Karlsruhe Institute of Technology

Dynamische Bayes‘sche Netze

Prof. Dr.-Ing. J. Marius Zöllner

Übersicht

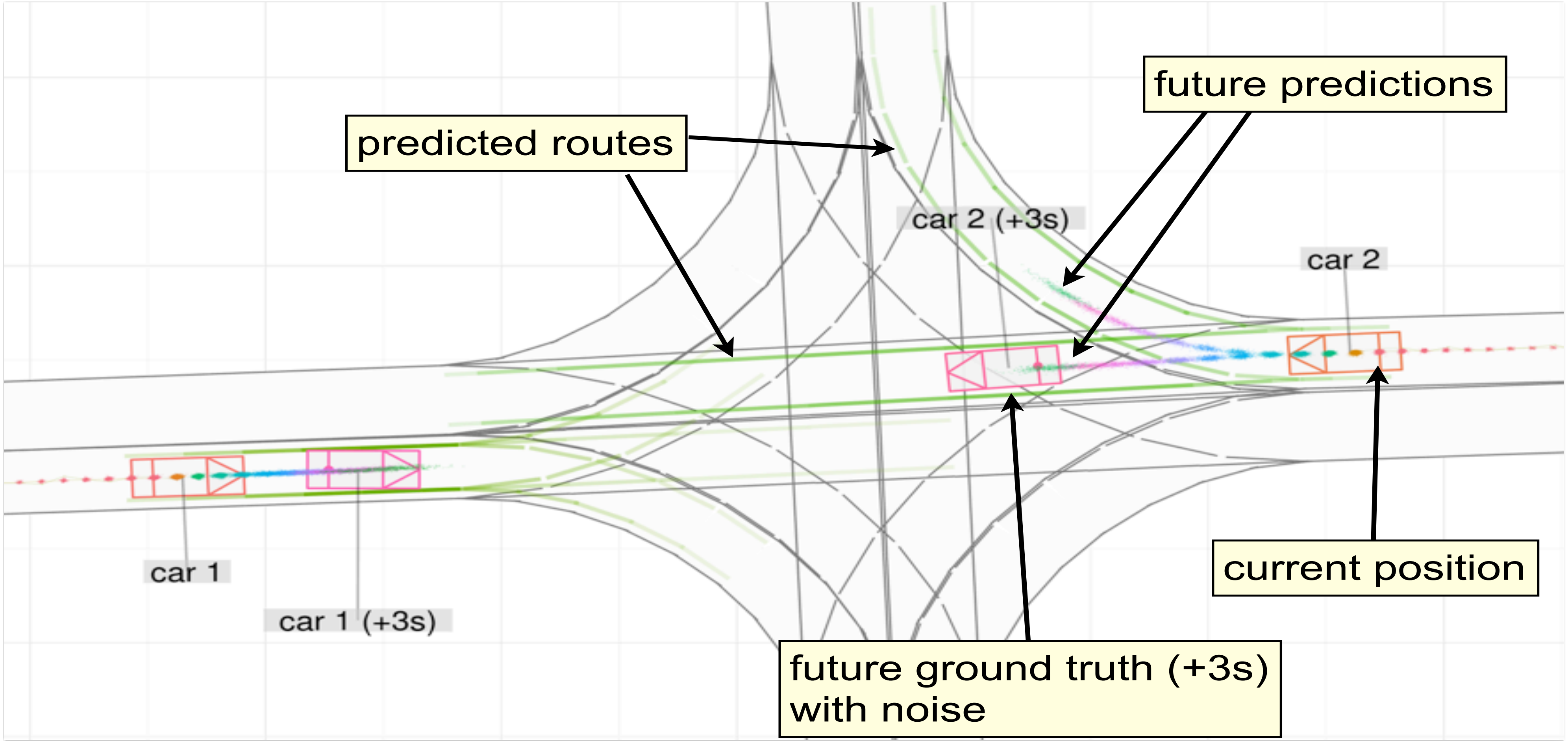
- Motivation
- Probabilistische Graphische Modelle
- Bayes‘sche Netze – Verstehen
  - Bedingte Unabhängigkeit
  - Inferenz
  - .....
- Dynamische Bayes‘sche Netze
- Zusammenhang mit Filtern
- Zusammenfassung

Motivation

- Bsp.-Problem 1: Automatisches Diagnosesystem
  - Patient zeigt Symptome einer Krankheit
  - An welcher Krankheit leidet der Patient?
  - Mit welcher Wahrscheinlichkeit kommen einzelne Krankheiten als Ursache in Frage?
  - ➢ Modellierung als Bayes‘sches Netz

- Bsp.-Problem 2: Fahrzeugverfolgung
  - Lage von Verkehrsteilnehmern fortlaufend verrauscht messbar
  - Wo befinden sich die Verkehrsteilnehmer jetzt?
  - Wie sieht die Verkehrssituation in 10s wahrscheinlich aus?
  - ➢ Modellierung als Dynamisches Bayes‘sches Netz

Bsp. - Problem 2: Fahrzeugverfolgung



[Gindele et al., ITS Magazine 2015]

Wahrscheinlichkeitstheorie

- Multiplikationssatz (Chain Rule)
  - Die Verbundwahrscheinlichkeitsverteilung (Joint Probability Distribution, JPD) mehrerer Zufallsvariablen lässt sich wie folgt faktorisieren:

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

$$P(X_1, X_2, \ldots X_n) = P(X_1) \prod_{k=2}^{n} P(X_k | X_{k-1}, \ldots, X_1)$$

- Theorem der totalen Wahrscheinlichkeit: Für sich gegenseitig ausschließende Ereignisse $A_1, \ldots, A_n$ mit $\sum_{i=1}^{n} P(A_i) = 1$ gilt für ein beliebiges Ereignis B

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i) = \sum_{i=1}^{n} P(B, A_i)$$

Marginalisierung

Theorem von Bayes

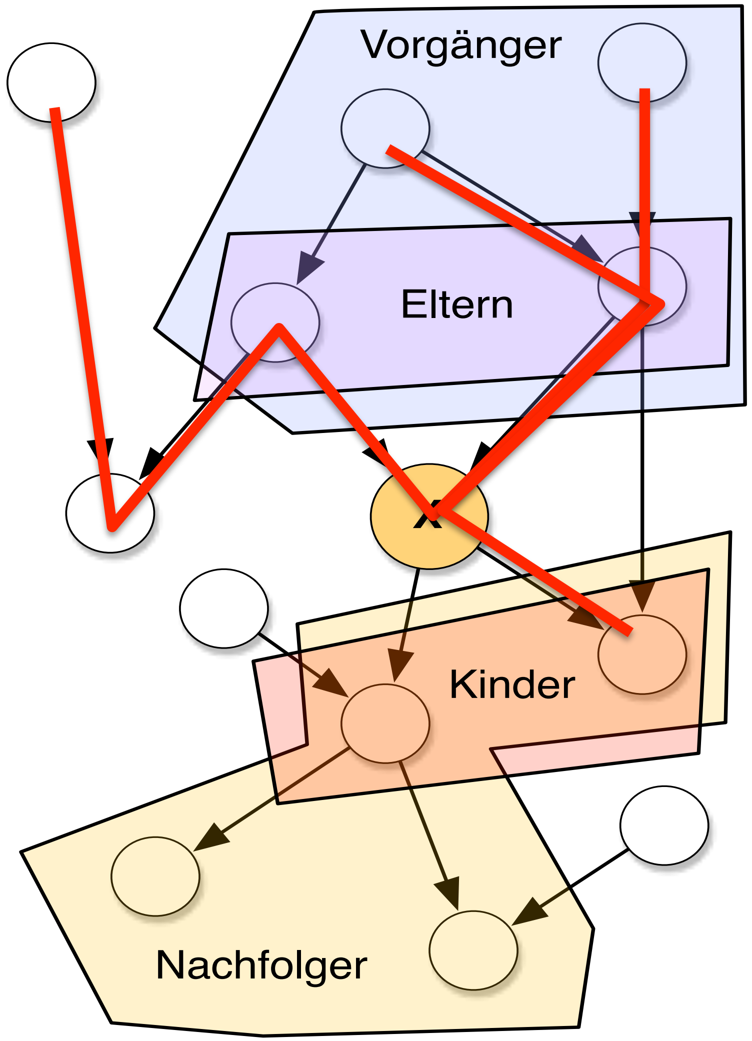
$$\underbrace{P(X|Y)}_{\text{a posteriori}} = \frac{\overbrace{P(Y|X)}^{\text{likelihood}}\overbrace{P(X)}^{\text{a priori}}}{P(Y)}$$

$P(X)$	A priori Verteilung über X (Prior) (z.B. Hypothesen)
$P(Y)$	A priori Verteilung über Y (Prior) (z.B. Daten)
$P(Y\|X)$	Verteilung über Y gegeben X (Likelihood)
$P(X\|Y)$	Verteilung von X gegeben Y (Posterior)

Graphen Theorie (gerichtete Graphen)

- Graph: Knoten + Kanten
- Eltern- / Kindknoten
- Nachfolger- / Vorgängerknoten

- *Weg*: Knotenfolge mit verbindenden Kanten (keine doppelten)
- *Gerichteter Weg*: Weg dessen Kanten nur in eine Richtung gehen
- *Pfad*: Weg ohne doppelte Knoten
- *Zyklus*: Gerichteter Weg mit identischem Start- und Endknoten



Probabilistische Graphische Modelle

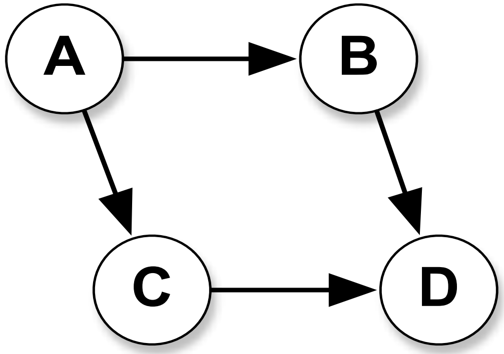
- Michael Jordan, 1998:

> *„Graphical models are a marriage between probability theory and graph theory. [...]. Fundamental to the idea of a graphical model is the notion of modularity -- a complex system is built by combining simpler parts. Probability theory provides the glue whereby the parts are combined, ensuring that the system as a whole is consistent, and providing ways to interface models to data.“*

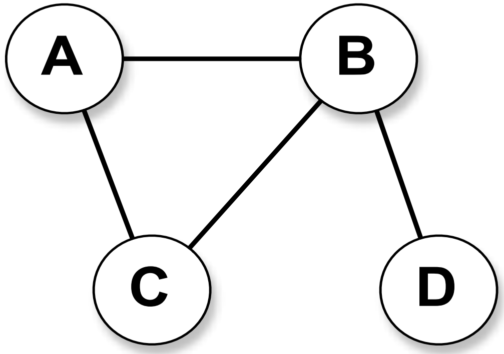
- Probabilistische Graphische Modelle (PGM) beschreiben Verbundwahrscheinlichkeitsverteilungen über Mengen von Zufallsvariablen
- Knoten repräsentieren Zufallsvariablen
- Graph kodiert Abhängigkeiten zwischen Zufallsvariablen

Arten von Graphischen Modellen

- Bayes'sche Netze (BN):
  - Gerichteter azyklischer Graph (DAG)
  - Erlaubt induzierte (bewirkende) Abhängigkeiten
  - Keine zyklischen Abhängigkeiten darstellbar
  - ➢ Geeignet zur Modellierung kausaler Zusammenhänge
  - ➢ Einfache Interpretation bedingter Wahrscheinlichkeitsverteilungen



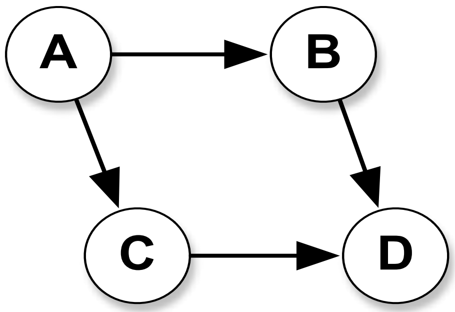
- Markov Random Field (MRF):
  - Ungerichteter Graph
  - Erlaubt zyklische Abhängigkeiten
  - Keine induzierten Abhängigkeiten darstellbar
  - ➢ Geeignet zur Modellierung von Korrelationen



