Bayesnetze

Seminar “Kognitive Robotik”

Markus Kaiser

Technische Universität München

12-12-12

Motivation

Wissen ist selten sicher oder vollständig.

Beispiel (Medizinische Diagnose)

Software soll mögliche Ursachen für Beschwerden finden.

- Vorgeschichte
- Symptome
- Testergebnisse

→

- Diagnose
- Therapie

Mögliche Unsicherheiten

- Vorgeschichte unvollständig, Symptome vage
- Symptome und Tests lassen mehrere Alternativen zu
- Das Wissen ist fehlerbehaftet

Definition (Wahrscheinlichkeitsraum)

Ein Wahrscheinlichkeitsraum P besteht aus einer Menge von Elementarereignissen Ω, Ereignissen A und einem Wahrscheinlichkeitsmaß Pr.

- A ist σ-Algebra über Ω
- $\Pr : A \mapsto [0, 1]$
- $P = (\Omega, A, \Pr)$

- $\Pr[\Omega] = 1$
- $A \cap B = \emptyset \Rightarrow \Pr[A \cup B] = \Pr[A] + \Pr[B]$

Definition (Bedingte Wahrscheinlichkeit)

A und B seien Ereignisse mit $\Pr[B] > 0$. Die bedingte Wahrscheinlichkeit von A unter der Bedingung B ist dann

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}$$

Definition (Unabhängigkeit)

Zwei Ereignisse A und B sind unabhängig wenn gilt

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B] = \Pr[A \mid B] \cdot \Pr[B]$$

Satz (Produktsatz)

Für unabhängige Ereignisse A_i gilt

$$\Pr\left[\bigcap_{i=1}^{n} A_i\right] = \prod_{i=1}^{n} \Pr\left[A_i \,\middle|\, \bigcap_{k=1}^{i-1} A_k\right]$$

Also zum Beispiel

$$\Pr[A, B, C] = \Pr[A \mid B, C] \cdot \Pr[B \mid C] \cdot \Pr[B]$$

und

$$\Pr[A \mid B] = \frac{\Pr[A, B]}{\Pr[B]} = {\color{red}\alpha} \cdot \Pr[A, B]$$

Satz (Satz über die totale Wahrscheinlichkeit)

Sind die Ereignisse A_i paarweise disjunkt und möglich, dann gilt für ein Ereignis B mit $\Pr[B] > 0$

$$\Pr[B] = \sum_{i=1}^{n} \Pr[B \mid A_i] \cdot \Pr[A_i]$$

Satz (Satz von Bayes)

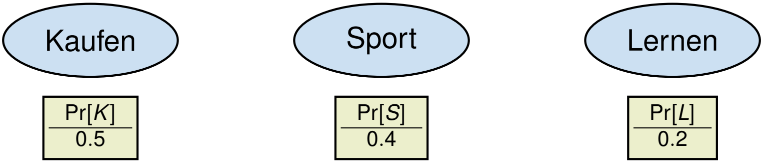
Für zwei Ereignisse A und B mit $\Pr[B] > 0$ *ist*

$$\Pr[A \mid B] = \frac{\Pr[B \mid A] \cdot \Pr[A]}{\Pr[B]}$$

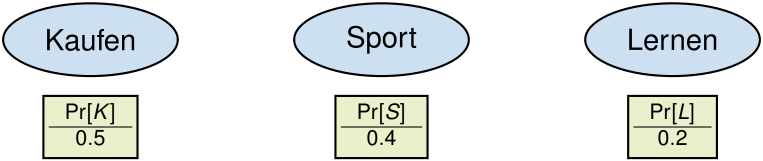
Nachmittagsplanung

Beispiel (Nachmittagsplanung)

Ich habe einen Nachmittag lang Zeit. Ich könnte Lebensmittel kaufen, Sport machen oder Lernen. Ich habe genug Zeit für alle drei. Da ich ein sehr unentschlossener Mensch bin würfle ich mittags mit einem W20.



