

Probabilistic Graphical Models in Computer Vision (IN2329)

Csaba Domokos

Summer Semester 2015/2016

Announcement: Computer Vision Group *	2
1. Introduction	3
Agenda for today's lecture *	4
Administration	5
The course: IN2329 *	6
April *	7
May *	8
June *	9
July *	10
Exam *	11
Bonus *	12
Recommended literature & prerequisites *	13
Course Page *	14
Overview	15
Overview of the lecture *	16
Human-pose estimation *	17
Binary image segmentation *	18

Semantic image segmentation *	19
Stereo matching *	20
Stereo matching *	21
Object detection *	22
Probability theory	23
Reasoning under uncertainty	24
Experiment, event space, event	25
Basic notations	26
Discrete probability space	27
Example: throwing two “fair dice” *	28
σ -algebra, measure, measure space	29
Probability space	30
Example: throwing a dart *	31
Some simple consequences of the axioms	32
Conditional Probability	33
Conditional probability	34
Example *	35
The chain rule	36
Bayes’ rule	37
Independence	38
Conditional independence	39
Summary	40
Literature *	41
A brain teaser *	42

Announcement: Computer Vision Group *

Inquiries for Bachelor and Master projects are always welcome!

We are currently working on the following research topics:



3D reconstruction



Optical flow



Shape analysis



Robot vision



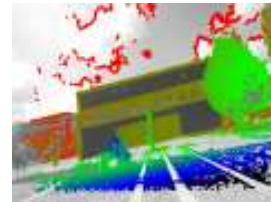
RGB-D vision



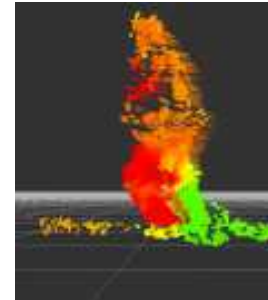
Image segmentation



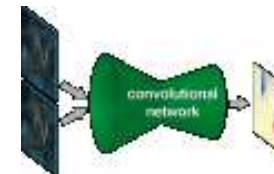
Convex relaxation



Visual SLAM



Scene flow



Deep learning

Please complete the form: <https://vision.in.tum.de/application>

Agenda for today's lecture *

1. Administration
2. Overview of the course
3. Introduction to Probability theory
 - Basic definitions
 - Conditional probability, Bayes' rule
 - Independence, conditional independence

The course: IN2329 *

The course **Probabilistic Graphical Models in Computer Vision** will be organized as follows:

- Lectures: on Tuesdays at 10.00–12.00 in Room 00.13.036
- Tutorials: on Tuesdays at 14.00–16.00 in Room 02.05.014

The tutorials combines *theoretical* and *programming* assignments:

- **Assignment distribution:** Tuesday 11.00–11.15 in Room 00.13.036
- **Theoretical assignment due:** Tuesday 11.00–11.15 in Room 00.13.036
- **Assignment presentation:** Tuesday 14.00–16.00 in Room 02.05.014

Lecturer

Dr. Csaba Domokos (csaba.domokos@in.tum.de)

Teaching assistant (TA)

Lingni Ma (lingni@in.tum.de)

Feel free to contact us!

April *

April 2016						
Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
				1	2	3
4	5	6	7	8	9	10
11	12 Lecture 1 Tutorial 1	13	14	15	16	17
18	19 Lecture 2 Tutorial 2	20	21	22	23	24
25	26 Lecture 3 Tutorial 3	27	28	29	30	

May *

May 2016						
Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
2	3 Lecture 4 Tutorial 4	4	5	6	7	8
9	10 Lecture 5 Tutorial 5	11	12	13	14	15
16	17	18	19	20	21	22
23	24 Lecture 6 Tutorial 6	25	26	27	28	29
30	31 Lecture 7 Tutorial 7					

June *

June 2016						
Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
		1	2	3	4	5
6	7 Lecture 8 Tutorial 8	8	9	10	11	12
13	14 Lecture 9 Tutorial 9	15	16	17	18	19
20	21 Lecture 10 Tutorial 10	22	23	24	25	26
27	28 Lecture 11 Tutorial 11	29	30			

July *

July 2016						
Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
				1	2	3
4	5 Lecture 12 Tutorial 12	6	7	8	9	10
11	12 Lecture 13 Tutorial 13	13	14	15	16	17
18	19	20	21	22	23	24
25	26 Week of the exam	27	28	29	30	31

Exam *

- The exam will be oral.
- According to our schedule, the exam will be held in the **last week of July, 25th–29th**.
- Students **need to be registered** prior to the exam: **May, 16th–June, 30th** via TUM online.

