SIMULACIÓN DE SISTEMAS
Dra. Elisa Schaeffer

# Reporte práctica 1: Movimiento Browniano

Eduardo Valdés
8 de agosto de 2017

## 1. INTRODUCCIÓN

Movimiento Browiano se refiere al movimiento de una partícula que cambia su posición uniformemente al azar. Los movimientos pueden ser de muchos distintos tipos, pero en esta práctica nos limitamos a un caso sencillo donde la particula mueve en pasos discretos, es decir, cada paso mide lo mismo, y las únicas posibles direcciones de movimiento son las direcciones paralelas a los ejes cardinales del sistema de coordenadas en el cual se realiza el movimiento. Vamos a utilizar pasos unitarios (es decir, el paso mide uno), teniendo como la posición inicial de la partícula el origen.

El objetivo de está práctica es examinar de manera sistemática como depende el número de veces que la particula regresa al origen de los siguientes tres factores:

- la dimensión sobre la que se mueve la partícula
- la duración de la caminata (número total de pasos)
- número de caminatas distintas que se realizan

## 2. DISEÑO DEL EXPERIMENTO

Para determinar el efecto que tienen la *dimensión*, la *duración* y número de *repeticiones* sobre el número de **cruces** que hace la particula al origen, se realiza un diseño de experimentos. El número de *repeticiones* funciona como las réplicas del experimento, aunque en

un sentido estricto no podríamos llamarlo así porque es uno de los factores a estudiar el efecto.

Para el experimento, la dimensión varía en $\{1,2,\ldots,8\}$; la duración de la caminata en $\{100,150,\cdots,400\}$ pasos y; las repeticiones en $\{50,100,\cdots,300\}$.

```
dimensiones<-1:8
replicas<-seq(50,300,50)
duraciones<-seq(100,400,50)
```

El experimento se realiza en R mediante las siguientes instrucciones:

```
cluster <- makeCluster(detectCores() - 1)
clusterExport(cluster, "experimento")
for(replica in replicas){
for (dimension in dimensiones) {
for (duracion in duraciones){
clusterExport(cluster, "replica")
clusterExport(cluster, "dimension")
clusterExport(cluster, "duracion")
resultado <- parSapply(cluster, 1:replica, experimento)
for(i in 1:replica){
datos <- rbind(datos, c(replica,dimension,duracion,resultado[i]))
temp<-c(temp,paste(replica,"_",dimension,"_",duracion))
}
}
}
}
stopCluster(cluster)

names(datos)<-c("repeticiones","dimension","duracion","cruces")
datos$repeticiones<-as.factor(datos$repeticiones)
datos$dimension<-as.factor(datos$dimension)
datos$duracion<-as.factor(datos$duracion)
datos$unidos<-as.factor(temp)
```

## 3. Pruebas de normalidad

El primer paso es realizar una prueba para comprobar si los datos provienen de una distribución normal. La prueba se realiza sobre los *residuales* de un ajuste lineal

```
lin<-lm(datos$cruces~datos$repeticiones+datos$dimension+datos$duracion)
residuales<-resid(lin)
```

Sin embargo; ninguna prueba de normalidad de R puede realizarse sobre un conjunto de observaciones mayor a 5000, por lo que se procede a hacer la prueba sobre una muestra.

```
muestra=datos[sample(nrow(datos),5000),]
linM=lm(muestra$cruces~muestra$repeticiones+muestra$dimension+muestra$duracion)
residualesM<-resid(linM)
```

Como apoyo, se graficaron sobrepuestas las densidades de los residuales sobre los datos originales y los de la muestra. La Figura 3.1 muestra que las densidades son muy similares.

