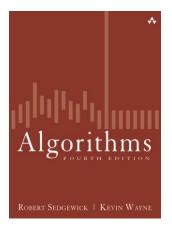
ID1020:elementary sorts

Ch 2.1



Slides adapted from Algorithms 4th Edition, Sedgewick.

A sorting problem

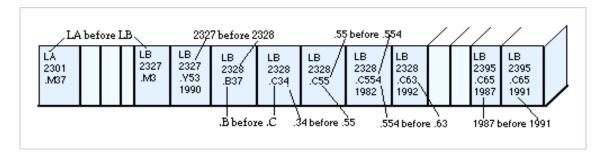
Eg. Student records at a university.



• Sorting. Order an array of N recods in ascending order.

Andrews	3	А	664-480-0023	097 Little
Battle	4	С	874-088-1212	121 Whitman
Chen	3	А	991-878-4944	308 Blair
Furia	1	А	766-093-9873	101 Brown
Gazsi	4	В	766-093-9873	101 Brown
Kanaga	3	В	898-122-9643	22 Brown
Rohde	2	А	232-343-5555	343 Forbes

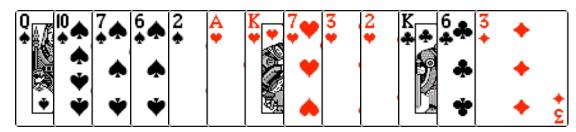
Sort applications



Library of Congress numbers



FedEx packages



playing cards



contacts

Exemple: sorting client (1)

- Target. Sort any possible types of data.
- Eg.1. Sort random numbers in ascending order.

```
public class Experiment
{
   public static void main(String[] args)
   {
      int N = Integer.parseInt(args[0]);
      Double[] a = new Double[N];
      for (int i = 0; i < N; i++)
          a[i] = StdRandom.uniform();
      Insertion.sort(a);
      for (int i = 0; i < N; i++)
          StdOut.println(a[i]);
   }
}</pre>
```

```
% java Experiment 10
0.08614716385210452
0.09054270895414829
0.10708746304898642
0.21166190071646818
0.363292849257276
0.460954145685913
0.5340026311350087
0.7216129793703496
0.9003500354411443
0.9293994908845686
```

Exemple: sorting client (2)

- Target. Sort any possible data types.
- Eg. 2. Sort all strings alphabetically.

```
public class StringSorter {
   public static void main(String[] args)
   {
     String[] a = StdIn.readAllStrings();
     Insertion.sort(a);
     for (int i = 0; i < a.length; i++)
        StdOut.println(a[i]);
   }
}</pre>
```

```
% more words3.txt
bed bug dad yet zoo ... all bad yes
% java StringSorter < words3.txt
all bad bed bug dad ... yes yet zoo [suppressing newlines]
```

Exempel: sorting client (3)

- Target. Sort any possible types of data.
- Eg. 3. Sort all files in a directory alphabetically by file name.

% java FileSorter.

Insertion.class

Insertion.java

InsertionX.class

InsertionX.java

Selection.class

Selection.java

Shell.class

Shell.java

ShellX.class

ShellX.java

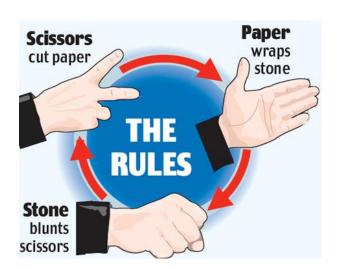
Total order

- Goal. Sort all possible data types (when sorting is well-defined).
- A total order is a binary relation

 which satisfy:
 - antisymmetric: IF $(v \le w \text{ and } w \le v)$ THEN v = w.
 - transitive: IF ($v \le w$ and $w \le x$) THEN $v \le x$.
 - total: either $v \le w$ or $w \le v$ or both.

• Eg.

- Common order for integer and real numbers.
- Chronological order for dates and times.
- Alphabetic order for strings.
- Non-transitive. Stone-scissors-paper.



Callbacks

- Goal. Sort all possible data types (when sorting is well-defined).
- How can sort() know how to compare data of different types, such as Double, String, and java.io.File, without any knowledge of the data types keys?
- Callback = a reference to executable code.
 - The client passes an array of objects as argument to the sort() function.
 - The sort() function calls the compareTo() method of the object when needed.
- Implementing callbacks.
 - Java: interfaces.
 - C: function pointers.
 - C++: class-type functors.
 - C#: delegates.
 - Python, Perl, ML, Javascript: first-class functions.