Nachmittagsplanung



Wie groß ist die Wahrscheinlichkeit,

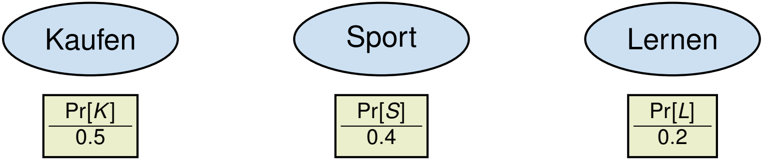
- dass ich lerne, einkaufen gehe, aber keinen Sport mache?

$$\Pr[K, \neg S, L] = \Pr[K] \cdot \Pr[\neg S] \cdot \Pr[L] = 0.5 \cdot 0.6 \cdot 0.2 = 0.06$$

- dass ich nicht lerne, aber Sport mache?

$$\Pr[S, \neg L] = \sum_k \Pr[k] \cdot \Pr[S] \cdot \Pr[\neg L] = \Pr[S] \cdot \Pr[\neg L] = 0.32$$

Nachmittagsplanung



- Alle möglichen Belegungen lassen sich in einer gemeinsamen Verteilung (joint distribution) darstellen.
- Für n binäre Zufallsgrößen hat diese aber 2^n Einträge.

K	S	L	Pr	K	S	L	Pr
F	F	F	0.24	T	F	F	0.24
F	F	T	0.06	T	F	T	0.06
F	T	F	0.16	T	T	F	0.16
F	T	T	0.04	T	T	T	0.04

Beispiel (Erweiterte Nachmittagsplanung)

Ich möchte das Haus nicht verlassen müssen, wenn es regnet. Deshalb halbiere ich bei Regen die Wahrscheinlichkeit für Aktivitäten außer Haus.
Wenn ich weder Sport mache noch lerne bekomme ich ein schlechtes Gewissen. Nach meinem Würfelergebnis möchte ich wissen, wie groß die Wahrscheinlichkeit ist, am Abend schlecht gelaunt zu sein.

Probleme:

- $\Pr[R, S] \neq \Pr[R] \cdot \Pr[S]$
- Alle Sätze brauchen Unabhängigkeit

Es sind immernoch Unabhängigkeiten vorhanden:

- Regen beeinflusst das Lernen nicht
- Einkaufen und Sport beeinflussen sich nicht

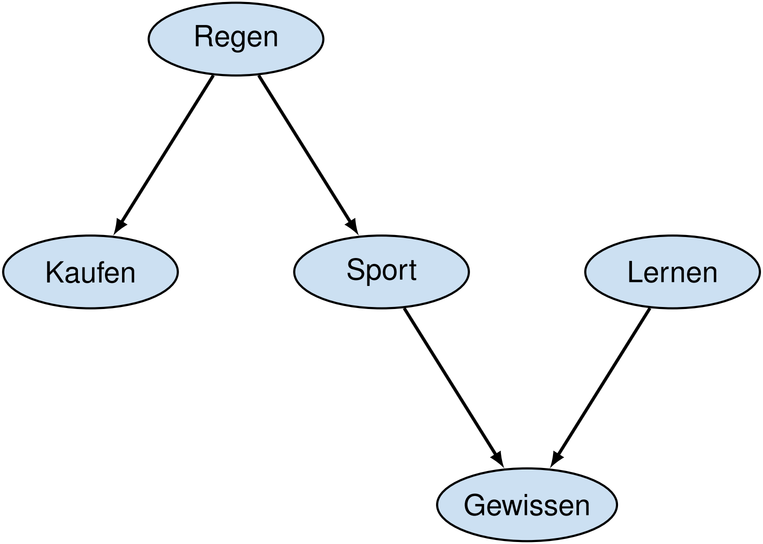
Definition (Bedingte Unabhängigkeit)

