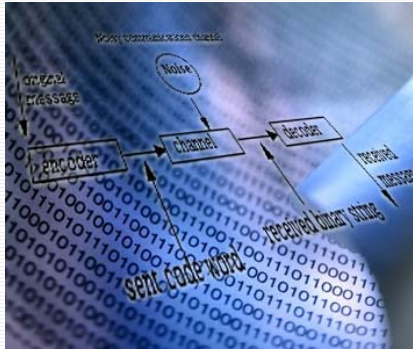


Kapitel 2: Stochastische Prozesse



Bedingte Verteilungen

- Ebenso kann die Verbundwahrscheinlichkeit von n Zufallsvariablen über bedingte Wahrscheinlichkeiten ausgedrückt werden

$$P(X_1 \cdots X_n) = P(X_1 \cdots X_{n-1})P(X_n | X_1, \dots, X_{n-1})$$

$$P(X_1 \cdots X_n) = \prod_{i=1}^n P(X_i | X_1, \dots, X_{i-1})$$

- Wiederum kommt eine Produktregel zur Anwendung



Summen- und Produktregeln sind in der Wahrscheinlichkeitsrechnung von zentraler Bedeutung

Beispiel: Verbunds WS

- Bedingte Verbundwahrscheinlichkeiten
- Sei $A = X_1, B = X_2, C = X_3$
- $n=2$: Gesetz von Bayes

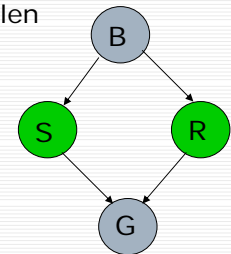
$$P(A, B) = P(A) \cdot P(B | A)$$

- $n=3$:

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

Bayes – Netz (1)

- Vier binäre (True, False) Zufallsvariablen
 - B: Bewölkt
 - S: Sprinkler
 - R: Regen
 - G: Gras ist nass



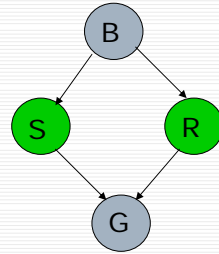
$P(B = T)$	$P(B = F)$
0.5	0.5

B	$P(S = F)$	$P(S = T)$
F	0.5	0.5
T	0.9	0.1

B	$P(R = F)$	$P(R = T)$
F	0.8	0.2
T	0.2	0.8

Bayes – Netz (2)

S	R	P(G = F)	P(G = T)
F	F	1.0	0.0
F	T	0.1	0.9
T	F	0.1	0.9
T	T	0.01	0.99



- Verbundwahrscheinlichkeit

$$\begin{aligned} P(S, B, R, G) &= P(B) \cdot P(S | B) \cdot P(R | B, S) \cdot P(G | B, S, R) \\ &= P(B) \cdot P(S | B) \cdot P(R | B) \cdot P(G | S, R) \end{aligned}$$

da $P(R | S) = P(R)$

Bayes – Netz (3)

- Beobachtung: Gras ist nass.
- Frage: Nass vom Sprinkler oder Regen?

$$\begin{aligned} P(S = T | G = T) &= \frac{P(S = T, G = T)}{P(G = T)} \\ &= \frac{\sum_{b,r} P(B = b, S = T, R = r, G = T)}{\sum_{b,r,s} P(B = b, S = s, R = r, G = T)} \\ &= \frac{0.2781}{0.6471} = 0.4237 \end{aligned}$$

Bayes – Netz (4)

- Beobachtung: Gras ist nass.
- Frage: Nass vom Sprinkler oder Regen?

$$\begin{aligned} P(R = T | G = T) &= \frac{P(R = T, G = T)}{P(G = T)} \\ &= \frac{\sum_{b,s} P(B = b, S = s, R = T, G = T)}{\sum_{b,r,s} P(B = b, S = s, R = r, G = T)} \\ &= \frac{0.4581}{0.6471} = 0.706 \end{aligned}$$

