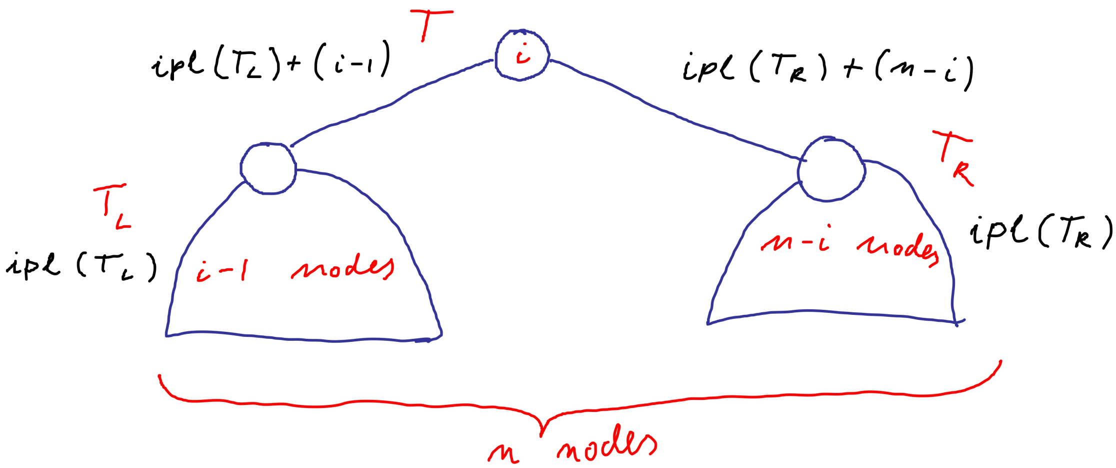
Average ipl (n) and epl (n) in any BS tree on n nodes

Let T be a BS tree created by n consecutive insertions of numbers
1 through n (not necessarily in that order) into initially empty tree.

There are n! orders (a.k.a. permutations) of numbers
1 through n. We assume that they are all equally likely.

Let i be the first number inserted.



$$\text{ipl}(T) = \text{ipl}(T_L) + (i-1) + \text{ipl}(T_R) + (n-i) =$$
$$= \text{ipl}(T_L) + \text{ipl}(T_R) + (n-1)$$

i can be equal to any of the numbers 1 through n, each with probability $\frac{1}{n}$.

Here is the reccurrence relation
(refer to your knowledge of Discrete Mathematics on recurrence relations):

$$\text{ipl}_{avg}[n\_] = \frac{1}{n}\sum_{i=1}^{n}\left(\text{ipl}_{avg}[i-1] + \text{ipl}_{avg}[n-i] + (n-1)\right) =$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left(\text{ipl}_{avg}[i-1] + \text{ipl}_{avg}[n-i]\right) + \frac{1}{n}\sum_{i=1}^{n}(n-1) =$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left(\text{ipl}_{avg}[i-1] + \text{ipl}_{avg}[n-i]\right) + \frac{1}{n}n(n-1) =$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left(\text{ipl}_{avg}[i-1] + \text{ipl}_{avg}[n-i]\right) + (n-1)$$

So,

$$ipl_{avg}[n\_] := n - 1 + \frac{1}{n} \sum_{i=1}^{n} \left(ipl_{avg}[i - 1] + ipl_{avg}[n - i]\right)$$

