Graphical models and inference

Kristjan Kalm *kristjan.kalm@mrc-cbu.cam.ac.uk*

Medical Research Council, Cognition & Brain Sciences Unit

October 22, 2013

Overview

- Multivariate probability distributions

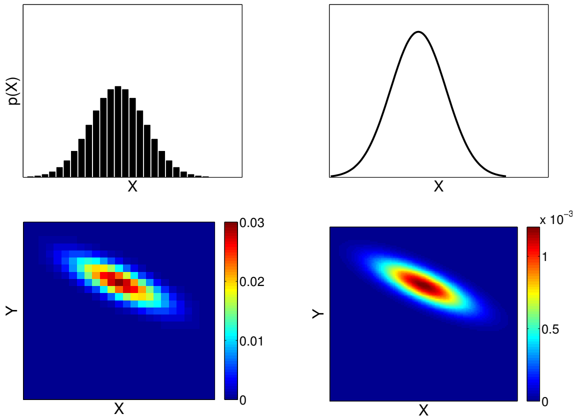
Overview

- Multivariate probability distributions
- Bayes Nets

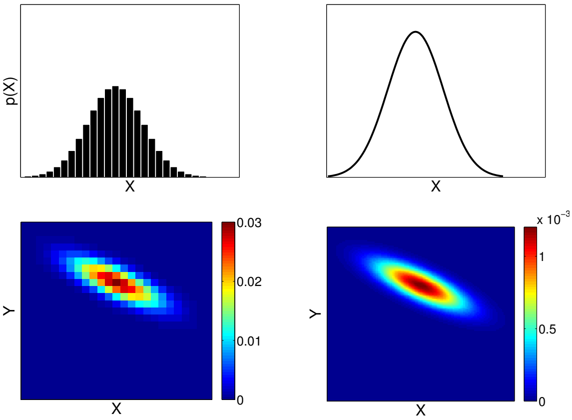
Overview

- Multivariate probability distributions
- Bayes Nets
- Complex graphical models

Multivariate probability distributions



Multivariate probability distributions



Joint probability $p(X, Y)$

Conditional probability $p(X|Y)$

Multivariate probability distributions

Two types of distributions

Joint probability	$p(X, Y)$
Conditional probability	$p(X|Y)$

Multivariate probability distributions

Two types of distributions

Joint probability	$p(X, Y)$
Conditional probability	$p(X\|Y)$

Are X and Y **independent**?

Multivariate probability distributions

Two types of distributions

Joint probability	$p(X, Y)$
Conditional probability	$p(X \mid Y)$

Are X and Y **independent**?

Independent *iff* $p(X, Y) = p(X)p(Y)$

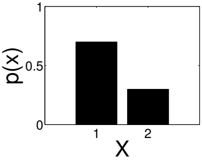
Multivariate probability distributions

Two types of distributions

Joint probability	$p(X, Y)$
Conditional probability	$p(X\|Y)$

Are X and Y **independent**?

Independent *iff* $p(X, Y) = p(X)p(Y)$



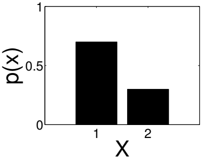
Multivariate probability distributions

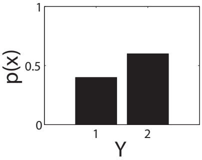
Two types of distributions

Joint probability	$p(X, Y)$
Conditional probability	$p(X\|Y)$

Are X and Y **independent**?

Independent *iff* $p(X, Y) = p(X)p(Y)$





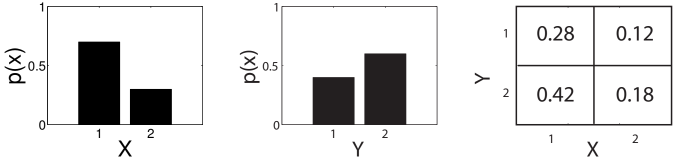
Multivariate probability distributions

Two types of distributions

Joint probability	$p(X, Y)$
Conditional probability	$p(X\|Y)$

Are X and Y **independent**?

Independent *iff* $p(X, Y) = p(X)p(Y)$



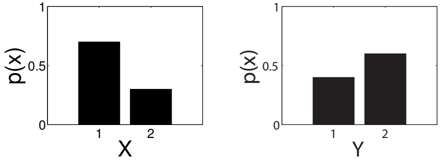
Multivariate probability distributions

Two types of distributions

Joint probability	$p(X, Y)$
Conditional probability	$p(X \mid Y)$

Are X and Y **independent**?

Independent *iff* $p(X, Y) = p(X)p(Y)$



Y \ X	1	2
1	0.28	0.12
2	0.42	0.18

Y \ X	1	2
1	0	0.22
2	0.78	0

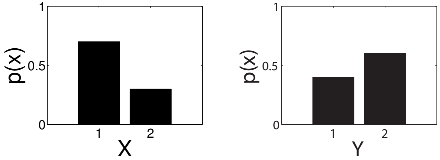
Multivariate probability distributions

Two types of distributions

Joint probability	$p(X, Y)$
Conditional probability	$p(X\|Y)$

Are X and Y **independent**?

