TUM / WINTER SEMESTER 2025/26

Fundamentals of Artificial Intelligence: Probabilistic Reasoning

Comprehensive Summary of Bayesian Networks & Hidden Markov Models

---

Based on Lecture Notes 9 & 10 and Summary Note 5

Detailed Summary & Mathematical Reference

Probability Fundamentals & Bayes' Rule

Sample Space (Ω):
The set of all possible outcomes (e.g., $\Omega = \{\text{heads, tails}\}$). Elements denoted by $\omega \in \Omega$.

Event Space ($\mathcal{F}$):
The powerset of Ω containing all possible combinations of outcomes.

Random Variable:
A mapping $X : \Omega \to D$ from the sample space to a domain D.

Math Container

Conditional Probability:

$$P(Y = y \mid X = x) = \frac{P(X = x \mid Y = y)P(Y = y)}{P(X = x)}$$

Normalization constant α

Product Rule:

$$P(X = x, Y = y) = P(X = x \mid Y = y)P(Y = y)$$

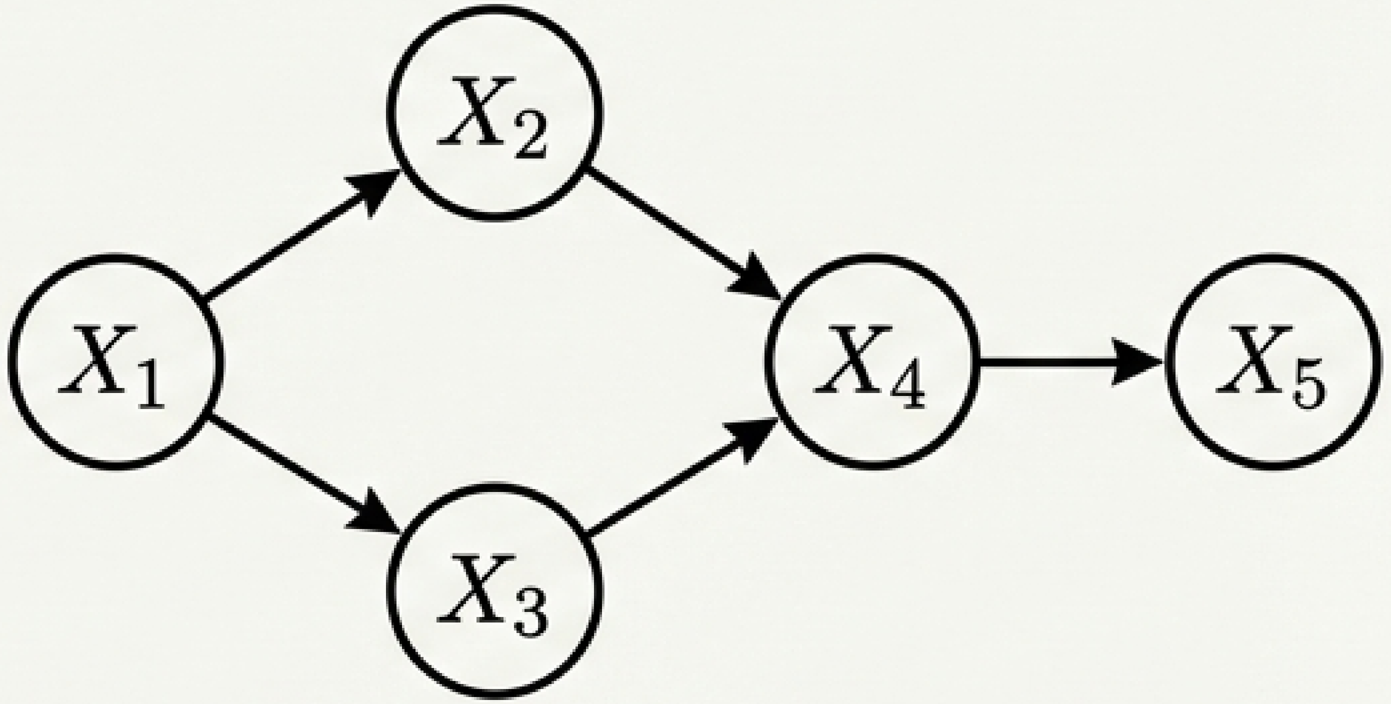
Bayes' Rule:

$$P(Y = y \mid X = x) = \frac{P(X = x \mid Y = y)P(Y = y)}{\sum_{y \in D_y} P(X = x \mid Y = y)P(Y = y)}$$

Source: Summary Note 5.1; Lecture 9.