$$ipl_{avg}[0] := 0$$

$$ipl_{avg}[1] := 0$$

$$ipl_{avg}[2]$$

1

$$ipl_{avg}[3]$$

$$\frac{8}{3}$$

$$ipl_{avg}[4]$$

$$\frac{29}{6}$$

It may be verified by direct calculation that the function

$$ipl_{avg}[n] = 2\,(n + 1) \left(\sum_{i=1}^{n} \frac{1}{i}\right) - 4\,n$$

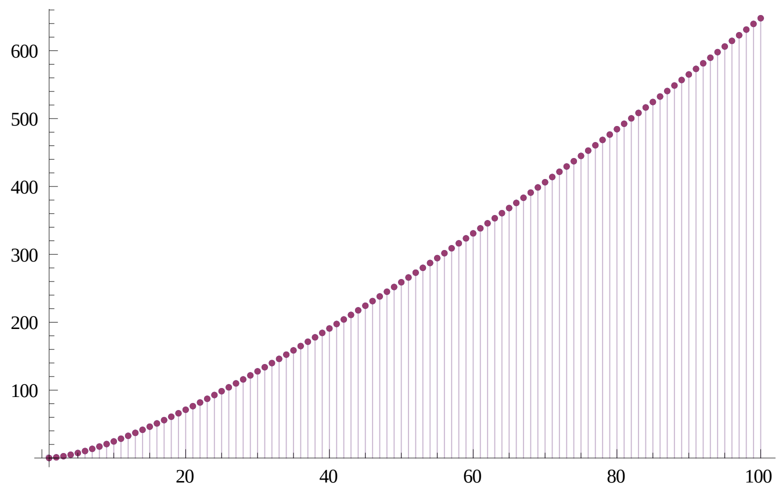
satisfies the above recurrence relation for every $n \geq 1$.

To that end it suffices to show that for every $n \geq 1$ **:**

$$2\,(n + 1) \left(\sum_{j=1}^{n} \frac{1}{j}\right) - 4\,n ==$$
$$n - 1 + \frac{1}{n} \sum_{i=0}^{n-1} \left(2\,(i + 1) \left(\sum_{j=1}^{i} \frac{1}{j}\right) - 4\,i + 2\,(n - i) \left(\sum_{j=1}^{n-1-i} \frac{1}{j}\right) - 4\,(n - 1 - i)\right)$$

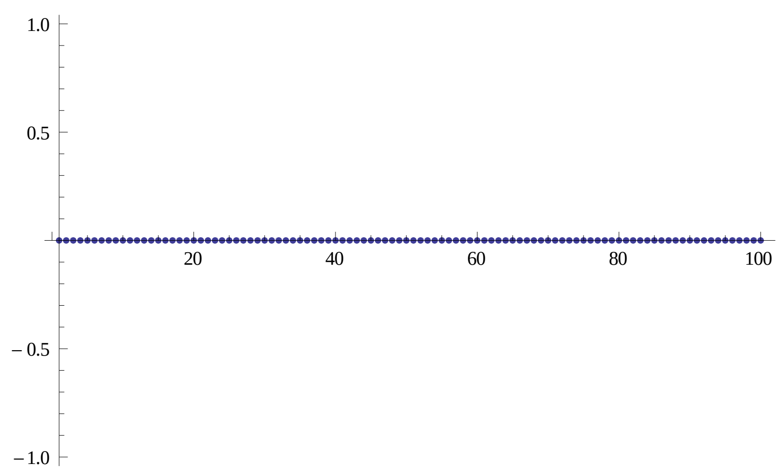
Here is an experimental verification by graphing

$$\text{DiscretePlot}\left[\left\{2\,(n+1)\left(\sum_{j=1}^{n}\frac{1}{j}\right)-4\,n,\ n-1+\frac{1}{n}\sum_{i=0}^{n-1}\left(2\,(i+1)\left(\sum_{j=1}^{i}\frac{1}{j}\right)-4\,i+2\,(n-i)\left(\sum_{j=1}^{n-1-i}\frac{1}{j}\right)-4\,(n-1-i)\right)\right\},\ \{n,1,100\}\right]$$



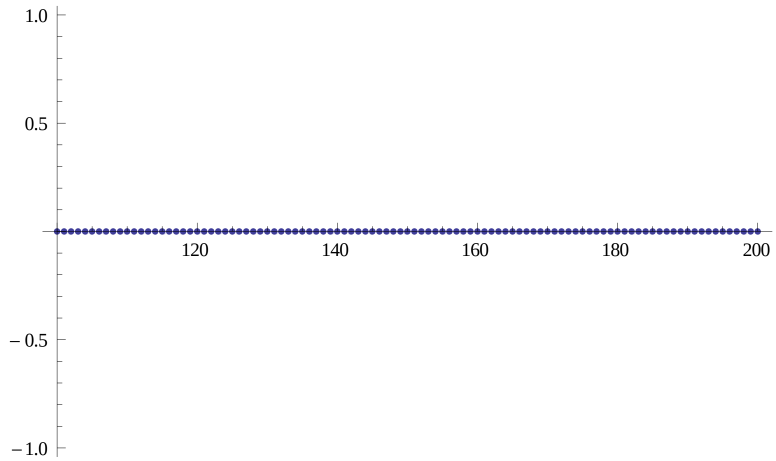
The differencebetween the sides of the equationis 0

