

Tarea 1

Juan Cantero Jimenez

2/26/2022

Caraga de datos y exploración

```
data <- read.csv("datos_investigacion.csv")

head(data)

##   X         L       FD Atractores           Id
## 1 0  0.16250242 0.6518040          2      YVMPMRMJXTLY
## 2 1 -0.02183638 1.5843295          3      FEMSSNFWNXVPUNYNPM
## 3 2  0.20041153 1.3935737          4 JHEQSAKIYOBMKVXJRJBCLKKL
## 4 3          NA 0.0000000          2      UCYIJKTYGEYX
## 5 4 -0.67532628 0.7113785          3      PHTMMOPVKKLVUHTGKH
## 6 5  0.30137549 2.2329936          4 AFKJPWBQDTQKJHSYHCYNIKIG

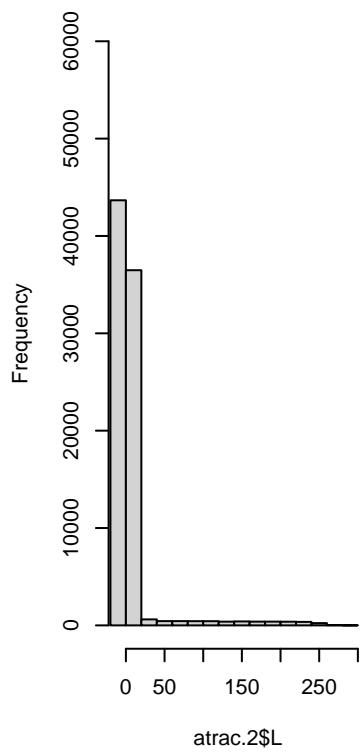
str(data)

## 'data.frame': 300000 obs. of 5 variables:
## $ X : int 0 1 2 3 4 5 6 7 8 9 ...
## $ L : num 0.1625 -0.0218 0.2004 NA -0.6753 ...
## $ FD : num 0.652 1.584 1.394 0 0.711 ...
## $ Atractores: int 2 3 4 2 3 4 2 3 4 2 ...
## $ Id : chr "YVMPMRMJXTLY" "FEMSSNFWNXVPUNYNPM" "JHEQSAKIYOBMKVXJRJBCLKKL" "UCYIJKTYGEYX" ...

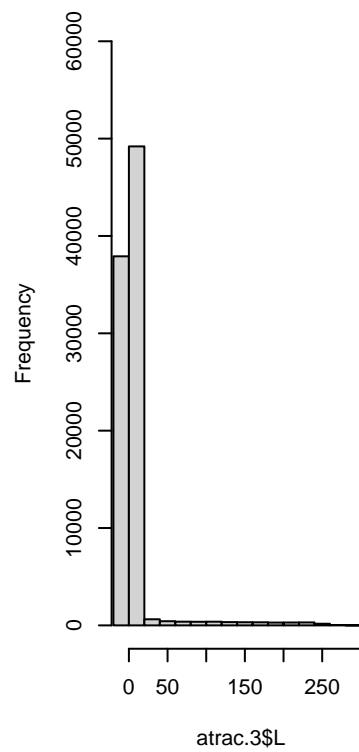
atrac.2 <- data[which(data$Atractores==2),]
atrac.3 <- data[which(data$Atractores==3),]
atrac.4 <- data[which(data$Atractores==4),]

par(mfrow=c(1,3))
hist(atrac.2$L,xlim = c(-10,300),ylim=c(0,60000))
hist(atrac.3$L,xlim = c(-10,300),ylim=c(0,60000))
hist(atrac.4$L,xlim = c(-10,300),ylim=c(0,60000))
```

