Stephen Au

Geog. 172

Fall 2015

**Term Project**

Research Question

Basketball has become increasingly popularized since its birth in 1891. It started as a way to keep teenagers conditioned during the cold months of winter. Since then, the sport has spawned professional leagues and amassed worldwide recognition. The game itself has evolved over time as different eras called for different play styles. We are currently in the era of the three-point shot; increased value is placed on three pointers therefore three point shooters are more valuable than ever before. Players like Kyle Korver, who were bench players in years past now play important roles within a team’s offensive structure. Players like the Golden State Warriors’ Stephen Curry and Klay Thompson have led this revolutionary change throughout the NBA landscape. These two players have helped set a trend that many NBA teams are beginning to follow; teams employ sharpshooters to get more points out of an offensive possession simply because three points are more than two. However, certain players are better at certain parts of the court at knocking down three-pointers (hotspots). This increased efficiency can be attributed to various reasons: certain shots are easier to make, players are more open, game flow dictates certain spots on the floor are easier to make, or just being more open but the end result (an increased efficiency at certain spots beyond the three-point line) is the same: a make or a miss. These three aforementioned players and their subsequent play styles drastically differ from the fourth player I want to include in this analysis. Kobe Bryant has made a living since entering the league in 1996 as the heir apparent to whom most NBA fans will dub the greatest player of all time, Michael Jordan. Kobe Bryant is known to be efficient at scoring in all aspects of the game. The question I am trying to test is whether or not underlying processes on the basketball court can describe NBA players (in this case, Stephen Curry, Klay Thompson, Kyle Korver, and Kobe Bryant) three-point shots. I will focus my analysis on Stephen Curry since he is the poster boy for the modern day three-point revolution. Are his made shots the result of favoring certain spots or does he make these shots randomly following an Independent Random Process.

DataS

The data I used comes from the 2014-2015 NBA regular season. I compare three other basketball players to Stephen Curry’s shot chart. I recorded all the individual three-pointers made for Stephen Curry, Kobe Bryant, Klay Thompson, and Kyle Korver; these players are known for their shooting abilities. I took a diagram of a basketball court and found the players’ shot charts online. I only recorded makes because I was only interested in whether or not their makes congregate in certain ‘hotspots’ beyond the three-point line. Although Kobe Bryant is a well-known scorer, he mostly does his damage from mid-range around the free-throw line and baselines. Stephen Curry and his teammate, Klay Thompson, have combined to make more three-pointers than NBA teams have in a season. Kyle Korver is a three-point specialist known and paid to make three-pointers at a high percentage. Not recording makes against misses dismisses plenty of factors and clouds up questions about efficiency from the three-point line, but I am only interested in whether or not there is a pattern among these three-point shooters. I will provide percentages for these players in five different three-point regions of the court, the corners, the wings, and the top of the key. Stephen Curry made about 3.6 three-pointers per game during the 2014-2015 season knocking them down at a 44% clip. His backcourt companion, Klay Thompson averaged 3.1 per game making about 44% as well. Kyle Korver made about 2.9 three’s last year but made almost half of his attempts (49.2%). I wanted to include Kobe here to compare how well older players with older play styles have adapted to the modern game. Kobe made about 1.5 three’s per game last year and knocked 37% of his threes in 35 games. The other three players played in all 82 games of the season. A noteworthy percentage to take down would be that Stephen Curry knocked down 40/58(69%) of his three-point attempts from the left corner. Everyone else shot about 40-50% or below this mark(Kobe).

Methods

I used the JPEG library to embed the NBA court into a plot and used the locator function to individually map out and record all of three-pointers made in the season. I recorded the data into a csv file and later used this csv file to run a quadrat analysis using 400 quadrat cells (20 rows, 20 columns). I tested the data against three different statistical models: Poisson (random), clustered (negative binomial), and binomial (regular). I also used the ppp function to look for intensity of three-pointers made for these three players. I used the K and G functions for the players point patterns to further test for clustering against an IRP/CSR process. The IRP/CSR values generated are skewed because I couldn’t set the boundary for the JPEG to only analyze makes beyond the three point line. The envelopes created here also don’t incorporate this so the plotted G and K envelopes for Stephen Curry are off. I tried testing for correlation between players based on the data generated, but the corr function in R didn’t yield numerical values when I tested Stephen Curry against Klay Thompson (a graph was produced though).