Callbacks in Java

client

```
public class StringSorter
{
   public static void main(String[] args)
   {
      String[] a = StdIn.readAllStrings();
      Insertion.sort(a);
      for (int i = 0; i < a.length; i++)
            StdOut.println(a[i]);
   }
}</pre>
```

Comparable interface (built-in of Java)

```
public interface Comparable<Item>
{
    public int compareTo(Item that);
}
```

No dependencies of the String data type

Data type implementation

```
public class String
implements Comparable<String>
{
    ...
    public int compareTo(String b)
    {
        ...
        return -1;
        ...
        return +1;
        ...
        return 0;
    }
}
```

sort implementation

```
public static void sort(Comparable[] a)
{
   int N = a.length;
   for (int i = 0; i < N; i++)
      for (int j = i; j > 0; j--)
       if (a[j].compareTo(a[j-1]) < 0)
            exch(a, j, j-1);
      else break;
}</pre>
```

Comparable API

- Implement compareTo() so that v.compareTo(w):
 - It defines a total order;
 - returns a negative integer, zero, or a positive integer IF v is respectively less than, equal to, or greater than w.
 - Throws an exception if the types are incompatible or if either is null.



Less than (return -1)



Equal (return 0)



Greater than (return +1)

- Built-in comparable types. Integer, Double, String, Date, File, ...
- User-defined comparable types. Implement the Comparable interface instances of the class (object) could be compared by the CompareTo methoden.

Why should compareTo() return negative,0,positive

- Comparisons are often effectively implemented by subtraction
- Eg: String comparisons are common
- Simple and fast implementation:
 - Subtract the ASCII-values of each character until the difference is non-zero or the end of one or both strings are reached.

Implement the Comparable interface

• Date data type. Simplified version of java.util.Date

```
public class Date implements Comparable<Date>{
  private final int month, day, year;
  public Date(int m, int d, int y) {
     month = m;
     day = d;
     year = y;
  public int compareTo(Date that)
      if (this.year < that.year ) return -1;
     if (this.year > that.year ) return +1;
      if (this.month < that.month) return -1;
      if (this.month > that.month) return +1;
     if (this.day < that.day ) return -1;
      if (this.day > that.day ) return +1;
     return 0;
```

Compare Date objects only to other Date objects

A non-transitive order

Why does not compareTo() implement a total order?

```
public class Double implements Comparable<Double> {
    private double x;
    ...

public int compareTo(Double that) {
    if (this.x < that.x) return -1;
    else if (this.x > that.x) return +1;
    else return 0;
}
```

```
compareTo() is non-transitiv!
It does not handle: -0.0 vs. 0.0 och Double.NaN
```

Elementary sorting

Two usefull operations

- Operations. Comparison and swap operations on elements (objects).
- Less-than. Is the element (object) v less than w?

```
private static boolean less(Comparable v, Comparable w) {
    return v.compareTo(w) < 0;
}</pre>
```

• Swap (exchange). Swap the element in the array a[] at index j to the element at index k.

```
private static void exch(Comparable[] a, int j, int k)
{
   Comparable swap = a[j];
   a[j] = a[k];
   a[k] = swap;
}
```

Mathematical analysis

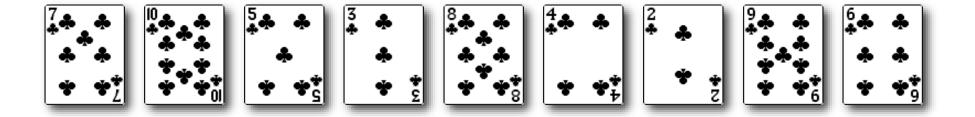
- We use the following cost model for the elementary sorting algorithms:
 - 1. Number of comparisons
 - 2. Number of exchanges (swaps)
- If the algorithm does not swap elements, then we count the number of array accesses.