Bayesian Networks: Structure & Semantics

A **Bayesian Network** is a directed acyclic graph (DAG) where nodes represent random variables and arrows represent parental influence.



The Chain Rule:

The full joint distribution is the product of local conditional distributions:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(X_i))$$

Complexity Comparison:

Structure: Bayesian Network (max k parents)	$O(n \cdot 2^k)$ (Linear)
Structure: Full Joint Distribution	$O(2^n)$ (Exponential)

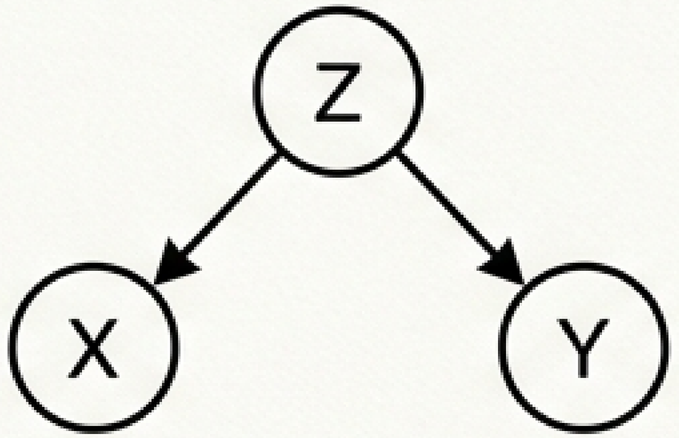
Source: Lecture 9, Slide 36; Summary Note 5.2.1.

Conditional Independence in Bayesian Networks

Diagram 1 (Chain)

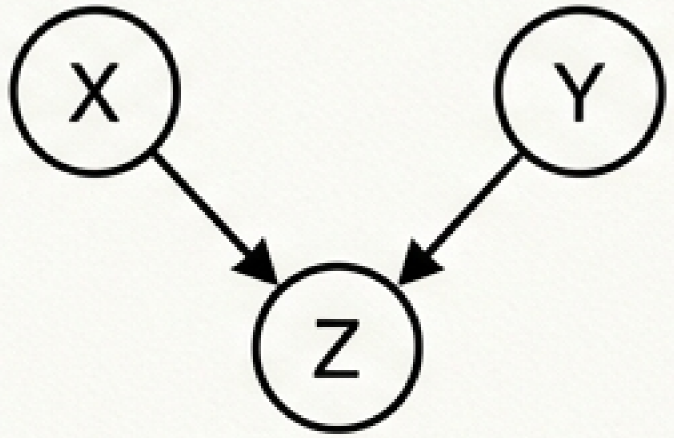
Blocked if Z is evidence.

Diagram 2 (Fork)



Blocked if Z is evidence.

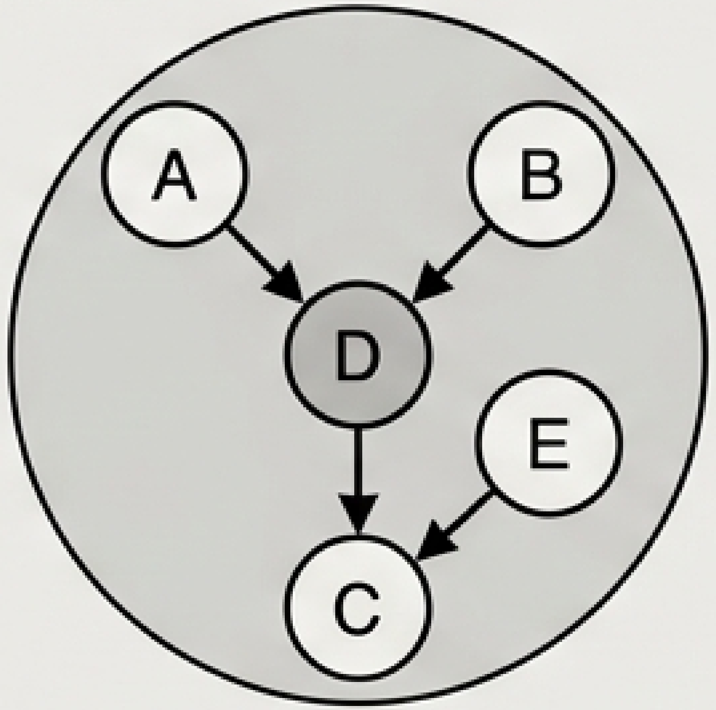
Diagram 3 (Collider/V-Structure)



Open if Z (or descendant) is evidence. **Blocked** if Z is **NOT** evidence.

