preliminaries

probability theory

Burak Galip ASLAN, PhD

probability theory

set theoretic point of view

probability distributions are defined on **sets of elements** that represent **events** *(e.g. the possible outcomes of rolling a die)*

outcomes of an experiment are often coded by using a random variable (statistical/stochastic variable) which is a function from one sample space to another space

events

all possible outcomes of an experiment are given by the sample space Ω and each event A is a subset of Ω.

a probability measure/function/distribution is a function from events to {0,1} interval.

events are sets so combinations of events can be represented by union(∪) or intersection (∩).

propositions

in the Probabilistic Network community, statistical variables are taken to be functions from one sample space to another sample space

For a set of variables V defined on outcomes «true» and «false», possible value assignments are;

for $V = \{V_1,\ldots,V_n\}$, $n \geq 1$

V=true or V=false

$v_i \equiv V_i$=true $\neg v_i \equiv V_i$=false

all compound propositions are constructed from these by applying the binary operators adhering to axioms of propositional logic

binary operators: $\wedge$ (conjunction) $\vee$ (disjunction)

unary operator: $\neg$ (negation)

joint probability distribution

the Probabilistic Network community approaches probability theory from an algebraic point of view by **associating probabilities with logical propositions** instead of sets

restricting discussion to ***binary variables*** *in text for ease of exposition*

joint probability distribution

let V be a set of statistical variables

let x be a logical proposition

let γ be the Boolean algebra of propositions spanned by V

let Pr: $\gamma \rightarrow [0,1]$ be a function such that;

$\Pr(x) \geq 0$, for all $x \in \gamma$, and
Pr(False)=0, Pr(True)=1

$\Pr(x \vee y) = \Pr(x) + \Pr(y)$, for all $x, y \in \gamma$
such that $x \wedge y \equiv$ False

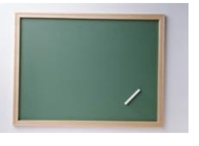
Pr is called a *joint probability distribution* on V for each $x \in \gamma$ the function value $\Pr(x)$ is termed
«the probability of x».

certainty

a probability Pr(*x*) for a logical proposition ***x*** expresses the amount of **certainty** concerning the truth of ***x***

Pr: certainty concerning the truth of an event i.e. the *probability of the truth* of the proposition asserting the occurence of the event

degree of truth is something different than the *probability of truth*.



strictly positive distribution

there are no points of complete certainty in «strictly positive» a.k.a. «non-extreme» probability distributions.

i.e. a strictly positive distribution cannot take values 0 and 1.

i.e. Pr is strictly positive if

$$\Pr(x)=0 \text{ implies } x \equiv \text{False}$$

overview

the concept of conditional probability

the concept of independence of propositions

chain rule

marginalisation property

conditioning property

Bayes' Theorem

the concept of independence of variables

conditional probability

let V be a set of statistical variables

let x be a logical proposition

let γ be the Boolean algebra of propositions spanned by V

let Pr be a joint probability distribution on V

for each $x,y \in \gamma$ with $\Pr(y)>0$, the conditional probability of x given y is denoted as $\Pr(x|y)$ and defined as:

$\Pr(x|y) = \Pr(x \wedge y) / \Pr(y)$

i.e. the amount of certainty on the truth of x given that <u>y is known with certainty</u>.

conditional probability

Pr(x|y)=p does <u>**not**</u> mean whenever y is known to be true, the probability of x equals p.

it means; **the probability of *x* equals p if *y* is known and <u>nothing else known may effect the certainty</u> concerning the truth of *x***.

The conditional probability distribution given *y* is denoted as Pr^{y} i.e. it can also be referred as <u>posterior probability distribution</u>.

overview

the concept of conditional probability

the concept of independence of propositions

chain rule

marginalisation property

conditioning property

Bayes' Theorem

the concept of independence of variables

independence of propositions

***(mutual)* independence:**

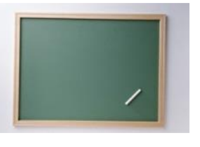
two propositions $x,y\in\gamma$ are called *(mutually)* independent in Pr **if** $\Pr(x \wedge y)=\Pr(x).\Pr(y)$ **else** x and y are called dependent in Pr.

independence of propositions

<u>***(mutual)* independence:**</u>

two propositions $x,y \in \gamma$ are called *(mutually)* <u>independent</u> in Pr **if** $\Pr(x \wedge y) = \Pr(x).\Pr(y)$ **else** x and y are called <u>dependent</u> in Pr.

<u>**conditional independence:**</u>

two propositions $x,y \in \gamma$ are called <u>conditionally independent</u> given the proposition $z \in \gamma$ in Pr **if** $\Pr(x \wedge y|z) = \Pr(x|z).\Pr(y|z)$ **else** x and y are called <u>conditionally dependent</u> given the proposition $z \in \gamma$ in Pr. 

configuration template

A collection of variables $C_{\mathbf{w}}$ is called a configuration template of W set where $W \subseteq V$, and it is shown as $C_{\mathbf{w}} = \wedge_{V_{\mathbf{i}} \in W} V_{\mathbf{i}}$ of all variables in W instead of writing explicity:

$W_{\mathbf{1}}$=value $\wedge$ $W_{\mathbf{2}}$=value $\wedge$ … $\wedge$ $W_{\mathbf{m}}$=value

$C_{\mathbf{w}}$, $W=(W_{\mathbf{1}},\ldots,W_{\mathbf{m}})$ explains all possible assignments to the variables in W, so any configuration $C_{\mathbf{w}}$ of W can be obtained by filling in or updating the necessary values.

For $W=\emptyset$ $C_{\mathbf{w}}$=True which will be used later.

overview

the concept of conditional probability

the concept of independence of propositions

chain rule

marginalisation property

conditioning property

Bayes’ Theorem

the concept of independence of variables

chain rule

Let $V=\{V_1,\ldots,V_n\}$ $n\geq 1$ be a set of statistical variables and let Pr be a joint probability distribution on V.

$Pr(C_V)=Pr(V_1 \wedge \ldots \wedge V_n)$
$=Pr(Vn|V_1 \wedge \ldots \wedge V_{n-1}).\ \ldots\ Pr(V_2|V_1).Pr(V_1)$

V_i can be either true(V_i) or false($\neg Vi$)
so there are 2^n equalities.

overview

the concept of conditional probability

the concept of independence of propositions

chain rule

marginalisation property 

conditioning property

Bayes' Theorem

the concept of independence of variables

overview

the concept of conditional probability

the concept of independence of propositions

chain rule

marginalisation property

conditioning property 

Bayes' Theorem

the concept of independence of variables

overview

the concept of conditional probability

the concept of independence of propositions

chain rule

marginalisation property

conditioning property

Bayes' Theorem 

the concept of independence of variables

overview

the concept of conditional probability

the concept of independence of propositions

chain rule

marginalisation property

conditioning property

Bayes’ Theorem

the concept of independence of variables

independence of variables

let $X,Y,Z \subseteq V$

the set of variables X is *conditionally independent* of the set of variables Y given the set of variables Z **if**

$$\Pr(C_{\mathbf{x}} | C_{\mathbf{y}} \wedge C_{\mathbf{z}}) = \Pr(C_{\mathbf{x}} | C_{\mathbf{z}})$$

i.e. ***once information about Z is available, information about Y is irrelevant w.r.t. X***.

otherwise

X is called *conditionally dependent* on Y given Z.

references

Probabilistic Reasoning

(Probabilistisch Redeneren)

authors: Linda van der Gaag
Silja Renooij

Fall 2017

Universiteit Utrecht

WIKIPEDIA

Google™