Einführung in die

Künstliche Intelligenz

SoSe 19

Prof. Dr. J. Fürnkranz, Prof. Dr. K. Kersting

TECHNISCHE UNIVERSITÄT DARMSTADT

Beispiellösung für das 6. Übungsblatt

Aufgabe 1 Alltagswahrscheinlichkeiten

a) Da es sich um eine ideale Münze handelt, ist jede mögliche Münzwurf-Sequenz gleich wahrscheinlich, und zwar $(0.5)^{10}$. Von allen möglichen 10-er Sequenzen existiert nur eine, in der zehnmal Kopf geworfen wird. Für das Ereignis 5-mal Kopf und 5-mal Zahl gibt es offensichtlich mehr als eine Möglichkeit und ist somit klar wahrscheinlicher. Alice wird die Wette gewinnen.

b) Es ist hier zu beachten, dass die Wahrscheinlichkeit, dass im nächsten Wurf Zahl fällt 0.5 ist. Das Ereignis wird weder wahrscheinlicher noch unwahrscheinlicher, wenn zuvor bereits 9-mal die Zahl geworfen wurde (siehe auch „Gambler's fallacy"). Aus a) wissen wir, dass Bob seinen Einsatz x sicher verlieren wird, falls er die zweite Wette verneint. Nimmt er die Wette an, verliert er zu 50% nichts und zu 50% verdoppelt sich der Verlust zu $2x$. Das bedeutet für den Erwartungswert $E_{Wette2}(Verlust) = x$. Da der Erwartungswert für beide Möglichkeiten identisch ist, sollte Bob indifferent bezüglich der Wahl sein (falls er *risikoneutral* ist).
In der Entscheidungstheorie gibt es für solche Fälle Konzepte, die Risikofreudigkeit als Parameter in den Entscheidungsprozess mit heranziehen. Falls Bob *risikoavers/risikoscheu* ist, d.h. eine geringere risikobehaftete Entscheidung präferiert (jedoch mit gleichem Erwartungswert) sollte er nicht auf die Zusatzwette eingehen.

c) In diesem Fall würde Bob die Wette gewinnen, da sein verstandenes Ereignis wie bei der zehnfachen Kopf Sequenz exakt einmal unter allen möglichen Sequenzen vorkommt. Somit ist die Wahrscheinlichkeit beider Ereignisse gleich und keines wahrscheinlicher als das andere. Bezüglich der Zusatzwette ändert sich nur, dass die Erwartungswerte der Entscheidungen nun beide einen Gewinn anstatt einem Verlust von x entsprechen.

Aufgabe 2 Wahrscheinlichkeitsrechnung

a) Wir zeigen (1) $\Leftrightarrow$ (2) und (1) $\Leftrightarrow$ (3).

- (1) $\Leftrightarrow$ (2)

$$P(a,b|c) = P(a|c)\cdot P(b|c) \Leftrightarrow \frac{P(a,b,c)}{P(c)} = P(a|c)\cdot \frac{P(b,c)}{P(c)} \Leftrightarrow \frac{P(a,b,c)}{P(c)}\cdot\frac{P(c)}{P(b,c)} = P(a|c)$$

$$\Leftrightarrow \frac{P(a,b,c)}{P(b,c)} = P(a|c) \Leftrightarrow P(a|b,c) = P(a|c)$$

- (1) $\Leftrightarrow$ (3)

$$P(a,b|c) = P(a|c)\cdot P(b|c) \Leftrightarrow \frac{P(a,b,c)}{P(c)} = P(a|c)\cdot P(b|c) \Leftrightarrow \frac{P(a,b,c)}{P(a,c)} = P(b|c) \Leftrightarrow P(b|a,c) = P(b|c)$$

b) Bei (2) ist ausreichend Information vorhanden, um die gemeinsame Wahrscheinlichkeitsverteilung $P(e_1,e_2,h)$ zu rekonstruieren. Wendet man den Satz von Bayes an, nämlich

