

Algoritmo Genético

Liliana Saus

Práctica 10

15 de octubre de 2018

Introducción

La práctica 9 es sobre el uso de algoritmos; genético y exacto , en particular para el problema de la mochila, donde la tarea consiste en:

- Seleccionar un subconjunto de objetos de tal forma que no se exceda la capacidad de la mochila en términos de la suma de los pesos de los objetos incluidos.
- El valor total de los objetos incluidos sea lo máximo posible

Objetivo

Se paraleliza el algoritmo genético y se estudia los efectos en su tiempo de ejecución con pruebas estadísticas y visualizaciones, variando el número de objetos en la instancia. Además se determina con qué tamaño de instancia el algoritmo genético es mejor que el algoritmo exacto en términos de valor total obtenido por segundo de ejecución.

Datos experimentales

Se realizó la simulación para instancias con $n = 50, 100, 200$ durante 50 iteraciones y con 10 réplicas. Se tomaron esa cantidad de réplicas por características de la computadora empleada. Para poder llevar acabo el objetivo, en la versión paralela del algoritmo genético se paralelizaron las fases:

- Mutación
- Reproducción
- Factibilidad
- Objetivo

Resultados

Se ejecutaron la versión secuencial y paralela, para comparar los tiempos, la figura 1 muestra los resultados. Se ve que los tiempos de ejecución para la versión secuencial son más grandes que para la versión paralela.