Bayes‘sche Netze

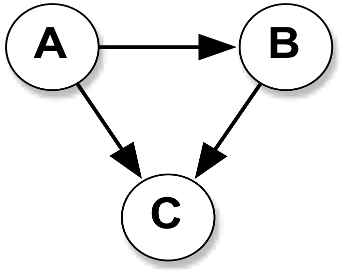
- Faktorisierte Darstellung einer JPD

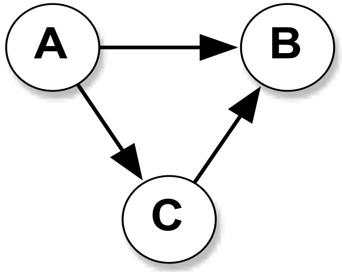
$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i | X_{parents(i)})$$

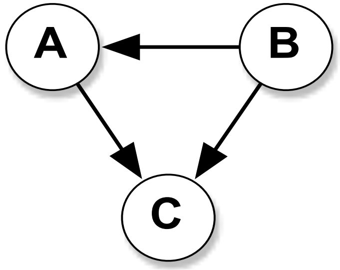


- Kanten stellen direkte Abhängigkeiten dar
- Abwesenheiten von Kanten kodieren bedingte Unabhängigkeiten

- Bsp.: $P(A, B, C)$ Multiplikationssatz liefert:







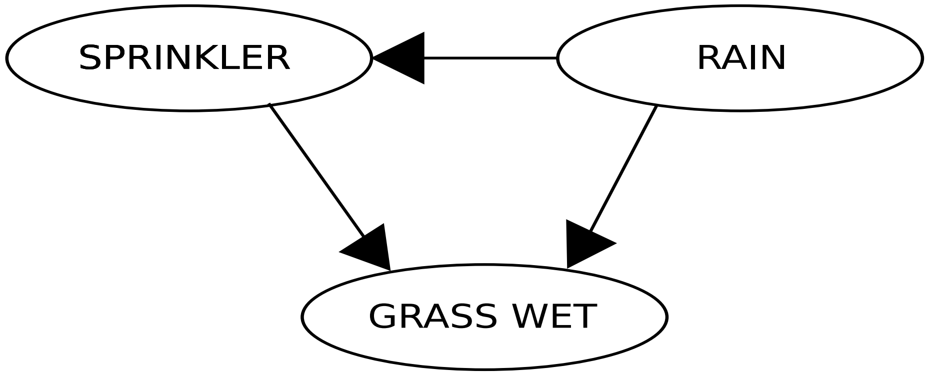
...

$P(A)P(B|A)P(C|B, A)$ $P(A)P(C|A)P(B|C, A)$ $P(B)P(A|B)P(C|A, B)$

Wassersprenger Szenario

$$P(W,S,R) = P(R)P(S|R)P(W|S,R)$$

RAIN	SPRINKLER T	SPRINKLER F
F	0.4	0.6
T	0.01	0.99



RAIN T	RAIN F
0.2	0.8

SPRINKLER	RAIN	GRASS WET T	GRASS WET F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

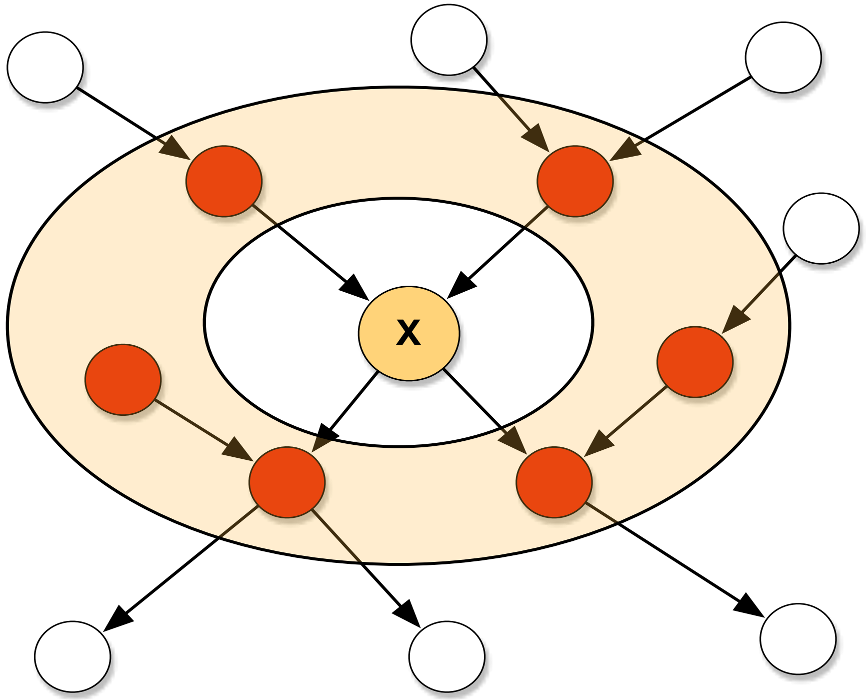
Bsp.: $P(W=t, S=t, R=f) = 0.8 \times 0.4 \times 0.9 = 0.288$

Inferenz und Lernen - Fragestellungen

- Gegeben Struktur + partielle Beispiele wie z.B.: $\{\ (W{=}t,\ S{=}t),\ \ (...),\ ...\}$
- Gesucht sind z.B. Wahrscheinlichkeitstabellen (allgemein Verteilungen)
  → Parameterlernen: Typischerweise über EM – Expectation Maximization
  → Problem: Iteratives Verfahren, benötig in jedem Schritt Inferenz

- Was für Inferenz - Anfragen sind, wie möglich?
  - Wie geht man „effizient“ mit BN um?

- Wichtige Aspekte dabei z.B.:
  - Reduzierung auf relevante Teile des Netzes hilfreich
    - Schnelle Inferenz
    - Reduzierung von Parametern
  - Wann ist ein Netz gut aufgebaut?
  - ...



Bedingte Unabhängigkeit

- Unabhängigkeit $(A \perp B)$:

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

  - Bsp: Würfeln mit zwei Würfeln

- Bedingte Unabhängigkeit $(A \perp B|C)$:

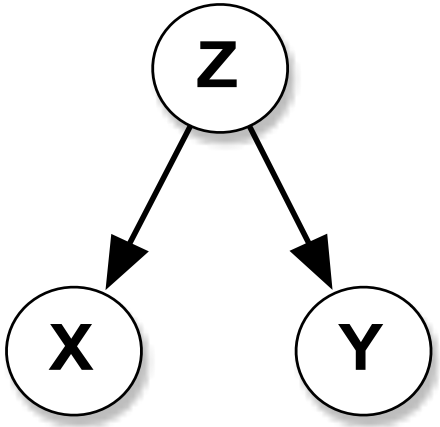
  $$P(A,B|C) = P(A|B,C)P(B|C) = P(A|C)P(B|C)$$

- Muss für alle Belegungen entsprechender Zufallsvariablen A,B und C gelten

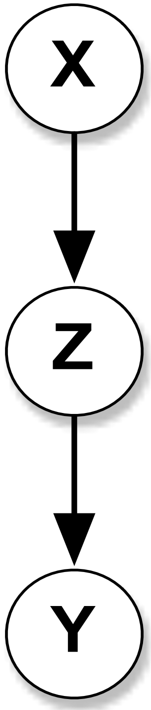
- Unabhängigkeiten können beim probabilistischen Schlussfolgern (Inferenz) und Lernen ausgenutzt werden (s. F. 32 ff)

Bedingte Unabhängigkeit in BNs

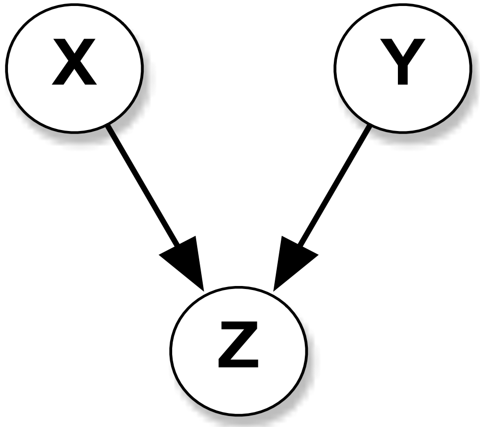
- Betrachtung von 3 Fällen
  - Gemeinsame Ursache
  - Indirekter Effekt
  - Gemeinsamer Effekt



(1) Gemeinsame Ursache



(2) Indirekter Effekt

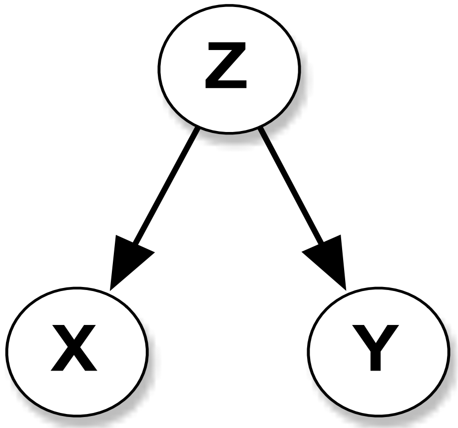


(3) Gemeinsamer Effekt

Fall 1: Gemeinsame Ursache

- JPD: $P(X,Y,Z) = P(Z)P(X|Z)P(Y|Z)$
- Untersuchung ob X und Y unabhängig sind, wenn keine Variable beobachtet wird
  - Marginalisierung nach Z und Nutzung der JPD:

$$P(X,Y) = \sum_Z P(Z)P(X|Z)P(Y|Z)$$



  - Faktorisiert im Allg. nicht zu $P(X)P(Y)$
  - ➢ $X \not\perp Y | \emptyset$
  - Bedeutet, dass X und Y nicht bedingt unabhängig sind gegeben der leeren Menge $\emptyset$

Fall 1: Gemeinsame Ursache

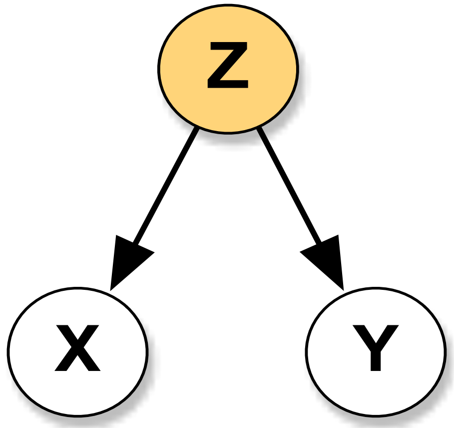
- JPD: $P(X,Y,Z) = P(Z)P(X|Z)P(Y|Z)$

- Untersuchung ob X und Y unabhängig sind, wenn Z beobachtet wird
  - Allgemeiner Multiplikationssatz

$$P(X,Y|Z) = \frac{P(X,Y,Z)}{P(Z)}$$
$$= P(X|Z)P(Y|Z)$$

➢ $(X \perp Y|Z)$

  - X und Y sind bedingt unabhängig gegeben Z
  - Wenn die gemeinsame Ursache bekannt ist, verrät X nichts über Y und umgekehrt