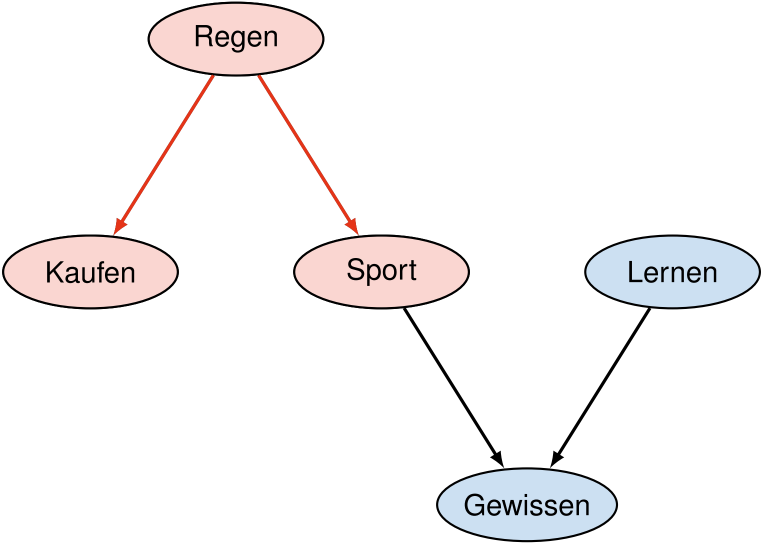
Zwei Ereignisse A und B sind bedingt unabhängig unter einem Ereignis C wenn gilt

$$\Pr[A, B \mid C] = \Pr[A \mid C] \cdot \Pr[B \mid C]$$

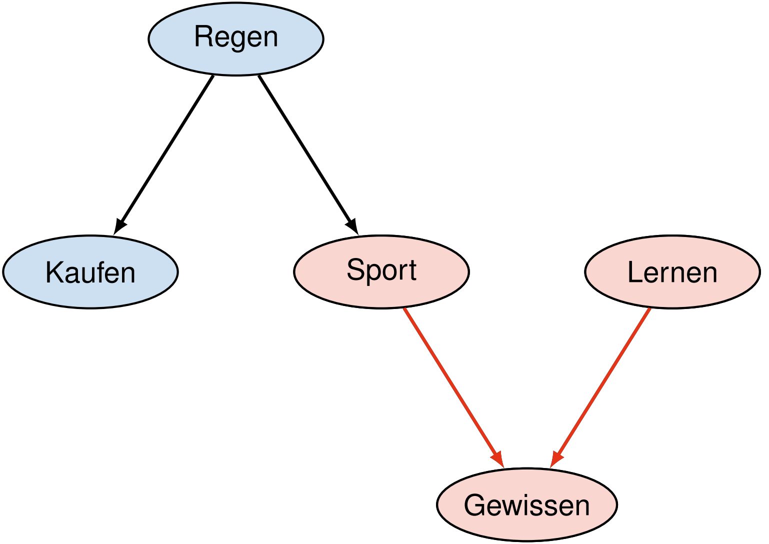
Einkaufen und Sport sind bedingt unabhängig unter Regen.

Bayesnetze

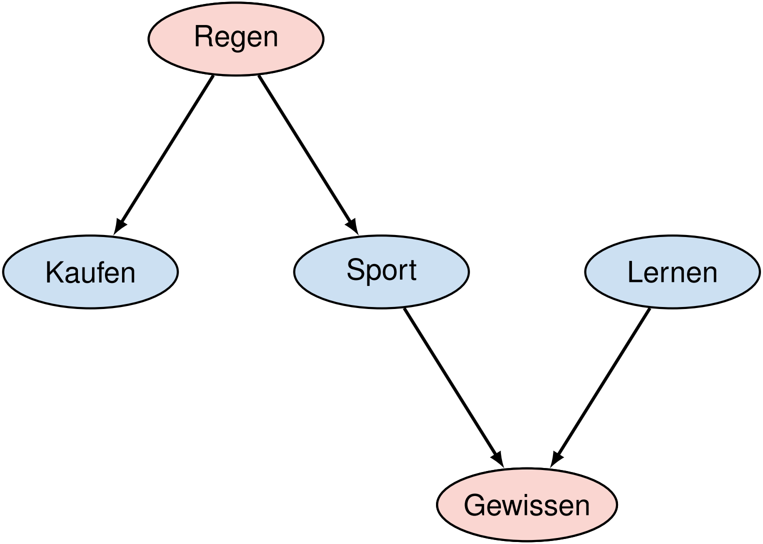



Regen
Kaufen
Sport
Lernen
Gewissen

Bayesnetze



Bayesnetze



Definition (d-Separation)

