**Algorithm 12.2 Likelihood-weighted particle generation**

**Procedure** LW-Sample (
$\mathcal{B}$, // Bayesian network over $\mathcal{X}$
$\boldsymbol{Z} = \boldsymbol{z}$ // Event in the network
)
Let $X_1, \ldots, X_n$ be a topological ordering of $\mathcal{X}$
$w \leftarrow 1$
**for** $i = 1, \ldots, n$
$\boldsymbol{u}_i \leftarrow \boldsymbol{x}\langle \mathrm{Pa}_{X_i} \rangle$ // Assignment to $\mathrm{Pa}_{X_i}$ in $x_1, \ldots, x_{i-1}$
**if** $X_i \notin \boldsymbol{Z}$ **then**
Sample $x_i$ from $P(X_i \mid \boldsymbol{u}_i)$
**else**
$x_i \leftarrow \boldsymbol{z}\langle X_i \rangle$ // Assignment to $X_i$ in $\boldsymbol{z}$
$w \leftarrow w \cdot P(x_i \mid \boldsymbol{u}_i)$ // Multiply weight by probability of desired value
**return** $(x_1, \ldots, x_n), w$

$I = i^1$ and force $S = s^1$ should be worth 80 percent of a sample, whereas one where we have $I = i^0$ and force $S = s^1$ should only be worth 5 percent of a sample.

When we have multiple observations and we want our sampling process to set all of them to their observed values, we need to consider the probability that each of the observation nodes, had it been sampled using the standard forward sampling process, would have resulted in the observed values. The sampling events for each node in forward sampling are independent, and hence the weight for each sample should be the product of the weights induced by each evidence node separately.

**Example 12.3**

*Consider the same network, where our evidence set now consists of $l^0, s^1$. Assume that we sample $D = d^1$, $I = i^0$, set $S = s^1$, sample $G = g^2$, and set $L = l^0$. The probability that, given $I = i^0$, forward sampling would have generated $S = s^1$ is $0.05$. The probability that, given $G = g^2$, forward sampling would have generated $L = l^0$ is $0.4$. If we consider the standard forward sampling process, each of these events is the result of an independent coin toss. Hence, the probability that both would have occurred is simply the product of their probabilities. Thus, the weight required for this sample to compensate for the setting of the evidence is* $0.05 \cdot 0.4 = 0.02$. ■

likelihood weighting

Generalizing this intuition results in an algorithm called *likelihood weighting (LW)*, shown in algorithm 12.2. The name indicates that the weights of different samples are derived from the likelihood of the evidence accumulated throughout the sampling process.

weighted particle

This process generates a *weighted particle*. We can now estimate a conditional probability $P(\boldsymbol{y} \mid \boldsymbol{e})$ by using LW-Sample $M$ times to generate a set $\mathcal{D}$ of weighted particles $\langle \xi[1], w[1] \rangle, \ldots, \langle \xi[M], w[M] \rangle$. We then estimate:

$$\hat{P}_{\mathcal{D}}(\boldsymbol{y} \mid \boldsymbol{e}) = \frac{\sum_{m=1}^{M} w[m] \boldsymbol{1}\{\boldsymbol{y}[m] = \boldsymbol{y}\}}{\sum_{m=1}^{M} w[m]}. \tag{12.6}$$

estimator

This *estimator* is an obvious generalization of the one we used for unweighted particles in