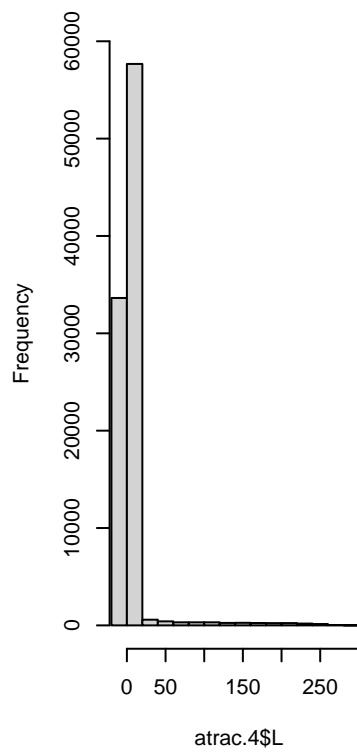
Histogram of atrac.2\$L



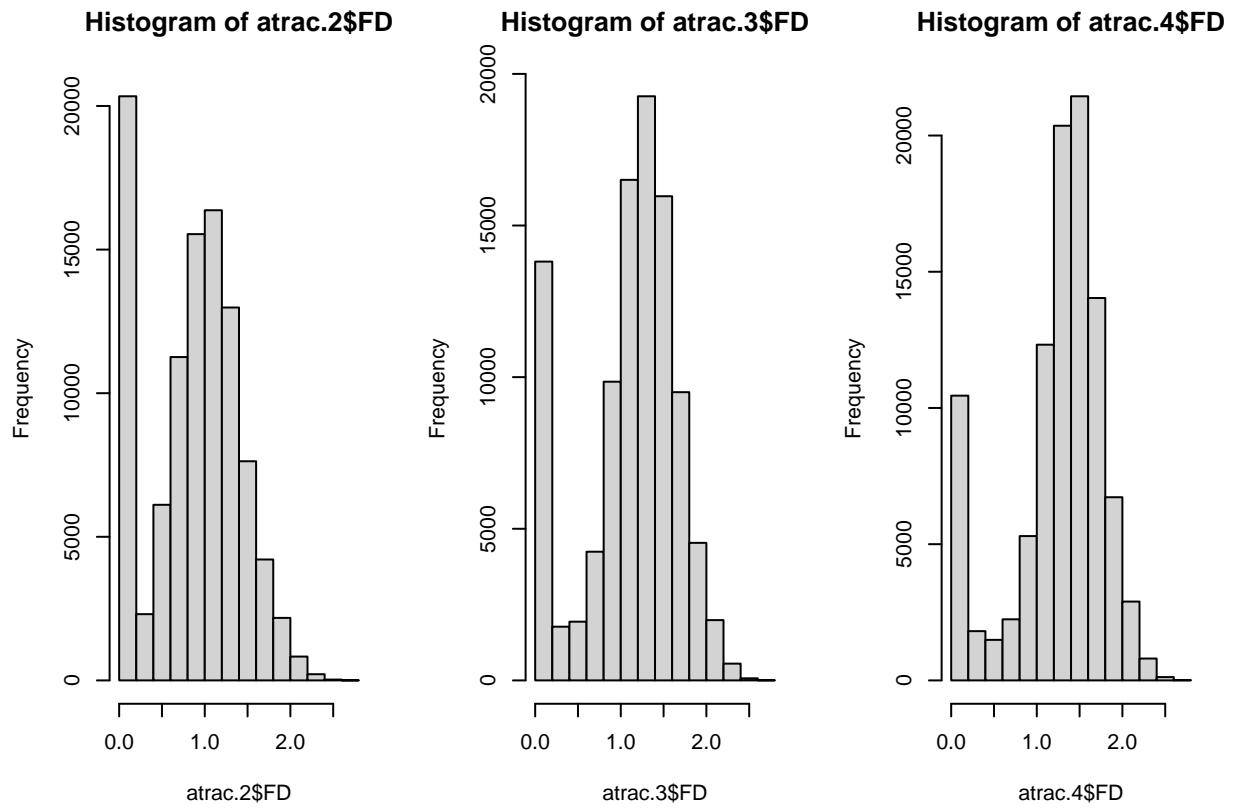
Histogram of atrac.3\$L



Histogram of atrac.4\$L

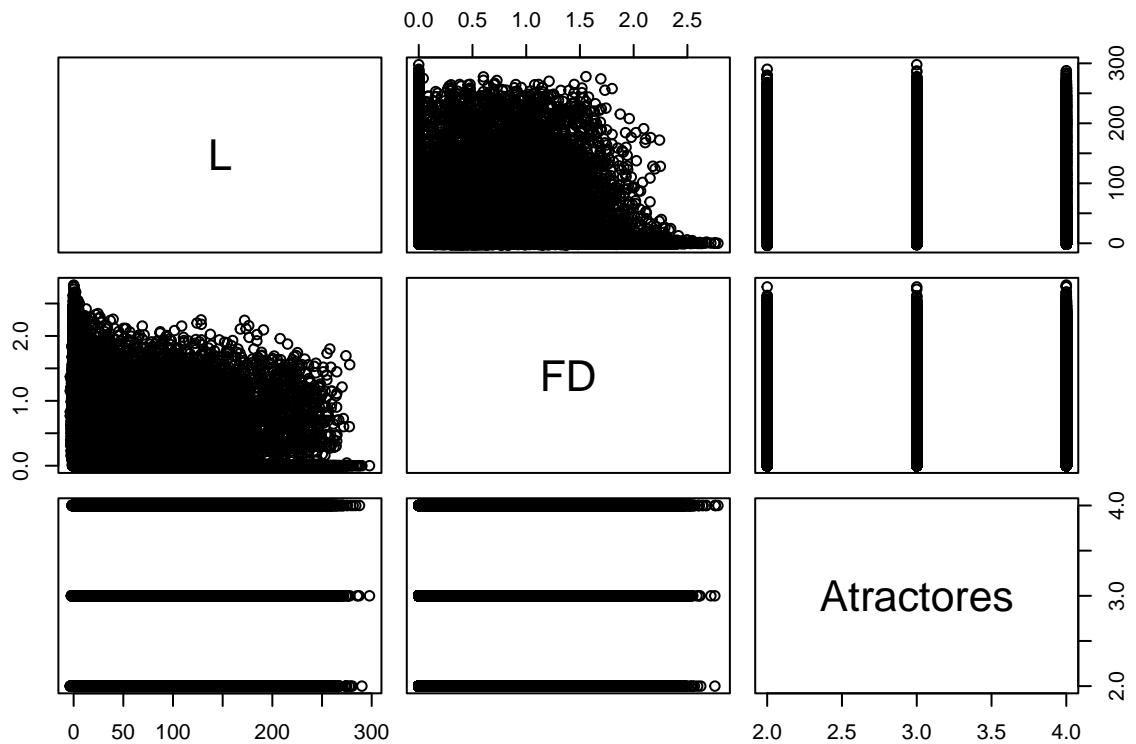


```
par(mfrow=c(1,3))
hist(atrac.2$FD)
hist(atrac.3$FD)
hist(atrac.4$FD)
```



```
data$Atractores.factor <- as.factor(data$Atractores)

pairs(data[,c("L", "FD", "Atractores")])
```



```

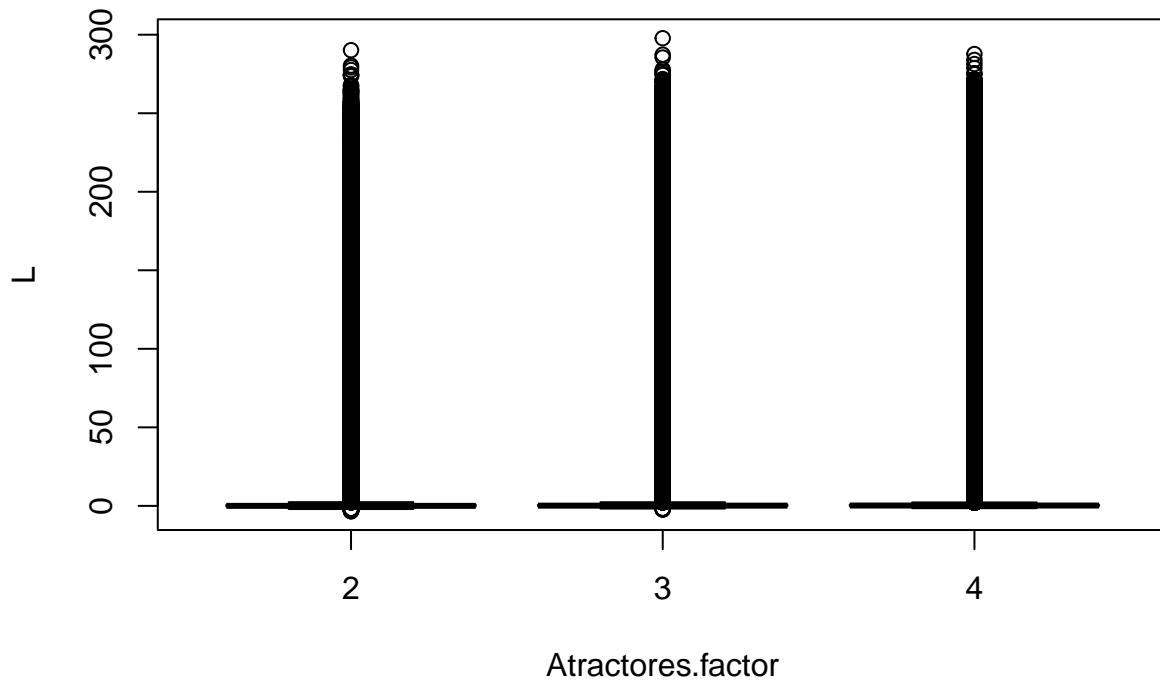
boxplot(L ~ Atractores.factor, data)

## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out = z$out[z$group
## == : Outlier (Inf) in boxplot 1 is not drawn

## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out = z$out[z$group
## == : Outlier (Inf) in boxplot 2 is not drawn

## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out = z$out[z$group
## == : Outlier (Inf) in boxplot 3 is not drawn

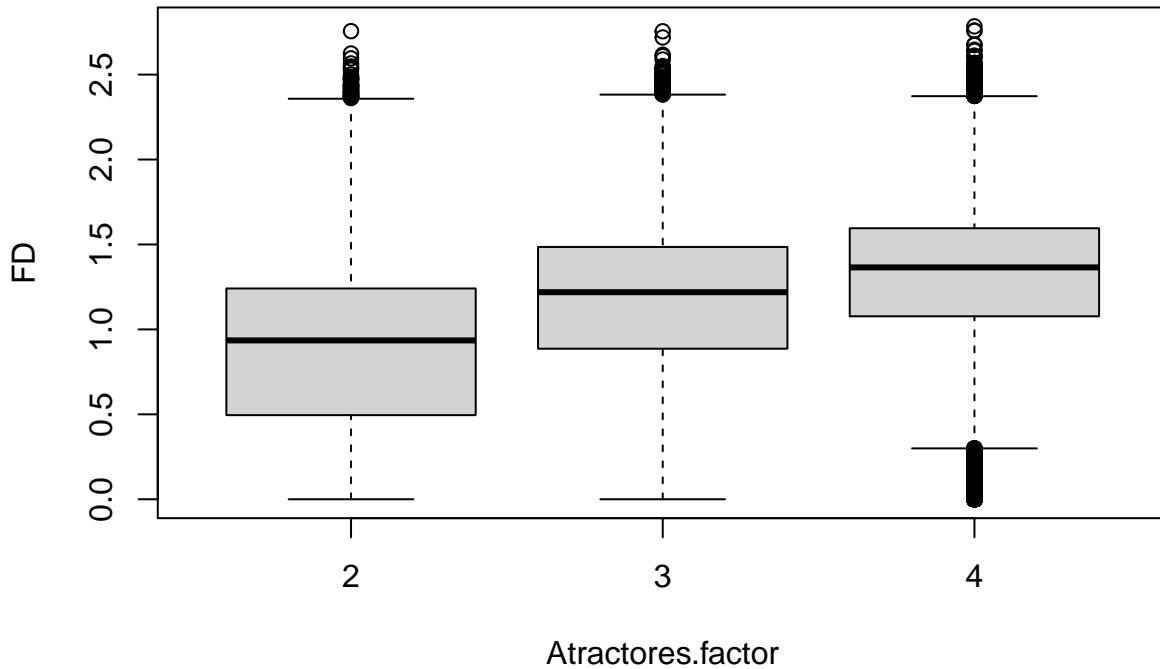
```



```

boxplot(FD ~ Atractores.factor, data)

