Bemerkung zu probabilistischer Modellierung

Zwischen korrelierten Variablen besteht oft

KEIN

direkter kausaler Zusammenhang!

Beispiele

- Mitarbeiterzufriedenheit und Profit/Mitarbeiter
- Fernsehkonsum und Schulleistung
- Videospielhäufigkeit und Gewaltbereitschaft
- Kirchenbindung und Einkommen
- … und Tausende weitere – schauen Sie in Ihre Zeitung! („Vermischtes“ oder „Wissenschaft“)

Die Verursachung wirkt über andere, verborgene Variablen!

Inferenz über (vollst.) gemeinsame W'keiten

Kommt ein Mann zum Arzt: „Herr Doktor, ich hab so Rückenschmerzen!"

Angenommen, **wir suchen**:

die (vollständige) gemeinsame Posteriori-Verteilung

- der Variablen $\mathbf{Y}$ (Ursache, z.B. Krankheit),
- gegeben Werte $\mathbf{e}$ der Variablen $\mathbf{E}$ („Evidenz", z.B. Symptom)

$$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha \mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}) = \alpha \sum_{\mathbf{h}} \mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}, \mathbf{H}=\mathbf{h})$$

(α Normalisierung, $\mathbf{H}$ „verborgene" (*hidden*) Variablen)

Ginge über Ausrechnen aus der Tabelle der vollständigen gemeinsamen Verteilung, aber …

😐 das ist ein wenig Rechnerei

☹ **wer sagt einem all die $O(m^{|\mathbf{Y}|+|\mathbf{E}|+|\mathbf{H}|})$ W'keitswerte??**

Ausnutzen von Unabhängigkeit

s.o.: Ereignisse a,b sind **unabhängig**, gdw: $P(a \cap b) = P(a) \cdot P(b)$

… also: <u>für unabhängige Variablen</u> A und B gilt:

$\mathbf{P}(A|B) = \mathbf{P}(A)$ oder $\mathbf{P}(B|A) = \mathbf{P}(B)$ oder $\mathbf{P}(A,B) = \mathbf{P}(A)\,\mathbf{P}(B)$

z.B. $\mathbf{P}(Sonnig|Zahnloch) = \mathbf{P}(Sonnig)$
$\mathbf{P}(A,B,C,D) = \mathbf{P}(A,C,D)\,\mathbf{P}(B)$

… also **Reduktion des Aufwandes** beim Rechnen und v.a. beim <u>Repräsentieren:</u>

- wenige kleinere W'keitstabellen (z.B. $\mathbf{P}(A,C,D)$, $\mathbf{P}(B)$)
- statt einer großen für die vollständige gemeinsame Verteilung (z.B. $\mathbf{P}(A,B,C,D)$)

… doch volle Unabhängigkeit ist ein (seltener) Spezialfall!

Bedingte Unabhängigkeit, Beispiel

Bspl: Heiße *Zahnloch*: Ich habe ein Loch im Zahn
Schmerz: Ich habe Zahnschmerzen
Haken: Der Stahlhaken beim Zahnarzt greift

⤷ gemeinsam nicht unabhängig!

Doch es gilt:

$$\mathbf{P}(Haken \mid Schmerz, Zahnloch) = \mathbf{P}(Haken \mid Zahnloch)$$

(ob der Haken greift, hängt nicht vom Zahnschmerz ab)

Haken ist **bedingt unabhängig** von *Schmerz*, **gegeben** *Zahnloch*

entsprechend:

$$\mathbf{P}(Schmerz \mid Haken, Zahnloch) = \mathbf{P}(Schmerz \mid Zahnloch)$$

$$\mathbf{P}(Schmerz, Haken \mid Zahnloch) = \mathbf{P}(Schmerz \mid Zahnloch)\mathbf{P}(Haken \mid Zahnloch)$$

Bedingte Unabhängigkeit, Definition

Zwei Ereignisse a, b sind **bedingt unabhängig**, gegeben c, gdw.:

$$P(a,b \mid c) = P(a \mid c)\, P(b \mid c)$$

Zwei Zufallsvariable A, B sind **bedingt unabhängig**, geg. C, gdw.:

$$\mathbf{P}(A, B \mid C) = \mathbf{P}(A \mid C)\, \mathbf{P}(B \mid C)$$

Damit äquivalent sind die Formulierungen

$$\mathbf{P}(A \mid B,C) = \mathbf{P}(A \mid C) \text{ und } \mathbf{P}(B \mid A,C) = \mathbf{P}(B \mid C)$$

↪ **kleinere Verteilungen ohne volle Unabhängigkeit!!**

Die Bayessche Regel („Satz von Bayes“)

Erinnerung Produktregel: $P(a \cap b) = P(a|b)P(b) = P(b|a)P(a)$

Satz von Bayes

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)} = \alpha \cdot P(a \mid b)P(b)$$

Üblicher Trick: „Verwurste“ Konstante bei **Normalisierung**!