Erwartungswert

- Der Erwartungswert $E(X)$ einer diskreten Zufallsvariablen X ist gegeben durch

$$E(X) = \sum_i x_i p(x_i)$$

- Dieser wird oft auch Mittelwert von X genannt
- Schwerpunkt der Häufigkeitsfunktion
- Erwartungswerte sind statistische Momente **erster Ordnung**
- Aufgrund der Linearität erhalten wir (Superposition)

$$Y = aX + b \Rightarrow E(Y) = aE(X) + b$$

Geometrische Verteilung

ETH

- Produkt ist defekt mit Wahrscheinlichkeit p
- Wie viele Produkte müssen untersucht werden, bis ein defektes Produkt gefunden wird?
- X : Anzahl inspizierter Produkte

$$P(X = k) = P_X(k) = q^{k-1} \cdot p \quad \text{wobei } q = 1-p$$

$$E(X) = \sum_{k=1}^{\infty} k \cdot p \cdot q^{k-1} = p \cdot \sum_{k=1}^{\infty} k \cdot q^{k-1} \quad \text{wobei } k \cdot q^{k-1} = \frac{d}{dq} \cdot q^k$$

$$= p \cdot \frac{d}{dq} \cdot \sum_{k=1}^{\infty} q^k = p \cdot \frac{d}{dq} \cdot \frac{q}{1-q}$$

$$= \frac{p}{(1-q)^2} = \frac{1}{p}$$

Varianz

ETH

- Die **Varianz** $\text{Var}(X)$ einer diskreten Zufallsvariablen X ist gegeben durch

$$\text{Var}(X) = E\{[X - E(X)]^2\}$$

- Vorausgesetzt, dass E existiert
- Für eine diskrete Zufallsvariable X mit $p(x)$ ergibt sich

$$\text{Var}(X) = \sum_i (x_i - E(X))^2 p(x_i)$$

- Ebenso erhalten wir

$$Y = aX + b \Rightarrow \text{Var}(Y) = a^2 \text{Var}(X)$$

Bernoulli - Verteilung

ETH

- X hat Bernoulli - Verteilung

$$X = 0 \rightarrow 1-p$$

$$X = 1 \rightarrow p$$

$$E(X) = 0 \cdot (1-p) + 1 \cdot p = p$$

$$\text{Var}(X) = (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p = p \cdot (1-p)$$

$$\rightarrow \text{Maximum bei } p = \frac{1}{2}$$

Standardabweichung

ETH

- Die **Standardabweichung** σ ist gegeben durch

$$\sigma = \sqrt{\text{Var}(X)},$$

und beschreibt eine Art mittlere Abweichung (Streuung) der Daten vom Mittelwert

- Die Varianz wird oft zur Fehlerberechnung in Messungen verwendet
- Man definiert dazu den mittleren quadratischen Fehler

$$MSE = E[(X - x_0)^2]$$

- Die **Kovarianz** misst die Verbundvariabilität zweier Zufallsvariablen X und Y

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

- Oftmals wird folgende Notation verwendet

$$E(X) = \mu_X \quad E(Y) = \mu_Y$$

- Es gilt auch

$$\text{Cov}(X, Y) = E(XY) - E(X) \cdot E(Y)$$

oder auch

$$\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$$

- Und vieles mehr... (siehe Rice pp.131ff)

- Die **Korrelation** zweier Zufallsvariablen X und Y wird über Varianz und Kovarianz definiert

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

- Sie ist von grosser praktischer Bedeutung
- Schliesslich bleibt der bedingte Erwartungswert

$$E(Y|X = x) = \sum_y y p_{Y|X}(y|x)$$

- Es gilt auch

$$E(Y) = \sum_x E(Y|X = x) p_X(x) = E(E(Y|X))$$

- Vergleiche Marginalisierung

- Ein **stochastischer Prozess** $\{X(t_i)\}$ ist ein Wahrscheinlichkeitsprozess, welcher eine Sequenz von Zufallsvariablen $X(t_i)$ erzeugt.
- Zu jedem Zeitpunkt t_i existiert also eine Zufallsvariable, die einer bestimmten Verteilung unterliegt (kontinuierlich) oder bestimmte Werte annehmen kann (diskret).



$$\{X(t_0) = x_0, X(t_1) = x_1, X(t_2) = x_2, X(t_3) = x_3 \dots\}$$

- Diskret** bezieht sich also auf Zeit t_i sowie auf mögliche Werte x_k
- Beispiel: Folge von Münzwürfen mit Wahrscheinlichkeiten p für Kopf und $q=1-p$ für Zahl**
- Hierbei gilt die statistische Unabhängigkeit der einzelnen Ergebnisse:

$$x \in \{Kopf, Zahl\}$$

- Es gilt also für zwei beliebige diskrete Zeiten t_i und t_j

$$P(X(t_i)|X(t_j)) = P(X(t_i))$$

- Das muss nicht so sein (Markov-Ketten)!

