Average number of comps in successful and unsuccessful search in average binary search tree

Below is the average number of comps of Quicksort from file Average-case_Quicksort_new.pdf

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

Here is its really good approximation experimentally derived with Mathematica

$$2(n+1)\left(\text{Log}[n]+\text{EulerGamma}+\frac{1}{2n}-\frac{1}{12n^2}+\frac{1}{120n^4}-\frac{1}{252n^6}+\frac{1}{240n^8}-\frac{1}{132n^{10}}+691\Big/\left(32760n^{12}\right)-\frac{1}{12n^{14}}+\frac{3617}{8160n^{16}}-43867\Big/\left(14364n^{18}\right)+\frac{174611}{6600n^{20}}-\frac{77683}{276n^{22}}+236364091\Big/\left(65520n^{24}\right)-\frac{657931}{12n^{26}}+3392780147\Big/\left(3480n^{28}\right)-1723168255201\Big/\left(85932n^{30}\right)+7709321041217\Big/\left(16320n^{32}\right)\right)-4n$$

where

In[21]:= N[EulerGamma]

Out[21]= 0.577216

0.5772156649015329`

We will use a rougher,
but still good (derived experimentally in file Summationa.nb) approximation

$$2(n+1)\left(\text{Log}[n]+0.5772156649015329+\frac{1}{2n}-\frac{1}{12n^2}\right)-4n$$

Since QuickSort is simulated by
BS-treeSort in that QuickSort builds a
BS tree T of pivots while sorting and makes
ipl(T) comparisons of keys, the average number of comparisons
done by QuickSort is the same as the average ipl in an average BS tree.

So, the average epl in an average BS tree is the above number pus 2 n, that is:

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

$$2(n+1)\left(\text{Log}[n]+0.5772156649015329+\frac{1}{2n}-\frac{1}{12n^2}\right)-2n$$

The average number of comps in a successful search in an average BS tree is

$$c_n=\frac{\text{ipl}}{n}+1=$$

$$\frac{1}{n}\left(2\,(n+1)\left(\text{Log}[n]+0.5772156649015329+\frac{1}{2\,n}-\frac{1}{12\,n^2}\right)-4\,n\right)+1$$

In[22]:= $$\text{Expand}\left[\frac{1}{n}\left(2\,(n+1)\left(\text{Log}[n]+0.5772156649015329+\frac{1}{2\,n}-\frac{1}{12\,n^2}\right)-4\,n\right)+1\right]$$

Out[22]= $$-1.84557-\frac{1}{6\,n^3}+\frac{5}{6\,n^2}+\frac{2.15443}{n}+2\,\text{Log}[n]+\frac{2\,\text{Log}[n]}{n}$$

$$-1.8455686701969345`-\frac{1}{6\,n^3}+\frac{5}{6\,n^2}+2.1544313298030655`/n+2\,\text{Log}[n]+\frac{2\,\text{Log}[n]}{n}$$

Putting

$$2\,\text{Log}[n]\approx 1.3862943611198906`\text{Log2}[n]$$

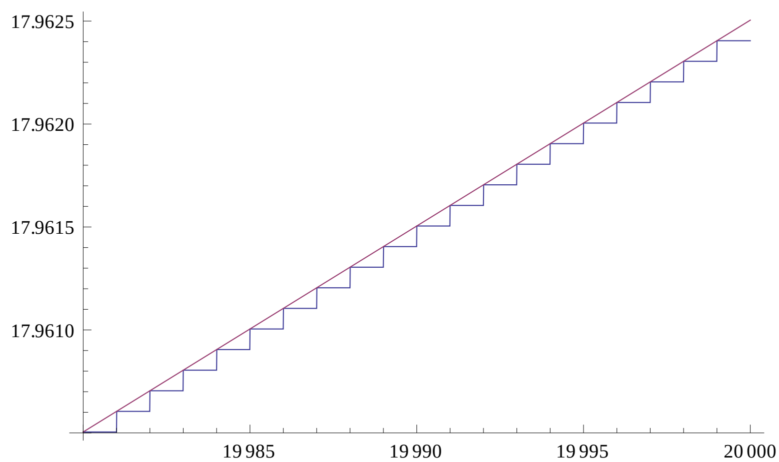
this yields the approximation from the file
2 trees.pdf

$$1.3862943611198906\text{Log2}[n]-1.8455686701969345+\frac{1}{n}\,1.3862943611198906\text{Log2}[n]+2.1544313298030655/n+\frac{5}{6\,n^2}-\frac{1}{6\,n^3}$$

