

Sperimentazione e Interpretazione

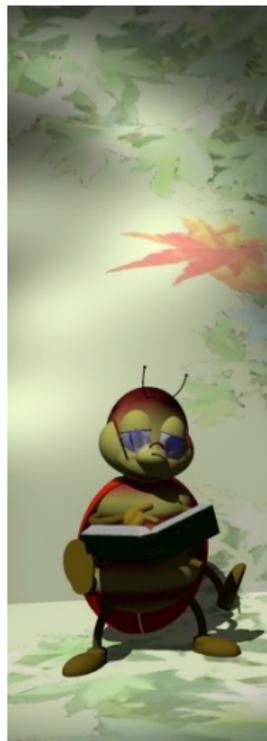
Claudio Miolo

Dipartimento di Scienze Matematiche, Informatiche e Fisiche
Università di Udine

Laboratorio PLS

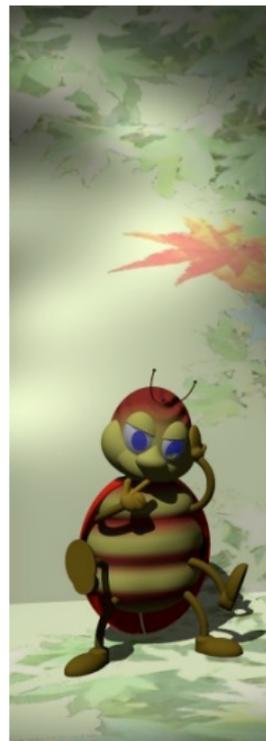
Sommario

- 1 **Retrospettiva**
 - nei panni di un matematico
 - nei panni di uno scienziato
- 2 **Indagine**
 - osservazioni da interpretare
 - osservazioni da spiegare
- 3 **Epilogo**
 - uno scienziato è curioso
 - nei panni di un ingegnere



Trama

- 1 **Retrospettiva**
 - nei panni di un matematico
 - nei panni di uno scienziato
- 2 **Indagine**
 - osservazioni da interpretare
 - osservazioni da spiegare
- 3 **Epilogo**
 - uno scienziato è curioso
 - nei panni di un ingegnere





Nei panni di un matematico

InsertionSort: Sitma del numero di confronti

- Caso medio (in “media” $\frac{1}{2}k$ confronti per l’inserimento fra k oggetti già ordinati):

$$\frac{1}{2} \cdot \frac{n(n-1)}{2} \text{ (circa)} \approx \gamma \cdot n^2$$

QuickSort: Sitma del numero di confronti

- Caso medio, bipartizioni in proporzione $\frac{1}{\rho}$ e $1 - \frac{1}{\rho}$ (ripartizione “media”, dove $1 < \rho < 2$):

$$\log_{\rho} n \text{ (livelli)} \cdot n \text{ (confronti/livello)} \approx \gamma \cdot n \log n$$



Nei panni di un matematico

InsertionSort: Sitma del numero di confronti

- Caso medio (in “media” $\frac{1}{2}k$ confronti per l’inserimento fra k oggetti già ordinati):

$$\frac{1}{2} \cdot \frac{n(n-1)}{2} \text{ (circa)} \approx \gamma \cdot n^2$$

QuickSort: Sitma del numero di confronti

- Caso medio, bipartizioni in proporzione $\frac{1}{\rho}$ e $1 - \frac{1}{\rho}$ (ripartizione “media”, dove $1 < \rho < 2$):

$$\log_{\rho} n \text{ (livelli)} \cdot n \text{ (confronti/livello)} \approx \gamma \cdot n \log n$$



Nei panni di un matematico

InsertionSort: Sitma del numero di confronti

- Caso medio (in “media” $\frac{1}{2}k$ confronti per l’inserimento fra k oggetti già ordinati):

$$\frac{1}{2} \cdot \frac{n(n-1)}{2} \text{ (circa)} \approx \gamma \cdot n^2$$

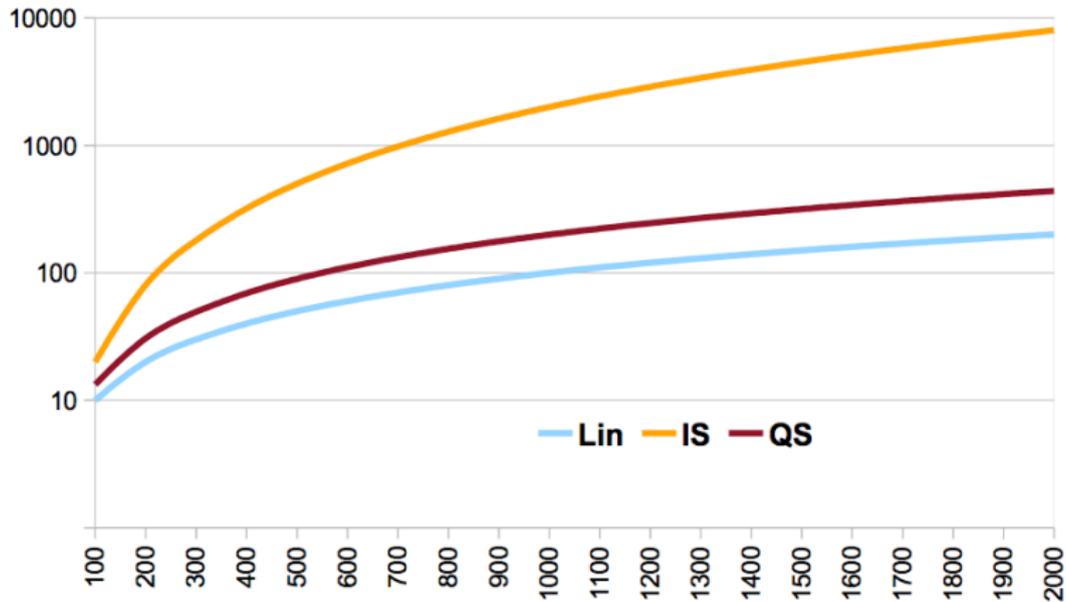
QuickSort: Sitma del numero di confronti

