# STA2700 - Graphical Models - Assignment 2

Luis Correia - Student No. 1006508566

October 21st 2020

## Question 1

We again consider binary (i.e., {0,1}-valued) sequences of length $N$, in which no two 1's are adjacent. for $1 \leq i < N$ and for adjacent variables $x_i, x_{i+1}$, let

$$f_i(x_i, x_{i+1}) = \begin{cases} 0 \text{ , if } x_i = x_{i+1} = 1 \\ 1 \text{ , otherwise} \end{cases} \tag{1}$$

and

$$p(\boldsymbol{x}) \propto \prod_{i=1}^{N-1} f_i(x_i, x_{i+1}) \tag{2}$$

$$p(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{Z} \prod_{i=1}^{N-1} f_i(x_i, x_{i+1}) \prod_{i=1}^{N-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_i - x_i)^2}{2}} \tag{3}$$

### Item(a)

Let $N = 100$. Draw one random sample according to $p(\boldsymbol{x})$ and call it the *input* - denoted by $\boldsymbol{x}$.

{*Solution.*}

The sample of size $N = 100$ generated according $p(\boldsymbol{x})$ is given below:

```
##   [1] 0 0 1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 1 0
##  [38] 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 1 0
##  [75] 1 0 0 0 1 0 0 1 0 0 0 0 1 0 0 1 0 1 0 0 0 1 0 0 0 1
```

### Item(b)

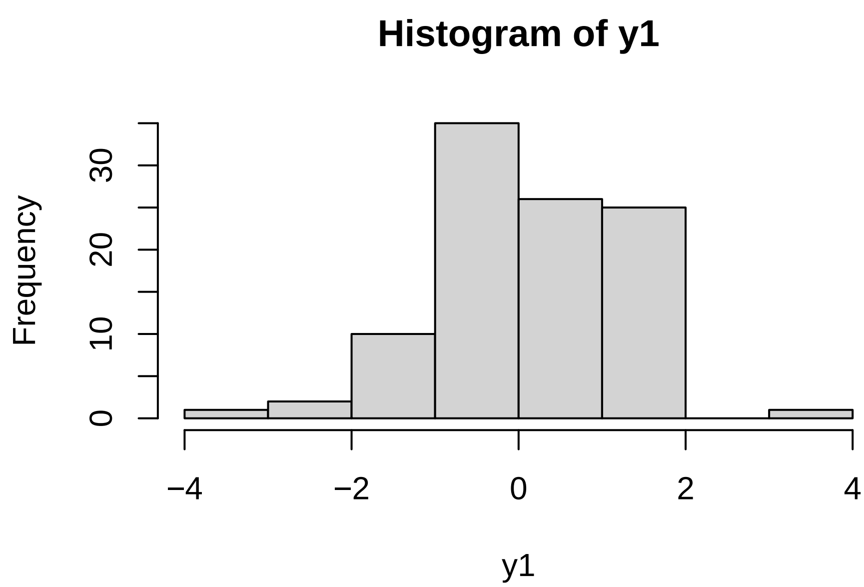
From the generated input $\boldsymbol{x}$, create the *output* $\boldsymbol{y}$.