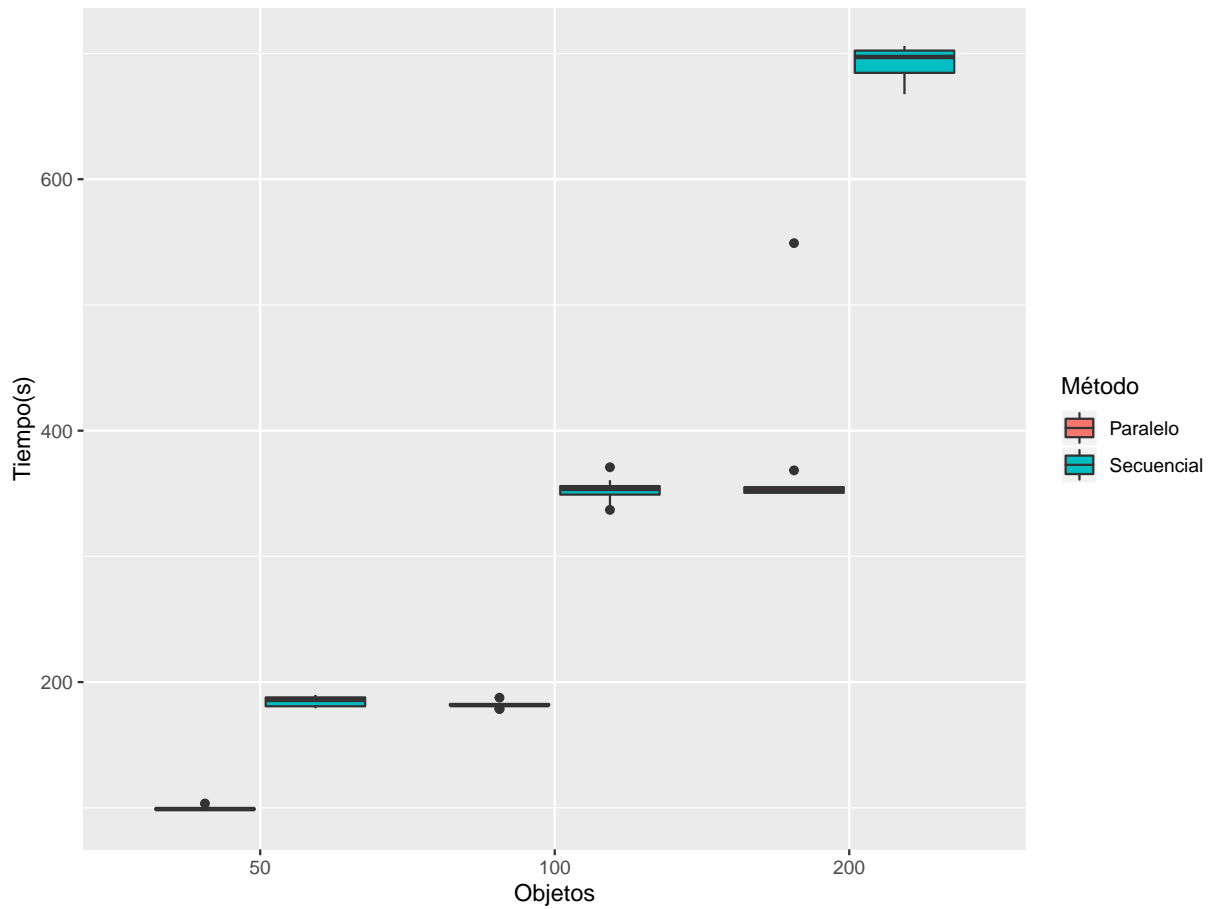


Figura 1: Tiempo de ejecución paralela y secuencial.

Se realiza la prueba de *Shapiro – Wilk* para examinar si los datos cumplen con normalidad.

```
1 > Shapiro.test(resultados$tiempo)
2
3     Shapiro-Wilk normality test
4
5 data:  resultados$Tiempo
6 W = 0.82121, p-value = 4.754e-07
```

Como el p-valor es menor que 0.05, se concluye que los datos no cumplen normalidad. Así que se realiza una prueba estadística no paramétrica *Kruskal – Wallis*

```
1 > kruskal.test(tiempo ~ M todo , data = Resultados)
2
3     Kruskal-Wallis rank sum test
4
5 data:  Tiempo by M todo
6 Kruskal-Wallis chi-squared = 15.582, df = 1,
7 p-value = 7.899e-05
```

Como el p-valor es menor que 0.05 se rechaza la hipótesis nula, es decir que hay diferencias significativas en los tiempos.

Por otro lado se calculó el beneficio entre el método exacto y el genético, usando

$$\text{beneficio} = \frac{\text{objetivo}}{\text{suma total de los objetivos}} \quad (1)$$

En la figura 2 se muestran los resultados, es fácil ver que hay un mayor beneficio en el algoritmo exacto que en el genético.

