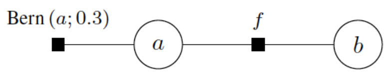
Homework 4

Deadline: December 5, 15:00

Exercise 1

Consider the factor graph below, where a and b can be either true or false. The definition of the factor f is provided in the table, and the prior distribution for a follows the Bernoulli distribution $Bern(0.3)$.



a	b	$f(a, b)$
true	true	0.9
true	false	0.1
false	true	0.2
false	false	0.8

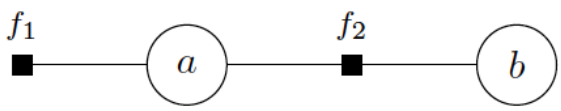
The factor f can be formulated as

$$f(a, b) = 0.9\delta(a)\delta(b) + 0.2\delta(\bar{a})\delta(b) + 0.1\delta(a)\delta(\bar{b}) + 0.8\delta(\bar{a})\delta(\bar{b})$$

Assume that we have observed that b is true. Compute the distribution of a using message passing.

Exercise 2

Consider the factor graph below, where a and b are continuous scalar random variables.



Suppose that $f_1(a) = \mathcal{N}(a; \mu_1, \sigma_1^2)$ and $f_2(a, b) = \mathcal{N}(b; \alpha a, \sigma_2^2)$. Using message passing:

Q1 Compute the marginal distribution of a;

Q2 Compute the marginal distribution of b.

Exercise 3

A company using two different machines produces random amounts of clips and pins each day. Suppose that the production follows a Poisson distribution $P(\lambda)$, where the rate λ depends on the quality of the steel the company is using on a specific day. The Poisson distribution is given by:

$$x \sim P(\lambda) \Leftrightarrow p(x) = P(x; \lambda) = \frac{\lambda^x exp(-\lambda)}{x!}$$

For high-quality steels, $\lambda = 10$, and if the steel is of low quality, $\lambda = 7$. The probability that the company will receive high-quality steel on a specific day is 0.25.
Suppose that at the end of the day the company has produced 10 clips and 8 pins.

Q1 Identify the conditional distributions in the model and draw the Bayesian Network.

Q2 Transform the Bayesian network into a factor graph.

Q3 Using message passing, compute the probability that the company was using high-quality steel.

Q4 Write a program to verify the calculated probability using Monte Carlo simulation.

Exercise 4

Figure 1 shows a model for capturing the inter dependencies between 5 discrete random variables related to a hypothetical student taking a class. The "student" Bayesian network models the dependencies between the following variables:

- D: Difficulty of the class ({Easy, Hard}),
- I: Intelligence of the student ({Low, High}),
- G: Grade of the student in the class ({A, B, C}),
- S: SAT score of the student ({Bad, Good}),
- L: Letter of recommendation ({Bad, Good}).

The conditional probability tables (CPTs) and graph structure for this network are shown in Figure 4.2. The joint probability distribution can be expressed as:

$$p(L, S, G, D, I) = p(L \mid G) \cdot p(S \mid I) \cdot p(G \mid D, I) \cdot p(D) \cdot p(I).$$

Q1 Write the full joint probability $p(L, S, G, D, I)$ using the given graph and CPTs.

Q2 Identify two sets of variables that are conditionally independent given another variable in the network.

Q3 Calculate the probability that a student is intelligent ($I =$ High) given that the student received a grade of C ($G = C$) using the CPTs provided. That is, compute:

$$p(I = \text{High} \mid G = C).$$

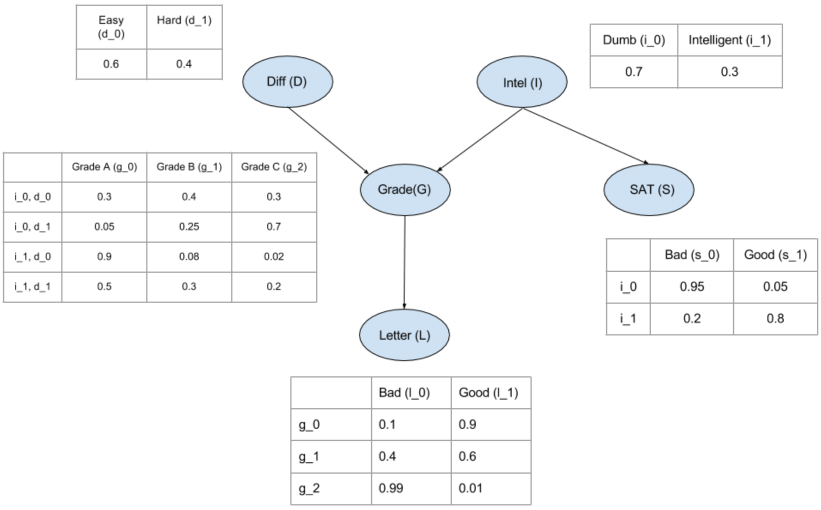


Figure 1: The (simplified) student network. "Diff" is the difficulty of the class. "Intel" is the intelligence of the student. "Grade" is the grade of the student in this class. "SAT" is the score of the student on the SAT exam. "Letter" is whether the teacher writes a good or bad letter of recommendation. The circles (nodes) represent random variables, the edges represent direct probabilistic dependencies. Th

