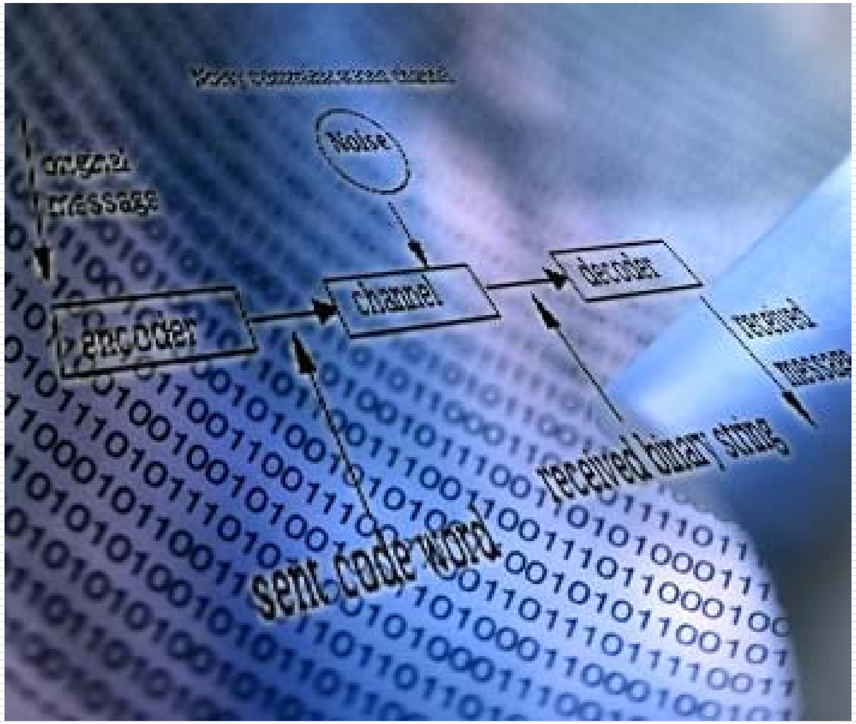
Kapitel 2:
Stochastische Prozesse



Bedingte Verteilungen

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

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

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

- Wiederum kommt eine Produktregel zur Anwendung

Summen- und Produktregeln sind in der Wahrscheinlichkeitsrechnung von zentraler Bedeutung

Beispiel: Verbunds WS

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

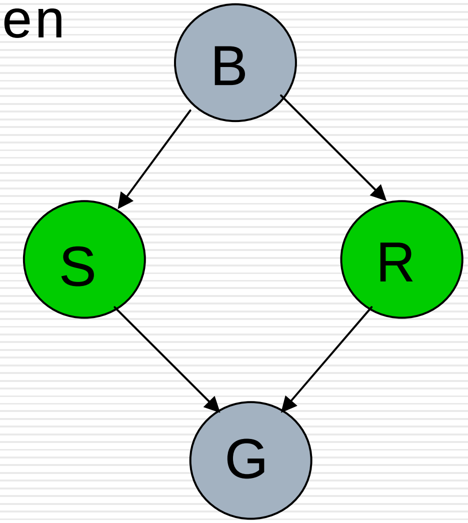
$$P(A,B) = P(A) \cdot P(B \mid A)$$

- $n=3$:

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

Bayes – Netz (1)

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



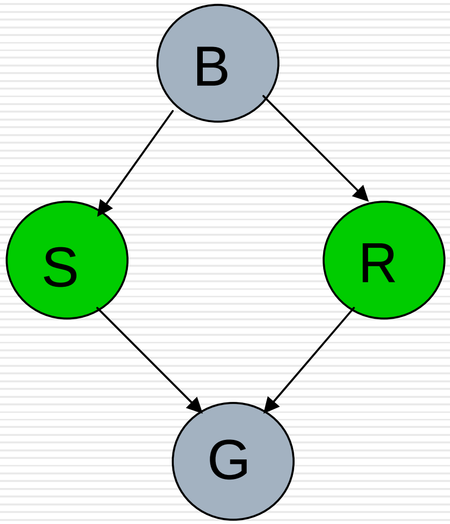
$P(B=T)$	$P(B=F)$
0.5	0.5

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



- Verbundswahrscheinlichkeit

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

$$\text{da } P(R\mid S) = P(R)$$

Bayes – Netz (3)