{*Solution.*}

For this part, we just generated $N = 100$ samples from $y_i \sim N(x_i, 1)$ with $i = 1, \ldots, N$.

```
##
##
## The sample Y obtained is:
```

```
##   [1]  0.63889623 -0.86228942 -0.72124013  1.98113737 -0.99645488 -1.18409311
##   [7] -0.45989308 -0.16845544 -0.05823319 -0.86689151  0.28110749  1.25743632
##  [13]  3.00991530  0.37721293 -0.09506081  0.06373301  1.66278878 -0.52913875
##  [19]  0.80193048 -0.69083437  1.58843866  0.38744844 -0.65244898  0.76576102
##  [25] -0.71262777 -0.32097771  0.49746443 -0.73553497  0.63117694 -1.50837571
##  [31] -1.01170831 -0.38626015  0.38280445 -0.79611609 -0.45716620  0.77741750
##  [37]  1.39943010 -1.57328745 -0.25916270 -2.52119238  1.49283580  0.82906687
##  [43]  1.10477542  0.44507361 -0.86669931 -3.27623917  0.30695650  0.22791645
##  [49]  0.56714186  1.48025317  1.24079966  1.54415483  1.19761206 -0.13397438
##  [55]  1.55677715  0.76925063 -0.32164714  1.51653200 -1.18441568  0.73046364
##  [61] -0.57869124  1.33791008 -0.26047867  1.75761254  1.08640695 -0.12922324
##  [67] -0.48424158 -1.84141702 -0.17828195  0.09532114  0.65978589  0.54712641
##  [73]  1.05423456 -0.80914054  1.20714357 -1.38468640 -2.45606972  1.04867530
##  [79]  0.80643969 -0.89054127  1.20200748  0.80420424 -0.59545917 -0.22603283
##  [85] -1.30950375  0.78573206  1.21665671  1.05355100 -0.30308436  1.09583207
##  [91]  0.39478234  1.97639397 -0.68571440 -1.51669457 -0.41984570  1.02635521
##  [97] -0.08587071 -0.11851469 -1.02609039  0.70714256
```




Similarly, we could obtain the same result by setting $y_i = x_i + \eta_i$ with $\eta_i \sim N(0,1)$ (*white noise*).

## Item(c)

Apply the sum-product algorithm to compute the probability of observing this particular $\boldsymbol{y}$ as the output.

{*Solution.*}

Note that our graph satisfies the key conditional independence property that $x_{n-1}$ and $x_{n+1}$ are independent given $x_n$, so that

$$x_{n+1} \perp\!\!\!\perp x_{n-1} | x_n \tag{4}$$

This implies that the probability of observing this particular $\boldsymbol{y}$ as our output can be calculated by:

$$\begin{aligned}
p(\boldsymbol{y}) &= \frac{1}{Z_N} \sum_{\boldsymbol{x} \in \mathcal{X}^n} p(\boldsymbol{x}, \boldsymbol{y}) \\
&= \frac{1}{Z_N} \sum_{\boldsymbol{x} \in \mathcal{X}^n} p(\boldsymbol{y}|\boldsymbol{x}) p(\boldsymbol{x}) \\
&= \frac{1}{Z_N} \sum_{\boldsymbol{x} \in \mathcal{X}^n} p(y_1|x_1) p(y_2|x_2) \dots p(y_n|x_n) p(x_1) p(x_2|x_1) p(x_3|x_2) \dots p(x_n|x_{n-1}) \\
&= \frac{1}{Z_N} \sum_{\boldsymbol{x} \in \mathcal{X}^n} \overbrace{\prod_{i=1}^{N} p(y_i|x_i)}^{p(\boldsymbol{y}|\boldsymbol{x})} \overbrace{p(x_1) \prod_{i=2}^{N} p(x_i|x_{i-1})}^{p(\boldsymbol{x})}
\end{aligned}$$

$$\implies p(\boldsymbol{y}) = \frac{1}{Z_N} \sum_{\boldsymbol{x} \in \mathcal{X}^n} \Big( \prod_{i=1}^{N} p(y_i|x_i) \times p(x_1) \prod_{i=2}^{N} p(x_i|x_{i-1}) \Big) \tag{5}$$

As we can see, (5) consists in applying the sum-product algorithm, using the marginals calculated in item(a), multiplying by the conditional probability of $\boldsymbol{y}|\boldsymbol{x}$.

The constant $Z_N$ is *normalization constant* calculated using the routine developed in Assignment 1.

After calculating the probabilities and running the sum-product algorithm, we obtained the following result for $p(\boldsymbol{y})$:

```
##
## The probability of observing the particular sample y obtained in item(a) is  1.380916e-85
```

# Question 2

Let $X$ be a binary variable (i.e., $\mathcal{X} = \{0, 1\}$), and let $Y_1$ and $Y_2$ be two independent rea-valued measurements of $X$. Assume that the process is modeled by the conditional densities $f(y_1, y_2|x) = f(y_1|x)f(y_2|x)$ with:

$$f(y_k|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_k - x)^2}{2\sigma^2}}, \text{ for } k = 1, 2$$

We want to compute $p(x|y_1, y_2)$ given two measurements $Y_1 = y_1$ and $Y_2 = y_2$.