Results

The quadrat analysis resulted in the following graphs for Kobe Bryant, (fig.1) Stephen Curry, (fig.2) Kyle Korver, (fig.3) and Klay Thompson (fig.4). When tested for random, clustered, and regular processes, these players yielded different results that can be interpreted differently depending on the conditions set. For Stephen Curry, the p-value for poisson and binomial distributions was 0 but the negative binomial p-value is the only process that produces a value above 0. If we set the cut-off value at 0.05, this would also fall under rejection but since this is the only model that has a p-value, there is small evidence that Stephen Curry’s shot chart is clustered(0.1% chance) under this model, although we are 99.99% sure this isn’t the case. In Kobe’s quadrat analysis, the resulting p-values under the statistical models resulted in rejecting all three models but his shot chart produced a p-value of 0.013 for the poisson distribution, or a 1.3% that his shot chart is due to random processes. Kyle Korver’s quadrat analysis yielded a p-value of 0.159 for the clustering model (negative binomial) so we can’t reject this under the 0.05 condition. Kyle Korver’s shot chart shows evidence of clustering, especially above the top of the key. Klay Thompson’s quadrat analysis yielded p-values well below 0.05 so we can safely reject these models under this condition (p-value of 0.019 for clustering). Figures 5 and 6 show the point pattern density for all four player’s shot charts under a ‘magnification’ value of 50. Figures 7,8,9,and 10 show the observed G and K functions plotted against a randomly produced process with envelopes simulated 199 times. This envelope can be disregarded because I couldn’t set a boundary to strictly simulate the IRP/CSR process for points beyond the three-point line. It is worth noting that despite this, Kobe’s shot chart can be classified as randomly dispersed because the observed line plot falls under the expected line in both G and K functions until they reach a distance of about 16-18 where it turns we can assume the shot chart becomes more clustered. This can be explained by the small amount of point patterns because of Kobe’s small number of games played and consequently smaller sample size. Figure 11 shows the scatterplot I attempted to produce to compare Stephen Curry’s shots to his teammates. However, the correlation matrix produced NA values when comparing Klay’s y-values to Curry’s x and y-values. This plot is not a good test for linear correlation between variables because the points plotted and the data collected results in a negative arc, so linear regression is not a suitable option for the research question.

Discussion

Quadrat analysis and point pattern analyses provides the most useful data to answer this research question. Linear regression is not suitable for this test because the data stored is in the shape of –x^2 graph, and this is not linear. A Moran’s I value could have provided a value to show how strong spatial autocorrelation was and this could in turn show that Stephen Curry likes to get his shots off at certain spots on the three-point line. Again, I couldn’t code R to set a boundary so that I could test the actual outcomes against an IRP/CSR process for three-pointers. This could explain reasons as to why Kyle Korver’s quadrat analysis yielded a p-value of 0.159 for clustering: expected values will hover around the middle and a majority of his shots come near the middle of the plot, resulting in a p-value greater than our cutoff. The high number of shots recorded does show that these shots tend to cluster around each other, but I can’t conclude that this clustering is due to first-order inhomogeneity or second-order. Performing a quadrat analysis with 400 quadrat cells also may have been too many cells to work with leading to a less intense point pattern observation. Quadrat and point pattern analysis provides the most useful data for this research question and the K and G functions could help determine whether the clustering is the result of first-order inhomogeneity (easier shots have more makes) or second-order (game conditions and defenses cause the players to shoot in the ‘harder’ spots). If it is the first condition of inhomogeneity, than Stephen Curry benefits from shooting and making easier shots. If it is the latter case, then Stephen Curry is good enough at shooting three-pointers to make defenses pay for letting him shoot those shots. The latter seems to be the case, especially considering his percentages at the three point line and the sheer amount of makes he has when compared to the other players. The G and K functions probably serve the best to compare Curry’s shot chart to a random one but looking at the density plot, and the quadrat analysis, it seems as though Stephen Curry can make a lot of three-pointers from anywhere along the three-point line.