Independent *iff* $p(X, Y) = p(X)p(Y)$



Y \ X	1	2
1	0.28	0.12
2	0.42	0.18

Y \ X	1	2
1	0	0.22
2	0.78	0

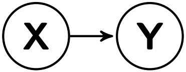
Y depends on X	$p(Y\|X)$
X depends on Y	$p(X\|Y)$

Graph notation

$p(X)$

Graph notation

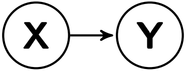
$p(X)$



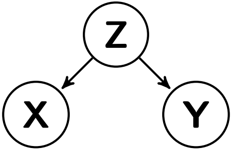
$p(X)$
$p(Y)$
$p(Y|X)$
$p(Y,X)$

Graph notation

$p(X)$



$p(X)$
$p(Y)$
$p(Y|X)$
$p(Y, X)$



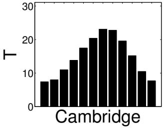
$p(X)$
$p(Y)$
$p(Z)$
$p(X|Z)$
$p(Y|Z)$
$p(Y, X|Z)$
$p(Y, X, Z)$

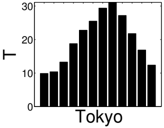
Establishing dependence

Weather in Cambridge and Tokyo

Establishing dependence

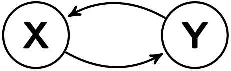
Weather in Cambridge and Tokyo

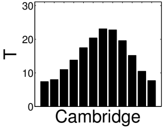


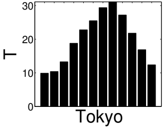


Establishing dependence

Weather in Cambridge and Tokyo

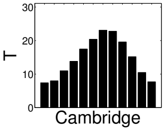


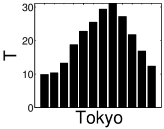


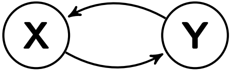


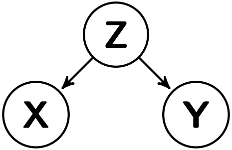
Establishing dependence

Weather in Cambridge and Tokyo



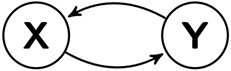


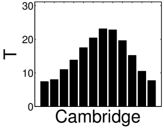


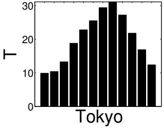


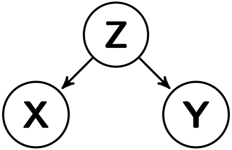
Establishing dependence

Weather in Cambridge and Tokyo





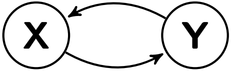


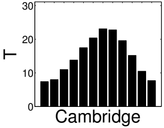


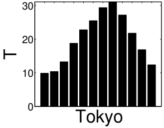
X	t Cambridge
Y	t Tokyo
Z	Month of the year

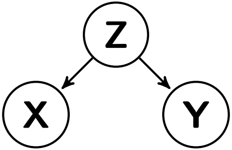
Establishing dependence

Weather in Cambridge and Tokyo







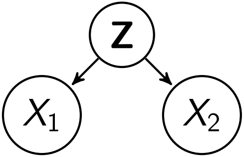


X	t Cambridge
Y	t Tokyo
Z	Month of the year

X and Y are **conditionally independent** *iff*

$$p(X, Y|Z) = p(X|Z)p(Y|Z)$$

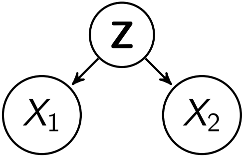
Conditional independence



X *t* Cambridge
Y *t* Tokyo
Z Month of the year

$p(X_1, X_2|Z) = p(X_1|Z)p(X_2|Z)$

Conditional independence

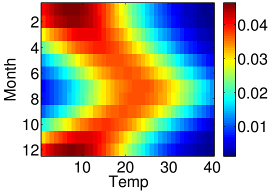


X $\quad$ t Cambridge

Y $\quad$ t Tokyo

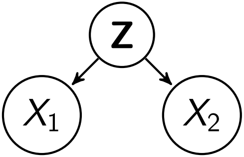
Z $\quad$ Month of the year

$$p(X_1, X_2|Z) = p(X_1|Z)p(X_2|Z)$$



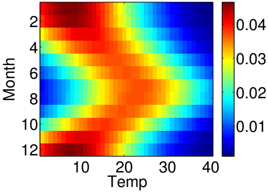
$p(X_1 = Camb|Z)\times$

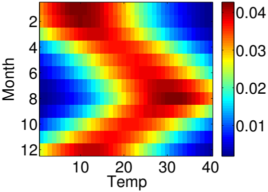
Conditional independence



X t Cambridge
Y t Tokyo
Z Month of the year

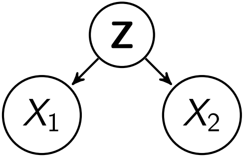
$$p(X_1, X_2|Z) = p(X_1|Z)p(X_2|Z)$$





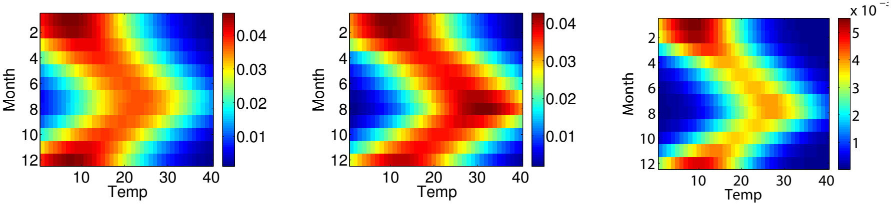
$p(X_1 = Camb|Z) \times$ $p(X_2 = Tokyo|Z)$

Conditional independence



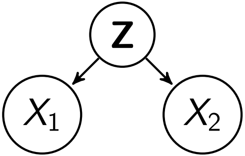
X t Cambridge
Y t Tokyo
Z   Month of the year

$$p(X_1, X_2|Z) = p(X_1|Z)p(X_2|Z)$$



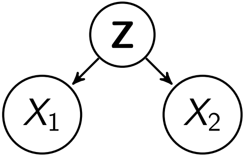
$p(X_1 = Camb|Z)\times$ $p(X_2 = Tokyo|Z)$ $= p(X_1, X_2|Z)$