## Item(a)

Draw the factor graph of this model and explain how the sum-product algorithm can be employed to carry out such computation.

{*Solution.*}

According with the description we can represent this graphical model as follows:

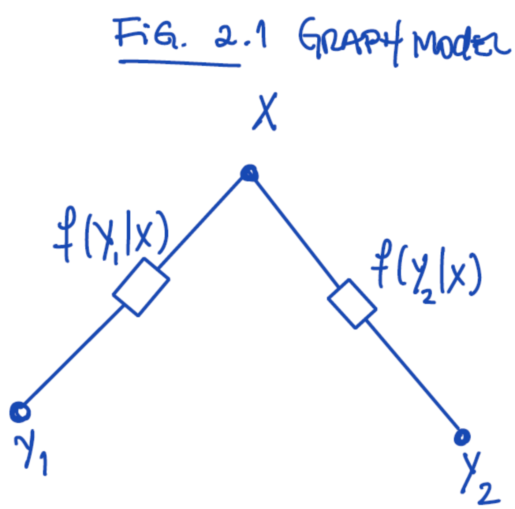



Figure 1: Graphical Model

We want to calculate $p(x|y_1, y_2)$ for given two measurements of $y_1$ and $y_2$.

This probability can be decomposed as follows:

$$\begin{aligned} p(x|y_1, y_2) &= \frac{p(x, y_1, y_2)}{p(y_1, y_2)} \\ &= \frac{p(y_1, y_2|x)p(x)}{p(y_1, y_2)} \end{aligned}$$

As $Y_1$ and $Y_2$ are independent, it follows that:

$$
\begin{aligned}
p(x|y_1,y_2) &= \frac{p(y_1|x)p(y_2|x)p(x)}{p(y_1)p(y_2)} \\
&= \frac{f(y_1|x)f(y_2|x)p(x)}{p(y_1)p(y_2)} \\
&= \frac{1}{Z}\sum_{x\in\mathcal{X}}\Big(\mu_{y_1}\to f(y_1|x)\times\mu_{y_2}\to f(y_2|x)\times p(x)\Big)
\end{aligned}
$$

... where

$$
Z = \Big(\sum_{x\in\mathcal{X}} p(y_1|x)\Big)\times\Big(\sum_{x\in\mathcal{X}} p(y_2|x)\Big)
$$

Then, the sum-product algorithm can be applied by calculating:

$$
p(x|y_1,y_2) = \frac{\sum_{x\in\mathcal{X}}\Big(\mu_{y_1}\to f(y_1|x)\times\mu_{y_2}\to f(y_2|x)\times p(x)\Big)}{\Big(\sum_{x\in\mathcal{X}}p(y_1|x)\Big)\Big(\sum_{x\in\mathcal{X}}p(y_2|x)\Big)} \tag{6}
$$

## Item(b)

Let $\sigma^2 = 2$. Also, assume that we are given the priori probability $p_X(1) = 0.1$, and two measurements $y_1 = -0.25$ and $y_2 = 0.94$. Apply the sum-product algorithm to compute $p(x|y_1,y_2)$ numerically.

{*Solution.*}

I applied the sum-product routine using the designed on (6) and obtained the following result:

```
##
## The probability p(x|y1,y2) is  0.2599199
```

# Question 3

Draw the factor graph of $f$ and $g$. If possible, give an example of a pairwise Markov property for each factor graph.

$$f(\boldsymbol{x}) = f_1(x_1, x_2)f_2(x_2, x_3, x_4, x_5)$$

and

$$g(\boldsymbol{x}) = g_1(x_1, x_2)g_2(x_2, x_3)g_3(x_3, x_4)g_4(x_1, x_4)g_5(x_1, x_3)g_6(x_2, x_4)$$

{*Solution.*}

We can represent the graphs as follows:

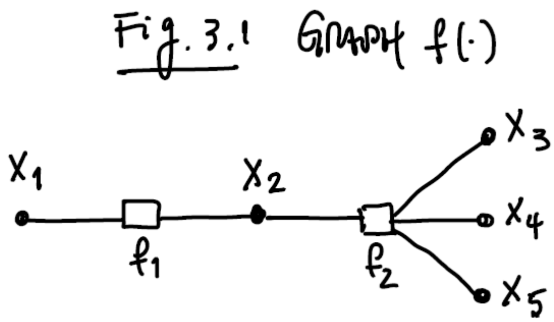



Figure 2: Graph of f(.)