```



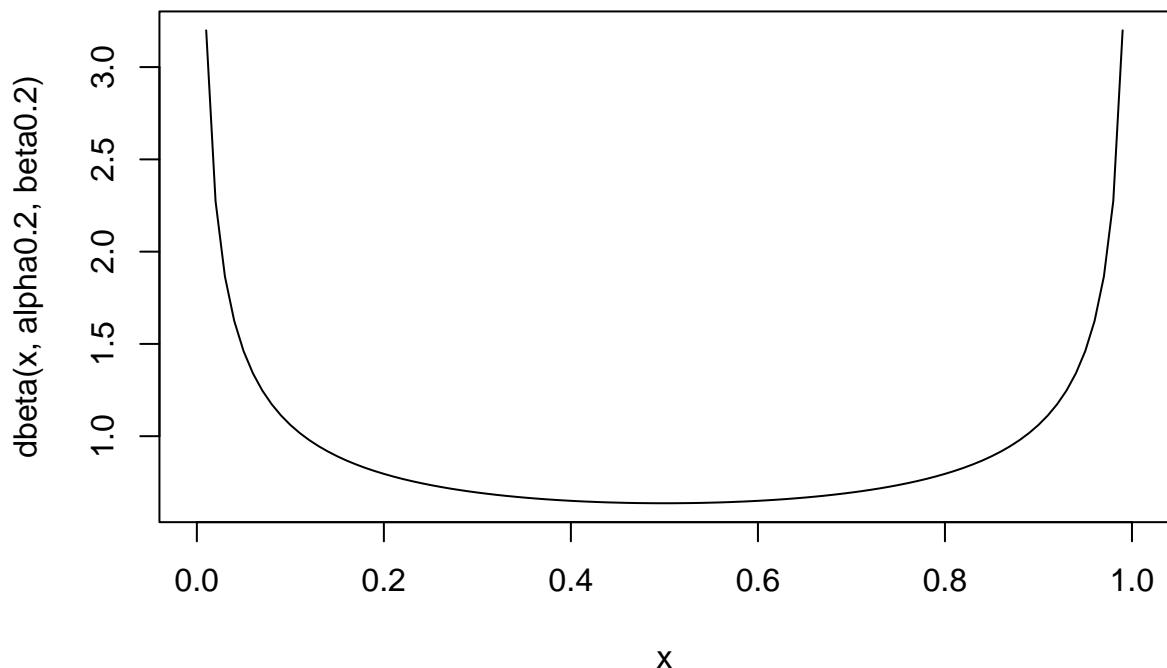
Inferencia sobre 2 atractores

```
n.2 <- nrow(atrac.2)
## atrac.2$FD != Inf & atrac.2$FD != -Inf & atrac.2$L != -Inf & atrac.2$L != Inf
exitos.2 <- length(which(atrac.2$L <= -0.2 & atrac.2$FD >= 1 & atrac.2$FD != Inf & atrac.2$FD != -Inf))
exitos.2
## [1] 13182
```

Distribución previa

Se ha escogido una distribución previa no informativa

```
alpha0.2 <- 0.5
beta0.2 <- 0.5
curve(dbeta(x, alpha0.2, beta0.2))
```