… oder für Verteilungen:

$$\mathbf{P}(Y \mid X) = \frac{\mathbf{P}(X \mid Y)\mathbf{P}(Y)}{\mathbf{P}(X)} = \alpha \cdot \mathbf{P}(X \mid Y)\mathbf{P}(Y)$$

… oder bei vorhandener Evidenz $\mathbf{e}$:

$$\mathbf{P}(Y \mid X, \mathbf{e}) = \frac{\mathbf{P}(X \mid Y, \mathbf{e})\mathbf{P}(Y \mid \mathbf{e})}{\mathbf{P}(X \mid \mathbf{e})}$$

4.2 Bayes-Netze

Diagnose mit bedingten W‘keiten

- Oft sind kausale Zusammenhänge (Ursache ↪ Wirkung) besser bekannt als diagnostische (Wirkung ↪ Ursache)
- Gesucht sind in der Regel aber **Diagnosen** (s. Arzt-Beispiel): gegeben Symptome (Evidenz, Wirkung), nenne mögliche Ursachen
- Daher nutze die Umformung:

$$P(Ursache \mid Wirkung) = \frac{P(Wirkung \mid Ursache) P(Ursache)}{P(Wirkung)}$$

Bedingte Unabhängigkeit + Bayessches Modell

… um W'tabellen weiter zu reduzieren

$$\mathbf{P}(Zahnloch, Schmerz, Haken)$$
$$= \alpha\, \mathbf{P}(Schmerz, Haken \mid Zahnloch)\, \mathbf{P}(Zahnloch)$$
$$= \alpha\, \mathbf{P}(Schmerz \mid Zahnloch)\, \mathbf{P}(Haken \mid Zahnloch)\, \mathbf{P}(Zahnloch)$$

Bed. W 'keit

Domänenwissen: Bed. Unabhängigkeit (s.o.)

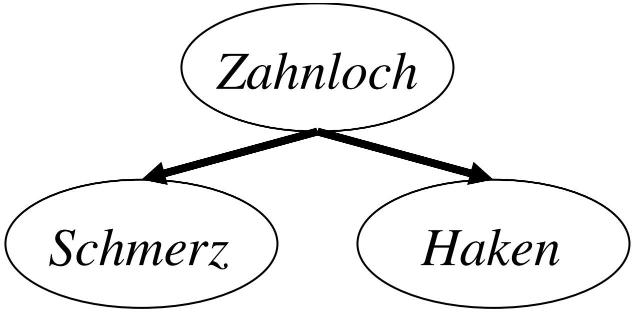
allgemein: **Naives Bayessches Modell**:

$$\mathbf{P}(Ursache, Effekt_1, \ldots, Effekt_n)$$
$$= \alpha\, \mathbf{P}(Ursache) \prod_i \mathbf{P}(Effekt_i \mid Ursache)$$

„Naiv" (= approximativ), wenn die E_i nicht wirklich/ nicht sicher unabhängig sind: **streng genommen falsch**!

Reduziert W'tabellen von $O(m^n)$ auf $O(m \cdot n)$!

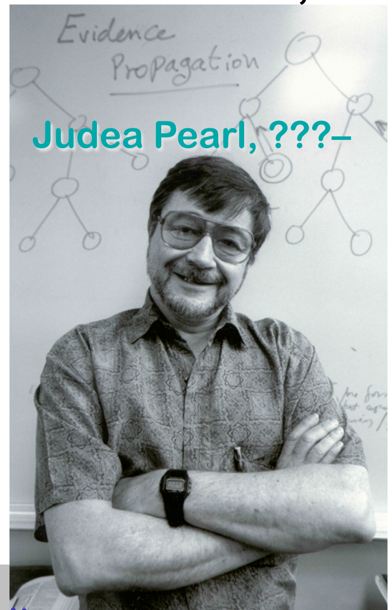
Bayes-Netze



… repräsentieren effizient gemeinsame W'verteilungen und (bzw. mit Hilfe von) Aussagen zur <u>bedingten Unabhängigkeit</u> von ZVn

