

Taller sobre Análisis de Complejidad Temporal y Experimentos sobre Algoritmos

Juan Ossa - A00358374

Estaban Yusunguaira - A00358393

Felipe García - A00233788

Samuel Satizabal - A00352296

1. Identificación componentes del Análisis y Diseño Experimental

a. Unidad Experimental

En el experimento planteado, las unidades experimentales per sé son el algoritmo QuickSort y el algoritmo RandomizeQuickSort. Los algoritmos se someterán a un número determinado de pruebas, con distintos factores de entrada como el tamaño de elementos de los arreglos a ordenar y el orden inicial de los elementos en el arreglo, comparando así cuál algoritmo tiene mejor desempeño temporal.

b. Variables de respuesta

La variable de respuesta escogida para este experimento es el tiempo de respuesta de cada algoritmo al finalizar de ejecutarse. De acuerdo a la cantidad de elementos de los arreglos y el orden de los mismos, el tiempo de ejecución del algoritmo debería variar.

c. Factores controlables

Los factores identificados que pueden ser controlados a voluntad para evidenciar variabilidad en el tiempo de ejecución de los dos algoritmos en cuestión son:

- **Número de elementos por arreglo:** Se harán pruebas con 10,100,1.000 y 10.000 elementos por arreglos.
- **Orden inicial de los elementos por arreglo:** Se harán pruebas con el número de elementos descrito anteriormente, ordenados de forma ascendente, descendente y aleatoria.

d. Factores no controlables

- **Limitaciones de hardware:** Limitaciones dadas por los componentes de hardware del equipo donde se realizan las pruebas como: La memoria ram, microprocesador, número de núcleos, entre otros.

- **Limitaciones espaciales y temporales:** Limitaciones dadas por el funcionamiento lógico y restricciones propias del lenguaje de programación.

e. Factores estudiados

En este experimento tomamos los factores de estudio a los factores controlados ya que son los factores que influyen directamente en la variación en el tiempo de ejecución de los algoritmos:

- Número de elementos por arreglo.
- Orden inicial de los elementos por arreglo.

f. Niveles

- Niveles utilizados según la cantidad de elementos por arreglo

Número de elementos de entrada	Nivel de elementos
10	1
10^2	2
10^3	3
10^4	4
10^5	5

- Niveles utilizados según el orden inicial de los elementos por arreglo

Número de elementos de entrada	Nivel de elementos
Ascendente	1
Descendente	2
Aleatorio	3

g. Tratamientos

Número de elementos de entrada	Orden de los elementos	Tratamiento	Repetición	Y
10	Ascendente	1	1	
			2	
			3	
10	Descendente	2	1	
			2	
			3	
10	Aleatorio	3	1	
			2	
			3	
10²	Ascendente	4	1	
			2	
			3	
10²	Descendente	5	1	
			2	

			3	
10 ²	Aleatorio	6	1	
			2	
			3	
10 ³	Ascendente	7	1	
			2	
			3	
10 ³	Descendente	8	1	
			2	
			3	
10 ³	Aleatorio	9	1	
			2	
			3	
10 ⁴	Ascendente	10	1	
			2	
			3	

10 ⁴	Descendente	11	1	
			2	
			3	
10 ⁴	Aleatorio	12	1	
			2	
			3	

2. Etapas del análisis y diseño de experimentos se han llevado a cabo hasta el momento

A continuación se enlistan y se explica de manera resumida las etapas realizadas hasta el momento durante el experimento realizado cuyo objetivo es demostrar cuál de los dos algoritmos entre el RandomizedQuickSort y QuickSort tiene mejor desempeño temporal

- **Etapla planeación y experimentación:** En esta etapa se delimitó el objeto del problema el cual fue el estudio del desempeño temporal de los algoritmos QuickSort y el algoritmo RandomizeQuickSort. Se establecieron los factores de estudio, la variable de respuesta, niveles y realización del experimento con un número de repeticiones establecidas que pudieran arrojar resultados evidentes.
- **Etapla análisis:** En esta etapa se realiza la técnica estadística central en el análisis de experimentos llamado el análisis de varianza o ANOVA. Esta prueba se hace con el objetivo de obtener efectos o diferencias muestrales, entendiendo que estamos viendo que los resultados experimentales son observaciones muestrales y no poblacionales.
- **Etapla interpretación:** En esta etapa a través del análisis de varianza se puede evidenciar que efectivamente se marca una tendencia que el algoritmo RandomizedQuickSort tiene mejor desempeño que el algoritmo QuickSort.