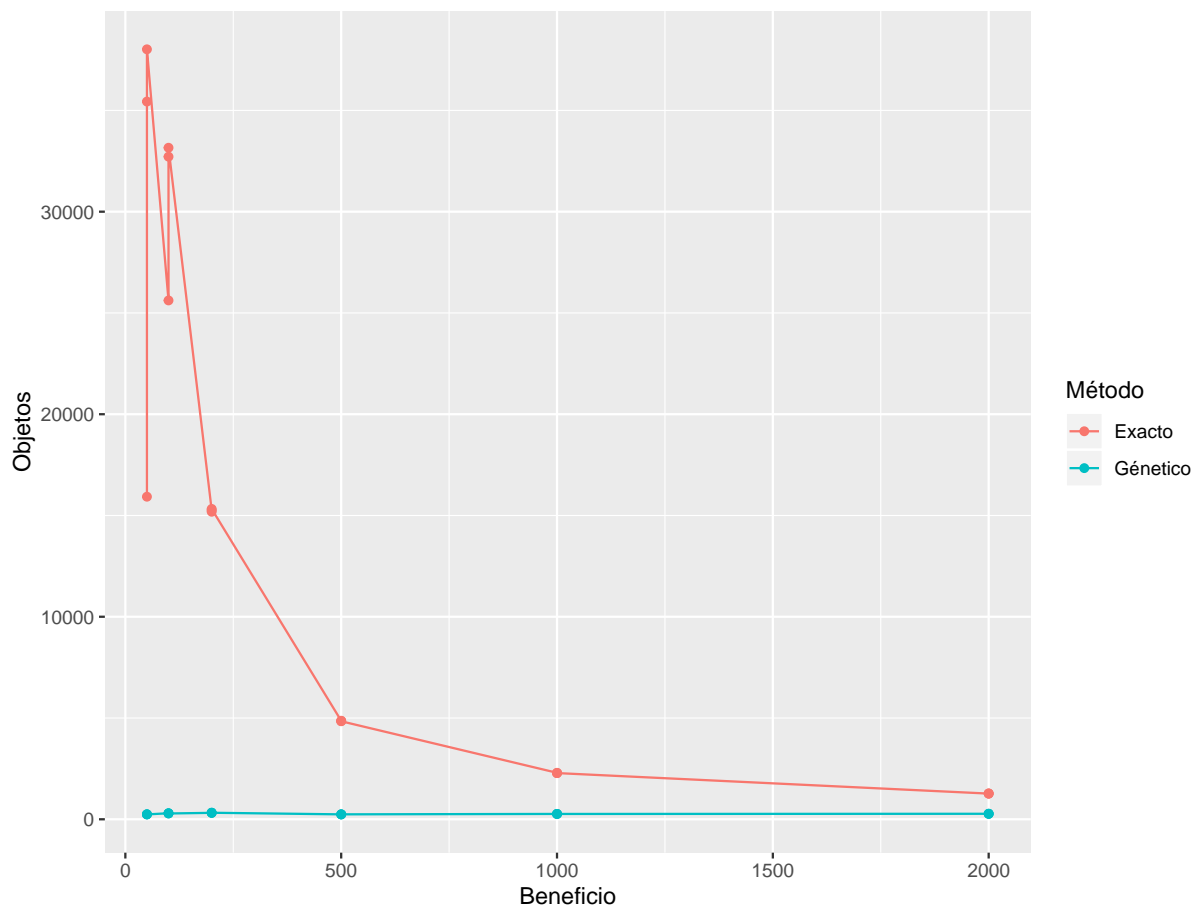


Figura 2: Beneficio entre algoritmo genético y exacto.

Estos resultados podrían cambiar si se realiza un ajuste en los parámetros.

Reto 1

El primer reto se cambia la selección de padres para reproducción a que use selección de ruleta: cada solución se selecciona como padre con una probabilidad que es proporcional a su valor de función objetivo y a su factibilidad, combinando los dos. Se estudia si este cambio produce una mejora estadísticamente significativa en la calidad de la solución.

La figura 3 muestra la función objetivo aplicando el método con ruleta y sin ruleta, lo que se puede observar que se alcanza un mejor valor por el método sin ruleta.