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

$$\mathrm{P}(S=\mathrm{T}\,|\,G=\mathrm{T}) = \frac{P(S=\mathrm{T},G=\mathrm{T})}{P(G=\mathrm{T})}$$

$$= \frac{\sum_{b,r} P(B=b,S=\mathrm{T},R=r,G=\mathrm{T})}{\sum_{b,r,s} P(B=b,S=s,R=r,G=\mathrm{T})}$$

$$= \frac{0.2781}{0.6471} = 0.4237$$

Bayes – Netz (4)

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

$$P(R=\mathrm{T} \mid G=\mathrm{T}) = \frac{P(R=\mathrm{T}, G=\mathrm{T})}{P(G=\mathrm{T})}$$

$$= \frac{\sum_{b,s} P(B=b, S=\mathrm{s}, R=\mathrm{T}, G=\mathrm{T})}{\sum_{b,r,s} P(B=b, S=s, R=r, G=\mathrm{T})}$$

$$= \frac{0.4581}{0.6471} = 0.706$$

Erwartungswert

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

$$\mathrm{E}(X) = \sum_i x_i \, p(x_i)$$

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

$$Y = aX + b \Rightarrow \mathrm{E}(Y) = a\mathrm{E}(X) + b$$

Geometrische Verteilung

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

$$\mathbf{P}(X = \mathbf{k}) = \mathbf{P}_X(k) = q^{k-1} \cdot p \qquad \textbf{wobei } q = 1\text{-}p$$

$$\mathbf{E}(X) = \sum_{k=1}^{\infty} k \cdot p \cdot q^{k-1} = p \cdot \sum_{k=1}^{\infty} k \cdot q^{k-1} \qquad \textbf{wobei } k \cdot q^{k-1} = \frac{\mathbf{d}}{\mathbf{d}q} \cdot q^k$$

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

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

Varianz

- Die Varianz Var(X) einer diskreten Zufallsvariablen X ist gegeben durch

$$\mathrm{Var}(X) = \mathrm{E}\left\{\left[X - \mathrm{E}(X)\right]^2\right\}$$

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

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

- Ebenso erhalten wir

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

Bernoulli - Verteilung

- *X* hat Bernoulli - Verteilung

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

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

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

Standardabweichung

- Die Standardabweichung σ ist gegeben durch

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

  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

$$MSE = \mathrm{E}\left[(X - x_0)^2\right]$$

Kovarianz

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

$$\mathrm{Cov}(X,Y) = \mathrm{E}\left[\left(X - \mathrm{E}(X)\right)\left(Y - \mathrm{E}(Y)\right)\right]$$

- Oftmals wird folgende Notation verwendet

$$\mathrm{E}(X) = \mu_X \qquad \mathrm{E}(Y) = \mu_Y$$

- Es gilt auch

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

oder auch

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

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

Korrelation

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

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

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

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

- Es gilt auch

$$\mathbf{E}(Y) = \sum_x \mathbf{E}(Y|X=x)\, p_X(x) = \mathbf{E}(\mathbf{E}(Y|X))$$

- Vergleiche Marginalisierung

Stochastische Prozesse

- 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).