Conditional independence

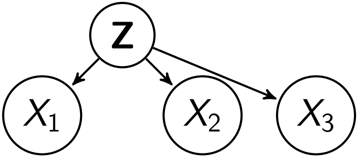


$$p(X_1, X_2|Z) = p(X_1|Z)p(X_2|Z)$$

Conditional independence

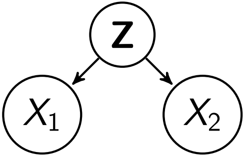


$$p(X_1, X_2|Z) = p(X_1|Z)p(X_2|Z)$$



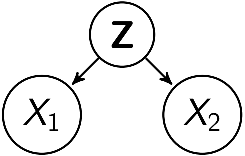
$$p(X_1, ..., X_n|Z) = p(X_1|Z)\cdot, ..., \cdot p(X_n|Z)$$

Conditional independence



$$p(X_1, ..., X_n|Z) = p(X_1|Z) \cdot, ..., \cdot p(X_n|Z)$$

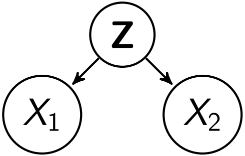
Conditional independence



$p(X_1, ..., X_n|Z) = p(X_1|Z) \cdot, ..., \cdot p(X_n|Z)$

since $p(X_1|X_2, Z) = p(X_1|Z)$ and $p(X_2|X_1, Z) = p(X_2|Z)$

Conditional independence

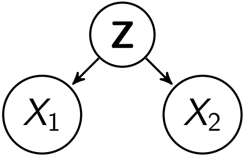


$p(X_1, ..., X_n|Z) = p(X_1|Z) \cdot, ..., \cdot p(X_n|Z)$

since $p(X_1|X_2, Z) = p(X_1|Z)$ and $p(X_2|X_1, Z) = p(X_2|Z)$

$p(X_1, X_2, Z) = p(X_1|Z) \cdot p(X_2|Z) \cdot p(Z)$

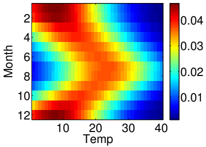
Conditional independence



$$p(X_1, ..., X_n|Z) = p(X_1|Z) \cdot, ..., \cdot p(X_n|Z)$$

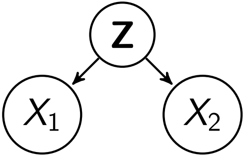
since $p(X_1|X_2, Z) = p(X_1|Z)$ and $p(X_2|X_1, Z) = p(X_2|Z)$

$$p(X_1, X_2, Z) = p(X_1|Z) \cdot p(X_2|Z) \cdot p(Z)$$



$p(X_1 = Camb|Z)$

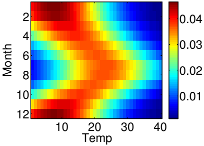
Conditional independence



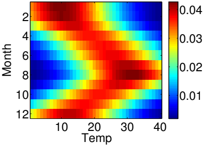
$$p(X_1, ..., X_n|Z) = p(X_1|Z) \cdot, ..., \cdot p(X_n|Z)$$

since $p(X_1|X_2, Z) = p(X_1|Z)$ and $p(X_2|X_1, Z) = p(X_2|Z)$

$$p(X_1, X_2, Z) = p(X_1|Z) \cdot p(X_2|Z) \cdot p(Z)$$

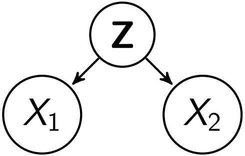


$p(X_1 = Camb|Z)$



$\times p(X_2 = Tokyo|Z)$

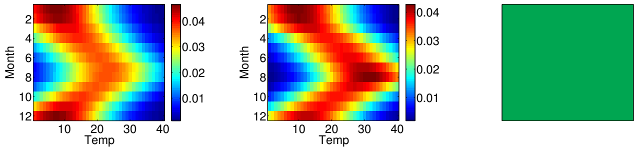
Conditional independence



$$p(X_1, ..., X_n|Z) = p(X_1|Z) \cdot, ..., \cdot p(X_n|Z)$$

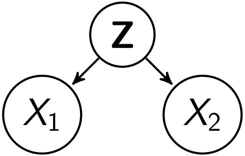
since $p(X_1|X_2, Z) = p(X_1|Z)$ and $p(X_2|X_1, Z) = p(X_2|Z)$

$$p(X_1, X_2, Z) = p(X_1|Z) \cdot p(X_2|Z) \cdot p(Z)$$



$p(X_1 = Camb|Z)$ $\times p(X_2 = Tokyo|Z)$ $\times p(Z)$

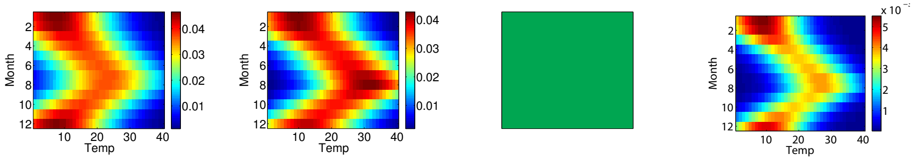
Conditional independence



$$p(X_1, \ldots, X_n|Z) = p(X_1|Z)\cdot, \ldots, \cdot p(X_n|Z)$$

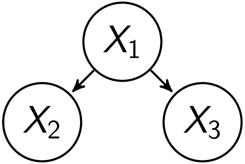
since $p(X_1|X_2, Z) = p(X_1|Z)$ and $p(X_2|X_1, Z) = p(X_2|Z)$

