ETH Zürich
Dept. Informatik

Prof. Markus Gross
WS 05/06

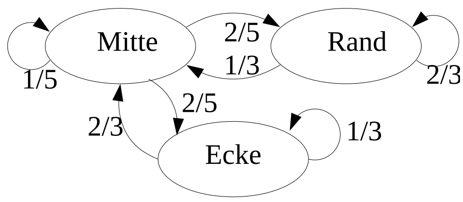
Informationstheorie

Lösung 5

5.1 Schach als Informationsquelle (aus dem Vordiplom Herbst 2000)

a) Der Läufer kann in 8 verschiedenen Feldern sein, damit gibt es 8 mögliche Zustände.

b) Durch die Symmetrien kann die Quelle wie folgt vereinfacht werden



Das resultierende lineare Gleichungssystem ist dann:

$$\begin{pmatrix} \frac{1}{5} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{5} & \frac{2}{3} & 0 \\ \frac{2}{5} & 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} p_m \\ p_r \\ p_e \end{pmatrix} = \begin{pmatrix} p_m \\ p_r \\ p_e \end{pmatrix}$$

Unter der Bedingung dass $p_m + p_r + p_e = 1$ erhält man als Lösung

$$p_m = \frac{5}{14} \quad p_r = \frac{3}{7} \quad p_e = \frac{3}{14}$$

Damit ergibt sich für die stationäre Verteilung $\bar{P}_X$ der Informationsquelle

$$\bar{P}_X(x) = \begin{cases} 5/28 & \text{falls } x \text{ ein Mittelfeld} \\ 3/28 & \text{falls } x \text{ ein Randfeld} \\ 3/28 & \text{falls } x \text{ ein Eckfeld} \end{cases}$$

c) Ist der Läufer in der Mitte gibt es 5 Zug-Möglichkeiten, in einer Ecke und am Rand 3, die alle gleichwahrscheinlich sind. Damit ist die Entropierate

$$\begin{aligned} \bar{H} &= p_m \cdot H[\frac{1}{5}; \frac{1}{5}; \frac{1}{5}; \frac{1}{5}; \frac{1}{5}] + p_r \cdot H[\frac{1}{3}; \frac{1}{3}; \frac{1}{3}] + p_e \cdot H[\frac{1}{3}; \frac{1}{3}; \frac{1}{3}] \\ &= p_m \cdot \log_2(5) + (p_r + p_e) \cdot \log_2(3) \approx 1.848 \end{aligned}$$

d) Da die zwei Läufer voneinander unabhängig sind erhalten wir $8 \times 8 = 64$ Zustände, die stationäre Verteilung ist $\bar{P}_{X_1 X_2}(x_1, x_2) = \bar{P}_X(x_1) \cdot \bar{P}_X(x_2)$ and die Entropierate ist $2\bar{H} \approx 2.268$.

5.2 Informationsdichte der Schweizer Presse

Dieses Methode wurde übrigens von Shannon vorgeschlagen. Seine Experimente für englische Texte ergaben eine Entropie von etwa 1.3 bit pro Buchstaben.

a) Der Codierer kann auch zum Decodieren verwendet werden, indem die Antworten auf seine Fragen aus der Zahlenfolge gebildet werden. Auf die ersten $(M_i - 1)$ Fragen nach dem i-ten Buchstaben antworten wir mit „Nein“, auf die folgende Frage (die uns den i-ten Buchstaben verrät) dann mit „Ja“.

b) Wir können aus dem Text die Zahlenfolge und aus der Zahlenfolge den Text berechnen. Deshalb müssen beide Entropien gleich sein!

Wir haben den folgenden Text aus dem Prolog zu Kants „Kritik der reinen Vernunft“ gewählt:

Die menschliche Vernunft hat das besondere Schicksal in einer Gattung ihrer Erkenntnisse: dass sie durch Fragen belästigt wird, die sie nicht abweisen kann; denn sie sind ihr durch die Natur der Vernunft selbst aufgegeben, die sie aber auch nicht beantworten kann; denn sie übersteigen alles Vermögen der menschlichen Vernunft.

