

PRÁCTICA 8: MODELO DE URNAS

FABIOLA VÁZQUEZ

11 de noviembre de 2020

1. Introducción

Se tiene una simulación sobre los fenómenos de coalescencia y fragmetación de partículas que forman cúmulos. El objetivo de esta práctica [4] es determinar el porcentaje de cúmulos que pueden ser filtrados si tienen un tamaño mayor o igual que un valor específico c , el cual es la media del tamaño de los cúmulos iniciales. El experimento se lleva a cabo en el software R [2] en un cuaderno de Jupyter [1].

2. Experimento

Para el experimento, se consideran los parámetros k , que es la cantidad de cúmulos que se tiene, n la cantidad de partículas y `duracion` que es la cantidad de pasos que se simula. El parámetro k varía su valor en $\{100, 1000, 10000\}$ y n , que se toma como un múltiplo de k , varía su valor en $\{10k, 20k, 50k, 100k, 1000k\}$ y el parámetro `duracion` varía en $\{10, 25, 50, 75, 100\}$.

Al código elaborado por Schaeffer [3], se le añade el código 1 para determinar el porcentaje de cúmulos que tienen un tamaño mayor o igual c y que pueden ser filtrados.

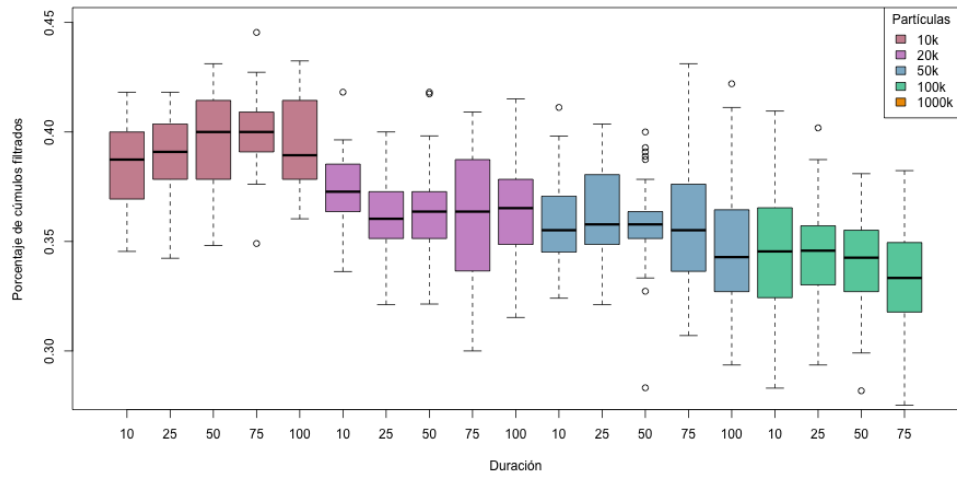
```
for (i in 1:length(freq$num)) {  
  if (freq$tam[i] >= c) {  
    cum <- c(cum, freq$num[i])  
  }  
}  
a <- sum(cum) / sum(freq$num)
```

Código 1: Cálculo del porcentaje de cúmulos filtrados.