$$p(X_1, X_2, Z) = p(X_1|Z) \cdot p(X_2|Z) \cdot p(Z)$$

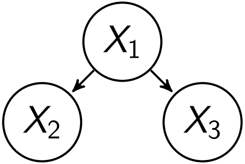


$p(X_1 = Camb|Z)$ $\times p(X_2 = Tokyo|Z)$ $\times p(Z)$ $= p(X_1, X_2, Z)$

Factoring the joint distribution

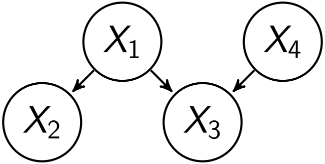


Factoring the joint distribution



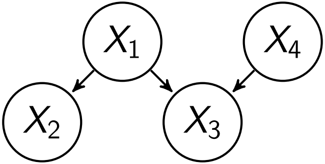
$$p(X_1, ..., X_n) = p(X_1|parents(X_1))\cdot, ..., \cdot p(X_n|parents(X_n))$$

Factoring the joint distribution



$$p(X_1, ..., X_n) = p(X_1|parents(X_1))\cdot, ..., \cdot p(X_n|parents(X_n))$$

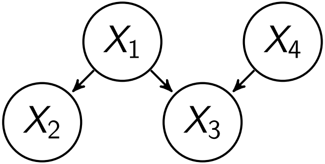
Factoring the joint distribution



$p(X_1, ..., X_n) = p(X_1|parents(X_1))\cdot, ..., \cdot p(X_n|parents(X_n))$

$p(X_1, ..., X_n) = \prod_{i=1}^{n} p(X_i|parents(X_i))$

Factoring the joint distribution

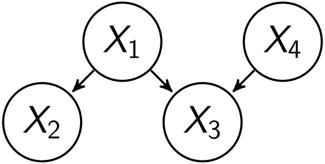


$$p(X_1, ..., X_n) = p(X_1 | parents(X_1)) \cdot, ..., \cdot p(X_n | parents(X_n))$$

$$p(X_1, ..., X_n) = \prod_{i=1}^{n} p(X_i | parents(X_i))$$

Factoring of the joint probability distribution is really important, since

Factoring the joint distribution



$p(X_1, ..., X_n) = p(X_1|parents(X_1))\cdot, ..., \cdot p(X_n|parents(X_n))$

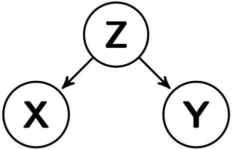
$p(X_1, ..., X_n) = \prod_{i=1}^{n} p(X_i|parents(X_i))$

Factoring of the joint probability distribution is really important, since

- $\log(x \cdot y) = \log(x) + \log(y)$
- taking the log gives an **additive model**
  $\log p(X_1, ..., X_n) =$
  $\log p(X_1|parents(X_1)) +, ..., + \log p(X_n|parents(X_n))$

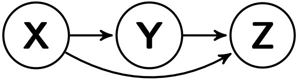
Bayes Nets

$$p(X_1, ..., X_n) = \prod_{i=1}^{n} p(X_i | parents(X_i))$$



$$p(x, y, z) = p(x|z)p(y|z)p(z)$$

$$p(x, y, z) = p(z|y)p(y|x)p(x)$$



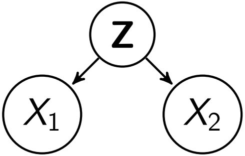
$$p(x, y, z) = p(z|y, x)p(y|x)p(x)$$

Inference with Bayes Nets

$$p(X_1, ..., X_n) = \prod_{i=1}^{n} p(X_i | parents(X_i))$$

Inference with Bayes Nets

$p(X_1, ..., X_n) = \prod_{i=1}^{n} p(X_i | parents(X_i))$



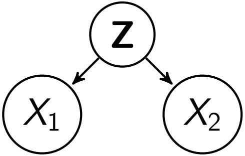
$p(X_1, X_2 | Z) = p(X_1 | Z) p(X_2 | Z)$

posterior $\propto$ likelihood $\times$ prior

$p(Z | X_1, X_2) \propto p(X_1, X_2 | Z) \times p(Z)$

Inference with Bayes Nets

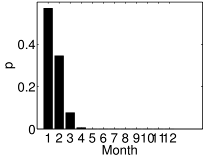
$$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i | parents(X_i))$$



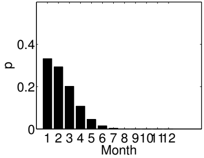
$$p(X_1, X_2 | Z) = p(X_1|Z)p(X_2|Z)$$

posterior $\propto$ likelihood $\times$ prior

$$p(Z|X_1, X_2) \propto p(X_1, X_2|Z) \times p(Z)$$



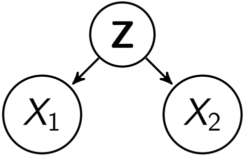
$p(X_1 = 7.4|Z)$



$\times p(X_2 = 9.9|Z)$

Inference with Bayes Nets

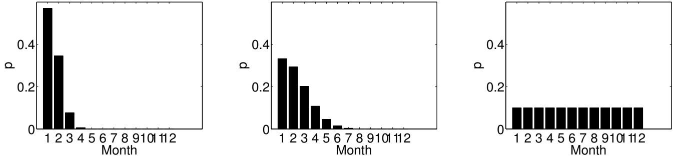
$$p(X_1, ..., X_n) = \prod_{i=1}^{n} p(X_i | parents(X_i))$$



$$p(X_1, X_2 | Z) = p(X_1 | Z) p(X_2 | Z)$$

posterior $\propto$ likelihood $\times$ prior

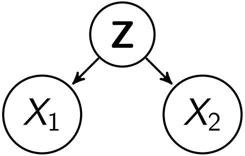
$$p(Z | X_1, X_2) \propto p(X_1, X_2 | Z) \times p(Z)$$



