Bayes’sche Netze: Grundlagen, Konstruktion, Anwendungen und Risikomodellierung

Ulrich Leuthäusser

Systemanalyse und Operations Research

Überblick

- **Grundlagen Bayes'scher Netze (BN)**
- **Bausteine von BNs**
- **Konstruktionsprinzipien und Lernen in BNs**
- **Anwendungen in der Situationsanalyse und Finanzmanagement**

Bill Gates 1996:
"Microsoft's competitive advantage is its expertise in Bayesian networks." **(Los Angeles Times, October 28, 1996)**

Explosionsartige Vermehrung der Anwendungen in den letzten Jahren:
VISTA system for NASA Mission Control
Air traffic management
Scene Analysis
Situation Assessment
Software Anwendungen: Lumiere Projekt von Microsoft (MSBN)
Technische Diagnose (Daimler)
Biostatistik – Medizinische Forschung (BUGS)
Spracherkennung (UC Berkeley)
Viele andere → Web

Kommerzielle Software: Toolbox für Matlab, Hugin, Netica

Bayes'sches Theorem

Die Wahrscheinlichkeit zweier Ereignisse A und B lässt sich mit Hilfe der bedingten Wahrscheinlichkeit P(A|B) für A, wenn B eingetreten ist, wie folgt ausdrücken:

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

so dass

$$P(A\,|\,B)=\frac{P(B\,|\,A)\cdot P(A)}{P(B)}$$

Interpretation:
P(A|B) beschreibt die Modifikation des Vorwissens P(A), wenn Evidenzen oder Daten B hinzukommen. Die Bayes'sche Formel erlaubt damit die Kombination von Vorwissen mit neuen beobachteten Daten.

- **Bayes, Thomas (1763)**
  An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions* of the Royal Society of London, 53:370-418

- **Subjektive Wahrscheinlichkeitsauffassung**

- **Unterschied: Ursache und Implikation**
  P(B|A) ist die Verallgemeinerung von A impliziert B: A → B und hat nichts zu tun mit “A verursacht B“
  (P(Dämmerung | Hahn kräht) > 0.99)

Beispiel – einfachstes Netz aus 2 Knoten

Krankheitstest

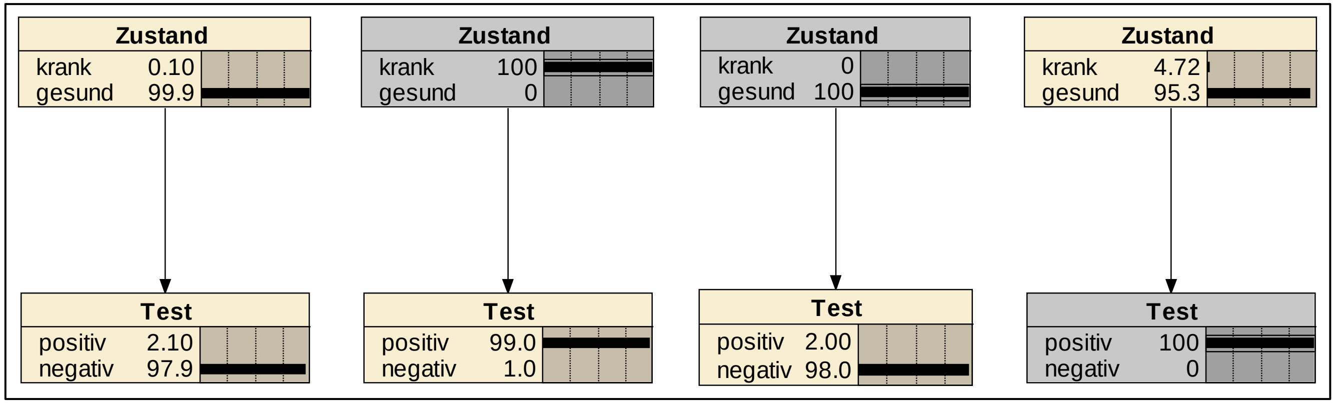
a priori Wahrscheinlichkeit erkrankt zu sein: **p(krank)=0.001**