Here is the graph of the exact value and its approximation

In[19]:= $$\text{Plot}\left[\left\{\frac{1}{n}\left(2\,(n+1)\left(\sum_{i=1}^{n}\frac{1}{i}\right)-4\,n\right)+1,\right.\right.$$
$$1.3862943611198906`\text{Log2}[n]-1.8455686701969343`+\frac{1}{n}\,1.3862943611198906`\text{Log2}[n]+$$
$$\left.\left.2.1544313298030655`/n+\frac{5}{6\,n^2}-\frac{1}{6\,n^3}\right\},\{n,19980,20000\}\right]$$

Out[19]=



Perfect match!

Here is the difference between the two at n = 10

In[17]:= $$\text{Table}\left[\frac{1}{n}\left(2\,(n+1)\left(\sum_{i=1}^{n}\frac{1}{i}\right)-4\,n\right)+1-\left(1.3862943611198906\text{`}\,\text{Log2}[n]-1.8455686701969343\text{`}+\frac{1}{n}\,1.3862943611198906\text{`}\,\text{Log2}[n]+2.1544313298030655\text{`}/n+\frac{5}{6\,n^2}-\frac{1}{6\,n^3}\right),\{n,10,10.0004\}\right]$$

Out[17]= $\{1.82469\times 10^{-6}\}$

The average number of comps in an unsuccessfulsearch in an average BS tree is

$$c'_n=\frac{\text{epl}}{n+1}=$$

$$\frac{1}{n+1}\left(2\,(n+1)\left(\text{Log}[n]+0.5772156649015329+\frac{1}{2\,n}-\frac{1}{12\,n^2}\right)-2\,n\right)$$

$$2\left(\text{Log}[n]+0.5772156649015329+\frac{1}{2\,n}-\frac{1}{12\,n^2}\right)-2\,\frac{n}{n+1}$$

In[23]:= $$\text{Expand}\left[2\left(\text{Log}[n]+0.5772156649015329+\frac{1}{2\,n}-\frac{1}{12\,n^2}\right)-2\,\frac{n}{n+1}\right]$$

Out[23]= $$1.15443-\frac{1}{6\,n^2}+\frac{1}{n}-\frac{2\,n}{1+n}+2\,\text{Log}[n]$$

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

$$\text{Together}\left[2-\frac{2}{1+n}\right]$$

$$\frac{2\,n}{1+n}$$

$$1.1544313298030657\text{`}-\frac{1}{6\,n^2}+\frac{1}{n}-2+\frac{2}{1+n}+2\,\text{Log}[n]$$

In[25]:= $1.1544313298030657\text{`}-2$

Out[25]= -0.845569

$$-0.8455686701969343\text{`}-\frac{1}{6\,n^2}+\frac{1}{n}+\frac{2}{1+n}+2\,\text{Log}[n]$$

N[2 Log[2]]

1.38629

1.3862943611198906`

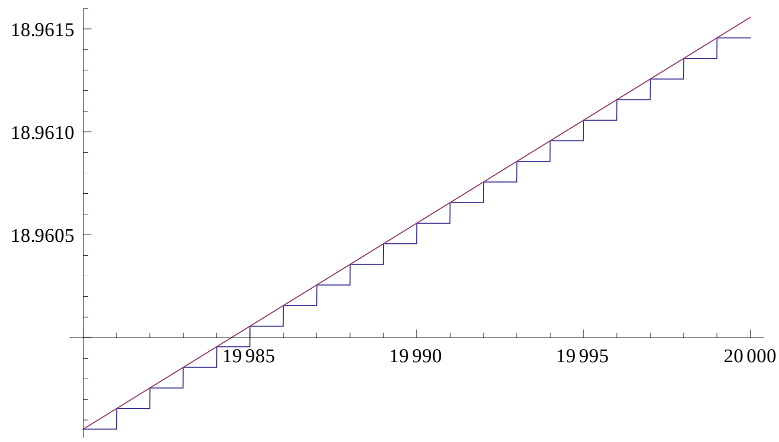
This yields the approximationfrom the file
2 trees.pdf