Zwei Knotenmengen X und Y sind in einem DAG G durch Z d-separiert $\langle X \mid Z \mid Y \rangle_G$ wenn es keinen (ungerichteten) Pfad zwischen X und Y gibt für den gilt

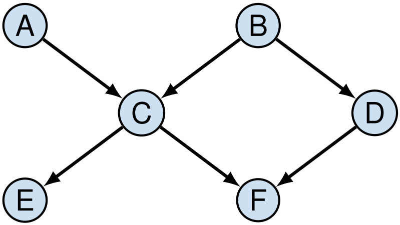
1. Jeder Knoten an dem Kanten zusammenlaufen ist selbst in Z oder hat einen Nachfahren in Z
2. Jeder andere Knoten ist nicht in Z

d-Separation

Definition (d-Separation)

Zwei Knotenmengen X und Y sind in einem DAG G durch Z d-separiert $\langle X \mid Z \mid Y \rangle_G$ wenn es keinen (ungerichteten) Pfad zwischen X und Y gibt für den gilt

1. Jeder Knoten an dem Kanten zusammenlaufen ist selbst in Z oder hat einen Nachfahren in Z
2. Jeder andere Knoten ist nicht in Z

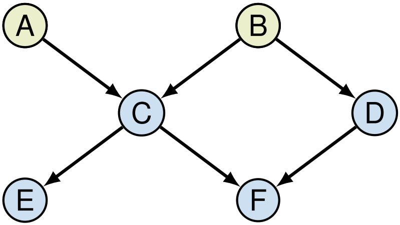


Definition (d-Separation)

Zwei Knotenmengen X und Y sind in einem DAG G durch Z d-separiert $\langle X \mid Z \mid Y \rangle_G$ wenn es keinen (ungerichteten) Pfad zwischen X und Y gibt für den gilt

1. Jeder Knoten an dem Kanten zusammenlaufen ist selbst in Z oder hat einen Nachfahren in Z
2. Jeder andere Knoten ist nicht in Z

$\langle A \mid \emptyset \mid B \rangle$

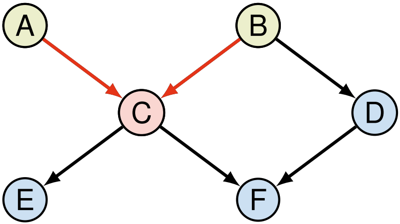


Definition (d-Separation)

Zwei Knotenmengen X und Y sind in einem DAG G durch Z d-separiert $\langle X \mid Z \mid Y \rangle_G$ wenn es keinen (ungerichteten) Pfad zwischen X und Y gibt für den gilt

1. Jeder Knoten an dem Kanten zusammenlaufen ist selbst in Z oder hat einen Nachfahren in Z
2. Jeder andere Knoten ist nicht in Z

$\neg\langle A \mid F \mid B \rangle$

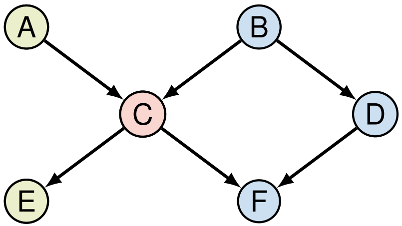


Definition (d-Separation)

Zwei Knotenmengen X und Y sind in einem DAG G durch Z d-separiert $\langle X \mid Z \mid Y \rangle_G$ wenn es keinen (ungerichteten) Pfad zwischen X und Y gibt für den gilt

1. Jeder Knoten an dem Kanten zusammenlaufen ist selbst in Z oder hat einen Nachfahren in Z
2. Jeder andere Knoten ist nicht in Z

$\langle A \mid C \mid E \rangle$



Satz (Unabhängigkeiten in Bayes-Netzen)