Fall 2: Indirekter Effekt

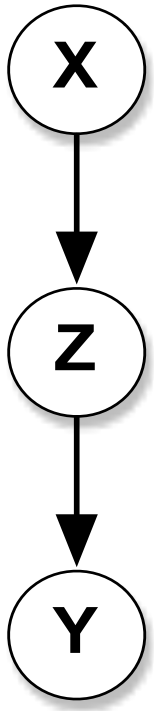
- JPD: $P(X,Y,Z) = P(X)P(Z|X)P(Y|Z)$

- Untersuchung ob X und Y unabhängig sind, wenn keine Variable beobachtet wird
  - Marginalisierung nach Z und Nutzung der JPD:

$$P(X,Y) = P(X)\sum_{Z} P(Z|X)P(Y|Z) = P(X)P(Y|X)$$

  - Faktorisiert im Allg. nicht zu $P(X)P(Y)$

➢ $X \not\perp Y | \emptyset$



Fall 2: Indirekter Effekt

- JPD: $P(X,Y,Z) = P(X)P(Z|X)P(Y|Z)$

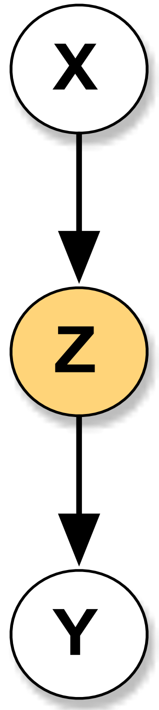
- Untersuchung ob X und Y unabhängig sind, wenn Z beobachtet wird

$$\begin{aligned} P(X,Y|Z) &= \frac{P(X,Y,Z)}{P(Z)} \\ &= \frac{P(X)P(Z|X)P(Y|Z)}{P(Z)} \\ &= P(X|Z)P(Y|Z) \end{aligned}$$

Satz von Bayes

➢ $(X \perp Y | Z)$

- X und Y sind bedingt unabhängig gegeben Z
- Z verrät bereits alles über die Verteilung von X und Y

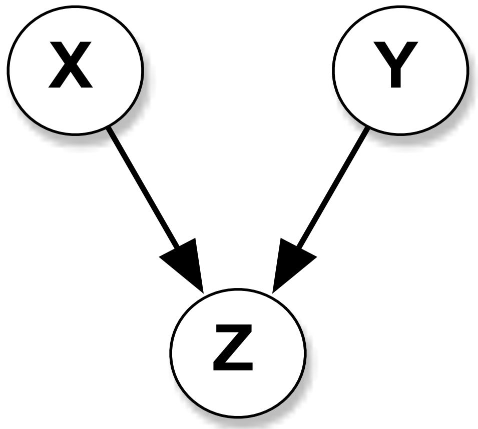


Fall 3: Gemeinsamer Effekt

- JPD: $P(X,Y,Z) = P(X)P(Y)P(Z|X,Y)$

- Untersuchung ob X und Y unabhängig sind, wenn keine Variable beobachtet wird
  - Marginalisierung nach Z und Nutzung der JPD:

$$P(X,Y) = P(X)P(Y)\sum_Z P(Z|X,Y) = P(X)P(Y)$$



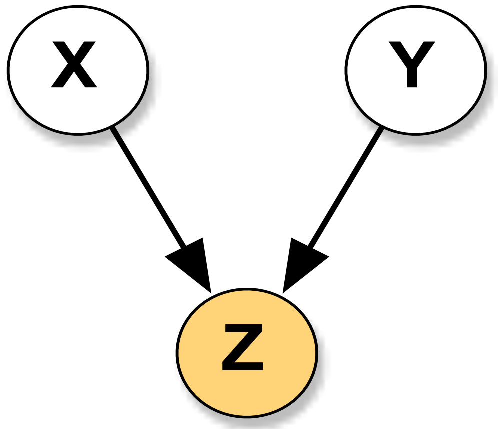
➢ $(X \perp Y | \emptyset)$

  - Bedeutet, dass X und Y bedingt unabhängig sind gegeben der leeren Menge $\emptyset$, im Unterschied zu Fall 1 und 2

Fall 3: Gemeinsamer Effekt

- JPD: $P(X,Y,Z) = P(X)P(Y)P(Z|X,Y)$

- Untersuchung ob X und Y unabhängig sind, wenn Z beobachtet wird

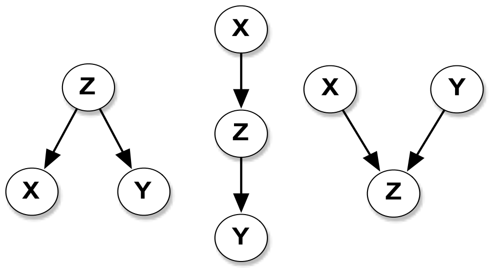


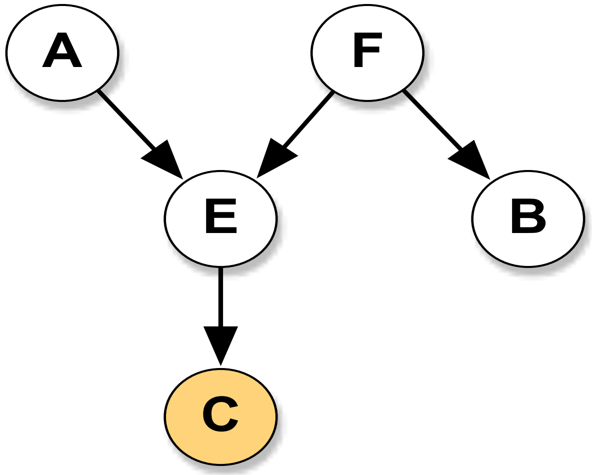
$$P(X,Y|Z) = \frac{P(X,Y,Z)}{P(Z)}$$
$$= \frac{P(X)P(Y)P(Z|X,Y)}{P(Z)}$$

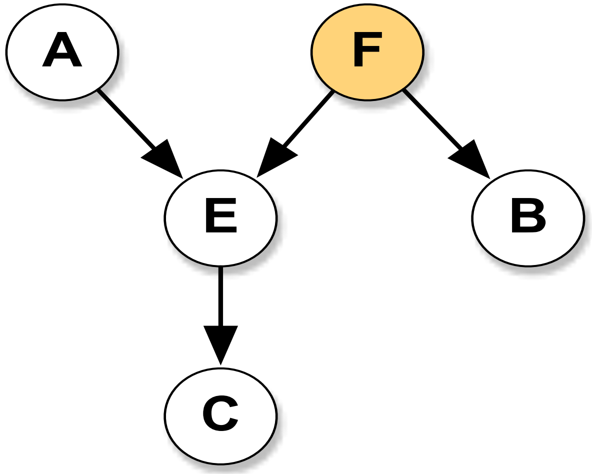
  - ➢ $(X \not\perp Y | Z)$
  - X und Y sind **nicht** bedingt unabhängig gegeben Z
  - Da sowohl X als auch Y den beobachteten Effekt Z beeinflussen, liefert X Information über Y gegeben Z und umgekehrt
  - Genau umgekehrt wie in Fall 1 und 2

D-Separation – „Blockieren der Abhängigkeit“

- Pfad zwischen einem Knoten A und einem Knoten B ist *blockiert durch beobachtete Knoten C bzw. F*, wenn für einen Pfadknoten
  - Fall 1 oder Fall 2 vorliegt und ein Pfadknoten C bzw. F ist
  - Fall 3 vorliegt und weder ein Pfadknoten noch einer seiner Nachfolgeknoten C bzw. F ist







→ $(A \not\perp B | C)$ $(A \perp B | F)$

D-Separation (allgemeine Definition)

- A,B und C sind disjunkte **Knotenmengen** eines gerichteten, azyklischen Graphen (DAG)

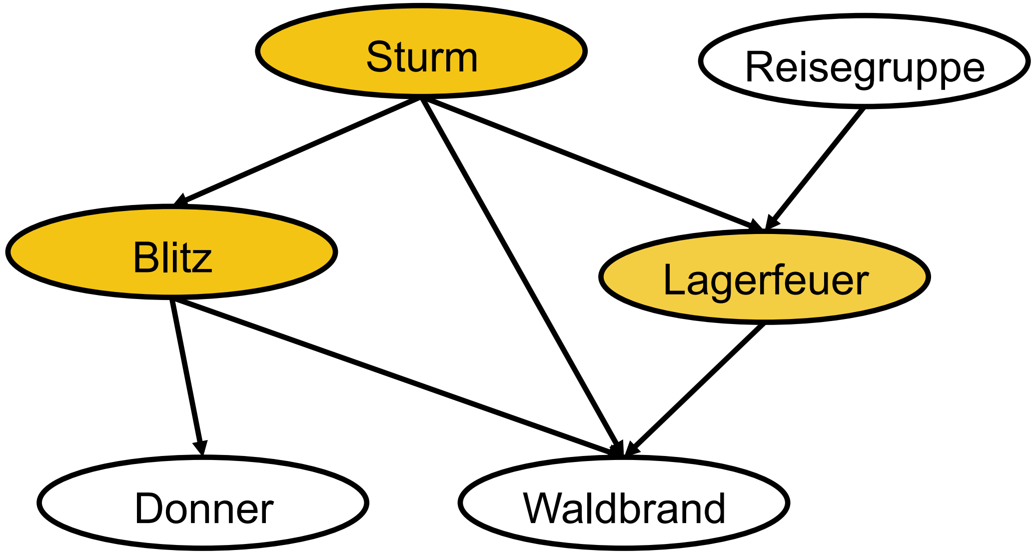
- Pfad zwischen einem Knoten aus A und einem Knoten aus B ist *blockiert*, wenn für einen Pfadknoten
  - Fall 1 oder Fall 2 vorliegt und Pfadknoten in C enthalten ist
  - Fall 3 vorliegt und weder der Pfadknoten noch einer seiner Nachfolgeknoten in C enthalten sind

- Wenn alle Pfade zwischen Knoten aus A und B blockiert sind, dann ist A *d-separiert* von B durch C

➢ Für die JPD gilt dann $(A \perp B | C)$

Übungsaufgabe

- Untersuche bedingte Unabhängigkeiten zwischen R (Reisegruppe) und D (Donner)



Pfade:

$$R \rightarrow L \leftarrow \underline{S} \rightarrow \underline{B} \rightarrow D$$
$$R \rightarrow L \leftarrow \underline{S} \rightarrow W \leftarrow \underline{B} \rightarrow D$$
$$R \rightarrow \underline{L} \rightarrow \underline{W} \leftarrow \underline{B} \rightarrow D$$
$$R \rightarrow \underline{L} \rightarrow \underline{W} \leftarrow \underline{S} \rightarrow \underline{B} \rightarrow D$$

Alle Pfade blockiert

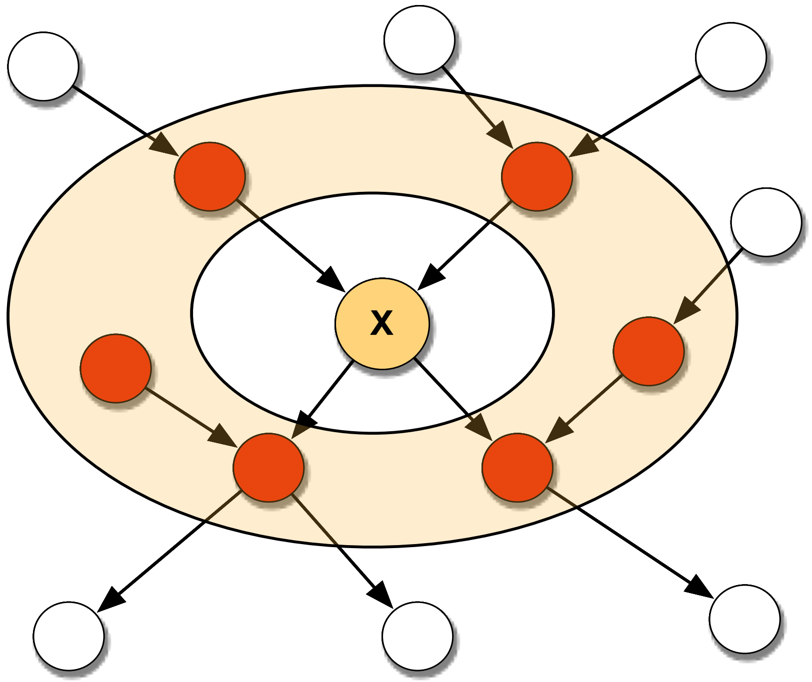
→ $(D \perp R | B, S, L)$

➔ Knoten sind unabhängig von allen Nicht-Nachfolgerknoten wenn die Elternknoten gegeben sind

➔ ....