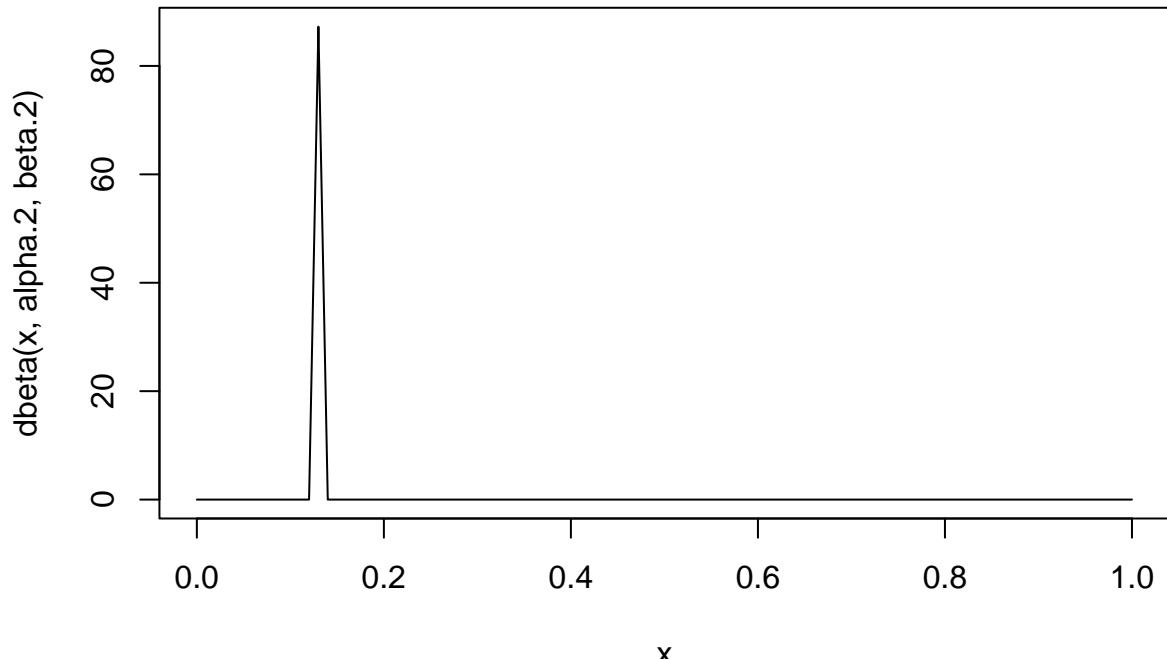


###

Distribución a posteriori de la proporción

```
alpha.2 <- alpha0.2 + exitos.2
beta.2 <- beta0.2 + n.2 - exitos.2

### Representación gráfica
curve(dbeta(x, alpha.2, beta.2))
```



```

### Esperanza

media.2 <- alpha.2 / (alpha.2 + beta.2)
varianza.2 <- media.2 * (1 - media.2) / (alpha.2 + beta.2 + 1)

### Intervalo de alta probabilidad

result <- data.frame(low=qbeta(0.025, alpha.2, beta.2),
                      media=media.2,
                      upp=qbeta(0.975, alpha.2, beta.2))
result

##           low      media      upp
## 1 0.1297339 0.1318237 0.1339274

```

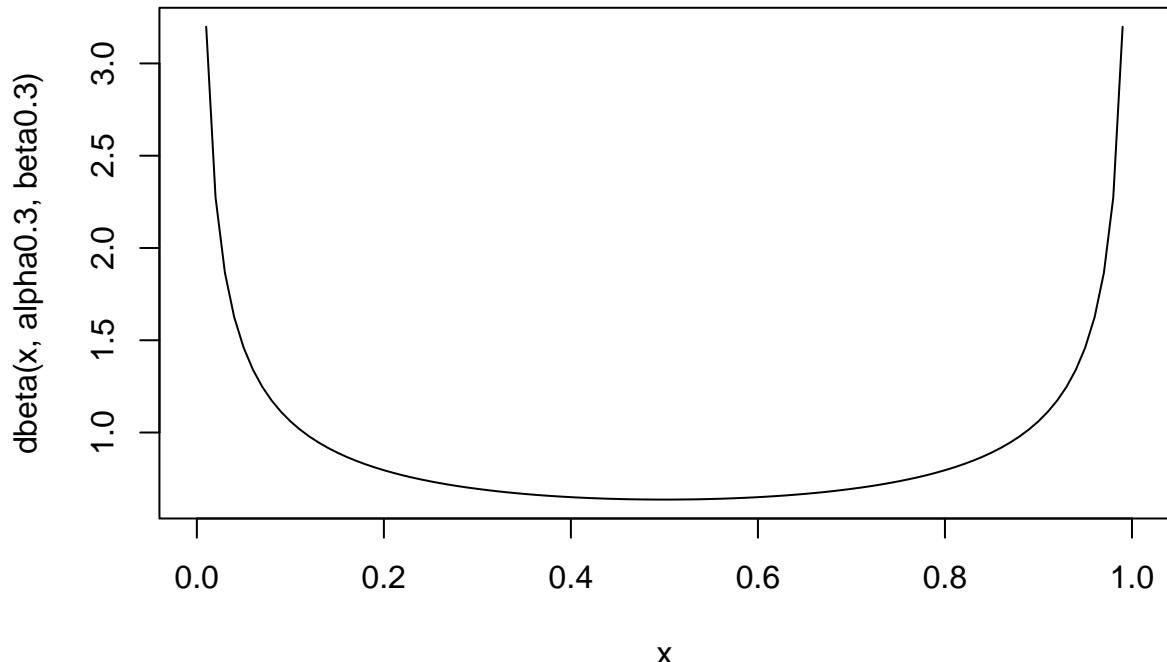
Inferecia sobre 3 atractores

```

n.3 <- nrow(atrac.3)
#& atrac.3$FD != Inf & atrac.3$FD != -Inf & atrac.3$L != -Inf & atrac.3$L != Inf
exitos.3 <- length(which(atrac.3$L <= -0.2 & atrac.3$FD >= 1 & atrac.3$FD != Inf & atrac.3$FD != -Inf))
exitos.3

## [1] 16967
alpha0.3 <- 0.5
beta0.3 <- 0.5
curve(dbeta(x, alpha0.3, beta0.3))