Etapas faltantes:

- **Etapas de control y conclusiones finales:** Esta etapa no se realizó en el presente experimento, falta llevar a cabo los controles respectivos y escribir la conclusión final de acuerdo al análisis e interpretación de los datos obtenidos junto con el trabajo futuro.

Clasificación de objetivos del experimento

El objetivo de este experimento se clasifica en el Diseño para estudiar el efecto de varios factores sobre la respuesta. Se escoge esta opción, por que se estudia la variable de respuesta de acuerdo a los dos factores controlables.

Complejidad Temporal

Línea	Instrucción	# veces que se ejecuta (Mejor, Promedio, Peor)	
	QuickSort(A,p,r)		
1	if p < r	n - 1	n - 1
2	q = Partition(A,p,r)	2n+1	4n+3
3	QuickSort(A,p,q-1)	T(n/2)	T(n-1)
4	QuickSort(A,q+1,r)	T(n/2)	T(0)

Mejor:

$$T(n) = 2T\left(\frac{n}{2}\right) + 3n = 2T\left(\frac{n}{2}\right) + O(n)$$

Evaluamos n/2

$$T\left(\frac{n}{2}\right) = 2T\left(\frac{\frac{n}{2}}{2}\right) + \frac{n}{2}$$

Sustituimos en la recursión inicial

$$T(n) = 2\left(2T\left(\frac{n}{2^2}\right) + \frac{n}{2}\right) + n$$

$$T(n) = 2^2 T\left(\frac{n}{2^2}\right) + 2n$$

Si continuamos con la sustitución se evidencia el patrón...

$$T(n) = 2^k T\left(\frac{n}{2^k}\right) + kn$$

Asumimos que :

$$T\left(\frac{n}{2^k}\right) = T(1)$$

$$\frac{n}{2^k} = 1 \quad ; \quad n = 2^k$$

$$k = \log \log n$$

$$T(n) = 2^k T(1) + kn$$

Reemplazamos k

$$T(n) = n * 1 + (\log \log n)n$$

$$T(n) = n + (\log \log n)n$$

Cuando $\log n > 1$, la parte derecha de la suma crecerá más rápido que n, por lo que

$$T(n) = O(n \log n)$$

Peor: $T(n) = T(n-1) + 5n + 2$

$$T(n) = T(n-1) + O(n)$$

$$T(n-1) = T(n-2) + n - 1$$

$$T(n) = (T(n-2) + n - 1) + n$$

$$T(n) = T(n-2) + 2n - 1$$

$$T(n-2) = T(n-3) + n - 2$$

$$T(n) = ((T(n-3) + n - 2) + n - 1) + n$$

$$T(n) = T(n-3) + 3n - 3$$

Si continuamos con la sustitución se evidencia el patrón...

$$T(n) = T(n-k) + (n - (k-1)) + (n - (k-2)) + \dots + (n-1) + n$$

Asumimos que :

$$n - k = 0$$

$$n = k$$

Reemplazamos k

$$T(n) = T(n-n) + (n - (n-1)) + (n - (n-2)) + \dots + (n-1) + n$$

$$T(n) = T(0) + (1) + (2) + \dots + (n-1) + n$$

Obtenemos T(0) + números naturales hasta n

$$T(n) = 0 + \frac{n(n-1)}{2}$$

$$T(n) = \frac{n(n-1)}{2}$$

Ya que llegamos a una función polinómica de grado dos, podemos afirmar que :

$$T(n) = O(n^2)$$

Línea	Instrucción	# veces que se ejecuta (Mejor, Promedio, Peor)	
	Partition(A,p,r)		
1	x = A[r]	1	1
2	i = p - 1	1	1

3	for j = p to r - 1	n-1	n-1
4	if A[j] ≤ x	n-2	n-2
5	i = i + 1	0	n-2
6	A[i] A[j]	0	n-2
7	A[i + 1] A[r]	1	1
8	return i + 1	1	1

