Intelligent Systems
Probability Theory

Carsten Rother, Dmitrij Schlesinger

WS2016/2017, 14.12.2016

COMPUTER VISION LAB DRESDEN

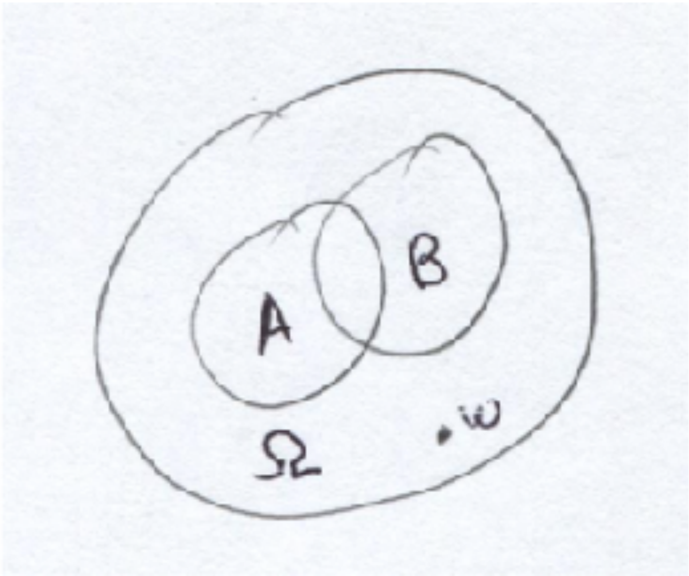
TECHNISCHE UNIVERSITÄT DRESDEN

Probability Space (simplified)

The set of **elementary events** $\omega \in \Omega$

May be of very different nature, e.g.:

- finite discrete set: e.g. $\Omega = \{human, monkey, app\}$
- the set of real numbers $\Omega = \mathbb{R}$
- the set of vectors $\Omega = \mathbb{R}^n$
- the set of graphs, of sequences, etc.



Note: there are sets that can not be a set of elementary events.

An **event** is a union of the elementary ones: $A \subseteq \Omega$

Probability measure

For discrete sets: just a number for each elementary event satisfying:

$$P(\omega) \geq 0, \quad \sum_{\omega \in \Omega} P(\omega) = 1$$

The probability for an event is

$$P(A) = \sum_{\omega \in A} P(\omega)$$

Note: for continuous sets the above definition does not work.

An example

Two people agree to meet in a time interval $[0, T]$. Both come to the meeting point equiprobable within this time interval. The person coming first waits for the second one a certain time a.

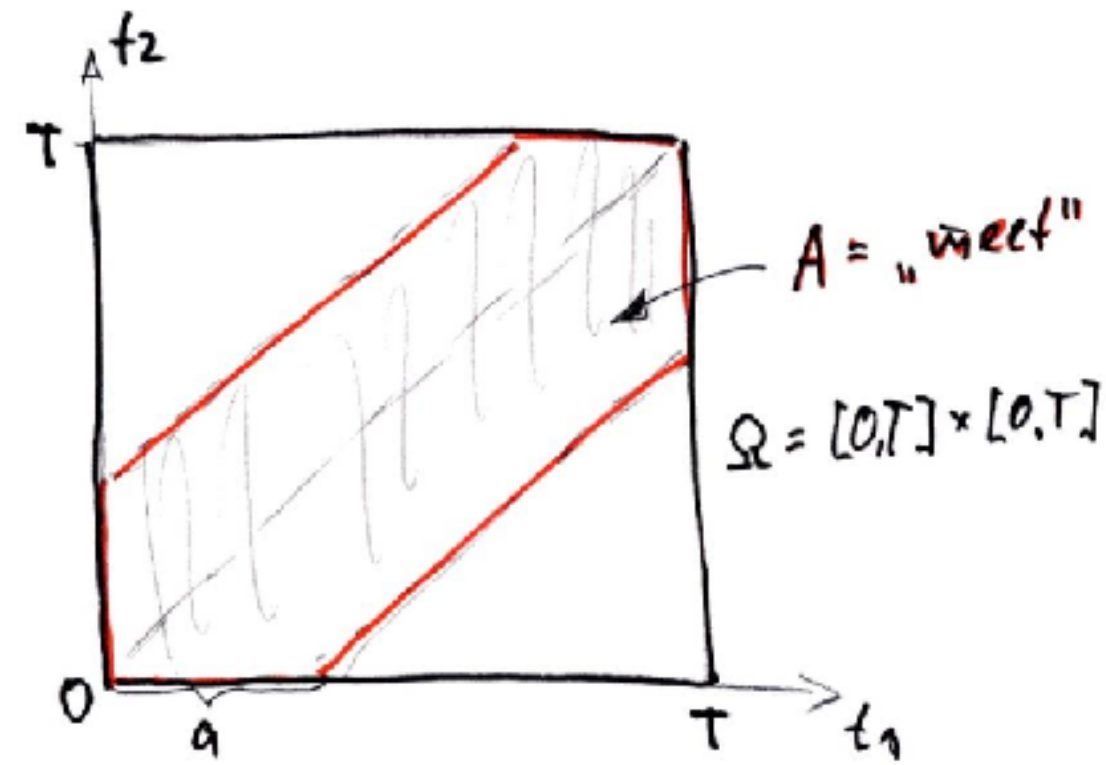
What is the probability that the people meet each other ?

