Average preformance of InsertionSort

Average number of comparisons while inserting (i + 1) st element x to a sorted sequence

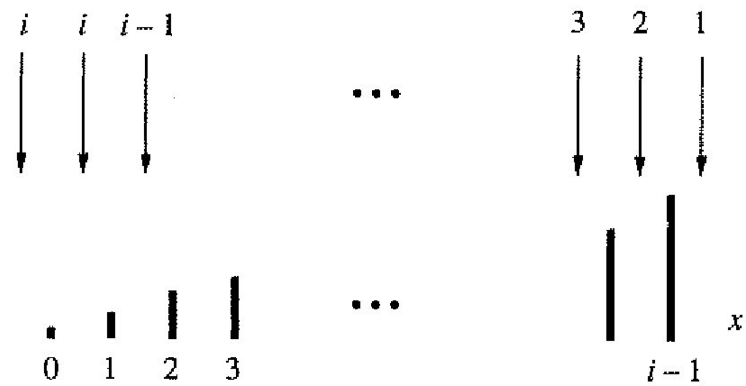


Figure 4.5 Number of comparisons needed to determine the position for x

E[0] < E[1] < ... < E[i - 1]   ← x

of i elements

The algorithm

Try at index i :

E[0] < E[1] < ... < E[i - 1] < _
↑?
x

If yes then

E[0] < E[1] < ... < E[i - 1] < x

(done after 1 comparison)

If no then try at index (i - 1) :

E[0] < E[1] < ... < E[i - 2] < _ < E[i]
↑?
x

If yes then

E[0] < E[1] < ... < E[i - 2] < x < E[i]

(done after 2 comps)

If no then try at index (i - 2) :

E[0] < E[1] < ... < E[i - 3] < _ < E[i - 1] < E[i]
↑?
x

If yes then

E[0] < E[1] < ... < E[i - 3] < x < E[i - 1] < E[i]

(done after 3 comps)

etc.

...

E[0] < E[1] < ... < E[j - 1] < x < E[j + 1] < ... < E[i - 1] < E[i]

(done after i - j + 1 comps ; we will derive this formula below)

etc.

...

If no then try at index 1 :

E[0] < _ < E[2] < ... < E[i]
↑?
x

If yes then

E[0] < x < E[2] < ... < E[i]

(done after i comps)

If no then

$x < E[1] < E[2] < ... < E[i]$

(done after how many comps?)

Let

• Pr (j) be the probability that the index of x after insertion in

$$E[0] < E[1] < ... < E[j-1] < \underset{\overset{\uparrow}{x}}{\_} < E[j+1] < ... < E[i-1] < E[i]$$

is j, where $j \in \{0, ..., i\}$

• # comp (j) be the number of comparisons necessary for the program to determine that x should be inserted at index j

We have (assuming that all i + 1 possible points of insertion are equally likely, see Figure 4.5 in your textbook) :

$$\Pr(j) = \frac{1}{(i+1)}$$

When j gets decremented after each "no" in the above algorithm,
comp (j) gets incremented, except for the last step when # comps (1) = # comps (0) = i.

Hence,

$$j + \#\text{comp}(j) = \text{constant}$$

for all $0 < j \leq i$.

Putting j = i (the first step of the algorithm) yields

$$i + 1 = \text{constant}$$

because # comps (i) = 1.

Therefore, for all $0 < j \leq i$,

$$j + \#\text{comp}(j) = i + 1$$

or, (subtracting j from both sides)

$$\#comp(j) = i - j + 1.$$

So,

$$\#comp(j) = \begin{cases} i - j + 1 & \text{for } 0 < j \le i \\ i & \text{for } j = 0 \end{cases}$$

The average number of comparisons while inserting $(i+1)$ st element (where $i \ge 0$) is :

$$\sum_{j=0}^{i} Pr(j) \times \#comp(j) = Pr(0) \times \#comp(0) + \sum_{j=1}^{i} Pr(j) \times \#comp(j) =$$

$$\frac{1}{(i+1)} \times i + \sum_{j=1}^{i} \frac{1}{(i+1)} \times (i - j + 1)$$

$$\frac{i}{2} + \frac{i}{1+i}$$

$$\frac{i}{2} + \frac{i}{1+i} = \frac{i}{2} + \frac{i+1}{1+i} - \frac{1}{1+i} = \frac{i}{2} + 1 - \frac{1}{1+i}$$

