Algorithm 12.4 Generating a Gibbs chain trajectory

Procedure Gibbs-Sample (
  $\boldsymbol{X}$ // Set of variables to be sampled
  Φ // Set of factors defining P_Φ
  $P^{(0)}(\boldsymbol{X})$, // Initial state distribution
  T // Number of time steps
)
  Sample $\boldsymbol{x}^{(0)}$ from $P^{(0)}(\boldsymbol{X})$
  for $t = 1, \ldots, T$
    $\boldsymbol{x}^{(t)} \leftarrow \boldsymbol{x}^{(t-1)}$
    for each $X_i \in \boldsymbol{X}$
      Sample $x_i^{(t)}$ from $P_\Phi(X_i \mid \boldsymbol{x}_{-i})$
        // Change X_i in $\boldsymbol{x}^{(t)}$
return $\boldsymbol{x}^{(0)}, \ldots, \boldsymbol{x}^{(T)}$

Example 12.4

Let us revisit example 12.3, recalling that we have the observations s^1, l^0. In this case, our algorithm will generate samples over the variables D, I, G. The set of reduced factors Φ is therefore: $P(I), P(D), P(G \mid I, D), P(s^1 \mid I), P(l^0 \mid G)$. Our algorithm begins by generating one sample, say by forward sampling. Assume that this sample is $d^{(0)} = d^1, i^{(0)} = i^0, g^{(0)} = g^2$. In the first iteration, it would now resample all of the unobserved variables, one at a time, in some predetermined order, say G, I, D. Thus, we first sample $g^{(1)}$ from the distribution $P_\Phi(G \mid d^1, i^0)$.

Note that because we are computing the distribution over a single variable given all the others, this computation can be performed very efficiently:

$$\begin{aligned} P_\Phi(G \mid d^1, i^0) &= \frac{P(i^0)P(d^1)P(G \mid i^0, d^1)P(l^0 \mid G)P(s^1 \mid i^0)}{\sum_g P(i^0)P(d^1)P(g \mid i^0, d^1)P(l^0 \mid g)P(s^1 \mid i^0)} \\ &= \frac{P(G \mid i^0, d^1)P(l^0 \mid G)}{\sum_g P(g \mid i^0, d^1)P(l^0 \mid g)}. \end{aligned}$$

Thus, we can compute the distribution simply by multiplying all factors that contain G, with all other variables instantiated, and renormalizing to obtain a distribution over G.

Having sampled $g^{(1)}$, we now continue to resampling $i^{(1)}$ from the distribution $P_\Phi(I \mid d^1, g^3)$, obtaining, for example, $i^{(1)} = i^1$; note that the distribution for I is conditioned on the newly sampled value $g^{(1)}$. Finally, we sample $d^{(1)}$ from $P_\Phi(D \mid g^3, i^1)$, obtaining d^1. The result of the first iteration of sampling is, then, the sample (i^1, d^1, g^3). The process now repeats. ■

Note that, unlike forward sampling, the sampling process for G takes into consideration the downstream evidence at its child L. Thus, its sampling distribution is arguably closer to the posterior. Of course, it is not the true posterior, since it still conditions on the originally sampled values for I, D, which were sampled from the prior distribution. However, we now resample I and D from a distribution that conditions on the new value of G, so one can imagine that their sampling distribution may also be closer to the posterior. Thus, perhaps the next sample of G,