Mejor: $T(n) = 2n + 1 = O(n)$

Peor: $T(n) = 4n - 3 = O(n)$

Línea	Instrucción	# veces que se ejecuta (Mejor, Promedio, Peor)		
	Random-QS(A, p, r)			
1	if p < r	n - 1	n - 1	n - 1
2	q = Rand-Parti(A, p, r)	2n+3	2n+3	4n-1
3	Randomized-QS(A, p, q-1)	T(i-1)		T(q)
4	Randomized-QS(A, q+1, r)	T(i-1)		T(n-1-q)

Para el **peor caso** tenemos que:

$$T(n) = T(q) + T(n - 1 - q) + \Theta(5n - 2), \text{ donde } 0 \leq q \leq n - 1$$

Asumimos que:

$$T(n) \leq cn^2$$

Para alguna constante c, entonces queda:

$$\begin{aligned}
 T(n) &\leq ((cq^2) + c(n - 1 - q)^2) + n \\
 &= c * ((q^2) + (n - 1 - q)^2) + n
 \end{aligned}$$

Entonces como la expresión alcanza su rango máximo $q=n-1$, por lo tanto, se cumple que:

$$\begin{aligned}
 ((q^2) + (n - 1 - q)^2) &\leq (n - 1)^2 \\
 n^2 - 2n + 1
 \end{aligned}$$

Ahora bien, así como se realizó en el caso anterior, podemos asumir que:

$$T(n) \leq cn^2 - c(2n - 1) + n$$

En el que la constante c puede llegar a ser un número tan elevado, que hace que n sea despreciable y por ende tendríamos que:

$$T(n) \leq cn^2$$

$$T(n) = O(n^2)$$

Para el **mejor caso** y el **caso promedio** se puede observar que tanto Randomized-QS como QuickSort son muy parecidos o hasta iguales, por lo tanto, para el mejor caso sería:

$$T(n) = O(n * \log(n))$$

Línea	Instrucción	# veces que se ejecuta (Mejor, Promedio, Peor)		
	Rand-Parti (A,p,r)			
1	i = Random(p,r)	1		1
2	A[r] <-> A[i]	1		1
3	return Partition(A,p,r)	2n+1		4n-3

Mejor: 2n + 3

Peor: 4n - 1

	Factores de Entrada		
# de Prueba	Variante del QS X ₁	Estado del Arreglo X ₂	Tiempo (en ms) Y
1	No Randomized	Ordenado No Descendente	$O(n^2)$
2	Randomized	Ordenado No Descendente	$O(n \log(n))$
3	No Randomized	Ordenado No Ascendente	$O(n^2)$
4	Randomized	Ordenado No Ascendente	$O(n \log(n))$
5	No Randomized	No ordenado (en orden aleatorio)	$O(n \log(n))$

6	Randomized	No ordenado (en orden aleatorio)	$O(n\log(n))$
---	------------	----------------------------------	---------------

Análisis ANOVA

En el siguiente análisis de datos se debe tener en cuenta lo siguiente:

1. **Planteamiento de hipótesis**
 - a. **Hipótesis nula:** Se basa en afirmar que los dos algoritmos tienen el mismo desempeño temporal.
 - b. **Hipótesis alternativa:** Se basa en afirmar que ambos algoritmos (muestras) se comportan de manera distinta, es decir tienen distinto desempeño temporal
2. **Interpretación de datos**
 - a. La muestra es equivalente a el tipo de algoritmo
 - b. Las columnas es equivalente a la cantidad de elementos por arreglo analizados
 - c. QS es equivalente al algoritmo QuickSort
 - d. RQS es equivalente al algoritmo RQS

Resultados análisis ANOVA para elementos ordenados no ascendentes

Análisis de varianza de dos factores con varias muestras por grupo

RESUMEN	10ND	100ND	1000ND	10000ND	Total
<i>QS</i>					
Cuenta	1000	1000	1000	1000	4000
Suma	8,0407	55,9121	4555,8567	395046,446	399666,256
Promedio	0,0080407	0,0559121	4,5558567	395,046446	99,9165639
Varianza	0,01400013	0,05276724	0,355288	306,820908	29121,3079