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

erhält man

$$P(h|e_1,e_2) = \frac{P(e_1,e_2|h)\cdot P(h)}{p(e_1,e_2)}$$

Bei (3) sind ebenfalls ausreichend Informationen vorhanden, denn wir können die Verbundwahrscheinlichkeit mit Hilfe der Kettenregel sowie mittels Marginalisierung rekonstruieren:

$$\begin{aligned} P(H,E_1,E_2) &= P(E_1|H,E_2)P(E_2|H)P(H) \\ &= P(E_1|H,E_2)P(E_2|H)\sum_{E_1} P(H|E_1)P(E_1) \end{aligned}$$

Anschließend kann die gewünschte bedingte Wahrscheinlichkeit mittels $P(h|e_1,e_2) = \frac{P(h,e_1,e_2)}{\sum_H P(H,e_1,e_2)}$ inferiert werden.

Für (1) kann ohne zusätzliche Annahmen die gemeinsame Verteilung aus den bedingten Wahrscheinlichkeiten nicht rekonstruiert werden.

c) Mit $P(e_1|h,e_2) = P(e_1|h)$ kann man folgendermassen umformen:

$$P(h|e_1,e_2) = \frac{P(h,e_1,e_2)}{P(e_1,e_2)} = \frac{P(e_1|h,e_2)\cdot P(e_2,h)}{P(e_1,e_2)} = \frac{P(e_1|h)\cdot P(e_2|h)\cdot P(h)}{P(e_1,e_2)}$$

Somit kann man nun auch aus den Angaben in (1) $P(h|e_1,e_2)$ bestimmen.

Aufgabe 3 Monty-Hall-Problem

Intuitiv:

In zwei von drei Fällen wird die erste Wahl auf ein Tor mit einer Niete gefallen sein. In diesen Fällen hat der Moderator die zweite Niete gezeigt, d.h. hinter dem dritten Tor ist sicher ein Gewinn. Im dritten Fall, in dem der Spieler zuerst den Gewinn besetzt hat, ist das verbleibende Tor eine Niete, und er verliert beim Wechsel. Die Entscheidung das andere Tor zu nehmen hat somit eine Gewinnwahrscheinlichkeit von $\frac{2}{3}$.

Tabellarisch:

Kandidat wählt Tor 1 und wechselt wenn anderes Tor geöffnet

Moderator möchte Tor 2 öffnen					Moderator möchte Tor 3 öffnen				
Fall	**Tor 1**	Tor 2	Tor 3	Resultat	Fall	**Tor 1**	Tor 2	Tor 3	Resultat
1	**A**	Z	Z	Verlust: Wechsel auf Tor 3	4	**A**	Z	Z	Verlust: Wechsel auf Tor 2
2	**Z**	A	Z	Gewinn: Wechsel auf Tor 2	5	**Z**	A	Z	Gewinn: Wechsel auf Tor 2
3	**Z**	Z	A	Gewinn: Wechsel auf Tor 3	6	**Z**	Z	A	Gewinn: Wechsel auf Tor 3

Formal:

- 3 Tore: A, B, C,
- je Möglichkeiten für jedes Tor, g (Gewinn), n (Niete)

Wir haben drei mögliche Ereignisse, die eine Wahrscheinlichkeit > 0 haben.

Ereignis	Tor A	Tor B	Tor C	$\Pr(e_i)$
e_A	g	n	n	$\frac{1}{3}$
e_B	n	g	n	$\frac{1}{3}$
e_C	n	n	g	$\frac{1}{3}$

Ohne Beschränkung der Allgemeinheit nehmen wir an, daß sich der Spieler für das Tor A entscheidet.

Weiters gibt es noch das Ereignis T, dass der Moderator ein Tor öffnet. Da sich der Spieler für A entschieden hat, kann der Moderator nur Tor B oder C öffnen, d.h. $\Pr(T = A) = 0$.