Insertion Sort (Insättningssortering)

Insertion sort

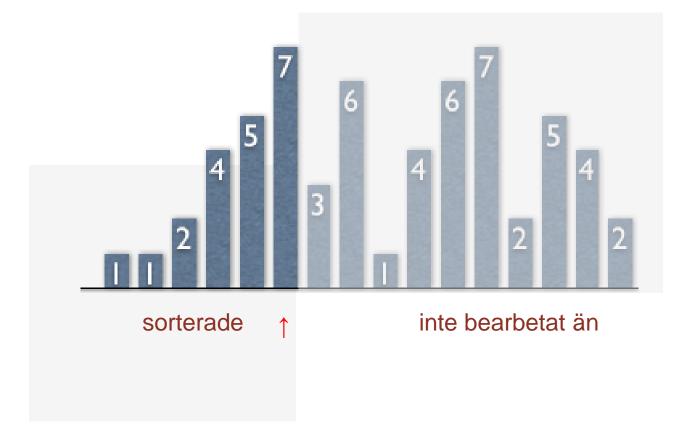
• During iteration i, a[i] swaps with every element that is larger on the left hand side.





Insertion sort

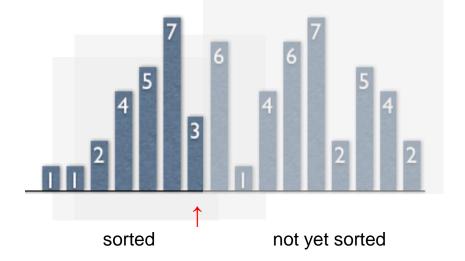
- Algorithm. ↑ scans from left to right.
- Invariants.
 - Elements to the left of ↑ (including ↑) are sorted in ascending order.
 - Elements to the right of ↑ has not been inspected (sorted) yet.



Insertion sort inner loop

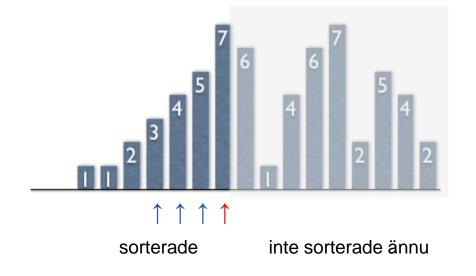
- Preserve the invariants:
 - Move the pointer (index) to the right.

```
i++;
```



 As an index is moved from right to left swap a[i] with every element on the left hand side which is larger.

```
for (int j = i; j > 0; j--)
  if (less(a[j], a[j-1]))
     exch(a, j, j-1);
  else break;
```



Insertion sort: Java implementation

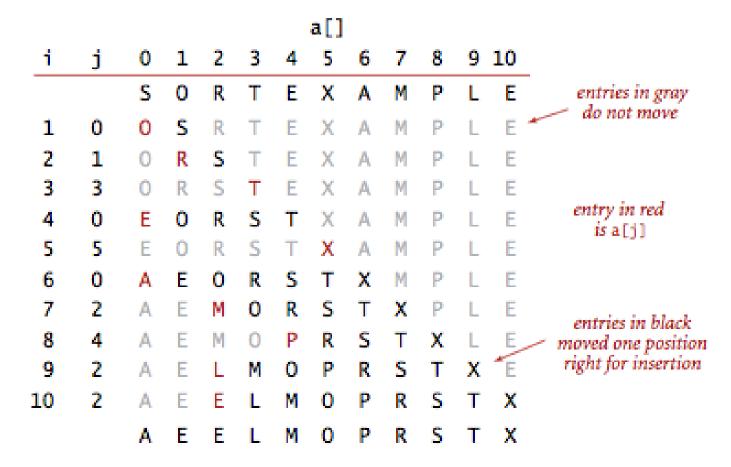
```
public class Insertion
  public static void sort(Comparable[] a)
      int N = a.length;
      for (int i = 0; i < N; i++) {
         for (int j = i; j > 0; j--) {
            if (less(a[j], a[j-1])) {
               exch(a, j, j-1);
            } else {
                 break;
   private static boolean less (Comparable v, Comparable w)
   {    /* as before */ }
   private static void exch(Comparable[] a, int i, int j)
   {    /* as before */ }
```

Insertion sort

https://en.wikipedia.org/wiki/Insertion_sort

Insertion sort: mathematical analysis

- Proposition. To sort an array of randomly ordered elements with unique keys, insertion sort do $\sim \frac{1}{4}$ N 2 comparisons and swaps $\sim \frac{1}{4}$ N 2 elements on average.
- Proof. Each of the N elements is compared to N/4 elements and is moved N/4 of steps on average.



