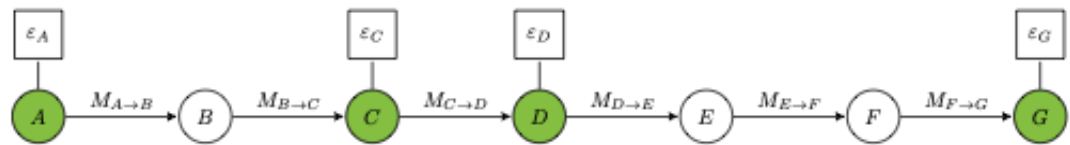


# Complete derivation of belief update rules for chains

## I. Formal definitions

### Evidence



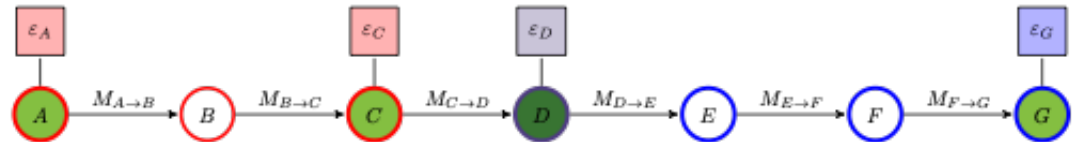
- **Direct evidence**  $\varepsilon_V$  for a node  $V$  is an observation  $V = v_*$  that determines the **local likelihood**  $\lambda_V^*(v) = [v = v_*]$ .
- **Indirect evidence**  $\varepsilon_V$  for a node  $V$  is a partial observation that determines the **local likelihood**  $\lambda_V^*(v) = \Pr[\varepsilon_V | V = v]$ .

### Example:

- For example, an indirect observation  $v \in \{0, 1\}$  leads to
 
$$\lambda_v^* = \begin{cases} 1, & \text{if } v = 0, \\ 1, & \text{if } v = 1, \\ 0, & \text{otherwise.} \end{cases}$$

## Evidence partitioning

- **Evidence** evidence is the summary evidence of all nodes in the chain.
- **Upstream evidence**  $\text{evidence}^+(V)$  is the evidence of all nodes preceeding  $V$  together with the evidence for  $V$ .
- **Downstream evidence**  $\text{evidence}^-(V)$  is the evidence of all nodes succeeding  $V$  together with the evidence for  $V$ .



In this figure, upstream and downstream evidence for node  $D$  are the following:

$$\text{evidence}^+(D) = \{\varepsilon_A, \varepsilon_C, \varepsilon_D\}$$

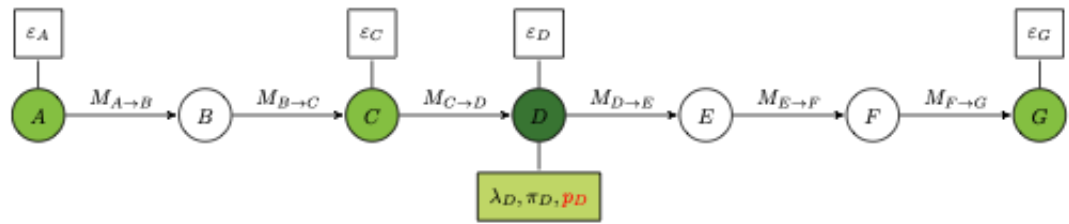
$$\text{evidence}^-(D) = \{\varepsilon_D, \varepsilon_G\} .$$

## Probabilities associated with nodes

- For a node  $V$  the **prior**  $\pi_V(v) = \Pr[V = v | \text{evidence}^+(V)]$ .
- For a node  $V$  the **likelihood**  $\lambda_V(v) = \Pr[\text{evidence}^-(V) | V = v]$ .
- For a node  $V$  the **marginal posterior probability**  
 $p_V(v) = \Pr[V = v | \text{evidence}^+(V), \text{evidence}^-(V)]$ .

