

Diseño y análisis de algoritmos

Comparación de rendimiento entre algoritmos de selección de medianas

Erik Regla

Universidad de Talca

30 de noviembre de 2016

Descripción de las pruebas

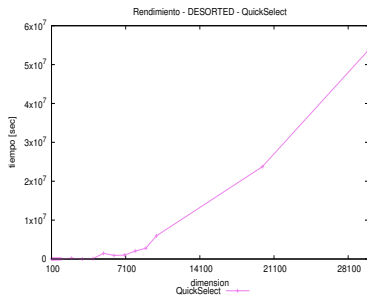
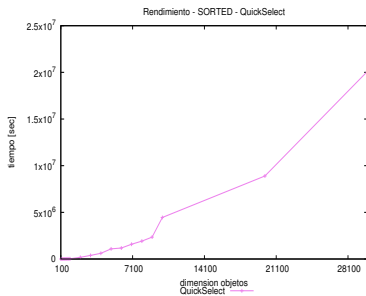
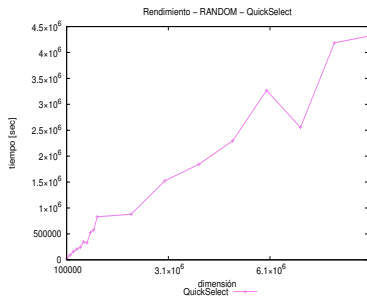
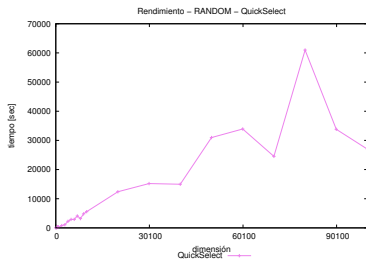
- ▶ Tamaño: 11GB
- ▶ 5 % de los elementos están repetidos
- ▶ Distribución uniforme
- ▶ Se realizaron pruebas para QuickSelect, BFPRT, iBFPRT

Descripción de las pruebas

- ▶ Tamaño: 11GB
- ▶ 5 % de los elementos están repetidos
- ▶ Distribución uniforme
- ▶ Se realizaron pruebas para QuickSelect, BFPRT, iBFPRTe
IntrospectiveQuickMedian

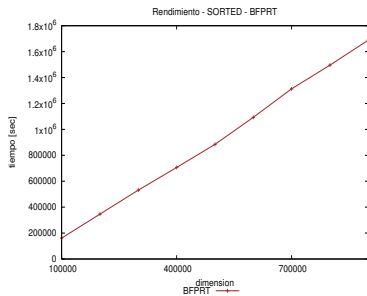
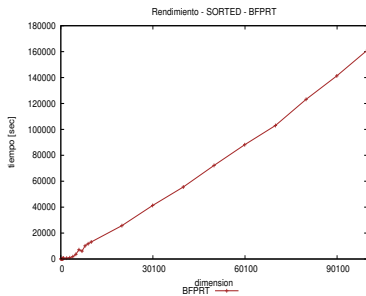
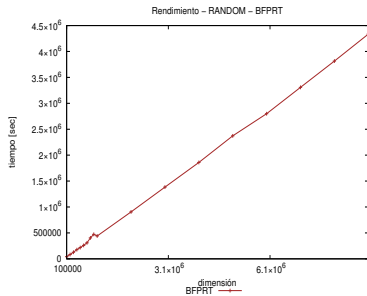
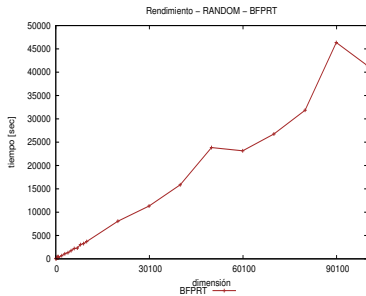
QuickSelect

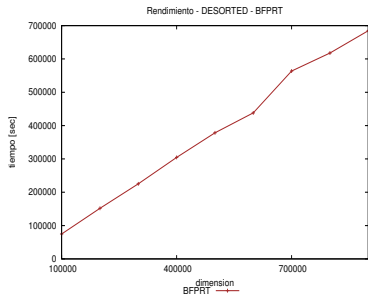
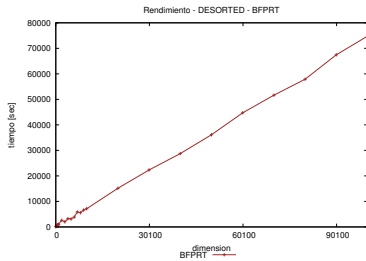
- ▶ Precisión: 100 % en todos los casos.
- ▶ Falla miserablemente en el peor caso ($O(n^2)$).
- ▶ Caso promedio $O(n)$.