- Caso medio, bipartizioni in proporzione $\frac{1}{\rho}$ e $1 - \frac{1}{\rho}$ (ripartizione “media”, dove $1 < \rho < 2$):

$$\log_{\rho} n \text{ (livelli)} \cdot n \text{ (confronti/livello)} \approx \gamma \cdot n \log n$$

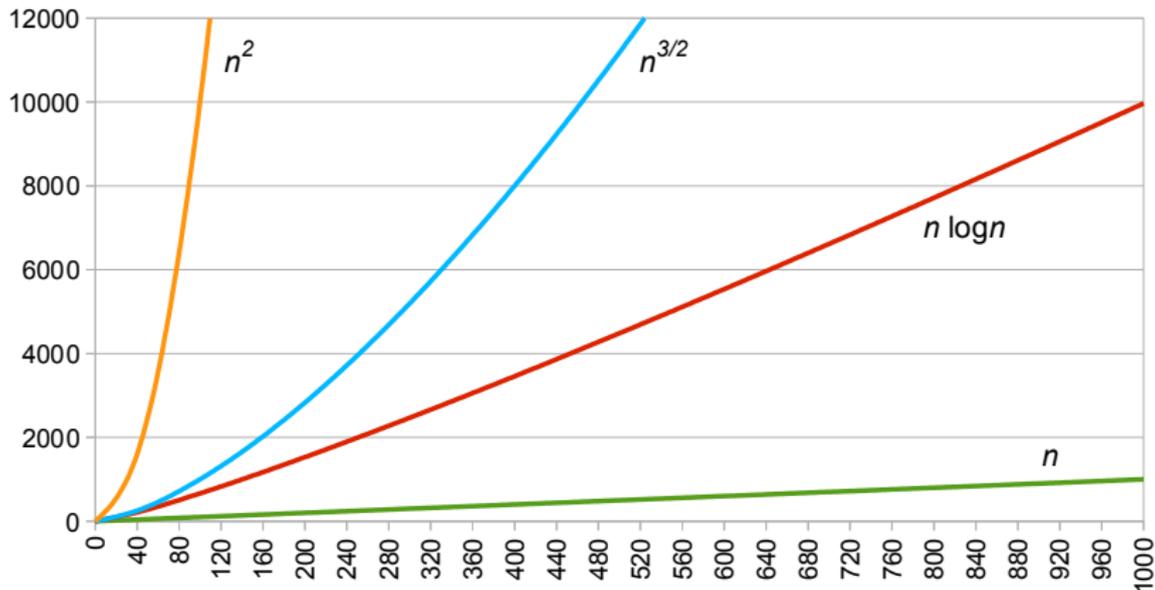


Confronto dei trend stimati





Oppure...





Domande...

- Come si possono misurare tempi di calcolo?
Quali precauzioni affinché i dati rilevati siano attendibili?
- I tempi misurati sperimentalmente confermano la teoria?
Quali aspetti sono generali e quali contingenti?
- *Quick Sort* è in tutti i casi migliore di *Insertion Sort*?
Si osserva qualcosa che un primo modello non coglie?



Domande... affrontate in laboratorio

- Come si possono misurare tempi di calcolo?
Quali precauzioni affinché i dati rilevati siano attendibili?
- I tempi misurati sperimentalmente confermano la teoria?
Quali aspetti sono generali e quali contingenti?
- *Quick Sort* è in tutti i casi migliore di *Insertion Sort* ?
Si osserva qualcosa che un primo modello non coglie?



Apprendista scienziato

Messa a punto di un adeguato “strumento di misura”

- *Ordine di grandezza* e unità di misura dei tempi di calcolo
- *Oggetto* delle misure: ordinamenti specifici? casi medi?
- *Risoluzione*, ripetibilità, interferenza di fattori estranei
- Misura dei tempi *medi* di ordinamento, generazione dei campioni, criteri
- Compensazione di *errori sistematici* della misurazione
- Errori *accidentali* cui non si può porre rimedio?



