

# Práctica 8: Modelo de Urnas

2 de octubre de 2017

## 1. Tarea

El objetivo de la tarea es paralelizar lo que sea posible y eficiente, y medir cuanto tiempo de computo se logro mejorar.

### 1.1. Descripción del experimento

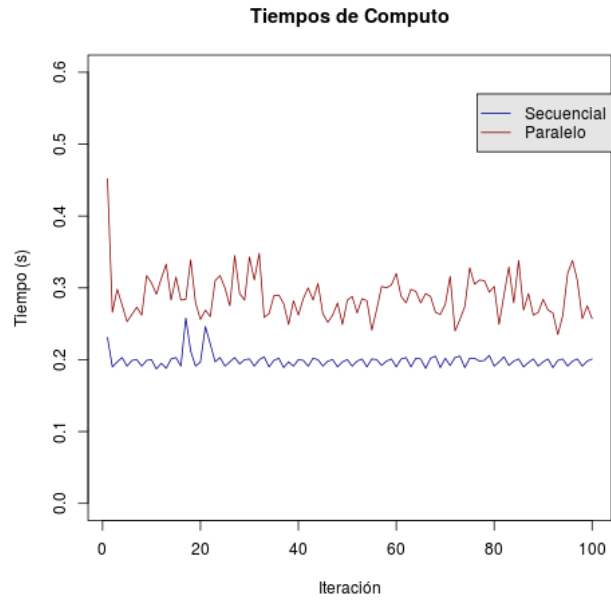
Para la paralelización del experimento se uso la librería *doParallel* y se crearon dos funciones, una función de que itera la fase de unión y una que itera la fase de fragmentación.

Se vario la cantidad de partículas 1,000,000 hasta 12,000,000 y el numero de cúmulos de 10,000 hasta 120,000, también se aumento la duración del experimento a cien replicas.

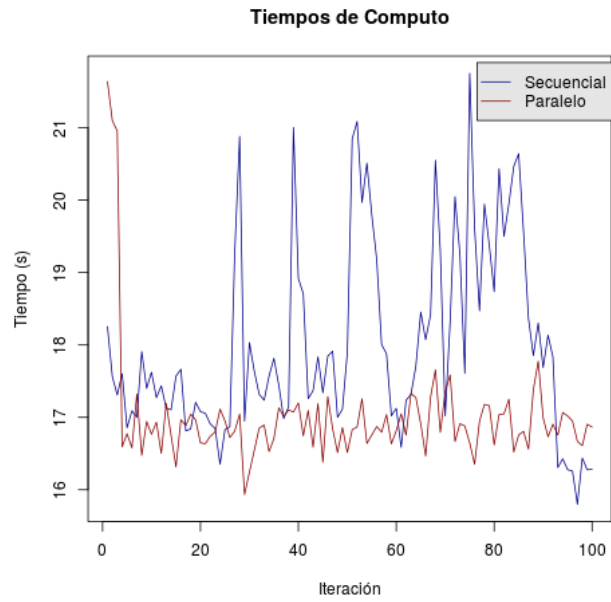
El experimento se ejecuto en se ejecutó en una laptop Lenovo Procesador: Core i5, Memoria RAM: 8Gb, Núcleos: 4.

### 1.2. Resultados

Como se puede observar en las gráficas de la Figura2 el tiempo de computo es menor cuando se realiza de forma paralela, sin embargo esto solo se da cuando los valores de la cantidad de cúmulos supera los 100,000 y las partículas los 10,000,000, antes de eso la forma secuencial dio mejores tiempos.