Insertion sort: trace

a[]

																			a[]																	
i	j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
		Α	S	О	М	Ε	W	Н	Α	Т	L	О	Ν	G	Ε	R	1	Ν	S	Ε	R	Т	1	О	Ν	S	О	R	Т	Ε	Х	Α	М	Р	L	Ε
0	0	Α	S	\circ	M	Ε	$\vee\!\!\vee$	Н	Α	Т	L	\circ	Ν	G	Ε	R	1	Ν	S	Ε	R	Т	1	0	Ν	S	\circ	R	Т	Ε	Х	А	M	Р	L	Ε
1	1	А	S	0	M	Ε	W	Н	А	Т	L	0	Ν	G	Ε	R	1	Ν	S	Ε	R	Т	1	0	Ν	S	0	R	Т	Ε	Х	А	M	Р	L	Ε
2	1	А	О	S	M	Ε	W	Н	А	Т	L	\circ	Ν	G	Ε	R	1	Ν	S	Ε	R	Т		0	Ν	S	\circ	R	Т	Ε	Х	А	M	Р	L	Ε
3	1	А	М	О	S	Ε	W	Н	А	Т	L	0	Ν	G	Ε	R	1	Ν	S	Ε	R	Т	1	0	Ν	S	0	R	Т	Ε	X	А	M	Р	L	Ε
4	1	А	Ε	М	О	S	W	Н	А	Т	L	\circ	Ν	G	Ε	R	1	Ν	S	Ε	R	Т		0	Ν	S	0	R	Т	Ε	X	А	M	Р	L	Ε
5	5	А	Ε	$[\vee]$	\circ	S	W	Н	А	Т	L	\circ	Ν	G	Ε	R	1	Ν	S	Ε	R	Т	1	0	Ν	S	\circ	R	Т	Ε	Х	А	M	Р	L	Ε
6	2	А	Ε	Н	М	О	S	W	А	Т	L	0	Ν	G	Ε	R	1	Ν	S	Ε	R	Т	1	0	Ν	S	0	R	Т	Ε	Х	А	M	Р	L	Ε
7	1	А	Α	Ε	Н	М	О	S	W	Т	L	\circ	Ν	G	Ε	R	1	Ν	S	Ε	R	Т		0	Ν	S	\circ	R	Т	Ε	Х	А	M	Р	L	Ε
8	7	А	Α	Ε	Н	M	0	S	Т	W	L	0	Ν	G	Ε	R	1	Ν	S	Ε	R	Т	1	0	Ν	S	0	R	Т	Ε	Х	А	M	Р	L	Ε
9	4	А	А	Е	Н	L	М	О	S	Т	W	\circ	Ν	G	Ε	R	1	Ν	S	Ε	R	Т		0	Ν	S	\circ	R	Т	Ε	Х	А	M	Р	L	Ε
10	7	А	Α	Ε	Н	L	M	0	О	S	Т	W	Ν	G	Ε	R	1	Ν	S	Ε	R	Т	1	0	Ν	S	0	R	Т	Ε	Х	А	M	Р	L	Ε
11	6	А	А	Е	Н	L	M	Ν	О	О	S	Т	W	G	Ε	R	1	Ν	S	Ε	R	Т		0	Ν	S	\circ	R	Т	Ε	Х	А	M	Р	L	Ε
12	3	А	Α	Ε	G	Н	L	М	Ν	О	О	S	Т	W	Ε	R	1	Ν	S	Ε	R	Т	1	0	Ν	S	0	R	Т	Ε	Х	А	M	Р	L	Ε
13	3	А	Α	Ε	Ε	G	Н	L	М	Ν	О	О	S	Т	W	R	1	Ν	S	Ε	R	Т		0	Ν	S	0	R	Т	Ε	X	А	M	Р	L	Ε
14	11	А	Α	Ε	Ε	G	Н	L	M	Ν	\circ	\circ	R	S	Т	W	-	Ν	S	Ε	R	Т		0	Ν	S	\circ	R	Т	Ε	Х	Α	M	Р	L	Ε
15	6	Α	Α	Ε	Ε	G	Н	1	L	М	Ν	О	О	R	S	Т	W	Ν	S	Е	R	Т		0	Ν	S	0	R	Т	Ε	X	Α	M	Р	L	Е