Ein **Bayes-Netz** ist ein gerichteter azyklischer Graph, wobei

- Knoten entsprechen Zufallsvariablen (diskret, kontinuierlich)
- Kanten entspr. <u>direkten</u> Abhängigkeiten zwischen Variablen; führt Kante von X nach Y, so heißt X ein **Elternknoten** von Y, die Menge aller Elternknoten ist $Parents(Y)$
- Jeder Knoten X trägt Anschrift $\mathbf{P}(X \mid Parents(X))$



Und was soll das?

- Berechnung beliebiger Verteilungen
- basierend auf stark reduzierten W'tabellen durch Nutzung bedingter Unabhängigkeit
- korrekt und praktisch effizient

Beispiel für Kalifornische Wissenschaffende

- Hausalarm geht an bei Einbruch und manchmal bei Erdbeben
- Nachbarn John und Mary sollen im Büro anrufen, wenn sie tagsüber Alarm hören
- John ruft im Büro an, aber nicht Mary. Wie wahrscheinlich ist es, dass gerade eingebrochen wird?

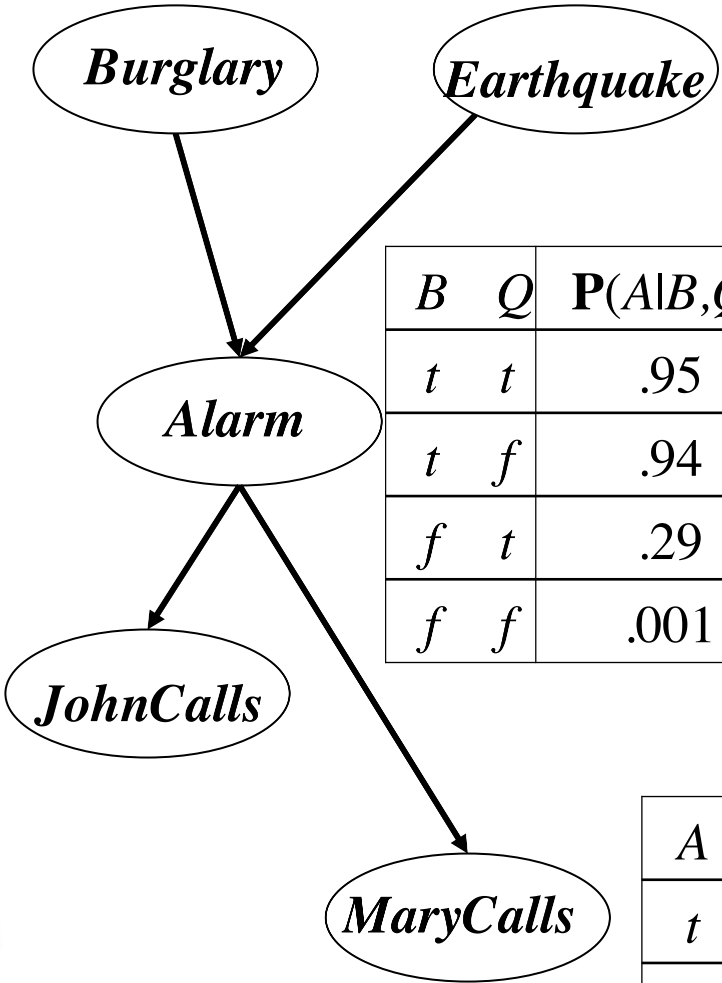
Modellierung

- Variablen *Burglary, EarthQuake, Alarm, JohnCalls, MaryCalls*
- Gesucht ist also $\mathbf{P}(B \mid J{=}t, M{=}f)$, kurz notiert $\mathbf{P}(B \mid j,{-}m)$
- *Alarm* hängt direkt ab von *Burglary, Earthquake*; *JohnCalls* und *MaryCalls* hängen direkt je nur ab von *Alarm*
- W'keiten wie nachfolgend angegeben

Beispiel-Bayes-Netz

P(B)
.001

P(Q)
.002



B	Q	**P**(A|B,Q)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	**P**(J|A)
t	.90
f	.05