Appendix

setwd("/Users/stephenau/Documents/CurryProject")

install.packages("jpeg")

library("jpeg")

par(mai=c(0,0,0,0))

graphics.off()

shots<-readJPEG("NBACOURT.jpg",native=TRUE)

res = dim(shots)[1:2] # get the resolution

plot(1,1,xlim=c(1,res[1]),ylim=c(1,res[2]),asp=1,type='n',xaxs='i',yaxs='i',xaxt='n',yaxt='n',xlab='',ylab='',bty='n')

rasterImage(shots,1,1,res[1],res[2])

#plots the jpg file as a matrix plot into R. This allows me to perform further functions on this data.

stephcurry<-locator()

kobebryant<-locator()

kylekorver<-locator()

klaythompson<-locator()

#Lets me click individual three-point shots made.

text(stephcurry,label=1:length(stephcurry$x),col="gold1")

text(kobebryant,label=1:length(kobebryant$x),col="purple1")

text(kylekorver,label=1:length(kylekorver$x),col="red1")

text(klaythompson,label=1:length(klaythompson$x),col="blue1")

#Labels the values from 1 to n of the made shots.

stephthrees<-data.frame(stephcurry)

kobethrees<-data.frame(kobebryant)

korverthrees<-data.frame(kylekorver)

thompsonthrees<-data.frame(klaythompson)

library('spatstat')

source('quadrat\_functions.r')

#stores the recorded shots and their x,y coordinates into a variable that can be brought up in R.

set.seed(3)

pp\_unif<-runifpoint(250)

pp\_cluster<-rMatClust(50,0.1,5)

pp\_CSR<-rpoispp(250)

pp\_inhibition<-rMaternI(200,0.028)

graphics.off()

curry<-read.csv("stephsthrees.csv")

kobe<-read.csv("kobesthrees.csv")

korver<-read.csv("korversthrees.csv")

thompson<-read.csv("thompsonsthrees.csv")

#reads the x and y-coordinates of the data recorded earlier in the shot chart.

pp\_curry<-data.frame(x=curry$x,y=curry$y)

quadrat(pp\_curry,20,20,TRUE,'CURRY')#the negative binomial p-value is the only

pp\_kobe<-data.frame(x=kobe$x,y=kobe$y)

quadrat(pp\_kobe,20,20,TRUE,'KOBE')#p-values show we cna reject all these

pp\_korver<-data.frame(x=korver$x,y=korver$y)

quadrat(pp\_korver,20,20,TRUE,'KORVER')

pp\_thompson<-data.frame(x=thompson$x,y=thompson$y)

quadrat(pp\_thompson,20,20,TRUE,'THOMPSON')

#these lines serve to perform quadrat analysis against statistical models that suggest clustering, random, and normal processes. The quadrat analysis is performed with 20 rows and 20 columns

bndry<-owin(xrange=c(0,599),yrange=c(0,800))

help(ppp)

curry.pp<-ppp(curry$x,curry$y,window=bndry)

kobe.pp<-ppp(kobe$x,kobe$y,window=bndry)

korver.pp<-ppp(korver$x,korver$y,window=bndry)

thompson.pp<-ppp(thompson$x,thompson$y,window=bndry)

#This stores point patterns processes into variables that are later used to plot

layout(matrix(1:4,2,2))

par(mai=rep(0.5,4))

plot(curry.pp,cex=0.5,main="CURRY")

plot(kobe.pp,meansize=2,main="KOBE")

plot(korver.pp,cex=0.5,main="KORVER")

plot(thompson.pp,cex=0.5,main="THOMPSON")

#plots the data from the csv files into a point pattern plot.

layout(matrix(1:4,2,2))

par(mai=rep(0.5,4))

plot(density.ppp(curry.pp,50))

plot(density.ppp(kobe.pp,50))

plot(density.ppp(korver.pp,50))

plot(density.ppp(thompson.pp,50))

#plots the intensity of the shot chart under a ‘magnification’ of 50. This shows where the players tend to take their most three-pointers.