Falschalarmwahrscheinlichkeit (Fehler 1. Art)

p(Test positiv|gesund) = 0.02

Übersehwahrscheinlichkeit (Fehler 2. Art)

p(Test negativ|krank) =0.01



Beispiel – einfachstes Netz aus 2 Knoten

Mit Hilfe des Bayes'schen Satzes erhält man das erstaunliche Ergebnis, dass trotz hoher Testgenauigkeit, die Wahrscheinlichkeit bei positivem Test krank zu sein, relativ klein ist:

$$p(k \mid \text{Tpos}) = \frac{p(\text{Tpos} \mid k)p(k)}{p(\text{Tpos} \mid k)p(k) + \left(1 - p(\text{Tneg} \mid g)\right) \cdot \left(1 - p(k)\right)} \approx 4{,}72\%$$

p(k|Tpos) hängt von der Falschalarmwahrscheinlichkeit **und** dem a priori p(k) ab, denn

$$p(k \mid \text{Tpos}) \approx \frac{p(k)}{\text{Fehler}\,1.\text{Art}}$$

(wenn p(k) << Fehler 1.Art)

Definition Bayes‘sches Netz

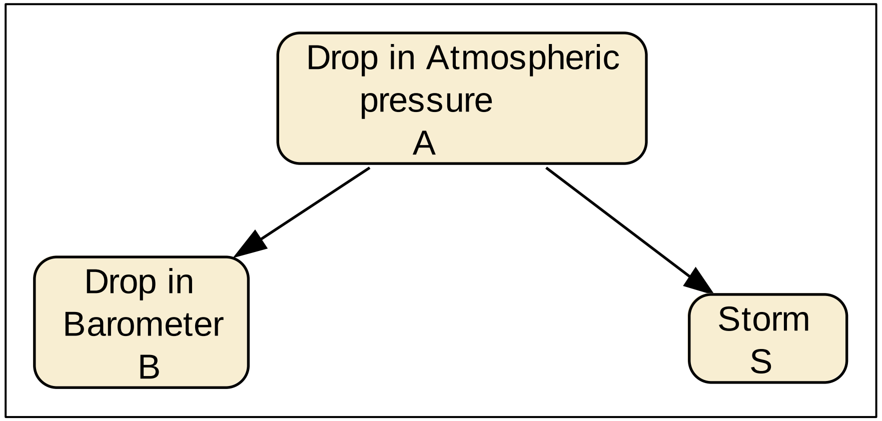
Ein Bayes’sches Netz stellt einen gerichteten azyklischen Graphen dar (DAG) und stellt die Wahrscheinlichkeitsverteilung des unsicheren Wissens graphisch dar.
Knoten repräsentieren Zufallsvariable und die Kanten direkte stochastische Abhängigkeiten.

Grundidee und Ziel
Sparsame Repräsentation von lokalen bedingten Wahrscheinlichkeiten mit möglichst wenig Kanten. Berechnung von Wahrscheinlichkeiten von nichtbeobachtbaren Variablen bei gegebenen Evidenzen.

Relation zu anderen Methoden
Viele (ältere) Methoden lassen sich in den BN-Formalismus einordnen.

Bausteine Bayes‘scher Netze

1. Divergierende Verknüpfung

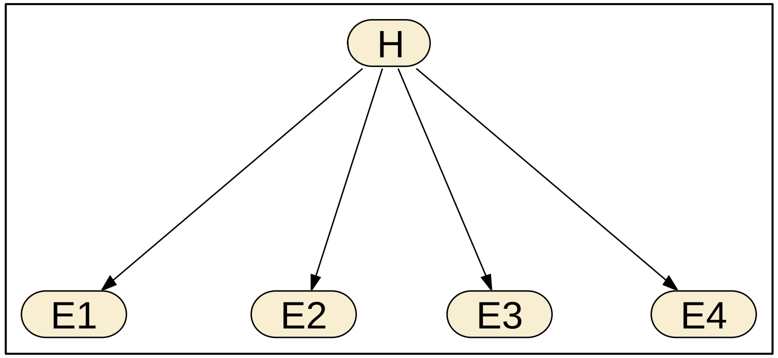


Kettenregel: P(A,B,S)=P(A)·P(B|A)·P(S|A)