Verification of the above

$$\frac{i}{2} + \text{Factor}\left[1 - \frac{1}{1+i}\right]$$

$$\frac{i}{2} + \frac{i}{1+i}$$

The average number of comparisons for all insertions is (remember that $Pr(0) \times \#comp(0) = 0$ so the first insertion at $i = 0$ may be ignored) :

$$T_{avg}(n) = \sum_{i=1}^{n-1} \sum_{j=0}^{i} Pr(j) \times \#comp(j) =$$

$$= \sum_{i=1}^{n-1} \left(\frac{i}{2} + \frac{i}{1+i}\right) =$$

$$= \sum_{i=1}^{n-1} \left(\frac{i}{2} + 1 - \frac{1}{i+1}\right) = \sum_{i=1}^{n-1} \left(\frac{i}{2} + 1\right) - \sum_{i=1}^{n-1} \frac{1}{1+i}$$

$$\sum_{i=1}^{n-1} \frac{1}{1+i} = (\text{put } j = i+1;\ j \in \{2, \ldots n\})$$

$$= \sum_{j=2}^{n} \frac{1}{j} = \sum_{j=1}^{n} \frac{1}{j} - \frac{1}{1} = \sum_{j=1}^{n} \frac{1}{j} - 1.$$

So,

$$T_{avg}(n) = \sum_{i=1}^{n-1}\left(\frac{i}{2}+1\right) - \left(\sum_{j=1}^{n}\frac{1}{j} - 1\right) = \sum_{i=1}^{n-1}\left(\frac{i}{2}+1\right) + 1 - \sum_{j=1}^{n}\frac{1}{j}.$$

$$\sum_{i=1}^{n-1}\left(\frac{i}{2}+1\right) + 1$$

$$1 + \frac{1}{4}\left(-4 + 3\,n + n^2\right)$$

$$\text{Expand}\left[1 + \frac{1}{4}\left(-4 + 3\,n + n^2\right)\right]$$

$$\frac{3\,n}{4} + \frac{n^2}{4}$$

$$\text{FullSimplify}\left[\frac{3\,n}{4} + \frac{n^2}{4}\right]$$

$$\frac{1}{4}\,n\,(3 + n)$$

Thus $\sum_{i=1}^{n-1}\left(\frac{i}{2}+1\right) + 1 = \frac{1}{4}\,n\,(n+3)$

Thus

$$T_{avg}(n) = \frac{1}{4}\,n\,(3+n) - \sum_{j=1}^{n}\frac{1}{j}$$

Recall that

$$\sum_{j=1}^{n}\frac{1}{j} \approx \text{Log}[n] + 0.577216$$

(Euler ' s formula for harmonic number, see file Summationa.nb)

Hence,

$$T_{avg}(n) \approx \frac{1}{4}\,n\,(n+3) - \text{Log}[n] - 0.577216 =$$

$$-0.577216\grave{} + 0.75\grave{}\,n + 0.25\grave{}\,n^2 - \text{Log}[n] \in \Theta\left(n^2\right)$$

Note. The difference between $\frac{1}{4}\,n\,(n+3)$ and textbook ' s lower bound for average number of comps $\frac{1}{4}\,n\,(n-1)$ is n.

```
Simplify[1/4 n (n + 3) - 1/4 n (n - 1)]
```