GSC<-Gest(curry.pp)

KSC<-Kest(curry.pp)

GKB<-Gest(kobe.pp)

KKB<-Kest(kobe.pp)

GKK<-Gest(korver.pp)

KKK<-Kest(korver.pp)

GKT<-Gest(thompson.pp)

KKT<-Kest(thompson.pp)

#makes use of the point pattern data from earlier and performs G and K functions for this data.

layout(matrix(c(1,2),1,2))

plot(GSC,cbind(theo,rs)~r,col=c("black","red"),ylab="G(r)",main="StephenCurry shotchart",lty=1,legendargs=list(cex=0.75),cex.main=0.8)

plot(KSC,sqrt(cbind(theo,border)/pi)-r~r,xlim=c(0,20),col=c("black","red"),ylab="L(r)",main="StephenCurry shotchart",lty=1,legendargs=list(cex=0.75),cex.main=0.8)

plot(GKB,cbind(theo,rs)~r,col=c("black","red"),ylab="G(r)",main="KOBE shotchart",lty=1,legendargs=list(cex=0.75),cex.main=0.8)

plot(KKB,sqrt(cbind(theo,border)/pi)-r~r,xlim=c(0,20),col=c("black","red"),ylab="L(r)",main="KOBE shotchart",lty=1,legendargs=list(cex=0.75),cex.main=0.8)

plot(GKK,cbind(theo,rs)~r,col=c("black","red"),ylab="G(r)",main="Kylekorver shotchart",lty=1,legendargs=list(cex=0.75),cex.main=0.8)

plot(KKK,sqrt(cbind(theo,border)/pi)-r~r,xlim=c(0,20),col=c("black","red"),ylab="L(r)",main="Kylekorver shotchart",lty=1,legendargs=list(cex=0.75),cex.main=0.8)

plot(GKT,cbind(theo,rs)~r,col=c("black","red"),ylab="G(r)",main="KlayThompson shotchart",lty=1,legendargs=list(cex=0.75),cex.main=0.8)

plot(KKT,sqrt(cbind(theo,border)/pi)-r~r,xlim=c(0,20),col=c("black","red"),ylab="L(r)",main="KlayThompson shotchart",lty=1,legendargs=list(cex=0.75),cex.main=0.8)

#these lines produce the basic G and K function plots from the earlier lines.

myKfun <- function(S, ..., lam) { Kinhom(S, lambda=lam[S], ...) }

GSCE<-envelope(curry.pp,Gest)

KSCE<-envelope(curry.pp,Kest)

layout(matrix(c(1,2),1,2))

plot(GSCE,cbind(theo,obs,lo,hi)~r,col=c("black","red","black","black"),lty=c(1,1,2,2),ylab="G(r)",main="Curryshots: G(r)",legendargs=list(cex=0.75),cex.main=0.8)

plot(KSCE,sqrt(cbind(theo,obs,lo,hi)/pi)-r~r,col=c("black","red","black","black"),lty=c(1,1,2,2),ylab="L(K(r))",main="Curryshots: K(r)",legendargs=list(cex=0.75),cex.main=0.8)

#produces K and G function plots with ‘envelopes’ that are randomly generated and used to observe whether or not our observed values fall within observed values.

curryVS<-read.csv('stephcurrycomposite.csv')

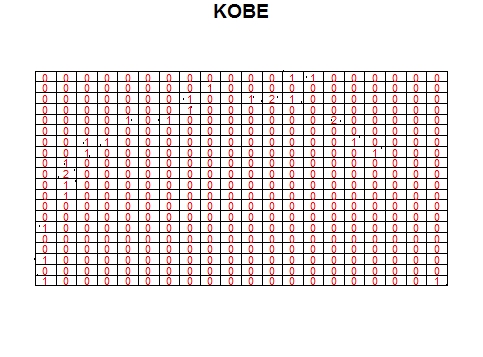
attach(curryVS)

names(curryVS)

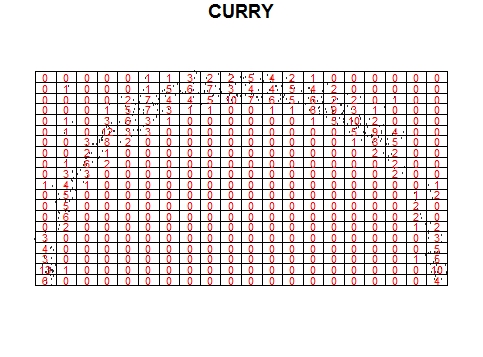
pairs(cbind(x,y,yKlay),main="StephvsKlay")

cor(cbind(x,y,xKlay,yKlay))