## II. Derivation of iterative update rules

### Marginal posterior probabilities



Mechanical application of Bayes rule yields

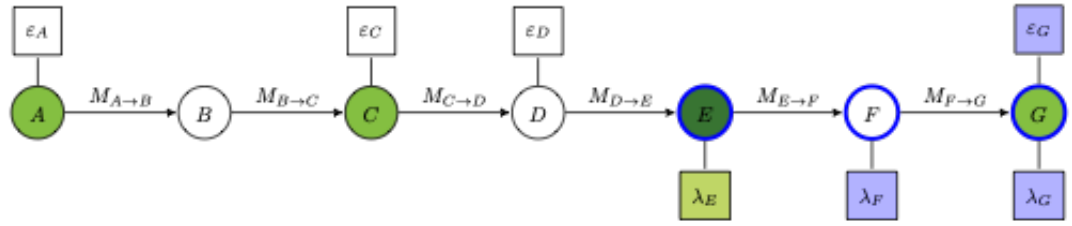
$$\begin{aligned}
 p_V(v) &= \Pr[V = v | \text{evidence}^+(V), \text{evidence}^-(V)] \\
 &= \frac{\Pr[\text{evidence}^-(V) | V = v, \text{evidence}^+(V)] \cdot \Pr[V = v | \text{evidence}^+(V)]}{\Pr[\text{evidence}^-(V) | \text{evidence}^+(V)]} \\
 &\propto \Pr[\text{evidence}^-(V) | V = v, \text{evidence}^+(V)] \cdot \Pr[V = v | \text{evidence}^+(V)] .
 \end{aligned}$$

As direct knowledge of the state  $V = v$  completely determines what happens with the next node, the knowledge  $V = v, \text{evidence}^+(V)$  is equivalent to the knowledge  $V = v$  and we can simplify:

$$\begin{aligned}
 p_V(v) &\propto \Pr[\text{evidence}^-(V) | V = v] \cdot \Pr[V = v | \text{evidence}^+(V)] \\
 &\propto \lambda_V(v) \cdot \pi_V(v) .
 \end{aligned}$$

As a result, if we know the likelihood  $\lambda_V(\cdot)$  and posterior  $\pi_V(\cdot)$  up to a constant then we can recover the marginal posterior  $p_V(\cdot)$  through normalisation. Up to a constant in this context means that we can omit all factors that do not depend on the value  $v$ .

## Likelihood update for a node without evidence



Let  $W$  be a successor node of  $V$ , then mechanical application of marginalisation rule yields

$$\begin{aligned}
 \lambda_V(v) &= \Pr[\text{evidence}^-(V)|V = v] \\
 &= \sum_{w \in W} \Pr[\text{evidence}^-(V) \wedge W = w|V = v] \\
 &= \sum_{w \in W} \Pr[\text{evidence}^-(V)|W = w, V = v] \cdot \Pr[W = w|V = v] .
 \end{aligned}$$

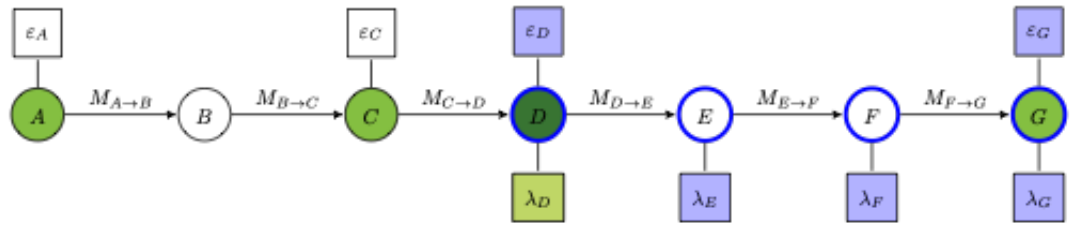
As the node  $V$  has no evidence, the downstream evidence must be in the successor nodes and thus  $\text{evidence}^-(V) = \text{evidence}^-(W)$ . The Markov property assures that knowledge of  $V = v$  is redundant when we know  $W = w$ . Consequently, we get

$$\begin{aligned}
 \lambda_V(v) &= \sum_{w \in W} \Pr[\text{evidence}^-(W)|W = w] \cdot \Pr[W = w|V = v] \\
 &= \sum_{w \in W} \lambda_W(w) M_{V \rightarrow W}[v, w] .
 \end{aligned}$$

