Probability terminology
○○○○○○○○○

Bayesian networks
○○○○○○○○○○○○○○○

Gaussian properties
○○○○

Exercise
○○○○

Lab 2 : Review on Bayesian Statistics

Jeroen Olieslagers
Richard-John Lin

Center for Data Science

09/13/2022

Competition results

Speed
As some of you noticed, a slightly more efficient way to compute the variance is to notice that :

$$\mathbb{V}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

As a general rule, try using numpy functions and numpy broadcasting as much possible. To compute $\mathbb{E}[X^2]$, we have 2 efficient solutions : np.dot and @ (same as np.matmul). For 1D vector dot product, it is recommended to use np.dot, while for 2D matrix product you should use @.

▶ Tina Wan : Using np.dot
▶ Yirong Bian and Hanyuan Zhang : Computing using $X.T@X$ (note that it is equivalent to $X@X$ for 1D array)
▶ Yuhao Zheng : Naive variance computation using numba's njit decorator

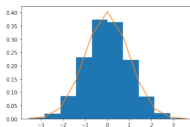| Probability terminology	Bayesian networks	Gaussian properties	Exercise
○○○○○○○○○	○○○○○○○○○○○○○○	○○○○	○○○○

Competition results

Originality

▶ Jiayun Wang : Fitting a normal distribution on this particular dataset


```
mu, std = norm.fit(x_np)
_, bins, _ = plt.hist(x_np, density=True)
plt.plot(bins, norm.pdf(bins, mu, std))
```

[<matplotlib.lines.Line2D at 0x7fa6cad6f940>]

According to the shape of the histogram we observe that the data is a fair fit of Gaussian distribution, so we can fit a normal curve to the data using scipy to find out the mean and variance.

▶ Ken Zeng : Defining the mean as $\underset{\theta}{\text{argmin}}\ \mathbb{E}[(X - \theta)^2]$ and minimizing this cost via gradient descent on a linear layer.

▶ Boyu Hu and Mingxuan Wu : Online mean computation

Table of Contents

Introduction

▶ Experiment : Repeatable operation carried out under controlled conditions, with well defined outcomes.

▶ Sample space : Set of possible outcomes, often denoted Ω
    Rolling a dice : $\Omega = \{1, 2, 3, 4, 5, 6\}$.

▶ Event : A subset of the sample space
    Event of obtaining a number greater than $4 = \{5, 6\}$.

▶ Random variable : Mapping from the possible outcomes in the sample space to a measurable space (often a real number)
    $X =$"The number rolled, mod 2" $\in \{0, 1\}$

▶ Probability : How likely an event is to occur.
    We often denote $p(x) = \mathbb{P}(X = x)$.

Independence

Two random variables A, B are said to be independent if any of the following is satisfied :

1. $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
2. $\mathbb{P}(A|B) = \mathbb{P}(A)$

Mutual independence

Let I be an ensemble and $(A_i)_{i \in I}$ a set of events. The $(A_i)_{i \in I}$ are said to be **mutually independent** iif

$$\forall J \in \mathcal{P}(I), \quad \mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

A set of events is said to be mutually independent if the probability of each event in the set is the same no matter which of the other events has occurred.

A remark

In particular when $I = J$:

$$\mathbb{P}\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} \mathbb{P}(A_i).$$

This formula is NOT equivalent to mutual independence.

Counter example

Assume we are doing coin flips.
A : The first flip is T
B : At most one T in the first 3 flips
C : Same as B
We have $\mathbb{P}(A \cap B \cap C) = \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ but the events are clearly not mutually independent !

Computing a probability

There a 3 fundamental theorems for computing probabilities :

▶ Law of total probability (aka marginalization or sum rule)

$$\mathbb{P}(A) = \sum_{b \in B} \mathbb{P}(A \cap B = b)$$

▶ Chain rule (aka product rule)

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|A, B)$$

▶ Bayes' theorem

$$\mathbb{P}(A \cap B) = \frac{\mathbb{P}(B \cap A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

| Probability terminology | Bayesian networks | Gaussian properties | Exercise |
| 000000●00 | 00000000000000 | 0000 | 0000 |

Law of total probability

Marginalization or process of "forgetting irrelevant information".