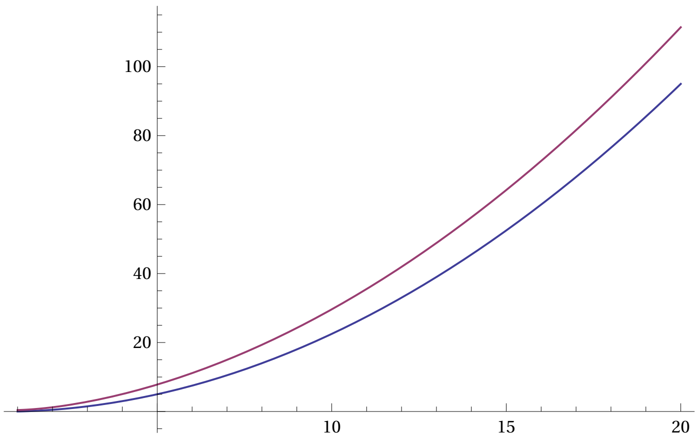
n

Thus the difference between the upper bound and textbook's lower bound is

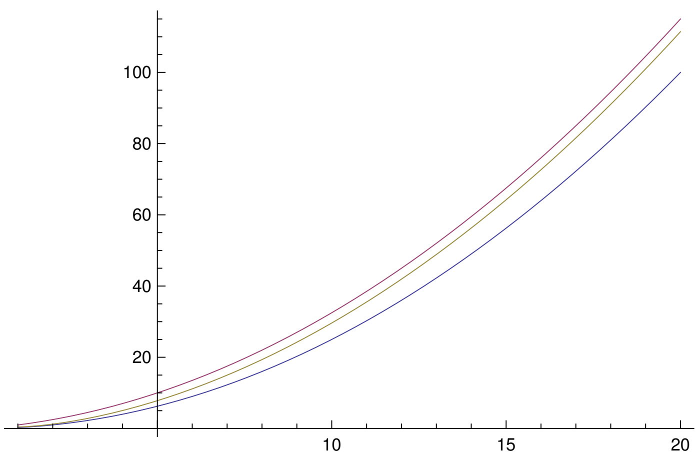
$n - \sum_{j=1}^{n} \frac{1}{j} \approx n - (\text{Log}[n] + 0.577216)$

In[5]:=
```
Plot[Tooltip[{1/4 n (n - 1), 1/4 n (3 + n) - Log[n] - 0.577216}],
 {n, 1, 20}, PlotTheme → "Classic"]
```

Out[5]=




```
Plot[Tooltip[{0.25` n^2, 1/4 n (3 + n), 1/4 n (3 + n) - Log[n] - 0.577216}], {n, 1, 20}]
```

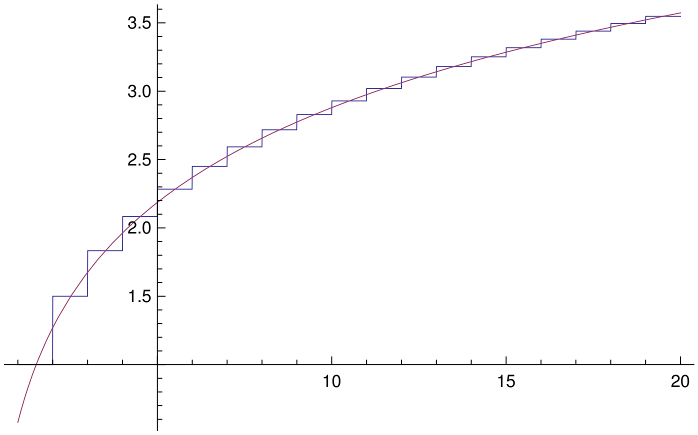



```
Limit[1/n^2 (-0.577216` + 0.75` n + 0.25` n^2 - 1.` Log[n]), n → ∞]
```

0.25

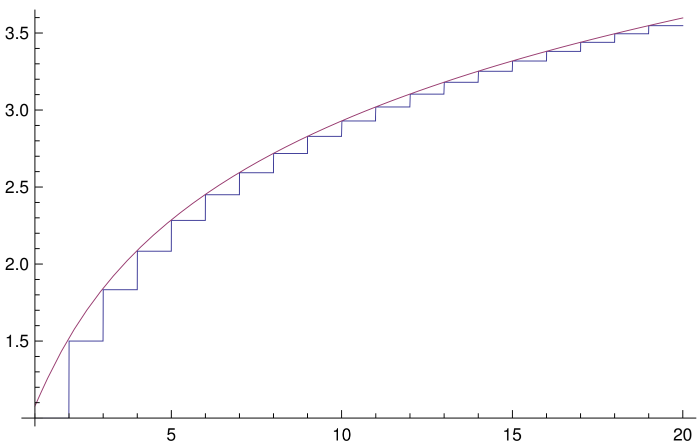
Euler's formula, again:

$$\text{Plot}\left[\left\{\sum_{j=1}^{n} \frac{1}{j}, \text{Log}[n] + 0.577216\right\}, \{n, 1, 20\}\right]$$