Representing  $\lambda_W(\cdot)$  as a column vector allows us to compact the equation in matrix notation:

$$\lambda_V \propto M_{V \rightarrow W} \lambda_W .$$

## Likelihood update for a node with direct evidence



Let  $W$  be a successor node of  $V$  and let  $V = v_*$  be a direct evidence associated with the node  $V$ . Then evidence decomposition yields

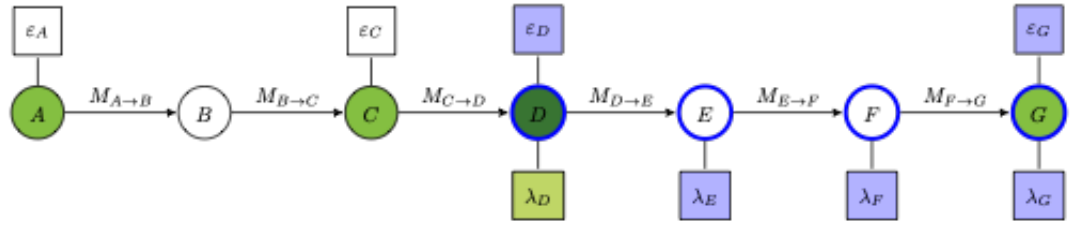
$$\begin{aligned}
 \lambda_V(v) &= \Pr[\text{evidence}^-(V) | V = v] \\
 &= \Pr[\text{evidence}^-(W) \wedge V = v_* | V = v] \\
 &= \Pr[\text{evidence}^-(W) | V = v_*, V = v] \cdot \Pr[V = v_* | V = v] \\
 &= \Pr[\text{evidence}^-(W) | V = v_*] \cdot [v_* = v] \quad .
 \end{aligned}$$

Note that  $\lambda_V(v)$  is nonzero only for a single value  $v_*$ . Thus by multiplying  $\lambda_V(v)$  with a constant value  $\lambda_V(v_*)^{-1}$ , we get an indicator:

$$\lambda_V(v) \propto [v = v_*] \quad .$$

Note that  $\lambda_V(v_*)^{-1}$  depends on  $v_*$  but remains constant if we consider different values of  $v \in V$ .

## Likelihood update for a node with indirect evidence



Let  $W$  be a successor node of  $V$  and let  $\varepsilon_V$  be an indirect evidence associated with the node  $V$ . Then evidence decomposition yields

$$\begin{aligned}\lambda_V(v) &= \Pr[\text{evidence}^-(V)|V = v] \\ &= \Pr[\text{evidence}^-(W) \wedge \varepsilon_v|V = v] \\ &= \Pr[\text{evidence}^-(W)|\varepsilon_V, V = v] \cdot \Pr[\varepsilon_V|V = v] \ .\end{aligned}$$

Again the direct knowledge  $V = v$  subsumes the partial knowledge  $\varepsilon_V$  and we get

$$\lambda_V(v) = \Pr[\text{evidence}^-(W)|V = v] \cdot \lambda_V^*(v)$$

where  $\lambda_V^*(v) = \Pr[\varepsilon_V|V = v]$  is the local likelihood. Now there can be several states for which the likelihood is nonzero and thus we must separately compute the left term. Again the mechanical application of the marginalisation rule yields

$$\begin{aligned}\lambda_1(v) &= \Pr[\text{evidence}^-(W)|V = v] \\ &= \sum_{w \in W} \Pr[\text{evidence}^-(W) \wedge W = w|V = v] \\ &= \sum_{w \in W} \Pr[\text{evidence}^-(W)|W = w, V = v] \cdot \Pr[W = w|V = v] \\ &= \sum_{w \in W} \Pr[\text{evidence}^-(W)|W = w] \cdot M_{V \rightarrow W}[v, w] \\ &= \sum_{w \in W} \lambda_W(w) \cdot M_{V \rightarrow W}[v, w] \ .\end{aligned}$$