Participation at the tutorial:

- **Not mandatory, but highly recommended:**

Theoretical assignments will help to understand the topics of the lecture. *Programming assignments* will help to apply the theory to practical Computer Vision problems.

- **Bonus:** Active students who solve *60% of the assignments* earn a bonus.

If someone receives a mark between 1.3 and 4.0 in the *final exam*, the mark will be **improved by 0.3 and 0.4, respectively**.

Note that marks of 1.0 and 5.0 cannot be improved!

Bonus *

To achieve the bonus, the following requirements have to be fulfilled:

Theory

- **60%** of all theoretical assignments have to be solved.
(Note that submissions happen **only** on Tuesdays at 11.00–11.15)
- The theoretical exercises have to *be presented in front of the class*.
(The TA randomly selects a student who presents an exercise.)

Programming

- **60%** of all programming assignments have to be presented during the tutorial.
- The programming exercises should *be explained to the TA*.

To promote team work, please form **groups of two or three students** in order to solve and submit the assignments.

Recommended literature & prerequisites *



- D. Koller, N. Friedman. Probabilistic Graphical Models: Principles and Techniques, MIT Press, 2009.
- S. Nowozin, C. H. Lampert. Structured Learning and Prediction in Computer Vision, Foundations and Trends in Computer Graphics and Vision, 2011. Download
- A. Blake, P. Kohli, C. Rother. Markov Random Fields for Vision and Image Processing, MIT Press, 2011.

In addition, we will mention recent *conference* and *journal papers*.

Prerequisites: the course is intended for *Master students*.

- **Basic Mathematics:** multivariate analysis and linear algebra.
- **Basic Computer Science:** dynamic programming and basic data structures.

Course Page *

On the **internal site** of the course page you could access to extra course material:

https://vision.in.tum.de/teaching/ss2016/lecture_graphical_models/material

Password: PGMCV:SS16

Course materials:

- Printer friendly **slides for each lecture** (available prior to the lecture)
- **Assignment sheets** (available after the lecture)
- **Solution sheets** (available after the tutorial)

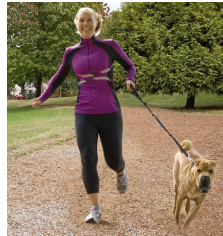
The course page will also be used for extra announcements.

Overview of the lecture *



Human–pose estimation *

The goal is to recognize an *articulated object* (i.e. human body) with different connecting parts (e.g., head, torso, left arm, ...).



Input image



Human pose estimation

An object is composed of a number of *rigid parts*, where each part is modeled as a **rectangle**. The *connections* encode generic relationships such as “close to”, “to the left of”.

Binary image segmentation *

The goal is to give a binary label $y_i \in \mathbb{B} \triangleq \{0, 1\}$ for each pixel i , where 0 and 1 mean the *background* (a.k.a. ground) and the *foreground* (a.k.a. figure), respectively.