```

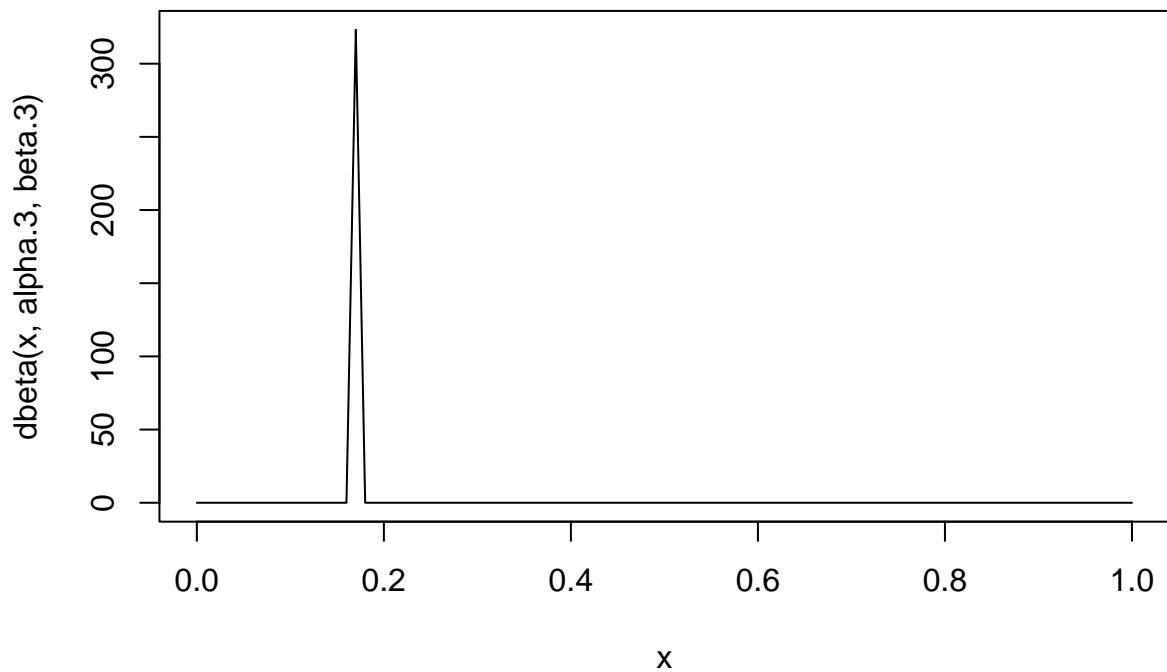


```

alpha.3 <- alpha0.3 + exitos.3
beta.3 <- beta0.3 + n.3 - exitos.3

### Representación gráfica
curve(dbeta(x, alpha.3, beta.3))

```



```

### Esperanza

media.3 <- alpha.3 / (alpha.3 + beta.3)
varianza.3 <- media.3*(1-media.3)/(alpha.3 + beta.3 + 1)

```

```
### Intervalo de alta probabilidad

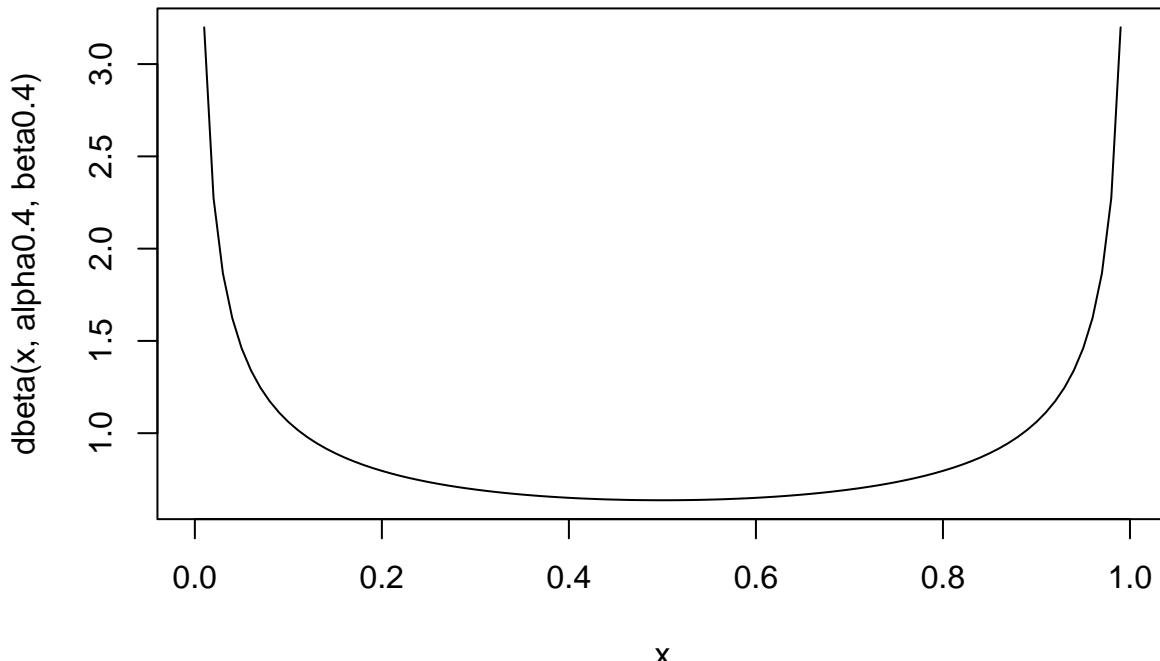
result <- data.frame(low=qbeta(0.025, alpha.3, beta.3),
                      media=media.3,
                      upp=qbeta(0.975, alpha.3, beta.3))
result

##      low     media      upp
## 1 0.1673532 0.1696733 0.1720059
```