... and...

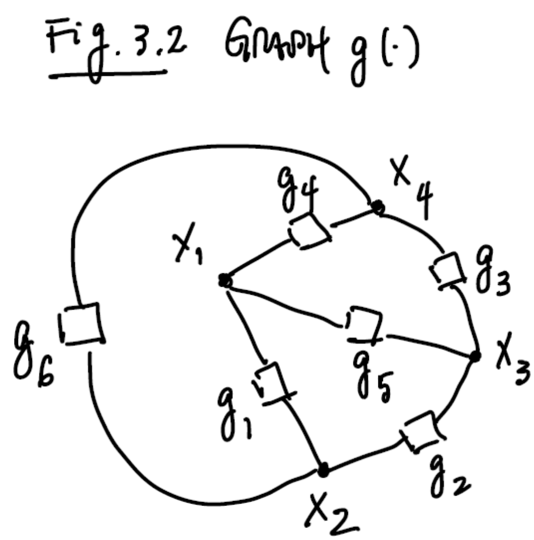



Figure 3: Graph of g(.)

The **Pairwise Markov Property** dictates that $\mu(.)$ *satisfies the Pairwise Markov property with respect to a graph* $G$ *if any* $i, j \in \mathcal{V}$ *not connected by an edge, we have*

$$\mu(x_i, x_j|x_{rest}) = \mu(x_i|x_{rest})\mu(x_j|x_{rest})$$

In the case of grapg $f$ we clearly can split it in 02 sub-graphs linked by $X_2$ such that the pairwise Markov property holds as follows:

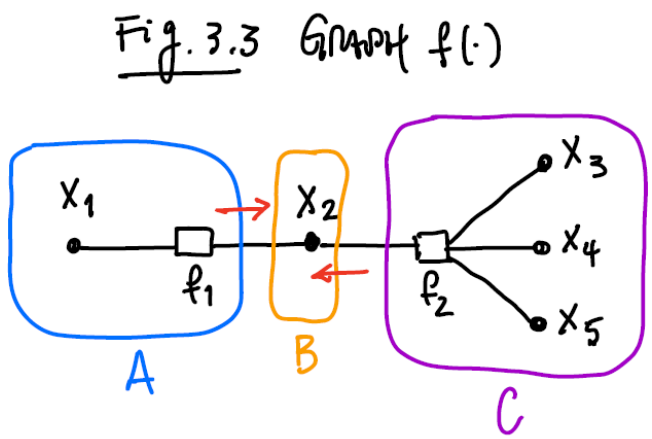



Figure 4: Graph of f(.)

... as we can see the property holds for any pair of *disconnected* vertices:

$$\mu(x_1, x_3|x_2) = \mu(x_1|x_2)\mu(x_3|x_2)$$
$$\mu(x_1, x_4|x_2) = \mu(x_1|x_2)\mu(x_4|x_2)$$
$$\mu(x_1, x_5|x_2) = \mu(x_1|x_2)\mu(x_5|x_2)$$

In case of graph $g$ **the property doesn't holds as all vertices are interconnected**.

