Made', lwd = 2)
a <- 4; b <- 8
plot(x, h.bayes(x,fg1,a,b), main = "FG KDE, Bandwidth Length, Alpha = 4, Beta = 8",
     ylab = 'Bandwidth', xlab = 'Percent Field Goals Made', lty = 2, lwd = 2, col="red", type = 'l')
a <- 1; b <- 2
```

```
plot(x, h.bayes(x,fg1,a,b), main = "FG KDE, Bandwidth Length, Alpha = 1, Beta = 2",
     ylab = 'Bandwidth', xlab = 'Percent Field Goals Made', lty = 3, lwd = 2, col = "blue", type = 'l')

###########################################
### Group Data By Decades
f1 <- A.sub(A2$o_fgm,1946:1955)/A.sub(A2$o_fga,1946:1955)*100
f2 <- A.sub(A2$o_fgm,1956:1965)/A.sub(A2$o_fga,1956:1965)*100
f3 <- A.sub(A2$o_fgm,1966:1975)/A.sub(A2$o_fga,1966:1975)*100
f4 <- A.sub(A2$o_fgm,1976:1985)/A.sub(A2$o_fga,1976:1985)*100
f5 <- A.sub(A2$o_fgm,1986:1995)/A.sub(A2$o_fga,1986:1995)*100
f6 <- A.sub(A2$o_fgm,1996:2005)/A.sub(A2$o_fga,1996:2005)*100


### Get Estimated Density for each Decade
x <- seq(24,55,.1); l <- length(x)

a <- 2; b <- 4

j1 <- KDE(x,h.bayes(x,f1,a,b),f1)
j2 <- KDE(x,h.bayes(x,f2,a,b),f2)
j3 <- KDE(x,h.bayes(x,f3,a,b),f3)
j4 <- KDE(x,h.bayes(x,f4,a,b),f4)
j5 <- KDE(x,h.bayes(x,f5,a,b),f5)
j6 <- KDE(x,h.bayes(x,f6,a,b),f6)

j <- c(j1,j2,j3,j4,j5,j6)

###Decade labels
yr <- c(rep("`46-`55",l),rep("`56-`65",l),rep("`66-`75",l),rep("`76-`85",l),rep("`86-`95",l),rep("`96-`05",l))

plot_ly(y = rep(x,6), z = j, x = yr, type = "scatter3d", mode = "lines", split = yr) %>%
  layout(
    title = "Estimated Percent Field Goal Made Density - Alpha = 2, Beta = 4",
    scene = list(
      yaxis = list(title = "Percent Field Goals Made"),
      xaxis = list(title = "Years"),
      zaxis = list(title = "Density")
    ))

#########################################
###Repeat for Free Throw Percentage
###Observation Level: Teams, Response: Overall Free Throw Percentage in a season

########################
#Histogram of Field Goal Percentage
ft1 <- A.sub(A2$o_ftm,1996:2005)/A.sub(A2$o_fta,1996:2005)*100

hist(ft1, main = 'Histogram of Free Throw Percentage for Teams From 1996 - 2005', xlab = "Free Throw
Percentage")


########################
### Density Estimation for Points Scored
#Try 3 different parameters for h
#1. a = 2, b = 2
#2. a = 4, b = 4
#3. a = 1, b = 1
x <- seq(66,84,.1); l <- length(x)

a <- 2; b <- 2
plot(x, KDE(x,h.bayes(x,ft1,a,b),ft1), type = 'l', main = "Effect of Prior Hyperparameters on Density Estimation
- Free Throw Percent",
     ylab = 'Density', ylim = c(-0.001,0.20), xlab = 'Percent Free Throws Made', lwd = 2)
a <- 4; b <- 4
lines(x, KDE(x,h.bayes(x,ft1,a,b),ft1), lty = 2, lwd = 2, col="red")
a <- 1; b <- 1
lines(x, KDE(x,h.bayes(x,ft1,a,b),ft1), lty = 3, lwd = 2, col = "blue")
legend('topleft',bty='n',lty=c(3,1,2), col = c("blue","black","red"), lwd = 2,
       legend=c('Alpha = 1, Beta = 1','Alpha = 2, Beta = 2','Alpha = 4, Beta = 4'), seg.len=4, cex=1)

### Bandwidth Values for Points Scored
#Try 3 different parameters for h
#1. a = 2, b = 2
#2. a = 4, b = 4
#3. a = 1, b = 1

a <- 2; b <- 2
plot(x, h.bayes(x,ft1,a,b), type = 'l', main = "FT KDE, Bandwidth Length, Alpha = 2, Beta = 2",
     ylab = 'Bandwidth', xlab = 'Percent Free Throws Made', lwd = 2)
```

```
a <- 4; b <- 4
plot(x, h.bayes(x,ft1,a,b), main = "FT KDE, Bandwidth Length, Alpha = 4, Beta = 4",
     ylab = 'Bandwidth', xlab = 'Percent Free Throws Made', lty = 2, lwd = 2, col="red", type = 'l')
a <- 1; b <- 1
plot(x, h.bayes(x,ft1,a,b), main = "FT KDE, Bandwidth Length, Alpha = 1, Beta = 1",
     ylab = 'Bandwidth', xlab = 'Percent Free Throws Made', lty = 3, lwd = 2, col = "blue", type = 'l')

###########################################
### Group Data By Decades
h1 <- A.sub(A2$o_ftm,1946:1955)/A.sub(A2$o_fta,1946:1955)*100
h2 <- A.sub(A2$o_ftm,1956:1965)/A.sub(A2$o_fta,1956:1965)*100
h3 <- A.sub(A2$o_ftm,1966:1975)/A.sub(A2$o_fta,1966:1975)*100
h4 <- A.sub(A2$o_ftm,1976:1985)/A.sub(A2$o_fta,1976:1985)*100
h5 <- A.sub(A2$o_ftm,1986:1995)/A.sub(A2$o_fta,1986:1995)*100
h6 <- A.sub(A2$o_ftm,1996:2005)/A.sub(A2$o_fta,1996:2005)*100


### Get Estimated Density for each Decade
x <- seq(58,84,.1); l <- length(x)

a <- 2; b <- 2

m1 <- KDE(x,h.bayes(x,h1,a,b),h1)
m2 <- KDE(x,h.bayes(x,h2,a,b),h2)
m3 <- KDE(x,h.bayes(x,h3,a,b),h3)
m4 <- KDE(x,h.bayes(x,h4,a,b),h4)
m5 <- KDE(x,h.bayes(x,h5,a,b),h5)
m6 <- KDE(x,h.bayes(x,h6,a,b),h6)

m <- c(m1,m2,m3,m4,m5,m6)

###Decade labels
yr <- c(rep("`46-`55",l),rep("`56-`65",l),rep("`66-`75",l),rep("`76-`85",l),rep("`86-`95",l),rep("`96-`05",l))

plot_ly(y = rep(x,6), z = m, x = yr, type = "scatter3d", mode = "lines", split = yr) %>%
  layout(
    title = "Estimated Percent Free Throws Made Density - Alpha = 2, Beta = 2",
    scene = list(
      yaxis = list(title = "Percent Free Throws Made"),
      xaxis = list(title = "Years"),
      zaxis = list(title = "Density")
    ))
```