Práctica 8: Modelo de urnas

1445183

26 de marzo de 2019

1. Objetivo

Paralelizar el código proporcionado [1] y medir el tiempo que se ahorra con la paralelización, observar si el ahorro es estadísticamente significativo para diferentes combinaciones de k y n, donde k es tamaño de cúmulo y n número de partículas.

2. Descripción

Para medir el tiempo se hace uso de `Sys.time()`, después se hacen vectores para los diferentes valores de k y n con 30 réplicas, recopilando los datos obtenidos en un `data.frame` llamado `resultado`:

```
resultado<-data.frame()
k <- 10000
n <- 1000000
#cumulo <- c(1000, 10000)
#particula <- c(1000000, 10000000, 100000000)
#for (k in cumulo) {
# for (n in particula) {
for (replicas in 1:30) {
  inicial<- Sys.time()
  originales <- rnorm(k)
  cumulos <- originales - min(originales) + 1
  cumulos <- round(n * cumulos / sum(cumulos))
  assert(min(cumulos) > 0)
  diferencia <- n - sum(cumulos)
  if (diferencia > 0) {
    for (i in 1:diferencia) {
      p <- sample(1:k, 1)
      cumulos[p] <- cumulos[p] + 1
    }
  } else if (diferencia < 0) {
    for (i in 1:-diferencia) {
      p <- sample(1:k, 1)
      if (cumulos[p] > 1) {
        cumulos[p] <- cumulos[p] - 1
      }
    }
  }
```

Para la prueba estadística se hace uso de `qqplot` para normalizar los valores.

.

3. Resultados

Se puede ver en la figura 1 la diferencia de tiempo usando los valores dados por la práctica de k y n siendo el código paralelizado el que presenta menor tiempo de compilación.

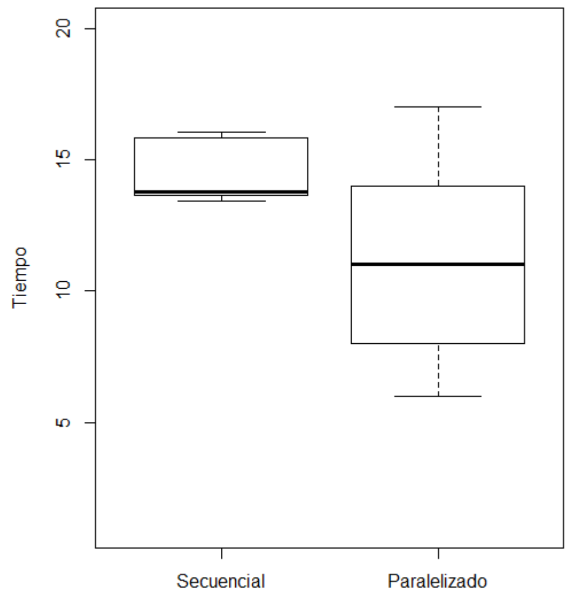


Figura 1: comparación de tiempos

En la figura 2 se puede observar que el tiempo es menor cuando se tiene una relación de k y n con los valores más bajos, las combinaciones de k y n se proporcionan en el cuadro 1.

Cuadro 1: Combinaciones de k y n

Combinación	k	n
1	1000	1000000
2	1000	10000000
3	1000	100000000
4	10000	1000000
5	10000	10000000
6	10000	100000000

En la figura 3 se observa la normalización del experimento paralelizado, cúmulos (k) de 1000 y 10000 con las 3 combinaciones de n correspondientes. Donde la relación de k y n con valores mayores dan una mejor normalización.

