Probabilistic Graphical Models

Lecture 3

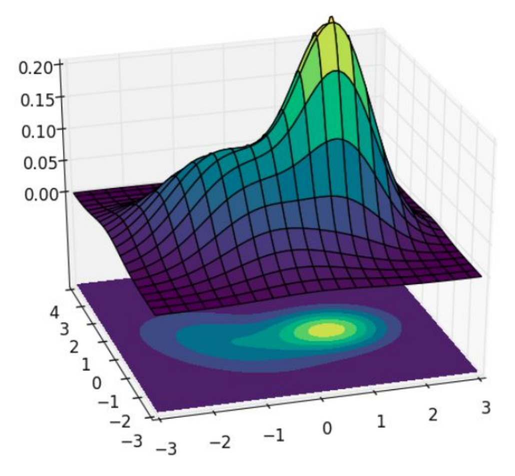
Independence Reduces Complexity

Remember: multivariate prediction

Remember: The joint probability distribution

K. N. Toosi
University of Technology

- The probability of co-occurrence.
- Probability mass function (Discrete Variables)
    - p(x,y) = Pr(X=x AND Y=y) = Pr(X=x, Y=y)
- Probability Density function (Continuous Variables)
    - p(x,y)



Remember: Probabilistic Modelling

K. N. Toosi
University of Technology

- System variables $X_1, X_2, \dots, X_N$
- Generative Model: Joint distribution $p(x_1, x_2, \dots, x_N)$
- If you have the joint distribution, you have everything

- Prediction:
  - Having $p(x,y,z) = Pr(X=x, Y=y, Z=z)$, predict x,y,z
  - Find the most likely configuration of system variables

$$x^*, y^*, z^* = \arg\max_{x,y,z} p(x, y, z)$$

Remember: Probabilistic Modelling

K. N. Toosi University of Technology

- System variables $X_1, X_2, \dots, X_N$
- Generative Model: Joint distribution $p(x_1, x_2, \dots, x_N)$
- If you have the joint distribution, you have everything

- Prediction:
  - Having p(x,y,z) = Pr(X=x, Y=y, Z=z)
  - If we know Z = z_0, predict x,y

$$x, y = \arg\max_{x,y} p(x, y, z_0)$$

Probabilistic Modelling

1. learning/modeling:
   - find $p(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)$
2. prediction/testing

$$y_1^*, y_2^*, \ldots, y_n^* = \arg\max_{y_1,\ldots,y_n} p(x_1, \ldots, x_m, y_1, \ldots, y_n)$$

$y_i \in \{0,1\}$, $n = 100$

- argmax over $2^{100} \approx 10^{30}$ different combinations of $y_1, y_2, \ldots, y_n$
- 10^{12} = 1000G iterations per second $\rightarrow$ 32 billion years!

Model Representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- How to store in computer?
- How much storage is needed?

Tabular representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- m = 3 ---> 8 entries
-

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

K. N. Toosi
University of Technology

Tabular representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- m = 3 ---> 8 entries, 7 parameters
-

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

K. N. Toosi
University of Technology

Tabular representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- m = 3 ---> 8 entries, 7 parameters
- 2^m entries
-

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

K. N. Toosi
University of Technology

Tabular representation

$p(x_1, x_2, \ldots, x_m)$ $\qquad x_i \in 0, 1$

- m = 3 ---> 8 entries, 7 parameters
- 2^m entries
- m = 100 --->

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

K. N. Toosi
University of Technology

Tabular representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- m = 3 ---> 8 entries, 7 parameters
- 2^m entries
- m = 100 ---> 2^{100} entries ≈ 10^{21} G entries

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

K. N. Toosi
University of Technology

Tabular representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- m = 3 ---> 8 entries, 7 parameters
- 2^m entries
- m = 100 ---> 2^{100} entries ≈ 10^{21} G entries
- 1 byte per entry -> 10^{21} GB of RAM!

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

K. N. Toosi University of Technology

Example:

The joint probability of

- **having a rainfall in an hour**, and
- **the sky being cloudy at the moment**
    - p(r,c) = Pr(R = r, C = c)

r (rain)	c (cloudy)	$Pr(R = r, C = c)$
0	0	0.75
0	1	0.10
1	0	0.05
1	1	0.10

Example:

The joint probability of

- **having a rainfall in an hour** and
- **the sky being cloudy at the moment**
  - $p(r,c) = Pr(R = r, C = c)$

r (rain)	c (cloudy)	$Pr(R = r, C = c)$
0	0	0.75
0	1	0.10
1	0	0.05
1	1	0.10

$Pr(R{=}r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

Question

- Having the joint distribution $Pr(R = r, C = c)$, what is $Pr(R = r)$?