16	10	А	Α	Ε	Ε	G	Н		L	M	Ν	Ν	О	О	R	S	Т	W	S	Ε	R	Т		0	Ν	S	0	R	Т	Ε	Х	Α	M	Р	L	Ε
17	15	А	Α	Ε	Ε	G	Н	-	L	M	Ν	Ν	0	0	R	S	S	Т	W	Ε	R	Т		0	Ν	S	0	R	Т	Ε	Х	А	M	Р	L	Ε
18	4	А	Α	Ε	Ε	Ε	G	Н	-1	L	М	Ν	Ν	О	О	R	S	S	Т	W	R	Т	1	0	Ν	S	0	R	Т	Ε	Х	А	M	Р	L	Ε
19	15	А	Α	Ε	Ε	Ε	G	Н	1	L	M	Ν	Ν	0	0	R	R	S	S	Т	W	Т	1	0	Ν	S	0	R	Т	Ε	Х	Α	M	Р	L	Ε
20	19	А	А	Ε	Ε	Ε	G	Н		L	M	Ν	Ν	0	0	R	R	S	S	Т	Т	W		0	Ν	S	0	R	Т	Ε	Х	А	M	Р	L	Ε
21	8	Α	Α	Ε	Ε	Ε	G	Н	1	1	L	М	Ν	Ν	О	О	R	R	S	S	Т	Т	W	0	Ν	S	0	R	Т	Ε	Х	Α	M	Р	L	Ε
22	15	Α	Α	Ε	Ε	Ε	G	Н			L	M	Ν	Ν	0	0	O	R	R	S	S	Т	Т	W	Ν	S	0	R	Т	Ε	Χ	Α	M	Р	L	Ε
23	13	А	Α	Е	Ε	Ε	G	Н	1		L	M	Ν	Ν	Ν	О	О	О	R	R	S	S	Т	Т	W	S	0	R	Т	Ε	Х	Α	M	Р	L	Ε
24	21	А	Α	Е	Ε	Ε	G	Н			L	M	Ν	Ν	Ν	0	0	0	R	R	S	S	S	Т	Т	W	0	R	Т	Ε	Х	Α	M	Р	L	Ε
25	17	А	Α	Е	Ε	Е	G	Н			L	M	Ν	Ν	Ν	0	0	0	O	R	R	S	S	S	Т	Т	W	R	Т	Ε	Х	Α	M	Р	L	Ε
26	20	А	Α	Е	Ε	Е	G	Н			L	M	Ν	Ν	Ν	0	0	0	0	R	R	R	S	S	S	Т	Т	W	Т	Ε	Х	Α	M	Р	L	Е
27	26	Α	Α	Е	Ε	Ε	G	Н			L	M	Ν	Ν	Ν	0	0	0	0	R	R	R	S	S	S	Т	Т	Т	W	Ε	X	Α	M	Р	L	Ε
28	5	Α	Α	Е	Ε	Е	Ε	G	Н	ı	ı	L	М	N	Ν	N	О	О	О	О	R	R	R	S	S	S	Т	Т	Т	W	Х	Α	M	Р	L	Е
29	29	Α	Α	Ε	Ε	Ε	Е	G	Н			L	M	Ν	Ν	Ν	0	0	0	0	R	R	R	S	S	S	Т	Т	Т	W	Х	Α	M	Р	L	Ε
30	2	Α	Α	Α	Ε	Ε	Ε	Ε	G	Н	ı	ı	L	М	Ν	N	Ν	О	О	О	О	R	R	R	S	S	S	Т	Т	Т	w	Х	M	Р	L	Е
31	13	Α	Α	А	Ε	Ε	Е	Ε	G	Н			L	M	М	N	Ν	Ν	О	О	О	О	R	R	R	S	S	S	Т	Т	Т	W	Х	Р	L	Е
32	21	Α	Α	Α	Е	Ε	Е	Е	G	Н			L	M	M	N	Ν	Ν	0	0	0	0	Р	R	R	R	S	S	S	Т	Т	Т	W	Х	L	Е
33	12	Α	Α	Α	E	Ε	E	Е	G	Н	1		L	L	М	М	N	N	N	0	0	0	0	Р	R	R	R	S	S	S	Т	Т	Т	W	Х	Е
34	7	Α	Α	Α	E	E	Е	E	E	G	Н	I	I	L	L	М	М	N	N	N	0	О	0	О	Р	R	R	R	S	S	S	Т	Т	Т	W	Х
		Α	Α	Α	Ε	Ε	Ε	Ε	Ε	G	Н	I	I	L	L	М	М	Ν	Ν	Ν	О	О	О	О	Р	R	R	R	S	S	S	Т	Т	Т	W	Х

