Probabilistic Graphical Models

Lecture 4

Conditional Independence

Previous topics

1928
K. N. Toosi University of Technology

- Probabilistic Models
  - Generative Model: joint distribution $p(x_1, x_2, \ldots, x_m)$
  - Representation, Prediction (inference), Learning, Sampling
- Marginal Distribution

$$p(m) = \Pr(M = m) = \sum_n p(m, n)$$

- Conditional Distribution

$$\Pr(N = n_0 \mid M = m) = \frac{\Pr(N{=}n_0, M{=}m)}{\sum_n \Pr(N{=}n, M{=}m)} = \frac{\Pr(N{=}n_0, M{=}m)}{\Pr(M{=}m)}$$

Previous topics

1928

K. N. Toosi University of Technology

- Probabilistic Independence (meaning?)
    - $Pr(M = m \mid N = n) = Pr(M = m)$
    - $P(M = m, N = n) = Pr(N = n)\ Pr(M = m)$

- Independence reduces complexity
    - $p(x_1, x_2, \dots, x_m) = p(x_1)\ p(x_2) \dots p(x_m)$

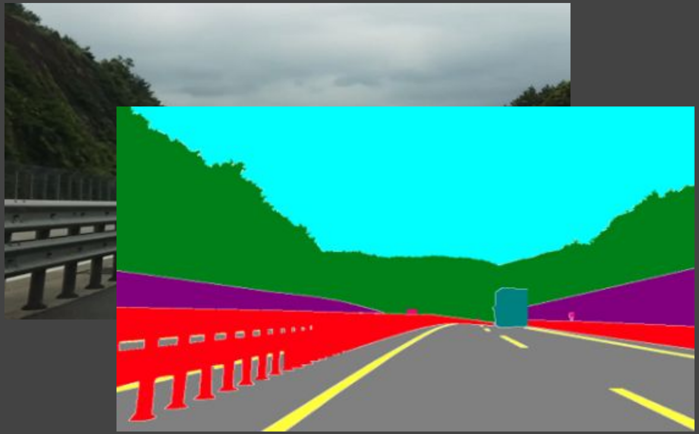
More than two variables

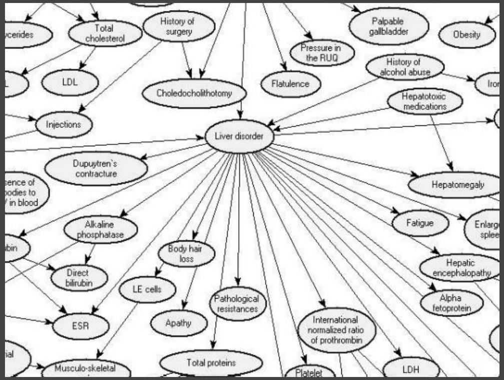
1928
K. N. Toosi University of Technology

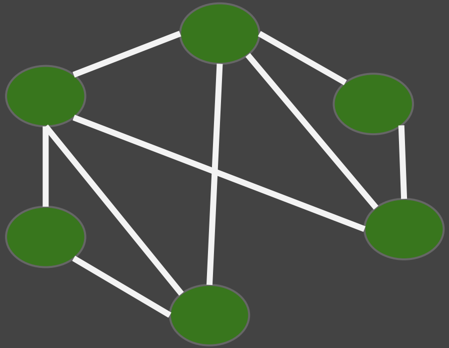
- $p(x_1, x_2, x_3, \ldots, x_m)$
- Pairwise independence
  - Every pair of variables x_i, x_j are independent
- Mutual Independence
  - $p(x_i$ | any **subset** of other variables $) = p(x_i)$
  - $p(x_1, x_2, \ldots, x_m) = p(x_1)\, p(x_2) \ldots p(x_m)$

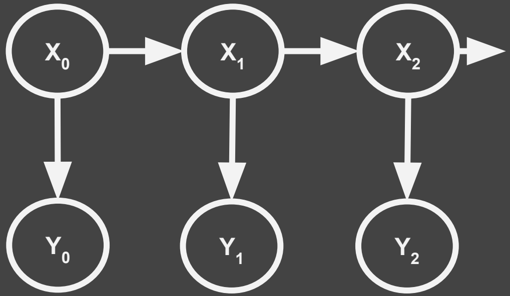
Using structure

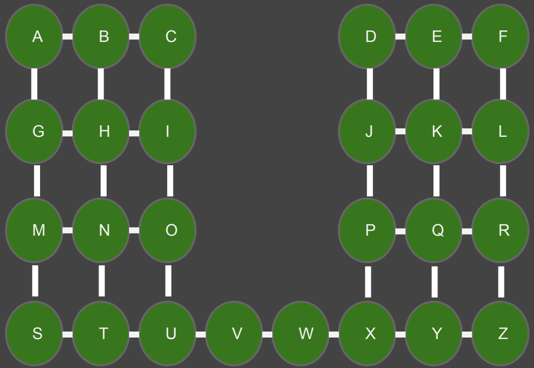
1928
K. N. Toosi University of Technology