$$Pr(R = r) = ?$$

Question

- Having the joint distribution $Pr(R = r, C = c)$, what is $Pr(R = r)$?

$$Pr(R = r) = Pr\left((R = r \text{ AND } C = 0) \text{ OR } (R = r \text{ AND } C = 1) \right)$$

K. N. Toosi
University of Technology

Question

- Having the joint distribution $Pr(R = r, C = c)$, what is $Pr(R = r)$?

$$Pr(R = r) = Pr\left((R = r \text{ AND } C = 0) \text{ OR } (R = r \text{ AND } C = 1) \right)$$

$$= Pr(R = r \text{ AND } C = 0) + Pr(R = r \text{ AND } C = 1) \qquad \text{(why?)}$$

Question

- Having the joint distribution $Pr(R = r, C = c)$, what is $Pr(R = r)$?

$$Pr(R = r) = Pr\,(\,(R = r \text{ AND } C = 0) \text{ OR } (R = r \text{ AND } C = 1)\,)$$

$$= Pr(R = r \text{ AND } C = 0) + Pr(R = r \text{ AND } C = 1) \qquad \text{(why?)}$$

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$Pr(R = r)$

R=0	R=1
0.85	0.15

Marginal Distribution

- Discrete: probability mass function p(m,n) = Pr(M=m, N=n)

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

- Continuous: probability density function p(x,y)

$$p(x) = \int p(x, y)\, dy$$

Marginal Probability

K. N. Toosi
University of Technology

$P(x, y)$	$x = 0$	$x = 1$	$x = 2$	row sum
$y = 0$	0.32	0.03	0.01	**0.36**
$y = 1$	0.06	0.24	0.02	**0.32**
$y = 2$	0.02	0.03	0.27	**0.32**
col sum	**0.40**	**0.30**	**0.30**	checksum = 1.0

image from http://stats.stackexchange.com

Marginal Probability

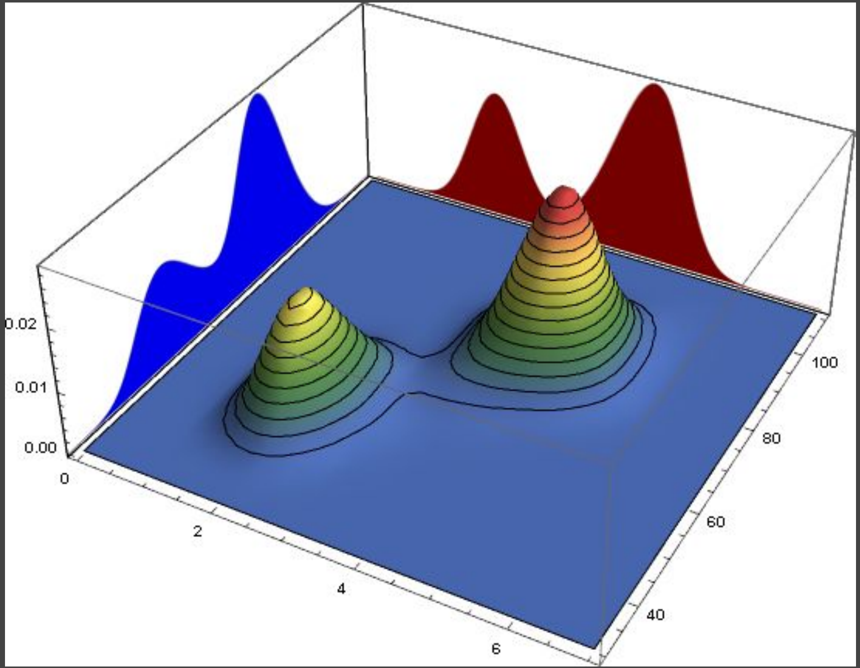


image from www.wolfram.com

Question

- What is the probability of having a rainfall today?

$$\Pr(R = r, C = c)$$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

K. N. Toosi
University of Technology

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1)$$

$\Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$Pr(R = r)$

R=0	R=1
0.85	0.15

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

- If we know the sky is cloudy, what is the probability of having a rainfall today?

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

- If we know the sky is cloudy, what is the probability of having a rainfall today?

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$$Pr(R = 1 \mid C = 1) =$$

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

- If we know the sky is cloudy, what is the probability of having a rainfall today?

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$$Pr(R = 1 \mid C = 1) =$$

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

- If we know the sky is cloudy, what is the probability of having a rainfall today?

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$$Pr(R = 1 \mid C = 1) = 0.10 / (0.10 + 0.10) = 0.5$$

$$= \frac{\Pr(R=1,C=1)}{\Pr(R=1,C=1)+\Pr(R=0,C=1)}$$

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

- If we know the sky is cloudy, what is the probability of having a rainfall today?