Die Wahrscheinlichkeit, mit der der Moderator eines der beiden anderen Tore öffnet, richtet sich nach den vorliegenden Ereignissen:

Ereignis	Tor A	Tor B	Tor C	$\Pr(T = B\|e_i)$	$\Pr(T = C\|e_i)$
e_A	g	n	n	$\frac{1}{2}$	$\frac{1}{2}$
e_B	n	g	n	0	1
e_C	n	n	g	1	0

Nehmen wir nun an, der Moderator öffnet Tür B. Was wir nun berechnen müssen, ist die Wahrscheinlichkeit, daß Ereignis e_A vorliegt, wenn wir wissen, dass B geöffnet wurde, also $\Pr(e_A|T = B)$.

$$\Pr(e_A|T = B) = \frac{\Pr(T = B|e_A) \cdot \Pr(e_A)}{\Pr(T = B)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

Analog wäre

$$\Pr(e_C|T = B) = \frac{\Pr(T = B|e_C) \cdot \Pr(e_C)}{\Pr(T = B)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Der Kandidat sollte somit das andere Tor wählen.

Aufgabe 4 Bayes'sches Netz

a) Nach Definition repräsentiert ein BN die Verbundwahrscheinlichkeitsverteilung wie folgt:

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

Im vorliegenden Fall gilt also:

$$P(L,W,D,G) = P(D)P(L)P(W|D,L)P(G|W)$$

Damit lässt sich die volle Tabellenrepräsentation von $P(L,W,D,G)$ berechnen:

L	W	D	G	$P(L,W,D,G)$
T	T	T	T	$0.2268 = P(G \mid W) \cdot P(W \mid D, L) \cdot P(D) \cdot P(L)$
T	T	T	F	$0.0252 = P(\neg G \mid W) \cdot P(W \mid D, L) \cdot P(D) \cdot P(L)$
T	T	F	T	$0.0864 = \ldots$
T	T	F	F	0.0096
T	F	T	T	0.0216
T	F	T	F	0.0864
T	F	F	T	0.0288
T	F	F	F	0.1152
F	T	T	T	0.108
F	T	T	F	0.012
F	T	F	T	0.0432
F	T	F	F	0.0048
F	F	T	T	0.024
F	F	T	F	0.096
F	F	F	T	0.0224
F	F	F	F	0.0896

b) Gesucht ist $P(W|\neg D, G)$. Wegen der Netzstruktur ist Wohlfühlen von Licht und Dünger abhängig, also P(W|L,D) und Gedeihen von Wohlfühlen, also P(G|W).

$$\begin{aligned}
P(W|\neg D, G) &= \frac{P(G, W, \neg D)}{P(G, \neg D)} \\
&= \frac{P(G, W, \neg D, L) + P(G, W, \neg D, \neg L)}{P(G, W, \neg D, L) + P(G, W, \neg D, \neg L) + P(G, \neg W, \neg D, L) + P(G, \neg W, \neg D, \neg L)} \\
&= \frac{0.1296}{0.1296 + 0.0512} \approx 0.7168
\end{aligned}$$

mit Benutzung von:

$$\begin{aligned}
P(G, W, \neg D, L) &= P(G|W)P(W|L, \neg D)P(\neg D)P(L) \\
&= 0.9 \cdot 0.4 \cdot (1 - 0.6) \cdot 0.6 = 0.0864 \\
P(G, W, \neg D, \neg L) &= P(G|W)P(W|\neg L, \neg D)P(\neg D)P(\neg L) \\
&= 0.9 \cdot 0.3 \cdot (1 - 0.6) \cdot (1 - 0.6) = 0.0432 \\
P(G, W, \neg D) &= 0.0864 + 0.0432 = 0.1296 \\
P(G, \neg W, \neg D, L) + P(G, \neg W, \neg D, \neg L) &= P(G|\neg W)P(\neg W|L, \neg D)P(L)P(\neg D) + P(G|\neg W)P(\neg W|\neg L, \neg D)P(\neg L)P(\neg D) \\
&= 0.0288 + 0.0224 = 0.0512
\end{aligned}$$

c) Die bedingten Unabhängigkeiten können z.B. über die Regeln der Wahrscheinlichkeitsrechnung hergeleitet werden:

- D und L sind zunächst unabhängig, werden aber abhängig, sobald W oder G bekannt sind. $P(D, L) = P(D)P(L) \implies D \perp\!\!\!\perp L$.
- Ist W bekannt, so ist G von L und D unabhängig, weil alle für G relevanten Informationen in W enthalten sind. $P(G|W, L, D) = P(G|W) \implies G \perp\!\!\!\perp L, D|W$.

Aufgabe 5 Inferenz in BNs

a) Berechnet werden soll $P(a)$. Wir zeigen einige unterschiedlich effiziente Varianten:

1. *Streng nach Schema*: Berechnung über die gemeinsame Verteilung, Berücksichtigung der Netzwerkstruktur, Aufsummieren über alle unbekannten Variablen (merke: Alarm (a) gilt).

$$P(a) = \sum_{B,E,J,M} P(a,B,E,J,M) = \sum_{B,E,J,M} P(E) \cdot P(B) \cdot P(a|B,E) \cdot P(J|a) \cdot P(M|a)$$

Die Summe läuft über **alle** $2 \times 2 \times 2 \times 2 = 16$ Möglichkeiten, die Variablen B, E, J, M mit Werten zu belegen. Der erste Term ist also

$$P(e) \cdot P(b) \cdot P(a|b,e) \cdot P(j|a) \cdot P(m|a) = 0.002 \cdot 0.001 \cdot 0.95 \cdot 0.9 \cdot 0.7 = 0.000001197$$

der zweite Term ist

$$P(e) \cdot P(\neg b) \cdot P(a|\neg b,e) \cdot P(j|a) \cdot P(m|a) = 0.002 \cdot 0.999 \cdot 0.29 \cdot 0.9 \cdot 0.7 = 0.00037$$

...

der letzte Term ist

$$P(\neg e) \cdot P(\neg b) \cdot P(a|\neg b,\neg e) \cdot P(\neg j|a) \cdot P(\neg m|a) = 0.998 \cdot 0.999 \cdot 0.001 \cdot 0.1 \cdot 0.3 = 0.00002991.$$

2. *Intelligent nach Schema*: Berechnung über die gemeinsame Verteilung, Berücksichtigung der Netzwerkstruktur, intelligente Berechnung der Summe über alle unbekannten Variablen.

Die oben stehende Summe lässt sich durch Herausheben von Termen auch umformen:

$$\begin{aligned}\sum_{B,E,J,M} P(E) \cdot P(B) \cdot P(a|B,E) \cdot P(J|a) \cdot P(M|a) &= \sum_J \sum_M \sum_B \sum_E P(E) \cdot P(B) \cdot P(a|B,E) \cdot P(J|a) \cdot P(M|a) \\ &= \left(\sum_J P(J|a)\right) \cdot \left(\sum_M P(M|a)\right) \cdot \sum_B \sum_E P(E) \cdot P(B) \cdot P(a|B,E)\end{aligned}$$

Es gilt natürlich $\sum_J P(J|a) = P(j|a) + P(\neg j|a) = 1$ und analog dazu $\sum_M P(M|a) = 1$.