- Ein stochastischer Prozess, dessen statistische Eigenschaften invariant unter Translation der Zeit bleiben, heisst **stationär**.
- Dies bezieht sich auf Momente verschiedener Ordnung
- In der Praxis oft auf Erwartungswerte beschränkt
- Dann gilt also:

$$E(X(t_j)) = E(X(t_{j+k}))$$

oder auch, dass die Verbundwahrscheinlichkeiten

$$P(X(t_i), X(t_j)) = f(t_i - t_j)$$

nur von der Zeitdifferenz abhängen

- Um einen **ergodischen**, stochastischen Prozess zu definieren, bedarf es einer neuen Variablen

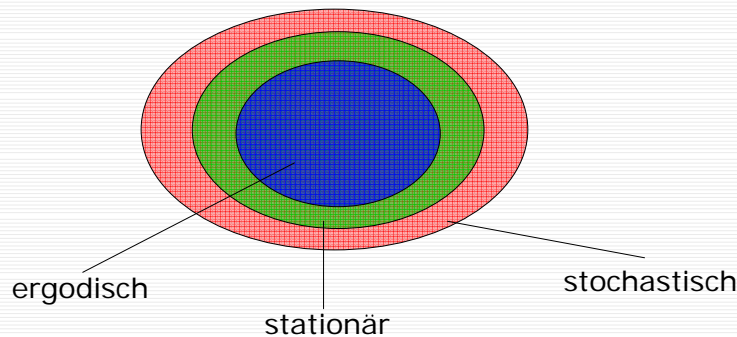
$$X_e = \lim_{r \rightarrow \infty} \frac{X(t_i) + X(t_{i+1}) + X(t_{i+2}) + \dots + X(t_{i+r})}{r+1}$$

- Wir mitteln also über die Werte der Zufallsvariablen einer immer grösseren Sequenz
- Konvergiert dieser Grenzwert gegen eine Konstante X_e gilt für eine beliebige Zufallsvariable $X(t_j)$

$$E(X(t_i)) = \langle E_s \rangle$$

- Der Erwartungswert über die Zeit ist also gleich dem Erwartungswert der Einzelvariablen

- Allgemein: Ergodische Prozesse sind Untermengen von stationären, stochastischen Prozessen

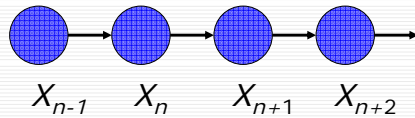


- Ein stochastischer Prozess, welcher eine Sequenz von Zufallsvariablen $\{X_j\}$ ist, heisst **Markov-Kette**, wenn gilt

$$P(X_n = x_k | X_1, \dots, X_{n-1}) = P(X_n = x_k | X_{n-1})$$

- Die bedingte Wahrscheinlichkeit für eine Zufallsvariable ist also nur von ihrem **direkten Vorgänger** abhängig
- Markov-Ketten werden mit einer Uebergangswahrscheinlichkeitsmatrix beschrieben, deren Elemente wie folgt gegeben sind

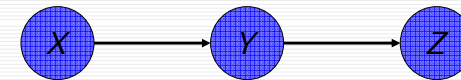
$$P_{k|l} = P(X_n = x_k | X_{n-1} = x_l)$$



- Die Berechnung der Verbundwahrscheinlichkeit erfolgt durch Produktbildung

$$P(X_{n-1}, X_n, X_{n+1}, X_{n+2}) = P(X_{n-1})P(X_n|X_{n-1})P(X_{n+1}|X_n)P(X_{n+2}|X_{n+1})$$