$$\mathbb{P}(A) = \sum_{b \in B} \mathbb{P}(A \cap B = b).$$

More generally, let $(B_i)_{i \in I}$ a collectively exhaustive family of events, i.e. $\sum_{i \in I} \mathbb{P}(B_i) = 1$,

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap B_i).$$

Conditional probability

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A).$$

Chain rule

By proceeding recursively, we can expand the formula.
Let $(A_i)_{i \in [\![1,n]\!]}$ a set of random variables, s.t. $\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) \neq 0$

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1) \prod_{i=2}^{n} \mathbb{P}\left(A_i \middle| \bigcap_{j=1}^{i-1} A_j\right).$$

Bayes Theorem

Bayes theorem derives directly from the definition of conditional probabilities.

Bayes Theorem

Let A, B two events such as $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$. Then,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

$$\texttt{posterior} = \frac{\texttt{likelihood} \times \texttt{prior}}{\texttt{evidence}}.$$

As we use it to estimate parameters, we sometimes drop the evidence:

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta).$$

► We can use these three theorems to take advantage of
  structure in models to make computations cheaper and
  inference easier

► Probabilistic models with any sort of conditional structure can
  be described as Bayesian networks

► Formally :

Bayesian network

A probabilistic graphical model representing the joint probability of
a set of random variables, capturing conditional dependencies via a
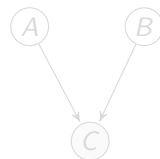Directed Acyclic Graph (DAG).

Probability terminology
000000000

Bayesian networks
00●000000000000

Gaussian properties
0000

Exercise
0000

Conditioning : 3 fundamental patterns
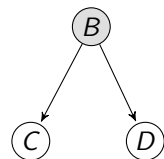
(a) Independent (b) Independent (c) Not independent

Figure 1: Three fundamental patterns

$$p(B, C, D) = p(B)p(C|B)p(D|B)$$
$$p(C, D|B) = p(C|B)p(D|B)$$
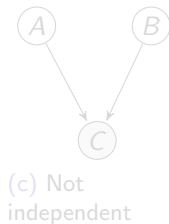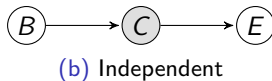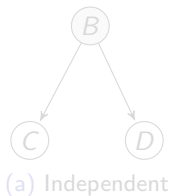
Conditioning : 3 fundamental patterns

$B \longrightarrow C \longrightarrow E$

(a) Independent   (b) Independent   (c) Not independent

Figure 1: Three fundamental patterns

$$p(B, C, E) = p(B)p(C|B)p(E|C)$$
$$= p(B)\frac{p(B|C)p(C)}{p(B)}p(E|C)$$
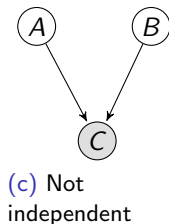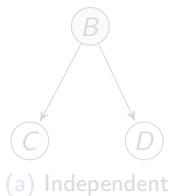$$p(B, E|C) = p(B|C)p(E|C)$$

Conditioning : 3 fundamental patterns

Figure 1: Three fundamental patterns

$$p(A, B, C) = p(A)p(B)p(C|A, B)$$
$$= \frac{p(C)^2 p(A|C)p(B|C)}{p(C|A)p(C|B)} p(C|A, B)$$
$$p(A, B|C) = p(A|C)p(B|C)\frac{p(C)p(C|A, B)}{p(C|A)p(C|B)}$$
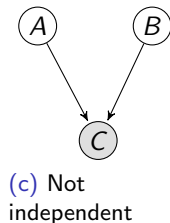
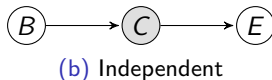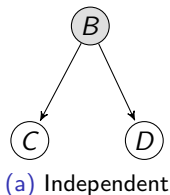## Conditioning : 3 fundamental patterns

Figure 1: Three fundamental patterns

The **Markov blanket** for an ensemble X consists in the **children**, **parents** and the **co-parents**.
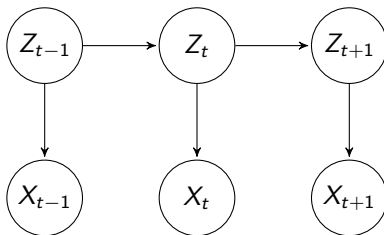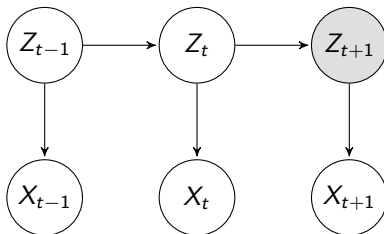
Example Bayesian network

Figure 2: Bayesian Network

Fact

Each variable is conditionally independent of all its nondescendants in the graph given the value of all its parents.