$p(X_1 = 7.4 | Z)$ $\times p(X_2 = 9.9 | Z)$ $\times p(Z)$

Inference with Bayes Nets

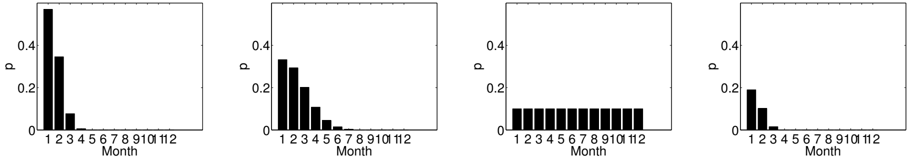
$$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i | parents(X_i))$$



$$p(X_1, X_2 | Z) = p(X_1|Z)p(X_2|Z)$$

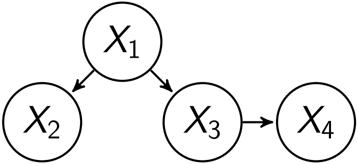
posterior $\propto$ likelihood $\times$ prior

$$p(Z|X_1, X_2) \propto p(X_1, X_2|Z) \times p(Z)$$



$p(X_1 = 7.4|Z)$ $\quad \times p(X_2 = 9.9|Z)$ $\quad \times p(Z)$ $\quad \propto p(Z|X_1, X_2)$

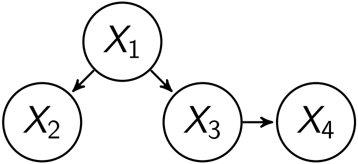
Formal definition



- Bayes Net (BN) is an annotated acyclic graph B that represents the joint probability distribution over a set of random variables V.

$$B = \langle G, \Theta \rangle$$

Formal definition

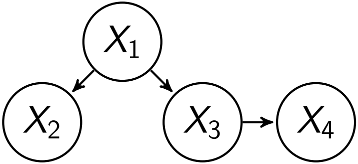


- Bayes Net (BN) is an annotated acyclic graph B that represents the joint probability distribution over a set of random variables V.

  $B = \langle G, \Theta \rangle$

- G is a graph with nodes $X_1, ..., X_n$ whose edges represent the dependencies.

Formal definition



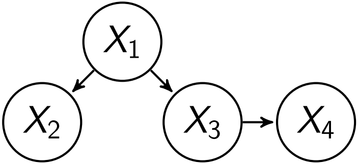
- Bayes Net (BN) is an annotated acyclic graph B that represents the joint probability distribution over a set of random variables V.

$$B = \langle G, \Theta \rangle$$

- G is a graph with nodes $X_1, ..., X_n$ whose edges represent the dependencies.
- B defines a unique JPD over V

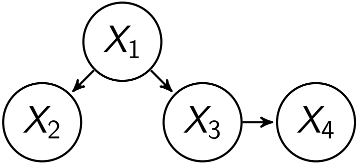
$$p(X_1, ..., X_n) = \prod_{i=1}^{n} p(X_i | \pi_i) = \prod_{i=1}^{n} \Theta_{x_i | \pi_i}$$

Recap



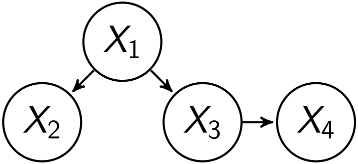
- Bayes Net (BN) is a directed acyclic graph (DAG)

Recap



- Bayes Net (BN) is a directed acyclic graph (DAG)
- which sets up conditional independence between variables

Recap



- Bayes Net (BN) is a directed acyclic graph (DAG)
- which sets up conditional independence between variables
- resulting in a factored joint probability distribution

Vision and touch

© 2002 Macmillan Magazines Ltd

NATURE | VOL 415 | 24 JANUARY 2002 | www.nature.com

Humans integrate visual and haptic information in a statistically optimal fashion

Marc O. Ernst* & Martin S. Banks

Vision Science Program/School of Optometry, University of California, Berkeley 94720-2020, USA

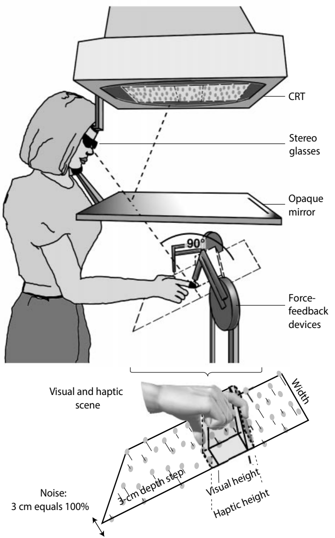
When a person looks at an object while exploring it with their hand, vision and touch both provide information for estimating the properties of the object. Vision frequently dominates the integrated visual–haptic percept, for example when judging size, shape or position[1–3], but in some circumstances the percept is clearly affected by haptics[4–7]. Here we propose that a general principle, which minimizes variance in the final estimate, determines the degree to which vision or haptics dominates. This principle is realized by using maximum-likelihood estimation[8–15] to combine the inputs. To investigate cue combination quantitatively, we first measured the variances associated with visual and haptic estimation of height. We then used these measurements to construct a maximum-likelihood integrator. This model behaved very similarly to humans in a visual–haptic task. Thus, the nervous system seems to combine visual and haptic information in a fashion that is similar to a maximum-likelihood integrator. Visual dominance occurs when the variance associated with visual estimation is lower than that associated with haptic estimation.