A	**P**(M|A)
t	.70
f	.01

Beispiel

$$
\begin{aligned}
&P(j,-m,a,b,-q)\\
&= P(j|a)P(-m|a)P(a|b,-q)P(b)P(-q)\\
&= 0.9\times 0.3\times 0.94\times 0.001\times 0.998\\
&= 0.0002532924
\end{aligned}
$$

Doch eigentlich suchen wir ja
$\mathbf{P}(B \mid j,-m)$! (Fortsetzung folgt)

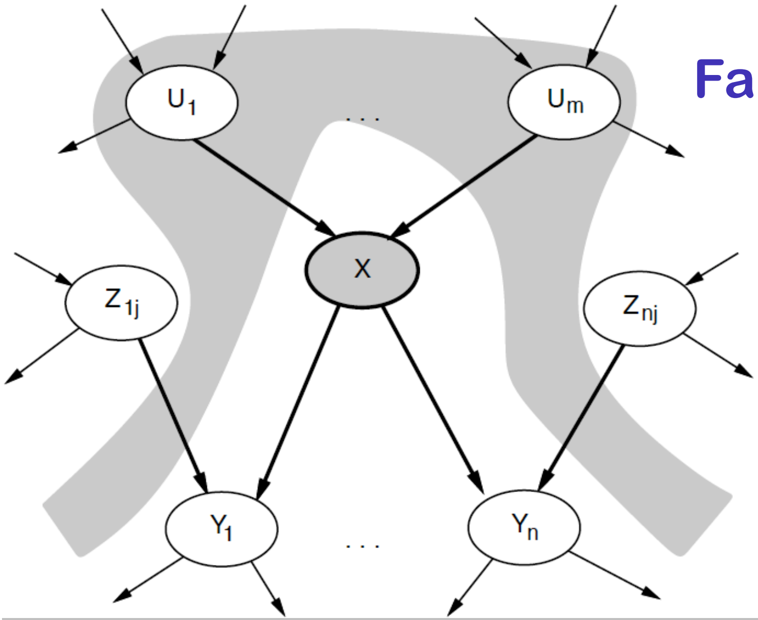
Lokalitätseigenschaften in Bayes-Netzen

Ist es korrekt, gemeinsame W'keit $P(j,-m,a,b,-q)$ über Produkt $P(j|a)P(-m|a)P(a|b,-q)P(b)P(-q)$ lokaler W'keiten auszurechnen?

↪ ja, nach Bayes-Netz-Konstruktion! (Ertel 7.4.7, R./N. 14.2)

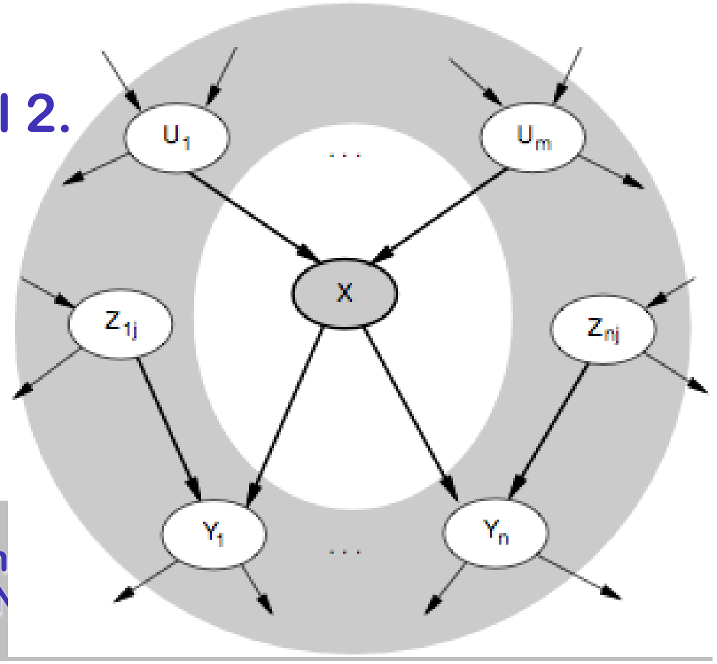
Zusätzlich gilt Satz: Knoten X ist bedingt unabhängig von

1. allen Nicht-Nachfolgern, gegeben seine Eltern
2. allen anderen Knoten, gegeben seine **Markow-Hülle** (die Eltern von X, die direkten Nachfolger von X und deren (andere) Eltern)