Probability terminology
000000000

Bayesian networks
0000●00000000000

Gaussian properties
0000

Exercise
0000

Example Bayesian network

$$X_{t+1} \perp\!\!\!\perp Z_{t-1}, X_{t-1}, Z_t, X_t \qquad |Z_{t+1}$$

$$X_t \quad \perp\!\!\!\perp Z_{t-1}, X_{t-1}, Z_{t+1}, X_{t+1} \quad |Z_t$$

$$Z_{t+1} \perp\!\!\!\perp Z_{t-1}, X_{t-1}, X_t \qquad |Z_t$$

$$X_{t-1} \perp\!\!\!\perp Z_t, X_t, Z_{t+1}, X_{t+1} \qquad |Z_{t-1}$$

$$Z_t \quad \perp\!\!\!\perp X_{t-1} \qquad\qquad\qquad |Z_{t-1}$$

Probability terminology
000000000

Bayesian networks
000000000000000

Gaussian properties
0000

Exercise
0000

Example Bayesian network

$$X_{t+1} \perp\!\!\!\perp Z_{t-1}, X_{t-1}, Z_t, X_t \qquad |Z_{t+1}$$

$$X_t \quad \perp\!\!\!\perp Z_{t-1}, X_{t-1}, \textcolor{red}{Z_{t+1}}, X_{t+1} \qquad |Z_t$$

$$\textcolor{gray}{Z_{t+1} \perp\!\!\!\perp Z_{t-1}, X_{t-1}, \textcolor{red}{X_t}} \qquad \textcolor{gray}{|Z_t}$$

$$\textcolor{gray}{X_{t-1} \perp\!\!\!\perp \textcolor{green}{Z_t}, X_t, Z_{t+1}, X_{t+1}} \qquad \textcolor{gray}{|Z_{t-1}}$$

$$\textcolor{gray}{Z_t \quad \perp\!\!\!\perp \textcolor{green}{X_{t-1}}} \qquad \textcolor{gray}{|Z_{t-1}}$$

Probability terminology
000000000

Bayesian networks
00000●00000000

Gaussian properties
0000

Exercise
0000

Example Bayesian network

$$X_{t+1} \perp\!\!\!\perp Z_{t-1}, X_{t-1}, Z_t, X_t \qquad |Z_{t+1}$$
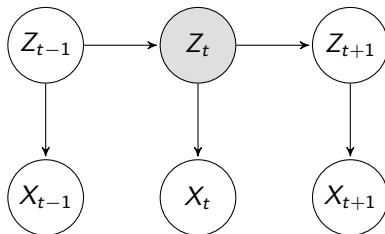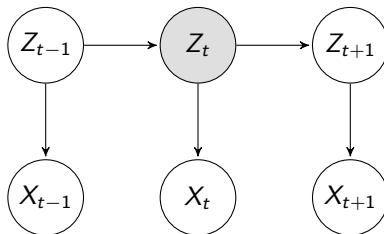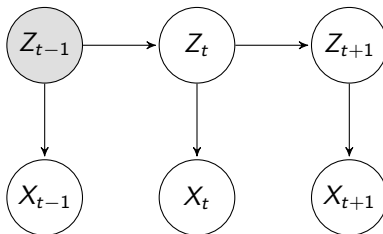$$X_t \quad \perp\!\!\!\perp Z_{t-1}, X_{t-1}, Z_{t+1}, X_{t+1} \quad |Z_t$$
$$Z_{t+1} \perp\!\!\!\perp Z_{t-1}, X_{t-1}, X_t \qquad |Z_t$$
$$X_{t-1} \perp\!\!\!\perp Z_t, X_t, Z_{t+1}, X_{t+1} \qquad |Z_{t-1}$$
$$Z_t \quad \perp\!\!\!\perp X_{t-1} \qquad\qquad\qquad |Z_{t-1}$$