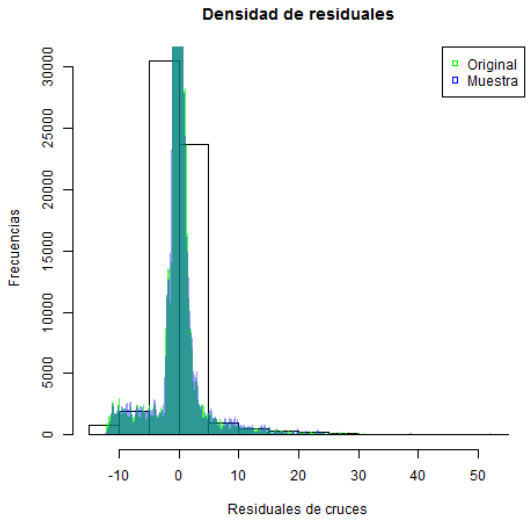



Figura 3.1: Densidad de residuales datos originales vs muestra

Además, se hizó una grafica cuantil-cuantil normal de la población y de la muestra. Como se aprecia en la figura 3.2, ninguna de las distribuciones se asemeja a la normal.

En base a las comparaciones de la muestra y la población con las Figuras 3.1 y 3.2 hacemos la prueba de normalidad de Shapiro sobre la muestra. Los resultados son los siguientes:

```
Shapiro-Wilk normality test

data:  residualesM
W = 0.6122, p-value < 2.2e-16
```

Por el valor del estadístico afirmamos que los datos no provienen de una distribución normal, por lo que tendremos que aplicar pruebas no paramétricas para determinar la significancia de los factores a estudio.

## 4. Prueba de Kruskal y Wallis

La primer prueba de Kruskal y Wallis que se realiza es el equivalente a la prueba ANOVA. Se desea determinar si las diferentes muestras de los datos provienen de la misma distribución; en este caso, una muestra es un conjunto de datos obtenidos con una terna (repeticiones, duración, dimensión), a la que llamaremos configuración. La hipótesis de la prueba es si existe diferencia significativa entre las muestras obtenidas por las diferentes configuraciones. Para esto representamos las configuraciones mediante la variable `datos$unidos`. Los resultados de la prueba son los siguientes:

```
Kruskal-Wallis rank sum test
```

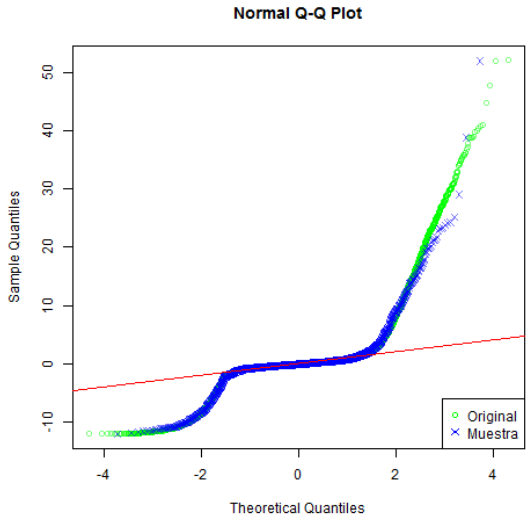



Figura 3.2: Gráfica cuantil-cuantil normal para las distribuciones de los datos originales y muestra

```
data:  datos$cruces by datos$unidos
Kruskal-Wallis chi-squared = 31121, df = 335, p-value < 2.2e-16
```

El valor *p* indica que si existe diferencia significativa entre las configuraciones; en otras palabras, al menos un factor esta provocando cambios significativos en el número de cruces en el origen. Averigüemos cuál es... a continuación se muestran los resultados de tres pruebas Kruskal y Wallis, una sobre cada factor. La hipotesis de las pruebas es si, para cada factor, existe diferencia significativa entre las muestras de cruces obtenidas por los niveles correspondientes.

## 4.1. PRUEBA PARA EL FACTOR *repeticiones*

```
Kruskal-Wallis rank sum test

data:  datos$cruces by datos$repeticiones
Kruskal-Wallis chi-squared = 1.8821, df = 5, p-value = 0.8652
```

Como puede observarse, no existe diferencia significativa entre los niveles del factor repeticiones. Por tanto podemos concluir que el número de repeticiones no afecta al cantidad de veces que la partícula cruza el origen.