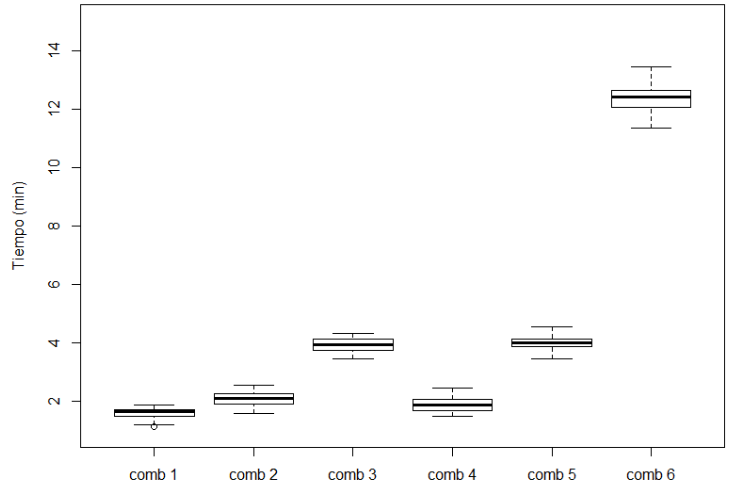
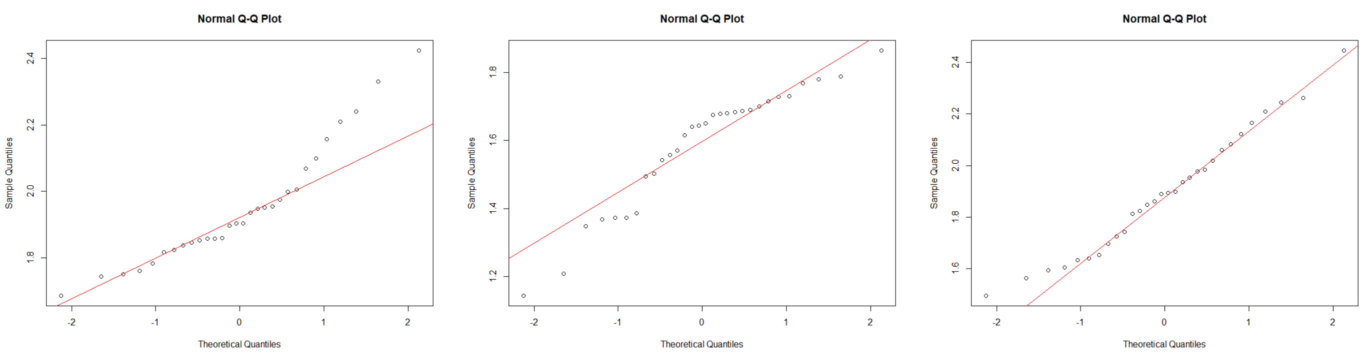


Figura 2: tiempo vs combinaciones de k y n



(a) Paralelizado

(b) k de 1000 con los 3 valores de n

(c) k de 10000 con los 3 valores de n

Figura 3: Normalización de resultados

4. Conclusiones

Al tener una relación con valores de k y n pequeños el tiempo es menor pero los valores no son muy significativos, no se normalizan, y al tener una relación de k y n con valores grandes mejora la normalización pero el tiempo es mayor. Si se paraleliza el código aún más, se podrían obtener mejores tiempos al igual que un resultado más significativo.

5. Descripción

Para el *reto 1* se necesita saber en cuál paso los cúmulos son suficientemente grandes, usando los valores para k y n proporcionados por el código de la práctica [1] se obtienen los valores máximos de los cúmulos usando `max(cumulos)` y por medio de los histogramas se observan los pasos, se hacen 30 réplicas y se obtienen los datos estadísticos con `qqplot`.

6. Resultados

En la figura 4 se observan los valores máximos de cúmulos normalizados y en la figura 5 se puede observar que en el paso 3 hay mayor cantidad de cúmulos grandes que ya pueden filtrarse.