Probability terminology
000000000

Bayesian networks
0000●00000000000

Gaussian properties
0000

Exercise
0000

Example Bayesian network

$$X_{t+1} \perp\!\!\!\perp Z_{t-1}, X_{t-1}, Z_t, X_t \qquad |Z_{t+1}$$

$$X_t \quad \perp\!\!\!\perp Z_{t-1}, X_{t-1}, \textcolor{red}{Z_{t+1}}, X_{t+1} \quad |Z_t$$

$$Z_{t+1} \perp\!\!\!\perp Z_{t-1}, X_{t-1}, \textcolor{red}{X_t} \qquad |Z_t$$

$$X_{t-1} \perp\!\!\!\perp \textcolor{green}{Z_t}, X_t, Z_{t+1}, X_{t+1} \quad |Z_{t-1}$$

$$\textcolor{gray}{Z_t \quad \perp\!\!\!\perp X_{t-1}} \qquad\qquad \textcolor{gray}{|Z_{t-1}}$$

Example Bayesian network

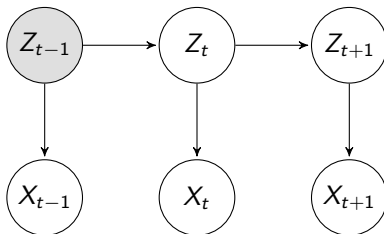$$X_{t+1} \perp\!\!\!\perp Z_{t-1}, X_{t-1}, Z_t, X_t \qquad |Z_{t+1}$$
$$X_t \quad \perp\!\!\!\perp Z_{t-1}, X_{t-1}, Z_{t+1}, X_{t+1} \quad |Z_t$$
$$Z_{t+1} \perp\!\!\!\perp Z_{t-1}, X_{t-1}, X_t \qquad |Z_t$$
$$X_{t-1} \perp\!\!\!\perp Z_t, X_t, Z_{t+1}, X_{t+1} \qquad |Z_{t-1}$$
$$Z_t \quad \perp\!\!\!\perp X_{t-1} \qquad\qquad\qquad\quad |Z_{t-1}$$

Computational costs in Bayesian networks

- ▶ We can take advantage of the structure in Bayesian networks to simplify calculations and reduce the computational cost.
- ▶ The probability $p(X_{t+1}, X_t, X_{t-1})$ (the evidence) is often one of interest.

$$p(X_{t+1}, X_t, X_{t-1}) = \sum_{Z_{t+1}} \sum_{Z_t} \sum_{Z_{t-1}} p(X_{t+1}, X_t, X_{t-1}, Z_{t+1}, Z_t, Z_{t-1})$$

- ▶ Assuming Z can take one of 3 values, and not using the conditional structure of the model, this has a computational cost of 3^3

Using the conditional structure of Bayesian networks

▶ Using the conditional structure, we can rewrite the
  $p(X_{t+1}, X_t, X_{t-1}, Z_{t+1}, Z_t, Z_{t-1})$ as the product of
  conditionals:
  $p(X_{t+1}|Z_{t+1})p(Z_{t+1}|Z_t)p(X_t|Z_t)p(Z_t|Z_{t-1})p(X_{t-1}|Z_{t-1})p(Z_{t-1})$
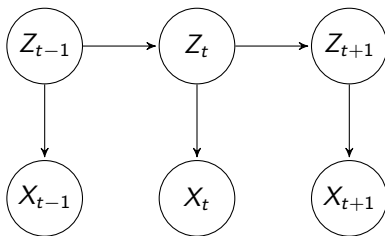
Using the conditional structure of Bayesian networks

▶ Using the conditional structure, we can rewrite the $p(X_{t+1}, X_t, X_{t-1}, Z_{t+1}, Z_t, Z_{t-1})$ as the product of conditionals:
$p(X_{t+1}|Z_{t+1})p(Z_{t+1}|Z_t)p(X_t|Z_t)p(Z_t|Z_{t-1})p(X_{t-1}|Z_{t-1})p(Z_{t-1})$