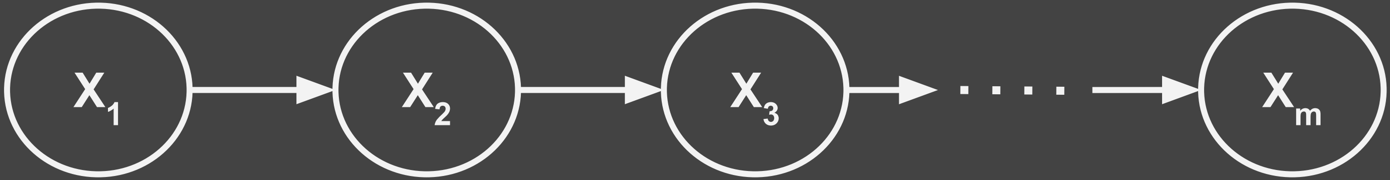






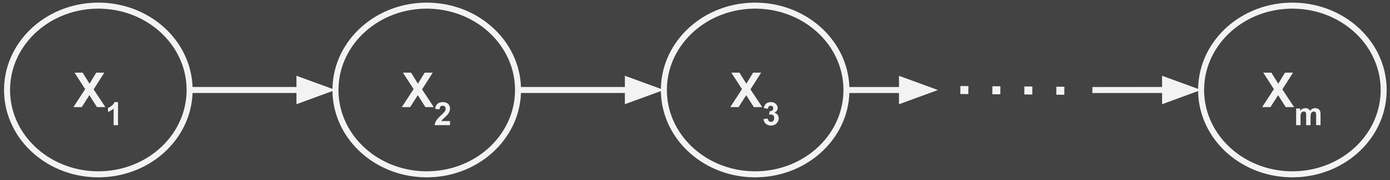


Example: faulty push-button



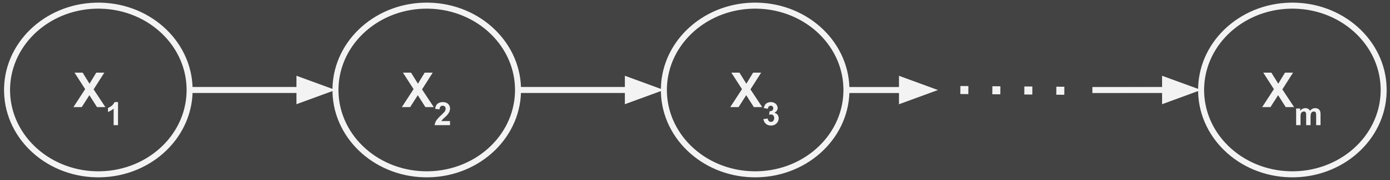
- x_n device on or off (1/0) after n times pressing the button
- button works with probability p

Example: faulty push-button



- x_n device on or off (1/0) after n times pressing the button
- button works with probability p if device is on, and
  with probability q if device is off

Example: faulty push-button

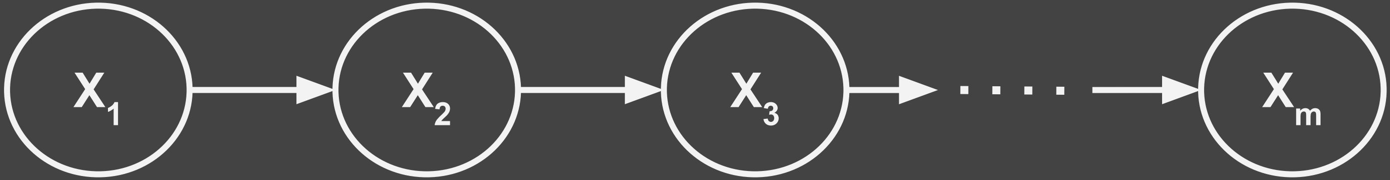


- x_n device on or off (1/0) after n times pressing the button
- button works with probability p_t if device is on, and
  with probability q_t if device is off

$p(x_1, x_2, \dots, x_m)$ $\quad x_i \in 0, 1$ $\quad$ (2^m-1 parameters)

$p(x_1, x_2, \dots, x_m) = p(x_1) \cdots p(x_m)$ $\quad$ (m parameters)

Example: faulty push-button



- x_n device on or off (1/0) after n times pressing the button
- button works with probability p_t if device is on, and
  with probability q_t if device is off

Example

- Are "Having a cloudy morning" and "getting wet" dependent?
- $P(W \mid C) = P(W)$?

Example

- Are "Having a cloudy morning" and "getting wet" dependent?
- $P(W \mid C) \neq P(W)$

Example

- Knowing that we had a rainfall
  - Are "Having a cloudy morning" and "getting wet" dependent?

Conditional Independence

- Knowing that we had a rainfall
  - Are "Having a cloudy morning" and "getting wet" dependent?
  - $P(W \mid R, C) = P(W \mid R)$

Conditional Independence

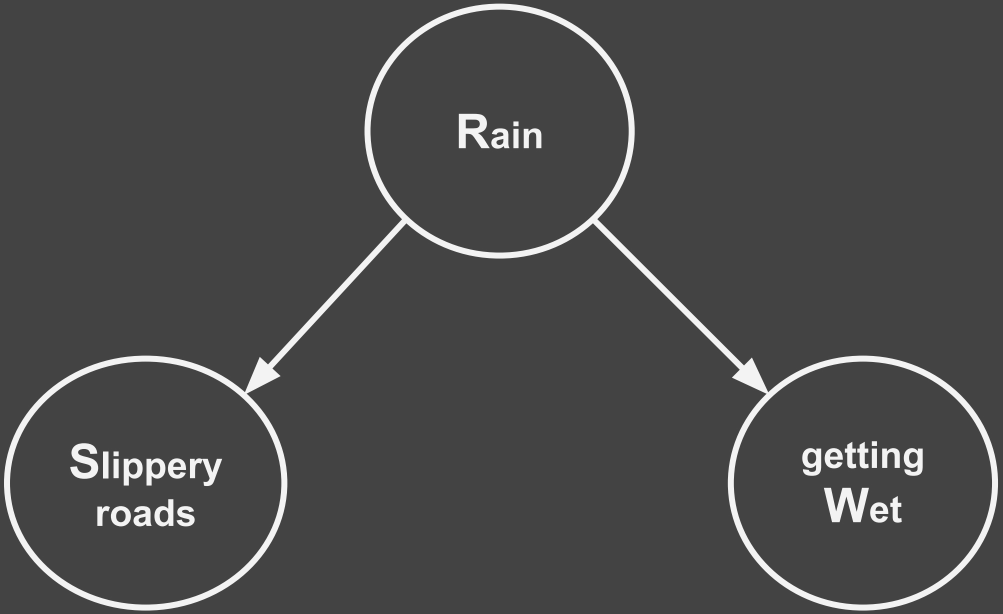
- Knowing that we had a rainfall
  - Are "Having a cloudy morning" and "getting wet" dependent?
  - $P(W \mid R, C) = P(W \mid R)$
  - $P(W, C \mid R) = P(W \mid R)\ P(C \mid R)$

Conditional Independence

- Knowing that we had a rainfall
  - Are "Having a cloudy morning" and "getting wet" dependent?
  - $P(W \mid R, C) = P(W \mid R)$
  - $P(W, C \mid R) = P(W \mid R)\ P(C \mid R)$
- W and C are conditionally independent given R