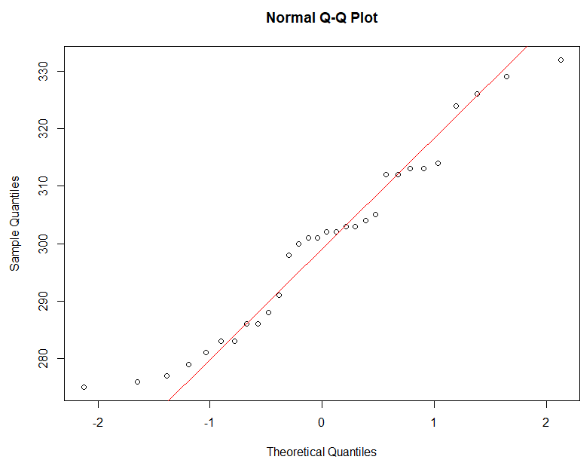
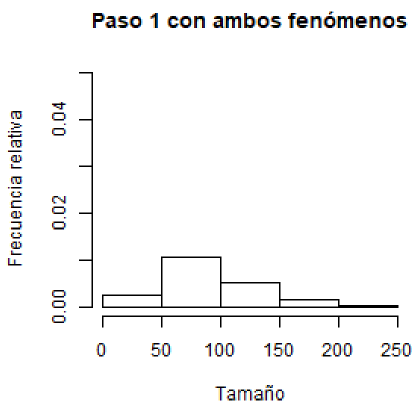
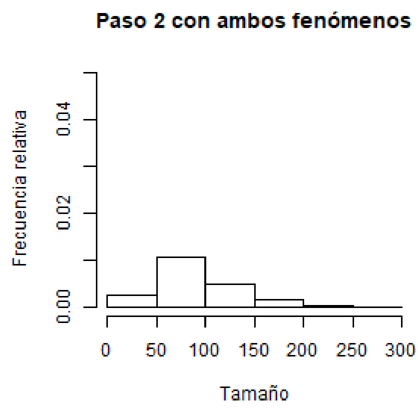


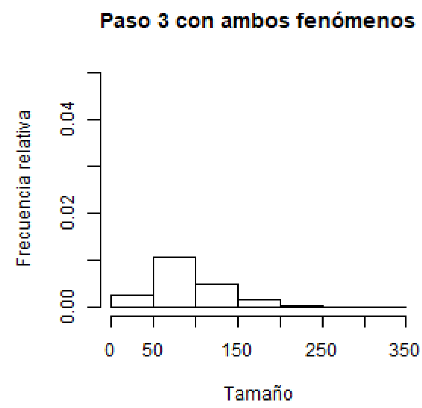
Figura 4: normalización de valores máximos de cúmulos



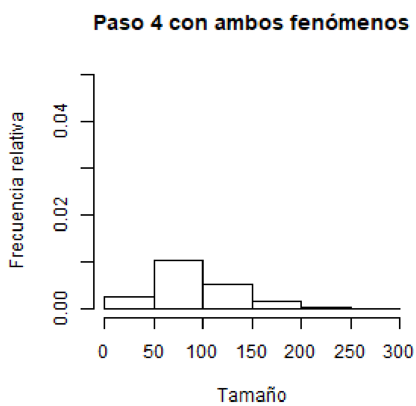
(a)



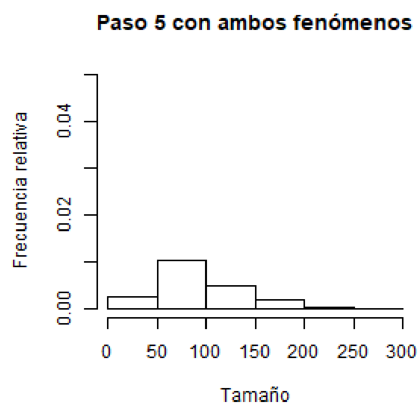
(b)



(c)



(d)



(e)

Figura 5: Histograma de pasos
(frecuencia vs tamaño de cúmulos)

Reto 2

7. Descripción

Para determinar la importancia del valor de c (valor crítico), en el código modificado para el *Reto 1* se cambia su valor original dándole un valor de 40 y se grafica en una curva sigmoidal.

8. Resultados

En la figura 6 se observa que al darle a c un valor de 40, la probabilidad de que se formen cúmulos que ya no se rompan. En la figura 7 se confirma en crecimiento de cúmulos a mayor cantidad de pasos.

