Expectations of the Pure Birth Tree

Mansi

February 27, 2018

1 Formulae and Expectations in the Pure Birth Process

In the pure birth model, each species has a constant probability, r (birth rate), of producing a new species at each point in time and extinction never occurs (i.e. extinction rate is 0).

Consider a case in which we start with N_0 species. The number of species is expected to grow exponentially over time. If N(t) is the function that relates the number of species at a given time to the birth rate, N(t) can be expressed as follows:

$$N(t) = N_0 e^r t$$

2 Controlling the Total Length of the Pure Birth Tree

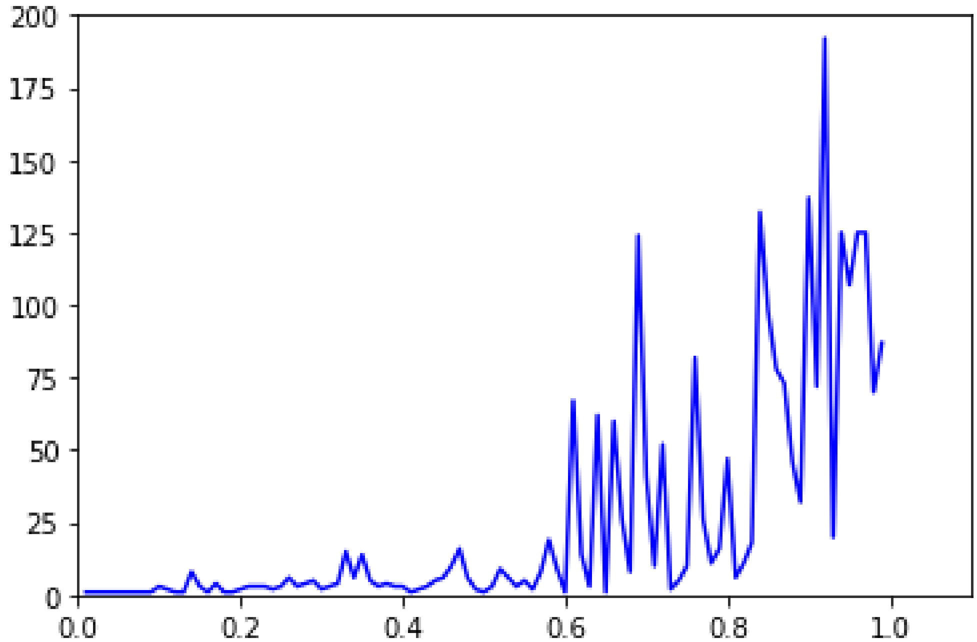
In the first set of iteration, multiple trees are generated with different birth rates but the same total length. The number of nodes in the trees are recorded. Finally, the number of nodes are plotted against the corresponding birth rates.

Since the total length (t) and the N_0 is the same for each tree, the number of nodes in the trees will depend only on the birth rate (r). If the number of nodes is N then for a given birth rate:

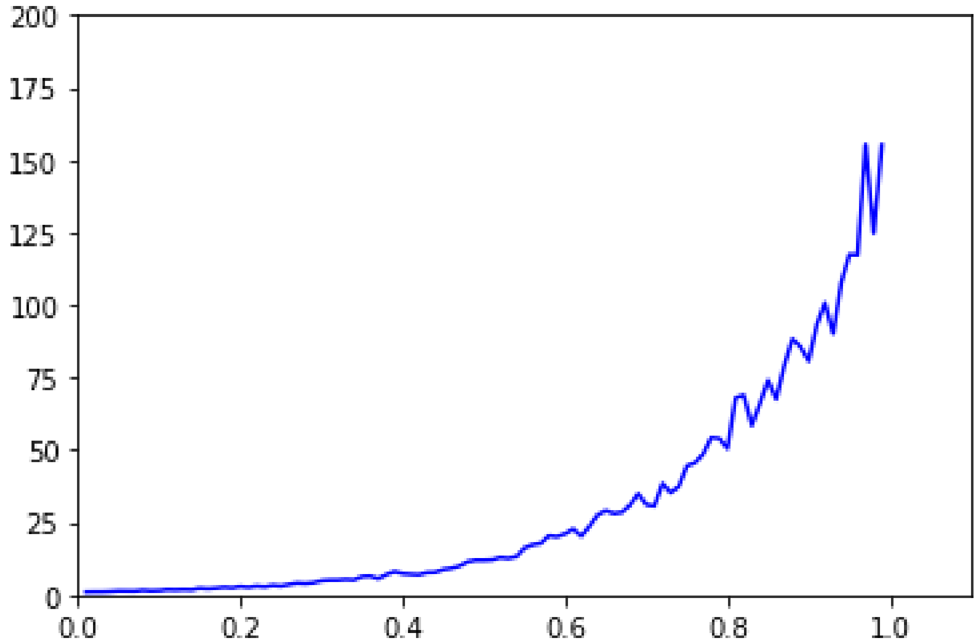
$$N = N_0 e^{rt}$$
$$log(N) = log(N_0) + rt$$