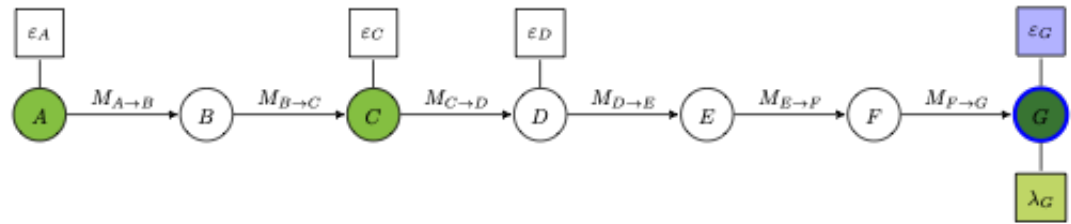
Thus matrix algebra allows us to compact the update rule:

$$\lambda_1 = M_{V \rightarrow W} \lambda_W$$

$$\lambda_V = \lambda_1 \otimes \lambda_V^*$$

where  $\otimes$  represents pointwise multiplication of vector entries.

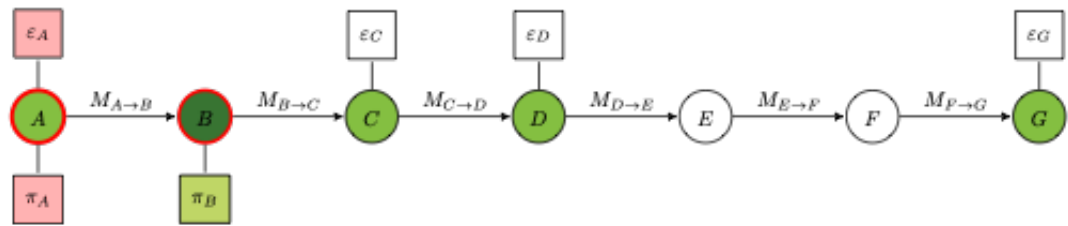
## Likelihood update for a node without successors



The rules for updating the likelihood are applicable for nodes that do have successors. Hence, we must address nodes without successors explicitly. Without loss of generality, we can assume that for such a node  $V$  there is evidence  $\varepsilon_V$ . If the evidence is missing we can treat this as a partial observation  $v \in V$  that creates a local likelihood  $\lambda_V^* = 1$ . If the evidence is direct then it creates a local likelihood  $\lambda_V^*(v) = [v = v_*]$ . As a result, we get

$$\begin{aligned}\lambda_V(v) &= \Pr[\text{evidence}^-(V) | V = v] \\ &= \Pr[\varepsilon_V | V = v] \\ &= \lambda_V^*(v) \ .\end{aligned}$$

## Prior update for a node without evidence



Let  $U$  be the predecessor node of  $V$  then mechanical application of marginalisation rule yields

$$\begin{aligned}
 \pi_V(v) &= \Pr[V = v | \text{evidence}^+(V)] \\
 &= \sum_{u \in U} \Pr[V = v \wedge U = u | \text{evidence}^+(V)] \\
 &= \sum_{u \in U} \Pr[V = v | U = u, \text{evidence}^+(V)] \cdot \Pr[U = u | \text{evidence}^+(V)] .
 \end{aligned}$$

As the node  $V$  has no evidence, the upstream evidence must be in the predecessor nodes and thus  $\text{evidence}^+(V) = \text{evidence}^+(U)$ . The Markov property assures that knowledge of  $\text{evidence}^+(U)$  is redundant when we know  $U = u$ . Consequently, we get