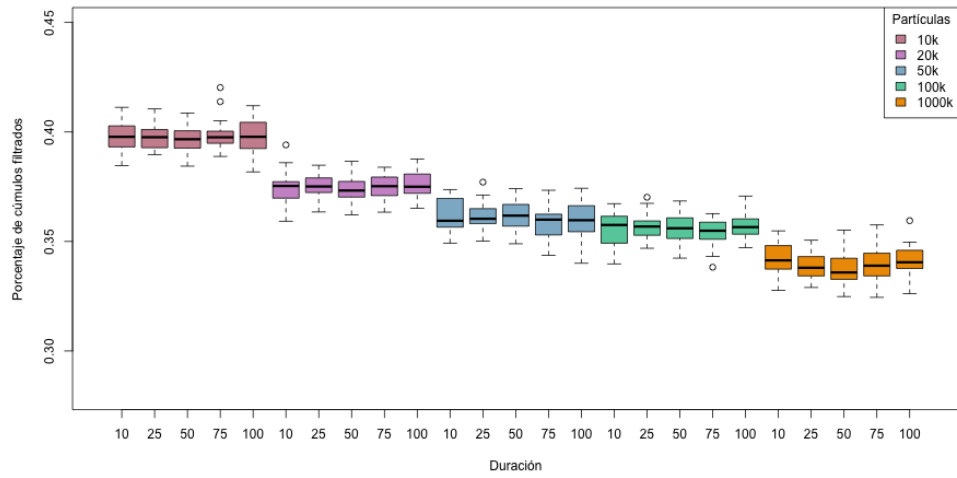
En la figura 1 se muestran las gráficas de caja-bigote de las diferentes combinaciones de los parámetros k , n y `duracion`. En la figura 1a se omitió la combinación de $k=100$ cúmulos con $n=1000k$ partículas, ya que el porcentaje de cúmulos filtrados en dicha combinación es de cero para los cinco valores de `duracion`. Como se aprecia en la figura 1 para los diferentes valores de k , la diferencia es que las cajas se hacen más pequeñas pero oscilan entre los mismos valores.

2.1. Pruebas de correlación

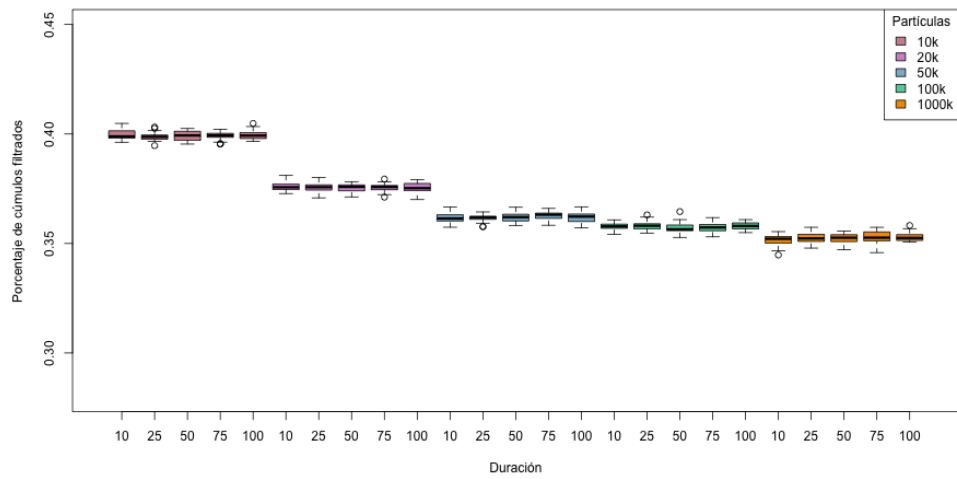
Para comprobar si alguno de los parámetros influye en la cantidad de cúmulos filtrados se realizan pruebas de correlación. El cuadro 1 muestra los valores p obtenidos en las pruebas, de los cuales solo



(a) Gráficas de caja-bigote con $k = 100$.



(b) Gráficas de caja-bigote con $k = 1000$.



(c) Gráficas de caja-bigote con $k = 10000$.

Figura 1: Gráficas de caja.

Cuadro 1: Resultados de las pruebas de correlación.

Parámetro	valor p	Coeficiente de correlación
Cúmulos k	2.2×10^{-16}	0.2310
Partículas n	0.1809	0.0309
duracion	0.9211	-0.0022

el del parámetro **k** es menor que 0.05, es decir que existe correlación entre la cantidad de cúmulos y el porcentaje de cúmulos filtrados, pero dado que el coeficiente de correlación es 0.2310 la correlación que existe es débil. Los parámetros **n** y **duracion** no afectan al porcentaje de cúmulos filtrados.

3. Reto 1

En el reto 1 se pide determinar en cuál iteración el porcentaje de cúmulos filtrados alcanza un punto máximo, para esto se modifica la función añadiendo el código 2.

```
cum <- c()
for (i in 1:length(freq$num)) {
  if (freq$tam[i] >= c) {
    cum <- c(cum, freq$num[i])
  }
}
a <- sum(cum) / sum(freq$num)
porcentajes[paso] <- a
}
return(which.max(porcentajes))
```

Código 2: Iteración en la cuál el porcentaje de cúmulos filtrados es máximo.