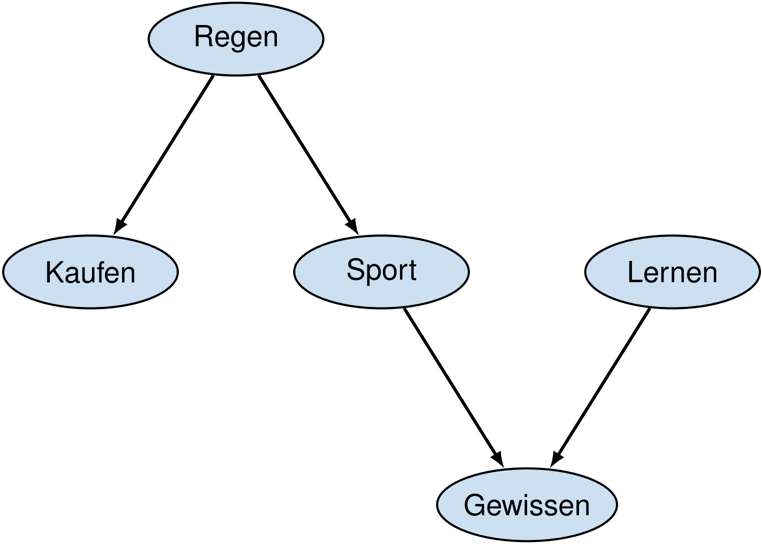
Zufallsvariablen X und Y sind in einem Bayes-Netz bedingt unabhängig unter E wenn $\langle X \mid E \mid Y \rangle$.
Speziell ist jeder Knoten bedingt unabhängig von

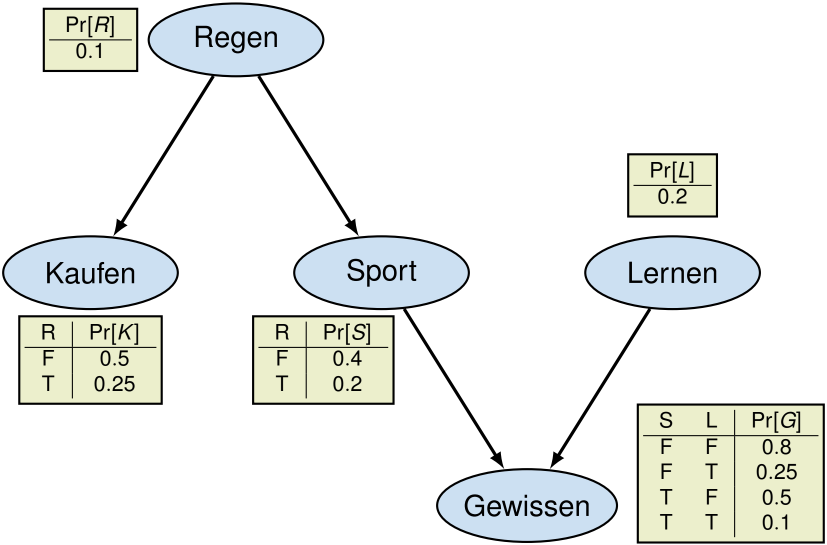
- *allen nicht-Nachfahren gegeben seine Eltern*
- *allen Knoten gegeben seine Markov-Einbettung*

Definition (Markov-Einbettung)

Die Markov-Einbettung (Markov blanket) eines Knotens ist die Menge all seiner Eltern, Kinder und den Eltern aller Kinder.

Bayesnetze





Pr[R]
0.1

Pr[L]
0.2

R	Pr[K]
F	0.5
T	0.25

R	Pr[S]
F	0.4
T	0.2

S	L	Pr[G]
F	F	0.8
F	T	0.25
T	F	0.5
T	T	0.1

Gemeinsame Verteilung

Satz (Gemeinsame Verteilung)

Für einen Eintrag $\Pr[x_1, \ldots, x_n]$ *in der gemeinsamen Verteilung eines Bayesnetzes gilt*

$$\Pr[x_1, \ldots, x_n] = \prod_{i=1}^{n} \Pr[x_i \mid \mathrm{parents}(X_i)]$$

Was ist die Wahrscheinlichkeit, dass ich ein schlechtes Gewissen bei schönem Wetter habe, obwohl ich Sport gemacht habe nachdem ich einkaufen war?