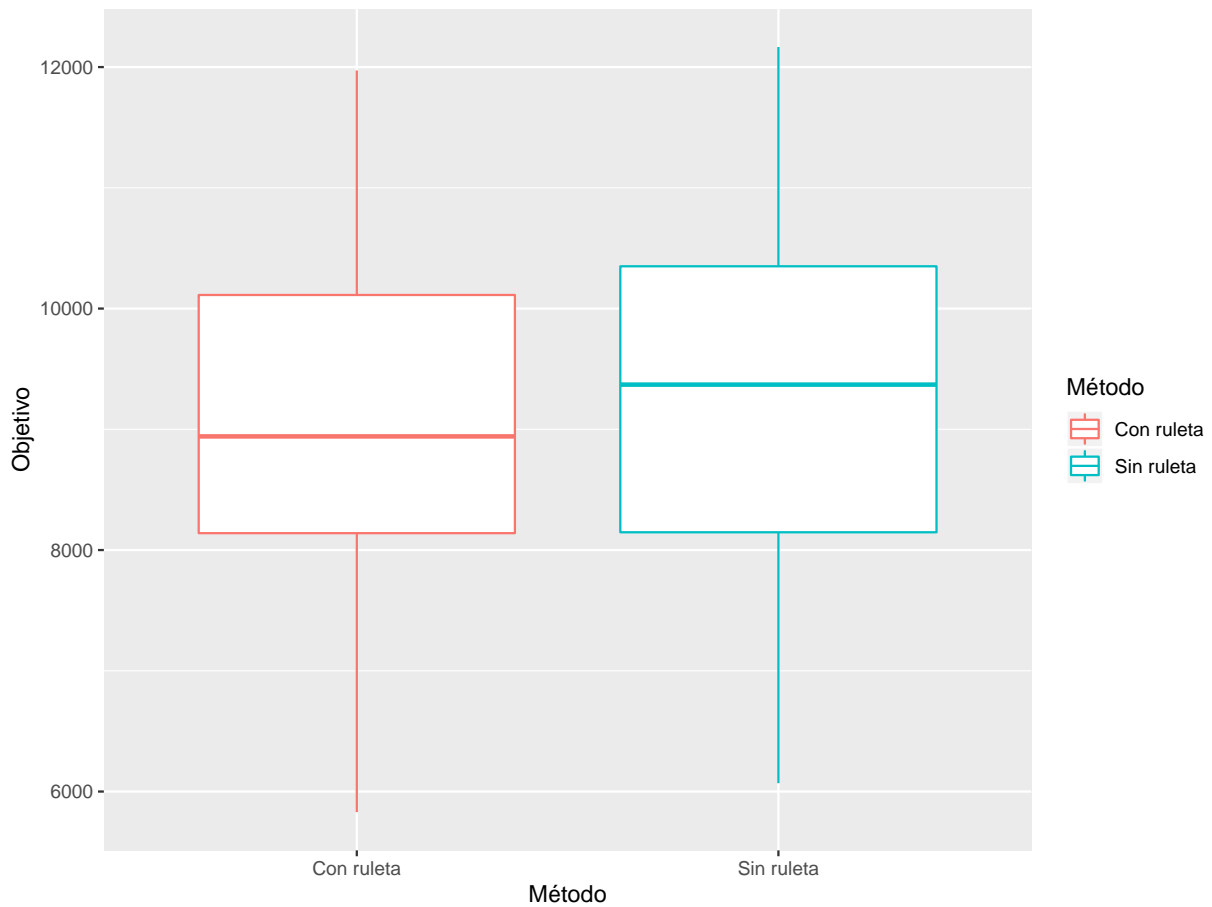


Figura 3: Función objetivo con ruleta y sin ruleta

Para probar si hay diferencia significativa entre los datos se realiza una prueba *Kruskal – Wallis*

```
1 > kruskal.test(tiempo ~ M todo , data = Resultados)
2
3     Kruskal-Wallis rank sum test
4 data:  datos$Objetivo by M todo
5 Kruskal-Wallis chi-squared = 0.41333, df = 1,
6 p-value = 0.5203
```

Como el p-valor es mayor que 0.05, aceptamos la hipótesis nula, por lo que se concluye que no hay nivel de significancia entre los datos. Se puede decir que haciendo un ajuste de parámetros se llegaría a otros resultados.

Reto 2

El segundo reto es extender la selección de ruleta a la fase de supervivencia: en vez de quedarse con las mejores soluciones, cada solución tiene una probabilidad de entrar a la siguiente generación que es proporcional a su valor de la función objetivo, incorporando el sí o no es factible la solución en cuestión, permitiendo que los k mejores entre las factibles entren siempre. Se estudia nuevamente el efecto de este cambio en la calidad de la solución. La figura 4 muestra la función objetivo aplicando el método con ruleta y sin ruleta, lo que se puede observar que se alcanza un mejor valor por el método sin ruleta.