Insertion sort: analysis

• Best case. If the elements of the array are ordered in ascending order, then insertion sort will perform N-1 comparisons and swap 0 elements.

AEELMOPRSTX

• Worst case. If the elements of the array are ordered in descending order (unique keys), then insertion sort will perform $\sim \frac{1}{2} \, N^2$ comparisons abd swap $\sim \frac{1}{2} \, N^2$ elements.

XTSRPOMLFEA

Insertion sort: partially sorted arrays

Def. An inversion is a pair of keys that are not ordered (in wrong order).

A E E L M O T R X P S

T-R T-P T-S R-P X-P X-S

(6 inversions)

- Def. An array is partially sorted if the number of inversions are $\leq c N$.
 - Ex. 1. A fully sorted array has 0 inversions.
 - Ex. 2. A sub-array of length 10 apended to a sorted array of length N.

- Proposition. Insertion sort sorts any partially sorted array in linear time (O(N)).
- Proof. The number of swaps equals the number of inversions.

Number of comparisions = number of swaps + (N - 1)

Insertion sort: improvements

- Binary insertion sort. One can use binary serach to find the "insertion point" in the sorted part of the array.
 - Number of comparisions ~ N lg N.
 - Still $\sim N^2$ array accesses.



Binary search for the first key > K

Elementary sort summary

algorithm	best	average	worst
selection sort	N^2	N^2	N^2
insertion sort	N	N^2	N^2
Shellsort (3x+1)	N log N	?	$N^{3/2}$
Goal	N	N log N	N log N

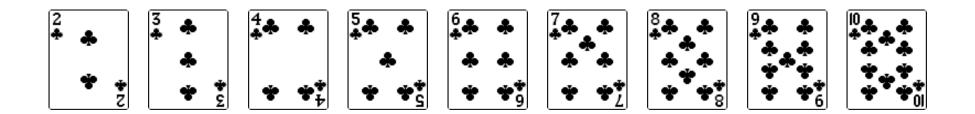
Time complexity to sort an array of N elements

Shuffling (Blandning)

How do you shuffle the elements of an array?

• Target. Shuffle the elements of an array so that the result is a random permutation of the elements.

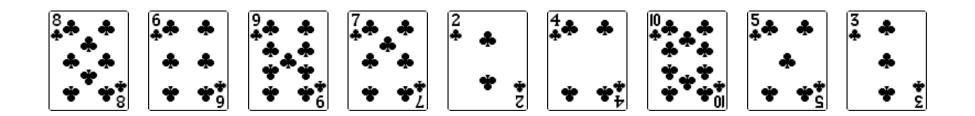
all permutations should be equally probable



How do you shuffle the elements of an array?

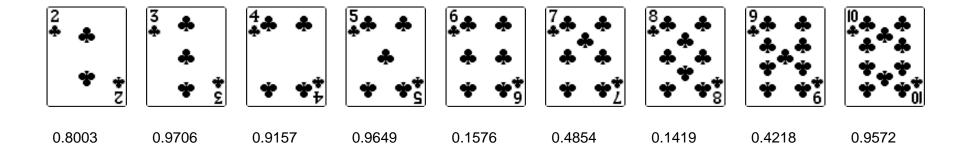
• Target. Shuffle the elements of an array so that the result is a random permutation of the elements.

all permutations should be equally probable



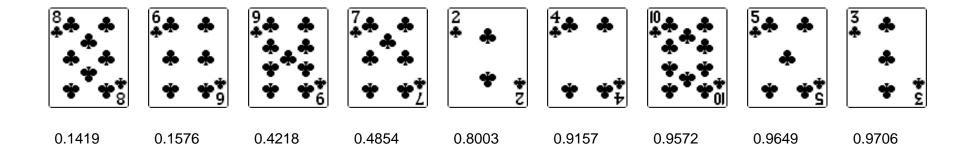
Shuffle sort

- Generate a random number associated with each element in the array.
- Sort the array.



Shuffle sort

- Generate a random number associated with each element in the array.
- Sort the array.



 Proposition. Shuffle sort results in a uniformly random permutation (depending on the quality of the random number generator).

