

can then be sampled in this order. In other words, we first sample $\mathbf{x}_1 \sim P(\mathbf{x}_1)$, then sample $P(\mathbf{x}_2 \mid Pa_{\mathcal{G}}(\mathbf{x}_2))$, and so on, until finally we sample $P(\mathbf{x}_n \mid Pa_{\mathcal{G}}(\mathbf{x}_n))$. So long as each conditional distribution $p(\mathbf{x}_i \mid Pa_{\mathcal{G}}(\mathbf{x}_i))$ is easy to sample from, then the whole model is easy to sample from. The topological sorting operation guarantees that we can read the conditional distributions in equation 16.1 and sample from them in order. Without the topological sorting, we might attempt to sample a variable before its parents are available.

For some graphs, more than one topological ordering is possible. Ancestral sampling may be used with any of these topological orderings.

Ancestral sampling is generally very fast (assuming sampling from each conditional is easy) and convenient.

One drawback to ancestral sampling is that it only applies to directed graphical models. Another drawback is that it does not support every conditional sampling operation. When we wish to sample from a subset of the variables in a directed graphical model, given some other variables, we often require that all the conditioning variables come earlier than the variables to be sampled in the ordered graph. In this case, we can sample from the local conditional probability distributions specified by the model distribution. Otherwise, the conditional distributions we need to sample from are the posterior distributions given the observed variables. These posterior distributions are usually not explicitly specified and parametrized in the model. Inferring these posterior distributions can be costly. In models where this is the case, ancestral sampling is no longer efficient.

Unfortunately, ancestral sampling is applicable only to directed models. We can sample from undirected models by converting them to directed models, but this often requires solving intractable inference problems (to determine the marginal distribution over the root nodes of the new directed graph) or requires introducing so many edges that the resulting directed model becomes intractable. Sampling from an undirected model without first converting it to a directed model seems to require resolving cyclical dependencies. Every variable interacts with every other variable, so there is no clear beginning point for the sampling process. Unfortunately, drawing samples from an undirected graphical model is an expensive, multi-pass process. The conceptually simplest approach is **Gibbs sampling**. Suppose we have a graphical model over an n -dimensional vector of random variables \mathbf{x} . We iteratively visit each variable \mathbf{x}_i and draw a sample conditioned on all of the other variables, from $p(\mathbf{x}_i \mid \mathbf{x}_{-i})$. Due to the separation properties of the graphical model, we can equivalently condition on only the neighbors of \mathbf{x}_i . Unfortunately, after we have made one pass through the graphical model and sampled all n variables, we still do not have a fair sample from $p(\mathbf{x})$. Instead, we must repeat the