BFPRT (Mediana de medianas no iterada)

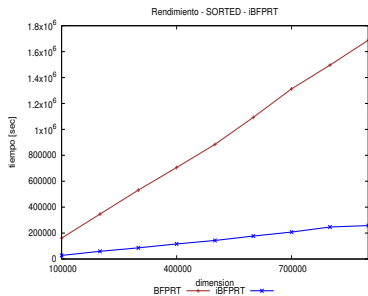
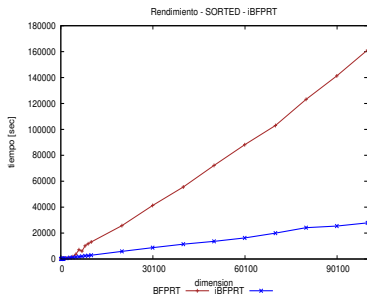
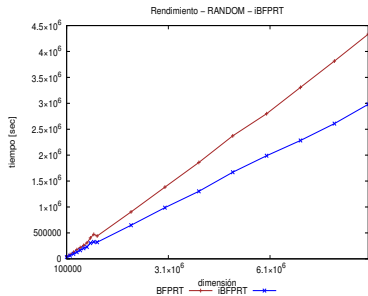
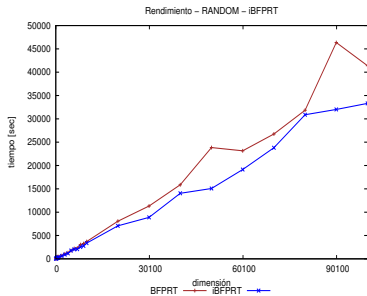
- ▶ Precisión: $98 \sim 100\%$ en todos los casos.
- ▶ Reduce el conjunto muestral a $\frac{n}{5}$ para obtener una mediana aproximada en $(O(n \lg(n)))$.
- ▶ Dependiendo del algoritmo de ordenamiento subyacente BFPRT puede variar al peor caso $(O(n^2))$.

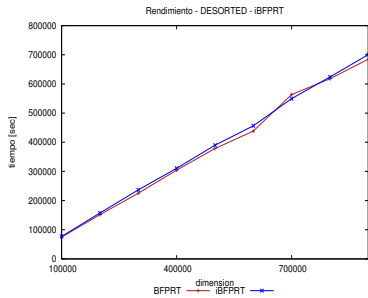
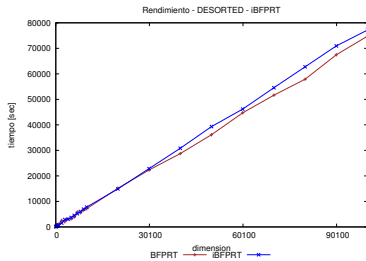




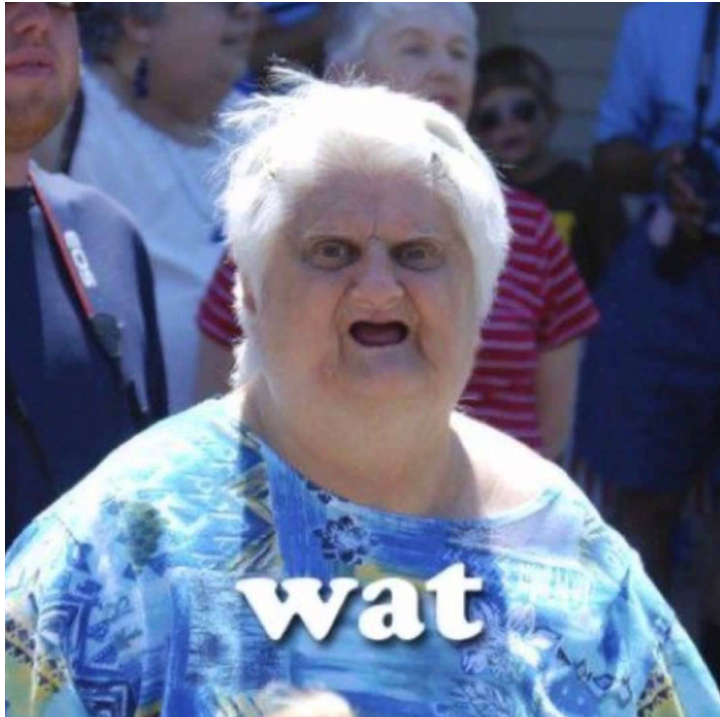
iBFPR (Mediana de medianas iterada)

- ▶ Precisión: 94 % a 99 % para el caso promedio y entre 84 % a 99 % para el peor caso.
- ▶ Iterativamente reduce el conjunto muestral en $\frac{n}{5}$ y luego los divide en $\frac{n}{k}$ conjuntos extrayendo la mediana de un conjunto de tamaño k fijo.
- ▶ Provee de una mediana aproximada en tiempo ($O(n)$) en peor caso.