- Ebenso können bedingte Wahrscheinlichkeiten zu Markov-Feldern erweitert werden

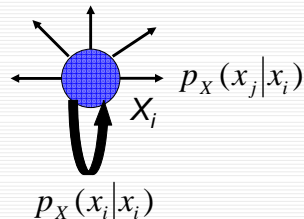


- $P(X, Y, Z) = P(X) \cdot P(Y | X) \cdot P(Z | Y)$
- Bedingte Wahrscheinlichkeit

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

- Spezialfall eines Bayes-Netztes

- Markov-Prozesse** eignen sich zur verfeinerten Modellierung einzelner diskreter Zufallsvariablen
- Dazu verwendet man einen **probabilistischen, endlichen Automaten**, welcher alle Zustände $\{x_j\}$ einer Variablen Z beschreibt
- Jeder Zustand x_i bleibt entweder erhalten, oder wird in Richtung eines anderen Zustandes x_j verlassen



- Dies wird mit Übergangswahrscheinlichkeiten ausgedrückt
- Es gilt die folgende Bedingung für jeden Knoten

$$\sum_{j=1}^N p_X(x_j | x_i) = 1$$

- Ferner gilt für die Wahrscheinlichkeit des Zustandes x_i

$$p_X(x_i) = \sum_{j=1}^N p_X(x_i | x_j) p_X(x_j)$$

- Dies ist ein Eigenwertproblem der Zustandsübergangsmatrix



Summen- und Produktregeln sind in der Wahrscheinlichkeitsrechnung von zentraler Bedeutung

Markovsche Modelle

- Allgemeine Formulierung

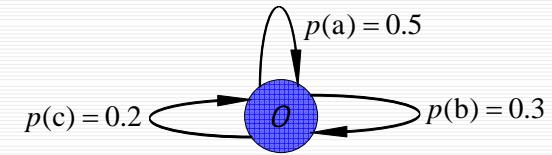
$$\begin{pmatrix} \vdots \\ p_X(x_i) \\ \vdots \end{pmatrix} = \begin{pmatrix} \ddots & & \\ & p_X(x_i|x_j) & \\ & \ddots & \ddots \end{pmatrix} \begin{pmatrix} \vdots \\ p_X(x_j) \\ \vdots \end{pmatrix}$$

- Eigenwertproblem



Fazit: Die Verbundwahrscheinlichkeit einer Sequenz hängt vom Wissen über ihre statistischen Eigenschaften ab.

Markov Modell 0-ter Ordnung ETH



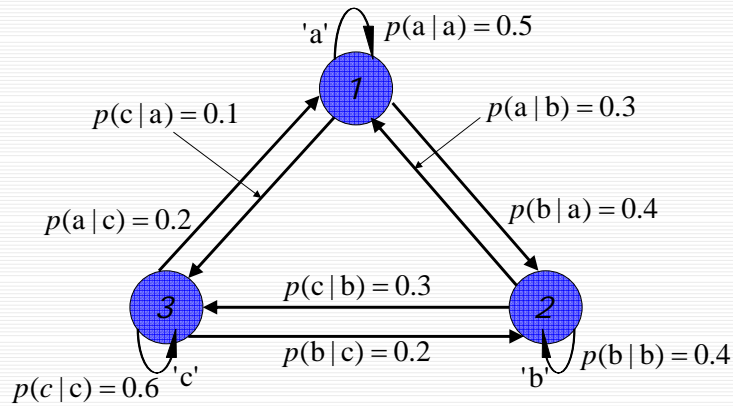
- Gegeben: Alphabet mit 3 Buchstaben
- 3 mögliche Werte der Zustandsvariable X

$$X \in \{a, b, c\}$$

- Gesucht: Wahrscheinlichkeit des Strings "acb"

$$\begin{aligned} p(aacb) &= p(a) \cdot p(a) \cdot p(c) \cdot p(b) \\ &= 0.5 \cdot 0.5 \cdot 0.2 \cdot 0.3 = 0.015 \end{aligned}$$

Markov Modell 1-ter Ordnung ETH



- $p(aacb) = p(a|a) \cdot p(a|a) \cdot p(c|a) \cdot p(b|c)$
 $= 0.5 \cdot 0.5 \cdot 0.1 \cdot 0.2 = 0.005$