Markov Blanket

- Minimale Menge an beobachteten Knoten, die notwendig ist um einen Knoten vom Rest des Graphen zu separieren



- Markov Blanket eines Knoten besteht aus dessen Elternknoten, Kindknoten und Elternknoten der Kindknoten

➔ Wichtig um effizient zu Inferieren und zu Lernen

Anfragen an BNs

- Zufallsvariablen (RV): $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$
  - Gegeben: BN und damit JPD: $P(X_1, ..., X_n)$
  - ggf. Evidenz: Instanziierungen der Zufallsvariablen $E \subseteq \mathcal{X}$ mit e

- Wahrscheinlichkeitsanfragen
  - Anfrage Variablen: $X, Y \subseteq \mathcal{X}$ mit und ohne Evidenzen z.B.
    - $P(X)$ (ohne Evidenzen)
    - $P(X \mid Y, E=e)$ (mit Evidenzen)
    - $P(Y \mid E=e)$ (auch Posterior genannt)

  - Maximum a-posteriori (MAP) Anfragen

$$\mathrm{MAP}(Y|e) = \operatorname*{argmax}_y P(y|e)$$

Inferenz in BNs

- Inferenz **ohne** Evidenz (Marginalisierung)
  - Gesucht: $P(X)$
  - Gegeben: $P(X,Y)$ (JPD durch BN)

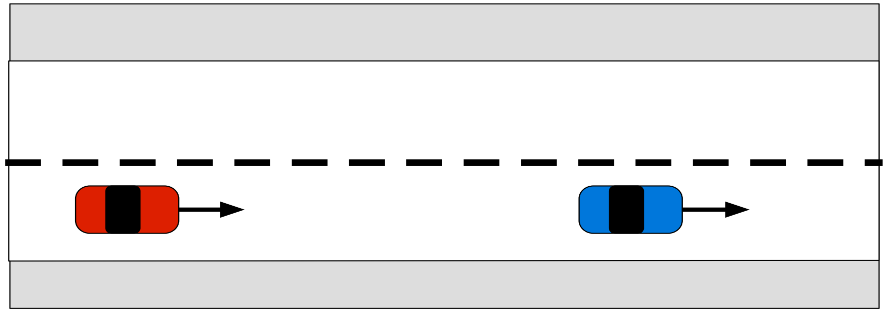
- Naiver Ansatz: Theorem der totalen Wahrscheinlichkeit

$$P(X) = \int_Y P(X,Y)$$

  - Erfordert Integral über alle möglichen Instanziierungen von Y
  - Im diskreten Fall wird das Integral durch eine Summe ersetzt

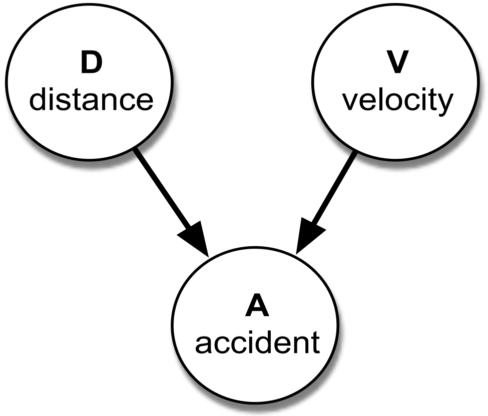
Inferenz in BNs

Bsp.: zwei fahrende Autos



- Variablen:
  - D: Distanz {*nah, weit*}
  - V: Geschwindigkeit {*langsam, mittel, schnell*}
  - A: Unfall {*Unfall, kein Unfall*}
- JPD:

$$P(A, D, V) = P(D)P(V)P(A|D, V)$$



- Mögliche Fragestellung:
  1. Verteilung der Distanzwahrscheinlichkeiten
  2. Verteilung der Geschwindigkeitswahrscheinlichkeiten
  3. Verteilung der Unfallwahrscheinlichkeiten

Inferenz in BNs

- Frage 1 & 2 – „trivial“
  weil D unabhängig ist von V und A ohne Evidenzen (bzw. V unabhängig ist von D und A)

- Frage 3 (Lösung über Marginalisierung):

$$
\begin{aligned}
P(A) &= \sum_d \sum_v P(d, v, A) \\
&= \sum_d \sum_v P(d) P(v) P(A|d, v)
\end{aligned}
$$

  - Ergebnis:

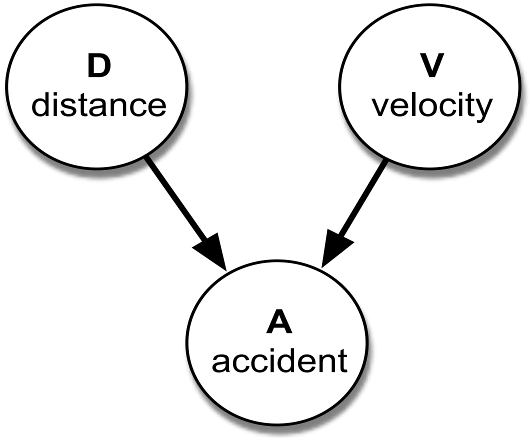
	accident	no accident
P(A)	0.2	0.8

P(D)

	close	far
P(D)	0.3	0.7

P(V)

	slow	average	fast
P(V)	0.2	0.5	0.3



P(A|D,V)

D	V	accident	no accident
close	slow	0.2	0.8
close	average	0.3	0.7
close	fast	0.8	0.2
far	slow	0.0	1.0
far	average	0.1	0.9
far	fast	0.2	0.8

Inferenz in BNs

- Inferenz **mit** Evidenz
  - Gesucht: $P(X|E=e)$
  - Gegeben: $P(X,Y,E)$ durch BN und Evidenz e

- Naiver Ansatz: Theorem der totalen Wahrscheinlichkeit zusammen mit Multiplikationssatz

$$P(X|e) = \int_Y P(X,Y|e) = \int_Y \frac{P(X,Y,e)}{P(e)}$$

  - Aufwand wie vorher nur zusätzliche Normalisierung notwendig

Inferenz in BNs

- Frage 1: $P(intelligent|college)$

$$P(int.|college) = \sum_s \frac{P(int., s, college)}{P(college)}$$

$$\approx \frac{0,52}{0,72} \approx 0,71$$

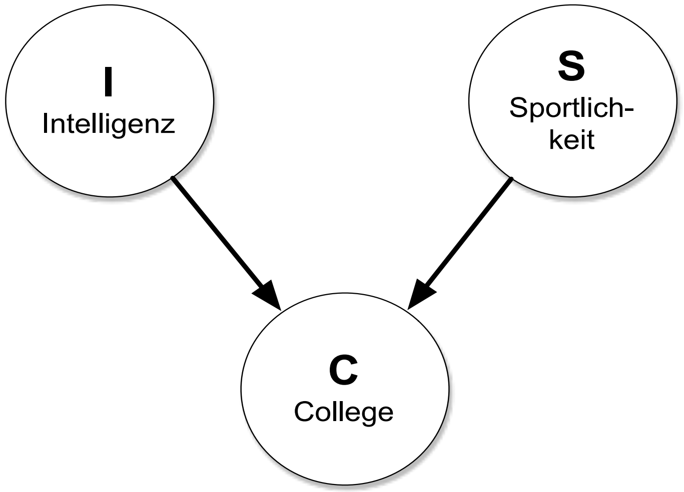
- Frage 2: $P(intelligent|sportlich, college)$

$$P(int.|college, sport.) = \frac{P(int., sport., college)}{P(college, sport.)}$$

$$\approx \frac{0,32}{0,52} \approx 0,63$$

	sportlich	*nicht sportlich*
P(S)	0,6	0,4

	intelligent	*nicht intelligent*
P(I)	0,6	0,4



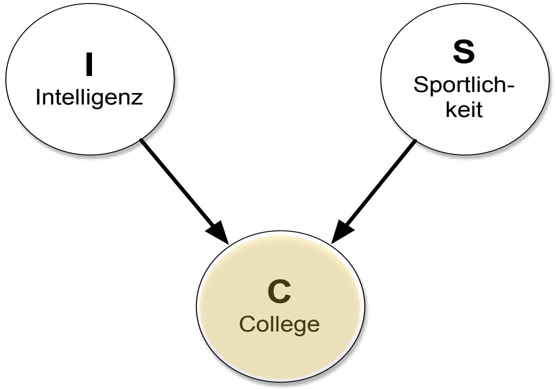
I	S	*college*	*kein college*
intelligent	*sportlich*	0,9	0,1
intelligent	*nicht sportlich*	0,8	0,2
nicht intelligent	*sportlich*	0,8	0,2
nicht intelligent	*nicht sportlich*	0,1	0,9

Explaining Away

- Ergebnisse der Inferenz:

$$P(intelligent|college) \approx 0.71$$
$$P(intelligent|sportlich, college) \approx 0.63$$

➢ Die Beobachtung, dass jmd. auf ein College geht und er sportlich ist, lässt diesen automatisch weniger intelligent erscheinen



- Die Beobachtung einer Ursache, die ausreicht um einen Effekt zu erklären, lässt andere mögliche Ursachen automatisch unwahrscheinlicher werden
  - ➢ Die Beobachtung des gemeinsamen Effekts macht die Ursachen bedingt abhängig (Fall 3)
- Kommt häufig in der Realität vor (z.B. Medizin)

Naive Bayes revisited

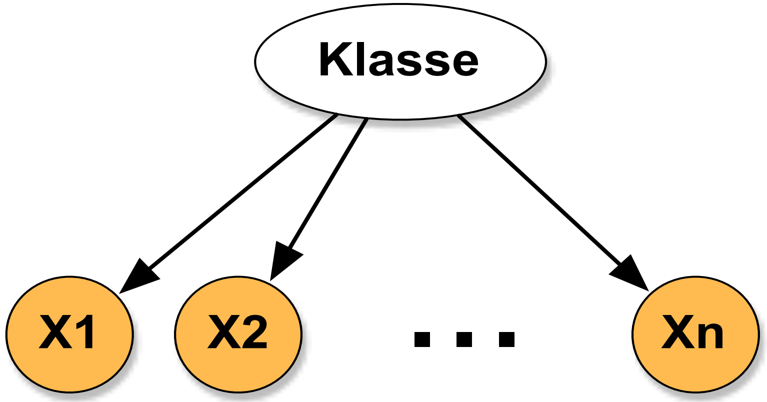
- Klassifikation C anhand von Attributen $X_1, X_2, \ldots, X_n$
- Naive Bayes Annahme:
  - Alle Attribute sind untereinander unabhängig gegeben C, d.h.

$$\forall i : \ (\ X_i \perp \boldsymbol{X}_{\{1 \ldots n\} \setminus i} \mid C \ )$$

- Vereinfachte Formel:

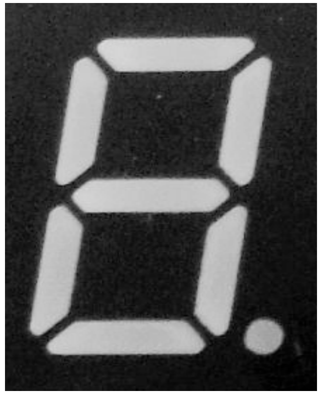
$$P(C|X_1, \ldots, X_n) \propto P(C) P(X_1, \ldots, X_n|C) = \ P(C) \prod_{i=1}^{n} P(X_i|C)$$

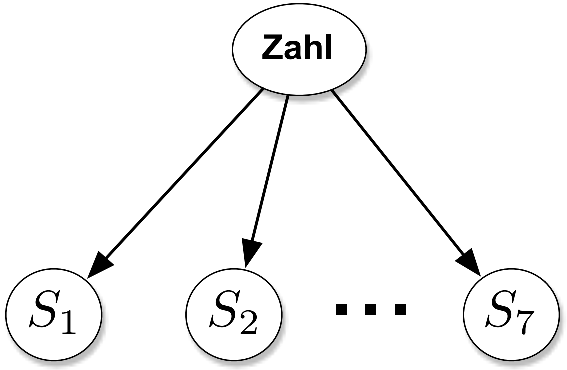
- Bayes‘sches Netz:



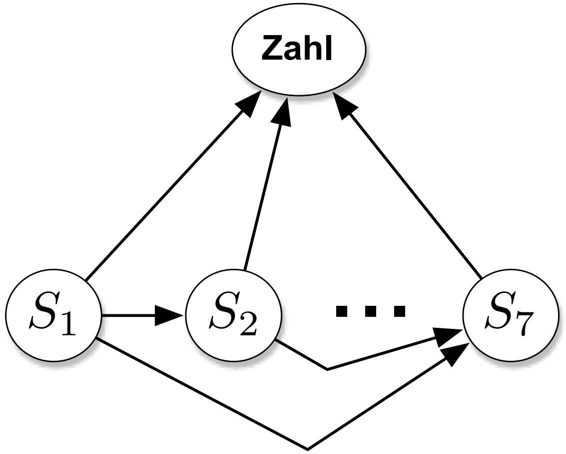
Kausalität in BNs

- Multiplikationssatz erlaubt beliebige Faktorisierungen
- Beziehungen in BNs müssen nicht kausal im Sinne Klasse → Merkmal sein!
- Vorteile kausaler BNs:
  - i.d.R. weniger Kanten
  - i.d.R. einfachere Modelle (weniger Parameter → besser lernbar)
- Bsp.: Modellierung einer digitalen Anzeige





Kausal ☺



Nicht kausal ☹

Inferenz in BNs - Komplexität

- Schlechte Nachricht
  - Exakte Inferenz in allgemeinen BNs ist NP-hart
  - Approximative Inferenz in allgemeinen BNs ist NP-hart

- Gute Nachricht
  - Dies betrifft nur den Worst-Case
  - Inferenzkomplexität in realen BNs oft nur linear oder polynomiell

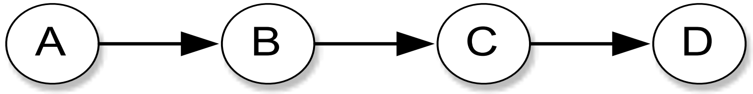
- Bsp. Exakte Inferenz mit diskreten Zufallsvariablen A,B,C,D

$$P(D) = \sum_a \sum_b \sum_c P(a, b, c, D)$$

  - Exponentieller Aufwand $O(k^n)$ bei n Zufallsvariablen mit k Werten

Inferenz in BNs

- Geht oft besser, wenn wir die Struktur des BNs ausnutzen
- Angenommen BN ist ein Kettengraph
- Ohne Struktur („naiver“ Ansatz)



$$
\begin{aligned}
P(D) &= \sum_a \sum_b \sum_c P(a,b,c,D) \\
&= \sum_a \sum_b \sum_c P(a)P(b|a)P(c|b)P(D|c) \qquad\qquad O(k^n) \\
&= \sum_c P(D|c) \sum_b P(c|b) \sum_a P(a)P(b|a)
\end{aligned}
$$

- Aufwandsvergleich bei Nutzung der Struktur
  - Einzelschritt $P(X_{i+1}) = \sum_{x_i} P(X_{i+1}|x_i)P(x_i)$ hat den Aufwand $O(k^2)$
  - Gesamt: $O(nk^2)$
  - Linearer Aufwand im Gegensatz zu exponentiellem

Inferenz in BNs (nur zur Übersicht)

- Verfahren lässt sich verallgemeinern
  - Variable Elimination
  - Conditioning
  - Belief Propagation
- Weiterer Effizienzgewinn durch dynamische Programmierung

- Approximative Inferenzverfahren
  - Loopy Belief Propagation
  - Importance Sampling (z.B. Likelihood Weighting)
  - Markov-Chain Monte Carlo Verfahren (z.B. Gibbs Sampling)

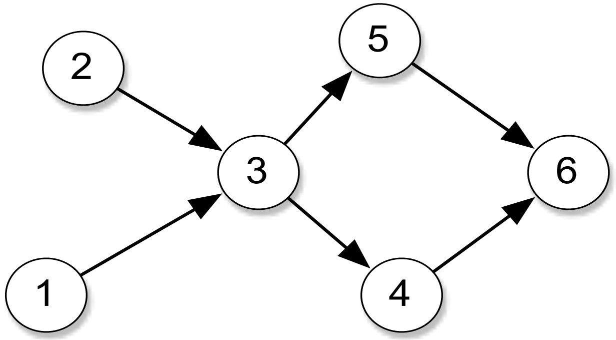
Generatives Modell

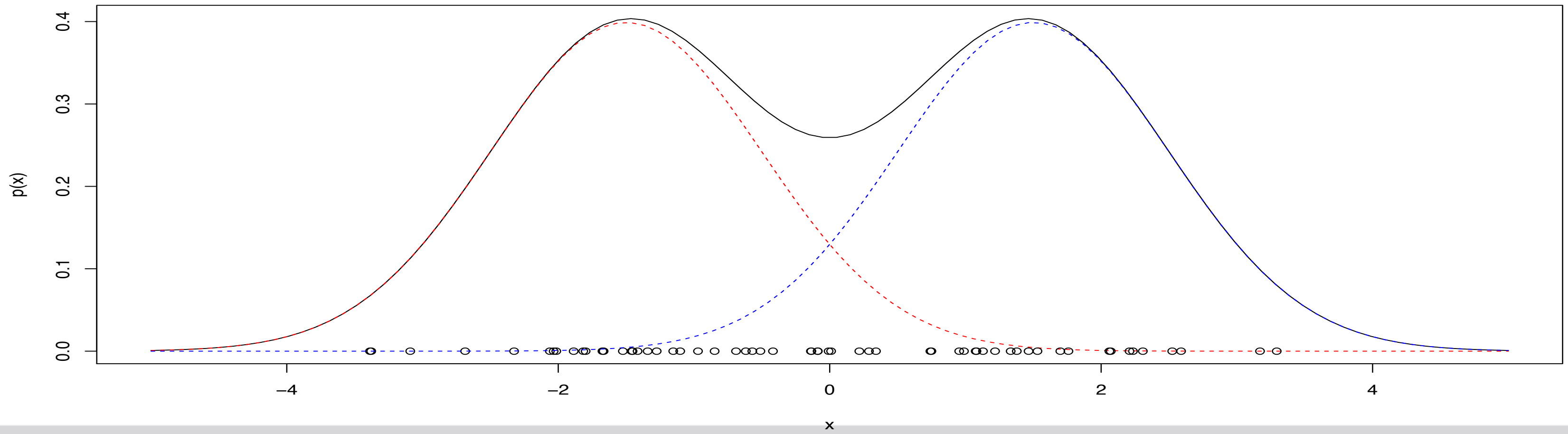
Karlsruhe Institute of Technology

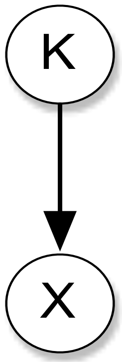
- Jede JPD ist ein generatives Modell, das es erlaubt Stichproben zu ziehen
- Viele Anwendungsgebiete
  - Approximative Inferenz
  - Visualisierung
  - Etc.
- Viele Verfahren existieren
- Oft ist es schwierig direkt aus der JPD zu ziehen, aber einfach Stichproben aus den Faktoren zu erzeugen

Generatives Modell - Ancestral Sampling

- Ancestral Sampling:
  - Knoten des BN topologisch sortieren
  - Werte für Prior-Knoten ziehen
  - Werte für bedingte Knoten anhand der Vorgängerwerte ziehen
- Auch zum Samplen marginaler Verteilungen geeignet
- Bsp.: Stichproben aus einer Gauß-Mixtur erzeugen







Dynamische Bayes‘sche Netze

Dynamische Bayes‘sche Netze

- Motivation
  - Gesucht: Position und Geschwindigkeit eines Fahrzeugs (Zustand)
  - Gegeben
    - Verrauschte Messungen
    - Vorwissen wie sich Fahrzeuge bewegen können
- Modellierung als BN
  - Zustand ändert sich über der Zeit
  - Verwendung von Zufallsvariablen zur Beschreibung der Fahrzeugzustände zu bestimmten Zeitpunkten

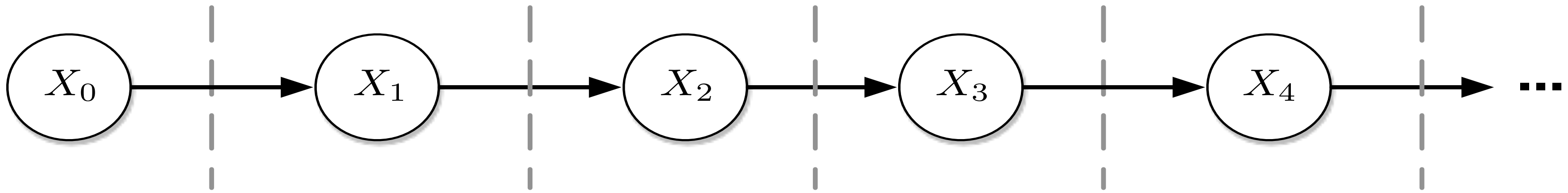
$$P(X_0, X_1, \ldots, X_T)$$

Dynamische Bayes‘sche Netze

- Dynamische Bayes‘sche Netze (DBNs) sind BNs zur Beschreibung dynamischer Prozesse
- Eigenschaften
  - Zustandsraum: Menge aller möglichen Systemzustände
  - Diskrete Zeitschritte zwischen Zustandsübergängen
  - Markov‘sche Annahme: Zukünftige Zustände unabhängig von vergangen Zuständen gegeben den aktuellen Zustand

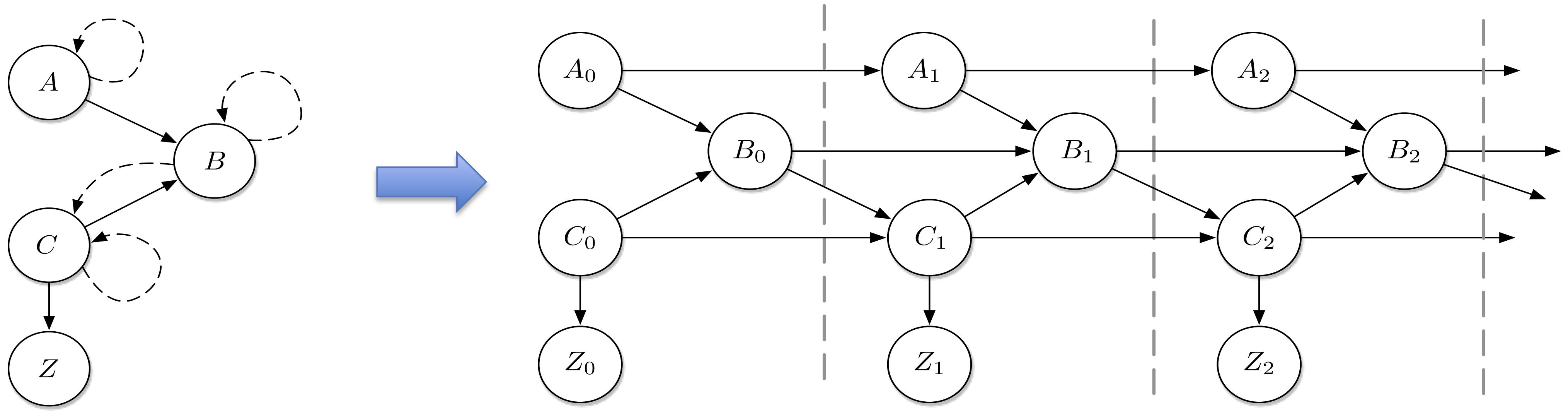
$$(\, X_{t+1} \perp X_{t-1}, \ldots, X_0 \mid X_t)$$

$$\Rightarrow P(X_0, X_1, \ldots, X_T) = P(X_0) \prod_{t=0}^{T-1} P(X_{t+1}|X_t)$$