▶ Moving the terms that don't appear out of each sum, we can simplify the computation

$$
p(X_{t+1}, X_t, X_{t-1}) = \sum_{Z_{t+1}} \Bigg(p(X_{t+1}|Z_{t+1})
$$
$$
\sum_{Z_t} \Bigg(p(Z_{t+1}|Z_t)p(X_t|Z_t)
$$
$$
\sum_{Z_{t-1}} p(Z_t|Z_{t-1})p(X_{t-1}|Z_{t-1})p(Z_{t-1}) \Bigg) \Bigg)
$$

Computational cost of computing the evidence

▶ To compute the inner most sum, we must create a table of Z_{t-1} and Z_t values, and sum over Z_t, this has a cost of 3^2

$p(Z_t \mid Z_{t-1})$	$Z_t = 0$	$Z_t = 1$	$Z_t = 2$
$Z_{t-1} = 0$	$p(Z_t = 0 \mid Z_{t-1} = 0)$	$p(Z_t = 1 \mid Z_{t-1} = 0)$	$p(Z_t = 2 \mid Z_{t-1} = 0)$
$Z_{t-1} = 1$	$p(Z_t = 0 \mid Z_{t-1} = 1)$	$p(Z_t = 1 \mid Z_{t-1} = 1)$	$p(Z_t = 2 \mid Z_{t-1} = 1)$
$Z_{t-1} = 2$	$p(Z_t = 0 \mid Z_{t-1} = 2)$	$p(Z_t = 1 \mid Z_{t-1} = 2)$	$p(Z_t = 2 \mid Z_{t-1} = 2)$

$\times$

$p(X_t \mid Z_{t-1})$	$Z_t = 0$	$Z_t = 1$	$Z_t = 2$
$Z_{t-1} = 0$	$p(X_{t-1} \mid Z_{t-1} = 0)$	$p(X_{t-1} \mid Z_{t-1} = 0)$	$p(X_{t-1} \mid Z_{t-1} = 0)$
$Z_{t-1} = 1$	$p(X_{t-1} \mid Z_{t-1} = 1)$	$p(X_{t-1} \mid Z_{t-1} = 1)$	$p(X_{t-1} \mid Z_{t-1} = 1)$
$Z_{t-1} = 2$	$p(X_{t-1} \mid Z_{t-1} = 2)$	$p(X_{t-1} \mid Z_{t-1} = 2)$	$p(X_{t-1} \mid Z_{t-1} = 2)$

$\times$

$p(Z_{t-1})$	$Z_t = 0$	$Z_t = 1$	$Z_t = 2$
$Z_{t-1} = 0$	$p(Z_{t-1} = 0)$	$p(Z_{t-1} = 0)$	$p(Z_{t-1} = 0)$
$Z_{t-1} = 1$	$p(Z_{t-1} = 1)$	$p(Z_{t-1} = 1)$	$p(Z_{t-1} = 1)$
$Z_{t-1} = 2$	$p(Z_{t-1} = 2)$	$p(Z_{t-1} = 2)$	$p(Z_{t-1} = 2)$

Figure 3: Product of inner sum

Computational cost of computing the evidence

▶ To compute the inner most sum, we must create a table of Z_{t-1} and Z_t values, and sum over Z_t, this has a cost of 3^2

$$=$$

$p(Z_t \mid Z_{t-1})p(X_{t-1} \mid Z_{t-1})p(Z_{t-1})$	$Z_t = 0$	$Z_t = 1$	$Z_t = 2$
$Z_{t-1} = 0$	A	B	C
$Z_{t-1} = 1$	D	E	F
$Z_{t-1} = 2$	G	H	I

Figure 4: Result of product

Computational cost of computing the evidence

▶ To compute the inner most sum, we must create a table of Z_{t-1} and Z_t values, and sum over Z_t, this has a cost of 3^2

$p(Z_t\|Z_{t-1})p(X_{t-1}\|Z_{t-1})p(Z_{t-1})$	$Z_t = 0$	$Z_t = 1$	$Z_t = 2$
$Z_{t-1} = 0$	A	B	C
$Z_{t-1} = 1$	D	E	F
$Z_{t-1} = 2$	G	H	I

$$\sum_{Z_{t-1}} \rightarrow$$

	$Z_t = 0$	$Z_t = 1$	$Z_t = 2$
$\sum_{Z_{t-1}} p(Z_t\|Z_{t-1})p(X_{t-1}\|Z_{t-1})p(Z_{t-1})$	$A+D+G$	$B+E+H$	$C+F+I$

Figure 5: Summing Z_{t-1}

Computational cost of computing the evidence

▶ Now do the same thing for the second sum, where the sum we
computed on the previous slide is equal to $p(X_{t-1}, Z_t)$. The
cost is again 3^2.