P(B|S)>P(B) keine Unabhängigkeit: ¬I(B,S)

P(B|S,A)=P(B|A) aber bedingte Unabhängigkeit: I(B,S|A)

Bausteine Bayes‘scher Netze: naiver Klassifikator



Gegeben seien eine Hypothese H mit P(H)
und n Evidenzen E_i mit den bedingten Wahrscheinlichkeiten

$$p_i = P(E_i \mid H) \qquad q_i = P(E_i \mid \overline{H})$$

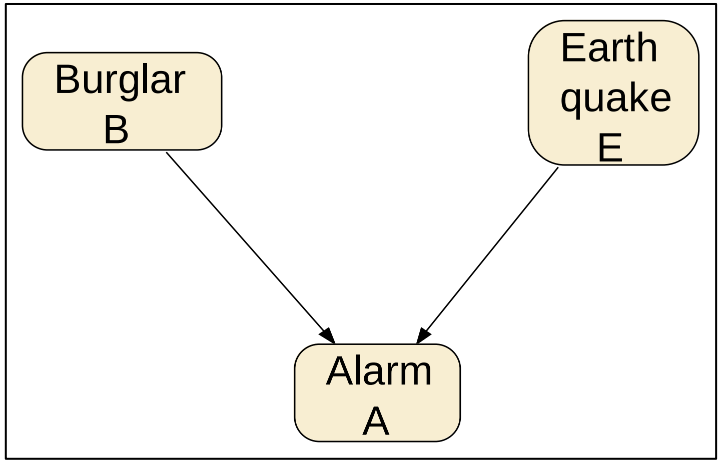
Kettenregel und Bayes’scher Satz ergeben

$$P(H \mid E1, E2, \ldots) = \frac{1}{1 + \frac{1 - p(H)}{p(H)} \prod_i \frac{q_i}{p_i}} \approx \frac{1}{1 + \frac{(F1.Art)^n}{p(H)}}$$

Einfluss des Vorwissens verschwindet. **Anwendung:** Spamfilter

Bausteine Bayes'scher Netze

2. Konvergierende Verbindung



$P(B|E)=P(B)$ Unabhängigkeit $I(B,E)$

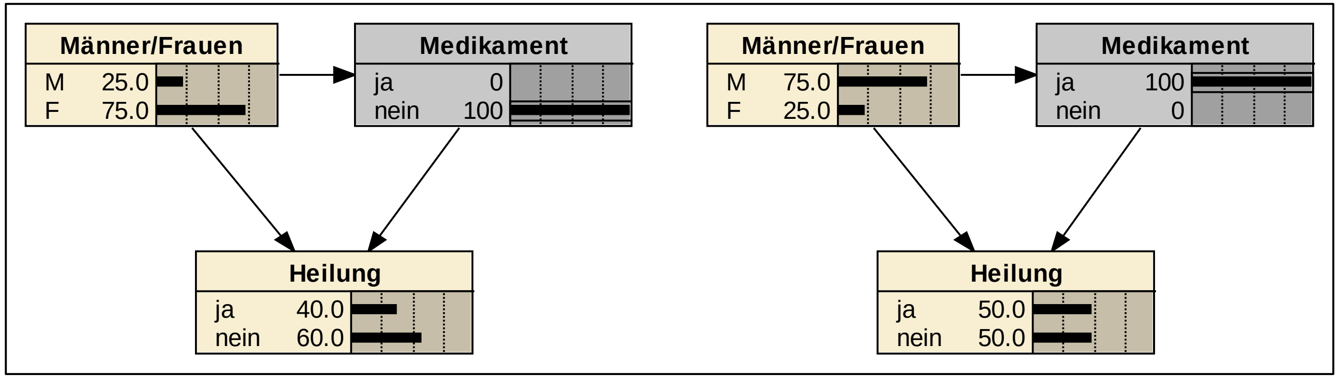
$P(B \mid E, A) < P(B \mid A)$, aber keine bedingte Unabhängigkeit $\neg I(B,E \mid A)$

Wenn die Auswirkung A bekannt ist und eine Ursache E gefunden wurde, dann fällt die Wahrscheinlichkeit der anderen Ursache B ("Explaining Away").