$1.3862943611198906`\text{Log2}[n] + -0.8455686701969343` + \frac{1}{n} + \frac{2}{n+1} - \frac{1}{6\,n^2}$

Here is a graph of the exact value and approximation

In[27]:= $\text{Plot}\left[\left\{\frac{1}{n+1}\left(2\,(n+1)\left(\sum_{i=1}^{n}\frac{1}{i}\right) - 2\,n\right),\; 1.3862943611198906`\text{Log2}[n] + -0.8455686701969343` + \frac{1}{n} + \frac{2}{n+1} - \frac{1}{6\,n^2}\right\}, \{n, 19980, 20000\}\right]$

Out[27]=



Again, perfect match!

Here is the difference between the two at $n = 10$

In[28]:= $\text{Table}\left[\frac{1}{n+1}\left(2\,(n+1)\left(\sum_{i=1}^{n}\frac{1}{i}\right) - 2\,n\right) - \left(1.3862943611198906`\text{Log2}[n] + -0.8455686701969343` + \frac{1}{n} + \frac{2}{n+1} - \frac{1}{6\,n^2}\right), \{n, 10, 10.0004\}\right]$

Out[28]= $\{1.65881 \times 10^{-6}\}$

Here is the difference between c'_n and c_n

In[36]:= $\text{Simplify}\left[\left(\frac{1}{n+1}\left(2\,(n+1)\left(\sum_{i=1}^{n}\frac{1}{i}\right) - 2\,n\right)\right) - \left(\frac{1}{n}\left(2\,(n+1)\left(\sum_{i=1}^{n}\frac{1}{i}\right) - 4\,n\right) + 1\right)\right]$

Out[36]= (n (3 + n) - 2 (1 + n) HarmonicNumber[n]) / (n (1 + n))

In[39]:= $\sum_{i=1}^{n}\frac{1}{i}$

Out[39]= HarmonicNumber[n]

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

$$\frac{n+3}{n+1} - \frac{2\sum_{i=1}^{n}\frac{1}{i}}{n}$$

$$\frac{n+3}{n+1} - \frac{1}{n}\,2\left(\text{Log}[n] + 0.5772156649015329 + \frac{1}{2\,n} - \frac{1}{12\,n^2}\right)$$

In[43]:= $\text{Expand}\left[-\frac{1}{n}\,2\left(\text{Log}[n] + 0.5772156649015329 + \frac{1}{2\,n} - \frac{1}{12\,n^2}\right)\right]$

Out[43]= $\frac{1}{6\,n^3} - \frac{1}{n^2} - \frac{1.15443}{n} - \frac{2\,\text{Log}[n]}{n}$

$$\frac{n+1}{n+1} + \frac{2}{n+1} - \frac{2\,\text{Log}[n]}{n} - 1.1544313298030657` / n - \frac{1}{n^2} + \frac{1}{6\,n^3}$$

$$1 - \frac{2\,\text{Log}[n]}{n} - 1.1544313298030657` / n + \frac{2}{n+1} - \frac{1}{n^2} + \frac{1}{6\,n^3}$$

or, approximately

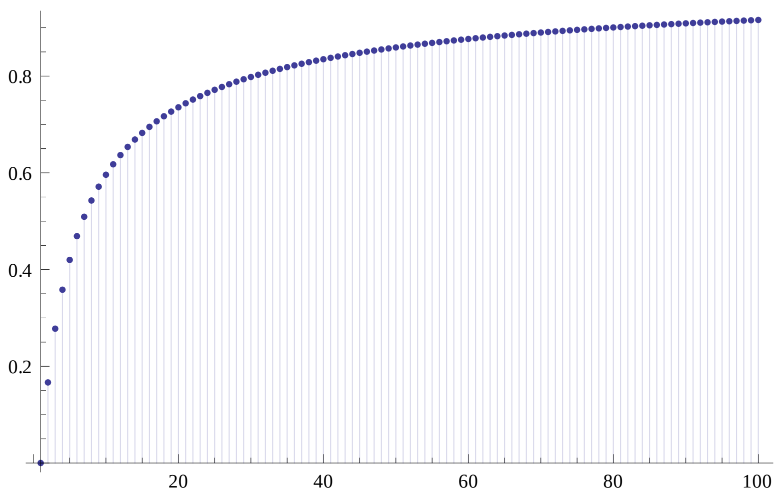
$$1 - \frac{2\,\text{Log}[n]}{n}$$

The red part converges to 0, so the entire expression converges to 1.