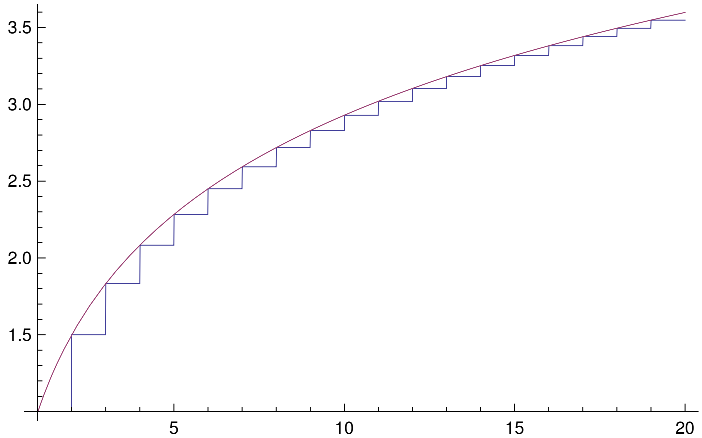
More accurate :

$$\text{Plot}\left[\left\{\sum_{j=1}^{n} \frac{1}{j}, \text{Log}[n] + 0.577216 + \frac{1}{2\,n}\right\}, \{n, 1, 20\}\right]$$



Even more accurate

$$\text{Plot}\left[\left\{\sum_{j=1}^{n}\frac{1}{j},\ \text{Log}[n]+0.577216+\frac{1}{2\,n}-\frac{1}{12\,n^2}\right\},\ \{n,\ 1,\ 20\}\right]$$

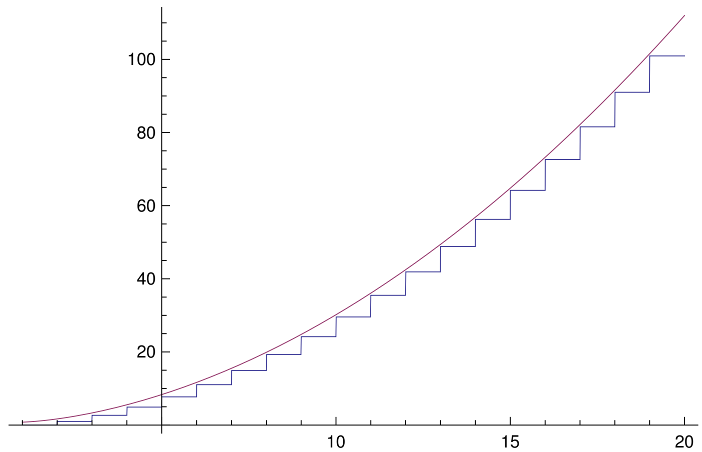


They yield more precise (optional for all students) aproximation of the average number of comparisons

$$T_{avg}\ (n)\ \approx\ \frac{1}{4}\,n\,(n+3)\ -\ \text{Log}[n]\ -\ 0.577216-\frac{1}{2\,n}+\frac{1}{12\,n^2}$$

Here is a visual verification of the above :

$$\text{Plot}\left[\left\{\sum_{i=1}^{n-1}\left(\frac{i}{2}+\frac{i}{1+i}\right),\ \frac{1}{4}\,n\,(n+3)-\left(\text{Log}[n]+0.577216+\frac{1}{2\,n}-\frac{1}{12\,n^2}\right)\right\},\ \{n,\ 1,\ 20\}\right]$$



In[11]:=
```
Plot[{1/4 n (n - 1), Sum[(i/2 + i/(1 + i)), {i, 1, n - 1}], 1/4 n (n + 3) - (Log[n] + 0.577216 (+ 1/(2 n)) - 1/(12 n^2))},
 {n, 1, 20}, PlotTheme → "Classic"]
```

Out[11]=