Prozess

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

Stochastische Prozesse

- 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

$$\mathrm{P}(X(t_i) \mid X(t_j)) = \mathrm{P}(X(t_i))$$

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

Stationarität

- 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

Ergodizität

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

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

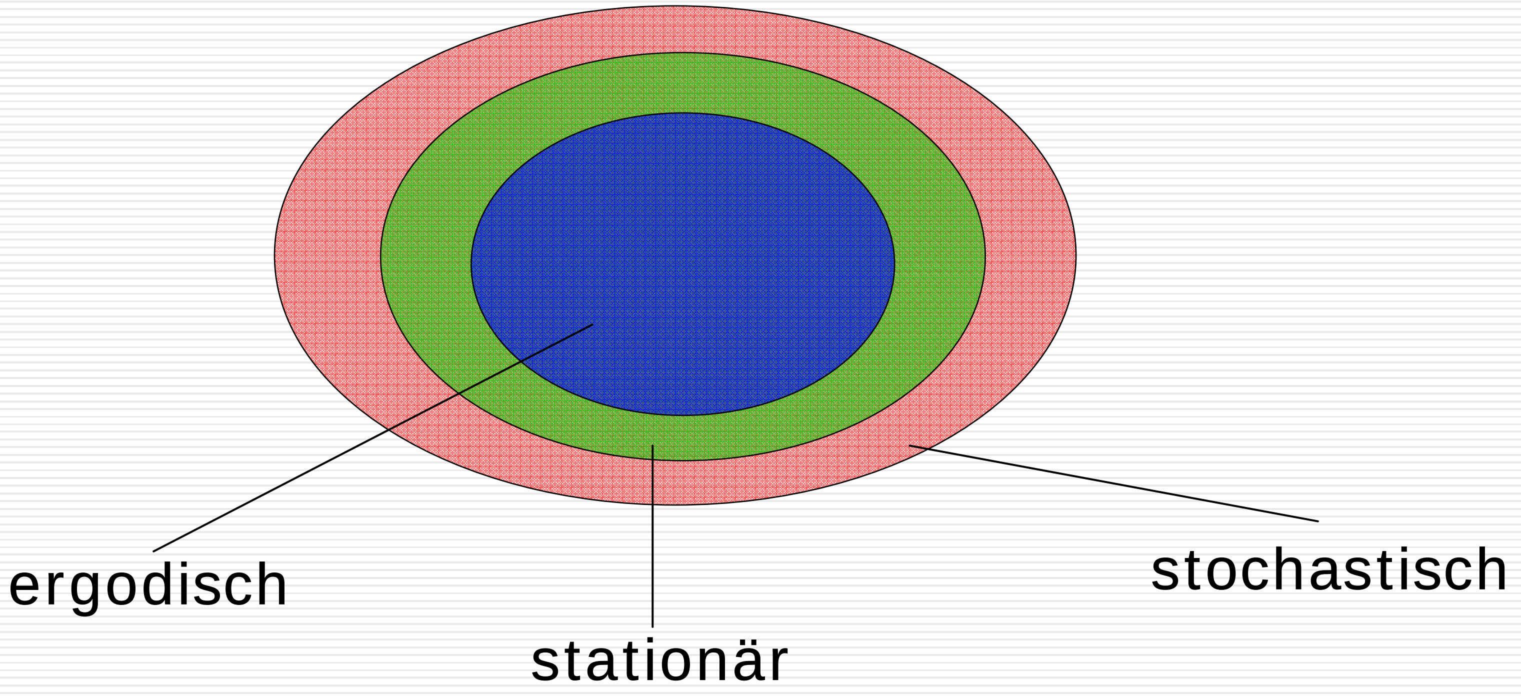
- Wir mitteln also über die Werte der Zufallsvariablen einer immer grösseren Sequenz
- Konvergiert dieser Grenzwert gegen eine Konstante X_e gilt für eine beliebige Zufallsvariable $X(t_i)$

$$\mathrm{E}(X(t_i)) = \langle \mathrm{E}_s \rangle$$

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

Stochastische Prozesse

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



Markov-Prozesse

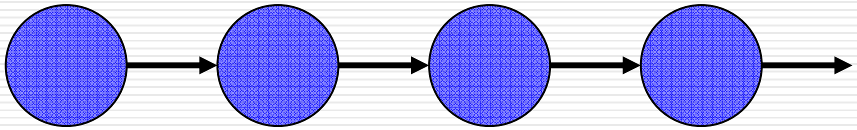
- Ein stochastischer Prozess, welcher eine Sequenz von Zufallsvariablen $\{X_i\}$ ist, heisst Markov-Kette, wenn gilt