The estimate of an environmental property by a sensory system can be represented by

$$\hat{S}_i = f_i(S) \qquad (1)$$

where S is the physical property being estimated and f is the operation by which the nervous system does the estimation. The subscripts refer to the modality (i could also refer to different cues within a modality). Each estimate, $\hat{S}_i$, is corrupted by noise. If the noises are independent and gaussian with variance σ_i^2, and the bayesian prior is uniform, then the maximum-likelihood estimate

* Present address: Max Planck Institute for Biological Cybernetics, Tübingen 72076, Germany.



Vision and touch

Ernst and Banks (2002) asked subjects which of two sequentially presented blocks was the taller. Subjects used either vision alone, touch alone or a combination of the two.

Vision and touch

Ernst and Banks (2002) asked subjects which of two sequentially presented blocks was the taller. Subjects used either vision alone, touch alone or a combination of the two.

If vision v and touch t information are independent given an object x then we have

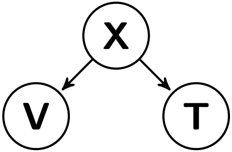
$$p(v, t, x) = p(v|x)p(t|x)p(x)$$

Vision and touch

Ernst and Banks (2002) asked subjects which of two sequentially presented blocks was the taller. Subjects used either vision alone, touch alone or a combination of the two.

If vision v and touch t information are independent given an object x then we have

$$p(v, t, x) = p(v|x)p(t|x)p(x)$$

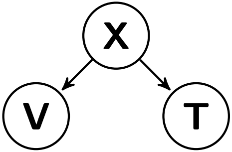


Vision and touch

Ernst and Banks (2002) asked subjects which of two sequentially presented blocks was the taller. Subjects used either vision alone, touch alone or a combination of the two.

If vision v and touch t information are independent given an object x then we have

$$p(v, t, x) = p(v|x)p(t|x)p(x)$$



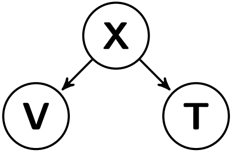
Bayesian fusion of sensory information then produces a posterior density

$$p(x|v, t) = \frac{p(v|x)p(t|x)p(x)}{p(v,t)}$$

Vision and touch

$$p(v|x) = \mathcal{N}(\mu, \sigma^2)$$

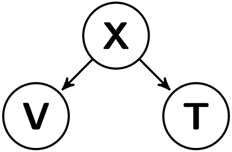
$$p(t|x) = \mathcal{N}(\mu, \sigma^2)$$



Vision and touch

$p(v|x) = \mathcal{N}(\mu, \sigma^2)$

$p(t|x) = \mathcal{N}(\mu, \sigma^2)$

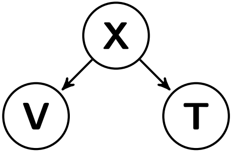


Ernst and Banks use precision instead of variance. Precision is inverse variance
$\lambda = \frac{1}{\sigma^2}$

Vision and touch

$$p(v|x) = \mathcal{N}(\mu, \sigma^2)$$

$$p(t|x) = \mathcal{N}(\mu, \sigma^2)$$



Ernst and Banks use precision instead of variance. Precision is inverse variance
$\lambda = \frac{1}{\sigma^2}$

For a Gaussian likelihood with mean m_d and precision λ_d and a Gaussian prior with mean m_0 and precision λ_0
the posterior is a Gaussian with

$$m = \frac{\lambda_d}{\lambda} m_d + \frac{\lambda_0}{\lambda} m_0$$

$$\lambda = \lambda_d + \lambda_0$$

Vision and touch

For a Gaussian likelihood with mean m_d and precision λ_d and a Gaussian prior with mean m_0 and precision λ_0 the posterior is a Gaussian with

$$m = \frac{\lambda_d}{\lambda} m_d + \frac{\lambda_0}{\lambda} m_0$$

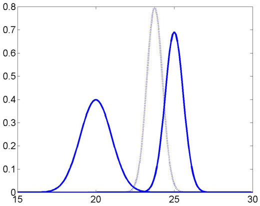
$$\lambda = \lambda_d + \lambda_0$$

Vision and touch

For a Gaussian likelihood with mean m_d and precision λ_d and a Gaussian prior with mean m_0 and precision λ_0 the posterior is a Gaussian with

$$m = \frac{\lambda_d}{\lambda} m_d + \frac{\lambda_0}{\lambda} m_0$$

$$\lambda = \lambda_d + \lambda_0$$



The two solid curves show the probability densities for the prior $m_0 = 20$, $\lambda_0 = 1$ and the likelihood $m_d = 25$ and $\lambda_d = 3$. The dotted curve shows the posterior distribution with $m = 23.75$ and $\lambda = 4$. The posterior is closer to the likelihood because the likelihood has higher precision.