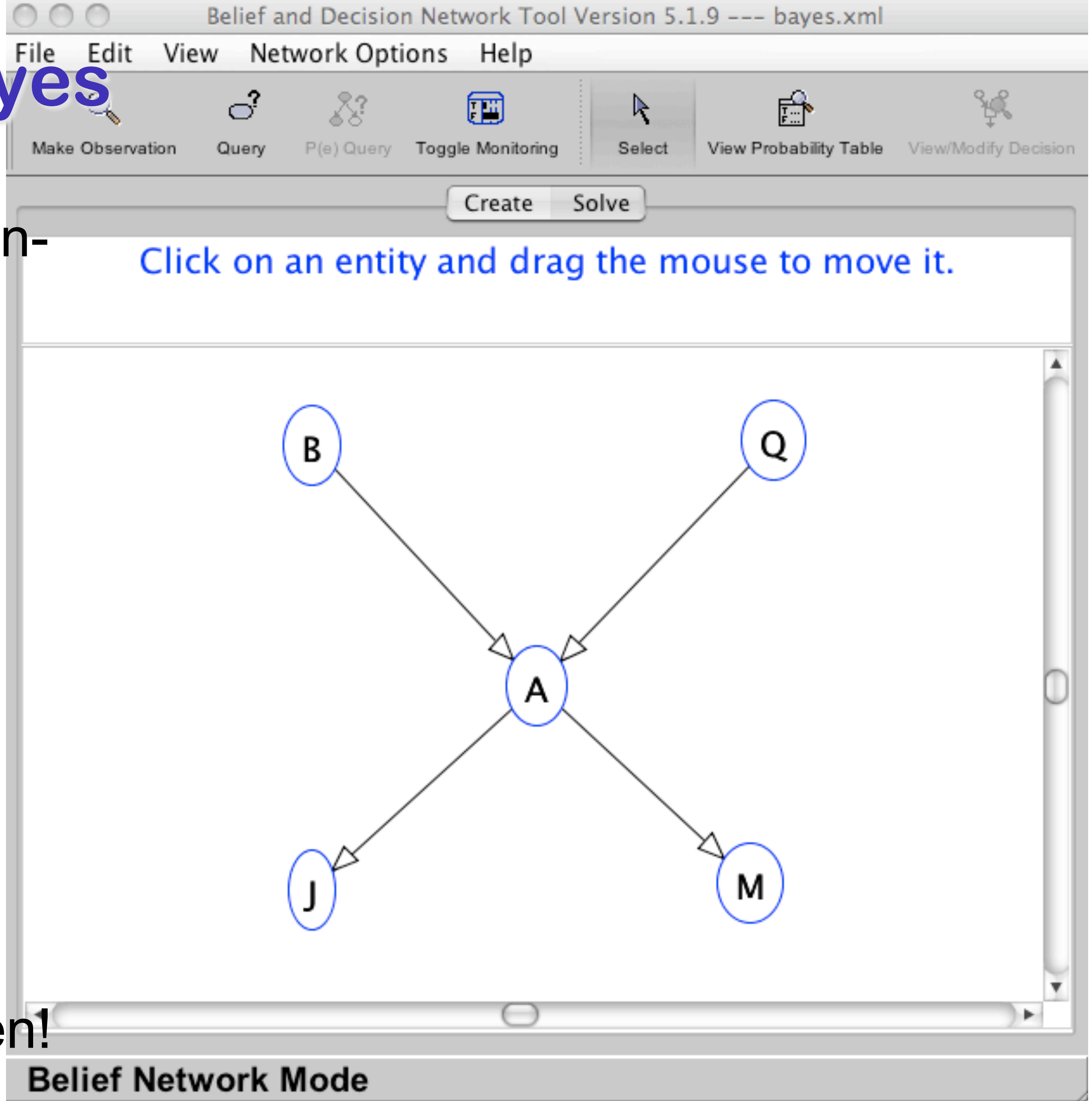


Fall 1.

Fall 2.



aispace.org/bayes



- Tool zum Experimentieren mit kleinen Bayes-Netzen
- Modus 1: Erstellen
- Modus 2: Rechnen
- Möglichkeiten zum Abfragen und Monitoring von Variablen
- Möglichkeit zum Einfügen von „Beobachtungen"
- siehe auch Übungen!