## 4.2. PRUEBA PARA EL FACTOR *duración*

```
Kruskal-Wallis rank sum test
```

```
data:  datos$cruces by datos$duracion
Kruskal-Wallis chi-squared = 26.934, df = 6, p-value = 0.000149
```

Aquí el valor de $p$ es un tanto confuso, pero con un valor de significancia suficientemente discriminante podríamos decir que no existe diferencia significativa en los niveles. Esta aseveración no está mal fundamentada si consideramos los valores $p$ de las pruebas anteriores con los que hemos aceptado las pruebas de hipótesis. El diagrama de bigotes para este factor (vease Figura 4.1) nos da una idea visual de la similitud de los cruces para los diferentes niveles de este factor. Por tanto, consideramos que la duración de la caminata tampoco influye sobre

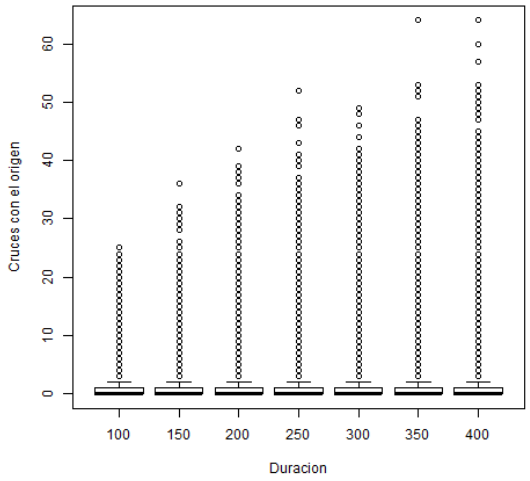



Figura 4.1: Diagrama de bigotes del número de cruces en dependencia de la duración de la caminata

el número de cruces con el origen.

## 4.3. Prueba para el factor *dimensión*

```
Kruskal-Wallis rank sum test

data:  datos$cruces by datos$dimension
Kruskal-Wallis chi-squared = 30882, df = 7, p-value < 2.2e-16
```

Recalcando los valore $p$ con que se han aceptado las preubas anteriores, aceptamos la hipotesis de que existe diferencia significativa entre los niveles del factor *dimensión*. Como intuitivamente era de esperarse, entre mas dimensiones tenga la partícula, más difícil es regresar al origen. Como muestra, el diagrama de bigotes correspondiente (véase Figura 4.2), refleja como a partir de la cuarta dimensión la probabilidad de regresar al origen es prácticamente nula.

La última prueba va encaminada a determinar una transición de fase del número de cruces

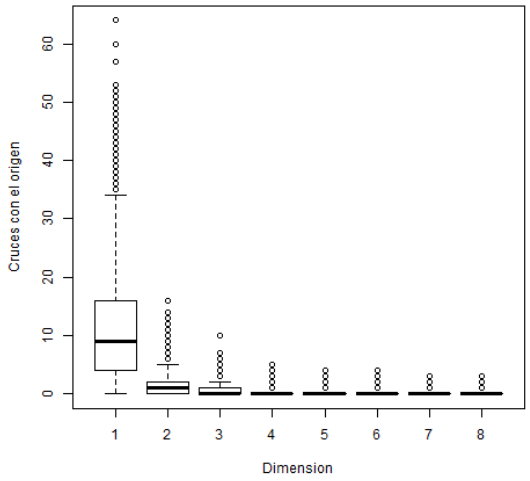



Figura 4.2: Diagrama de bigotes del número de cruces al origen para el factor dimensión.

en dependencia de su dimensión, para esto realizamos múltiples pruebas a pares para determinar si existe diferencia significativa entre cada par de dimensiones. La prueba que se realiza tiene el nombre de prueba de Dunn[1], realizada mediante los siguientes comandos:

```
library(FSA)
PT = dunnTest(datos$cruces~datos$dimension,data=datos,
method="bh")
PT = PT$res
```

La variable PT contiene los valores $p$ (p.adj en este caso) de cada prueba. Los pares que que tienen diferencia estadística en el número de cruces son