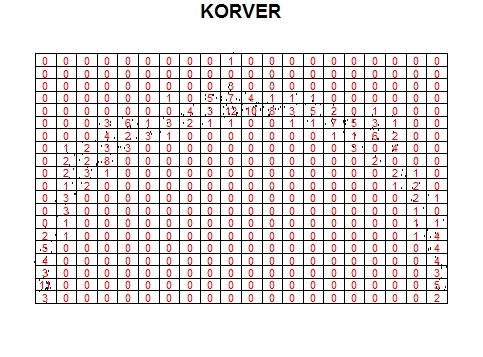
#this was a correlation test between Stephen Curry and Klay Thompson’s shots. The x and y coordinates between Curry’s shots yielded a value close to 0



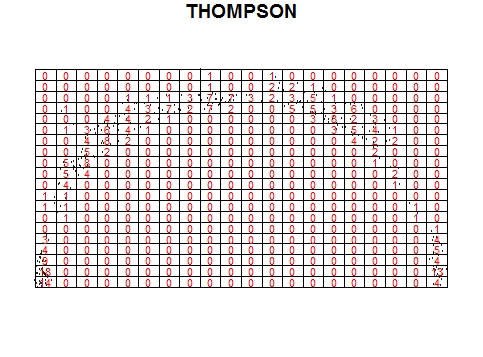
Figure



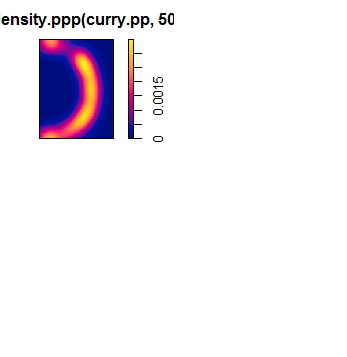
Figure



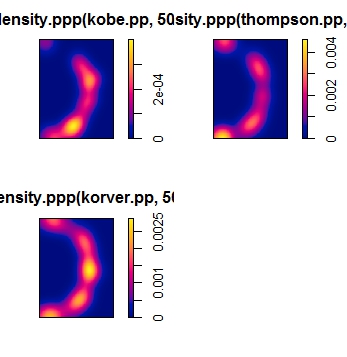
Figure



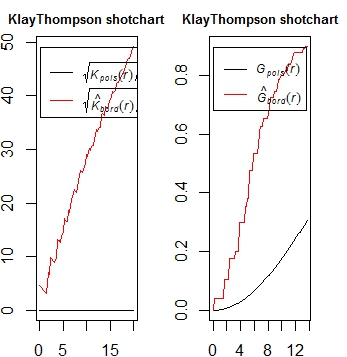
Figure



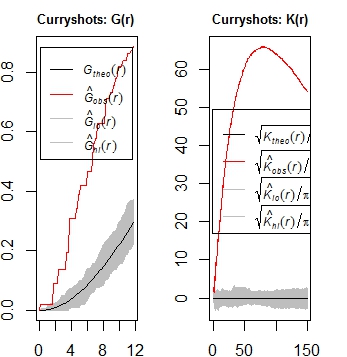
Figure



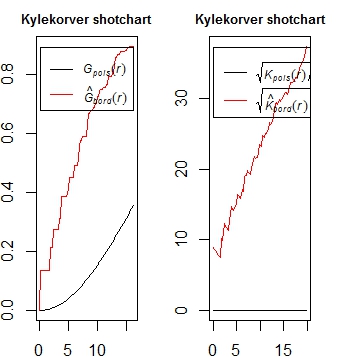
Figure



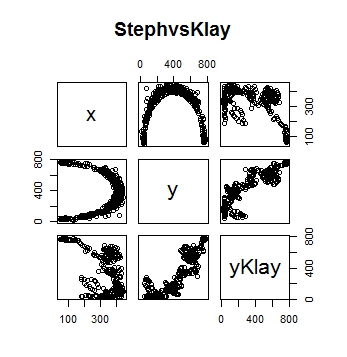
Figure



Figure



Figure



Figure