Markov Blanket

A node is independent of the rest of the network given its parents, children, and children's parents.



Source: Summary Note 5.2.2 - 5.2.3; Lecture 9.

Exact Inference: **Enumeration**

Objective: Compute $P(Query \mid Evidence)$ by summing out hidden variables.

Query: $P(B \mid j, m)$

$$
\begin{aligned}
&= \alpha P(B, j, m) \\
&= \alpha \sum_{e} \sum_{a} P(B, e, a, j, m) \\
&= \alpha P(B) \sum_{e} P(e) \sum_{a} P(a \mid B, e) P(j \mid a) P(m \mid a)
\end{aligned}
$$

Complexity

Time Complexity: $O(n \cdot d^n)$ in the worst case.

Problem: Inefficient because it repeats calculations for intermediate terms.

Source: Lecture 9; Summary Note 5.2.6.1.

Exact Inference: **Variable Elimination**

- **Idea:** Interleave summation and product operations to avoid re-computation.
- **Mechanism:** Evaluate right-to-left and store intermediate results as **Factors (f_i)**.
- **Optimization:** Remove irrelevant leaf nodes (nodes not in query or evidence) before processing.

$$f_6(A,B) = \sum_e f_2(E) \times f_3(A,B,E) \times f_4(A) \times f_5(A)$$

$$\Big\downarrow \text{ Move independent factors out}$$

$$= f_4(A) \times f_5(A) \times \boxed{\left[\sum_e (f_2(E) \times f_3(A,B,E))\right]}$$

Computed once & stored

Source: Summary Note 5.2.6.1; Lecture 9.

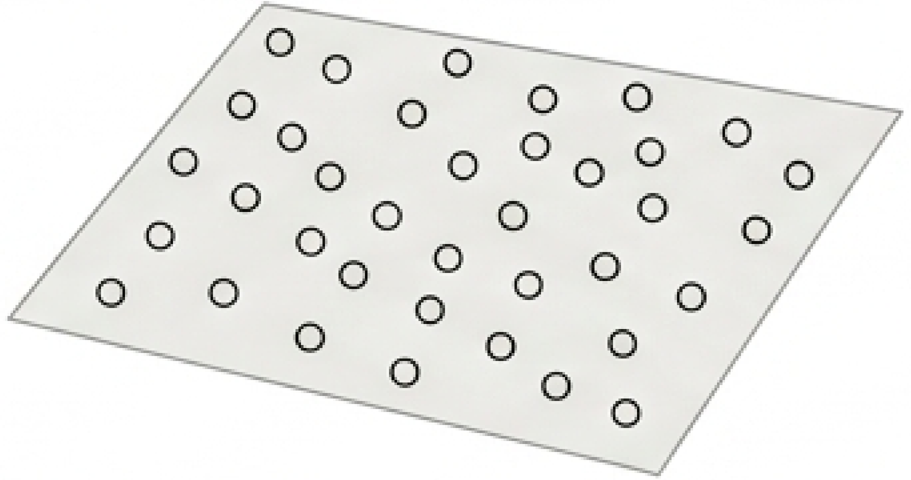
Approximate Inference Methods

When exact inference is NP-hard (large networks), we use Monte Carlo simulation.