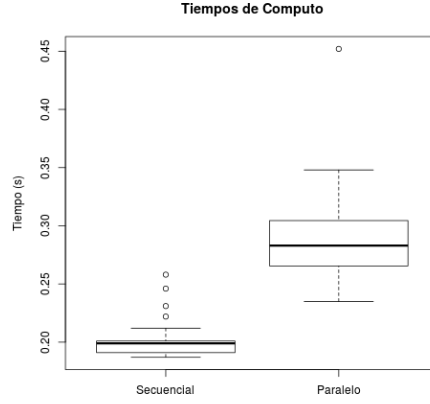


(a) cúmulos: 10,000, partículas: 1,000,000

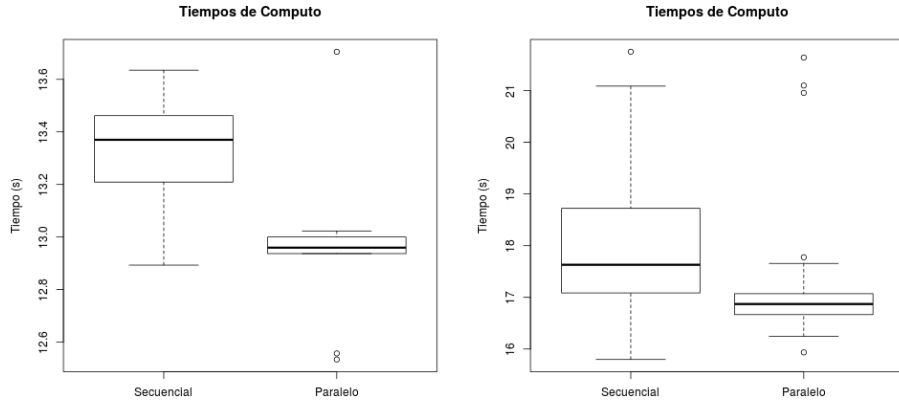


(b) cúmulos: 120,000, partículas: 12,000,000

Figura 1: Tiempos de computo durante las 100 iteraciones.



(a) cúmulos: 10,000, partículas: 1,000,000



(b) cúmulos: 100,000, partículas: 10,000,000    (c) cúmulos: 120,000, partículas: 12,000,000

Figura 2: Visualización de los tiempos de computo obtenidos en el experimento de forma secuencial y paralela con diferente numero de cúmulos y partículas.

## 2. Reto 1

El objetivo del reto 1 es variar el numero de cúmulos y de partículas donde la cantidad de partículas es treinta veces mas que la cantidad de cúmulos y demostrar que el tiempo ahorrado de forma paralela es significativo mediante una prueba estadística.

## 2.1. Descripción del experimento

Para el experimento se uso el mismo código de la tarea solo se agrego un ciclo que fuera variando el valor de los cúmulos y de las partículas.

Se generaron dos matrices de tiempos, uno de manera secuencial y uno de forma paralelo.

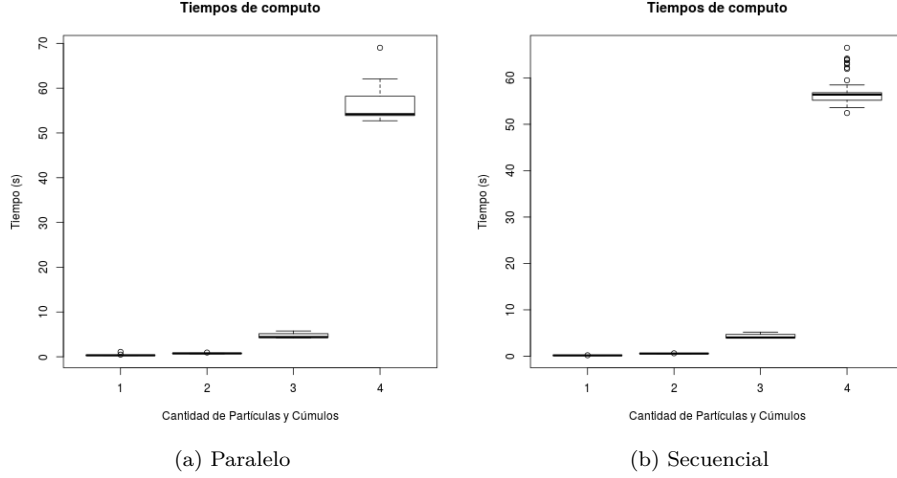


Figura 3: Tiempos de computo durante las 50 iteraciones, cada uno de los números del 1 – 4 representan diferentes valores de cúmulos iniciando en 10,000 hasta 40,000 y las partículas iniciando en 300,000 hasta 1,200,000.