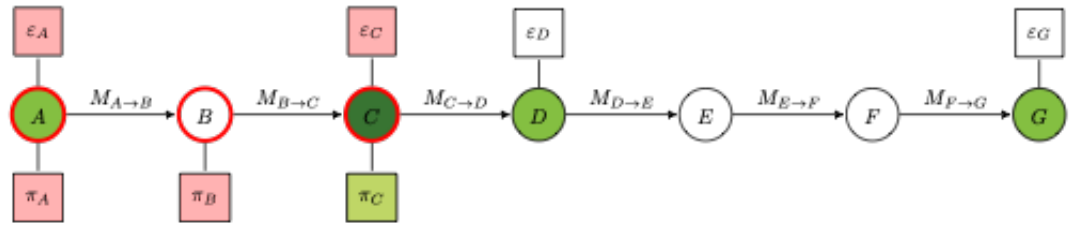
$$\begin{aligned}
 \pi_V(v) &= \sum_{u \in U} \Pr[V = v | U = u] \cdot \Pr[U = u | \text{evidence}^+(U)] \\
 &= \sum_{u \in U} M_{U \rightarrow V}[u, v] \cdot \pi_U(u) .
 \end{aligned}$$

Representing  $\pi_U(\cdot)$  as a row vector allows us to compact the equation in matrix notation:

$$\pi_V \propto \pi_U M_{U \rightarrow V} .$$



## Prior update for a node with direct evidence



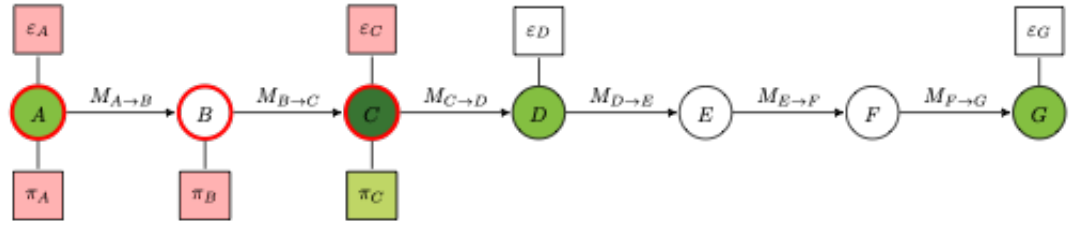
Let  $U$  be a predecessor node of  $V$  and let  $V = v_*$  be a direct evidence associated with the node  $V$ . Then the evidence decomposition yields

$$\begin{aligned}\pi_V(v) &= \Pr[V = v | \text{evidence}^+(V)] \\ &= \Pr[V = v | V = v_*, \text{evidence}^+(U)] .\end{aligned}$$

Again the evidence  $V = v_*$  is the most direct information about  $V$ , the remaining evidence  $\text{evidence}^+(U)$  is irrelevant unless  $\text{evidence}^+(U)$  directly contradicts  $V = v_*$ . In this case, nothing can be done and prior is not defined at all. Thus we can simplify and get an indicator prior:

$$\begin{aligned}\pi_V(v) &= \Pr[V = v | V = v_*] \\ &= [v = v_*] .\end{aligned}$$

## Prior update for a node with indirect evidence



Let  $U$  be a predecessor node of  $V$  and let  $\varepsilon_V$  be an indirect evidence associated with the node  $V$ . Then the evidence decomposition yields

$$\begin{aligned}\pi_V(v) &= \Pr[V = v | \text{evidence}^+(V)] \\ &= \Pr[V = v | \varepsilon_V, \text{evidence}^+(U)] .\end{aligned}$$

Mechanical application of Bayes rule yields

$$\begin{aligned}\pi_V(v) &= \Pr[V = v | \varepsilon_V, \text{evidence}^+(U)] \\ &= \frac{\Pr[\varepsilon_V | V = v, \text{evidence}^+(U)] \cdot \Pr[V = v | \text{evidence}^+(U)]}{\Pr[\varepsilon_V | \text{evidence}^+(U)]} \\ &\propto \Pr[\varepsilon_V | V = v, \text{evidence}^+(U)] \cdot \Pr[V = v | \text{evidence}^+(U)] .\end{aligned}$$