| Comparison | Z | P.unadj | P.adj |
|---|---|---|---|
| 1 - 2 | 64.008420 | 0.000000e+00 | 0.000000e+00 |
| 1 - 3 | 105.084461 | 0.000000e+00 | 0.000000e+00 |
| 2 - 3 | 41.076042 | 0.000000e+00 | 0.000000e+00 |
| 1 - 4 | 120.105302 | 0.000000e+00 | 0.000000e+00 |
| 2 - 4 | 56.096882 | 0.000000e+00 | 0.000000e+00 |
| 3 - 4 | 15.020841 | 5.362324e-51 | 8.341393e-51 |
| 1 - 5 | 126.347287 | 0.000000e+00 | 0.000000e+00 |
| 2 - 5 | 62.338868 | 0.000000e+00 | 0.000000e+00 |
| 3 - 5 | 21.262826 | 2.508464e-100 | 4.131588e-100 |
| 4 - 5 | 6.241985 | 4.320514e-10 | 5.259756e-10 |
| 1 - 6 | 129.905180 | 0.000000e+00 | 0.000000e+00 |
| 2 - 6 | 65.896760 | 0.000000e+00 | 0.000000e+00 |
| 3 - 6 | 24.820718 | 5.357283e-136 | 9.375245e-136 |
| 4 - 6 | 9.799878 | 1.127222e-22 | 1.502962e-22 |
| 1 - 7 | 132.092610 | 0.000000e+00 | 0.000000e+00 |

[1] Idea tomada de http://rcompanion.org/rcompanion/d_06.html

```
2 - 7  68.084191  0.000000e+00  0.000000e+00
3 - 7  27.008149 1.185616e-160 2.213149e-160
4 - 7  11.987308  4.141452e-33  5.798032e-33
5 - 7   5.745323  9.174574e-09  1.070367e-08
1 - 8 133.346410  0.000000e+00  0.000000e+00
2 - 8  69.337990  0.000000e+00  0.000000e+00
3 - 8  28.261949 1.015074e-175 2.030148e-175
4 - 8  13.241108  5.079388e-40  7.485414e-40
5 - 8   6.999123  2.575703e-12  3.278168e-12
```

y los no significativos son:

```
Comparison         Z       P.unadj         P.adj
15       5 - 6 3.557892 0.0003738427 0.0004187038
21       6 - 7 2.187431 0.0287110976 0.0297744716
27       6 - 8 3.441230 0.0005790751 0.0006236194
28       7 - 8 1.253800 0.2099148022 0.2099148022
```

Como puede observar, el valor de significancia es como digo discriminante, (0.00001). Sin embargo; es fácil darse cuenta de que existe una diferencia muy grande entre los valores $p$ de las pruebas que son aceptadas y las que no.
Podemos concluir que estadísticamente las dimensiones 5,6,7 y 8 son equivalentes, más aún, la mediana de cruces con el origen es 0. Es decir, existe una transición de fase a partir de la quinta dimensión.

## 5. Conclusiones

Los resultados obtenidos mediante pruebas no parametricas para detrminar si existe un efecto del número de repeticiones, la duración de la caminata y la dimensión de la particula sobre el numero de cruces con el origen son los siguientes:

- Sólo la dimensión afecta la cantidad de cruces con el origen
- A partir de la quinta dimensión, el número de cruces es cero

## 6. Reto 1

En el script `Reto1.R` se estudia de forma sistemática y automatizada el tiempo de ejecución de una caminata en términos del largo de la caminata y la dimensión.

El experimento tiene como factores la *dimensión* de la partícula y la *duración* de la caminata. La variable de respuesta es el **tiempo** de simulación de cada caminata. Se realizaron 200 replicas, los niveles del factor dimensión son $\{1,2,\cdots,8\}$ y del factor duración $\{100,150,\cdots,500\}$.
El esquema general a seguir para comprobar el efecto de estos dos factores sobre el tiempo es el siguiente:

```
Realizar el experimento
/* Revisar si los datos son normales                                    */
if numero de datos es mayor a 5000 then
    realizar prueba Shapiro sobre muestra
else
    Realizar prueba Shapiro sobre datos originales
end
if los datos son normales then
    Realizar prueba ANOVA
    if algún factor presenta significancia then
        Realizar prueba de Tukey para cada par de niveles
        Concluir niveles significativos
    end
else
    Realizar prueba Kruskall y Wallis para las configuraciones (dimension,duracion) if
     Existe diferencia significativa entre las configuraciones then
        Realizar prueba Kruskall y Wallis para cada factorif algun factor presenta
         significancia then
            Realizar prueba Dunn para cada par de niveles
            Concluir niveles significativos
        end
    end
end
```

De forma automática se ofrece como resultado que factores y que niveles presentan significancia, además de gráficas de apoyo, en dependencia del tipo de prueba que se realice.