IntrospectiveQuickMedian



Algorithm 1 Selección de mediana aproximada utilizando introspección

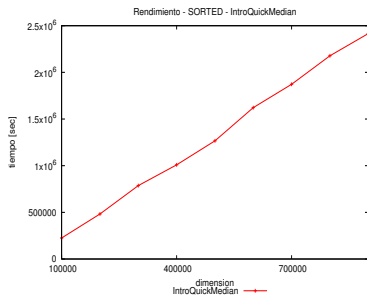
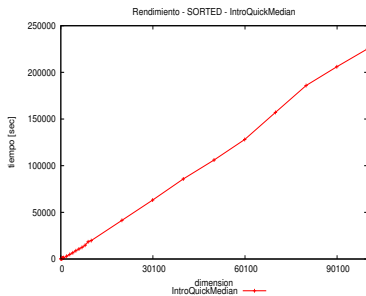
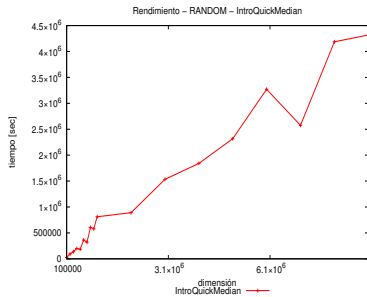
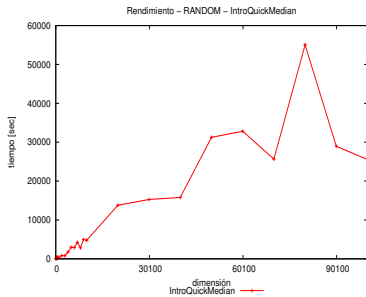
Require: J arreglo de datos desordenados, k el largo de la lista temporal K , i, j son los límites del arreglo J

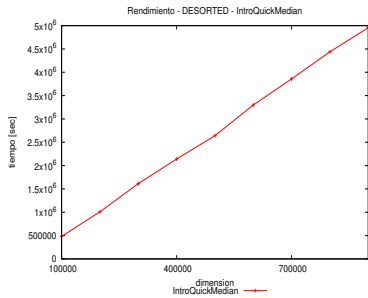
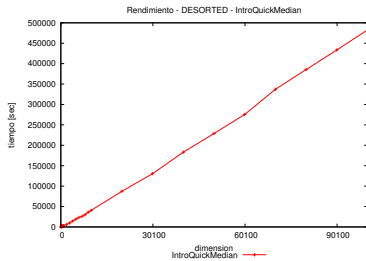
Ensure: $A_{\lfloor \frac{|A|}{2} \rfloor}$ es la mediana aproximada de J

```
1: if recursión en peor caso then
2:    $A \leftarrow []$ 
3:    $i \leftarrow 0$ 
4:   while  $i + k < |A|$  do
5:      $K \leftarrow A_{i \dots \min(i+k, |A|)}$ 
6:      $K \leftarrow \text{ordenar}(K)$ 
7:      $\text{insertar } K_{\lfloor \frac{|K|}{2} \rfloor} \text{ en } A$ 
8:      $i \leftarrow i + k$ 
9:   end while
10:   $A \leftarrow \text{ordenar}(A)$ 
11:  return  $\text{medianaDeMedianas}(A)$ 
12: else
13:   if  $p \neq \lfloor \frac{|A|}{2} \rfloor$  then
14:      $p \leftarrow \text{obtenerPivote}(A, i, j)$ 
15:      $p \leftarrow \text{particionar}(A, i, j)$ 
16:     if  $p > \lfloor \frac{|A|}{2} \rfloor$  then
17:        $i \leftarrow p$ 
18:     else
19:        $j \leftarrow p$ 
20:     end if
21:     return  $\text{IntrospectiveQuickMedian}(A, i, j)$ 
22:   else
23:     return  $p$ 
24:   end if
25: end if
```

IntrospectiveQuickMedian (Mediana de medianas iterada)