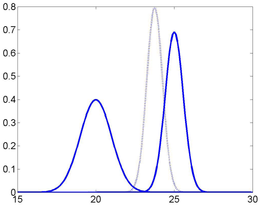
$$23.75 = \frac{3}{4} 25 + \frac{1}{4} 18$$

Vision and touch

For a Gaussian likelihood with mean m_d and precision λ_d and
a Gaussian prior with mean m_0 and precision λ_0
the posterior is a Gaussian with

$$m = \frac{\lambda_d}{\lambda} m_d + \frac{\lambda_0}{\lambda} m_0$$

$$\lambda = \lambda_d + \lambda_0$$



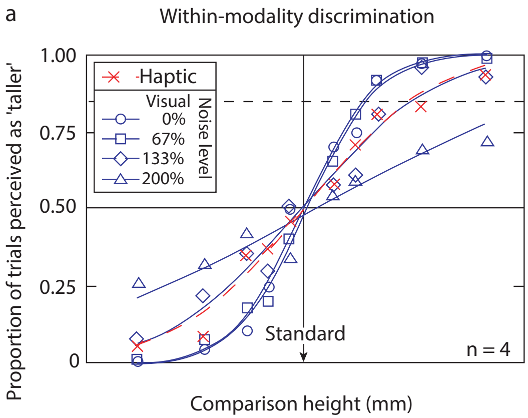
- Precisions add
- The posterior mean is the sum of the priorand data means, each weighted by their relative precision

Vision and touch

They recorded the accuracy with which discrimination could be made and plotted this as a function of difference in block height. This was first done for each condition alone. One can then estimate precisions, λ_v and λ_t.

Vision and touch

They recorded the accuracy with which discrimination could be made and plotted this as a function of difference in block height. This was first done for each condition alone. One can then estimate precisions, λ_v and λ_t.



Vision and touch

$$m_{vt} = \frac{\lambda_v}{\lambda_{vt}} m_v + \frac{\lambda_t}{\lambda_{vt}} m_t$$

$$\lambda_{vt} = \lambda_v + \lambda_t$$

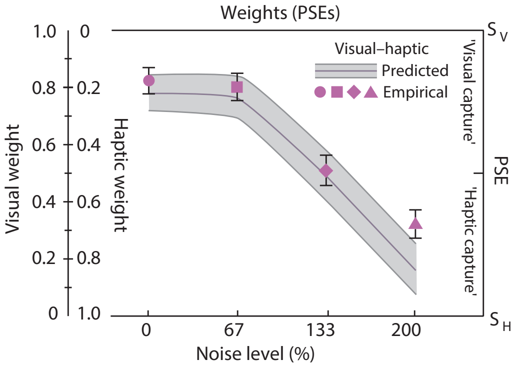
$$m_{vt} = w_v m_v + w_t m_t$$

Vision and touch

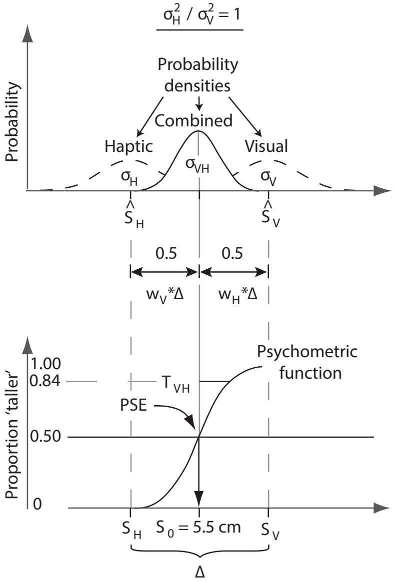
$$m_v t = \frac{\lambda_v}{\lambda_v t} m_v + \frac{\lambda_t}{\lambda_v t} m_t$$

$$\lambda_v t = \lambda_v + \lambda_t$$

$$m_v t = w_v m_v + w_t m_t$$



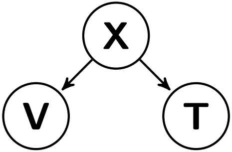
Vision and touch



Vision and touch

$$p(v|x) = \mathcal{N}(\mu, \sigma^2)$$

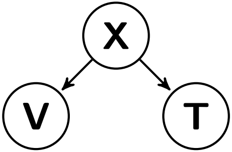
$$p(t|x) = \mathcal{N}(\mu, \sigma^2)$$



Vision and touch

$p(v|x) = \mathcal{N}(\mu, \sigma^2)$

$p(t|x) = \mathcal{N}(\mu, \sigma^2)$



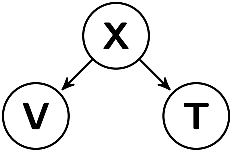
Bayesian fusion of sensory information produces a posterior density

$p(x|v, t) \propto p(v|x)p(t|x)$

Vision and touch

$$p(v|x) = \mathcal{N}(\mu, \sigma^2)$$

$$p(t|x) = \mathcal{N}(\mu, \sigma^2)$$



Bayesian fusion of sensory information produces a posterior density

$$p(x|v,t) \propto p(v|x)p(t|x)$$

$$m_v t = \frac{\lambda_v}{\lambda_v t} m_v + \frac{\lambda_t}{\lambda_v t} m_t$$