Se realiza el experimento variando el valor de **k** en {100, 1000, 100000} y **n** en {10k, 100k, 1000k} y se fija el parámetro **duracion** en 50.

Referencias

- [1] Thomas Kluyver, Benjamin Ragan-Kelley, Pérez, et al. Jupyter notebooks—a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing*, page 87. IOS Press, 2016.
- [2] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
- [3] Elisa Schaeffer. aggrFrag.R. <https://github.com/satuelisa/Simulation/blob/master/UrnModel/aggrFrag.R>.
- [4] Elisa Schaeffer. Práctica 8: modelo de urnas. <https://elisa.dyndns-web.com/teaching/comp/par/p8.html>.

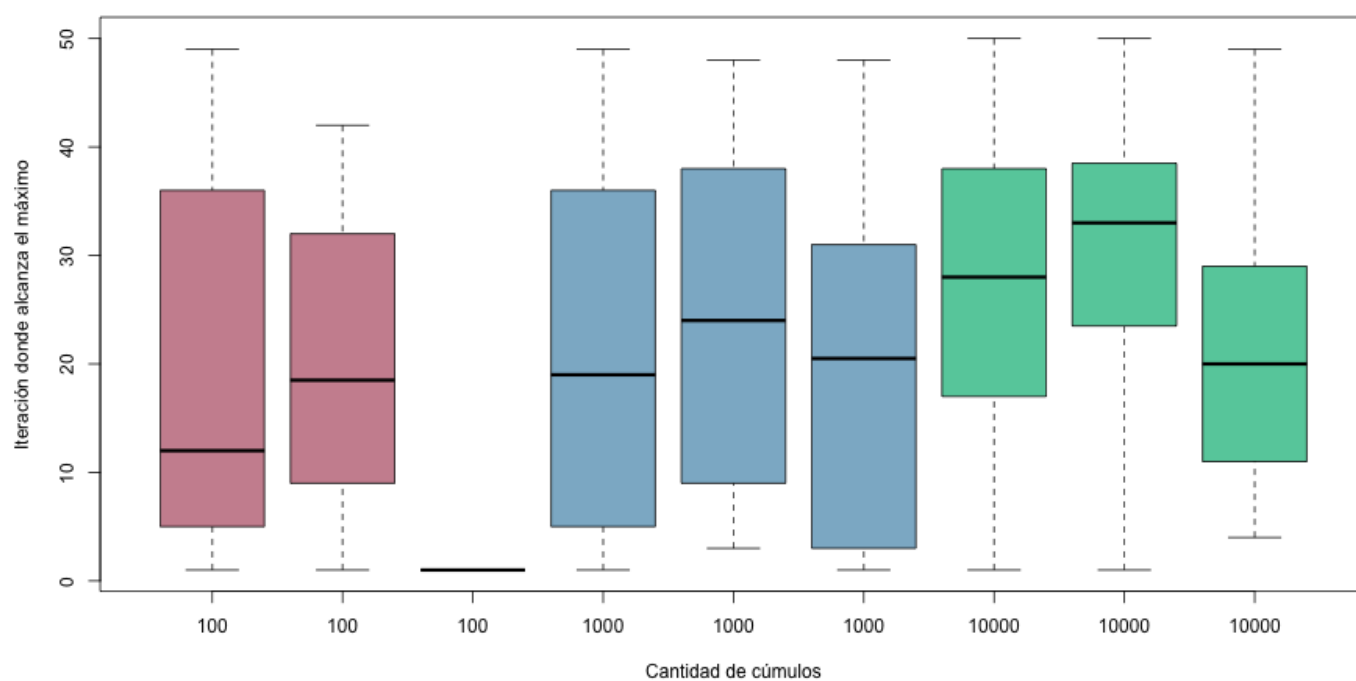


Figura 2: Gráficas de caja-bigote.