Approximations of minimal epl

If m is the number of external nodes then

$m \lg m$

is an "easy"
(proved in the textbook by a routine induction - see an "optional" file Ext_path _etc.pdf)
lower bound on epl and

$m\,(\lceil \text{Log2}[m] \rceil + 1) - 2^{\lceil \text{Log2}[m] \rceil}$

is the exact minimum epl., that has been derived in the file

http : // csc.csudh.edu / suchenek / CSC401 / Slides / 2 trees.pdf

As has been calculated in file LowerBoundAverageCaseSorting.nb,
this is the same as $m\,(\lg m + \epsilon\,(m))$.

So, the mimimum epl in any 2 - tree with m external nodes is :

$m\,(\text{Log2}[m] + \epsilon[m])$

where

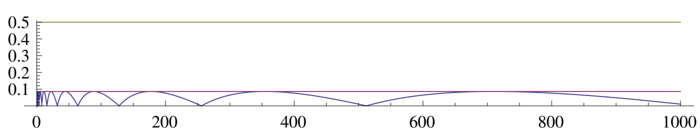
In[7]:= $\beta[x\_] := 1 + x - 2^x$

$\theta[x\_] := \lceil x \rceil - x$

$\epsilon[x\_] := \beta[\theta[\text{Log2}[x]]]$

Here is a plot of function ϵ[n]

```
Plot[{ϵ[n], .0861, .5}, {n, 1, 1000}, AspectRatio → .13]
```



Because ϵ oscillates between 0 and 0.08607133205593431, we conclude that

$m \lg m \le \text{epl} < m\,(\lg m + .0861)$

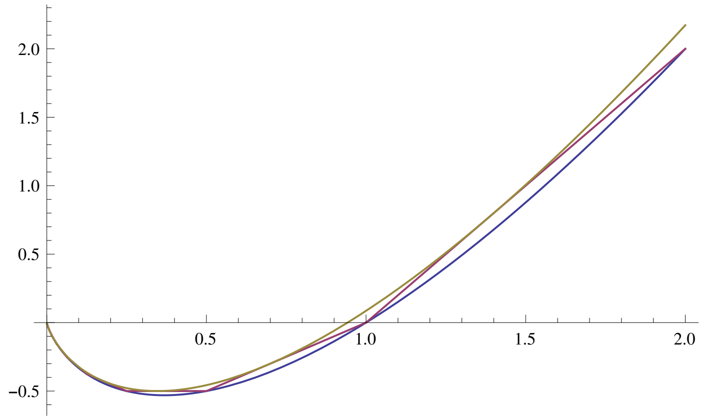
These are tight lower and upper approximations on $m\,(\lceil \text{Log2}[m] \rceil + 1) - 2^{\lceil \text{Log2}[m] \rceil}$.

See file
LowerBoundAverageCaseSorting.nb.
for detailed calculations.

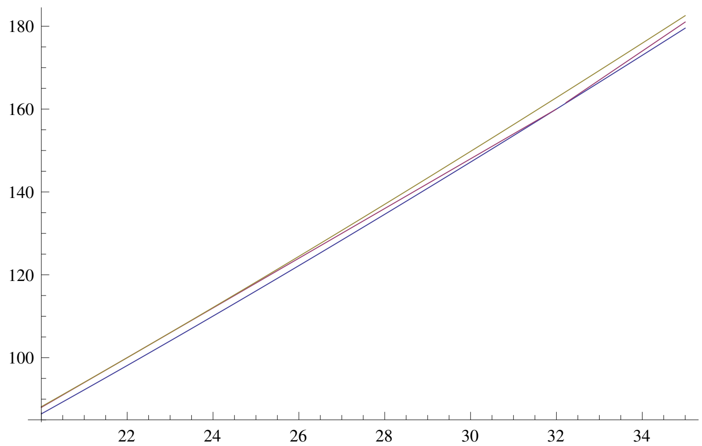
Here is a plot of the minimum epl against
the textbook ' s lower bound $m \lg m$ and our approximation :

```
In[13]:= Plot[Tooltip[{m Log2[m], m (Ceiling[Log2[m]] + 1) - 2^Ceiling[Log2[m]], m (Log2[m] + 0.0861)}],
  {m, 0.001, 2}, PlotTheme → "Classic"]
```