Bausteine Bayes'scher Netze

3. Vollständige Verknüpfung

Simpson's Paradox: Größenverhältnisse von Wahrscheinlichkeiten können umgekehrt werden, wenn man von Populationen zu Subpopulationen übergeht.



Medikamenteinnahme = C, Heilung = H, Männer/Frauen = M/F

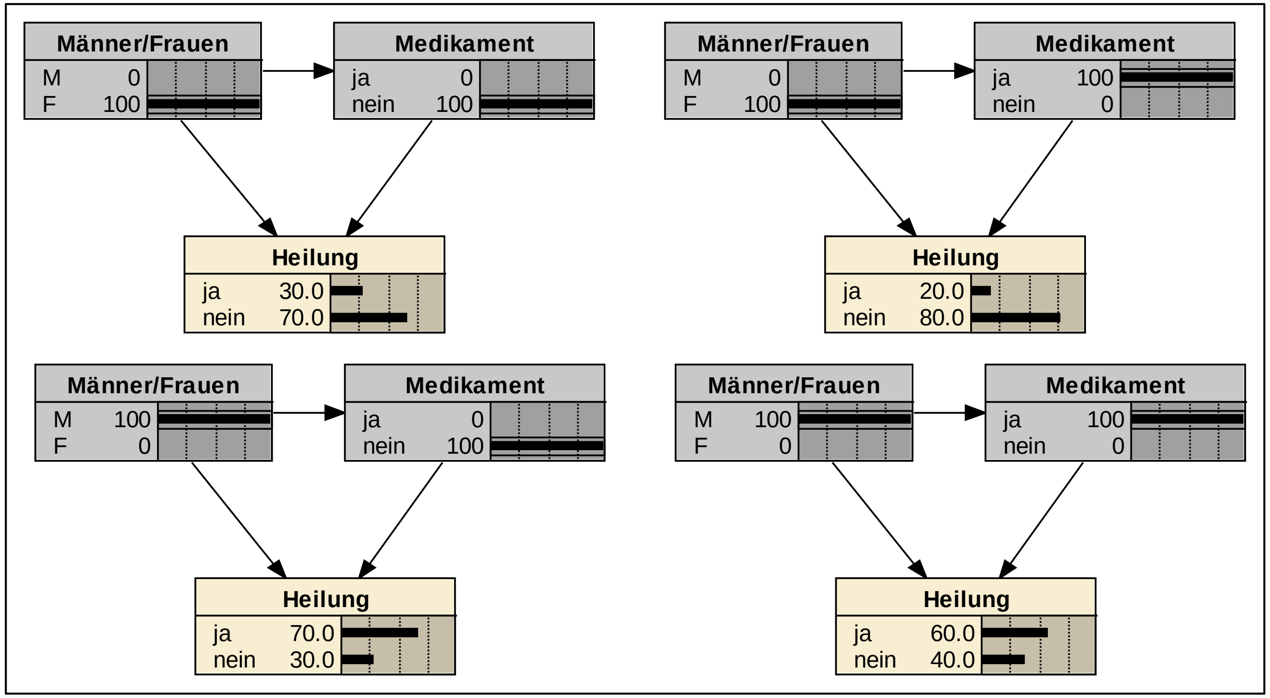
$$P(H \mid C) > P(H \mid \neg C)$$

Bausteine Bayes‘scher Netze

Kein Medikament ~C | Medikament C

Frauen F

Männer M



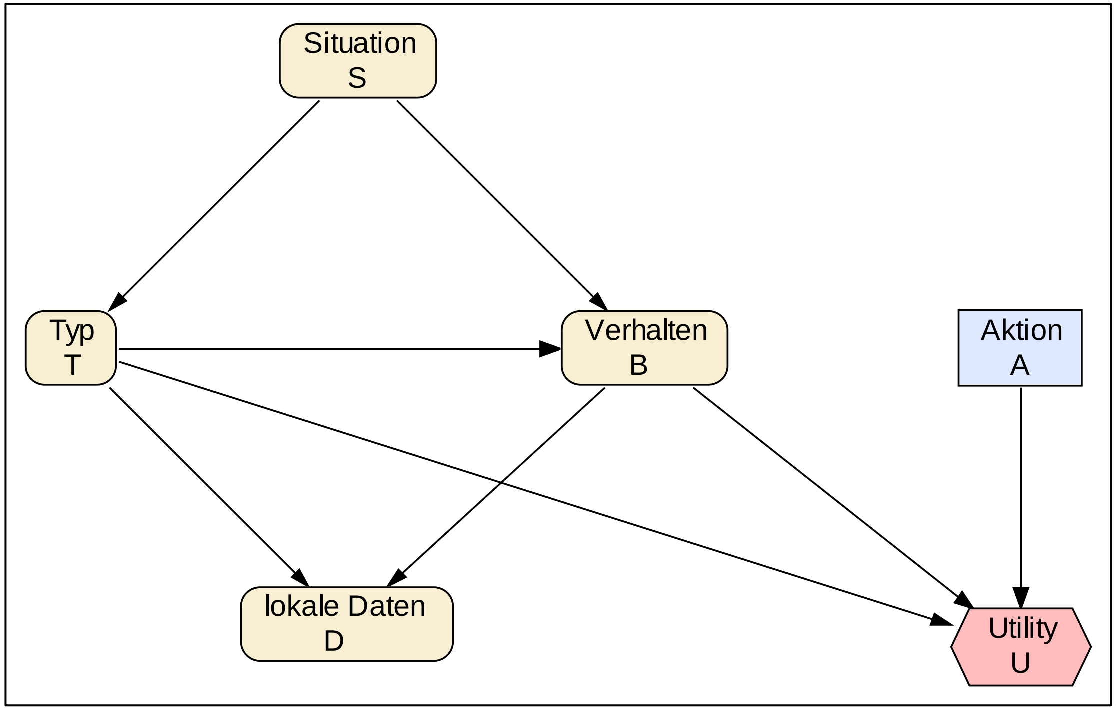
$$P(H \mid C,F) < P(H \mid \neg C,F)$$

$$P(H \mid C,M) < P(H \mid \neg C,M)$$

Da man auch $P(H \mid C) = P(H \mid \neg C)$ erzeugen kann, gibt es Verteilungen, die Unabhängigkeiten besitzen, die aus dem DAG nicht hervorgehen. Faithfulness Annahme: solche Verteilungen sollen nicht vorkommen, der DAG bestimmt alle I.

Anwendung: BN für Situationsanalysen

Situation Assessment and Time-Critical Decision Making (Teilnetzwerk)



$$P(S,T,B,D) = P(D \mid B,T) \cdot P(B \mid T,S) \cdot P(T \mid S) \cdot P(S)$$

anstatt allgemein

$$P(S,T,B,D) = P(D \mid B,T,S) \cdot P(B \mid T,S) \cdot P(T \mid S) \cdot P(S)$$

Anwendung: größeres BN

- Motiv: Einbau von Vorwissen in Kombination mit statistischen Daten
- Generischer Charakter des Problems
- Modularität, Skaleninvarianz
- Erweiterung zum Entscheidungsnetzwerk:
  Hinzugefügt wurde eine Utilityfunktion, die angibt, welchen Nutzen es hat, wenn ein bestimmtes T und B bei einer gewählten Aktion A vorliegt. Das Maximum des Erwartungswerts der Utility liefert die optimale Entscheidung.

Konstruktion und Lernen

Vorgehensweise bei der Konstruktion

1. Strukturlernen

- Versuch, eine optimale Struktur aus Daten abzuleiten
- Bestimmung der (bedingten) Unabhängigkeitsbeziehungen {I}
- Berücksichtigung der kausalen Ordnung (nicht trivial). Oft stimmt die kausale Ordnung mit den Unabhängigkeitsbeziehungen überein.
- Verwendung einfacher Teilstrukturen (z.B. “noisy or“ bei konvergierenden Verbindungen)

Daten → {I} +faithfulness+ keine verborgenen Ursachen → alle {kausale DAGs}

also Lernen von kausalen Beziehungen aus Daten unter Faithfulness
(sehr aktuell, interdisziplinär, umstritten).