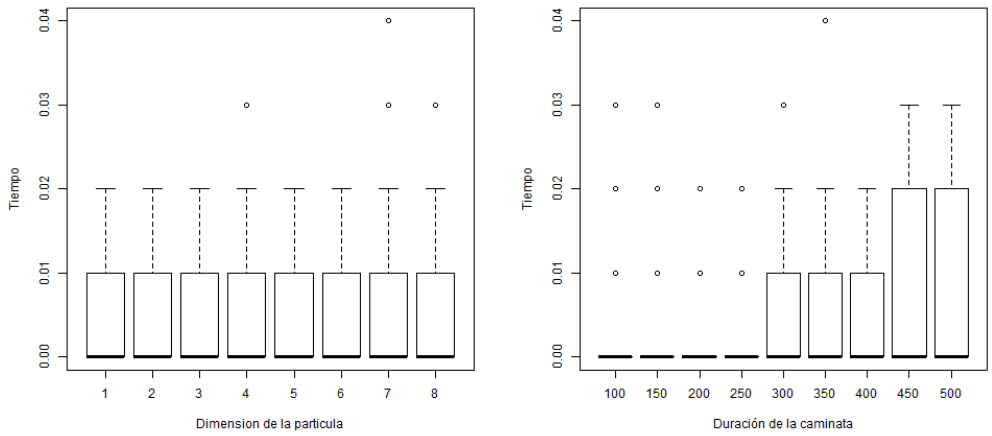
En particular, los resultados obtenidos para el experimento realizado son:

```
El factor dimensión no es significativo; es decir, el tiempo de ejecución no depende
de la dimensión de la particula

El factor duracion es significativo; es decir, el tiempo de ejecución depende
de la duración de la caminata

Los pares con diferencia son:"
"200-100" "250-100" "300-100" "350-100" "400-100" "450-100" "500-100" "300-150" "350-150"
"450-150" "500-150" "300-200" "350-200" "400-200" "450-200" "500-200" "350-250" "400-250"
"500-250" "400-300" "450-300" "500-300" "450-350" "500-350" "500-400"
```

Los diagramas de bigotes obtenidos de cada factor aparecen en la Figura 6.1

## 7. RETO 2

Para comparar una implementación que aprovecha o no paralelismos se realizó un experimento donde se fijó la dimensión de la partícula. Este factor sabemos por el experimento




(a) Dimensión (b) Duración

Figura 6.1: Diagramas de bigotes del tiempo de simulación contra los factores de dimensión y duración

anterior que no es significativo en el tiempo. Por el contrario, la duración de la caminata si lo es; sin embargo fue fijada en 300 para observar el efecto que tiene un enfoque paralelo contra uno secuencial. Para cada enfoque, se realizaron 100, 200, ..., 1000 caminatas. Para cada uno de estos valores se hicieron 10 replicas.

La Figura 7.1 muestra el diagrama de bigotes para ambos enfoques, donse se observa claramente la eficiencia de utilizar paralelismo

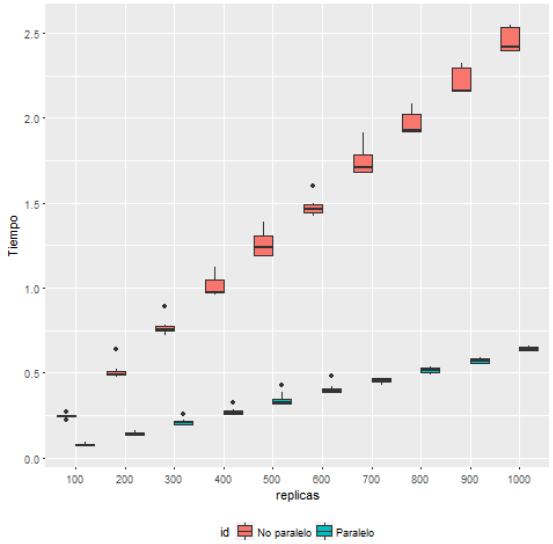



Figura 7.1: Diagrama de bigotes implementación paralela vs no paralela