Inferecia sobre 4 atractores

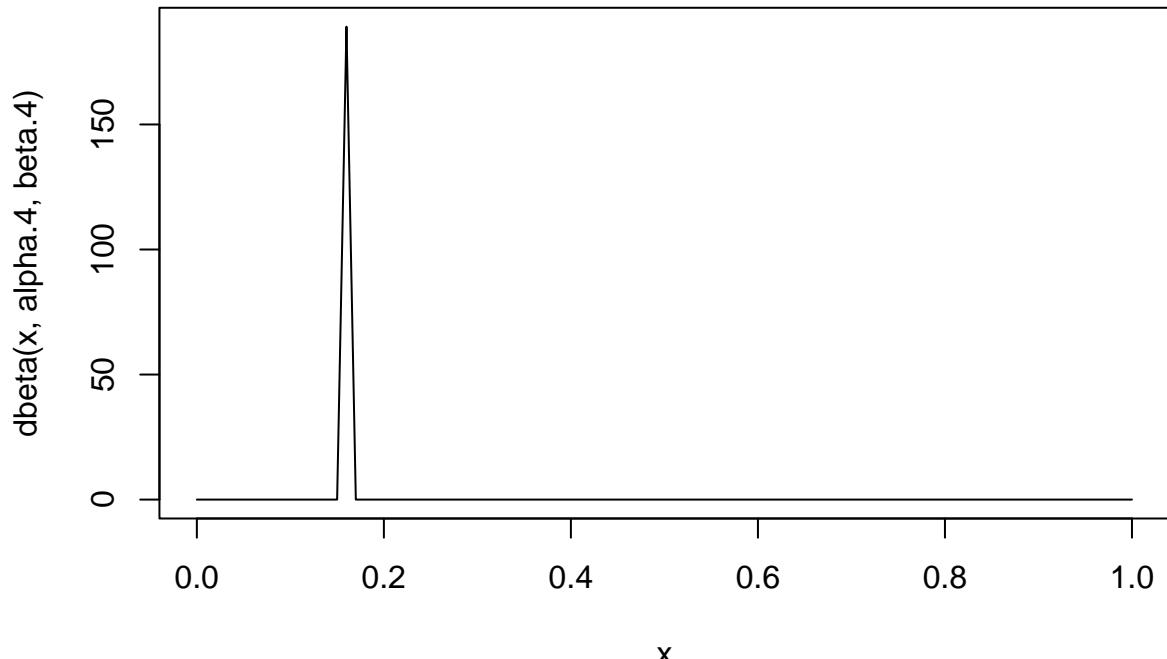
```
n.4 <- nrow(atrac.4)
## atrac.4$FD != Inf & atrac.4$FD != -Inf & atrac.4$L != -Inf & atrac.4$L != Inf
exitos.4 <- length(which(atrac.4$L <= -0.2 & atrac.4$FD >= 1 & atrac.4$FD != Inf & atrac.4$FD != -Inf))
exitos.4

## [1] 16127
alpha0.4 <- 0.5
beta0.4 <- 0.5
curve(dbeta(x, alpha0.4, beta0.4))
```



```
alpha.4 <- alpha0.4 + exitos.4
beta.4 <- beta0.4 + n.4 - exitos.4

### Representación gráfica
curve(dbeta(x, alpha.4, beta.4))
```



```
### Esperanza
```

```
media.4 <- alpha.4 / (alpha.4 + beta.4)
varianza.4 <- media.4 * (1 - media.4) / (alpha.4 + beta.4 + 1)
```

```
### Intervalo de alta probabilidad
```

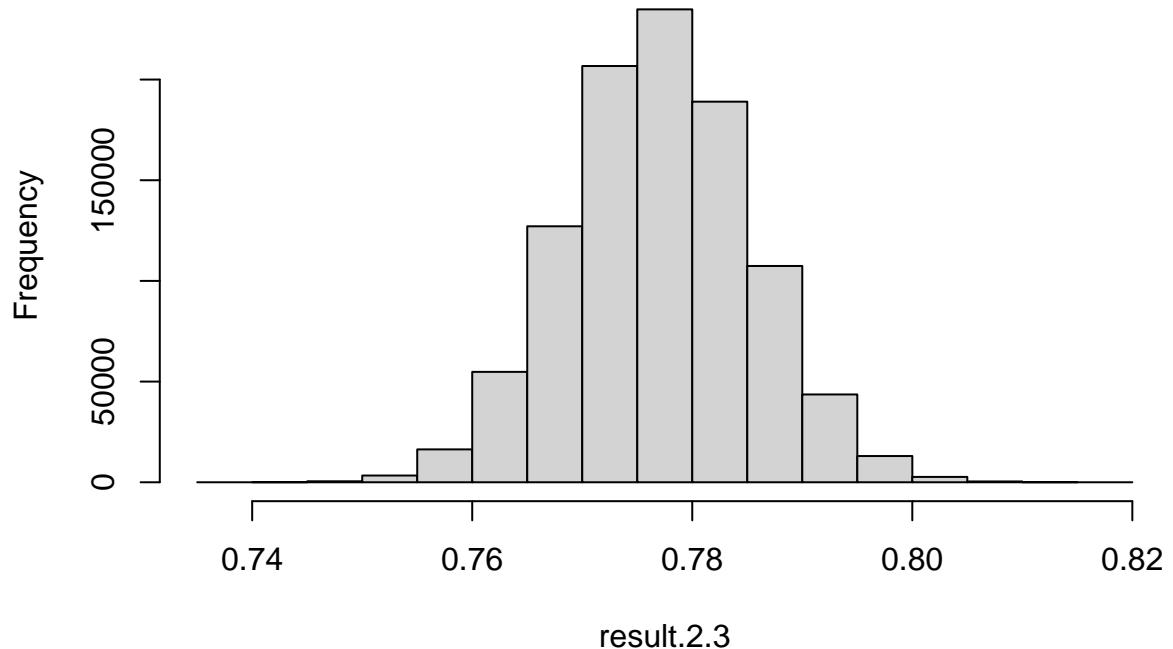
```
result <- data.frame(low=qbeta(0.025, alpha.4, beta.4),
                      media=media.4,
                      upp=qbeta(0.975, alpha.4, beta.4))
result
```

```
##           low      media      upp
## 1 0.1590003 0.1612734 0.1635593
```

inferencia sobre proporción 2 / 3 atractores

```
N <- 1000000
prop.2 <- rbeta(N, alpha.2, beta.2)
prop.3 <- rbeta(N, alpha.3, beta.3)
result.2.3 <- prop.2/prop.3
hist(result.2.3)
```