Apprendista scienziato

Messa a punto di un adeguato “strumento di misura”

- *Ordine di grandezza* e unità di misura dei tempi di calcolo
- *Oggetto* delle misure: ordinamenti specifici? casi medi?
- *Risoluzione*, ripetibilità, interferenza di fattori estranei
- Misura dei tempi *medi* di ordinamento, generazione dei campioni, criteri
- Compensazione di *errori sistematici* della misurazione
- Errori *accidentali* cui non si può porre rimedio?



Apprendista scienziato

Messa a punto di un adeguato “strumento di misura”

- *Ordine di grandezza* e unità di misura dei tempi di calcolo
- *Oggetto* delle misure: ordinamenti specifici? casi medi?
- *Risoluzione*, ripetibilità, interferenza di fattori estranei
- Misura dei tempi *medi* di ordinamento, generazione dei campioni, criteri
- Compensazione di *errori sistematici* della misurazione
- Errori *accidentali* cui non si può porre rimedio?



Apprendista scienziato

Messa a punto di un adeguato “strumento di misura”

- *Ordine di grandezza* e unità di misura dei tempi di calcolo
- *Oggetto* delle misure: ordinamenti specifici? casi medi?
- *Risoluzione*, ripetibilità, interferenza di fattori estranei
- Misura dei tempi *medi* di ordinamento, generazione dei campioni, criteri
- Compensazione di *errori sistematici* della misurazione
- Errori *accidentali* cui non si può porre rimedio?



Apprendista scienziato

Messa a punto di un adeguato “strumento di misura”

- *Ordine di grandezza* e unità di misura dei tempi di calcolo
- *Oggetto* delle misure: ordinamenti specifici? casi medi?
- *Risoluzione*, ripetibilità, interferenza di fattori estranei
- Misura dei tempi *medi* di ordinamento, generazione dei campioni, criteri
- Compensazione di *errori sistematici* della misurazione
- Errori *accidentali* cui non si può porre rimedio?



Apprendista scienziato

Messa a punto di un adeguato “strumento di misura”

- *Ordine di grandezza* e unità di misura dei tempi di calcolo
- *Oggetto* delle misure: ordinamenti specifici? casi medi?
- *Risoluzione*, ripetibilità, interferenza di fattori estranei
- Misura dei tempi *medi* di ordinamento, generazione dei campioni, criteri
- Compensazione di *errori sistematici* della misurazione
- Errori *accidentali* cui non si può porre rimedio?



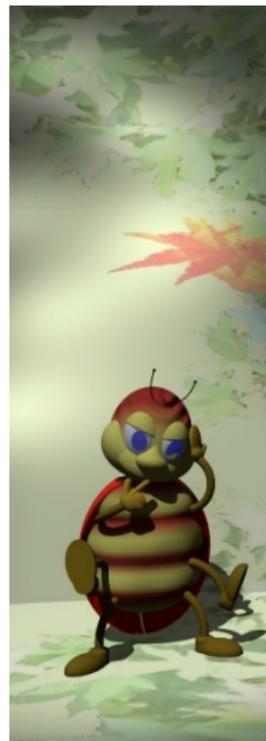
Apprendista scienziato

Messa a punto di un adeguato “strumento di misura”

- *Ordine di grandezza* e unità di misura dei tempi di calcolo
- *Oggetto* delle misure: ordinamenti specifici? casi medi?
- *Risoluzione*, ripetibilità, interferenza di fattori estranei
- Misura dei tempi *medi* di ordinamento, generazione dei campioni, criteri
- Compensazione di *errori sistematici* della misurazione
- Errori *accidentali* cui non si può porre rimedio?

Trama

- 1 Retrospettiva
 - nei panni di un matematico
 - nei panni di uno scienziato
- 2 Indagine
 - osservazioni da interpretare
 - osservazioni da spiegare
- 3 Epilogo
 - uno scienziato è curioso
 - nei panni di un ingegnere





Modelli e osservazioni

- Analisi dei dati sperimentali per scoprire proprietà oppure per verificare ipotesi e previsioni teoriche
- Confronto di dati rilevati su computer diversi: quali sono le proprietà “invarianti” ?
- Qual è la relazione fra stime teoriche e misure sperimentali?
- Uso di uno *spreadsheet* . . .



Modelli e osservazioni

- Analisi dei dati sperimentali per scoprire proprietà oppure per verificare ipotesi e previsioni teoriche
- Confronto di dati rilevati su computer diversi: quali sono le proprietà “invarianti” ?
- Qual è la relazione fra stime teoriche e misure sperimentali?
- Uso di uno *spreadsheet* . . .



Modelli e osservazioni

- Analisi dei dati sperimentali per scoprire proprietà oppure per verificare ipotesi e previsioni teoriche
- Confronto di dati rilevati su computer diversi: quali sono le proprietà “invarianti” ?
- Qual è la relazione fra stime teoriche e misure sperimentali?
- Uso di uno *spreadsheet* . . .