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$$Pr(R = 1 \mid C = 1) = 0.10 / (0.10 + 0.10) = 0.5$$

$$= \frac{\Pr(R{=}1, C{=}1)}{\Pr(R{=}1, C{=}1) + \Pr(R{=}0, C{=}1)}$$

$$= \frac{\Pr(R{=}1, C{=}1)}{\Pr(C{=}1)}$$

Question

K. N. Toosi
University of Technology

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = \boxed{0.15}$$

- If we know the sky is cloudy, what is the probability of having a rainfall today?

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$$Pr(R = 1 \mid C = 1) = 0.10 \,/\, (0.10 + 0.10) = \boxed{0.5}$$

$$= \frac{\Pr(R=1, C=1)}{\Pr(R=1, C=1) + \Pr(R=0, C=1)}$$

$$= \frac{\Pr(R=1, C=1)}{\Pr(C=1)}$$

Conditional Distribution

- Discrete: joint PMF p(m,n) = Pr(M=m, N=n)

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

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

- Continuous: joint PDF p(x,y)

Conditional Distribution

- Discrete: joint PMF $p(m,n) = Pr(M=m, N=n)$

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

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

- Continuous: joint PDF $p(x,y)$

$$p(y \mid x) = \frac{p(x,y)}{\int p(x,y)\, dy} = \frac{p(x,y)}{p(x)}$$

Question

- Pr(rain in 1hr) = .15
- Pr(rain in 1hr | cloudy now) = .5

- Pr(rain in 1hr) = .15
- Pr(rain in 1hr | I failed the PGM exam) = ?

Question

- Pr(rain in 1hr) = .15
- Pr(rain in 1hr | cloudy now) = .5

- Pr(rain in 1hr) = .15
- Pr(rain in 1hr | I failed the PGM exam) = .15

Probabilistic Independence

- $\Pr(M = m \mid N = n) = \Pr(M = m)$ for all m, n

K. N. Toosi
University of Technology

Probabilistic Independence

- $Pr(M = m \mid N = n) = Pr(M = m)$ for all m, n

  $P(M = m, N = n) / Pr(N = n) = Pr(M = m)$

Probabilistic Independence

- $Pr(M = m \mid N = n) = Pr(M = m)$ for all m, n

  $P(M = m, N = n) / Pr(N = n) = Pr(M = m)$

  $\Rightarrow P(M = m, N = n) = Pr(N = n) \; Pr(M = m)$

Probabilistic Independence

K. N. Toosi University of Technology

- $Pr(M = m \mid N = n) = Pr(M = m)$ for all m, n

  $P(M = m, N = n) / Pr(N = n) = Pr(M = m)$

  $\Rightarrow P(M = m, N = n) = Pr(N = n)\ Pr(M = m)$

- What does independence mean?
  - does "having a rainfall" depend on "people using umbrellas"?

Probabilistic Independence

K. N. Toosi
University of Technology

- $Pr(M = m \mid N = n) = Pr(M = m)$ for all m, n

  $P(M = m, N = n) / Pr(N = n) = Pr(M = m)$

  $\Rightarrow P(M = m, N = n) = Pr(N = n)\ Pr(M = m)$

- Continuous case:

  $$p(y \mid x) = p(y) \quad \Rightarrow \quad p(x,y) = p(x)\, p(y)$$

More than two variables

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

testing independence

$$p(x_1, x_2, x_3) \qquad x_i \in \{0, 1\}$$

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.06
0	0	1	0.04
0	1	0	0.24
0	1	1	0.16
1	0	0	0.06
1	0	1	0.04
1	1	0	0.24
1	1	1	0.16

K. N. Toosi
University of Technology

testing independence

K. N. Toosi
University of Technology

$$p(x_1, x_2, x_3) \qquad x_i \in \{0, 1\}$$

x_1	$p(x_1)$
0	?
1	?

x_2	$p(x_2)$
0	?
1	?

x_3	$p(x_3)$
0	?
1	?

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.06
0	0	1	0.04
0	1	0	0.24
0	1	1	0.16
1	0	0	0.06
1	0	1	0.04
1	1	0	0.24
1	1	1	0.16

testing independence

$$p(x_1, x_2, x_3) \qquad x_i \in \{0, 1\}$$

x_1	$p(x_1)$
0	?
1	?

x_2	$p(x_2)$
0	?
1	?

x_3	$p(x_3)$
0	?
1	?

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.06
0	0	1	0.04
0	1	0	0.24
0	1	1	0.16
1	0	0	0.06
1	0	1	0.04
1	1	0	0.24
1	1	1	0.16

testing independence

$$p(x_1, x_2, x_3) \qquad x_i \in \{0, 1\}$$

x_1	$p(x_1)$
0	0.5
1	?

x_2	$p(x_2)$
0	?
1	?

x_3	$p(x_3)$
0	?
1	?