Histogram of result.2.3



```
mean(result.2.3)
```

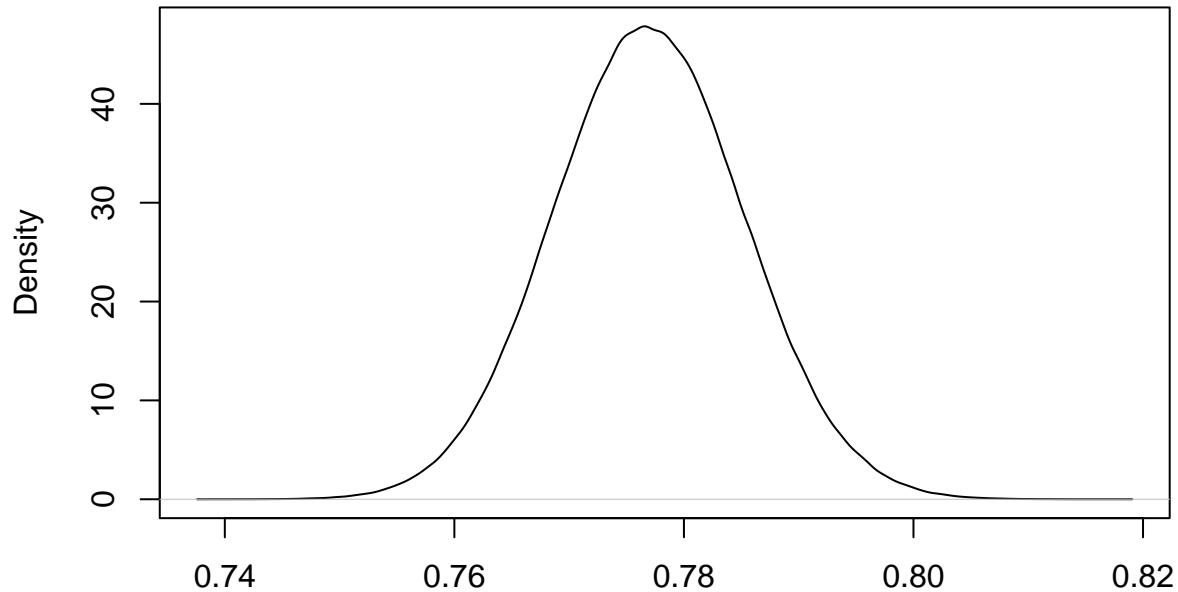
```
## [1] 0.7769601
```

```
var(result.2.3)
```

```
## [1] 6.949987e-05
```

```
plot(density(result.2.3))
```

density.default(x = result.2.3)



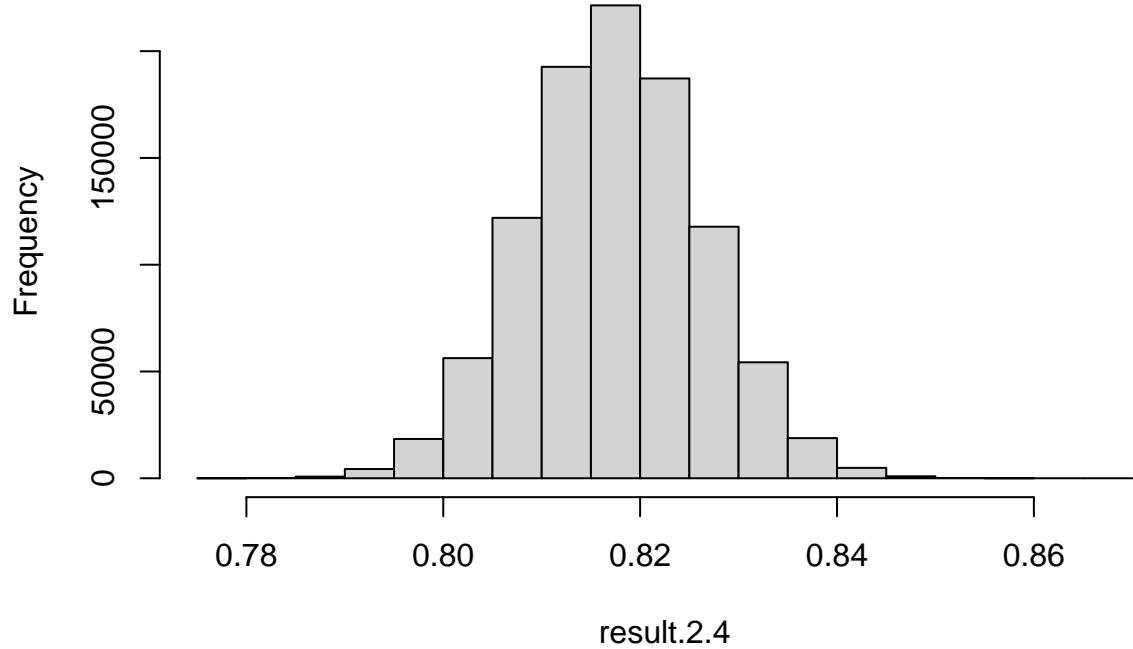
N = 1000000 Bandwidth = 0.0004734

```
confidence <- data.frame(low = quantile(result.2.3, c(0.025)),
                           media = mean(result.2.3),
                           upp = quantile(result.2.3, c(0.975)))
confidence
##          low      media       upp
## 2.5% 0.7607299 0.7769601 0.7934529
```

inferencia sobre proporción 2 / 4 atractores

```
N <- 1000000
prop.2 <- rbeta(N, alpha.2, beta.2)
prop.4 <- rbeta(N, alpha.4, beta.4)
result.2.4 <- prop.2/prop.4
hist(result.2.4)
```

Histogram of result.2.4



```
mean(result.2.4)
```

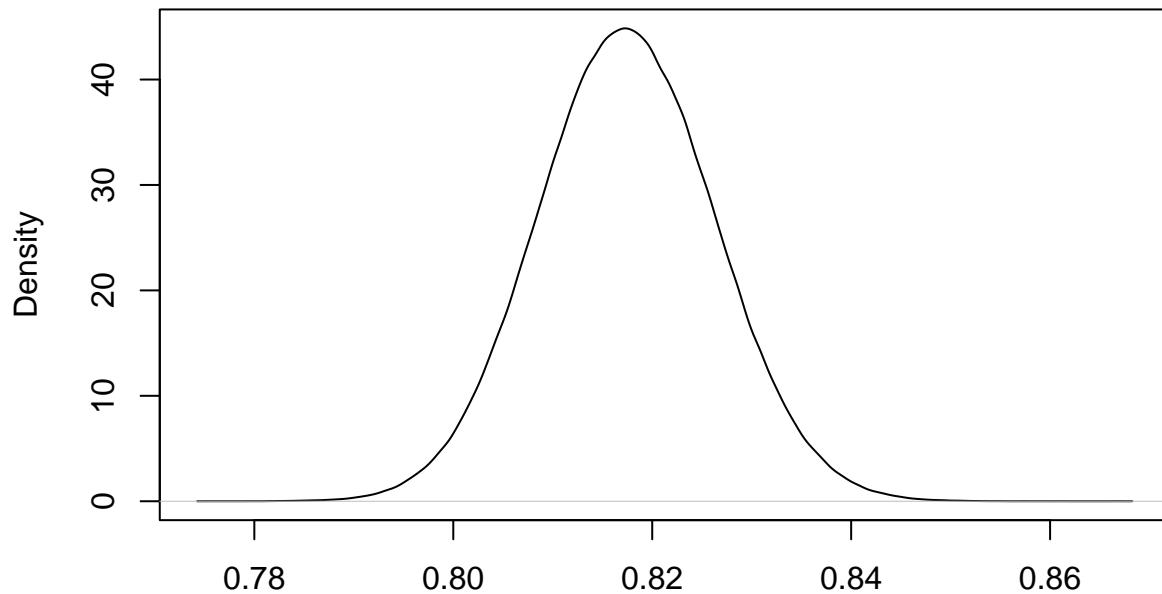
```
## [1] 0.817429
```

```
var(result.2.4)
```

```
## [1] 7.877092e-05
```

```
plot(density(result.2.4))
```

density.default(x = result.2.4)