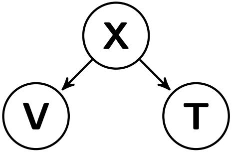
$$\lambda_v t = \lambda_v + \lambda_t$$

$$m_v t = w_v m_v + w_t m_t$$

Vision and touch

$$p(v|x) = \mathcal{N}(\mu, \sigma^2)$$

$$p(t|x) = \mathcal{N}(\mu, \sigma^2)$$



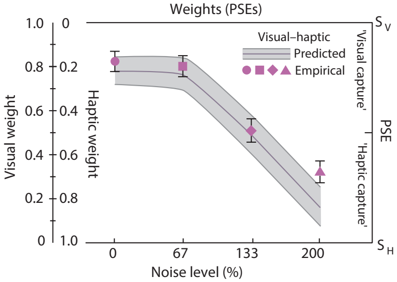
Bayesian fusion of sensory information produces a posterior density

$$p(x|v,t) \propto p(v|x)p(t|x)$$

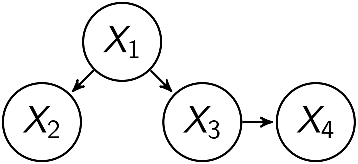
$$m_v t = \frac{\lambda_v}{\lambda_v t} m_v + \frac{\lambda_t}{\lambda_v t} m_t$$

$$\lambda_v t = \lambda_v + \lambda_t$$

$$m_v t = w_v m_v + w_t m_t$$

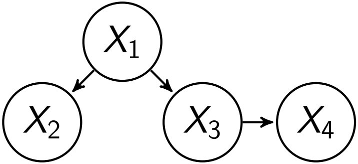


Learning with BN



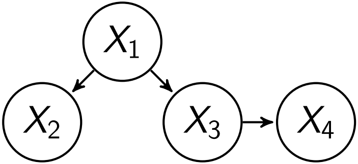
- $B = \langle G, \Theta \rangle$

Learning with BN



- $B = \langle G, \Theta \rangle$
- G is a graph with nodes $X_1, ..., X_n$ whose edges represent the dependencies.

Learning with BN



- $B = \langle G, \Theta \rangle$
- G is a graph with nodes $X_1, ..., X_n$ whose edges represent the dependencies.
- B defines a unique JPD over V

$$p(X_1, ..., X_n) = \prod_{i=1}^{n} p(X_i | \pi_i) = \prod_{i=1}^{n} \Theta_{x_i | \pi_i}$$

Complex inference with BN

- In many practical settings the BN is unknown and one needs to learn it from the data.

Complex inference with BN

- In many practical settings the BN is unknown and one needs to learn it from the data.
- Problem: Given data and prior information, estimate the graph topology G and the parameters Θ.

Complex inference with BN

- In many practical settings the BN is unknown and one needs to learn it from the data.
- Problem: Given data and prior information, estimate the graph topology G and the parameters Θ.

Four cases of BN learning problems

Case	**Structure**	**Observability**	**Learning method**
1	Known	Full	Maximum-likelihood estimation
2	Known	Partial	EM (or gradient descent), MCMC
3	Unknown	Full	Search through model space
4	Unknown	Partial	EM + Search through model space

Complex inference with BN

Case	Structure	Observability	Learning method
1	Known	Full	Maximum-likelihood estimation

Complex inference with BN

Case	Structure	Observability	Learning method
1	Known	Full	Maximum-likelihood estimation

- Goal: find the values of BN parameters (in each CPD) that maximise the (log)likelihood of the dataset.

Complex inference with BN

Case	Structure	Observability	Learning method
1	Known	Full	Maximum-likelihood estimation

- Goal: find the values of BN parameters (in each CPD) that maximise the (log)likelihood of the dataset.
- Dataset $X = x_1, ..., x_n$

Complex inference with BN

Case	Structure	Observability	Learning method
1	Known	Full	Maximum-likelihood estimation

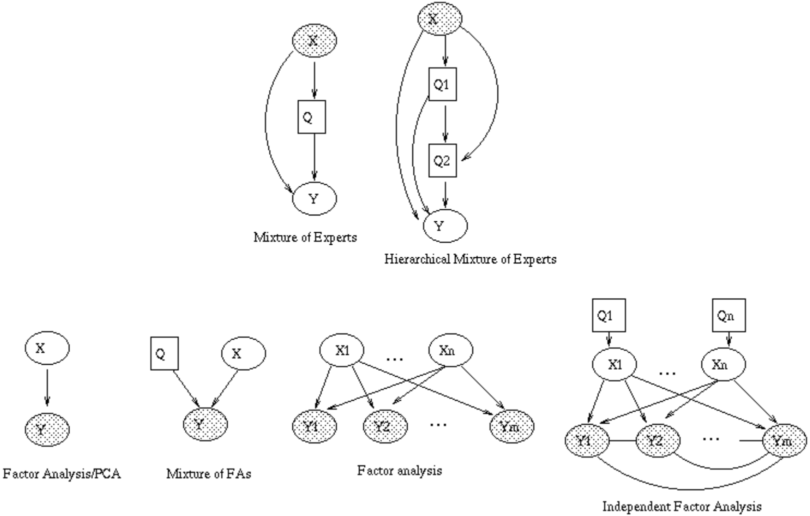
- Goal: find the values of BN parameters (in each CPD) that maximise the (log)likelihood of the dataset.
- Dataset $X = x_1, ..., x_n$
- Parameter set $\Theta = \theta_1, ..., \theta_n$, where θ_i is the vector of parameters for the CPD of x_i

Complex inference with BN

Case	Structure	Observability	Learning method
1	Known	Full	Maximum-likelihood estimation

- Goal: find the values of BN parameters (in each CPD) that maximise the (log)likelihood of the dataset.
- Dataset $X = x_1, ..., x_n$
- Parameter set $\Theta = \theta_1, ..., \theta_n$, where θ_i is the vector of parameters for the CPD of x_i
- $\log L(\Theta|X) = \sum_n \log P(x_i|\pi_i, \theta_i)$

Complex models with BN



A Unifying Review of Linear Gaussian Models, Sam Roweis & Zoubin Ghahramani. *Neural Computation* 11(2) (1999) pp.305-345