Here is a graph of the exact difference between c'_n and c_n

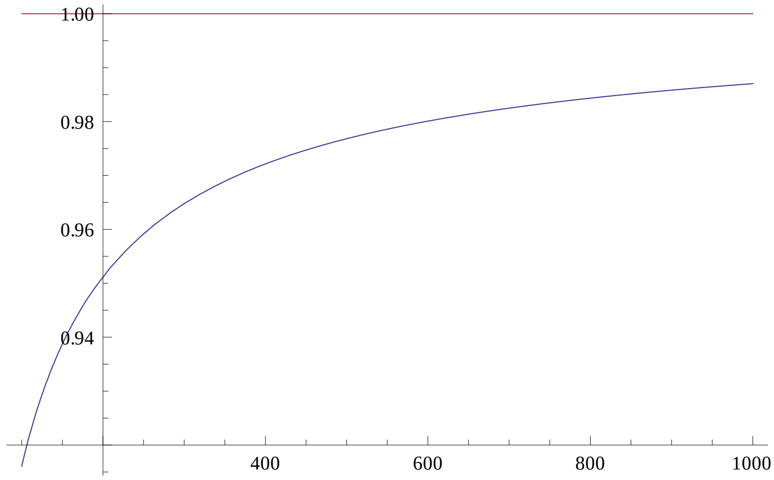
In[41]:= $\text{DiscretePlot}\left[\frac{n+3}{n+1} - \frac{2\sum_{i=1}^{n}\frac{1}{i}}{n}, \{n, 1, 100\}\right]$

Out[41]=



In[47]:= Plot$\left[\left\{\frac{n+3}{n+1} - \frac{2\sum_{i=1}^{n}\frac{1}{i}}{n}, 1\right\}, \{n, 100, 1000\}\right]$

Out[47]=

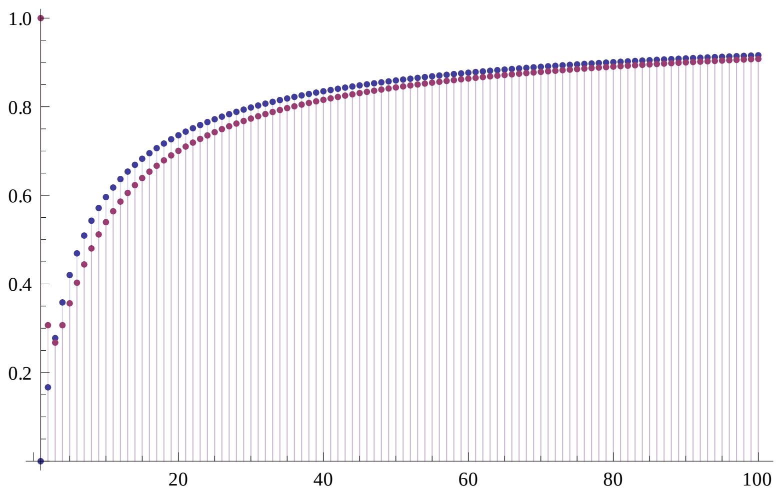


So, $c'_n \approx c_n + 1$ for large n.

For comparison, here is a graph of both exact and approximate difference

In[51]:= DiscretePlot$\left[\left\{\frac{n+3}{n+1} - \frac{2\sum_{i=1}^{n}\frac{1}{i}}{n}, 1 - \frac{2\,\text{Log}[n]}{n}\right\}, \{n, 1, 100\}\right]$

Out[51]=



So, it's pretty good for $n \geq 3$.

Here is a graph of the aproximate difference alone

In[52]:= Plot$\left[\left\{1-\frac{2\,\text{Log}[n]}{n}\right\}, \{n, 1, 100\}\right]$

Out[52]=