Figure 6: Product of second sum

Computational cost of computing the evidence

▶ Now do the same thing for the second sum, where the sum computed two slides ago is equal to $p(X_{t-1}, Z_t)$. The cost is again 3^2.

$p(Z_{t+1}\|Z_t)p(X_t\|Z_t)p(X_{t-1},Z_t)$	$Z_{t+1}=0$	$Z_{t+1}=1$	$Z_{t+1}=2$
$Z_t = 0$	J	K	L
$Z_t = 1$	M	N	O
$Z_t = 2$	P	Q	R

$$\sum_{Z_t} \rightarrow$$

	$Z_{t+1}=0$	$Z_{t+1}=1$	$Z_{t+1}=2$
$\sum_{Z_t} p(Z_{t+1}\|Z_t)p(X_t\|Z_t)p(X_{t-1},Z_t)$	$J+M+P$	$K+N+Q$	$L+O+R$
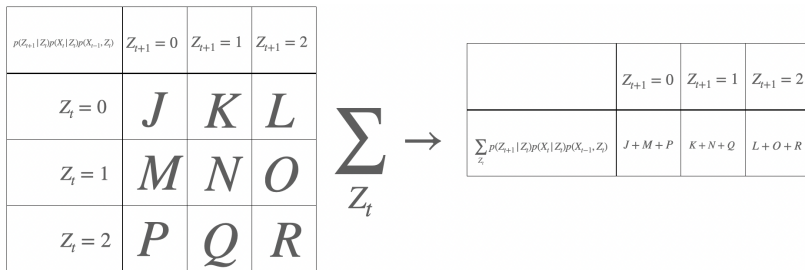
Figure 7: Summing Z_t

Computational cost of computing the evidence

▶ For the final sum, we only have two tables with only one row
  (for Z_{t+1}), so the cost for this will only be 3^1. The sum on
  the previous slide is now $p(X_t, X_{t-1}, Z_{t+1})$

	$Z_{t+1} = 0$	$Z_{t+1} = 1$	$Z_{t+1} = 2$
$p(X_{t+1}\|Z_{t+1})$	$p(X_{t+1}\|Z_{t+1} = 0)$	$p(X_{t+1}\|Z_{t+1} = 1)$	$p(X_{t+1}\|Z_{t+1} = 2)$

$\times$

	$Z_{t+1} = 0$	$Z_{t+1} = 1$	$Z_{t+1} = 2$
$p(X_t, X_{t-1}, Z_{t+1})$	$J + M + P$	$K + N + Q$	$L + O + R$

$=$

	$Z_{t+1} = 0$	$Z_{t+1} = 1$	$Z_{t+1} = 2$
$p(X_{t+1}\|Z_{t+1})$ $p(X_t, X_{t-1}, Z_{t+1})$	S	T	U

Figure 8: Product of final sum

Computational cost of computing the evidence

- As a the final result, we get that
  $\sum_{Z_{t+1}} p(X_{t+1}|Z_{t+1})p(X_t, X_{t-1}, Z_{t+1}) = S + T + U = p(X_{t+1}, X_t, X_{t-1})$
- The total cost of this computation is $2 \times 3^2 + 3^1$

Probability terminology
000000000

Bayesian networks
0000000000000●

Gaussian properties
0000

Exercise
0000

Generalizing the model

- ▶ If instead of only having $\{Z_i, X_i | i \in \{t-1, t, t+1\}\}$, we extend the model to $\{Z_i, X_i | i \in \{1, \ldots, T\}\}$ and we allow the dimensionality of Z_i to be generalized to d, we find that the computational cost is $T \times d^2 + d^1$

- ▶ The cost of computing the evidence naively is d^T

- ▶ We have thus reduced the cost from being exponential in time (where time is defined as the value of T) and polynomial in dimension (d) to linear in time and quadratic in dimension

| Probability terminology | Bayesian networks | Gaussian properties | Exercise |
| 000000000 | 00000000000000 | 0●00 | 0000 |