$$\begin{aligned}
\Pr[G, S, \neg L, K, \neg R] &= \Pr[G \mid S, \neg L, K, \neg R] \cdot \Pr[S \mid \neg L, K, \neg R] \cdot \ldots \\
&= \Pr[G \mid S, \neg L] \cdot \Pr[S \mid \neg R] \\
&\quad \cdot Pr[\neg L] \cdot \Pr[K \mid \neg R] \cdot \Pr[\neg R] \\
&= 0.5 \cdot 0.8 \cdot 0.4 \cdot 0.5 \cdot 0.9 = 0.072
\end{aligned}$$

Konstruktion

Wie konstruiert man zu einer Domäne ein möglichst kompaktes, lokal strukturiertes und natürliches Netz?

- Expertenwissen notwendig
- Knoten modellieren quantifizierbare Ereignisse
  - Unbekanntes in Verteilungen
  - Einfachheit vs. Genauigkeit
- Lokale Struktur schaffen
  - Direkte Einflüsse
  - Kleine Einflüsse ignorieren
- Kausale Reihenfolge abbilden

Knotenrepräsentation

- Grundsätzlich beliebige Verteilungen möglich
- Abwägung zwischen effizient und exakt

- Logische Verknüpfungen
- Kanonische Verteilungen

Beispiel (Noisy Or)

Ein Patient hat Fieber, wenn er eine Erkältung oder Grippe hat. Hat er aber eine Krankheit, könnte er auch kein Fieber haben.

$$\Pr[\neg B \mid E, \neg G] = 0.6, \quad \Pr[\neg B \mid \neg E, G] = 0.2$$

E	G	Pr[B]	Pr[¬B]	E	G	Pr[B]	Pr[¬B]
F	F	0	1	T	F	0.4	0.6
F	T	0.8	0.2	T	T	0.88	0.12

Knotenrepräsentation

- Grundsätzlich beliebige Verteilungen möglich
- Abwägung zwischen effizient und exakt

- Logische Verknüpfungen
- Kanonische Verteilungen

Beispiel (Noisy Or)

Ein Patient hat Fieber, wenn er eine Erkältung oder Grippe hat. Hat er aber eine Krankheit, könnte er auch kein Fieber haben.

$$\Pr[\neg B \mid E, \neg G] = 0.6, \quad \Pr[\neg B \mid \neg E, G] = 0.2$$

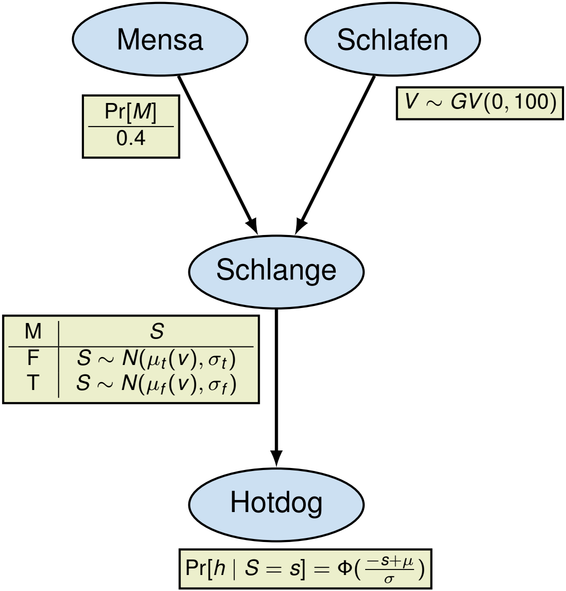
E	G	$\Pr[B]$	$\Pr[\neg B]$	E	G	$\Pr[B]$	$\Pr[\neg B]$
F	F	0	1	T	F	0.4	0.6
F	T	0.8	0.2	T	T	0.88	0.12

Beispiel (Hotdog zum Mittagessen)