Konstruktion und Lernen

Ist die Struktur bekannt folgt das

2. Parameterlernen

Verteilung P für die Parameter θ für eine gegebene Struktur S und Daten D:

$$P(\text{Parameter } \theta \,|\, \text{Struktur S, Daten D}) = \frac{P(D \,|\, \theta, S) \cdot P(\theta \,|\, S)}{P(D \,|\, S)}$$

Maximum Likelihood (ML) - Schätzung liefert rechte Seite:

$$P(D \,|\, \theta, S) = \prod_{i=1}^{N} P(y_i \,|\, \theta, S)$$

Gute Algorithmen stehen zur Verfügung, auch wenn die Daten nicht vollständig sind.

Konstruktion und Lernen

Bias Variance Dilemma

Einfaches Modell = hoher Bias (nicht genug Flexibilität) →
deshalb großer Fehler, aber geringe Abhängigkeit vom Trainingsset =
niedrige Varianz

Komplexes Modell = niedriger Bias (alle Anpassungen möglich) =
overfitting → starke Abhängigkeit vom Trainingsset = hohe Varianz →
deshalb großer Fehler

Verfahren zur Vermeidung dieses Dilemmas fügen zur ML-Schätzung
Terme hinzu, die zu viele Parameter bestrafen
(Bayesian Information Criterion BIC).

Dynamische BNs: Finanzmarktmodellierung

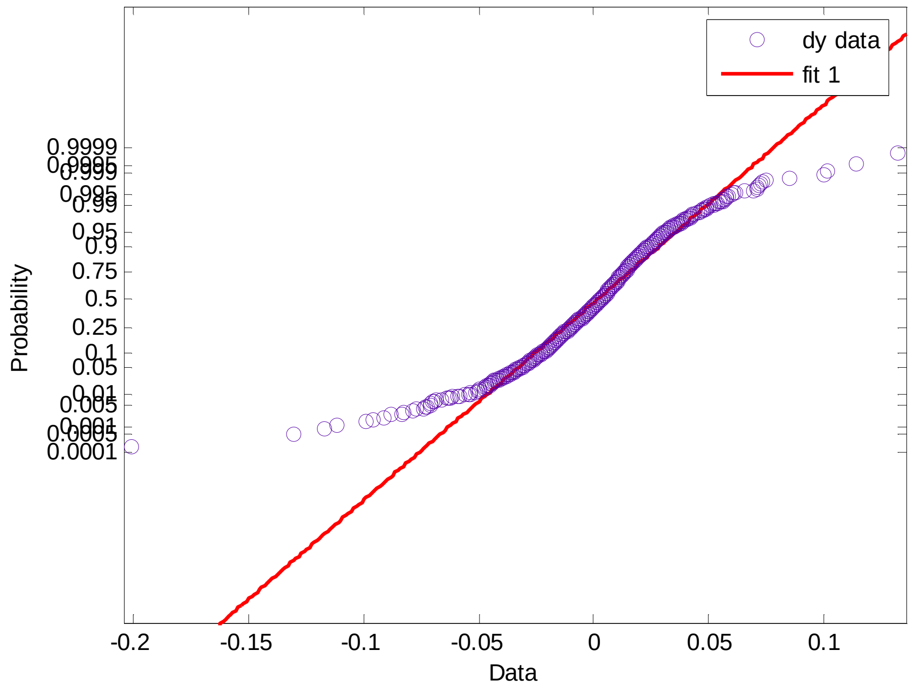
Modell eines Aktienmarkts:

- Es gibt eine nicht beobachtbare Variable, den Trend μ, die sich also mit der Zeit ändern kann (z.B. nach einem first-order Markov Modell)
- Geschätzt wird μ durch verrauschte v, den täglichen logarithmischen Preisdifferenzen
- Problem: sehr schlechtes Signal-Rausch Verhältnis

$$\frac{\sigma_\mu^2}{\sigma_v^2} \approx \left(\frac{1}{20}\right)^2$$

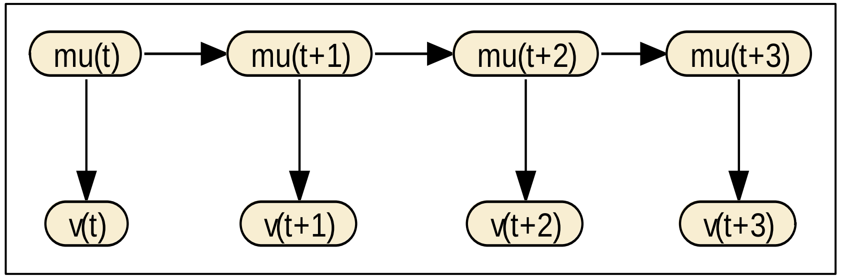
- Die Verteilung der v ist wegen der “fat tails“ keine Normalverteilung