Out[13]=

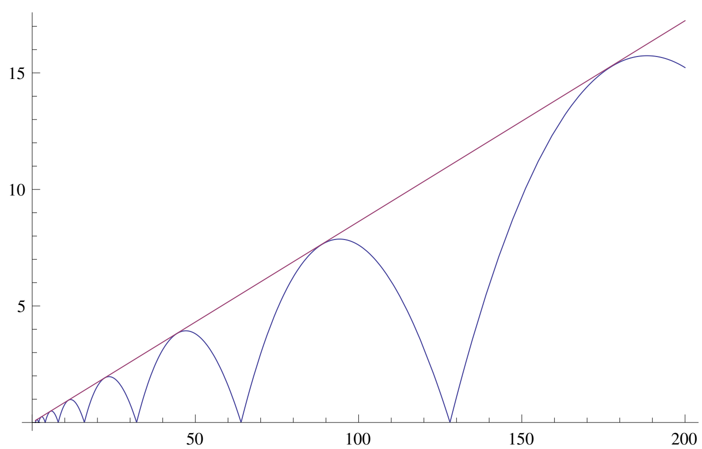



```
Plot[Tooltip[{m Log2[m] , m (⌈Log2[m]⌉ + 1) - 2^⌈Log2[m]⌉, m (Log2[m] + .0861)}], {m, 20, 35}]
```



Here is a plot of the difference between the first two against the line y = .0861 x.

```
Plot[{m (⌈Log2[m]⌉ + 1) - 2^⌈Log2[m]⌉ - m Log2[m], .0861 m}, {m, 1, 200}]
```



The rest of this file is optional for all students.

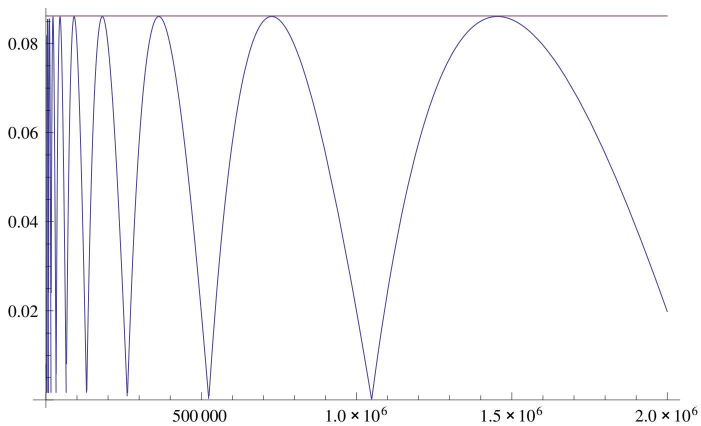
Here it is, the unsuccessful attempt :

$\text{Limit}\left[\left\{\left((\lceil \text{Log2}[m]\rceil + 1) - \frac{1}{m}2^{\lceil \text{Log2}[m]\rceil}\right) - \text{Log2}[m]\right\}, m \to \infty\right]$

{Interval[{-1, 1}]}

So, we plot of the difference divided by m against constant .0861 .

$\text{Plot}\left[\left\{\left((\lceil \text{Log2}[m]\rceil + 1) - \frac{1}{m}2^{\lceil \text{Log2}[m]\rceil}\right) - \text{Log2}[m] , .0861\right\}, \{m, 1, 2\,000\,000\}\right]$

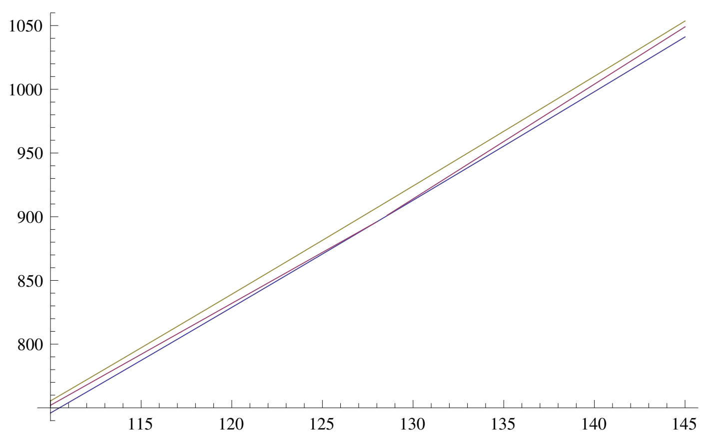


It ranges between 0 and .0861 .

The function that defines it (for x = ⌈Log2[m]⌉ - Log2[m]) is :

$\alpha(x) = 1 + x - 2^x$

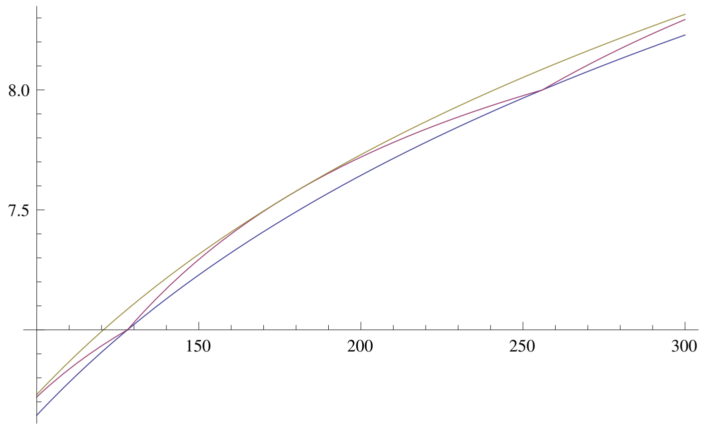
Plot[{1 + x - 2^x, 1}, {x, 0, 1}]