# Question 4

Consider the factor graph for the factorization $f(x_1, x_2, x_3) = f_1(x_1, x_2)f_2(x_2, x_3)f_3(x_1, x_3)$ with binary variables. Suppose, after convergence, the sum-product algorithm gives the following set of beliefs in the vector/matrix form

$$b_1(x_1) = b_2(x_2) = b_3(x_3) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

and

$$b_1(x_1, x_2) = \begin{bmatrix} 0.49 & 0.01 \\ 0.01 & 0.49 \end{bmatrix}$$

$$b_2(x_2, x_3) = \begin{bmatrix} 0.49 & 0.01 \\ 0.01 & 0.49 \end{bmatrix}$$

$$b_3(x_1, x_3) = \begin{bmatrix} 0.01 & 0.49 \\ 0.49 & 0.01 \end{bmatrix}$$

Show that the beliefs over variable nodes and factor nodes are locally consistent; but can not be the marginals of any global PMF $p(x_1, x_2, x_3)$.

Indeed, this example shows that there are locally consistent beliefs that do not correspond to any global distribution.

{*Solution.*}

## Part1 - Variables/Nodes locally consistent

For this part of demonstration, I will refer the definition of *locally consistent marginals*[1]:

**Definition**: a collection of distributions $b_i(.)$ over $\mathcal{X}$ for each $i \in \mathcal{V}$ and $b_a(.)$ over $\mathcal{X}^{|\partial_a|}$ for each $a \in \mathcal{F}$ is locally consistent if they satisfy:

- Normalization Condition

$$\sum_{x_i} b_i(x_i) = 1, \forall i \in \mathcal{V} \tag{7}$$

$$\sum_{\underline{x}_{\partial_a}} b_a(\underline{x}_{\partial_a}) = 1, \forall a \in \mathcal{F} \tag{8}$$

- Marginalization Condition

$$\sum_{\underline{x}_{\partial_a} i} b_a(\underline{x}_{\partial_a}) = b_i(x_i), \forall a \in \mathcal{F}, \forall i \in \mathcal{V} \tag{9}$$

We have the following graphical model configuration:

[1] Mézard, Montanari - Information, Physics, and Computation (2009), pag.306-307

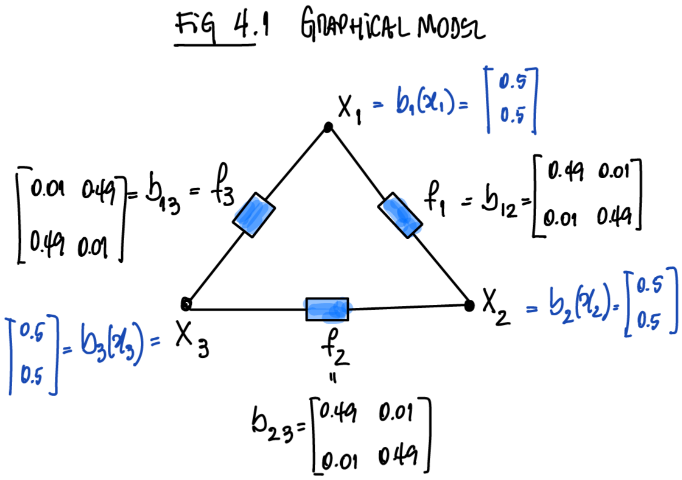



Figure 5: Graphical Model of f(.)

Applying (7) and (8), we clearly see from the normalization conditions are satisfied because all summations of variable nodes $b_i(x_i)$ and the factor $b_a(\underline{x}_{\partial_a})$ - here represented by $b_1(x_1, x_2)$, $b_2(x_2, x_3)$ and $b_3(x_1, x_3)$ - are equal to 1.

Now applying the Marginalization Condition on (9), we have that:

- For $i = 1$:

$$\sum_{\underline{x}_{\partial_a} \setminus 1} b_a(\underline{x}_{\partial_a}) = \frac{1}{2} \overbrace{\begin{bmatrix} 0.49 & 0.01 \\ 0.01 & 0.49 \end{bmatrix}}^{b_2(x_2, x_3)} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$
$$= b_1(x_1).$$

- For $i = 2$:

$$\sum_{\underline{x}_{\partial_a} \setminus 1} b_a(\underline{x}_{\partial_a}) = \frac{1}{2} \overbrace{\begin{bmatrix} 0.01 & 0.49 \\ 0.49 & 0.01 \end{bmatrix}}^{b_3(x_1, x_3)} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$
$$= b_2(x_2).$$

- For $i = 3$ we have $b_2(x_2, x_3) = b_1(x_1, x_2)$ so the condition also holds.