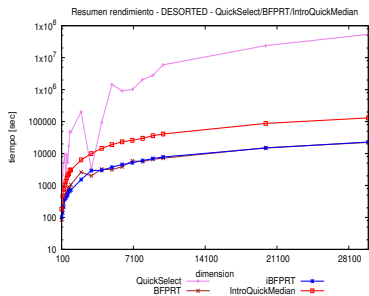
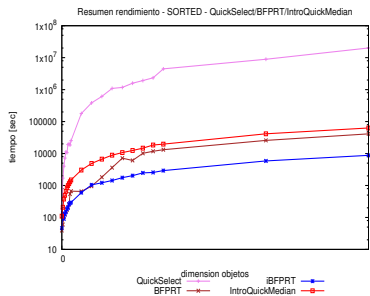
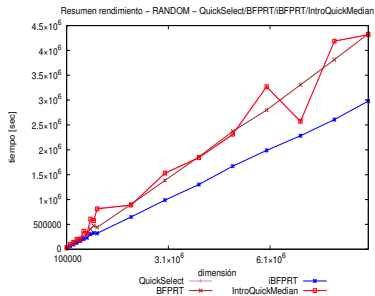
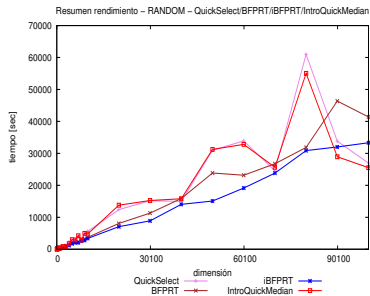
- ▶ Precisión: 99 % a 100 % para el caso promedio y 85 % a 99 % en el peor caso.
- ▶ Versión híbrida entre QuickSelect y iBFPRT.
- ▶ Provee de una mediana aproximada en tiempo ($O(n)$) en peor caso.
- ▶ Mejora la precisión del caso promedio de iBFPRT.





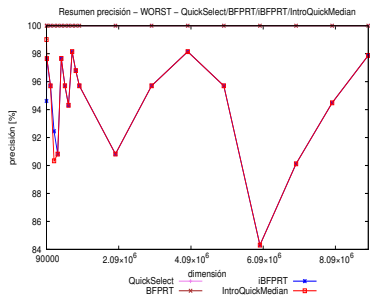
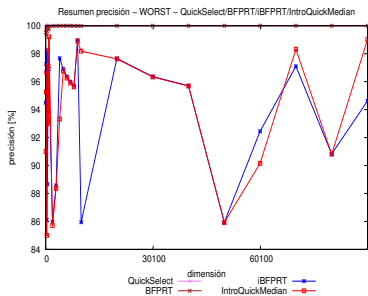
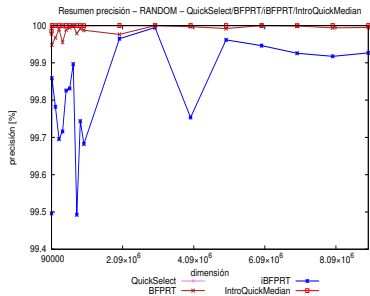
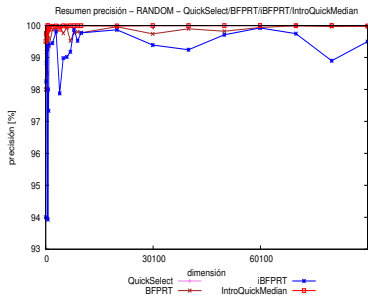
Resumen de los experimentos - Tiempo de ejecución

- ▶ BFPRT es el algoritmo más estable entre todos.
- ▶ iBFPRT es el algoritmo más rápido de todos.
- ▶ Los algoritmos basados en QuickSelect e iBFPRT son erráticos en el caso promedio.
- ▶ Todos los algoritmos describen un comportamiento uniforme para el peor caso.



Resumen de los experimentos - Precisión

- ▶ IntrospectiveQuickMedian presenta la mayor precisión en el caso promedio, junto con QuickSelect.
- ▶ IntrospectiveQuickMedian en la mayoría de los casos describe un comportamiento similar a iBFPRT.
- ▶ A medida que aumenta el tamaño del conjunto de datos, todos los algoritmos comienzan incrementar su precisión para el caso promedio.



Trabajo futuro

- ▶ Investigar el efecto que produce reducir el espacio muestral utilizando la técnica empleada en BFPRT no iterado sobre IntrospectiveQuickMedian
- ▶ Investigar el efecto que tiene este algoritmo de selección de medianas en algoritmos de selección basados en pivotes.
- ▶ Investigar el efecto que tiene el criterio de selección de espacio muestral sobre la condición de introspección.

FIN