Daraus ergibt sich

$$P(a) = \sum_B \sum_E P(E) \cdot P(B) \cdot P(a|B,E) \tag{1}$$

Zur Berechnung dieser Summe benötigt man lediglich $2 \times 2 = 4$ Terme.

$$\begin{aligned}P(a) &= \sum_B \sum_E P(E) \cdot P(B) \cdot P(a|B,E) \\ &= P(e) \cdot P(b) \cdot P(a|b,e) \\ &+ P(\neg e) \cdot P(b) \cdot P(a|b,\neg e) \\ &+ P(e) \cdot P(\neg b) \cdot P(a|\neg b,e) \\ &+ P(\neg e) \cdot P(\neg b) \cdot P(a|\neg b,\neg e) \\ &= 0.002 \cdot 0.001 \cdot 0.95 + 0.998 \cdot 0.001 \cdot 0.94 + 0.002 \cdot 0.999 \cdot 0.29 + 0.998 \cdot 0.999 \cdot 0.001 \\ &\approx 0.00252\end{aligned}$$

Wie in der Vorlesung gezeigt, könnte man die Berechnung durch die Definition von Faktoren in der Praxis noch weiter optimieren.

3. *Mit Überlegen:* Auf die Gleichung (1) kann man auch sofort kommen, wenn man ein bißchen überlegt.
   Es gilt natürlich

$$P(a,B,E) = P(a|B,E) \cdot P(B,E)$$

   und da wir aus der Netzwerkstruktur ablesen können, daß B und E unabhängig sind, gilt

$$P(B,E) = P(B) \cdot P(E),$$

   woraus (1) folgt.

b) Zu berechnen ist nun $P(a|j) = \frac{P(a,j)}{P(j)}$. Zur Berechnung von

$$P(a,j) = \sum_{B,E,M} P(E) \cdot P(B) \cdot P(a|B,E) \cdot P(j|a) \cdot P(M|a)$$

würden wir $2 \times 2 \times 2 = 8$ Terme benötigen.

Aber natürlich gilt

$$P(a,j) = P(j|a) \cdot P(a)$$

Da wir $P(a)$ in a) berechnet haben, und $P(j|a)$ gegeben ist, können wir direkt berechnen

$$P(a,j) \approx 0.00252 \cdot 0.9 \approx 0.00227$$

Die noch fehlende Berechnung von $P(j)$ könnte analog zu a) passieren.

Idee:

Natürlich gilt auch $P(j) = P(j,a) + P(j,\neg a)$, sodaß man anstatt $P(j)$ auch $P(j,\neg a)$ berechnen kann, und $P(a|j)$ wird dann direkt als

$$P(a|j) = \frac{P(a,j)}{P(j)} = \frac{P(a,j)}{P(a,j) + P(\neg a,j)}$$

berechnet.

$P(\neg a,j)$ können wir aber wieder einfach berechnen als

$$\begin{aligned} P(\neg a,j) &= P(j|\neg a) \cdot P(\neg a) = P(j|\neg a) \cdot (1 - P(a)) \\ &\approx 0.05 \cdot (1 - 0.00252) = 0.05 \cdot 0.99748 \approx 0.04987 \end{aligned}$$

Nun ergibt sich

$$P(a|j) = \frac{P(a,j)}{P(j)} = \frac{P(a,j)}{P(j,a) + P(j,\neg a)} \approx \frac{0.00227}{0.00227 + 0.04987} \approx 0.04353$$

Durch die zusätzliche Evidenz, dass John angerufen hat, ist die Wahrscheinlichkeit dafür, dass ein Alarm losgangen ist, ungefähr 17-mal so hoch wie ohne diese Evidenz.

Anmerkung 1:

Unsere Ableitung zu Beginn entspricht einfach der Anwendung des Satz von Bayes:

$$P(a|j) = \frac{P(j|a) \cdot P(a)}{P(j)}$$

Dieser Satz ist oft nützlich um eine unbekannte konditionale Wahrscheinlichkeit ($P(a|j)$) auf eine bekannte ($P(j|a)$) zurückzuführen.

c) Nun sollen wir $P(a|j,e)$ berechnen.