Q4 Suppose we observe that the student has a good SAT score ($S = \text{Good}$) in addition to a grade $G = C$. Calculate the probability that a student is intelligent ($I = \text{High}$).

Q5 by calculating $p(D = \text{Hard} \mid G = C, S = Good)$, explain what cause this observation influences our belief.

Q6 explain the difference between causality and correlation using an example from the figure 1.

Exercise 5

One application of Markov Random Fields (MRFs) is to denoise noisy images. In this task, we minimize the following energy function using the **Iterated Conditional Mode (ICM)**:

$$E(x, y) = h \sum_i x_i - \beta \sum_{i,j} x_i x_j - \eta \sum_i x_i y_i,$$

where:

- $x_i \in \{-1, +1\}$ represents the binary state of pixel i in the denoised image,
- $y_i \in \{-1, +1\}$ represents the binary state of pixel i in the noisy observed image,
- h, β, and η are scalar parameters controlling the trade-offs between the terms.

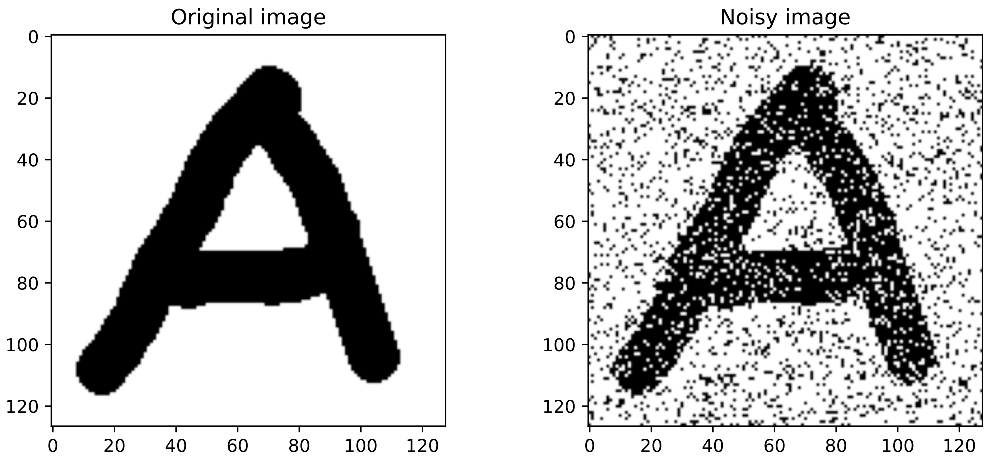


Figure 2: Original vs noisy image

Iterated Conditional Mode (ICM)

1. **Initialization**: Set $x_i = y_i$ for all i.
2. For each pixel $j = 1, \ldots, N$:
   - Compute the energy $E(x, y)$ for the states $x_j = +1$ and $x_j = -1$, keeping all other variables fixed.
   - Update x_j to the state that minimizes the energy.
3. Repeat until convergence or a stopping criterion is met.

You can download the noisy image for this exercise from the following link. The image is stored in a CSV file format. You can read it into Python as follows:

```
import pandas as pd
img = pd.read_csv('letterA.csv').to_numpy()
```

Tasks

Q1 Simulate noisy images with varying noise levels (e.g., Gaussian noise with different standard deviations or salt-and-pepper noise with varying probabilities).

Q2 Implement the ICM algorithm to minimize the energy function.

Q3 Plot the denoised image after iterations from the following list: $[0, 1, 5, 15, 50]$.

Q4 Compute and plot the **Normalized Mean Squared Error (NMSE)** between the denoised image (D) and the original ground truth image (G) for each noise level:

$$\text{NMSE} = \frac{\|D - G\|^2}{\|G\|^2}.$$

Q5 Analyze the performance of the denoising algorithm as the noise level increases.

Q6 Vary the parameters h, β, and η in the energy function and evaluate their impact on denoising performance.

Q7 For each parameter, plot:

- NMSE vs h,
- NMSE vs β,
- NMSE vs η.

Q8 Discuss and justify how each parameter influences the trade-offs between pixel similarity, smoothness, and adherence to the noisy image.