Dynamische BNs: Finanzmarktmodellierung



Dynamische BNs: Finanzmarktmodellierung

1. Modell: Kalman Filter Modell



$$P(\mu_{t+3},\mu_{t+2},\ldots,v_{t+3},v_{t+2},\ldots) = P(\mu_{t+3} \mid \mu_{t+2})P(v_{t+3} \mid \mu_{t+3})P(\mu_{t+2} \mid \mu_{t+1})P(v_{t+2} \mid \mu_{t+2})\ldots$$

$P(v_t \mid \mu_t)$ ≈ Gemisch aus Gaussverteilungen

$P(\mu_{t+1} \mid \mu_t)$ z.B. $\mu_t = f_t(\mu_{t-1}) + \eta_t$

Spezialfall: lineares Gauss'sches Modell:

$$\mu_{t+1} = \delta + \Phi \cdot \mu_t + \eta_t$$

$$v_t = \mu_t + \varepsilon_t$$

Dynamische BNs: Finanzmarktmodellierung

Exakte Schätzung erfolgt mit Kalmanfilter.

Vereinfachung: wenn sich μ_t innerhalb des Samplingintervals (z.B. n Wochen) nur wenig ändert, also $\mu_{t+1} \approx \mu_t$

$$\mu_t = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \frac{\sigma_v^2}{n}} \cdot m_t + \left(1 - \frac{\sigma_\mu^2}{\sigma_\mu^2 + \frac{\sigma_v^2}{n}}\right) \cdot \mu_0 \qquad \text{mit} \qquad m_t = \frac{1}{n}\sum_i v_{t-i}$$

$$\text{mit} \quad P(v \mid \mu) = N(\mu, \sigma_v) \qquad \text{und} \qquad P(\mu) = N(\mu_0, \sigma_\mu)$$

wenn $\sigma_\mu \to 0$ oder $\sigma_v \to \infty$ dann
$\mu_t = \mu_0$ d.h. Black-Scholes Fall = Random Walk =
Buy-and-Hold Strategie: wenn $\mu_0 > r_0$ (risikolose Rendite).

Dynamische BNs: Finanzmarktmodellierung

Wenn man dagegen an zeitlich veränderliche Trends glaubt, dann gilt:

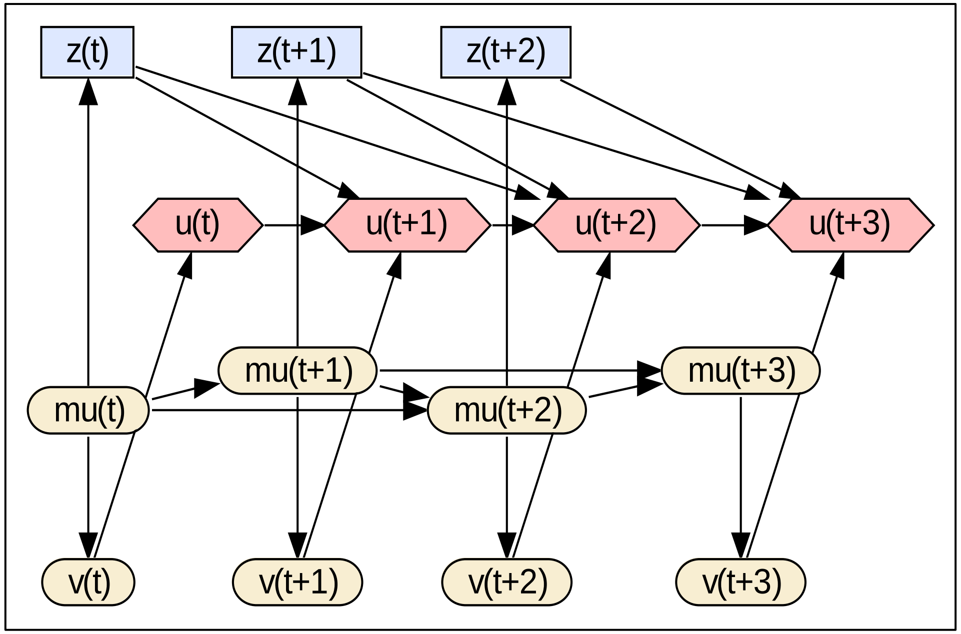
$$\left(m_t > r_0 - (\mu_0 - r_0) \frac{\sigma^2}{n \cdot \sigma_0^2} \right) \Rightarrow \text{"im Markt"}$$