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

- 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

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

Markov-Ketten



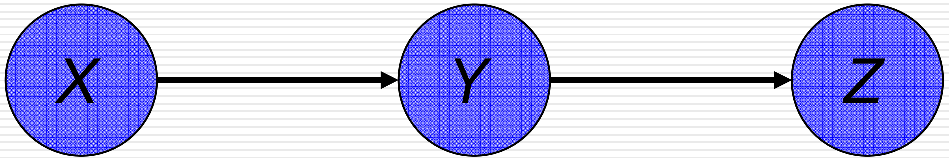
X_{n-1} X_n X_{n+1} X_{n+2}

- 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

3-stufige Markovkette



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

- Bedingte Wahrscheinlichkeit

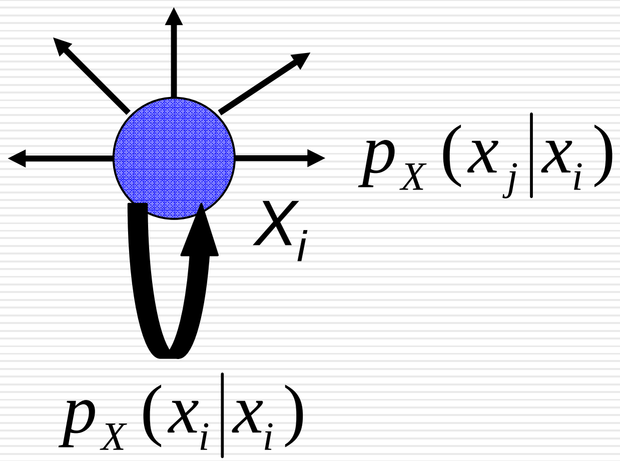
$$\mathrm{P}(X,Z\,|\,Y) = \frac{\mathrm{P}(X,Y,Z)}{\mathrm{P}(Y)} = \frac{\mathrm{P}(X,Y)\cdot \mathrm{P}(Z\,|\,Y)}{\mathrm{P}(Y)}$$

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

- Spezialfall eines Bayes-Netzes

Markov Zustandsautomaten

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



Markovsche Modelle

- 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

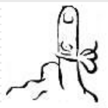
Summen- und Produktregeln sind in der Wahrscheinlichkeitsrechnung von zentraler Bedeutung

Markovsche Modelle

- Allgemeine Formulierung

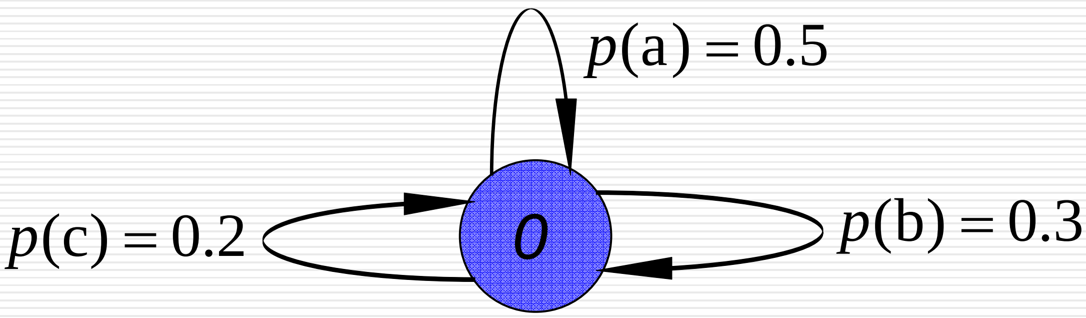
$$\begin{pmatrix} \vdots \\ p_X(x_i) \\ \vdots \end{pmatrix} = \begin{pmatrix} \ddots & & \\ & p_X(x_i \mid x_j) & \\ & \ddots & \end{pmatrix} \begin{pmatrix} \vdots \\ \\ p_X(x_j) \end{pmatrix}$$

