Algorithm 19.2 Expectation-maximization algorithm for BN with table-CPDs

Procedure Compute-ESS (
  $\mathcal{G}$, // Bayesian network structure over $X_1, \ldots, X_n$
  $\boldsymbol{\theta}$, // Set of parameters for $\mathcal{G}$
  $\mathcal{D}$ // Partially observed data set
)
    // Initialize data structures
  for each $i = 1, \ldots, n$
    for each $x_i, \boldsymbol{u}_i \in \mathit{Val}(X_i, \mathrm{Pa}^{\mathcal{G}}_{X_i})$
      $\bar{M}[x_i, \boldsymbol{u}_i] \leftarrow 0$
    // Collect probabilities from all instances
  for each $m = 1 \ldots M$
    Run inference on $\langle \mathcal{G}, \boldsymbol{\theta} \rangle$ using evidence $\boldsymbol{o}[m]$
    for each $i = 1, \ldots, n$
      for each $x_i, \boldsymbol{u}_i \in \mathit{Val}(X_i, \mathrm{Pa}^{\mathcal{G}}_{X_i})$
        $\bar{M}[x_i, \boldsymbol{u}_i] \leftarrow \bar{M}[x_i, \boldsymbol{u}_i] + P(x_i, \boldsymbol{u}_i \mid \boldsymbol{o}[m])$
  return $\{\bar{M}[x_i, \boldsymbol{u}_i] : \forall i = 1, \ldots, n, \forall x_i, \boldsymbol{u}_i \in \mathit{Val}(X_i, \mathrm{Pa}^{\mathcal{G}}_{X_i})\}$

Procedure Expectation-Maximization (
  $\mathcal{G}$, // Bayesian network structure over $X_1, \ldots, X_n$
  $\boldsymbol{\theta}^0$, // Initial set of parameters for $\mathcal{G}$
  $\mathcal{D}$ // Partially observed data set
)
  for each $t = 0, 1 \ldots,$ until convergence
      // E-step
    $\{\bar{M}_t[x_i, \boldsymbol{u}_i]\} \leftarrow$ Compute-ESS($\mathcal{G}, \boldsymbol{\theta}^t, \mathcal{D}$)
      // M-step
    for each $i = 1, \ldots, n$
      for each $x_i, \boldsymbol{u}_i \in \mathit{Val}(X_i, \mathrm{Pa}^{\mathcal{G}}_{X_i})$
        $\theta^{t+1}_{x_i \mid \boldsymbol{u}_i} \leftarrow \frac{\bar{M}_t[x_i, \boldsymbol{u}_i]}{\bar{M}_t[\boldsymbol{u}_i]}$
  return $\boldsymbol{\theta}^t$

Maximization (M-step): Treat the expected sufficient statistics as observed, and perform maximum likelihood estimation, with respect to them, to derive a new set of parameters. In other words, set

$$\boldsymbol{\theta}^{t+1}_{x \mid \boldsymbol{u}} = \frac{\bar{M}_{\boldsymbol{\theta}^t}[x, \boldsymbol{u}]}{\bar{M}_{\boldsymbol{\theta}^t}[\boldsymbol{u}]}.$$

M-step

This phase is called the *M-step* (*maximization step*), because we are maximizing the likelihood relative to the expected sufficient statistics.

A formal version of the algorithm is shown fully in algorithm 19.2.