Therefore the number of nodes v/s birth rate plot should be exponentially increasing. The logarithmic plot would be a straight line with slope equal to the maximum total length (t)

Following is a plot for $r\epsilon[0, 1]$ with tress simulated at increments of 0.01. The maximum length for the trees is 5. The maximum number of nodes in a tree should be e^5 i.e. 149

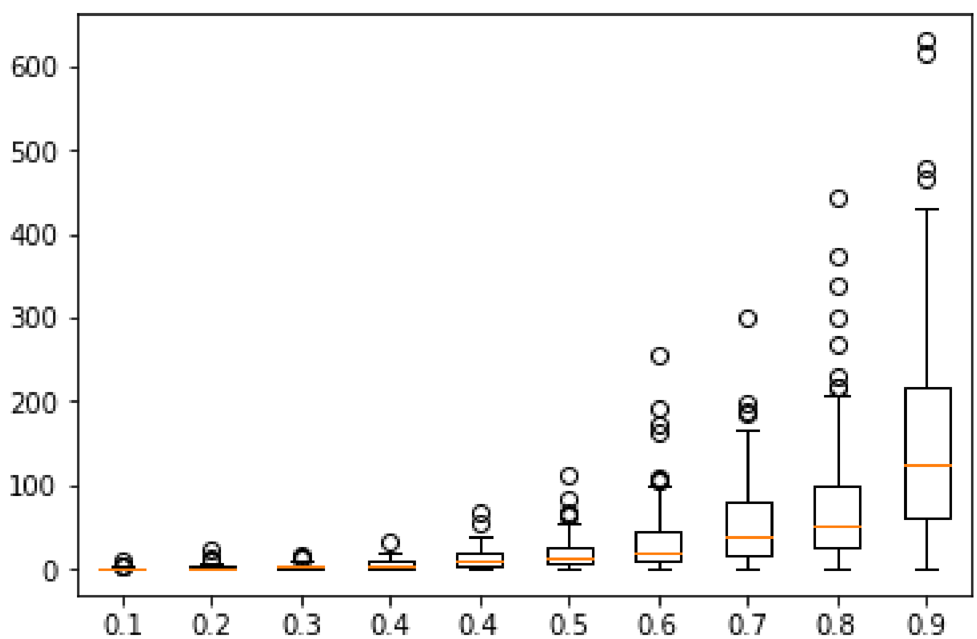


As seen in the graph, the increments are not exponential and rather random. This is because of the low number of data points per birth rate. This can be fixed by running each birth rate 100 times and taking the average at the end. This process results in the following graph:

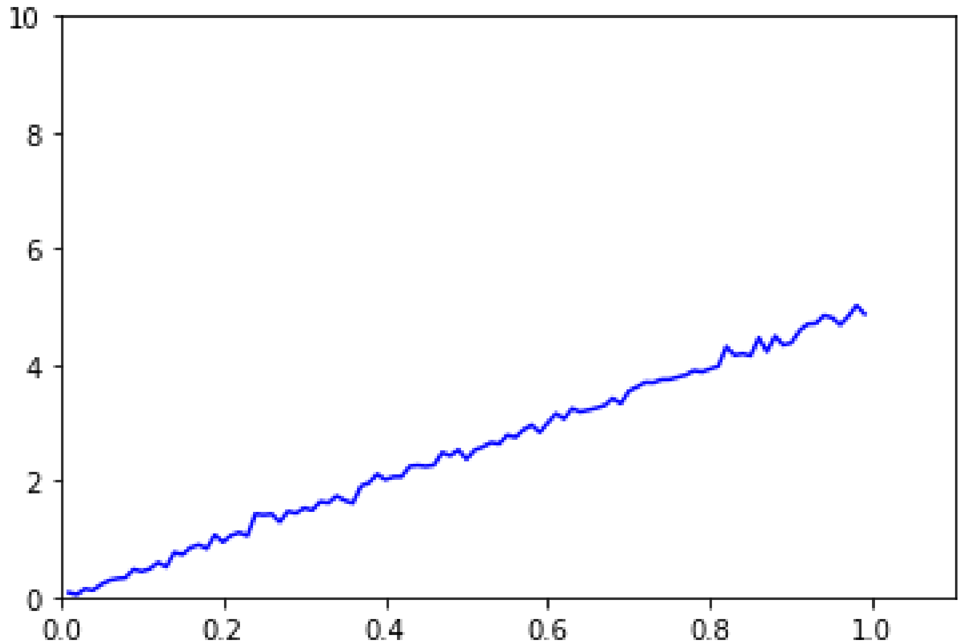


As seen above, this is a better exponential graph.The maximum number of nodes is around 150, as expected.

The variation graph is as follows:



The logarithmic plot is as follows:



As expected, the graph is a roughly linear graph with slope equal to 5.