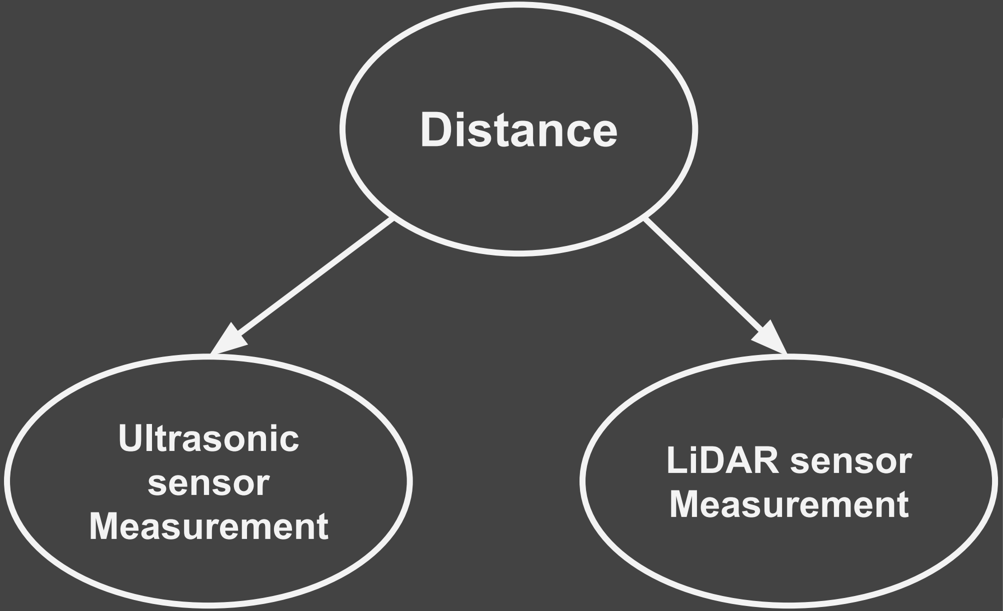
Conditional Independence

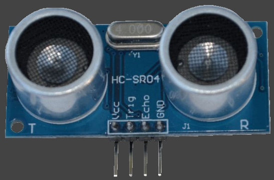
K. N. Toosi
University of Technology



Conditional Independence

K. N. Toosi
University of Technology

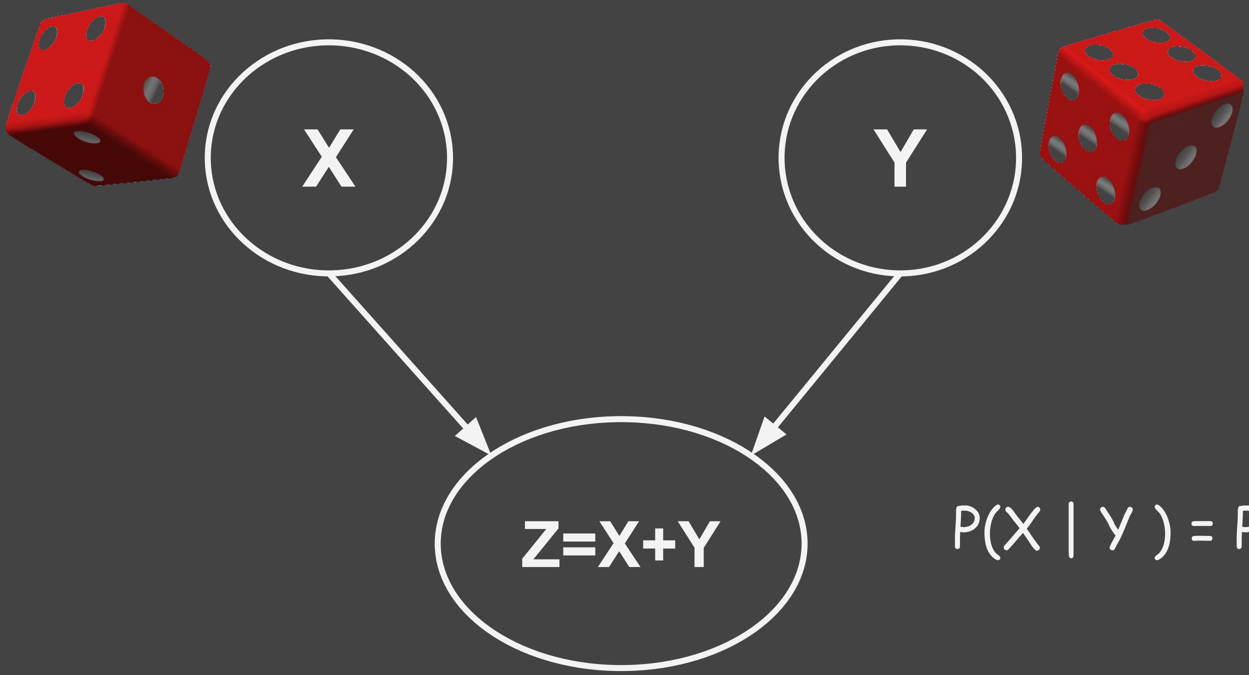




Conditioning can destroy independence

$$P(X \mid Y) = P(X)$$

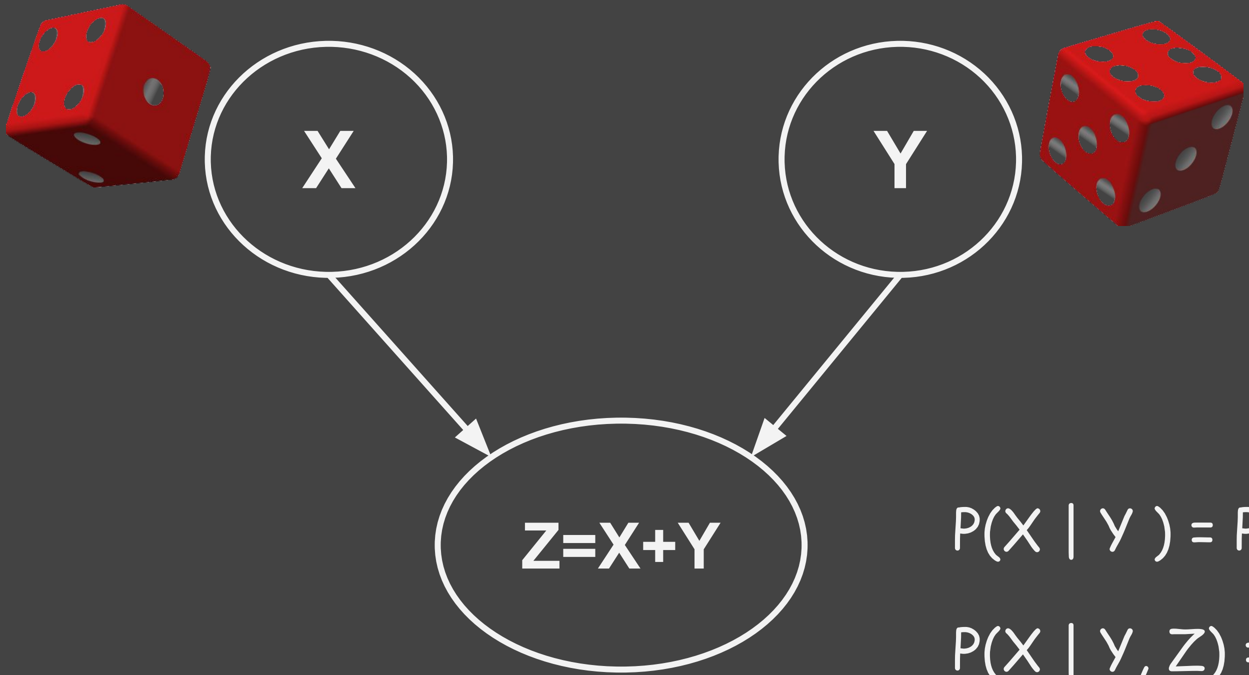
Conditioning can destroy independence



$$P(X \mid Y) = P(X)$$

Conditioning can destroy independence

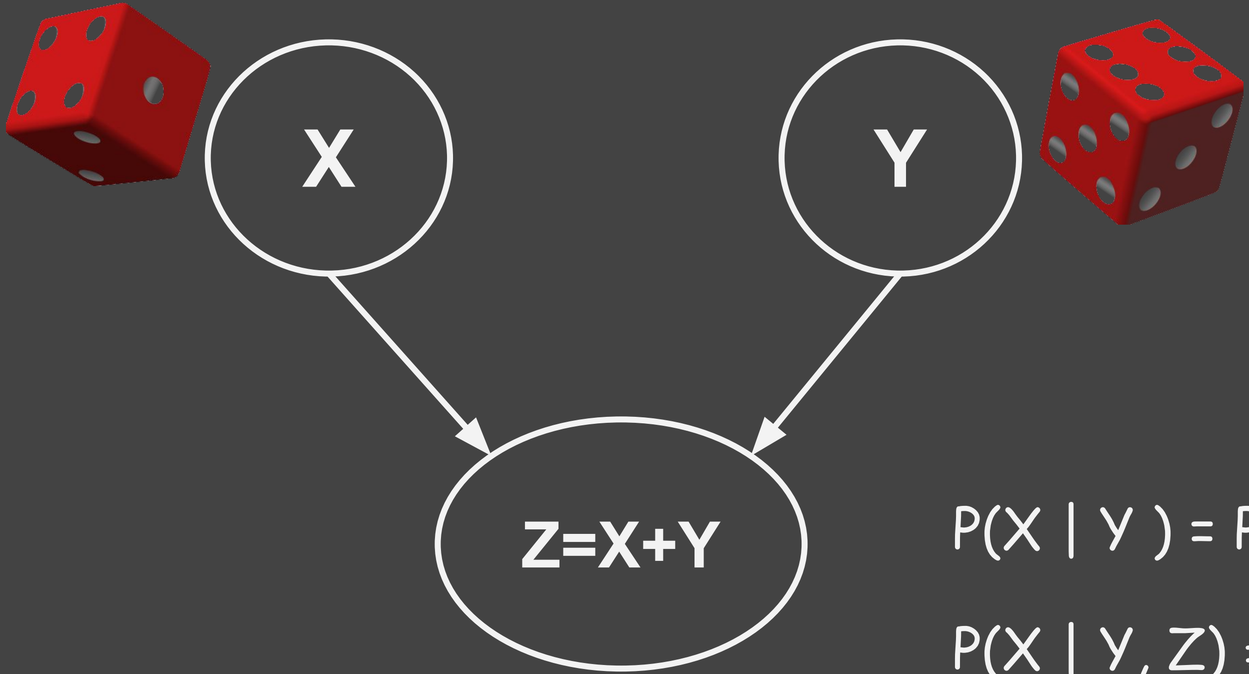
K. N. Toosi
University of Technology



$$P(X \mid Y) = P(X)$$

$$P(X \mid Y, Z) = P(X \mid Z)\ ?$$

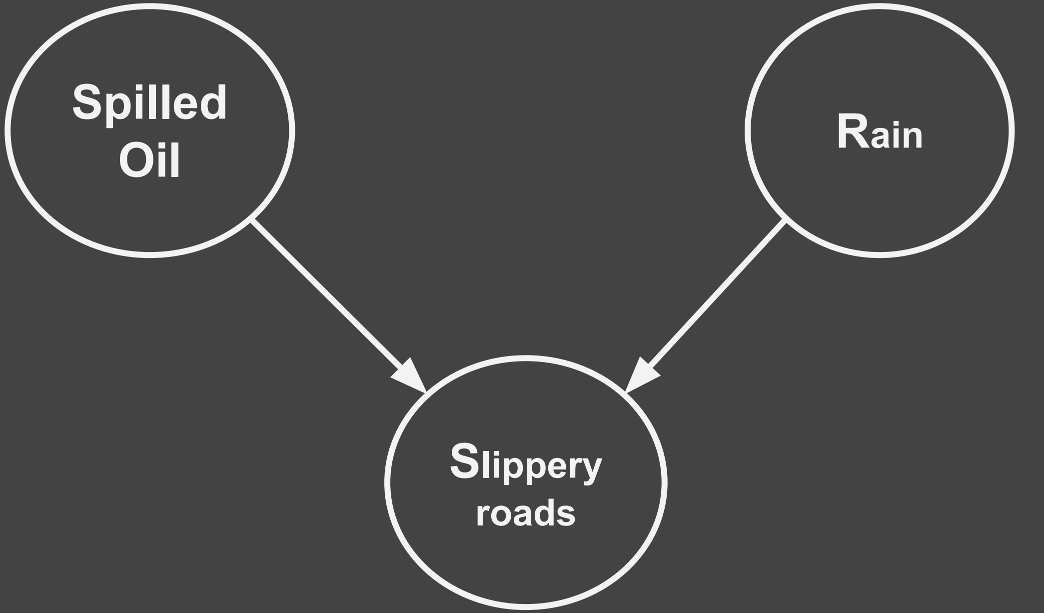
Conditioning can destroy independence



$$P(X \mid Y) = P(X)$$

$$P(X \mid Y, Z) \neq P(X \mid Z)$$

Conditioning can destroy independence



Conditional independence reduce complexity?

- Tabular representation P(C,R,W)
- General case: how many independent parameters in general?

Conditional independence reduce complexity?

- Tabular representation P(C,R,W)
- General case: how many independent parameters in general? 7

Conditional independence reduce complexity?

- Tabular representation P(C,R,W)
- General case: how many independent parameters in general? 7
- Fully independent case: P(C,R,W) = P(C) P(R) P(W)
  - How many parameters?

Conditional independence reduce complexity?

- Tabular representation P(C,R,W)
- General case: how many independent parameters in general? 7
- Fully independent case: P(C,R,W) = P(C) P(R) P(W)
    - How many parameters? 3

Conditional independence reduce complexity?

- Tabular representation P(C,R,W)
- General case: how many independent parameters in general? 7
- Fully independent case: P(C,R,W) = P(C) P(R) P(W)
  - How many parameters? 3
- Conditionally independent: P(C,R,W): P(W | R, C) = P(W | R)
  - How many parameters?