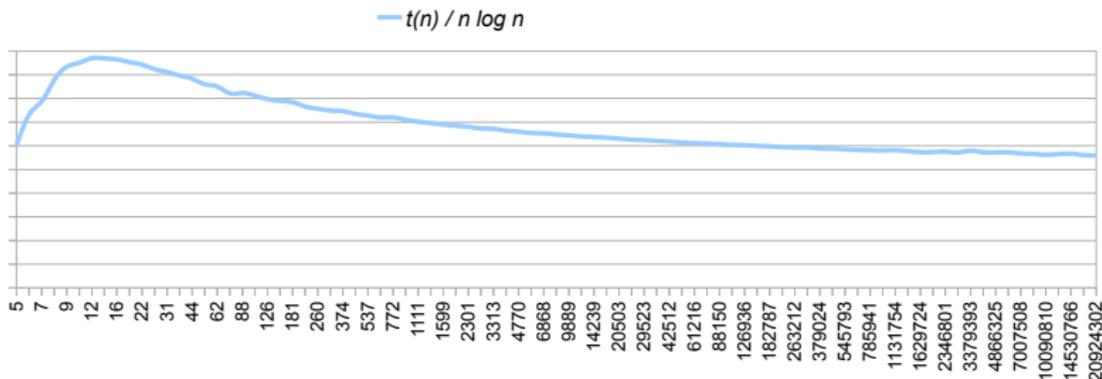


Modelli e osservazioni

- Analisi dei dati sperimentali per scoprire proprietà oppure per verificare ipotesi e previsioni teoriche
- Confronto di dati rilevati su computer diversi: quali sono le proprietà “invarianti” ?
- Qual è la relazione fra stime teoriche $n \log n$ e misure sperimentali $t(n)$?
- Uso di uno *spreadsheet* . . .

Trend dei costi di *QuickSort*

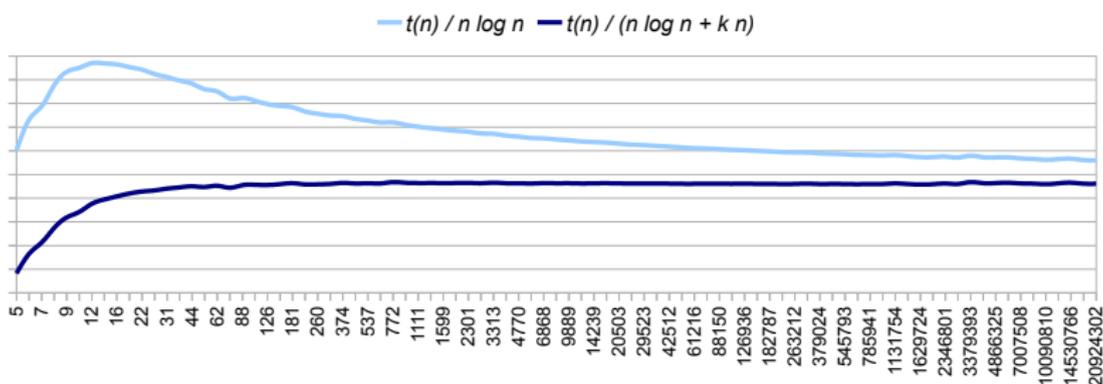
In rapporto al termine “dominante” (tramite spreadsheet) . . .
+ eventuale correzione determinata empiricamente:



Stima di k ? (casella dello spreadsheet)

Trend dei costi di *QuickSort*

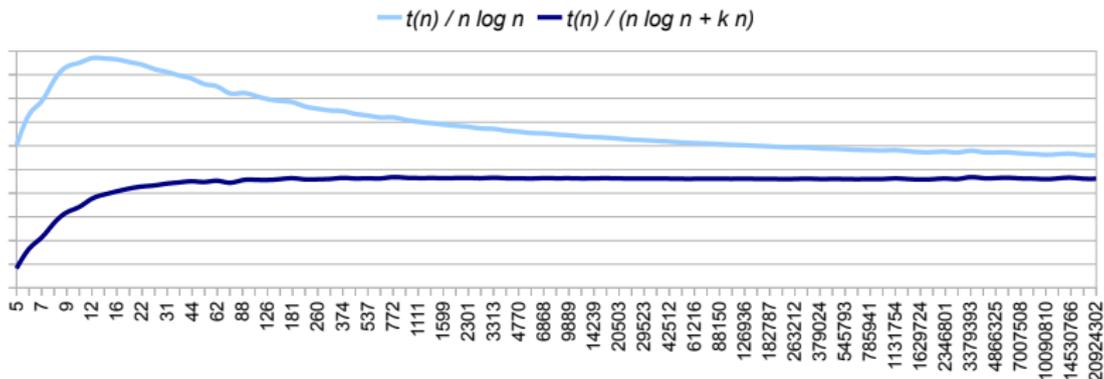
In rapporto al termine “dominante” (tramite spreadsheet)...
+ eventuale correzione determinata empiricamente:



Stima di k ? (casella dello spreadsheet)

Trend dei costi di *QuickSort*

In rapporto al termine “dominante” (tramite spreadsheet)...
+ eventuale correzione determinata empiricamente:



Stima di k ? (casella dello spreadsheet)



Interpretazione

Interpretazione di quanto osservato:

- $k \cdot n?$
- $\approx n$

L'interpretazione va sempre messa in discussione



Interpretazione

Interpretazione di quanto osservato:

- $k \cdot n$?
- $\approx n$

L'interpretazione va sempre messa in discussione



Interpretazione

Interpretazione di quanto osservato:

- $k \cdot n$?
- $\approx n$ è il numero di ricorsioni!