  - Eigenwertproblem



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

Markov Modell 0-ter Ordnung



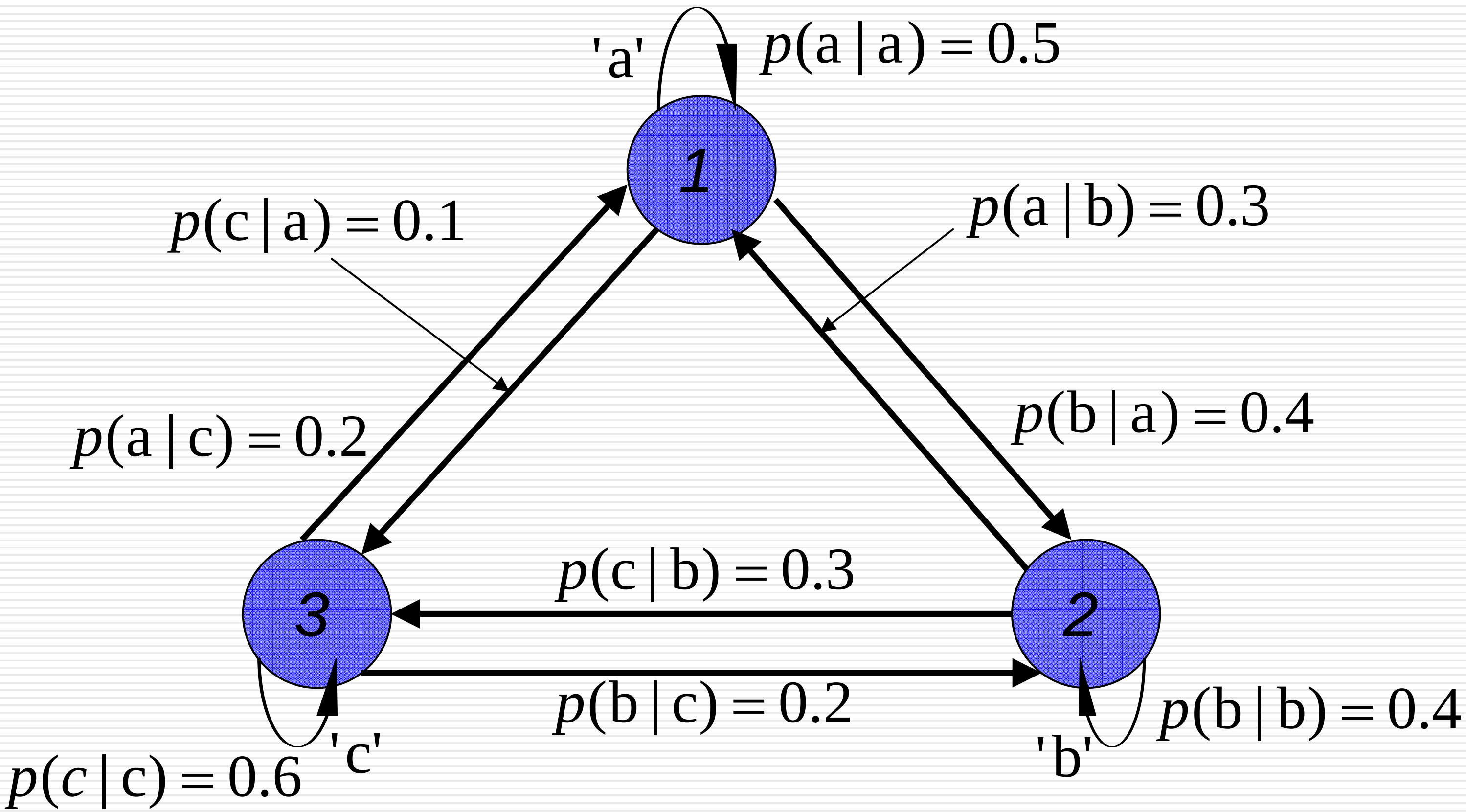
- Gegeben: Alphabet mit 3 Buchstaben
- 3 mögliche Werte der Zustandsvariable X

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

- Gesucht: Wahrscheinlichkeit des Strings “aacb“

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

Markov Modell 1-ter Ordnung ETH



- $p(\text{aacb}) = p(\text{a}|\text{a}) \cdot p(\text{a}|\text{a}) \cdot p(\text{c}|\text{a}) \cdot p(\text{b}|\text{c})$

$$= 0.5 \cdot 0.5 \cdot 0.1 \cdot 0.2 = 0.005$$