Gaussian identities

- ▶ Gaussians are the preferred distributions data scientists work with due to their nice properties
- ▶ The two most important identities are that the sum and product of two Gaussians is another Gaussian
- ▶ These give rise to the fact that the marginal of a Gaussian over two variables is another Gaussian
- ▶ This means inference can be done exactly with analytical solutions since the normalization constants can be computed exactly.
- ▶ https://cs.nyu.edu/∼roweis/notes/gaussid.pdf (replace the ∼ with the one from your keyboard for link to work)

Two important identities (1/2)

Let $\mathbf{z}$:
$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^\top & \mathbf{B} \end{bmatrix} \right).$$

Then
$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}\left(\mathbf{a} + \mathbf{C}\mathbf{B}^{-1}(\mathbf{y} - \mathbf{b}), \mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^\top \right)$$
$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}\left(\mathbf{b} + \mathbf{C}^\top\mathbf{A}^{-1}(\mathbf{x} - \mathbf{a}), \mathbf{B} - \mathbf{C}^\top\mathbf{A}^{-1}\mathbf{C} \right)$$

Two important identities (2/2)

Let $\mathbf{x}, \mathbf{y}$:

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$$
$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}).$$

Then,

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{\top})$$
$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\Sigma}\left[\mathbf{A}^{\top}\mathbf{L}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\right], \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma} = \left(\boldsymbol{\Lambda} + \mathbf{A}^{\top}\mathbf{L}\mathbf{A}\right)^{-1}.$$

Table of Contents

Probability terminology
○○○○○○○○○

Bayesian networks
○○○○○○○○○○○○○○○

Gaussian properties
○○○○

Exercise
○●○○

From Bishop (Chapter 8)
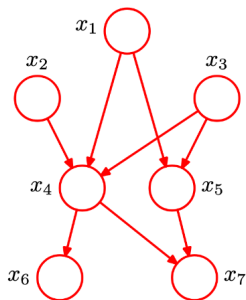
1. Markov blanket of x_5?
2. Joint distribution ?

Probability terminology
○○○○○○○○○
Bayesian networks
○○○○○○○○○○○○○○○
Gaussian properties
○○○○
Exercise
○●○○

From Bishop (Chapter 8)

1. Markov blanket of x_5? $\{x_7, x_1, x_3, x_4\}$
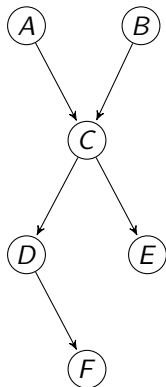2. Joint distribution ?

From Bishop (Chapter 8)
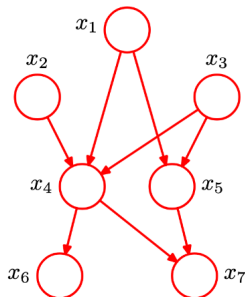
1. Markov blanket of x_5? $\{x_7, x_1, x_3, x_4\}$
2. Joint distribution ?
   $p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$
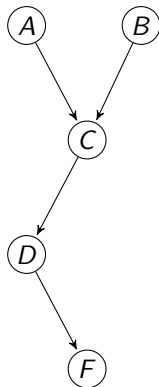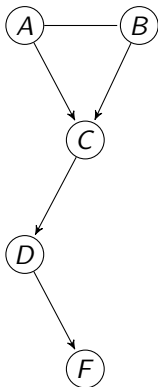
From http://web.mit.edu/jmn/www/6.034/d-separation.pdf

1. Are A and B conditionally independent given D and F ?
2. Are D and E conditionally independent given C?
3. Are D and E conditionally independent given A and B ?
4. $P(D|CEF) \stackrel{?}{=} P(D|C)$

From http://web.mit.edu/jmn/www/6.034/d-separation.pdf

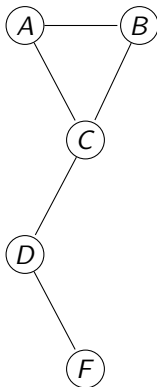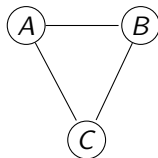Are A and B conditionally independent given D and F ? **No.**

(a) Ancestral graph

(b) Moralize

(c) Disorient

(d) Delete givens