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.06
0	0	1	0.04
0	1	0	0.24
0	1	1	0.16
1	0	0	0.06
1	0	1	0.04
1	1	0	0.24
1	1	1	0.16

testing independence

$$p(x_1, x_2, x_3) \qquad x_i \in \{0, 1\}$$

x_1	$p(x_1)$
0	0.5
1	0.5

x_2	$p(x_2)$
0	0.2
1	0.8

x_3	$p(x_3)$
0	0.6
1	0.4

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.06
0	0	1	0.04
0	1	0	0.24
0	1	1	0.16
1	0	0	0.06
1	0	1	0.04
1	1	0	0.24
1	1	1	0.16

K. N. Toosi
University of Technology

testing independence

K. N. Toosi
University of Technology

$$p(x_1, x_2, x_3) \qquad x_i \in \{0, 1\}$$

$$p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3)$$

x_1	$p(x_1)$
0	0.5
1	0.5

x_2	$p(x_2)$
0	0.2
1	0.8

x_3	$p(x_3)$
0	0.6
1	0.4

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.06
0	0	1	0.04
0	1	0	0.24
0	1	1	0.16
1	0	0	0.06
1	0	1	0.04
1	1	0	0.24
1	1	1	0.16

testing independence

$$p(x_1, x_2, x_3) \qquad x_i \in \{0, 1\}$$

$$p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3)$$

how many parameters?

x_1	$p(x_1)$
0	0.5
1	0.5

x_2	$p(x_2)$
0	0.2
1	0.8

x_3	$p(x_3)$
0	0.6
1	0.4

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.06
0	0	1	0.04
0	1	0	0.24
0	1	1	0.16
1	0	0	0.06
1	0	1	0.04
1	1	0	0.24
1	1	1	0.16

K. N. Toosi University of Technology

Representation

K. N. Toosi University of Technology

$$p(x_1, x_2, x_3) \qquad x_i \in \{0, 1\}$$

$$p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3)$$

how many parameters?

x_1	$p(x_1)$
0	0.5
1	0.5

x_2	$p(x_2)$
0	0.2
1	0.8

x_3	$p(x_3)$
0	0.6
1	0.4

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.06
0	0	1	0.04
0	1	0	0.24
0	1	1	0.16
1	0	0	0.06
1	0	1	0.04
1	1	0	0.24
1	1	1	0.16

Inference

$$p(x_1, x_2, x_3) \qquad x_i \in \{0, 1\}$$

$$p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3)$$

$$\max_{x_1, x_2, x_3} p(x_1, x_2, x_3)$$

$$= \max_{x_1, x_2, x_3} p(x_1)p(x_2)p(x_3)$$

$$= \max_{x_1} p(x_1) \max_{x_2} p(x_2) \max_{x_3} p(x_3)$$

x_1	$p(x_1)$
0	0.5
1	0.5

x_2	$p(x_2)$
0	0.2
1	0.8

x_3	$p(x_3)$
0	0.6
1	0.4

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.06
0	0	1	0.04
0	1	0	0.24
0	1	1	0.16
1	0	0	0.06
1	0	1	0.04
1	1	0	0.24
1	1	1	0.16

Inference

$$p(x_1, x_2, x_3) \qquad x_i \in \{0, 1\}$$

$$p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3)$$

$$\max_{x_1,x_2,x_3} p(x_1, x_2, x_3)$$

$$= \max_{x_1,x_2,x_3} p(x_1)p(x_2)p(x_3)$$

$$= \max_{x_1} p(x_1) \max_{x_2} p(x_2) \max_{x_3} p(x_3)$$

x_1	$p(x_1)$
0	0.5
1	0.5

x_2	$p(x_2)$
0	0.2
1	0.8

x_3	$p(x_3)$
0	0.6
1	0.4

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.06
0	0	1	0.04
0	1	0	0.24
0	1	1	0.16
1	0	0	0.06
1	0	1	0.04
1	1	0	0.24
1	1	1	0.16

Tabular representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- m = 3 ---> 8 entries, 7 parameters
- 2^m entries
- m = 100 ---> 2^{100} entries ≈ 10^{21} G entries
- 1 byte per entry -> 10^{21} GB of RAM!

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

- How many independent parameters?

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

Tabular representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- m = 3 ---> 8 entries, 7 parameters
- 2^m entries
- m = 100 ---> 2^{100} entries ≈ 10^{21} G entries
- 1 byte per entry -> 10^{21} GB of RAM!

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

- How many independent parameters? m

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

K. N. Toosi
University of Technology

How to do inference?

K. N. Toosi
University of Technology