The key idea is to consider the set of elementary events Ω in an appropriate way. Let (t_1, t_2) be the pair consisting of the arrival times. The event of interest is

$$A = \{(t_1, t_2) : |t_1 - t_2| \leq a\}$$

its probability is

$$P(A) = \frac{T^2 - (T - a)^2}{T^2}$$



Random Variables

Here a special case — **real-valued** random variables.

A random variable ξ is just a mapping $\xi : \Omega \to \mathbb{R}$, i.e. there is a real value $\xi(\omega)$ for each $\omega \in \Omega$.

Example: the set of elementary events is a set of balls in a bag. The random variable for this set might be e.g. the weight for each ball, or the size or radioactivity whatever.

Note: elementary events are not numbers – they are elements of a general set Ω

Random variables are in contrast numbers, i.e. they can be summed up, subtracted, squared etc.

Probability Distribution

Cumulative distribution function for a random variable ξ :

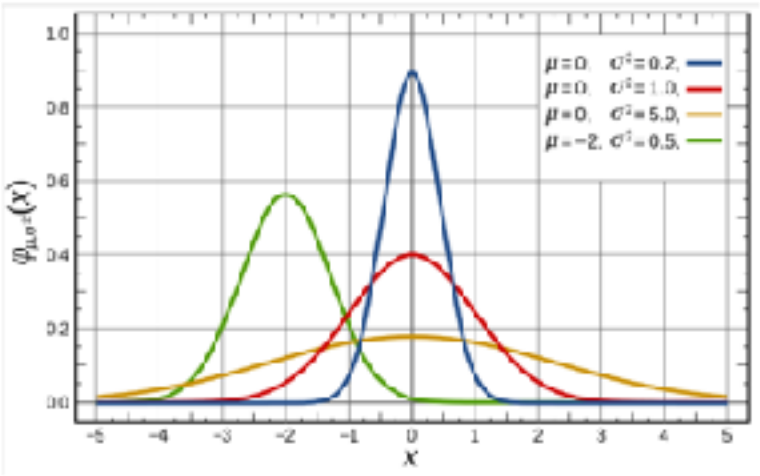
$$F_\xi(r) = P(\{w : \xi(\omega) \leq r\})$$

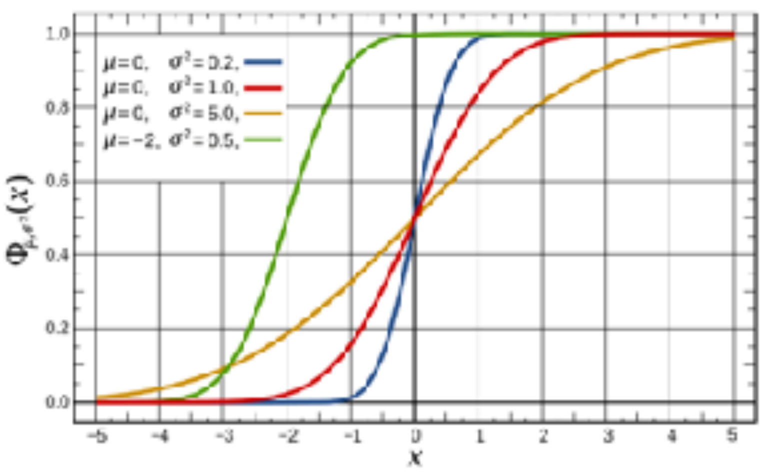
Probability distribution for a discrete random variable $\xi : \Omega \to \mathbb{Z}$:

$$p_\xi(r) = P(\{w : \xi(\omega) = r\})$$

Probability density for a continuous random variable $\xi : \Omega \to \mathbb{R}$:

$$p_\xi(r) = \frac{\partial F_\xi(r)}{\partial r}$$





Mean

A mean (expectation, average …) of a discrete random variable ξ is:

$$\mathbb{E}_P(\xi) = \sum_{\omega \in \Omega} P(\omega) \cdot \xi(\omega)$$

(e.g. the average weight of balls in a bag).

Note:

$$\mathbb{E}_P(\xi) = \sum_r \sum_{\omega:\xi(\omega)=r} P(\omega) \cdot r = \sum_r \left[r \cdot \sum_{\omega:\xi(\omega)=r} P(\omega) \right] = \sum_r r \cdot p_\xi(r)$$

Example

Compute the averaged number of a die (obviously $=3.5$).

Compute the averaged squared number of a die.

$$\mathbb{E}(x) = \frac{1}{6} \cdot (1 + 4 + 9 + 25 + \ldots) = 15\frac{1}{6}$$

Note: $\mathbb{E}(x^2) \neq \mathbb{E}^2(x)$

Consider two independent dice. Compute the probability distribution and the mean for the sum of the two dice numbers. What do you think — is the probability distribution uniform ?