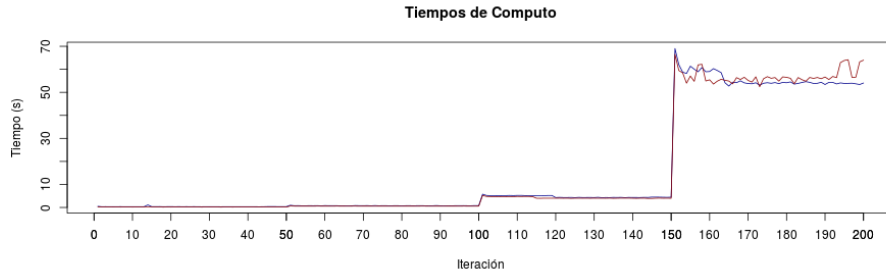


Figura 4: Tiempos de computo para diferentes tamaños de cúmulos y partículas de forma paralela (*azul*) y de manera secuencial (*rojo*), los tamaños cambian cada cincuenta iteraciones multiplicando el valor inicial de cúmulos y partículas por 1, 2, 3, y 4 iniciando en 10,000 el numero de cúmulos y en 300,000 el numero de partículas.

### 2.1.1. Pruebas de Normalidad

Se valida si las muestras son normales o no para poder decidir que prueba utilizar.

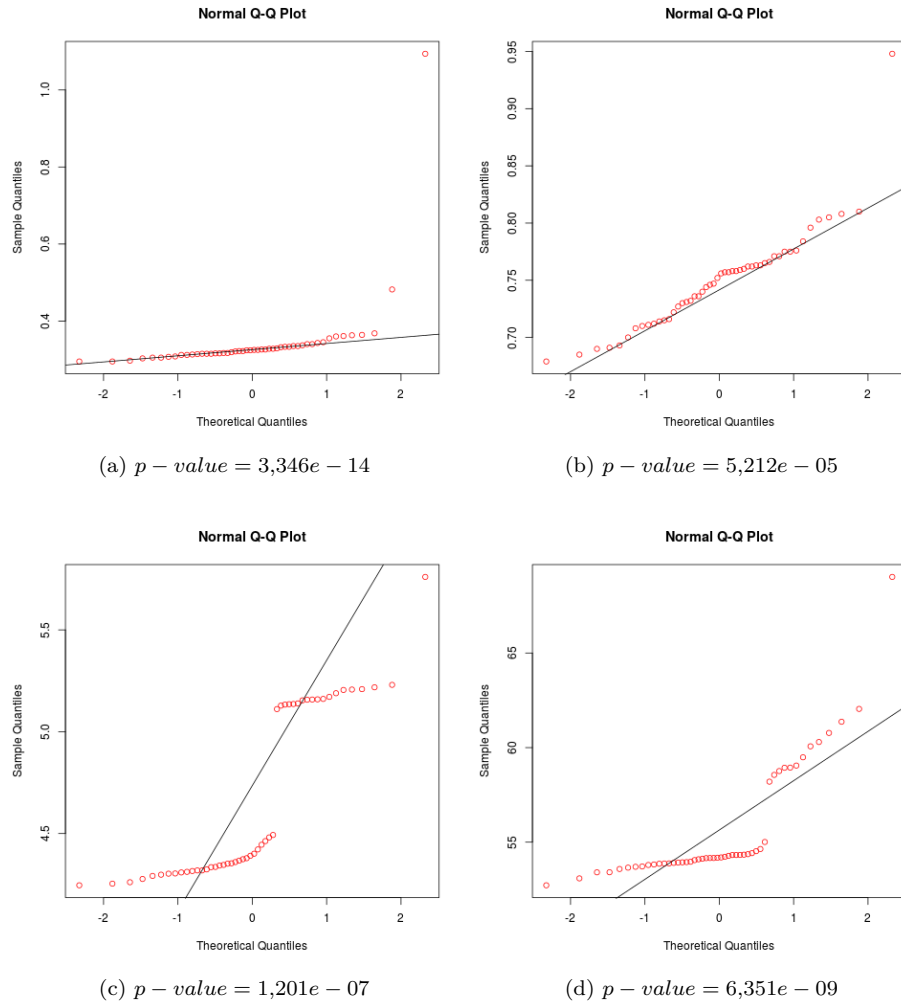


Figura 5: Visualización de los datos obtenidos de forma paralelo para ver si son normales.