Conditional independence reduce complexity?

1928
K. N. Toosi University of Technology

- Conditionally independent: P(C,R,W) = P(W | R, C) = P(W | R)
    - How many parameters?
    - P(C,R,W) = P(W | C, R) P(C,R) = P(W | R) P(C,R)

Conditional independence reduce complexity?

1928
K. N. Toosi University of Technology

- Conditionally independent: P(C,R,W) = P(W | R, C) = P(W | R)
    - How many parameters?
    - P(C,R,W) = P(W | C, R) P(C,R) = P(W | R) P(C,R)
    - P(W | R):
    - P(C,R):

Conditional independence reduce complexity?

- Conditionally independent: P(C,R,W) = P(W | R, C) = P(W | R)
  - How many parameters?
  - P(C,R,W) = P(W | C, R) P(C,R) = P(W | R) P(C,R)
  - P(W | R): P(W = 0| R = 0), P(W = 0| R = 1) : 2 parameters
  - P(C,R):

Conditional independence reduce complexity?

K. N. Toosi University of Technology

- Conditionally independent: P(C,R,W) = P(W | R, C) = P(W | R)
  - How many parameters?
  - P(C,R,W) = P(W | C, R) P(C,R) = P(W | R) P(C,R)
  - P(W | R): P(W = 0| R = 0), P(W = 0| R = 1) : 2 parameters
  - P(C,R): 3 parameters

Conditional independence reduce complexity?

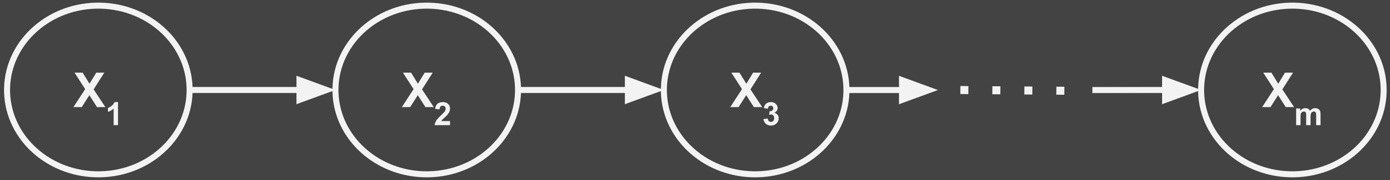
- Conditionally independent: P(C,R,W) = P(W | R, C) = P(W | R)
  - How many parameters?
  - P(C,R,W) = P(W | C, R) P(C,R) = P(W | C) P(C,R)
  - P(W | R): P(W = 0| R = 0), P(W = 0| R = 1) : 2 parameters
  - P(C,R): 3 parameters
  - P(C,R,W) = P(W | C, R) P(C,R) : 5 parameters

Conditional independence reduce complexity?

1928
K. N. Toosi University of Technology

- Tabular representation P(C,R,W)
- General case: 7 parameters
- Fully independent case: 3 parameters
- Conditionally independent: 5 parameters

Example: faulty push-button



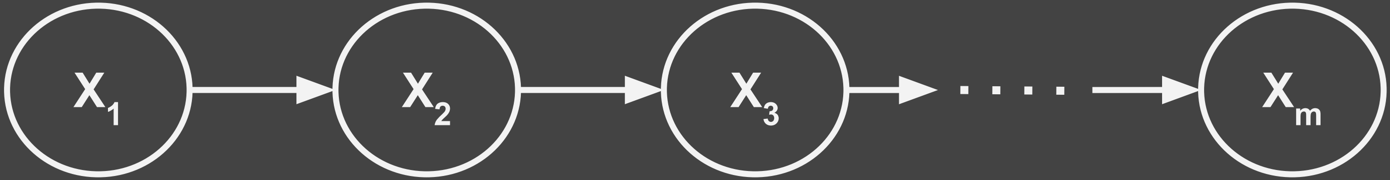
- x_n device on or off (1/0) after n times pressing the button
- button works with probability p_t if device is on, and
  with probability q_t if device is off

$p(x_1, x_2, \ldots, x_m)$ $\quad x_i \in 0, 1$ $\quad$ (2^m-1 parameters)

$p(x_1, x_2, \ldots, x_m) = p(x_1) \cdots p(x_m)$ $\quad$ (m parameters)

Example: faulty push-button

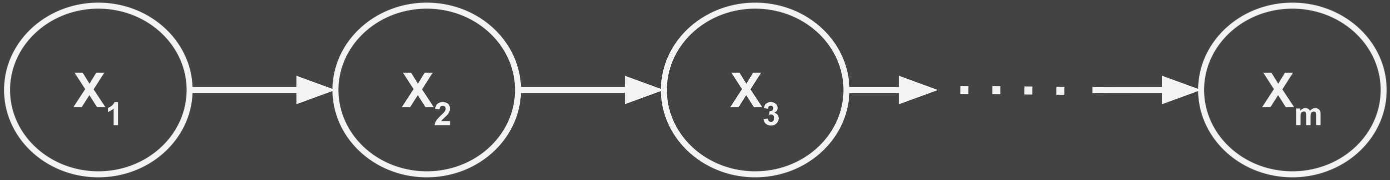
K. N. Toosi
University of Technology



Observation:

$$p(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_2, x_1) = ?$$

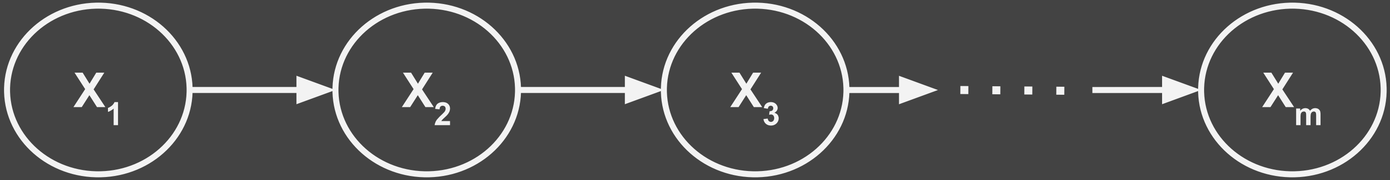
Example: faulty push-button



Observation:

$$p(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_2, x_1) = p(x_t \mid x_{t-1})$$