Direct Sampling

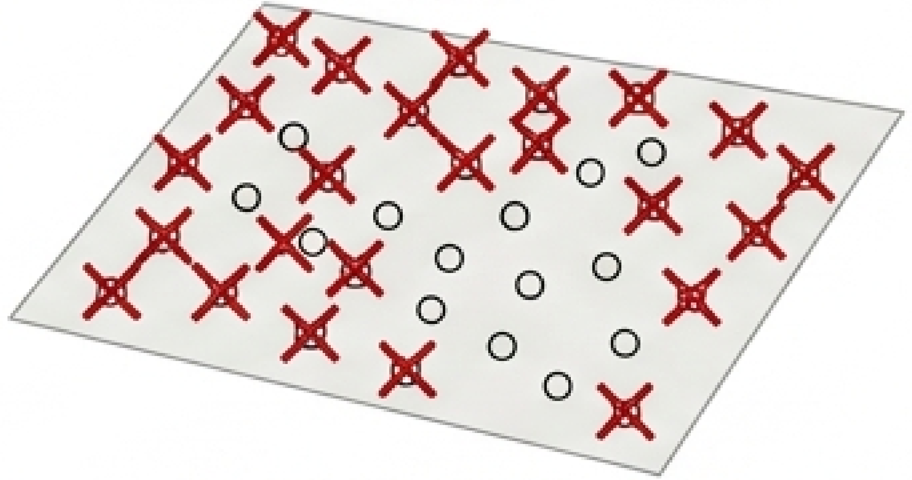
Generate samples from the prior distribution.

$$P(x) \approx \prod P(x_i|\text{parents})$$



Rejection Sampling

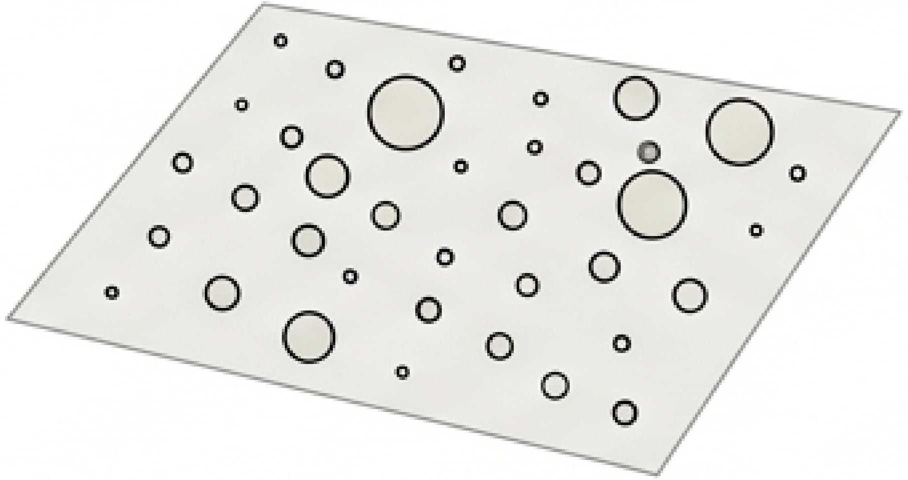
Generate samples, but **reject** those that do not match the evidence.



Likelihood Weighting

Fix evidence variables to observed values. Weight samples by likelihood.

$$w(x, e) = \prod P(e_i|\text{parents}(E_i))$$



Source: Summary Note 5.2.6.3; Lecture 9.

The Transition to Time: Stochastic Processes

Static

$$t-1 \rightarrow t \rightarrow t+1$$

Dynamic

Stochastic Process: A sequence of random variables $X_1, X_2, X_3, \ldots$

Markov Property: The future depends only on the present state, not the past history.

The Markov Assumption:

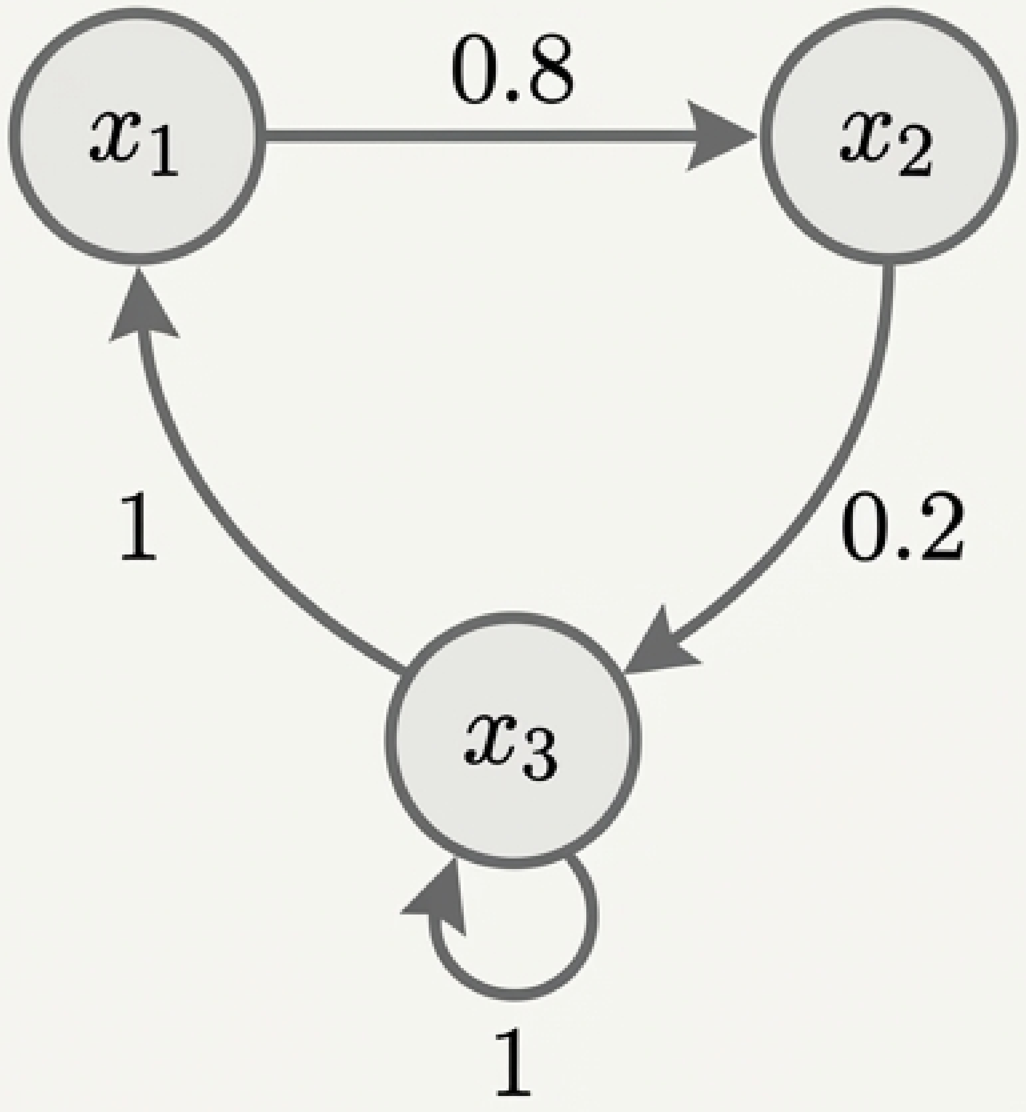
$$P(X_n = x_i | X_{n-1} = x_j, \ldots, X_0 = x_l) = P(X_n = x_i | X_{n-1} = x_j)$$

Stationary Process Assumption:
(Physics don't change over time)

$$\forall t : P(X_n = x_i | X_{n-1} = x_j) = P(X_{n+t} = x_i | X_{n-1+t} = x_j)$$

Source: Lecture 10; Summary Note 5.3.

Stationary Markov Chains



A discrete stationary process with the Markov property.

Law of Total Probability:

$$P(X_n = x_i) = \sum_{j=1}^{N} P(X_n = x_i | X_{n-1} = x_j) P(X_{n-1} = x_j)$$

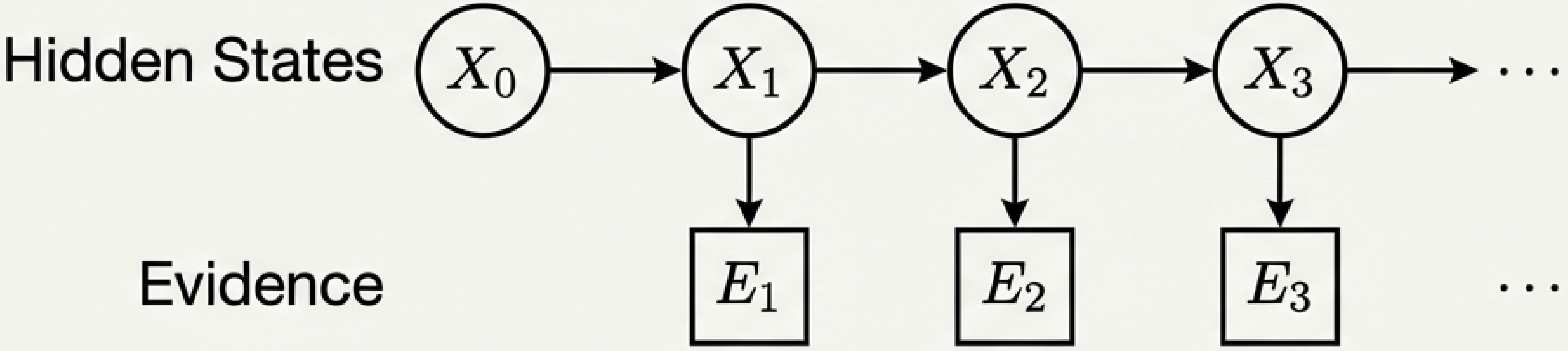
Matrix Notation:

- Transition Matrix T where $T_{i,j} = P(X_n = x_i | X_{n-1} = x_j)$.
- Update Equation: $p_n = T p_{n-1}$
- Long-term: $p_n = T^n p_0$

Source: Lecture 10; Summary Note 5.3.1.

Hidden Markov Models (HMM) Architecture

HMM structure (Trellis)



Transition Model (T)

$$T_{i,j} = P(X_n = x_i | X_{n-1} = x_j)$$

(System dynamics)

Sensor Model (H or O)

$$H_{i,j} = P(E_n = e_i | X_n = x_j)$$

(Observation probability)

Full Joint Distribution:

$$P(X_{0:t}, E_{1:t}) = P(X_0) \prod_{i=1}^{t} P(E_i | X_i) P(X_i | X_{i-1})$$

Source: Lecture 10; Summary Note 5.3.3.

Inference Task 1: Filtering

Goal

Computing Belief State $P(X_t \mid e_{1:t})$ given all evidence to date.

Prior Belief ($f_{1:t}$) → Prediction Step → Correction with Evidence (O_{t+1}) → Current Belief ($f_{1:t+1}$)

Mathematical Derivation

$$P(X_{t+1} \mid e_{1:t+1}) = \underbrace{\alpha P(e_{t+1} \mid X_{t+1})}_{\text{Sensor Model}} \underbrace{\sum_{x_t} P(X_{t+1} \mid x_t)}_{\text{Transition Model}} \underbrace{P(x_t \mid e_{1:t})}_{\text{Previous Result}}$$

Matrix Form (Highlight Box)

$$f_{1:t+1} = \alpha O_{t+1} T f_{1:t}$$

Note: O_{t+1} is the diagonal matrix of sensor probabilities.

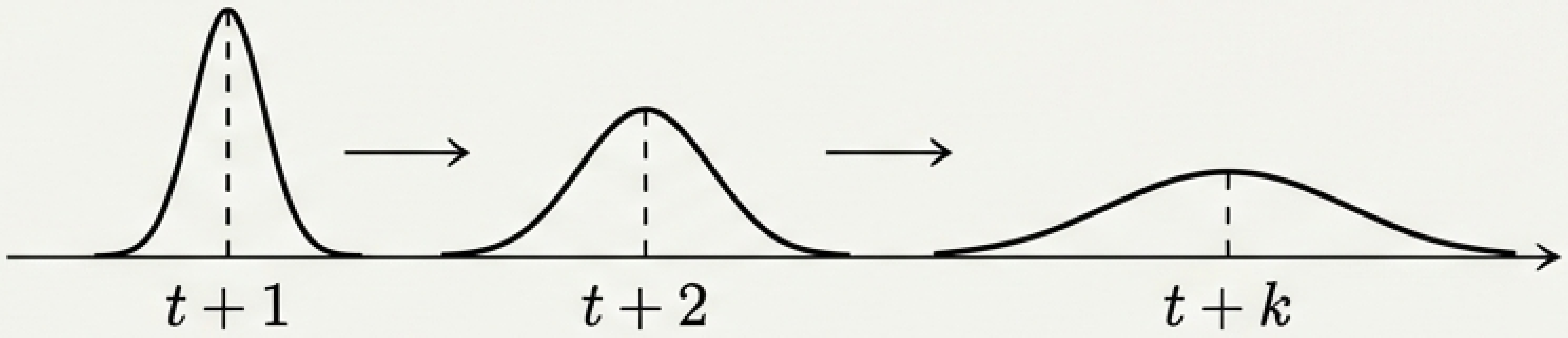
Source: Lecture 10; Summary Note 5.3.4.1.