- Vorheriges Abtrennen des risikolosen Anteils
- Vermeidung von Bearmärkten
- Erweiterung auf externe Variablen möglich

Dynamische BNs: Finanzmarktmodellierung

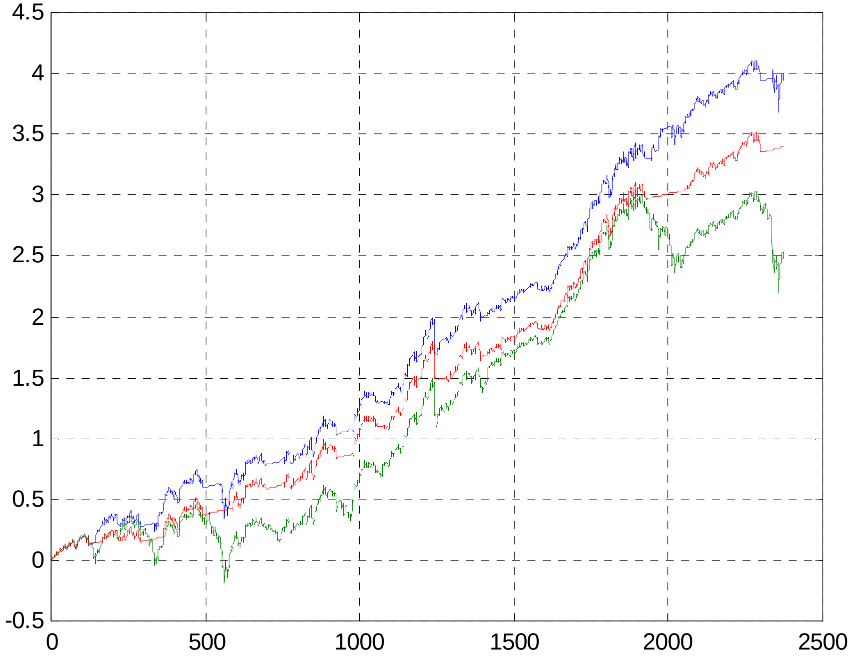
2. Modell

allgemeinere Graphenstrukturen und Utility- und Entscheidungsvariable



Dynamische BNs: Finanzmarktmodellierung

Ergebnisse



Beispiel S&P 500-Markt im Zeitraum Januar 1964 bis Juli 2009 auf Wochenbasis, Grüne Kurve: S&P 500 (Buy-and-Hold-Strategie), Rote Kurve: AR(2) für μ; Blaue Kurve: komplexeres $\mu_t = f_t(\mu_{t-1}, \mu_{t-2}) + \eta_t$

Fazit

- BNs sind ideal für die Kombination von Vorwissen mit beobachteten Daten

- Sehr dynamisches Gebiet: viele Netze sind veröffentlicht, leistungsfähige kommerzielle Software und Algorithmen stehen zur Verfügung

- Fokus geht in Richtung Entdeckung kausaler Strukturen, Zerlegbarkeit von BNs und Skalierbarkeit

- Man kann mit Bayes'schen Netzen sehr schnell mit Unsicherheit behaftete Systeme modellieren, ohne die großen, traditionellen Hürden in Wahrscheinlichkeitstheorie, Zeitreihenanalyse, etc. überwinden zu müssen.