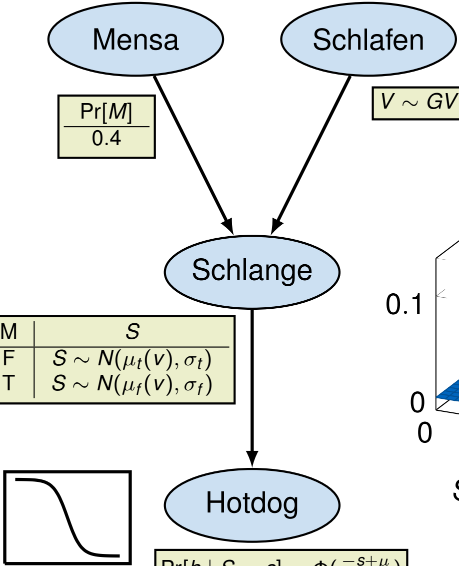
Im MI-Gebäude gibt es jeden Dienstag Hotdogs. Um genau 12 Uhr werfe ich einen Blick auf die Schlange und entscheide allein anhand ihrer Länge ob ich auf mein Mittagessen verzichte.
Die Länge der Schlange ist abhängig davon, ob es in der Mensa akzeptables Essen gibt und wieviele Studenten verschlafen haben.

- Der Anteil der schlafenden Studenten ist gleichverteilt.
- Die Länge der Schlange ist normalverteilt.

Hotdogs


Mensa
Schlafen
$\Pr[M]$
0.4
$V \sim GV(0, 100)$
Schlange
M | S
F | $S \sim N(\mu_t(v), \sigma_t)$
T | $S \sim N(\mu_f(v), \sigma_f)$
Hotdog
$\Pr[h \mid S = s] = \Phi(\frac{-s+\mu}{\sigma})$

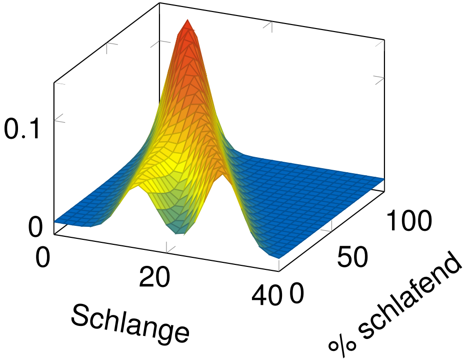
Hotdogs



Pr[M]
0.4

$V \sim GV(0, 100)$

M	S
F	$S \sim N(\mu_t(v), \sigma_t)$
T	$S \sim N(\mu_f(v), \sigma_f)$



$\Pr[h \mid S = s] = \Phi(\frac{-s+\mu}{\sigma})$

Definition (Inferenz)

Für ein Netz mit Knotenmenge $K = \{X\} \cup E \cup Y$ soll die Wahrscheinlichkeit ermittelt werden, dass eine Variable X einen Wert annimmt, unter der Bedingung, dass für die Evidenzvariablen E die Werte e beobachtet werden und die versteckten Variablen Y unbekannt sind.

$$\Pr[X \mid e] = \Pr[X = x \mid E_1 = e_1, \ldots, E_n = e_n] = ?$$

Wenn ich abends ein schlechtes Gewissen habe, wie groß ist die Wahrscheinlichkeit, dass ich nicht gelernt habe?

$$\Pr[\neg L \mid G] \approx 0.936$$

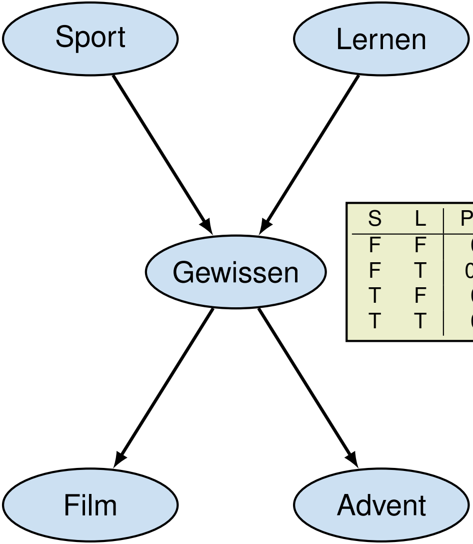
Beispiel (Bekämpfung des schlechten Gewissens)

Wenn ich ein schlechtes Gewissen habe werde ich unglücklich.
Unglück kann man bekämpfen, indem man sich mit einem Film ablenkt oder passend zur Jahreszeit Schokolade aus dem Adventskalender isst.
Zur Vereinfachung soll der Einfluss des Regens nun nicht mehr beachtet werden.