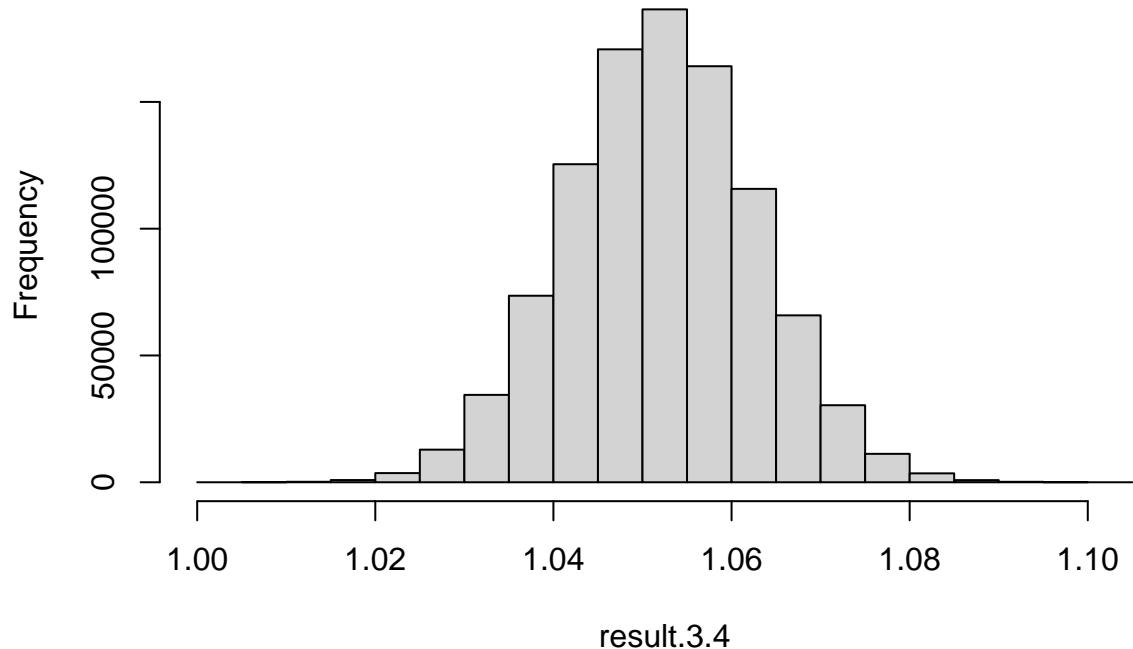
N = 1000000 Bandwidth = 0.000504

```
confidence <- data.frame(low = quantile(result.2.4, c(0.025)),
                           media = mean(result.2.4),
                           upp = quantile(result.2.4, c(0.975)))
confidence
##           low     media      upp
## 2.5% 0.8002143 0.817429 0.8349582
```

inferencia sobre proporción 3 / 4 atractores

```
N <- 1000000
prop.3 <- rbeta(N, alpha.3, beta.3)
prop.4 <- rbeta(N, alpha.4, beta.4)
result.3.4 <- prop.3/prop.4
hist(result.3.4)
```

Histogram of result.3.4



```
mean(result.3.4)
```

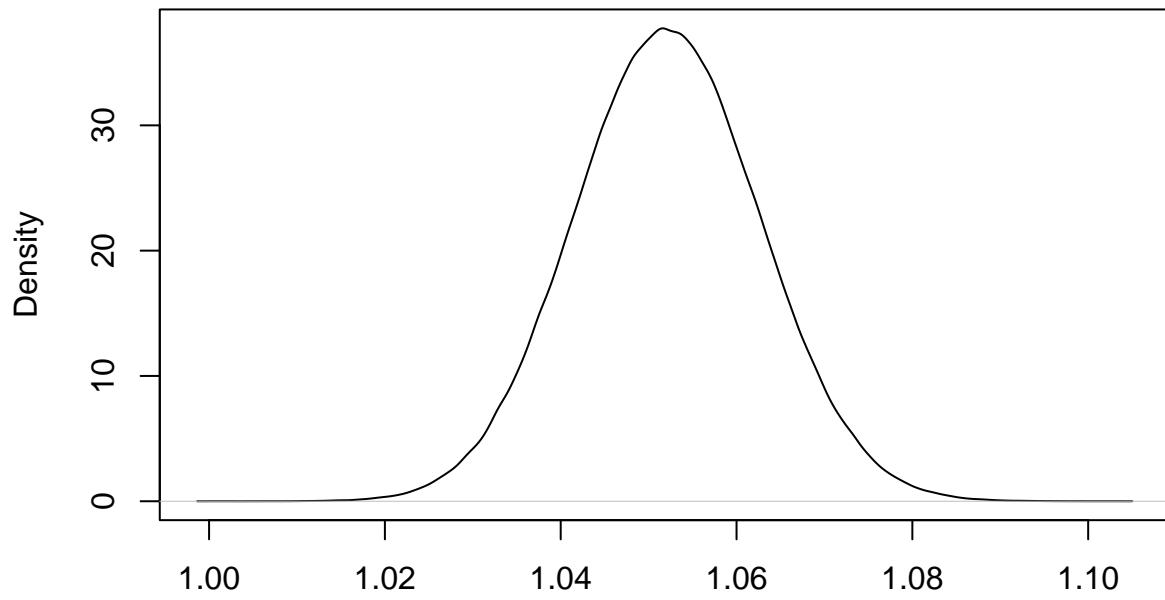
```
## [1] 1.052128
```

```
var(result.3.4)
```

```
## [1] 0.0001118195
```

```
plot(density(result.3.4))
```

density.default(x = result.3.4)



N = 1000000 Bandwidth = 0.0006005

```
confidence <- data.frame(low = quantile(result.3.4, c(0.025)),
                           media = mean(result.3.4),
                           upp = quantile(result.3.4, c(0.975)))
confidence
##           low     media      upp
## 2.5% 1.031565 1.052128 1.072997
```