**DiscreeTest Exploration**

I ran iterative SFA with the initial states that lead to our 8 attractors of interest in BoolNet (Leuk1, Leuk2, Apop1-6). The results were discretized using GEDpro tools with every method that split the data into 2 levels: **bikmeans2, i2, kmeans2, max25, max50, max75, mean, median, q2, TDT, top25, top50, and top75**. I binarized across the rows of the SFA results because each row is the activity of one gene over time. Then, we used DiscreeTest to determine which method of discretization is the most appropriate for our SFA data. DiscreeTest first uses a sign test to compare the distribution of the discretized data to the original data. For 3 of the 8 attractors (Leukemia 2, Apoptosis 1,3), none of the discretization methods passed the sign test. For the other 5 attractors, only 1 or 2 of the discretization methods passed the sign test, and there was not a consensus for which method was the best. The minimum mean area between curves is 24.67, and the maximum is 30.32.

|  |  |  |
| --- | --- | --- |
| **Attractor** | **Optimal Discretization Method** | **Mean Area Between Curves** |
| Leukemia 1 | Equal Width Discretization with 2 levels | 29.38 |
| Apoptosis 2 | Equal Width Discretization with 2 levels | 25.88 |
| Apoptosis 4 | Equal Width Discretization with 2 levels | 25.27 |
| Apoptosis 5 | Top 75% | 24.67 |
| Apoptosis 6 | Top 75% | 30.32 |

The Top75 method sorts the expression values in descending order, assigns the first 75% of the values to 1, and the rest of the values to 0.