$$\text{DiscretePlot}\left[\left\{2\,(n+1)\left(\sum_{j=1}^{n}\frac{1}{j}\right)-4\,n-\left(n-1+\frac{1}{n}\sum_{i=0}^{n-1}\left(2\,(i+1)\left(\sum_{j=1}^{i}\frac{1}{j}\right)-4\,i+2\,(n-i)\left(\sum_{j=1}^{n-1-i}\frac{1}{j}\right)-4\,(n-1-i)\right)\right)\right\},\ \{n,1,100\}\right]$$



```
DiscretePlot[{2 (n + 1) (Sum[1/j, {j, 1, n}]) - 4 n -
  (n - 1 + 1/n Sum[2 (i + 1) (Sum[1/j, {j, 1, i}]) - 4 i + 2 (n - i) (Sum[1/j, {j, 1, n - 1 - i}]) - 4 (n - 1 - i), {i, 0, n - 1}])}, {n,
  100, 200}]
```

$$\text{DiscretePlot}\left[\left\{2(n+1)\left(\sum_{j=1}^{n}\frac{1}{j}\right)-4n-\left(n-1+\frac{1}{n}\sum_{i=0}^{n-1}\left(2(i+1)\left(\sum_{j=1}^{i}\frac{1}{j}\right)-4i+2(n-i)\left(\sum_{j=1}^{n-1-i}\frac{1}{j}\right)-4(n-1-i)\right)\right)\right\},\{n,100,200\}\right]$$



Substitutingthe Euler' s approximation

Log[n] + .577

of

$$\sum_{i=1}^{n}\frac{1}{i}$$

in the above formula (see file Summationa nb) we obtain:

$ipl_{avg}[n] \approx 2 (n+1) (\text{Log}[n] + .577) - 4 n =$

Simplify[2 (n + 1) (Log[n] + .577) - 4 n]

1.154 – 2.846 n + 2 (1 + n) Log[n]

$$1.154\text{`} - 2.846\text{`}\, n + 2 (1+n) \frac{\text{Log}[n]}{\text{Log}[2]} \text{Log}[2]$$

2 N[Log[2]]

1.38629

1.154` – 2.846` n + 1.386 (1 + n) Log2[n]

$ipl_{avg}[n] \approx 1.386 (n+1) \text{Log2}[n] - 2.846 n + 1.154$

Note.
1.386 (n + 1) Log2[n] – 2.846 n + 1.154
is the number of comparisonsof keys that Quicksortwill perform
on average while sorting an n – element array with no duplicate keys.

$epl_{avg}[n] = ipl_{avg}[n] + 2\,n \approx 1.386\,(n+1)\,\text{Log2}[n] - 0.846\,n + 1.154$

$epl_{avg}[n] \approx 1.386\,(n+1)\,\text{Log2}[n] - 0.846\,n + 1.154$

Average number of comparisonsfor successfulsearch in an average BS tree :

$$c_n = \frac{ipl_{avg}[n]}{n} + 1 \approx \frac{1}{n}(1.386\,(n+1)\,\text{Log2}[n] - 2.846\,n + 1.154) + 1 \approx$$
$$1.386\,\text{Log2}[n] - 2.846 + 1 = 1.386\,\text{Log2}[n] - 1.846$$

$c_n \approx 1.386\,\text{Log2}[n] - 1.846$

Average number of comparisonsfor unsuccessfulsearch in an average BS tree :

$$c'_n = \frac{epl_{avg}[n]}{n+1} \approx (1.386\,(n+1)\,\text{Log2}[n] - 0.846\,n + 1.154) / (n+1) \approx$$
$$1.386\,\text{Log2}[n] - 0.846$$

$c'_n \approx 1.386\,\text{Log2}[n] - 0.846$

NB*