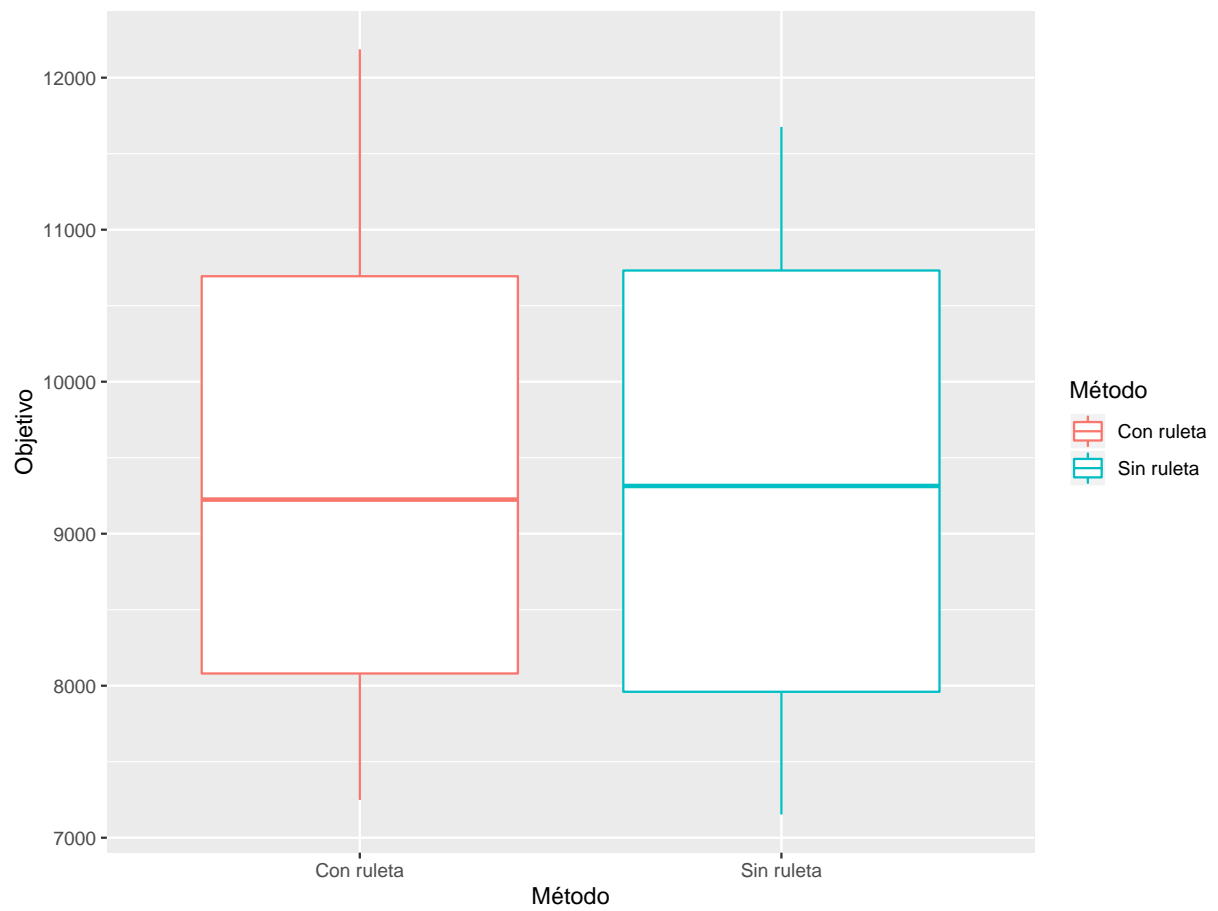


Figura 4: Función objetivo con ruleta y sin ruleta

Para probar si hay diferencia significativa entre los datos se realiza una prueba *Kruskal – Wallis*

```
1 kruskal.test(tiempo ~ M todo , data = Resultados)
2
3     Kruskal-Wallis rank sum test
4
5 data:  datos$Objetivo by M todo
6 Kruskal-Wallis chi-squared = 0.03694, df = 1,
7 p-value = 0.8476
```

Como el p-valor es mayor que 0.05, aceptamos la hipótesis nula, por lo que se concluye que no hay nivel de significancia entre los datos. Por lo que haciendo un ajuste de parámetros se podría llegar a otros resultados.

Referencias

- [1] ELISA SCHAEFFER *R paralelo: simulación and análisis de datos*, 2018.
<https://elisa.dyndns-web.com/teaching/comp/par/>