Once the model is learnt, it can be interesting to **query** the model, i.e. to ask something

* Example 1: the “**Two Box problem**”
Inference

- A possible interesting question can be: “What is the probability of extracting an apple?”
  - I.e. we are interested in $P(F = 'a')$
- This probability is obtained by exploiting the following facts
  - $P(F)$ can be obtained from $P(B,F)$ by marginalization
  - $P(B,F)$ – the joint probability – is given by the Bayesian Network, which also provides a factorization of it
The Bayesian Network provides a factorization of the joint probability

\[
P(F) = \sum_B P(F, B) = \sum_B P(B)P(F|B)
\]
Inference

\[
P(F = 'a') = \sum_{B} P(F = 'a', B)
= \sum_{B} P(B)P(F = 'a'|B)
= P(B = 'b')P(F = 'a'|B = 'b') + P(B = 'r')P(F = 'a'|B = 'r')
= \alpha(1 - \gamma) + (1 - \alpha)(1 - \beta)
= 0.5 \cdot 0.75 + 0.5 \cdot 0.25
= 0.5
\]
**Inference**

- **Exercise**: compute $p(x = 5)$ within a Gaussian Mixture Model with 2 one-dimensional Gaussians

\[
p(g = 1) = \pi_1 \quad p(g = 2) = \pi_2
\]

\[
p(x|g = 1) = \mathcal{N}(x|\mu_1, \sigma_1) \\
p(x|g = 2) = \mathcal{N}(x|\mu_2, \sigma_2)
\]

**Trained GMM**

$\pi_1 = 0.3, \pi_2 = 0.7$

$\mu_1 = 4.2, \sigma_1 = 2$

$\mu_2 = 5.2, \sigma_2 = 1$
Inference

\[ P(x = 5) = \sum_{g} P(x = 5, g) \]
\[ = \sum_{g} P(g)P(x = 5|g) \]
\[ = P(g = 1)P(x = 5|g = 1) + P(g = 2)P(x = 5|g = 2) \]
\[ = \pi_1 \mathcal{N}(x = 5|\mu_1, \sigma_1) + \pi_2 \mathcal{N}(x = 5|\mu_2, \sigma_2) \]

Knowing that
\[ \mathcal{N}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{||x-\mu||^2}{2\sigma^2}} \]

\[ P(x = 5) = 0.3 \frac{1}{\sqrt{2\pi\cdot 2^2}} e^{-\frac{||5-4.2||^2}{2^2}} + 0.7 \frac{1}{\sqrt{2\pi\cdot 1^2}} e^{-\frac{||5-5.2||^2}{2.1^2}} \]
\[ = 0.0721 + 0.2737 \]
\[ = 0.3458 \]
Inference

• Another interesting inference can be performed on the hidden variables
  • To understand the causes
• Two boxes example: knowing that I have extracted an apple, what is the probability that I've chosen the blu box?
  • $P(B = 'b' \mid F = 'a')$
• NOTE: $P(B = 'b' \mid F = 'a')$ is different from the conditional of the Bayesian network – $P( F = 'a' \mid B = 'b')$
Inference

- The procedure is similar

\[ P(B|F) = \frac{P(F, B)}{P(F)} \]

Definition of conditional probability

\[ P(F, B) = P(B)P(F|B) \]

The joint probability (and its factorization) is given by BN

\[ P(F) = \sum_B P(F, B) = \sum_B P(B)P(F|B) \]

P(F) is computed as before via marginalization
Other inferences

- Inference needed during **learning**
  - Example: the E-M for mixture of Gaussians

\[
\begin{align*}
  w_{i1} &= p(y_i = 1|\mathcal{X}, \theta^{(i-1)}) = p(y_i = \text{’blue’}|\mathcal{X}, \theta^{(i-1)})
\end{align*}
\]

- **Optimization**: which is the configuration of the hidden variables for which the joint probability is **maximum**? Which is the **most probable** configuration of the hidden variables?
  - Often called MAP (Maximum a Posteriori) estimation
Inference: summary

- Inference is used to extract interesting information from the Bayesian Network
  - Summaries, causes, optimization
- In general, performing inference is not always so easy (e.g. integrals)

Example: a GMM with an infinite number of components

\[ p(x) = \int g p(x, g) dg \]
Inference: summary

- Another aspect: the structural complexity of Bayesian Networks (e.g. cycles) may make the inference problem intractable.

- (Again) Needed tradeoff: **computability** vs **descriptivity**
Inference: summary

Many complex algorithms have been proposed to perform non-trivial inference (not seen here):

- Exact inference (variable elimination, belief propagation – for trees, ..)
- Variational inference (mean field)
- Monte Carlo inference (Gibbs sampling)

For more info see the Kevin Murphy's tutorial (further readings)
Other Probabilistic Graphical Models

- There are other two families of Probabilistic Graphical Models:

  - Markov Random Fields
  - Factor Graphs
Markov Random Fields

- **Undirected** graph of random variables
- Each variable is independent of all other variables, given its neighbors

![Diagram of Markov Random Fields](attachment://diagram.png)
Markov Random Fields

- Widely used to analyse images

They can model **continuity** (smoothness) and **spatial proximity**, crucial aspects in images

**Example:** noise removal
Factor Graphs

- Less investigated class of probabilistic Graphical Models (introduced by Frey)
- Main idea: to express a global function of several variables as a collection of factors (local functions) over a subset of those variables
- The graph has two kinds of node:
  - Nodes for **variables**
  - Nodes for functions (called **factors**)


Factor Graphs

$x_1, x_2, x_3$ are the variables
$f_a, f_b, f_c,$ and $f_d$ are factors

This factor graph encodes the factorization of a function $g(x_1, x_2, x_3)$ over all the variables

$$g(x_1, x_2, x_3) = f_a(x_1, x_2)f_b(x_3)f_c(x_2, x_3)f_d(x_2, x_3)$$
Factor Graphs

- This formalism is more general than the Bayesian Network formalism

Every Bayesian Network can be written as a Factor Graph

\[ P(B, F) = P(B)P(F|B) \]
This formalism can be used also for “non probabilistic” functions

Example: **Affinity Propagation** algorithm for Clustering

[Science 315, 972–976 (2007)]
Conclusions

- Probabilistic Graphical Models represent a powerful tool to model structured objects
  - Capability to capture the complexity
  - Different information can be extracted
  - Many algorithms / tools to perform training, inference, optimization

- Constraint: tradeoff between **computability** and **descriptivity**
Further Readings