Algorithm 10.1: Gibbs' sampling from a discrete undirected model

This algorithm generates samples from an undirected model with distribution

$$Pr(x_{1\ldots D}) = \frac{1}{Z}\prod_{c=1}^{C}\phi_c[\mathcal{S}_c],$$

where the c^{th} function $\phi_c[\mathcal{S}_c]$ operates on a subset $\mathcal{S}_c \subset \{x_1, x_2, \ldots, x_D\}$ of the D variables and returns a positive number. For this algorithm, we assume that each variable $\{x_d\}_{d=1}^{D}$ is discrete and takes values $x_d \in \{1, 2, \ldots, K\}$.

In Gibbs' sampling, we choose each variable in turn and update by sampling from its marginal posterior distribution. Since, the variables are discrete, the marginal distribution is a categorical distribution (a histogram), so we can sample from it by partitioning the range 0 to 1 according to the probabilities, drawing a uniform sample between 0 and 1, and seeing which partition it falls into.

Algorithm 10.1: Gibbs' sampling from undirected model

Input : Potential functions $\{\phi_c[\mathcal{S}_c]\}_{c=1}^{C}$
Output: Samples $\{\mathbf{x}_t\}_1^T$
begin
- // Initialize first sample in chain
- $\mathbf{x}_0 = \mathbf{x}^{(0)}$
- // For each time sample
- **for** t=1 **to** T **do**
  - $\mathbf{x}_t = \mathbf{x}_{t-1}$
  - // For each dimension
  - **for** d=1 **to** D **do**
    - // For each possible value of the dth variable
    - **for** k=1 **to** K **do**
      - // Set the variable to k
      - $x_{td} = k$
      - // Compute the unnormalized marginal probability
      - $\lambda_k = 1$
      - **for** c $s.t.$ $x_d \in \mathcal{S}_c$ **do**
        - $\lambda_k = \lambda_k \cdot \phi_c[\mathcal{S}_c]$
      - **end**
    - **end**
    - // Normalize the probabilities
    - $\boldsymbol{\lambda} = \boldsymbol{\lambda}/\sum_{k=1}^{K}\lambda_k$
    - // Draw from categorical distribution
    - $x_{td} = \mathsf{Sample}\left[\mathsf{Cat}_{x_{td}}[\boldsymbol{\lambda}]\right]$
  - **end**
- **end**

end

It is normal to discard the first few thousand entries so that the initial conditions are forgotten. Then entries are chosen that are spaced apart to avoid correlation between the samples.