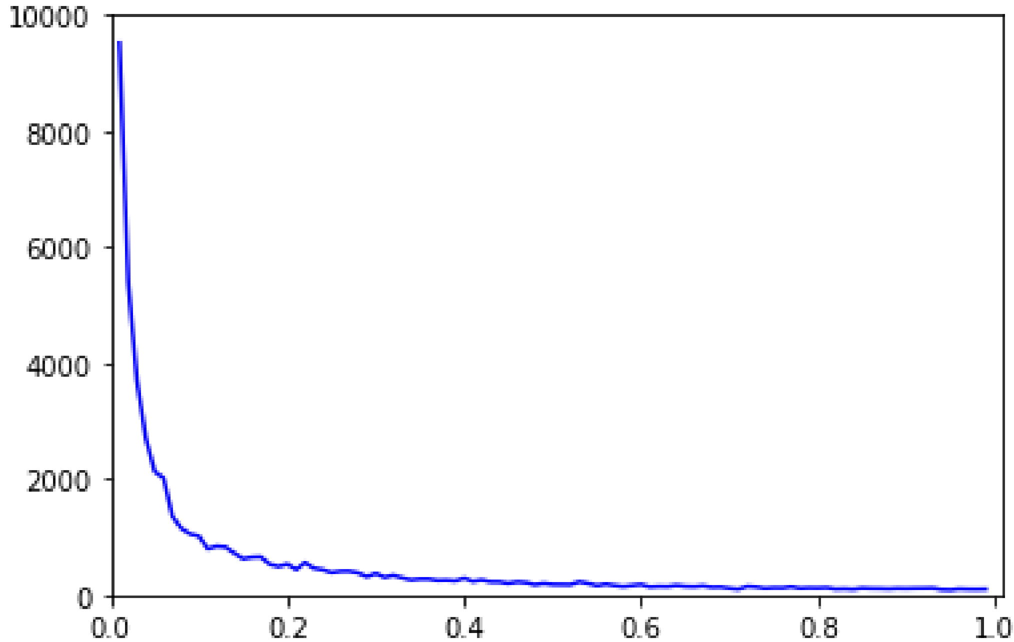
3 Controlling the Total Number of Nodes in the Pure Birth Tree

In the second set of iterations, the total number of nodes i.e. N(t) was kept constant. The total length (t) of the tree was plotted against the birth rate (r).

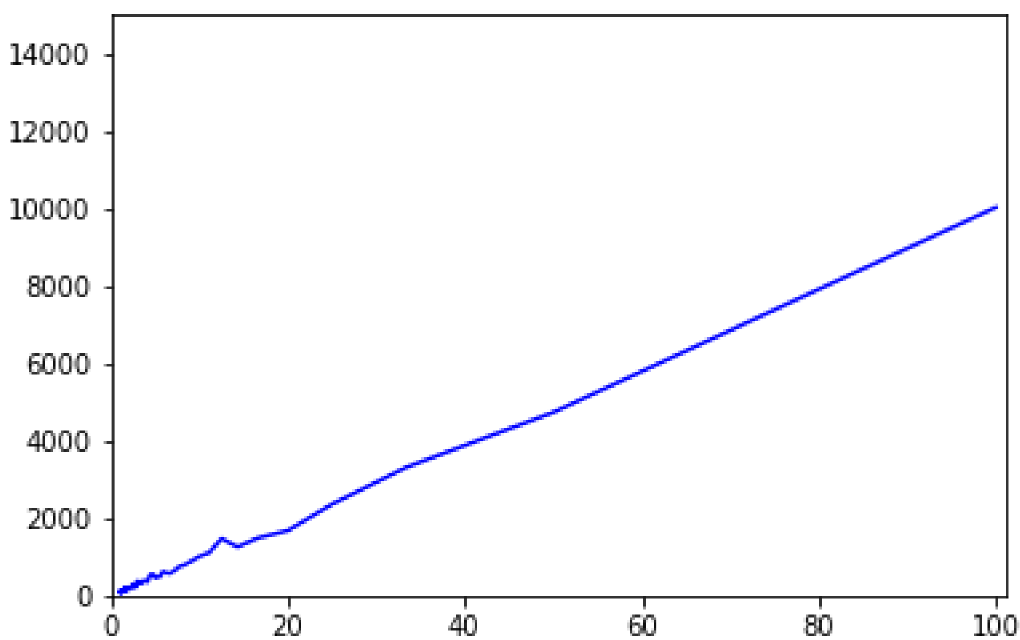
$$N = N_0 e^{rt}$$
$$log(N) = log(N_0) + rt$$
$$t = \frac{log(N) - log(N_0)}{r}$$

Therefore the total length v/s birth rate graph should be an inversely related graph. The graph of total length v/s $\frac{1}{birthrate}$ should be a straight line.

Following is a plot for $r\epsilon[0, 1]$ with tress simulated at increments of 0.01. The maximum number of nodes for the trees is 100.



As seen above, the graph shows an inverse relationship.
The graph of total length v/s $\frac{1}{birthrate}$ is as follows:



The graph is a straight line. The slope of the line is 100.