Random number generation

- What we normally produce is a deterministic sequence of numbers exhibiting statistical properties of a random sequence (white noise)
- This is called pseudo-random number generators
- The sequence is determined by an initial value called seed
- Running the same random number generator several times with the same seed will generate the exact same sequences of numbers
- The best statistical properties are often generated if the seed is chosen as a large prime number
- That the sequences can be repeated if the seeds are known is helpful for making experiments (simulations) reproducible
- The quality of the algorithm used by different random number generators may vary substantially

The Microsoft Shuffle

• Microsoft antitrust probe by EU. Microsoft agreed to show a random sequence of the available browsers on the screen in Windows 7.

http://www.browserchoice.eu

Select your web browser(s)



A fast new browser from Google. Try it now!



Safari for Windows from Apple, the world's most innovative browser.



Your online security is Firefox's top priority. Firefox is free, and made to help you get the most out of the



The fastest browser on Earth. Secure, powerful and easy to use, with excellent privacy protection.



Designed to help you take control of your privacy and browse with confidence. Free from Microsoft.

Was placed last 50% of the time

The Microsoft Shuffle

 The Microsoft solution? Implemented a shuffle sort by implementing a comparator which should generate a random answer.

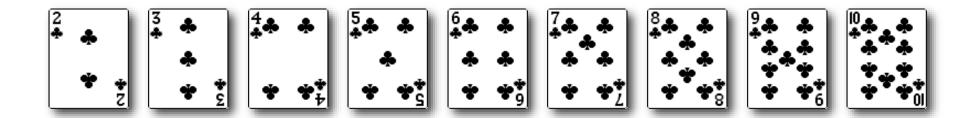
```
public int compareTo(Browser that)
{
   double r = Math.random();
   if (r < 0.5) return -1;
   if (r > 0.5) return +1;
   return 0;
}

browser comparator
(should implement a total order)
```

 Not reflexive, not antisymmetric, not transitive. If IE was the last in the nonpermitated array and the sorting is insertion sort it would end up last in 50% of the shuffles

Knuth shuffle

- In iteration i, generate a random integer r in the interval 0 to i.
- Swap a[i] and a[r].



• Proposition. [Fisher-Yates 1938] Knuth shuffle sort results in a uniformly random permutation of the input array in linear time.

Knuth shuffle

- In iteration i, generate a random integer r in the interval 0 to i.

- Swap a[i] and a[r].

```
common bug : r between 0 and N-1 correct: r between i and N-1
```

```
public class StdRandom
{
    ...
    public static void shuffle(Object[] a)
    {
        int N = a.length;
        for (int i = 0; i < N; i++)
        {
            int r = StdRandom.uniform(i + 1);
            exch(a, i, r);
        }
    }
    Between 0 and i
}
</pre>
```

Broken Knuth shuffle

- What if r is drawn from the interval 0 to N-1 ?
- No longer uniform distribution (random)!

Instead of 0 and i

АВС	1/6	4/27
АСВ	1/6	5/27
ВАС	1/6	5/27
ВСА	1/6	5/27
САВ	1/6	4/27
СВА	1/6	4/27

Probability of each permuation of (each shuffle of) { A, B, C }

Online Poker

Texas hold'em poker. The program should shuffle a deck of cards.



How We Learned to Cheat at Online Poker: A Study in Software Security http://www.datamation.com/entdev/article.php/616221

Online poker anecdote

Shuffling algorithm in FAQ at www.planetpoker.com

```
for i := 1 to 52 do begin
    r := random(51) + 1;
    swap := card[r];
    card[r] := card[i];
    card[i] := swap;
end;
```

- Bug 1. Random number r will never be 52 => $52^{:nd}$ card can never remain in the $52^{:nd}$ position.
- Bug 2. The shuffle is not uniform (should be between 1 and i).
- Bug 3. random() used a small 32-bit seed => 2³² possible shuffles.
- Bug 4. Seed = milli seconds since midnight => 86.4 million shuffles.
- Exploit. After having examined 5 cards and synchronized with the clock of the server, one can predict all future cards in real time.

— Robert R. Coveyou

[&]quot;The generation of random numbers is too important to be left to chance."

Online poker: best praxis

- Best praxis for shuffle.
 - Use a hardware solution to generate random numbers that has passed both FIPS 140-2 and NIST statistical tests.
 - But hardware random number generators may be fragile and could fail silently.
 - Use an independent generator such as random.org which use atmospheric noise to generate the random number sequence.





RANDOM.ORG

Summary

- Simple algorithms to sort arrays
 - Selection sort
 - Insertion sort
 - Shell sort
- Shuffle
 - To properly shuffle a deck of cards is hard!