As we have (7), (8) and (9) satisfied, we have **the beliefs over variable nodes and factor nodes are locally consistent.**

These results can be verified numerically (below).

```
##
## ---------------------------------
```

```
##
## Beliefs over Variables X1, X2, X3
##
## Blf-X1=
##      [,1]
## [1,]  0.5
## [2,]  0.5
##
## Blf-X2=
##      [,1]
## [1,]  0.5
## [2,]  0.5
##
## Blf-X3=
##      [,1]
## [1,]  0.5
## [2,]  0.5
##
##
## ---------------------------------
##
## Beliefs over Factor Nodes f1, f2, f3
##
## Blf-f1=
##      [,1]
## [1,]  0.5
## [2,]  0.5
##
## Blf-f2=
##      [,1]
## [1,]  0.5
## [2,]  0.5
##
## Blf-f3=
##      [,1]
## [1,]  0.5
## [2,]  0.5
```

## Part2 - Global PMF is inconsistent

A Global PMF $\mathbb{P}(\underline{x})$ is given by:

$$\mathbb{P}(\underline{x}) \propto \prod_{a=1}^{M} \psi_a(\underline{x}_{\partial_a}) \tag{10}$$

... and must satisfy

$$\sum_{\underline{x} \in \mathcal{X}} \mathbb{P}(\underline{x}) = 1 \tag{11}$$

In our present case, we can calculate the terms in (10) by using the factor node beliefs of our graph, such that we cover all variable nodes. Then we have:

$$\begin{aligned}
\mathbb{P}(\underline{x}) &\propto \prod_{a=1}^{3} \psi_a(\underline{x}_{\partial_a}) \\
&\propto \psi_1(\underline{x}_{\partial_1})\psi_2(\underline{x}_{\partial_2})\psi_3(\underline{x}_{\partial_3}) \\
&\propto b_1(x_1, x_2) b_2(x_2, x_3) b_3(x_1, x_3) \\
&\propto \overbrace{\begin{bmatrix} 0.49 & 0.01 \\ 0.01 & 0.49 \end{bmatrix}}^{b_1(x_1,x_2)} \times \overbrace{\begin{bmatrix} 0.49 & 0.01 \\ 0.01 & 0.49 \end{bmatrix}}^{b_2(x_2,x_3)} \times \overbrace{\begin{bmatrix} 0.01 & 0.49 \\ 0.49 & 0.01 \end{bmatrix}}^{b_3(x_1,x_3)} \\
&\propto \begin{bmatrix} 0.007204 & 0.117796 \\ 0.117796 & 0.007204 \end{bmatrix}
\end{aligned}$$

Using this result to calculate the overall probability in (11) by applying the sum-product algorithm, we have:

$$\begin{aligned}
\sum_{\underline{x} \in \mathcal{X}} \mathbb{P}(\underline{x}) &= \sum_{\underline{x} \in \mathcal{X}} \left( \frac{1}{Z} \prod_{a=1}^{3} \psi_a(\underline{x}_{\partial_a}) \right) \\
&= \frac{1}{Z} \sum_{\underline{x} \in \mathcal{X}} \left( \psi_1(\underline{x}_{\partial_1})\psi_2(\underline{x}_{\partial_2})\psi_3(\underline{x}_{\partial_3}) \right) \\
&= \frac{1}{2} \left( \begin{bmatrix} 1 & 1 \end{bmatrix} \times \begin{bmatrix} 0.007204 & 0.117796 \\ 0.117796 & 0.007204 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) \\
&= 0.125 < 1
\end{aligned}$$

In this case the condition to $\mathbb{P}(\underline{x})$ in (11) is not verified then **the marginals does not fit a global PMF**.

```
##
##
## ---------------------------------

##
## Global Belief

##
## Glb-bf=

##           [,1]     [,2]
## [1,] 0.007204 0.117796
## [2,] 0.117796 0.007204

##
## Total Probability of Global PMF

##       [,1]
## [1,] 0.125
```