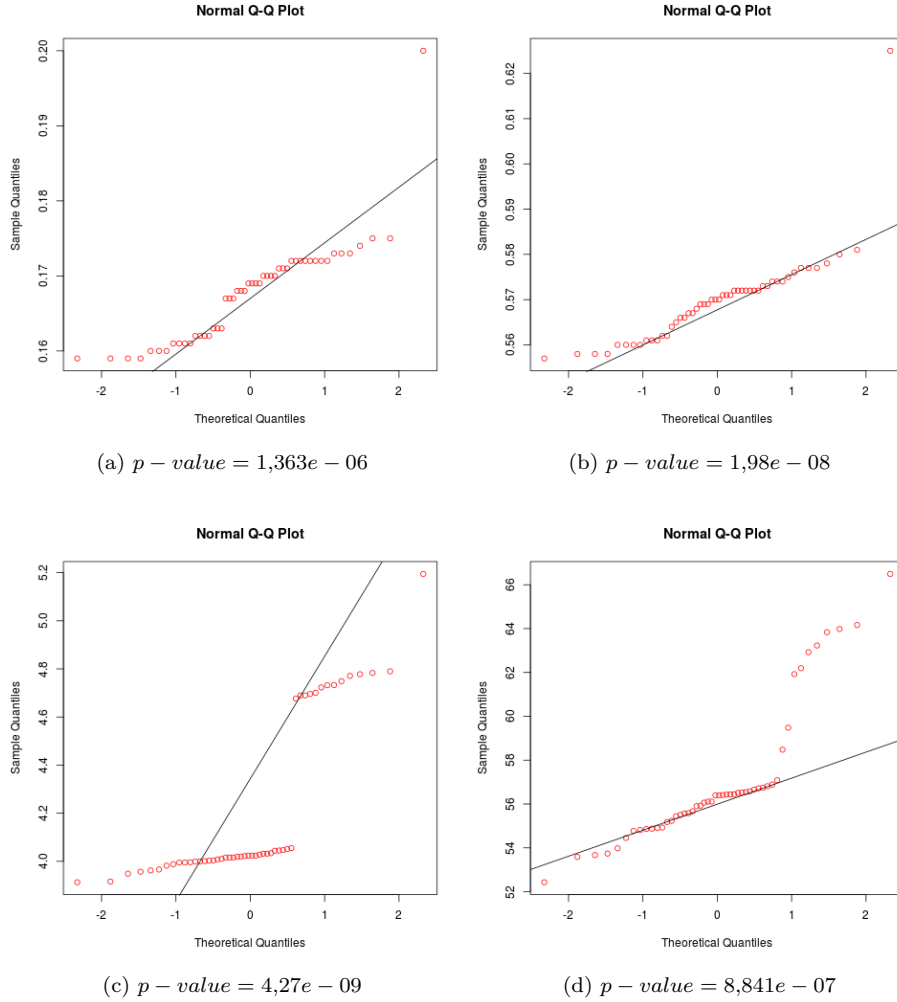


Figura 6: Visualización de los datos obtenidos de manera secuencial para ver si son normales.

Como se puede observar los datos no siguen una distribución normal por lo que se opta por pruebas no paramétricas.

## 2.2. Resultados

Para la comparación de medias, como los datos no siguen una distribución normal se optó por usar una prueba no paramétrica, en específico se usó la prueba de *kruskal Wallis* ya que nos permite usar datos no normales y poder comparar medias de más de dos muestras.

Las hipótesis de la prueba de *Kruskal Wallis* son:

- $H_0$ : Las medias son iguales.
- $H_1$ : Las medias son distintas.

Kruskal–Wallis

chi-squared = 182.2      df = 3      p-value < 2.2e-16

Como el valor es menor a 0,05 se rechaza la hipótesis nula y se asume que el realizar el experimento de forma paralela si mejora el tiempo de computo.