<i>RQS</i>					
Cuenta	1000	1000	1000	1000	4000
Suma	18,0353	141,2761	1135,9112	9012,2519	10307,4745
Promedio	0,0180353	0,1412761	1,1359112	9,0122519	2,57686863
Varianza	0,01776977	0,1227689	0,16733029	3,43984532	14,9322885

<i>Total</i>					
Cuenta	2000	2000	2000	2000	
Suma	26,076	197,1882	5691,7679	404058,698	
Promedio	0,013038	0,0985941	2,84588395	202,029349	
Varianza	0,01590199	0,08954682	3,18664796	37429,2897	

ANÁLISIS DE VARIANZA

<i>Origen de las variaciones</i>	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
Muestra	18950032,6	1	18950032,6	487475,258	0	6,63806691
Columnas	60638126,2	3	20212708,7	519956,648	0	3,7840707
Interacción	55567018,8	3	18522339,6	476473,179	0	3,7840707
Dentro del grupo	310679,687	7992	38,8738347			
Total	135465857	7999				

Resultados análisis ANOVA para elementos ordenados no descendentes

Análisis de varianza de dos factores con varias muestras por grupo

RESUMEN	10NA	100NA	1000NA	10000NA	Total
<i>QS</i>					
Cuenta	1000	1000	1000	1000	4000
Suma	7,9882	57,7316	3571,1271	295501,797	299138,644
Promedio	0,0079882	0,0577316	3,5711271	295,501797	74,784661
Varianza	0,00792059	0,05419512	0,43487885	204,779692	16296,1135

<i>RQS</i>					
Cuenta	1000	1000	1000	1000	4000
Suma	23,0156	147,4863	1231,01	10435,7916	11837,3035
Promedio	0,0230156	0,1474863	1,23101	10,4357916	2,95932588
Varianza	0,02253327	0,12442402	0,25000019	30,8604634	26,6664513

<i>Total</i>					
Cuenta	2000	2000	2000	2000	
Suma	31,0038	205,2179	4802,1371	305937,589	
Promedio	0,0155019	0,10260895	2,40106855	152,968794	
Varianza	0,01527579	0,09127988	1,71199009	20443,5809	

ANÁLISIS DE VARIANZA

<i>Origen de las variaciones</i>	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
Muestra	10317757,5	1	10317757,5	348964,727	0	6,63806691
Columnas	34722200,9	3	11574067	391455,324	0	3,7840707
Interacción	30316298,4	3	10105432,8	341783,531	0	3,7840707
Dentro del grupo	236297,573	7992	29,5667634			
Total	75592554,4	7999				

Resultados análisis ANOVA para elementos ordenados aleatoriamente

Análisis de varianza de dos factores con varias muestras por grupo

RESUMEN	10RA	100RA	1000RA	10000RA	Total
	<i>QS</i>				
Cuenta	1000	1000	1000	1000	4000
Suma	3,9868	14,8712	207,939	2381,8485	2608,6455
Promedio	0,0039868	0,0148712	0,207939	2,3818485	0,65216138
Varianza	0,00396197	0,014546423	0,16192253	0,30312048	1,12490322
	<i>RQS</i>				
Cuenta	1000	1000	1000	1000	4000
Suma	17,936	131,4733	1170,9468	9515,6419	10835,998
Promedio	0,017936	0,1314733	1,1709468	9,5156419	2,7089995
Varianza	0,017571089	0,116835491	0,18366325	2,49639905	16,3523644
	<i>Total</i>				
Cuenta	2000	2000	2000	2000	
Suma	21,9228	146,3445	1378,8858	11897,4904	
Promedio	0,0109614	0,07317225	0,6894429	5,9487452	
Varianza	0,010809813	0,069058808	0,40466843	14,1281762	

ANÁLISIS DE VARIANZA

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Muestra	8461,166145	1	8461,16614	20524,2307	0	6,638066908
Columnas	49141,94574	3	16380,6486	39734,5006	0	3,784070704
Interacción	17454,92532	3	5818,30844	14113,457	0	3,784070704
Dentro del grupo	3294,722258	7992	0,41225253			
Total	78352,75946	7999				

Análisis

Según los datos obtenidos en las tres formas de ordenar elementos con respecto a la cantidad de elementos en cada arreglo, al ser F mayor al Valor Crítico para F, se rechazaría la hipótesis nula.

Anexos

Experimentos en formato excel ordenados

[Link](#)