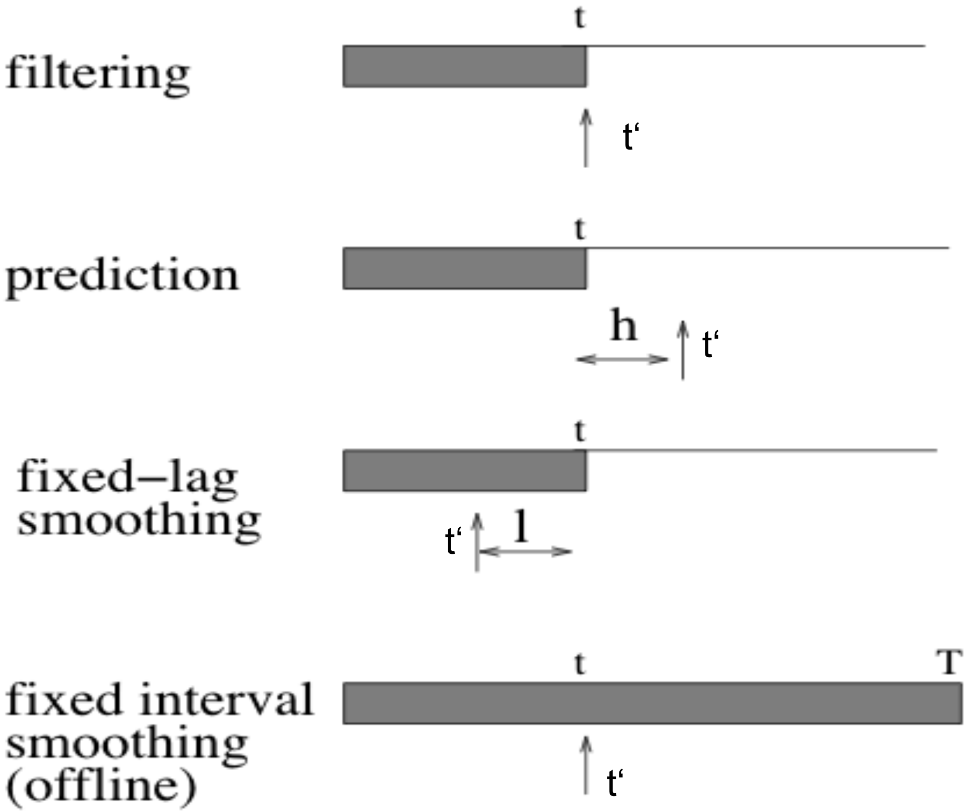
Dynamische Bayes‘sche Netze

- Beschreibung
  - Prior $P(X_0)$: Initiale Verteilung über den Zustand
  - Prozessmodell $P(X_t|X_{t-1})$: Bedingte Wahrscheinlichkeitsverteilung zur Beschreibung der Systemübergänge (Transitionsmodell)
- Zeitscheiben sind BN und definieren DBN
- DBN kann viele Knoten haben



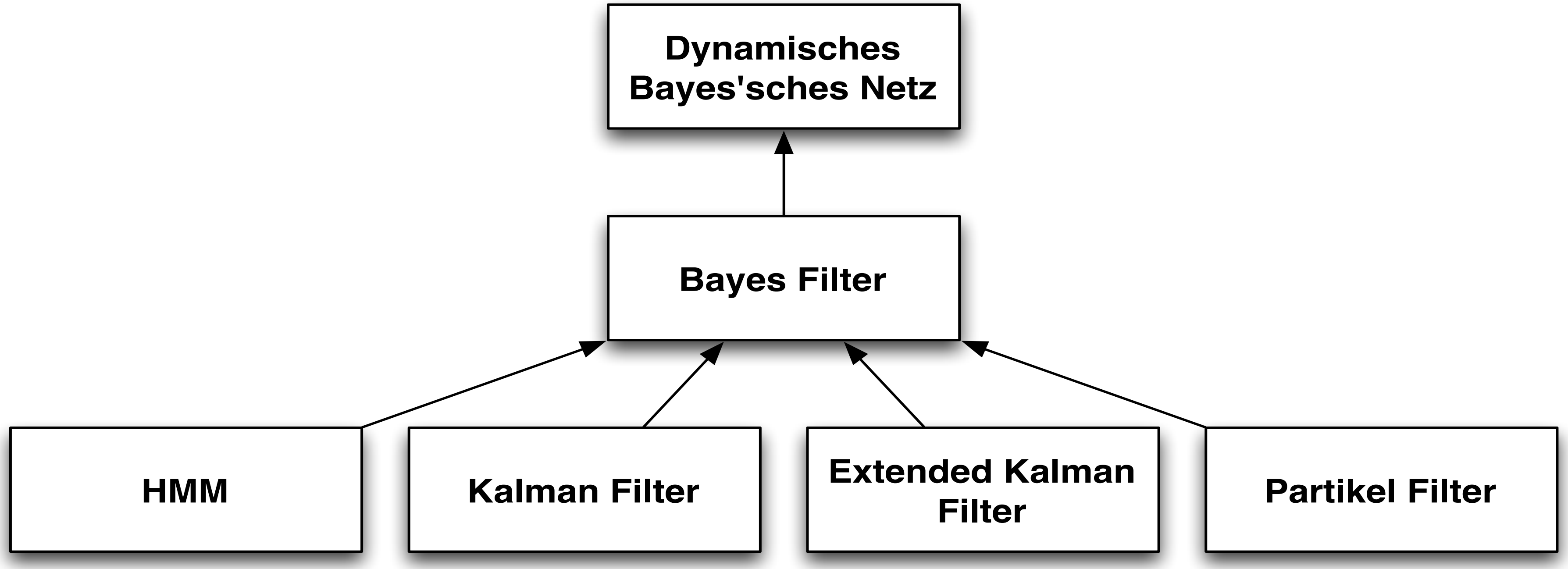
Anfragen

- Gegeben: Evidenzen bis zum Zeitpunkt t
- Gesucht: Verteilung über Zustand zu einem Zeitpunkt t‘
- Varianten (je nach t‘):



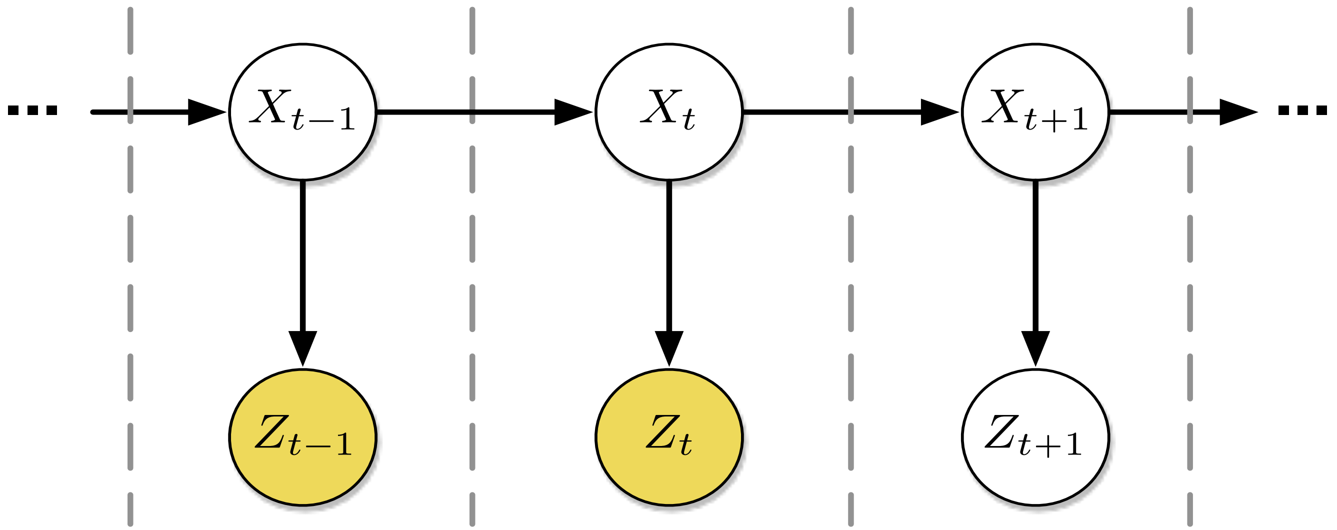
Filterverfahren im Überblick

- Unterscheidungsmerkmale
  - Zustandsraum (diskret, kontinuierlich, hybrid)
  - Modelle (linear z.B. $x_{t+1} = Ax_t + e_t$, nicht-linear)
  - Verteilungen und Inferenzverfahren



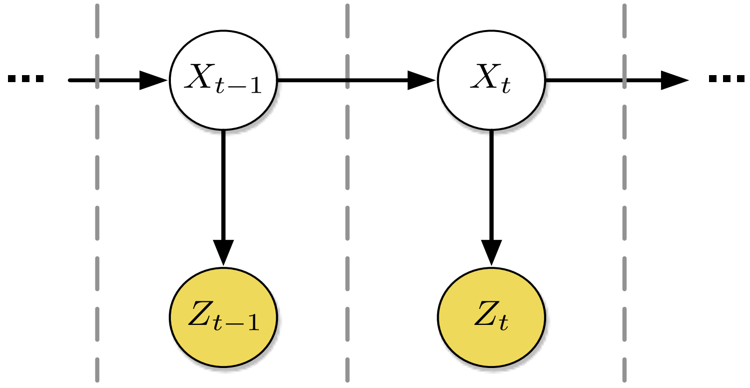
Bayes Filter

- DBN mit Eigenschaften
  - Zustandsraum: Systemzustand + Messung
  - Modelle: linear/nicht-linear (Markov 1. Ordnung)
  - Verteilungen und Inferenzverfahren: beliebig
  - Graphstruktur (Baum) ermöglicht effiziente Inferenzverfahren



Bayes Filter

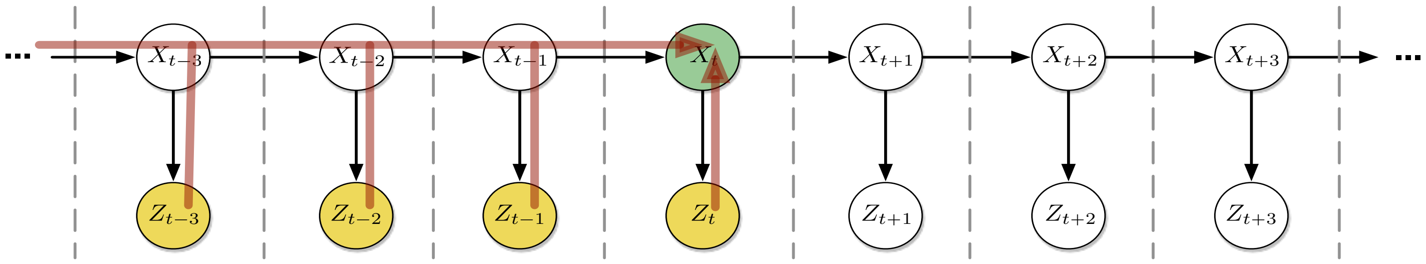
- Transitionsmodell $P(X_t|X_{t-1})$:
  - Beschreibt den Übergang des System von einem Zeitpunkt zum nächsten
- Messmodell $P(Z_t|X_t)$:
  - Beschreibt, welche Messungen ein Systemzustand zur Folge haben kann
- Markov‘sche Annahme
  - Gültig für die Zustände X