Inferenz durch Aufzählen

Pr[S]
0.4

Pr[L]
0.2



S	L	Pr[G]
F	F	0.8
F	T	0.25
T	F	0.5
T	T	0.1

G	Pr[M]
F	0.4
T	0.75

G	Pr[A]
F	0.6
T	0.9

Beispiel

Wie groß ist die Wahrscheinlichkeit, dass ich ohne Sport einen Film schaue und dabei Schokolade esse?

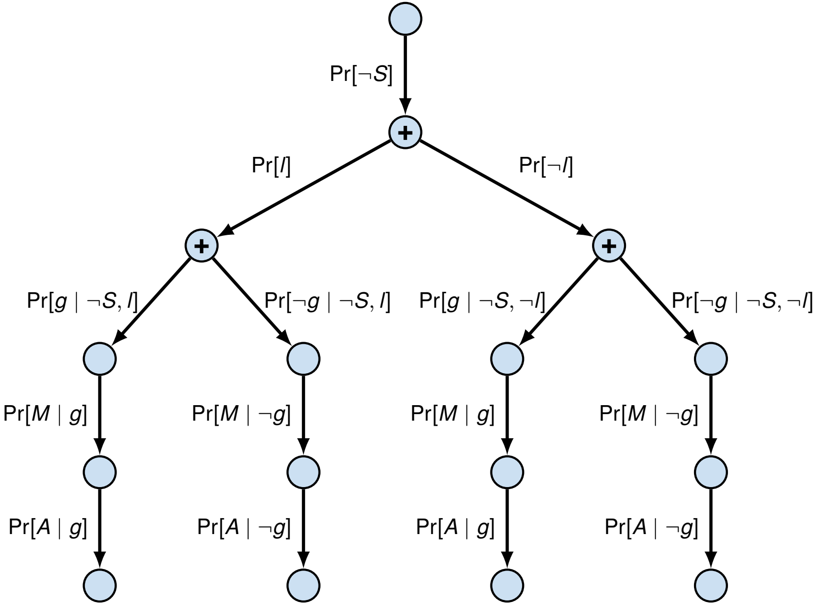
Aus dem Satz der totalen Wahrscheinlichkeit erhalten wir

$$\Pr[X \mid e] = \alpha \Pr[X, e] = \alpha \sum_{y} \Pr[X, e, y]$$

also

$$\begin{aligned}
\Pr[\neg S \mid M, A] &= \alpha \sum_{g} \sum_{l} \Pr[\neg S, M, A, g, l] \\
&= \alpha \cdot \Pr[\neg S] \cdot \sum_{l} \Pr[l] \\
&\quad \cdot \sum_{g} \Pr[g \mid \neg S, l] \cdot \Pr[M \mid g] \cdot \Pr[A \mid g] \\
&\approx 0.585
\end{aligned}$$

Berechnungsbaum



Alternative Algorithmen

- Inferenz durch Aufzählen hat exponentielle Laufzeit
- Mit dynamischer Programmierung linear auf Polytrees
- Aber: Exakte Interferenz ist NP-vollständig

Alternative: Approximative Algorithmen

- Direct/Rejection Sampling
- Likelihood weighting
- Markov chain Monte Carlo (MCMC)

Anwendungen

- Nachmittagsplanung
- Krankheitsdiagnose
- Kriminalitätsbekämpfung
- Genanalyse
- Entscheidungsfindung

Zusammenfassung

Bayesnetze . . .

- repräsentieren mehrdimensionale Zufallsvariablen
- haben eine kompakte Darstellung
- benötigen Expertenwissen
- modellieren Kausalitätszusammenhänge
- sparen Speicher
- erlauben Inferenz