Die gemeinsame Verteilung lässt sich schon recht schnell nach Schema analog zu a) berechnen, da nun weniger unbekannte Variablen vorhanden sind (nämlich nur noch B und M) und wie in a) bereits festgestellt $\sum_M P(M|a) = 1$ gilt:

$$\begin{aligned}
P(a,j,e) &= \sum_M \sum_B P(e) \cdot P(B) \cdot P(a|B,e) \cdot P(j|a) \cdot P(M|a) \\
&= P(j|a) \cdot P(e) \cdot \left(\sum_M P(M|a) \right) \sum_B P(B) \cdot P(a|B,e) \\
&= P(j|a) \cdot P(e) \cdot 1 \cdot \sum_B P(B) \cdot P(a|B,e) \\
&= P(j|a) \cdot P(e) \cdot (P(b) \cdot P(a|b,e) + P(\neg b) \cdot P(a|\neg b,e)) \\
&= 0.9 \cdot 0.002 \cdot (0.001 \cdot 0.95 + 0.999 \cdot 0.29) \\
&= 0.0018 \cdot (0.00095 + 0.28971) \approx 0.000523
\end{aligned}$$

Analog zu b) brauchen wir nun entweder $P(j,e)$ oder $P(\neg a,j,e)$. Wir nehmen letztes und berechnen wie oben

$$\begin{aligned}
P(\neg a,j,e) &= P(j|\neg a) \cdot P(e) \cdot (P(b) \cdot P(\neg a|b,e) + P(\neg b) \cdot P(\neg a|\neg b,e)) \\
&= 0.05 \cdot 0.002 \cdot (0.001 \cdot 0.05 + 0.999 \cdot 0.71) \\
&= 0.0001 \cdot (0.00005 + 0.70929) \approx 0.000071
\end{aligned}$$

Wir erhalten nun

$$P(a|j,e) = \frac{P(a,j,e)}{P(a,j,e) + P(\neg a,j,e)} \approx \frac{0.000523}{0.000523 + 0.000071} \approx 0.88057$$

Wenn wir wissen, dass John angerufen hat und ein Erdbeben stattfand, können wir also zu fast 90% sicher sein, dass der Alarm losgegangen ist.

Zum Vergleich: Wenn wir nur wüssten, dass ein Erdbeben stattfand (ohne den Anruf von John) liegt die Wahrscheinlichkeit, dass dabei ein Alarm losgegangen ist, nur bei ca. 0.29:

$$P(a|e) = P(a|e,b) \cdot P(b) + P(a|e,\neg b) \cdot P(\neg b) = 0.95 \cdot 0.001 + 0.29 \cdot .999 \approx 0.29$$

d) Nachdem wir nun erfahren, dass Mary nicht angerufen hat, sollen wir $P(a|e,j,\neg m)$ berechnen.

Man könnte nun $P(a,j,e,\neg m)$ und $P(\neg a,j,e,\neg m)$ analog zu c) berechnen (geht sogar noch schneller), mit etwas Überlegung kann man aber noch einfacher zum Ziel gelangen. Klarerweise gilt:

$$P(a,j,e,\neg m) = P(\neg m|a,j,e) \cdot P(a,j,e) \tag{2}$$

Nachdem in jedem Bayes'schen Netz gilt, daß die Belief Function in einem Knoten nur von seiner Markov-Decke (also den Eltern, den Kindern, und den anderen Eltern der Kinder) abhängt, und die Markov-Decke für M nur aus dem Knoten A besteht, gilt:

$$P(\neg m|a,j,e) = P(\neg m|a)$$

und (2) wird dann zu

$$P(a,j,e,\neg m) = P(\neg m|a) \cdot P(a,j,e) = 0.3 \cdot 0.000523 \approx 0.000157 \tag{3}$$

da wir ja $P(a,j,e) = 0.000523$ schon in c) berechnet haben.