Bayes Filter - Anfragen

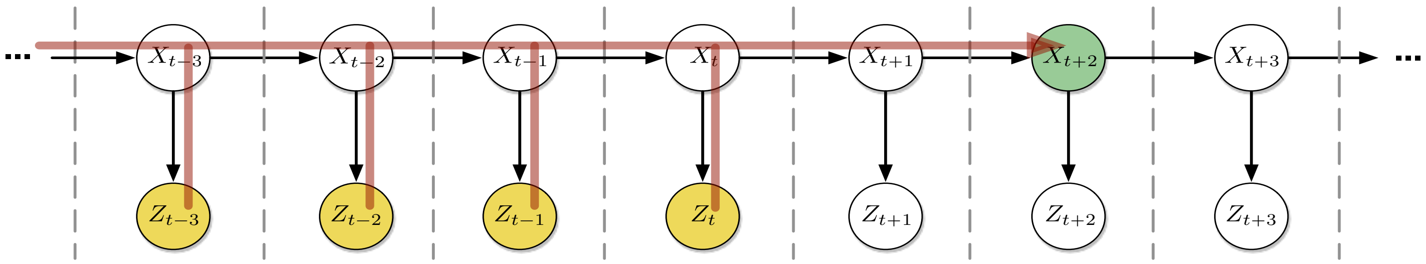
- Filtern

$$P(X_t|\underline{Z_t, \ldots, Z_0})$$



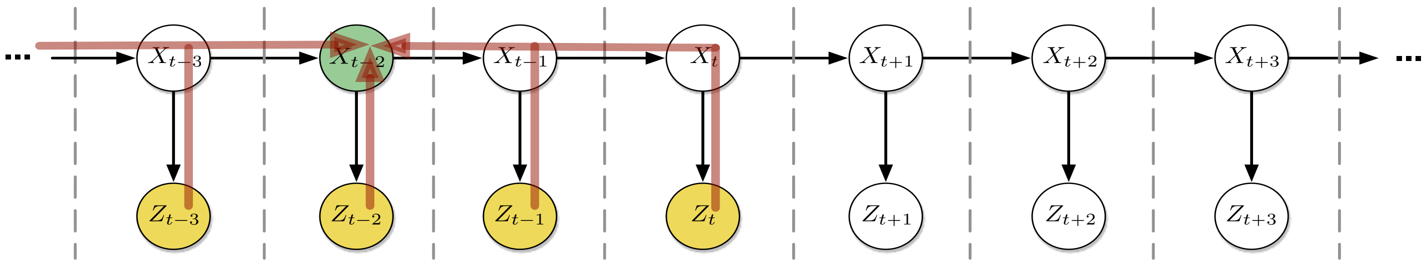
- Prädizieren

$$P(X_{t+k}|\underline{Z_t, \ldots, Z_0})$$



- Glätten

$$P(X_{t-k}|\underline{Z_t, \ldots, Z_0})$$



Kalman Filter

- Eigenschaften
  - Erfordert Verteilungen, die vollständig über die ersten beiden Momente (Erwartungswert und Kovarianz) beschrieben werden können (i.d.R. Normalverteilungen)
  - Modelle: Linear + additives, weißes Rauschen
  - Unimodalität der Verteilungen und Linearitätsannahme der Modelle sehr restriktiv
  - Wird in der Realität dennoch sehr häufig eingesetzt
- Erweiterungen für nicht-lineare Modelle
  - Extended Kalman Filter
  - Unscented Kalman Filter

Kalman Filter

- Bsp.: Fahrzeugverfolgung
- Zustand:
  - Position x_1, x_2
  - Geschwindigkeit v_1, v_2

$$x = \begin{pmatrix} x_1 \\ x_2 \\ v_1 \\ v_2 \end{pmatrix}$$

- Transitionsmodell:
  - Berechnet Folgeposition mit Unsicherheiten

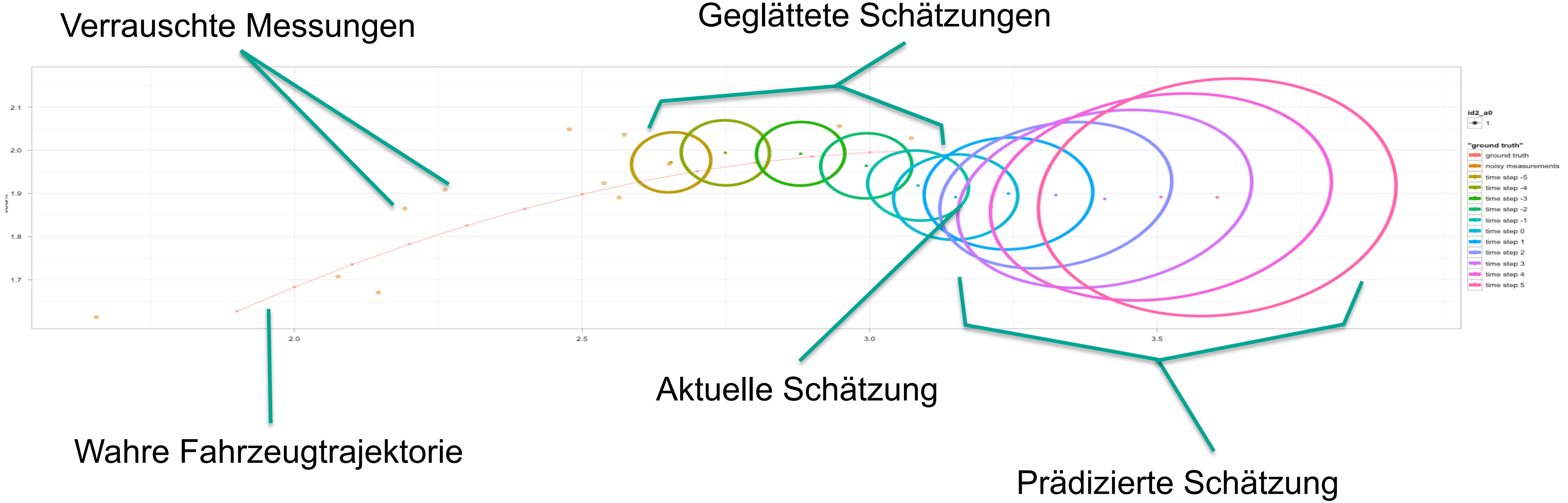
$$x_{t+1} = Ax_t + e_t = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ v_{1t} \\ v_{2t} \end{pmatrix} + e_t$$

- Messmodell:
  - Gemessen werden verrauschte Positionen (weißes Rauschen)

$$z_t = Ax_t + e_{zt} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ v_{1t} \\ v_{2t} \end{pmatrix} + e_{zt}$$

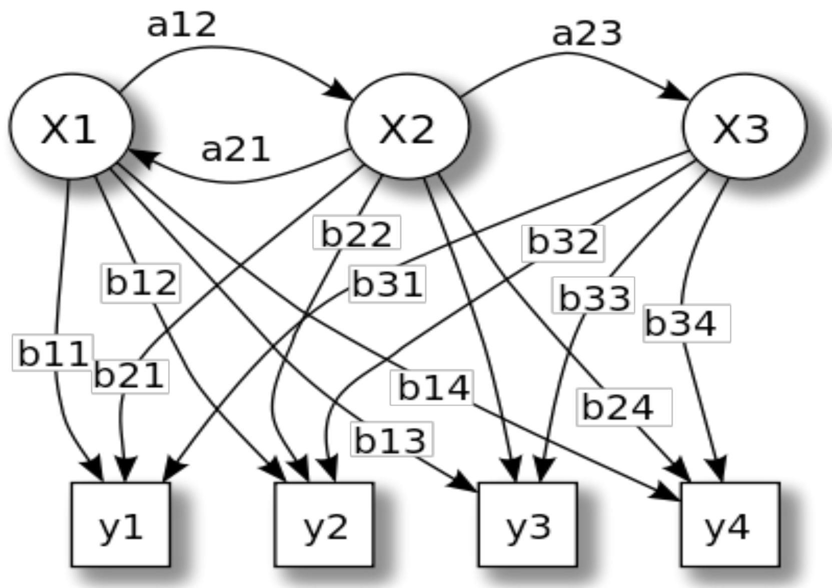
Kalman Filter

- Bsp.: Fahrzeugverfolgung



HMMs

- Eigenschaften
  - Zustandsraum: diskret
  - Modelle: bedingte Wahrscheinlichkeitstabellen
  - Verteilungen: Wahrscheinlichkeitstabellen
  - Transitionsmodell kann über endlichen Automaten dargestellt werden, dessen Kanten mit Wahrscheinlichkeiten annotiert sind



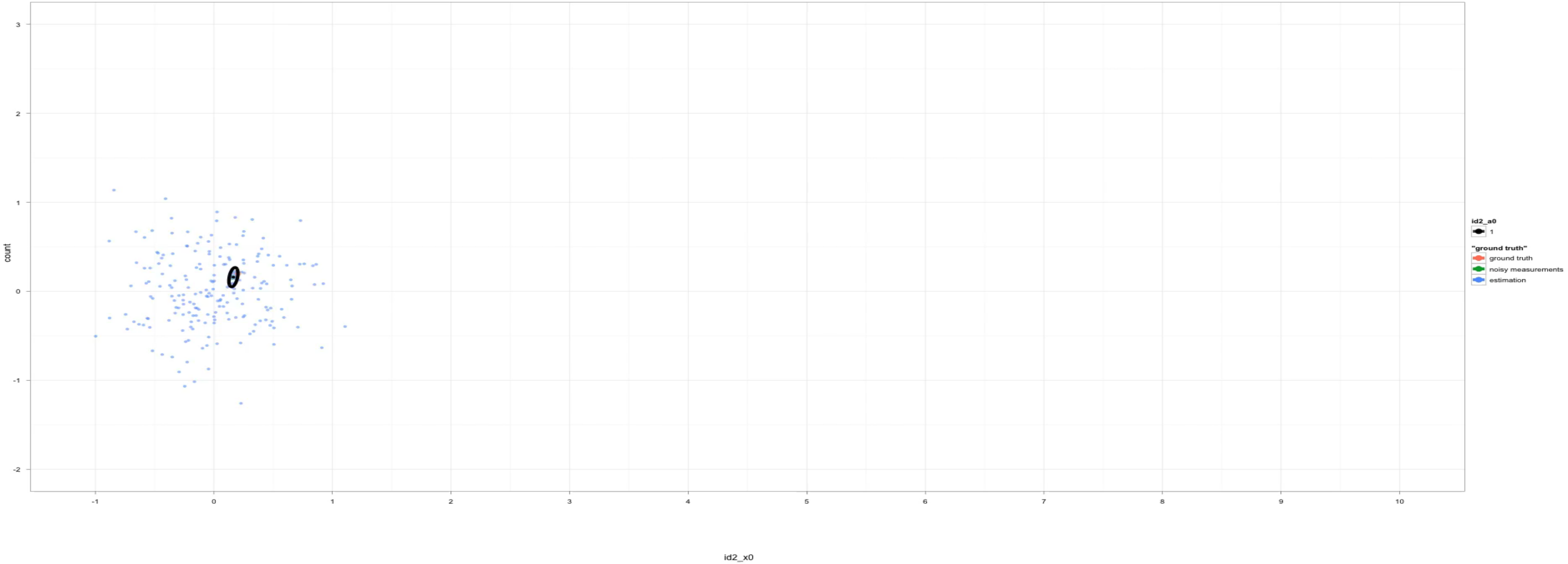
- Erweiterungen
  - kontinuierlichen Messungen
  - Hierarchische HMMs
  - Coupled HMMs