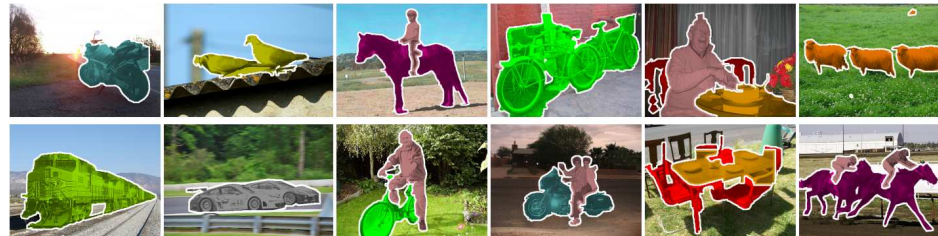
Input image



Figure–ground segmentation

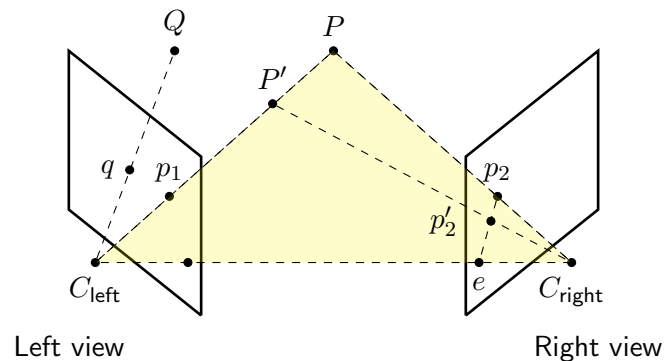
Semantic image segmentation *

The goal is to give a label $y_i \in \mathcal{L} = \{1, 2, \dots, c\}$ for each pixel i according to its *semantic meanings*.



Exemplar semantic segmentations

Stereo matching *

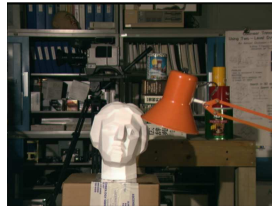


Given two images (i.e. left and right), an observed 2D point p_1 on the *left image*, corresponding to a 3D point P that is situated on a line in \mathbb{R}^3 . This line will be observed as a line on the *right image*.

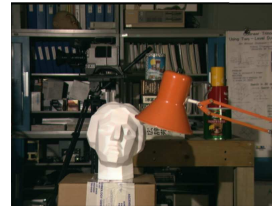
The pixels p_1 and p_2 corresponding to P should have similar intensities.

Stereo matching *

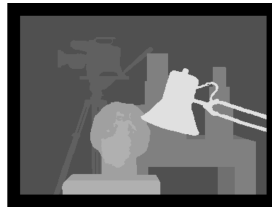
Usually we assume **rectified images** (i.e. the directions of the cameras are parallel).



Left view



Right view



Ground truth (depth map)



Result (depth map)

Object detection *

We address the problem of *binary image segmentation*, where we also assume non-local **parameters** that are **known a priori**. For example, one can assume prior knowledge about the **shape** of the foreground.



Exemplar binary segmentation of cars assuming shape prior

You may realize that we will mainly deal with labelling problems.

Reasoning under uncertainty

We often want to understand a system when we have *imperfect* or *incomplete* information due to, for example, noisy measurement.

There are *two main reasons* why we might **reason under uncertainty**:

- *Laziness*: modeling every detail of a complex system is *costly*.
- *Ignorance*: we may not completely understand.

Probability $P(A)$ refers to a degree of confidence that an event A with uncertain nature will occur.

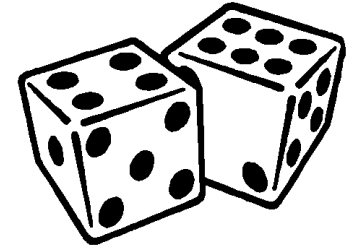
It is common to assume that $0 \leq P(A) \leq 1$:

- If $P(A) = 1$, we are certain that A occurs,
- while $P(A) = 0$ asserts that A will not occur.

Experiment, event space, event

An **experiment** is a (random) process that can be infinitely many times repeated and has a well-defined set of possible **outcomes**. In case of repeated experiments the individual repetitions are also called **trials**.

Example: throwing two “fair dice” (i.e. we assume equally likely chance of landing on any face) with six faces.



The **event space**, denoted by Ω , is the set of possible outcomes.

Example: $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$.

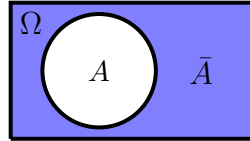
A set of outcomes $A \subseteq \Omega$ is called an **event**. An **atomic event** is an event that contains a single outcome $\omega \in \Omega$.