The probability distribution is not uniform anymore:

$$p_\xi \propto (1,2,3,4,5,6,5,4,3,2,1)$$

The mean value is $\mathbb{E}(x) = 7$

6	7	8	9	10	11	12
5	6	7	8	9	10	11
4	5	6	7	8	9	10
3	4	5	6	7	8	9
2	3	4	5	6	7	8
$\xi_2 = 1$	2	3	4	5	6	7
	$\xi_1 = 1$	2	3	4	5	6

Note:

$$\mathbb{E}(\xi_1 + \xi_2) = \sum_\omega P(\omega)\big(\xi_1(\omega) + \xi_2(\omega)\big) = \mathbb{E}(\xi_1) + \mathbb{E}(\xi_2)$$

Random variables of higher dimensions

To be short :-)

Joint probability distribution (for two randon variables x and y):

$$p_{xy}(r,s) = P(\{\omega : (x=r) \wedge (y=s)\})$$

We will use mainly shorthand just $p(x,y)$

Marginal probability distribution ("sum-rule"):

$$p(x) = \sum_y p(x,y)$$

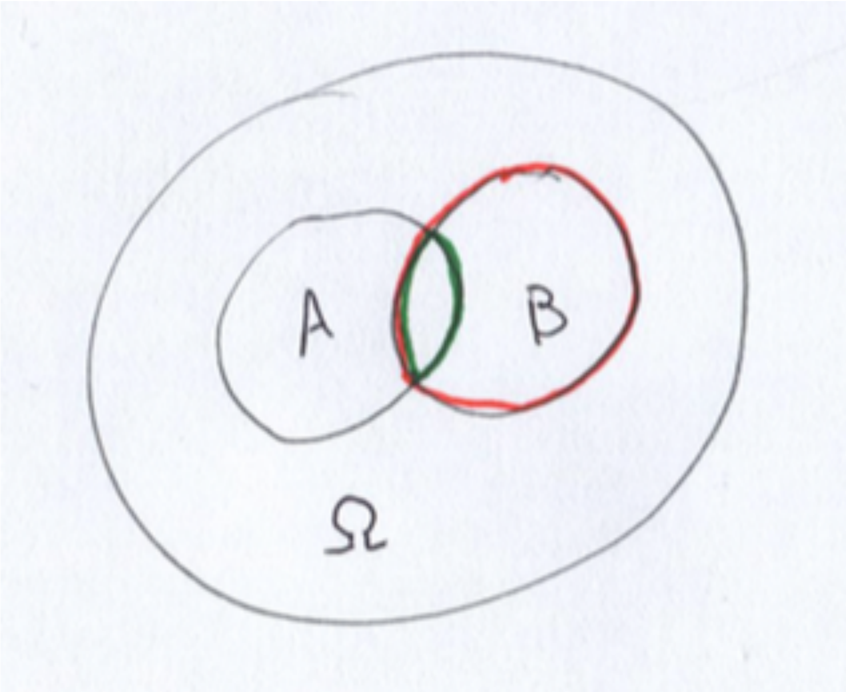
Two variables are independent if

$$p(x,y) = p(x) \cdot p(y)$$

Further definitions

Conditional probability distribution:

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(x,y)}{\sum_{x'} p(x',y)}$$



Product rule:

$$p(x,y) = p(y) \cdot p(x|y) = p(x) \cdot p(y|x)$$

Bayes' rule (formula, theorem …)

$$p(y|x) = \frac{p(x|y) \cdot p(y)}{p(x)}$$

Example

Let the probability to be taken ill be

$$p(ill) = 0.02$$

Let the conditional probability to have a temperature in that case is

$$p(temp|ill) = 0.9$$

However, one may have a temperature without any illness, i.e.

$$p(temp|\bar{ill}) = 0.05$$

What is the probability to be taken ill provided that one has a temperature?

Example

Bayes' rule:

$$p(ill|temp) = \frac{p(temp|ill) \cdot p(ill)}{p(temp)} =$$

$$\text{(marginal probability in the denominator)}$$

$$= \frac{p(temp|ill) \cdot p(ill)}{p(temp|ill) \cdot p(ill) + p(temp|i\bar{l}l) \cdot p(i\bar{l}l)} =$$

$$= \frac{0.9 \cdot 0.02}{0.9 \cdot 0.02 + 0.05 \cdot 0.98} \approx 0.27$$

— not so high as expected :-),

the reason – very low **prior** probability to be taken ill

"Recognition" models

Back to Machine Learning:

The **model**: two variables are usually present:

- the first one is typically discrete $k \in K$ and is called "class"

- the second one is often continuous $x \in X$ and is called "observation"

The **recognition** (inference) task:
Let the joint probability distribution $p(x,k)$ be "given".
Observe x, estimate k.

The (statistical) **learning** task: given a training set

$$L = \big((x_1,k_1),(x_2,k_2),\ldots,(x_l,k_l)\big)$$

"find" the corresponding probability distribution $p(x,k)$