If we observe multiple correlated variables such as words in a certain document, we need a way to compactly represent the joint distribution, as well as a way to use this distribution to infer one set of variables given another in a reasonable amount of computation time. We also need to learn the parameters of this distribution with a reasonable amount of data. How to resolve these questions are at the core of probabilistic modeling, inference, and learning.

The key to efficiently representing large joint distributions is to make some assumptions about conditional independence, also known as CI. To do this, we can use the Markov assumption, which tells us that the future is independent of the past given the present. Using this assumption plus the chain rule, we can obtain a first-order Markov chain, which can be characterized by an initial distribution over states plus a state transition matrix.

While first-order Markov assumptions are useful for defining distributions on 1d sequences, we can define distributions on 2d images, 3d videos, and in general arbitrary collections of variables such as genes on some biological pathway using graphical models. A graphical model, also known as GM, is a way to represent a joint distribution by making CI assumptions, particularly that the nodes in the graph represent random variables and the lack of edges represents CI assumptions. These models can also be thought of as independence diagrams in a way. There are several kinds of graphical model, depending on whether the graph is directed, undirected, or some combination of directed and undirected.