As  $\text{evidence}^+(U)$  is redundant if we know  $V = v$ , we get

$$\begin{aligned}\pi_V(v) &\propto \Pr[\varepsilon_V | V = v] \cdot \Pr[V = v | \text{evidence}^+(U)] \\ &\propto \lambda_V^*(v) \cdot \Pr[V = v | \text{evidence}^+(U)]\end{aligned}$$

where  $\lambda_V^*(v) = \Pr[\varepsilon_V | V = v]$  is the local likelihood. Let  $\pi_1(v)$  denote the second factor. Then mechanical application of the marginalisation rule yields

$$\begin{aligned}\pi_1(v) &= \Pr[V = v | \text{evidence}^+(U)] \\ &= \sum_{u \in U} \Pr[V = v \wedge U = u | \text{evidence}^+(U)] \\ &= \sum_{u \in U} \Pr[V = v | U = u, \text{evidence}^+(U)] \cdot \Pr[U = u | \text{evidence}^+(U)] \\ &= \sum_{u \in U} \Pr[V = v | U = u] \cdot \Pr[U = u | \text{evidence}^+(U)] \\ &= \sum_{u \in U} M_{U \rightarrow V}[u, v] \cdot \pi_U(u) .\end{aligned}$$

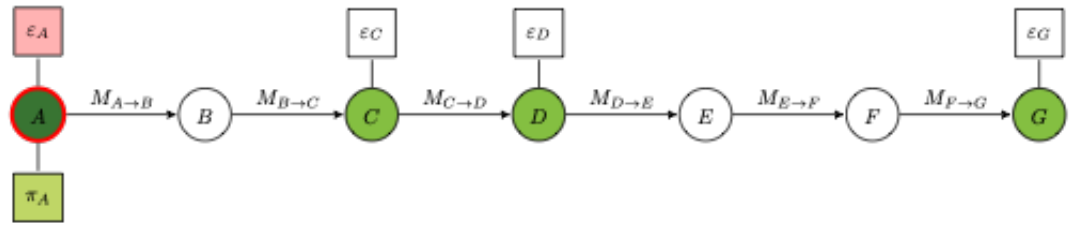
Thus matrix algebra allows us to compact the update rule:

$$\pi_1 = \pi_U M_{U \rightarrow V}$$

$$\pi_V = \pi_1 \otimes \lambda_V^*$$

where  $\otimes$  represents pointwise multiplication of vector entries.

## Prior update for a node without a predecessor



The rules for updating the prior are applicable for nodes that do have predecessors. Hence, we must address nodes without predecessors explicitly. Without loss of generality, we can assume that for such a node  $V$  there is evidence  $\varepsilon_V$ . If the evidence is missing we can treat this as a partial observation  $v \in V$  that creates a local likelihood  $\lambda_V^* = 1$ . If the evidence is direct then it creates a local likelihood  $\lambda_V^*(v) = [v = v_*]$ . As a result, we get an expression that can be further manipulated with the Bayes rule:

$$\begin{aligned}
 \pi_V(v) &= \Pr[V = v | \text{evidence}^+(V)] \\
 &= \Pr[V = v | \varepsilon_V] \\
 &= \frac{\Pr[\varepsilon_V | V = v] \cdot \Pr[V = v]}{\Pr[\varepsilon_V]} \\
 &\propto \lambda_V^*(v) \cdot M_V[v]
 \end{aligned}$$

where  $M_V$  is the vector of initial probabilities. The corresponding matrix formulation is

$$\pi_V \propto \lambda_V^* \otimes M_V \quad .$$

For a direct observation the expression simplifies to

$$\pi_V \propto [v = v_*]$$

as only one entry is nonzero.