Example: faulty push-button

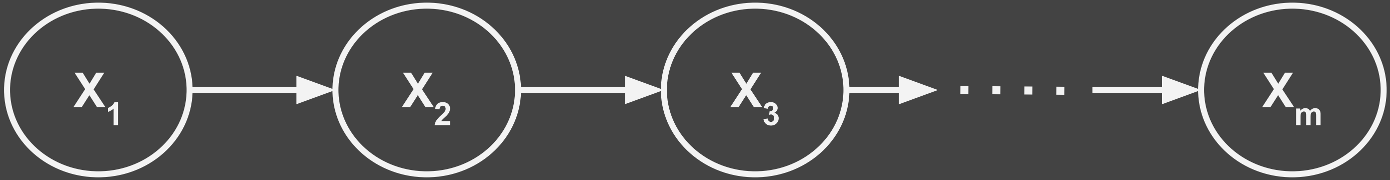


Observation:

$$p(x_t \mid x_{t-1}, x_{t-2}, \dots, x_2, x_1) = p(x_t \mid x_{t-1})$$

this is is meaning of arrows → (directional edges) in the above graph (more on this later)

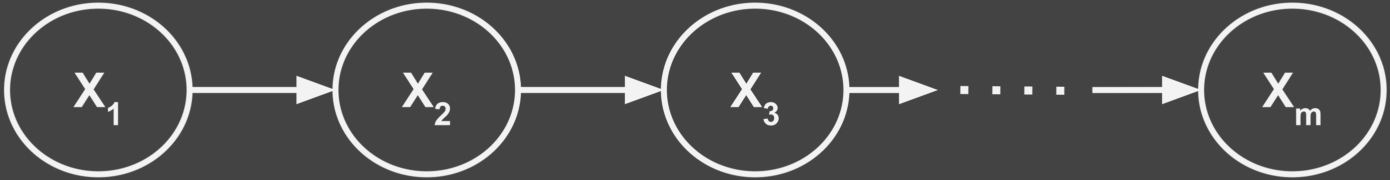
Example: faulty push-button



Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_2, x_1) = p(x_t \mid x_{t-1})$

$$p(x_1, x_2, \ldots, x_m) = p(x_m \mid x_1, x_2, \ldots, x_{m-1})\, p(x_1, x_2, \ldots, x_{m-1})$$

Example: faulty push-button

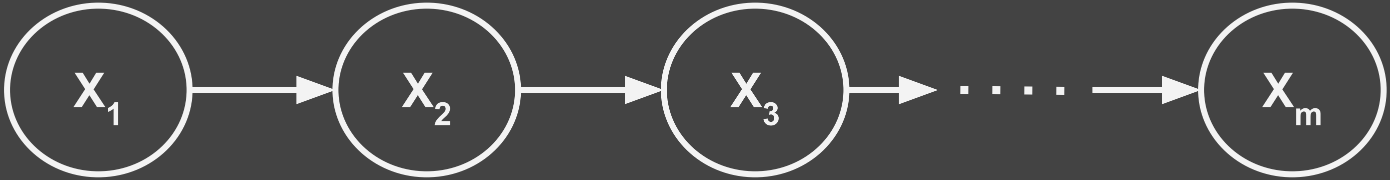


Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_2, x_1) = p(x_t \mid x_{t-1})$

$$p(x_1, x_2, \ldots, x_m) = p(x_m \mid x_1, x_2, \ldots, x_{m-1})\, p(x_1, x_2, \ldots, x_{m-1})$$

$$= p(x_m \mid x_{m-1})\, p(x_1, x_2, \ldots, x_{m-1})$$

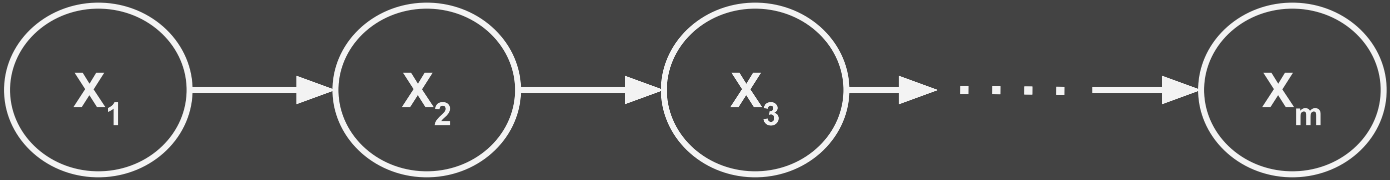
Example: faulty push-button



Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_2, x_1) = p(x_t \mid x_{t-1})$

$$
\begin{aligned}
p(x_1, x_2, \ldots, x_m) &= p(x_m \mid x_1, x_2, \ldots, x_{m-1})\, p(x_1, x_2, \ldots, x_{m-1}) \\
&= p(x_m \mid x_{m-1})\, p(x_1, x_2, \ldots, x_{m-1}) \\
&= p(x_m \mid x_{m-1})\, p(x_{m-1} \mid x_1, x_2, \ldots, x_{m-2})\, p(x_1, x_2, \ldots, x_{m-2})
\end{aligned}
$$

Example: faulty push-button



Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_2, x_1) = p(x_t \mid x_{t-1})$

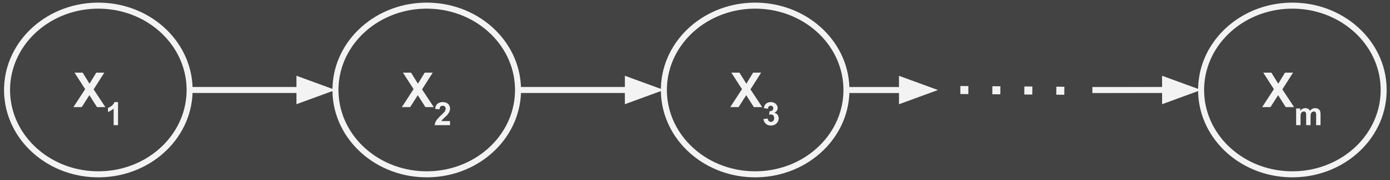
$$p(x_1, x_2, \ldots, x_m) = \ldots = p(x_m \mid x_{m-1})\, p(x_1, x_2, \ldots, x_{m-1})$$

$$= p(x_m \mid x_{m-1})\, p(x_{m-1} \mid x_1, x_2, \ldots, x_{m-2})\, p(x_1, x_2, \ldots, x_{m-2})$$

$$= p(x_m \mid x_{m-1})\, p(x_{m-1} \mid x_{m-2})\, p(x_1, x_2, \ldots, x_{m-2})$$