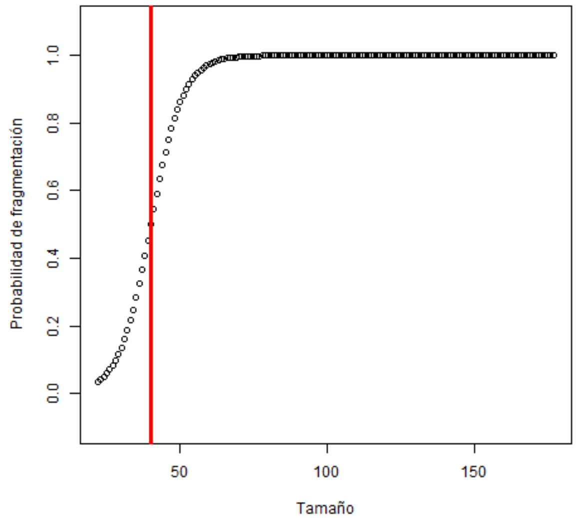
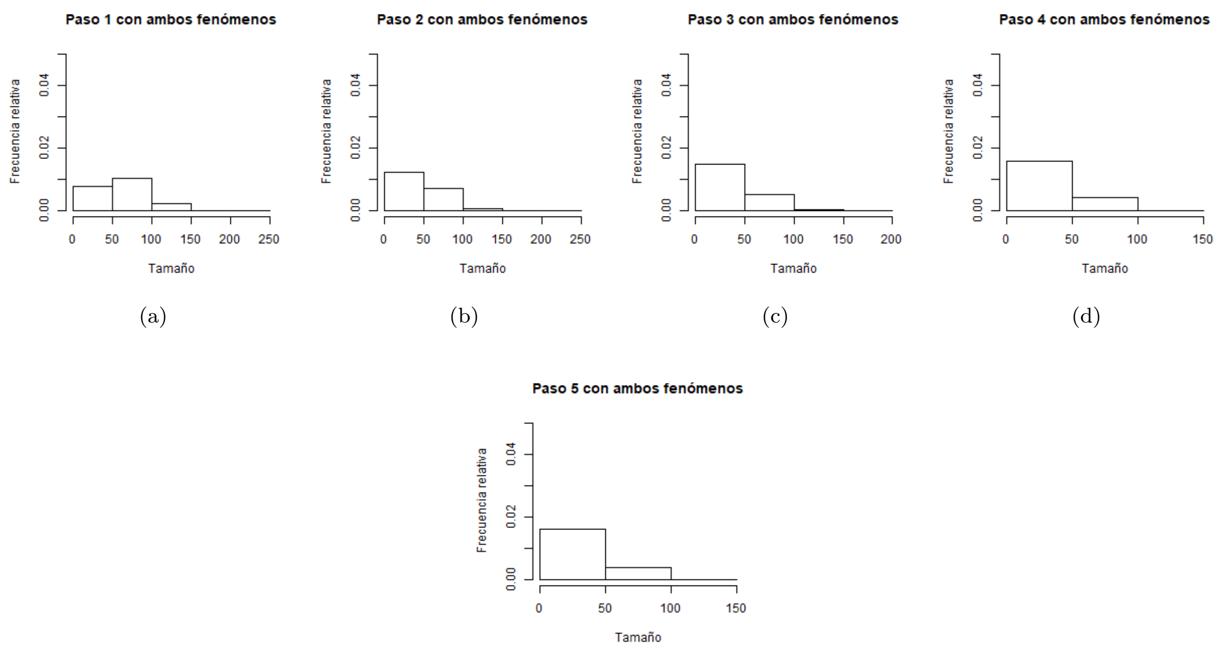


Figura 6: Curva sigmoidal



(a) (b) (c) (d) (e)

Figura 7: Histograma de pasos
(frecuencia vs tamaño de cúmulos)

9. Conclusiones

El valor de c afecta en la probabilidad de que los cúmulos se rompan o no, esto afecta en cuál paso se tendrán cúmulos de tamaño grande para poder filtrarlos y por lo tanto los tiempos varían también, por ejemplo, en este caso de c=40, hay mayor probabilidad de que los cúmulos no se rompan, por lo que disminuirá el tiempo ya que no se tendrán que unir cúmulos rotos y se podrán filtrar más rápido.

Referencias

[1] Elisa Schaeffer. Práctica 8: modelo de urnas, 2019. URL `https://elisa.dyndns-web.com/teaching/comp/par/p8.html`.