Example: $A = \{(i, j) : i + j = 11\}$, i.e. the sum of the numbers showing on the top is equal to eleven.

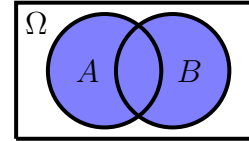
Basic notations

Let A and B be two events from an event space Ω . We will use the following notations:

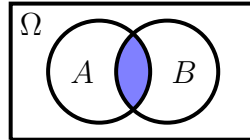
A does not occur: $\bar{A} = \Omega \setminus A$



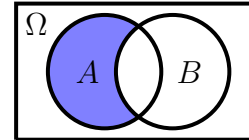
either A or B occur: $A \cup B$



both A and B occur: $A \cap B$



A occurs and B does not: $A \setminus B$



- The \emptyset is called the **impossible event**; and
- Ω is the **sure event**.

Discrete probability space

A *probability space* represents our **uncertainty** regarding an *experiment*.

A triple (Ω, \mathcal{A}, P) is called a **discrete probability space**, if

- Ω is not empty and **countable** (i.e. $\exists \mathcal{S} \subseteq \mathbb{N}$ such that $|\Omega| = |\mathcal{S}|$),
- \mathcal{A} is the **power set** $\mathcal{P}(\Omega)$ (i.e. the set of all subsets of Ω), and
- $P : \mathcal{A} \rightarrow \mathbb{R}$ is a function, called a **probability measure**, with the following properties:
 1. $P(A) \geq 0$ for all $A \in \mathcal{A}$
 2. $P(\Omega) = 1$
 3. **σ -additivity** holds: if $A_n \in \mathcal{A}, n = 1, 2, \dots$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

The conditions 1-3. are called **Kolmogorov's axioms**.

Example: throwing two “fair dice” *

For this case a *discrete probability space* (Ω, \mathcal{A}, P) is given by

- **Event space:** $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$.
- $\mathcal{A} = \mathcal{P}(\Omega) = \{\{(1, 1)\}, \dots, \{(1, 1), (1, 2)\}, \dots, \{(1, 1), (1, 2), (1, 3)\}, \dots\}$.
- The **probability measure**

$$P(A) = \frac{|A|}{36} = \frac{k}{36},$$

where k is the number of *atomic events* in A .

Example: Let A denote the event that “the sum of the numbers showing on the top is equal to eleven”, that is

$$A = \{(i, j) : i + j = 11\} = \{(5, 6), (6, 5)\}.$$

Hence

$$P(A) = P(\{(5, 6), (6, 5)\}) = \frac{2}{36}.$$

σ -algebra, measure, measure space

Assume an arbitrary set Ω and $\mathcal{A} \subseteq \mathcal{P}(\Omega)$. The set \mathcal{A} is a **σ -algebra over Ω** if the following conditions are satisfied:

1. $\emptyset \in \mathcal{A}$,
2. $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$ (i.e. it is *closed under complementation*),
3. $A_i \in \mathcal{A} (i \in \mathbb{N}) \Rightarrow \bigcup_{i=0}^{\infty} A_i \in \mathcal{A}$ (i.e. it is *closed under countable union*).

It is a consequence of this definition that $\Omega \in \mathcal{A}$ is also satisfied. (See exercise.)

Assume an *arbitrary set* Ω and a σ -algebra \mathcal{A} over Ω . A function $P : \mathcal{A} \rightarrow [0, \infty]$ is called a **measure** if the following conditions are satisfied:

1. $P(\emptyset) = 0$,
2. P is σ -additive.

Let \mathcal{A} be a σ -algebra over Ω and $P : \mathcal{A} \rightarrow [0, \infty]$ is a *measure*. (Ω, \mathcal{A}) is said to be a **measurable space** and the triple (Ω, \mathcal{A}, P) is called a **measure space**.

Probability space

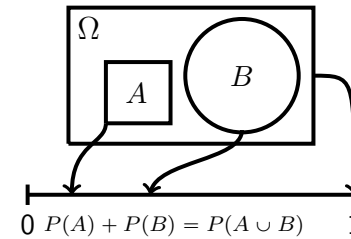
A **probability space** is a triple (Ω, \mathcal{A}, P) , where (Ω, \mathcal{A}) is a *measurable space*, and P is a *measure* such that $P(\Omega) = 1$, called a **probability measure**.