Example: faulty push-button

K. N. Toosi
University of Technology

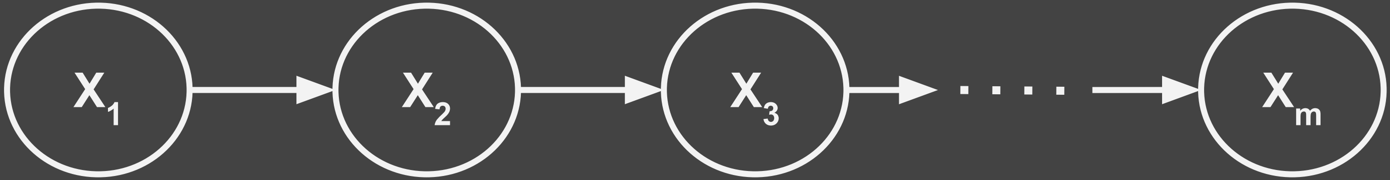


Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_2, x_1) = p(x_t \mid x_{t-1})$

$$p(x_1, x_2, \ldots, x_m) = \ldots$$

$$= p(x_m \mid x_{m-1})\, p(x_{m-1} \mid x_{m-2}) \ldots p(x_4 \mid x_3)\;\; p(x_3 \mid x_2)\, p(x_2, x_1)$$

Example: faulty push-button



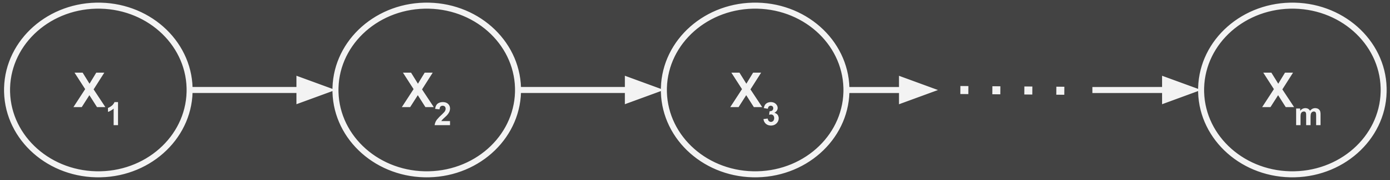
Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \dots, x_2, x_1) = p(x_t \mid x_{t-1})$

$p(x_1, x_2, \dots, x_m) = \dots$

$= p(x_m \mid x_{m-1})\, p(x_{m-1} \mid x_{m-2}) \dots p(x_4 \mid x_3)\ \ p(x_3 \mid x_2)\, p(x_2, x_1)$

$= p(x_m \mid x_{m-1})\, p(x_{m-1} \mid x_{m-2}) \dots p(x_4 \mid x_3)\ \ p(x_3 \mid x_2)\, p(x_2 \mid x_1)$
$p(x_1)$

Example: faulty push-button

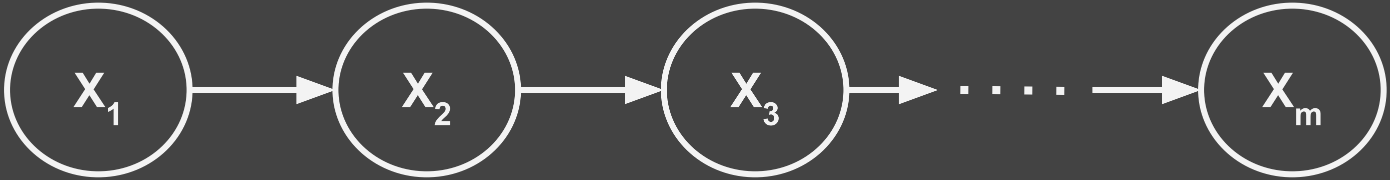


Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_2, x_1) = p(x_t \mid x_{t-1})$

$$p(x_1, x_2, \ldots, x_m) = \ldots$$

$$= p(x_1)\ p(x_2|x_1)\ p(x_3|x_2)\ \ldots\ p(x_{m-1}|x_{m-2})\ p(x_m|x_{m-1})$$

Example: faulty push-button



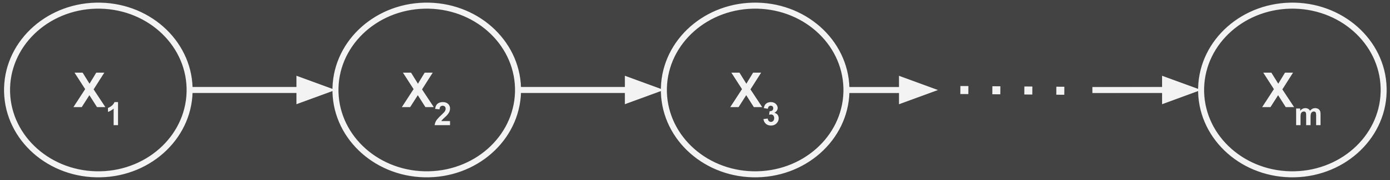
Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_2, x_1) = p(x_t \mid x_{t-1})$

$p(x_1, x_2, \ldots, x_m) = \ldots$

$= p(x_1) \; p(x_2|x_1) \; p(x_3|x_2) \; \ldots \; p(x_{m-1}|x_{m-2}) \; p(x_m|x_{m-1})$

How many parameters?

Example: faulty push-button



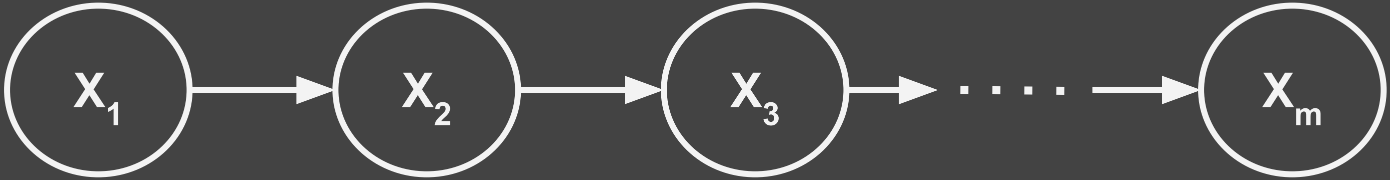
$p(x_1, x_2, \ldots, x_m) = \ldots$

$= p(x_1) \quad p(x_2 | x_1) \quad p(x_3 | x_2) \quad \ldots \quad p(x_{m-1} | x_{m-2}) \quad p(x_m | x_{m-1})$

$p(x_1)$:

$p(x_t | x_{t-1})$:

Example: faulty push-button



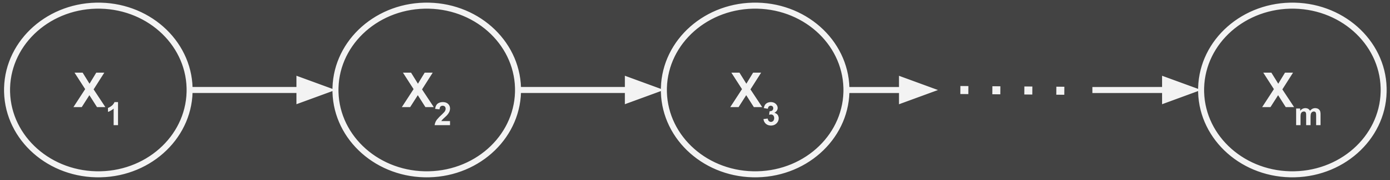
$p(x_1, x_2, \ldots, x_m) = \ldots$

$= p(x_1) \;\; p(x_2|x_1) \;\; p(x_3|x_2) \; \ldots \; p(x_{m-1}|x_{m-2}) \;\; p(x_m|x_{m-1})$

$p(x_1)$: 1 parameter

$p(x_t|x_{t-1})$:

Example: faulty push-button



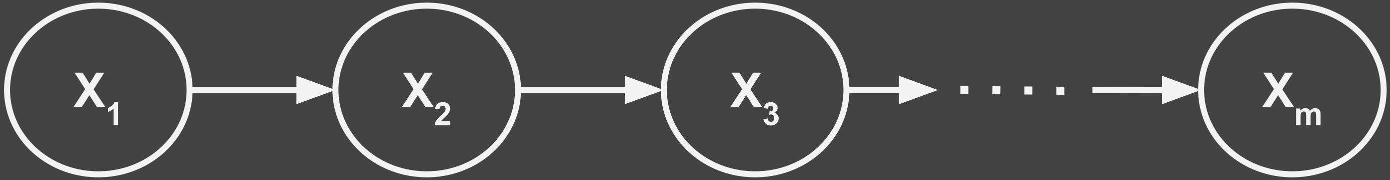
$p(x_1, x_2, \ldots, x_m) = \ldots$

$= p(x_1) \;\; p(x_2|x_1) \;\; p(x_3|x_2) \;\; \ldots \;\; p(x_{m-1}|x_{m-2}) \;\; p(x_m|x_{m-1})$

$p(x_1)$: 1 parameter

$p(x_t|x_{t-1})$: 2 parameters

Example: faulty push-button



$p(x_1, x_2, \dots, x_m) = \dots$

$= p(x_1) \;\; p(x_2|x_1) \;\; p(x_3|x_2) \;\dots\; p(x_{m-1}|x_{m-2}) \;\; p(x_m|x_{m-1})$

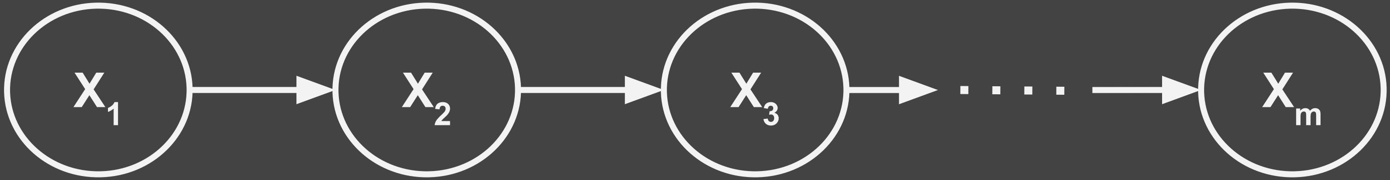
$p(x_1)$: 1 parameter

$p(x_t|x_{t-1})$: 2 parameters: p_t, q_t

$q_t = p(x_t = 1|x_{t-1} = 0)$

$p_t = p(x_t = 0|x_{t-1} = 1)$

Example: faulty push-button



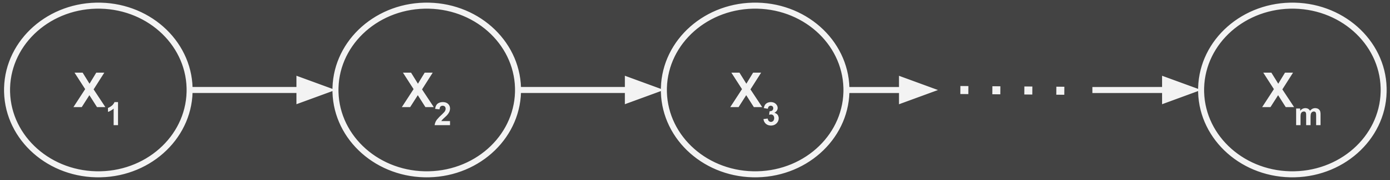
$p(x_1, x_2, \ldots, x_m) = \ldots$

$= p(x_1) \;\; p(x_2|x_1) \;\; p(x_3\,|x_2) \;\; \ldots \;\; p(x_{m-1}\,|x_{m-2}) \;\; p(x_m\,|x_{m-1})$

$p(x_1)$: 1 parameter, $p(x_t|x_{t-1})$: 2 parameters

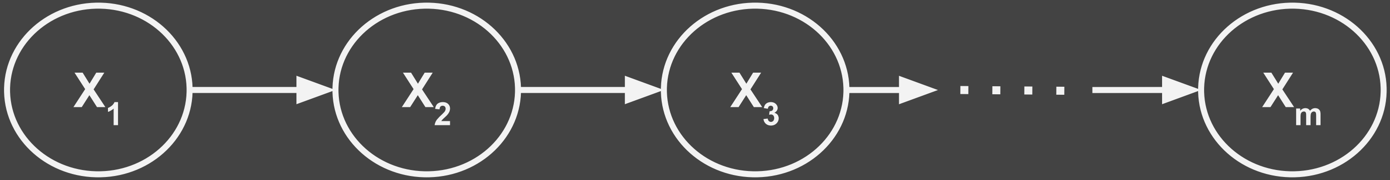
total: 2m-1 free parameters

Example: faulty push-button



- fully dependent: 2^m-1 free parameters (about 10^{30} for n=100)
- fully independent: m free parameters (100 for n=100)
  - $p(x_1, x_2, ..., x_m) = p(x_1)\ p(x_2)\ ...\ p(x_m)$
- conditionally independent: 2m-1 free parameters (199 for n=100)
  - $p(x_1, x_2, ..., x_m) = p(x_1)\ p(x_2 | x_1)\ ...\ p(x_m | x_{m-1})$

Example: faulty push-button



$p(x_1, x_2, \dots, x_m) = \dots$

$= p(x_1) \;\; p(x_2|x_1) \;\; p(x_3|x_2) \;\; \dots \;\; p(x_{m-1}|x_{m-2}) \;\; p(x_m|x_{m-1})$

total: 2m-1 free parameters

if $p_t = q_t$: m parameters

if $p_t = q_t = p$: 2 parameters (parameter sharing)