CS 4341

Trevor Rocks

Alex Ruggiero

Bayesian Network Analysis

For our project we utilized Python version 2.7 to create a program to build and analyze a Bayesian Network. The network was analyzed using both a rejection sampling method and a likelihood-weighting method, and we then compared these two methods.

The following data was collected from option A using query 1.

Rejection Sampling

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # of Samples | 200 | 400 | 600 | 800 | 1000 |
| T1 | 0.5800 | 0.8000 | 0.7567 | 0.7913 | 0.7420 |
| T2 | 0.4650 | 0.8000 | 0.7800 | 0.7475 | 0.7430 |
| T3 | 0.5050 | 0.7750 | 0.7833 | 0.7575 | 0.7890 |
| T4 | 0.4450 | 0.7750 | 0.7517 | 0.7863 | 0.7690 |
| T5 | 0.5050 | 0.7775 | 0.7450 | 0.7875 | 0.7820 |
| T6 | 0.4650 | 0.7675 | 0.7633 | 0.7888 | 0.7580 |
| T7 | 0.5750 | 0.7550 | 0.7767 | 0.7513 | 0.7820 |
| T8 | 0.5100 | 0.7350 | 0.7883 | 0.7700 | 0.7930 |
| T9 | 0.4550 | 0.7350 | 0.7800 | 0.7775 | 0.7720 |
| T10 | 0.4950 | 0.7325 | 0.7383 | 0.7713 | 0.7800 |
| Mean | 0.5000 | 0.7653 | 0.7663 | 0.7729 | 0.7710 |
| Variance | 0.0022 | 0.0006 | 0.0003 | 0.0003 | 0.0003 |

Likelihood-Weighting

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # of Samples | 200 | 400 | 600 | 800 | 1000 |
| T1 | 0.8365 | 0.7742 | 0.8104 | 0.7795 | 0.7523 |
| T2 | 0.7390 | 0.7561 | 0.7701 | 0.7835 | 0.7682 |
| T3 | 0.7082 | 0.7442 | 0.8140 | 0.7926 | 0.7732 |
| T4 | 0.7854 | 0.8032 | 0.7755 | 0.7717 | 0.7930 |
| T5 | 0.7493 | 0.7623 | 0.8055 | 0.7755 | 0.7813 |
| T6 | 0.8192 | 0.8058 | 0.8173 | 0.8032 | 0.7474 |
| T7 | 0.7320 | 0.7801 | 0.7562 | 0.7798 | 0.7570 |
| T8 | 0.7808 | 0.7006 | 0.7384 | 0.8051 | 0.7690 |
| T9 | 0.7734 | 0.7471 | 0.7647 | 0.7604 | 0.7806 |
| T10 | 0.7581 | 0.8218 | 0.7307 | 0.7568 | 0.7671 |
| Mean | 0.7682 | 0.7695 | 0.7783 | 0.7808 | 0.7689 |
| Variance | 0.0016 | 0.0013 | 0.0010 | 0.0003 | 0.0002 |

Comparisons of Means

Comparisons of Variance

The following data was collected from option A using query 2.

Rejection Sampling

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # of Samples | 200 | 400 | 600 | 800 | 1000 |
| T1 | 0.5800 | 0.5075 | 0.5200 | 0.5313 | 0.5190 |
| T2 | 0.4650 | 0.4825 | 0.4950 | 0.4775 | 0.5020 |
| T3 | 0.5050 | 0.4825 | 0.4783 | 0.5125 | 0.4980 |
| T4 | 0.4450 | 0.4750 | 0.5050 | 0.5038 | 0.4810 |
| T5 | 0.5050 | 0.4925 | 0.4970 | 0.5413 | 0.4730 |
| T6 | 0.4650 | 0.4950 | 0.5017 | 0.5163 | 0.4830 |
| T7 | 0.5750 | 0.5175 | 0.5150 | 0.4863 | 0.4700 |
| T8 | 0.5100 | 0.4850 | 0.5150 | 0.5263 | 0.5060 |
| T9 | 0.4550 | 0.4750 | 0.4750 | 0.5113 | 0.5240 |
| T10 | 0.4950 | 0.5400 | 0.5050 | 0.5050 | 0.4970 |
| Mean | 0.5000 | 0.4953 | 0.5007 | 0.5112 | 0.4953 |
| Variance | 0.0022 | 0.0004 | 0.0002 | 0.0004 | 0.0003 |

Likelihood-Weighting

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # of Samples | 200 | 400 | 600 | 800 | 1000 |
| T1 | 0.5200 | 0.4475 | 0.4817 | 0.5100 | 0.4880 |
| T2 | 0.5200 | 0.5150 | 0.5233 | 0.5088 | 0.4990 |
| T3 | 0.5150 | 0.4750 | 0.5033 | 0.4787 | 0.4840 |
| T4 | 0.5000 | 0.4600 | 0.5150 | 0.5038 | 0.5330 |
| T5 | 0.5500 | 0.4900 | 0.4970 | 0.4825 | 0.5060 |
| T6 | 0.4750 | 0.4925 | 0.4783 | 0.4950 | 0.5010 |
| T7 | 0.5300 | 0.4850 | 0.5217 | 0.5125 | 0.5130 |
| T8 | 0.4550 | 0.4825 | 0.4983 | 0.4787 | 0.4950 |
| T9 | 0.5450 | 0.5175 | 0.4917 | 0.5100 | 0.5200 |
| T10 | 0.4700 | 0.4925 | 0.5017 | 0.4837 | 0.4960 |
| Mean | 0.5080 | 0.4858 | 0.5012 | 0.4964 | 0.5035 |
| Variance | 0.0010 | 0.0005 | 0.0002 | 0.0002 | 0.0002 |

Comparisons of Means

Comparisons of Variance

Analysis

In Query 1 the rejection method converged onto a mean of .771. This is because as the rejection method works through the larger data set it is better able to reject data that might be outside of the desired range. While the likelihood-weighting method converged to a mean more quickly, if the sample size is increased high and higher the rejection method will become more and more accurate. In query 2 neither method converged, with both data sets showing greater variation than in the first query. This is likely due to the dataset itself.