To summarize:

A triple (Ω, \mathcal{A}, P) is called **probability space**, if

- the **event space** Ω is *not empty*,
- \mathcal{A} is a **σ -algebra** over Ω , and
- $P : \mathcal{A} \rightarrow \mathbb{R}$ is a function with the following properties:
 1. $P(A) \geq 0$ for all $A \in \mathcal{A}$
 2. $P(\Omega) = 1$
 3. **σ -additive**: if $A_n \in \mathcal{A}$, $n = 1, 2, \dots$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$



Example: throwing a dart *

Suppose a dart is thrown at a round board modeled as a unit circle. The **event space** contains the location of the dart if it lands in the board only. Hence it is given by

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$



We denote the **area** of an the **event** $A \subseteq \Omega$ by $\mu(A)$, which is defined as the *Riemann-integral* of the **characteristic function of A**

$$\mu(A) := \int_{\Omega} \chi_A(x) dx, \quad \text{where } \chi_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases}$$

The σ -**algebra** \mathcal{A} over Ω is defined as follows

$$\mathcal{A} = \{A \subseteq \Omega : \mu(A) \text{ exists}\}.$$

The **probability measure** $P : \Omega \rightarrow [0, 1]$ is given by $P(A) = \frac{\mu(A)}{\pi}$.

Some simple consequences of the axioms

The following rules are frequently used in applications:

■ $P(A) = 1 - P(\Omega \setminus A)$.

Proof. Note that A and $\Omega \setminus A$ are disjoint.

$$1 = P(\Omega) = P(A \cup (\Omega \setminus A)) = P(A) + P(\Omega \setminus A). \quad \square$$

■ $P(\emptyset) = 0$.

Proof. $P(\emptyset) = 1 - P(\Omega \setminus \emptyset) = 1 - P(\Omega) = 1 - 1 = 0. \quad \square$

■ If $A \subseteq B$, then $P(A) \leq P(B)$.

■ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

■ $P(A \cup B) \leq P(A) + P(B)$.

■ $P(A \setminus B) = P(A) - P(A \cap B)$.

■ ...

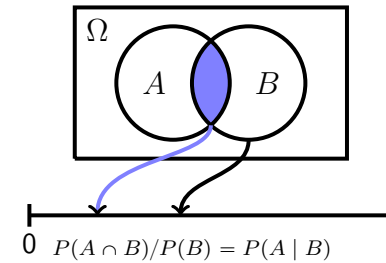
Conditional probability

Conditional probability allows us to reason with *partial information*.

If $P(B) > 0$, the **conditional probability of A given B** is defined as

$$P(A | B) \triangleq \frac{P(A \cap B)}{P(B)} .$$

This is the probability that A occurs, given we have observed B , i.e. we know the experiment's actual outcome will be in B .



Note that the *axioms and rules of probability theory are fulfilled* for the conditional probability. (e.g., $P(A | B) = 1 - P(\bar{A} | B)$).

Example *

Consider two producing machines creating identical product in a factory. Assume we are given the following table with probabilities

	Machine I	Machine II	
The product is good	0.56	0.41	0.97
The product is waste	0.01	0.02	0.03
	0.57	0.43	1

Question: What is the probability of a product was created by Machine I, when it is good?

Let A denote the event that “the product was created by Machine I” and let B denote the event that “the product is good”.

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{0.56}{0.97} \approx 0.58 .$$

The chain rule

Starting with the definition of *conditional probability* $P(B | A)$ and multiplying by $P(A)$ we get the **product rule**:

$$P(A \cap B) = P(A)P(B | A) .$$

The chain rule is given by

$$P(\cap_{i=1}^n A_i) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \cdots P(A_n | \cap_{i=1}^{n-1} A_i) . \quad (1)$$

Proof. By induction. For $n = 2$ we get the product rule. Let $n \in \mathbb{N}$ be given and suppose Eq. (1) is true for $k \leq n$. Then

$$P(\cap_{i=1}^{n+1} A_i) = P(A_{n+1} \cap (\cap_{i=1}^n A_i)) = P(A_{n+1} | \cap_{i=1}^n A_i)P(\cap_{i=1}^n A_i) .$$

□

The chain rule will become important later when we discuss *conditional independence*.