Inference Task 2: Prediction

Goal

Computing future state $P(X_{t+k} \mid e_{1:t})$ for $k > 0$ (without new evidence).

Visual Graph



Uncertainty increases over time.

Key Formula

$$P(X_{t+k+1} \mid e_{1:t}) = \sum_{x_{t+k}} P(X_{t+k+1} \mid x_{t+k}) P(x_{t+k} \mid e_{1:t})$$

Insight

- As $k \to \infty$, the probability converges to the stationary distribution of the Markov chain.
- Prediction is essentially filtering without the sensor update step.

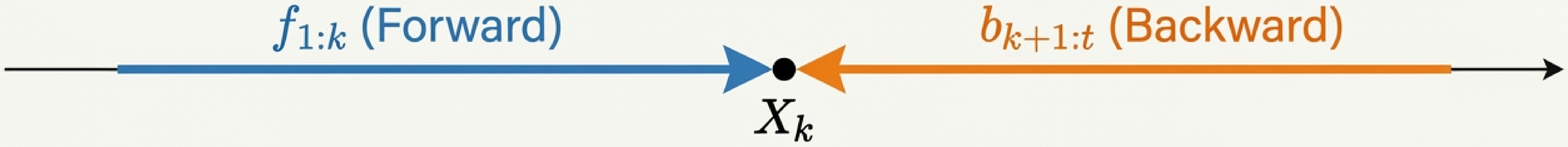
Source: Lecture 10; Summary Note 5.3.4.2.

Inference Task 3: Smoothing

Goal

Computing past state $P(X_k \mid e_{1:t})$ using evidence from the future.

Concept Visual



Key Formula

$$P(X_k \mid e_{1:t}) = \alpha P(X_k \mid e_{1:k}) P(e_{k+1:t} \mid X_k)$$
$$P(X_k \mid e_{1:t}) = \alpha f_{1:k} \times b_{k+1:t}$$

Note: × denotes pointwise vector product.

Source: Lecture 10; Summary Note 5.3.4.3.

The Backward Recursion Algorithm

Computing the "message" from future evidence back to the state of interest.

Algorithm (Pseudo-code style)

```
function Forward-Backward(ev, prior) returns a vector of probability distributions
   inputs: ev, a vector of evidence values for steps 1,...,t;
           prior, the prior distribution on the initial state, P(X0);
   local variables: fv, a vector of forward messages for steps 0,...,t;
                    b, backward message representation, initially all ones;
                    sv, a vector of smoothed estimates for steps 1,...,t
   fv[0] ← prior;
   for i = 1 to t do
      fv[i] ← Forward(fv[i-1], ev[i]);
   for i = t downto 1 do
      sv[i] ← Normalize(fv[i] × b);
      b ← Backward(b, ev[i]);
   return sv
```

Matrix Recursion (The Core Math)

$$b_{k+1:t} = T^T O_{k+1} b_{k+2:t}$$

The transition matrix T is transposed relative to standard definitions.

Initialization

Initialize $b_{t+1:t} = \mathbf{1}$ (vector of ones).

Most Likely Explanation (Viterbi Algorithm)

Goal

Find the single sequence of states that maximizes the probability: $\text{argmax}_{x1:t} P(x_{1:t} \mid e_{1:t})$.

Key Insight

The most likely *sequence* is not necessarily the sequence of most likely *individual states*.

The Recursive Step

The formula replaces Summation (Σ) with Maximization ($\max$). The core recursive step is:

$$\max_{x_{1:t}} P(\dots) = \alpha P(e_{t+1} \mid X_{t+1}) \max_{x_t}(P(X_{t+1} \mid x_t) \max_{x_{1:t-1}} \dots).$$

Complexity

Time: $O(t)$

Space: $O(t)$ (Requires keeping back-pointers to reconstruct the path).

Source: Lecture 10; Summary Note 5.3.4.4.