L'interpretazione va sempre messa in discussione



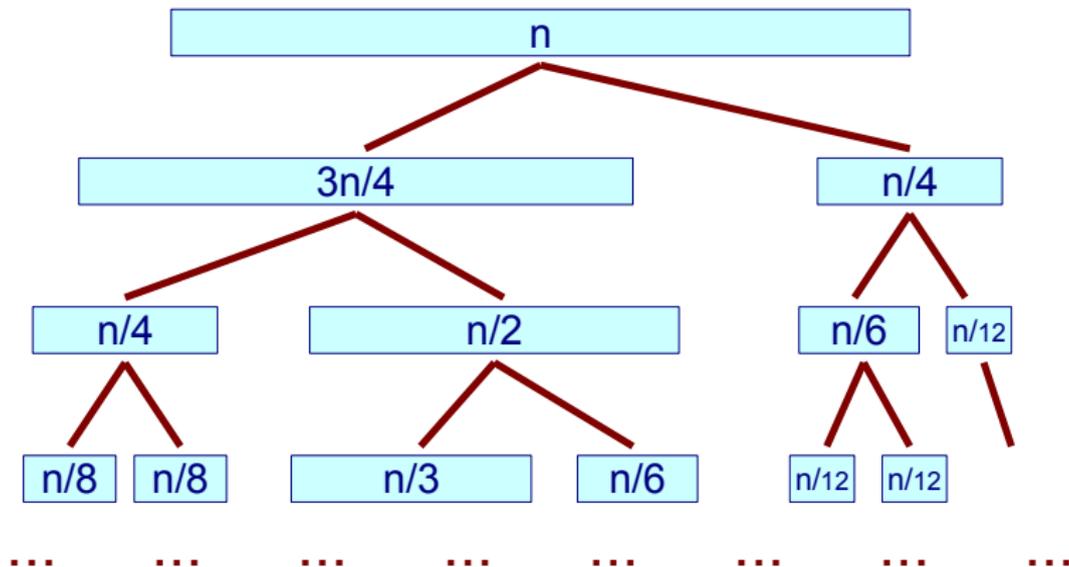
Interpretazione

Interpretazione di quanto osservato:

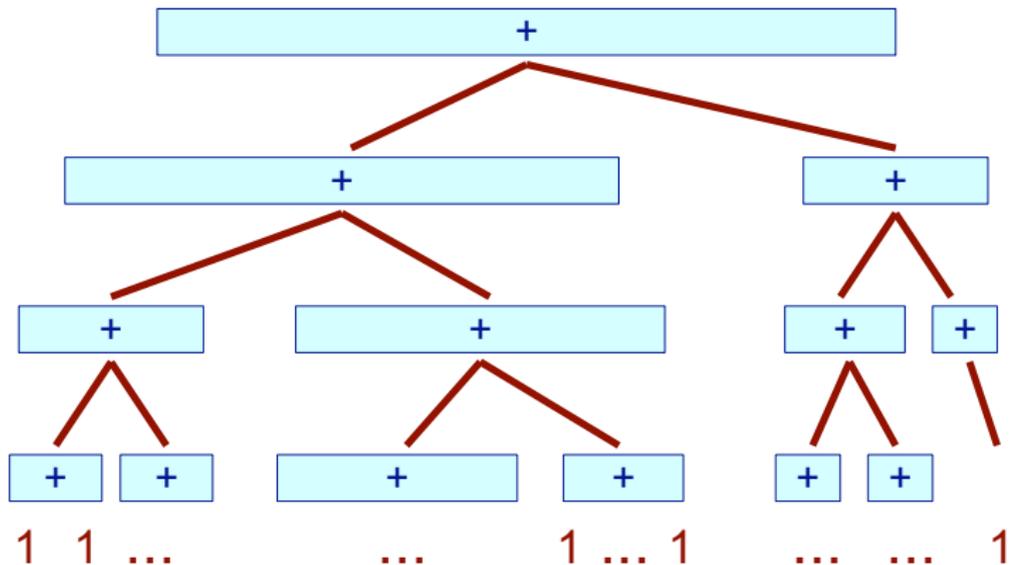
- $k \cdot n$?
- $\approx n$ è il numero di ricorsioni!

L'interpretazione va sempre messa in discussione

Numero complessivo di ricorsioni



Numero complessivo di ricorsioni





Numero complessivo di ricorsioni

Con la stessa struttura ricorsiva potrei “contare”,
ovvero sommare, il numero di elementi da ordinare
— per il resto è solo questione di parentesi:

- n unità e $n - 1$ somme, corrispondenti a $2n - 1$ ricorsioni:

$$1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$$

- albero molto sbilanciato:

$$(((((((1 + 1) + 1) + 1) + 1) + 1) + 1) + 1)$$

- albero perfettamente bilanciato:

$$(((1 + 1) + (1 + 1)) + ((1 + 1) + (1 + 1)))$$



Numero complessivo di ricorsioni

Con la stessa struttura ricorsiva potrei “contare”,
ovvero sommare, il numero di elementi da ordinare
— per il resto è solo questione di parentesi:

- n unità e $n - 1$ somme, corrispondenti a $2n - 1$ ricorsioni:

$$1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$$

- albero molto sbilanciato:

$$(((((((1 + 1) + 1) + 1) + 1) + 1) + 1) + 1)$$

- albero perfettamente bilanciato:

$$(((1 + 1) + (1 + 1)) + ((1 + 1) + (1 + 1)))$$



Numero complessivo di ricorsioni

Con la stessa struttura ricorsiva potrei “contare”, ovvero sommare, il numero di elementi da ordinare — per il resto è solo questione di parentesi:

- n unità e $n - 1$ somme, corrispondenti a $2n - 1$ ricorsioni:

$$1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$$

- albero molto sbilanciato:

$$(((((((1 + 1) + 1) + 1) + 1) + 1) + 1) + 1)$$

- albero perfettamente bilanciato:

$$(((1 + 1) + (1 + 1)) + ((1 + 1) + (1 + 1)))$$



Numero complessivo di ricorsioni

Con la stessa struttura ricorsiva potrei “contare”, ovvero sommare, il numero di elementi da ordinare — per il resto è solo questione di parentesi:

- n unità e $n - 1$ somme, corrispondenti a $2n - 1$ ricorsioni:

$$1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$$

- albero molto sbilanciato:

$$(((((((1 + 1) + 1) + 1) + 1) + 1) + 1) + 1)$$

- albero perfettamente bilanciato:

$$(((1 + 1) + (1 + 1)) + ((1 + 1) + (1 + 1)))$$



Interpretazione

Ulteriori osservazioni:

- Per n grande l'andamento è sostanzialmente stabile anche se si effettuano pochi ordinamenti. . .
- Per n grande le prestazioni relative a una sequenza casuale sono rappresentative del caso medio!



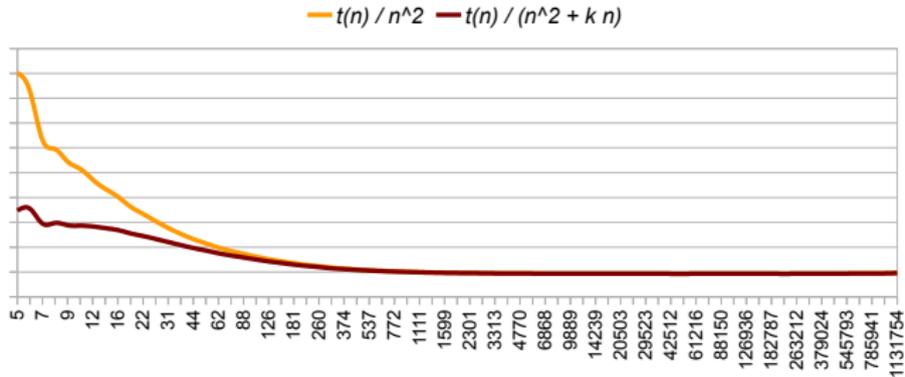
Interpretazione

Ulteriori osservazioni:

- Per n grande l'andamento è sostanzialmente stabile anche se si effettuano pochi ordinamenti. . .
- Per n grande le prestazioni relative a una sequenza casuale sono rappresentative del caso medio!

Trend dei costi di *InsertionSort*

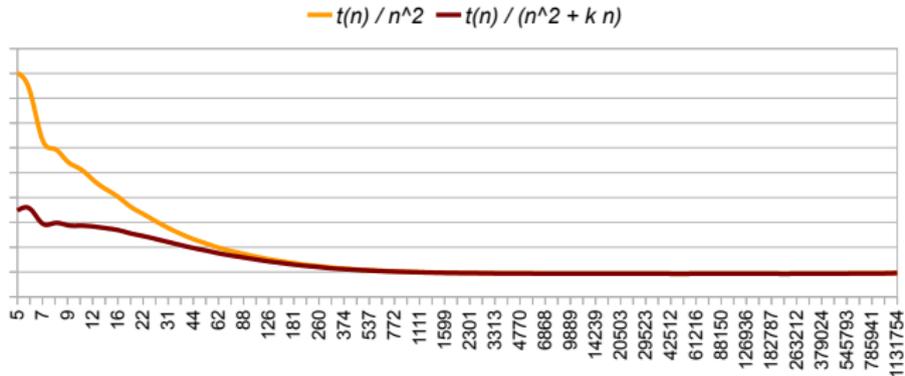
Simile analisi:



Stima di k ?

Trend dei costi di *InsertionSort*

Simile analisi:



Stima di k ?



Osservazioni inattese (?) ...