Bayes' rule

By making use of the product rule we can get

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

$P(A | B)$ is often called the **posteriori probability**, and $P(B | A)$ is called the **likelihood**, and $P(A)$ is called the **prior probability**.

A more general version of **Bayes' rule**, when we have a background event C :

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

Example: What is the probability that a product is good, if it was created by Machine I? We are given $P(A | B) = 0.58$, $P(A) = 0.57$ and $P(B) = 0.97$.

$$P(B | A) = \frac{P(A | B)P(B)}{P(A)} = \frac{0.58 \cdot 0.97}{0.57} \approx 0.98 .$$

Independence

Two events A and B are **independent**, denoted by $A \perp B$, if

$$P(A | B) = P(A)$$

or, equivalently, iff

$$P(A \cap B) = P(A)P(B) .$$

If A and B are **independent**, learning that B happened *does not make A more or less likely to occur*.

Example: Suppose we roll a die. Let us consider the events A denoting “the die outcome is even” and B denoting “the die outcome is either 1 or 2”.

If the die is fair, then $P(A) = \frac{1}{2}$ and $P(B) = \frac{1}{3}$. Moreover $A \cap B$ means the event that the outcome is two, so $P(A \cap B) = \frac{1}{6}$.

$$P(A \cap B) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = P(A)P(B) \quad \Rightarrow \quad A \text{ and } B \text{ are independent.}$$

Conditional independence

Let A , B and C be events. A and B are **conditionally independent** given C , denoted by $A \perp\!\!\!\perp B \mid C$, iff

$$P(A \mid C) = P(A \mid B \cap C),$$

or, equivalently, iff

$$P(A \cap B \mid C) = P(A \mid C)P(B \mid C).$$

A and B are **conditionally independent** given C means that *once we learned C , learning B gives us no additional information about A .*

Examples:

- The operation of a car's *starter motor* is conditionally independent its *radio* given the *status of the battery*.
- *Symptoms* are conditionally independent given the *disease*.

Summary

- A **probability space** is a triple (Ω, \mathcal{A}, P) , where (Ω, \mathcal{A}) is a *measurable space*, and P is a *measure* such that $P(\Omega) = 1$.
- Let $P(B) > 0$, then the **conditional probability of A given B** is defined as

$$P(A \mid B) \triangleq \frac{P(A \cap B)}{P(B)}.$$

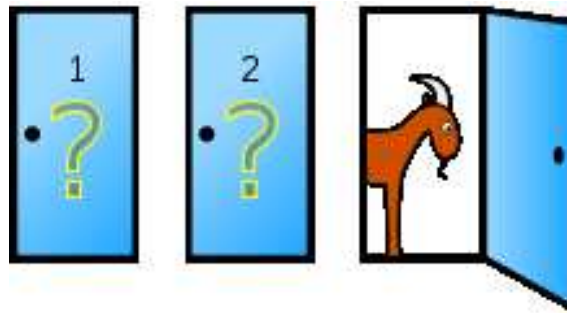
- If A and B are **independent** ($A \perp\!\!\!\perp B$), learning that B happened does not make A more or less likely to occur.
- A and B are **conditionally independent given C** , denoted by $A \perp\!\!\!\perp B \mid C$, means that once we learned C , learning B gives us no additional information about A .

In the **next lecture** we will learn about

- *Random variables*
- *Probability distributions*
- *Expectation-maximization algorithm*

A brain teaser *

Suppose you are on a game show and you are given the choice of three doors: Behind one door is a **car**; behind the others, **goats**. You pick a door, say No. 1, and the *host*, who knows what is behind the doors, opens another door, say No. 3, which has a goat.



He then says to you, "Do you want to pick door No. 2?"

Question: Is it to your advantage to switch your choice?