Analog erhalten wir

$$P(\neg a, j, e, \neg m) = P(\neg m|\neg a) \cdot P(\neg a, j, e) = 0.99 \cdot 0.000071 = 0.000070 \tag{4}$$

und

$$P(a|j, e, \neg m) = \frac{P(a, j, e, \neg m)}{P(a, j, e, \neg m) + P(\neg a, j, e, \neg m)} = \frac{0.000157}{0.000157 + 0.000070} \approx 0.6916$$

Die Tatsache, daß Mary nicht angerufen hat, führt also wieder zu einer leichten Beruhigung, die Wahrscheinlichkeit, daß der Alarm losgegangen ist, ist aber immer noch recht hoch.

Aufgabe 6 Density Networks

Für den Posterior $p(z|x)$ gilt allgemein:

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} = \frac{p(x|z)p(z)}{\int_z p(x|z)p(z)}.$$

$p(x|z)$ und $p(z)$ sind gegeben, problematisch ist jedoch das Integral. Das beschriebene Modell zeichnet sich dadurch aus, dass die bedingte Wahrscheinlichkeitsverteilung $p(x|z)$ außerordentlich komplex ist, weil sie durch ein neuronales Netz repräsentiert wird. Das Integral lässt sich daher im Allgemeinen nicht analytisch lösen und muss stattdessen approximiert werden. Ein Ansatz dazu sind Monte Carlo Methoden: Wir ziehen zufällig Werte für z aus $p(z)$, und approximieren den Wert des Integrals mit dem Durchschnitt der dabei berechneten Ergebnisse:

$$\int_z p(x|z)p(z) = \mathbb{E}_{p(z)}[p(x|z)] \approx \frac{1}{N}\sum_{i=1}^{N} p(x|z^i), \quad z^i \sim p(z)$$

Dies entspricht dem in der Vorlesung vorgestellten Likelihood Weighting: Die Samples z^i werden danach gewichtet, wie gut sie zur beobachteten Evidenz x passen. Auf diese Weise bekommt man zwar in der Theorie bei $N \to \infty$ ein exaktes Ergebnis, in der Praxis kann es aber sein, dass sehr viele Samples benötigt werden, um solche z^i zu finden, welche nach dem echten Posterior $p(z|x)$ wahrscheinlich sind. Dies ist insbesondere dann der Fall, wenn z hochdimensional ist. Eine Möglichkeit, diesem Problem zu begegnen, ist die in der Vorlesung vorgestellte Methode *Markov Chain Monte Carlo*. Hierbei werden die Samples nicht unabhängig gezogen, sondern schrittweise angepasst. Das kann zum Beispiel passieren, indem in jedem Schritt nur eine zufällig gewählte Variable z_k neu gezogen wird, während alle anderen $z_{\neq k}$ gleich bleiben, also

$$z_k^{i+1} \sim p(z_k|z_{\neq k}^t, x), \quad z_{\neq k}^{t+1} = z_{\neq k}^t$$

Mit auf diese Weise generierten Samples konvergiert das Verfahren immer noch gegen das korrekte Ergebnis, findet aber in der Regel schneller relevante Samples.

Zum Einsatz kommen solche Modelle (leicht abgewandelt) beispielsweise in tiefen generativen Modellen wie dem Variational Autoencoder (VAE) oder Generative Adversarial Networks (GANs). Sie haben zum Ziel, komplexe Wahrscheinlichkeitsverteilungen wie etwa Verteilungen über natürliche Bilder zu modellieren, und führen dazu latente Variablen z ein. Auch im Kontext von größeren Bayes'schen Netzen können Density Networks eingesetzt werden, um komplexe Verteilungen zu modellieren, für die keine einfache parametrische Form bekannt ist. Inferenz ist ein Kernproblem in solchen Modellen. In VAEs wird dazu *Variational Inference* eingesetzt, ein anderes approximatives Verfahren, welches in dieser Veranstaltung nicht näher behandelt werden kann.