Partikel Filter

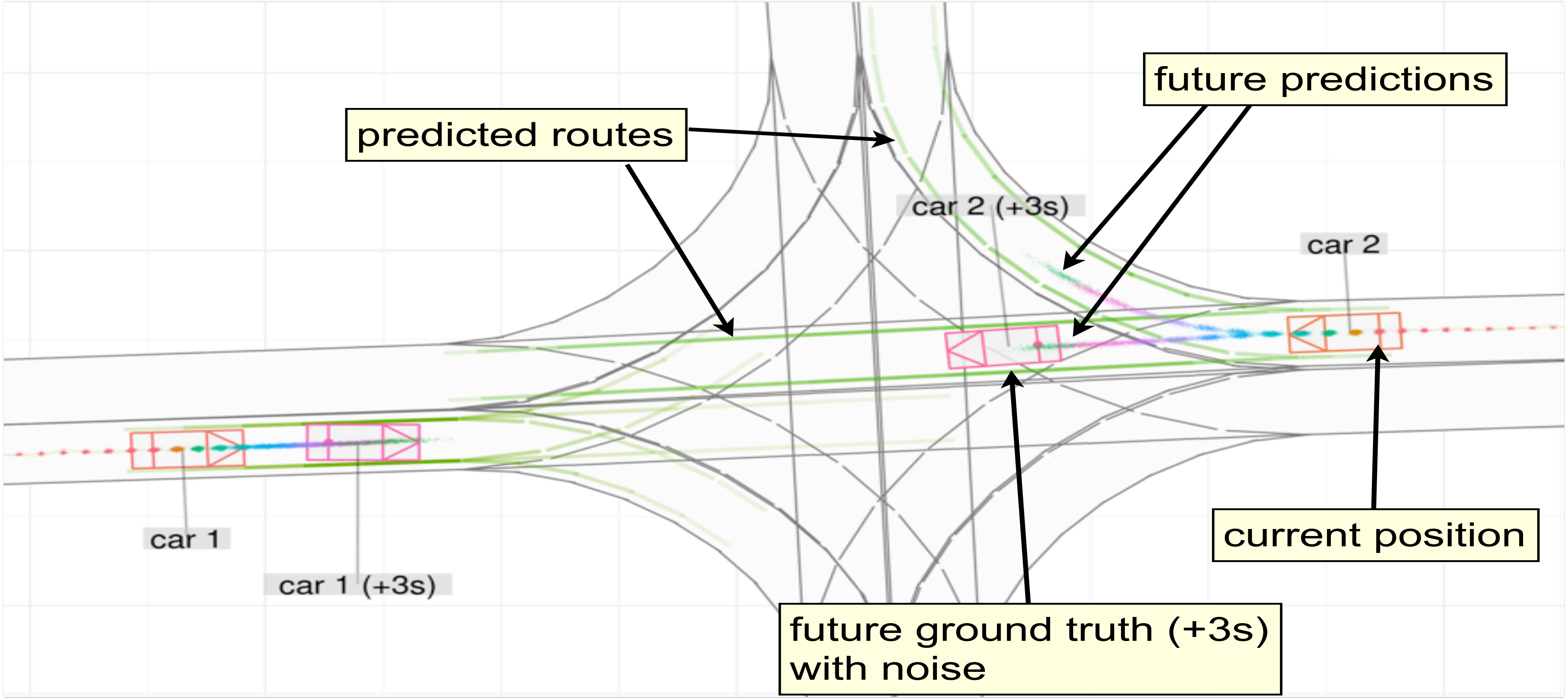
- Eigenschaften
  - Zustandsraum: diskret/kontinuierlich/hybrid
  - Modelle: linear/nicht-linear
  - Verteilungen: Partikelverteilungen (Dirac Mixturen), Multimodal
  - Inferenz: und partikelbasierte Inferenzverfahren (z.B. Likelihood Weighting)

- Erweiterungen
  - Rao-Blackwellized Partikel Filter (gemischt exakt-approximativ)

Partikel Filter - Video



Anwendung: Situationsanalyse und Prädiktion im Straßenverkehr



[Gindele et al., ITS Magazine 2015]

Situationsanalyse und Prädiktion

$$P(X|z,X^-) \propto P(z|X)\int_{X^-,L^-,LR^-,R^-,A^-} P(X,X^-,L^-,LR^-,R^-,A^-)$$

$$P(X_1 \ldots X_n|z,X_1^- \ldots X_n^-)$$

$P(Z|X)$ – Meassurement

$P(R|X)$ – Trafic. Part. Relations

$P(L|X)$ – Lane Matching

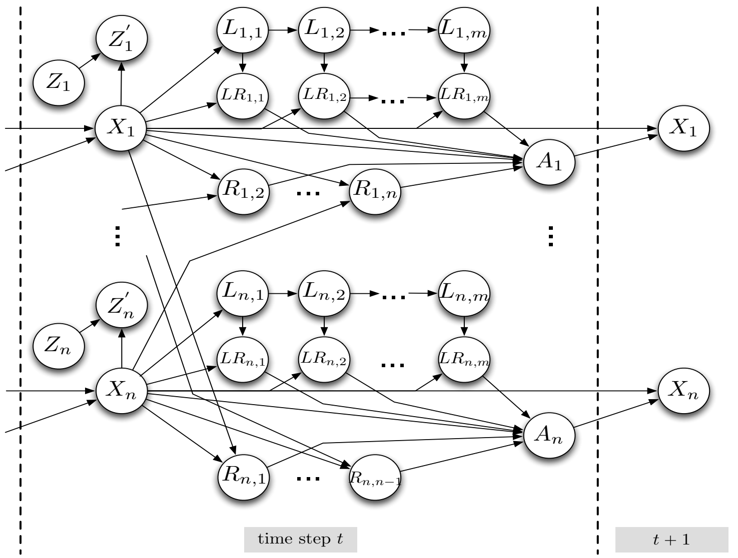
$P(L_{i+1}|L_i)$ – Lane Following

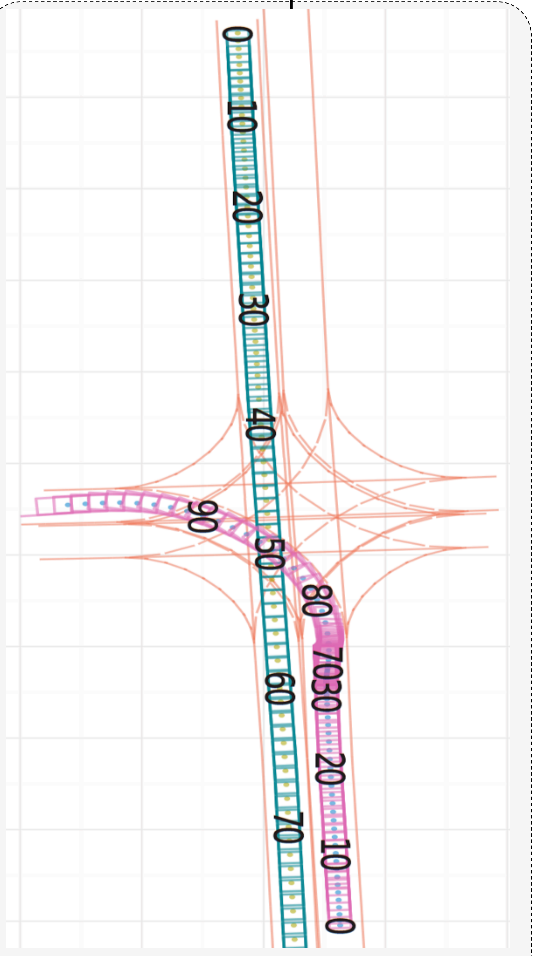
$P(LR|X,L)$ – Lane Relations

$P(X^{(t+1)}|X^{(t)},A^{(t)})$ – Motion

$P(A|X,R,LC)$ – Policy Modell

Lernen aus Daten möglich
z.B. mit Monte Carlo - EM





Perzeption, Wissen

$X=(X_1 \ldots X_n),\ X_i=(\vec{x},w,v)$

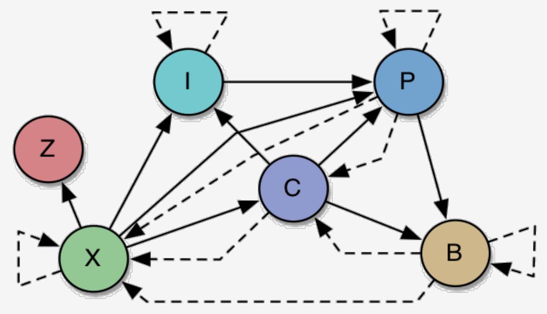
$Z=(Z_1 \ldots Z_n)$

$R=\{R_{i,j}\}, R_{i,j}=\Delta X_{ij}$

$L=(L_{1,1:m} \ldots L_{n,1:m})$

$LR=(LR_{1,1:m} \ldots LR_{n,1:m})$

$A=(A_1 \ldots A_n),\ A_i=(w,a)$

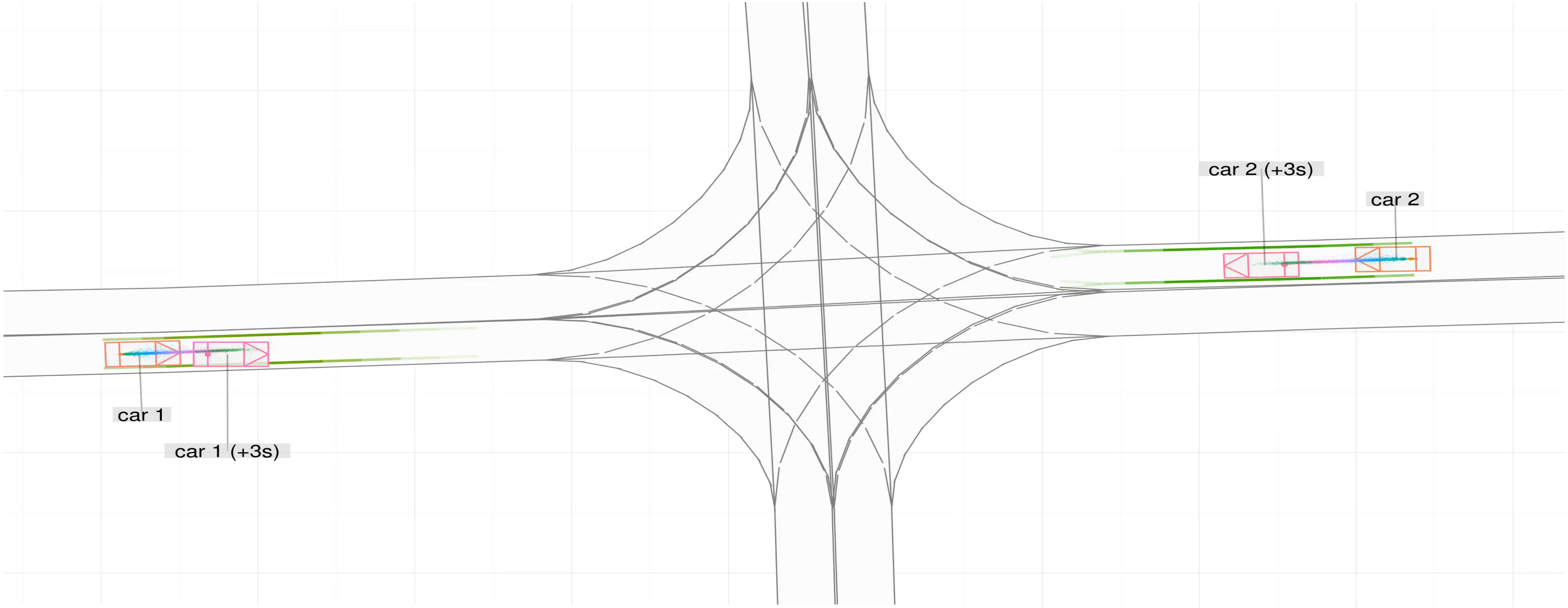


Probabilistsiches Modell (DBN)

Inferierte Situationsaussage

[Gindele et al., ITS Magazine 2015]

Situationsanalyse und Prädiktion



model	pos rmse (Δt: 0.3 s)	pos rmse (Δt: 6.0 s)	mll
learned model	0.19 m	6.48 m	2.032
single track model	0.28 m	12.83 m	0.988

[Gindele et al., ITS Magazine 2015]

Zusammenfassung

- Probabilistische Graphische Modelle kodieren JPDs inklusive bedingter Unabhängigkeiten zwischen Zufallsvariablen
- Bedingte Unabhängigkeiten ermöglichen effiziente Inferenzverfahren, einfachere Modelle und damit effizienteres Lernen
- DBNs erweitern BNs um den Aspekt der Zeit
- DBNs eignen sich hervorragend, um nicht beobachtbare Zustände von dynamischen Systemen zu schätzen

- Ausblick:
  - Lernen der Struktur und Modelle von BNs und DBNs

Literatur

- [1] D. Koller and N. Friedman. ***Probabilistic graphical models: principles and techniques***. The MIT Press, 2009.
- [2] C. Bishop. ***Pattern recognition and machine learning***, volume 4. Springer New York, 2006.
- [3] K. Murphy. ***Dynamic Bayesian Networks: Representation, Inference and Learning***. PhD thesis, University of California, 2002.
- [4] J. Pearl. ***Probabilistic reasoning in intelligent systems: networks of plausible inference***. Morgan Kaufmann, 1988.
- [5] S. Srihari. ***Machine Learning and Probabilistic Graphical Models Course***. Vorlesungsfolien, Department of Computer Science and Engineering, University at Buffalo, 2011.
- [5] T. Gindele, S. Brechtel. ***Learning Driver Behavior Models from Traffic Observations for Decision Making and Planning***. IEEE Intelligent Transportation Systems Magazine, Vol. 7, pp. 69-79.