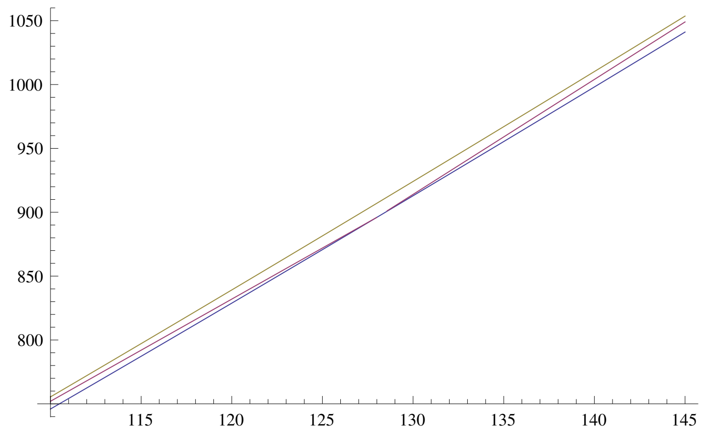


So the average length from the root to an external node in a balanced 2 - tree with m external nodes has these lower and upper approximations :

$$\text{Log2}[m] \le \lceil \text{Log2}[m]\rceil + 1 - \frac{1}{m}2^{\lceil \text{Log2}[m]\rceil} < \text{Log2}[m] + .0861$$

```
Plot[Tooltip[{ Log2[m] , ⌈Log2[m]⌉ + 1 - (1/m) 2^⌈Log2[m]⌉, Log2[m] + .0861}], {m, 100, 300}]
```

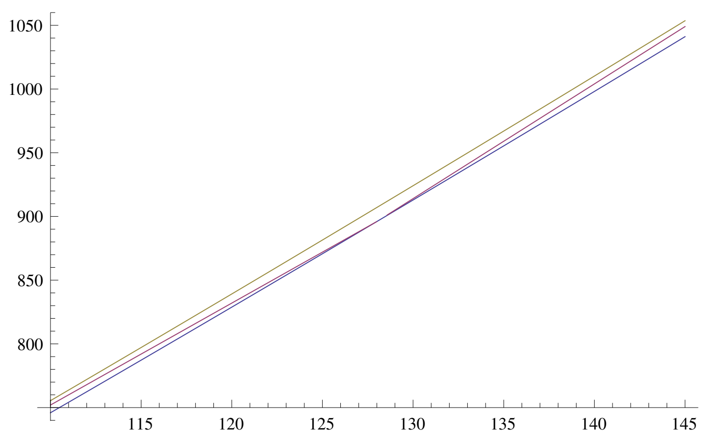




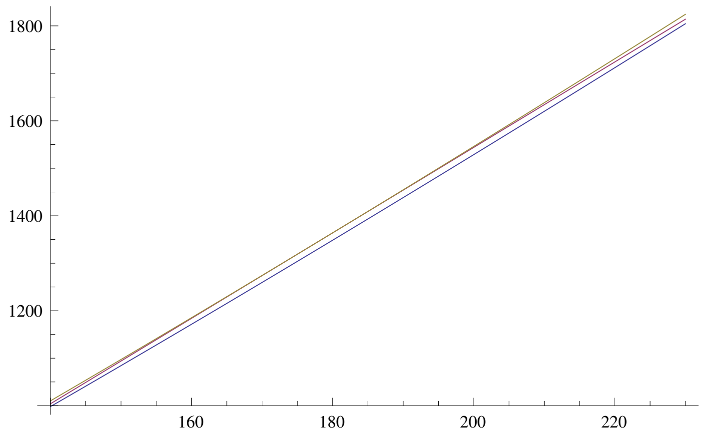
or

$$m \, \text{Log2}[m] \le m(\lceil \text{Log2}[m] \rceil + 1) - 2^{\lceil \text{Log2}[m] \rceil} < m(\text{Log2}[m] + .0861)$$

```
Plot[Tooltip[{ m Log2[m] , m (⌈Log2[m]⌉ + 1) - 2^⌈Log2[m]⌉, m (Log2[m] + .0861)}], {m, 110, 145}]
```



```
Plot[Tooltip[{ m Log2[m] , m (⌈Log2[m]⌉ + 1) - 2^⌈Log2[m]⌉, m (Log2[m] + .0861)}], {m, 140, 230}]
```



We will discover in file LowerBoundAverageCaseSorting.nb that the middle line above is the lower bound on the worst - case running time of any sorting program that sorts by a decision tree.