

This talk presents a proposed method for using machine learning techniques to discover the dependency relationships among the bit positions in a sample of n -bit blocks of binary data. It is computationally infeasible to exhaustively test for all possible dependency combinations, so a heuristic technique is needed. We use Bayesian Networks to represent possible dependence structures, and then use a Genetic Algorithm to evolve increasingly better-scoring Bayesian Networks that best identify the statistical dependencies in the data sample. A technique from probabilistic graphical models is used to estimate the min-entropy of the unknown underlying joint probability distribution based on the best-scoring Bayesian Network. This method has shown success for block sizes of several dozen bits. Future optimizations may allow for analysis of even larger block-size entropy sources whose output spaces cannot be easily decomposed into smaller spaces for independent analysis.