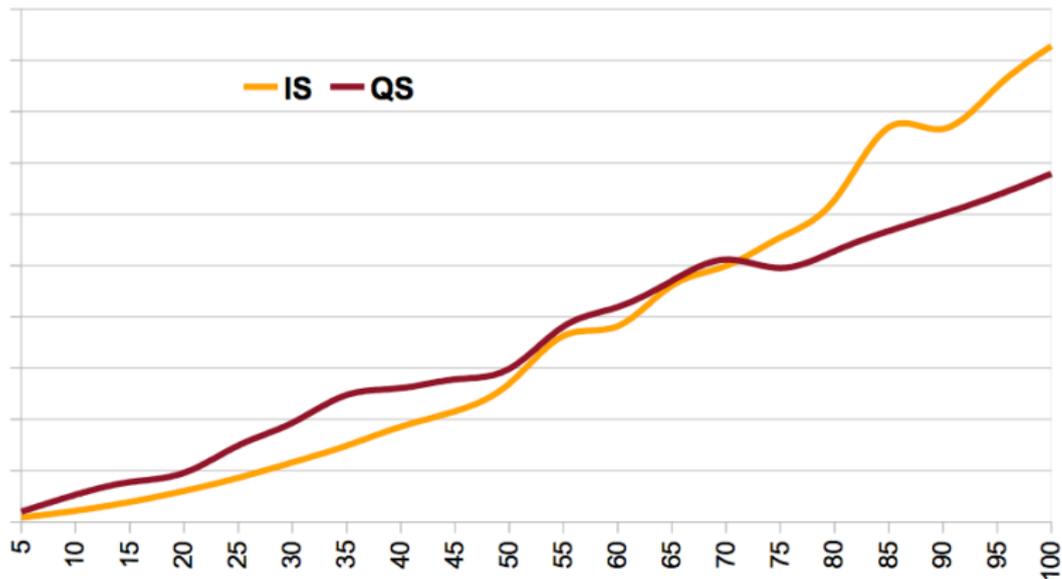


Osservazioni inattese (?) ...

Ogni nuova macchina che viene costruita è un esperimento [... e] interroga la natura; e noi ascoltiamo le risposte osservando la macchina in opera e analizzandola con tutti gli strumenti disponibili, analitici e di misura. Ciascun nuovo programma che viene costruito è un esperimento.

Newell & Simon, 1976

Osservazioni inattese (?) ... per sequenze corte





Ma, ad una analisi più accurata. . .

InsertionSort: In “media” $\frac{1}{2}k$ confronti per inserimento. . .

- Stima approssimativa nel caso medio
(circa metà delle volte va meglio e metà delle volte va peggio):

$$\frac{1}{2} \cdot \frac{n(n-1)}{2} \text{ (circa)} \approx \frac{1}{4} \cdot n^2$$

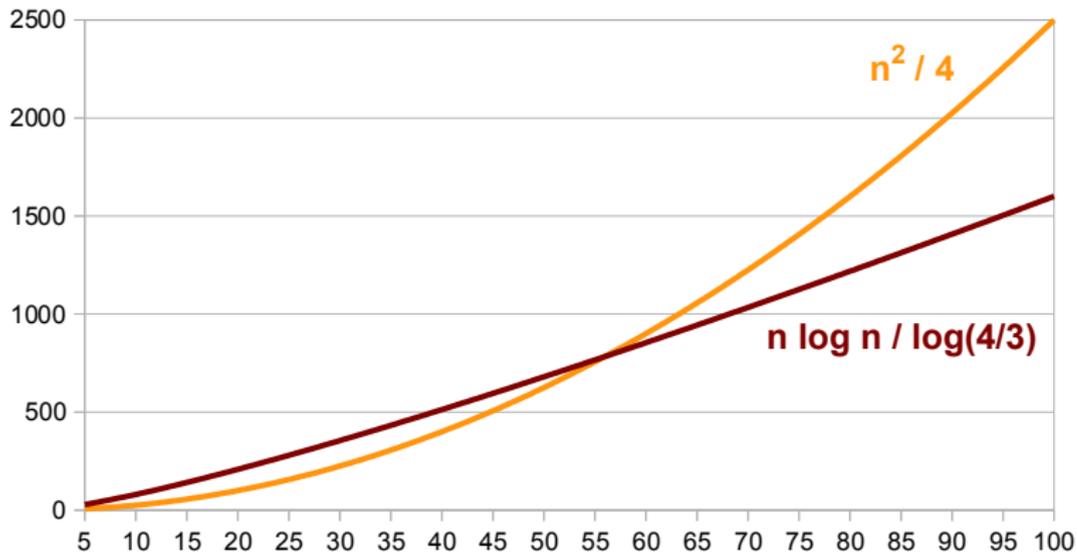
QuickSort: In “media” bipartizioni in proporzione $\frac{3}{4}$, $\frac{1}{4}$

- Stima approssimativa nel caso medio
(circa metà delle volte va meglio e metà delle volte va peggio):

$$n \cdot \log_{\frac{4}{3}} n = n \cdot \frac{1}{\log_2(\frac{4}{3})} \log_2 n \approx 2.41 \cdot n \log n$$

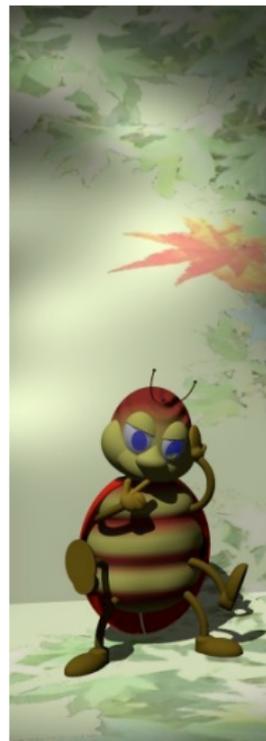


Raffinamento del modello



Trama

- 1 Retrospettiva
 - nei panni di un matematico
 - nei panni di uno scienziato
- 2 Indagine
 - osservazioni da interpretare
 - osservazioni da spiegare
- 3 Epilogo
 - uno scienziato è curioso
 - nei panni di un ingegnere





Cosa dire di *Selection Sort* ?

```
public void selectionSort( int[] seq ) {  
  
    int n = seq.length;  
  
    for ( int i=n-1; i>0; i-- ) {  
  
        int k = 0, x = seq[0];  
  
        for ( int j=1; j<=i; j++ ) {  
            if ( seq[j] > x ) {  
                k = j; x = seq[j];  
            }  
        }  
        seq[k] = seq[i]; seq[i] = x;  
    }  
}
```



Cosa dire di *Heap Sort* ?

```
public void heapSort( int[] seq ) {  
  
    int n = seq.length;  
  
    buildHeap( seq );  
  
    for ( int k=n; k>2; k-- ) {  
  
        int z = seq[0];  
        dropHeap( 1, k-1, seq[k-1], seq );  
        seq[k-1] = z;  
    }  
    int z = seq[0];  
    seq[0] = seq[1];  seq[1] = z;  
}
```



Ancora qualche domanda. . . da scienziato

- Codice di *SelectionSort* e *HeapSort* a disposizione
- Eventuale visualizzazione
- Strumento di misura e spreadsheet. . .



Ancora qualche domanda. . . da scienziato

- Codice di *SelectionSort* e *HeapSort* a disposizione
- Eventuale visualizzazione
- Strumento di misura e spreadsheet. . .



Ancora qualche domanda. . . da scienziato

- Codice di *SelectionSort* e *HeapSort* a disposizione
- Eventuale visualizzazione
- Strumento di misura e spreadsheet. . .



Domande da ingegnere

- C'è modo di *applicare* le conoscenze acquisite?
- È possibile migliorare le prestazioni combinando assieme i “blocchi funzionali” *QuickSort* e *InsertionSort* ?
- In quali condizioni le prestazioni di ciascuna tecnica sono più o meno vantaggiose? Quali *tradeoff* ?
- Come si può assicurare l'*affidabilità* del sistema che ne risulta? come organizzare il processo di sviluppo?
- Quali sono le condizioni “ottimali” di integrazione delle due componenti? come regolarle su base empirica?



Domande da ingegnere

- C'è modo di *applicare* le conoscenze acquisite?
- È possibile migliorare le prestazioni combinando assieme i “blocchi funzionali” *QuickSort* e *InsertionSort* ?
- In quali condizioni le prestazioni di ciascuna tecnica sono più o meno vantaggiose? Quali *tradeoff* ?
- Come si può assicurare l'*affidabilità* del sistema che ne risulta? come organizzare il processo di sviluppo?
- Quali sono le condizioni “ottimali” di integrazione delle due componenti? come regolarle su base empirica?



Domande da ingegnere

- C'è modo di *applicare* le conoscenze acquisite?
- È possibile migliorare le prestazioni combinando assieme i “blocchi funzionali” *QuickSort* e *InsertionSort* ?
- In quali condizioni le prestazioni di ciascuna tecnica sono più o meno vantaggiose? Quali *tradeoff* ?
- Come si può assicurare l'*affidabilità* del sistema che ne risulta? come organizzare il processo di sviluppo?
- Quali sono le condizioni “ottimali” di integrazione delle due componenti? come regolarle su base empirica?



Domande da ingegnere

- C'è modo di *applicare* le conoscenze acquisite?
- È possibile migliorare le prestazioni combinando assieme i “blocchi funzionali” *QuickSort* e *InsertionSort* ?
- In quali condizioni le prestazioni di ciascuna tecnica sono più o meno vantaggiose? Quali *tradeoff* ?
- Come si può assicurare l'*affidabilità* del sistema che ne risulta? come organizzare il processo di sviluppo?
- Quali sono le condizioni “ottimali” di integrazione delle due componenti? come regolarle su base empirica?



Domande da ingegnere

- C'è modo di *applicare* le conoscenze acquisite?
- È possibile migliorare le prestazioni combinando assieme i “blocchi funzionali” *QuickSort* e *InsertionSort* ?
- In quali condizioni le prestazioni di ciascuna tecnica sono più o meno vantaggiose? Quali *tradeoff* ?
- Come si può assicurare l'*affidabilità* del sistema che ne risulta? come organizzare il processo di sviluppo?
- Quali sono le condizioni “ottimali” di integrazione delle due componenti? come regolarle su base empirica?



Domandina. . .

Che differenza c'è fra l'approccio allo studio
di un ingegnere e di uno scienziato?

Risposta:



Domandina. . .

Che differenza c'è fra l'approccio allo studio
di un ingegnere e di uno scienziato?

Risposta:

Allo scienziato interessa capire, all'ingegnere applicare