Als Zahlenfolge erhielten wir:

[1, 3, 1, 1, 8, 2, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 10, 1, 1, 14, 4, 1, 1, 1, 1, 19, 1, 1, 1, 5, 5, 1, 1, 6, 1, 4, 5, 1, 1, 1, 1, 1, 1, 6, 8, 1, 4, 15, 2, 1, 1, 1, 1, 5, 2, 1, 2, 1, 1, 1, 2, 1, 10, 2, 5, 7, 2, 1, 1, 1, 3, 12, 1, 1, 1 , 1, 4, 4, 3, 1, 1, 1, 8, 2, 1, 1, 1, 4, 1, 1, 1, 14, 4, 1, 1, 1, 3, 5, 1, 1, 1, 1, 16, 1, 1, 1, 1, 1, 1, 6, 1, 15, 2, 5, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 3, 3, 1, 1, 1, 1, 1, 1, 1, 1, 4, 17, 18, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 5, 2, 1, 1, 2, 1, 1, 1, 2, 1, 3, 1, 1, 5, 12, 1, 2, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1, 12, 1, 5, 1, 2, 1, 2, 1, 1, 1, 9, 1, 1, 1, 1, 1, 1, 1, 1, 4, 6, 4, 1, 5, 1, 1, 5, 1, 1, 16, 1, 2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 2, 4, 2, 1, 4, 17, 5, 1, 1, 5, 1, 1, 1, 1, 12, 1, 1, 1, 1, 1, 6, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 4, 1, 1, 2, 2, 1, 1, 1, 12, 10, 1, 1, 1, 7, 5, 1, 2, 1, 2, 1, 1, 4, 4, 1, 1, 6, 1, 11, 4, 6, 8, 5, 1, 1, 1, 1, 1, 1, 1, 2, 1, 9, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]

c) Wir zählen, wie oft jeder Wert j in der Sequenz $M_1, \ldots, M_t$ auftritt und nennen diese Zahl n_j. Daraus lässt sich nun die Entropie abschätzen

$$\bar{H} \approx - \sum_{j:n_j \neq 0} \frac{n_j}{t} \cdot \log_2(\frac{n_j}{t})$$

In unserem Beispiel ergibt dies

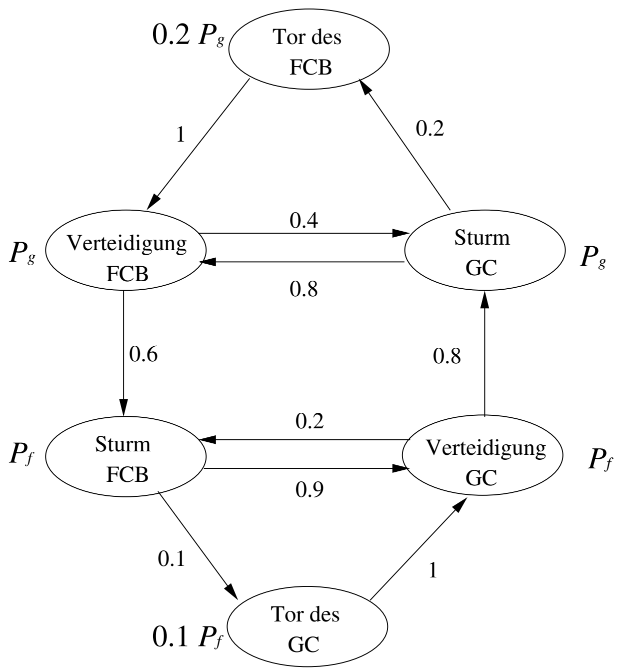
$$\bar{H} = 1.956.$$

Der obige Text scheint also relativ viel Entropie zu enthalten. Zumindest war die Redundanz für unseren Codierer nicht offensichtlich.

d) Es gibt häufig Folgen von Einsen, da der Codierer oft Worte oder Satzteile richtig errät. Wir können die Schätzung verbessern, indem wir für lange Einserfolgen neue Zeichen einführen.

5.3 Fussball als Informationsquelle

a) Die Zustände und Übergangswahrscheinlichkeiten sind in der Figur illustriert. Wir berechnen die stationäre Verteilung. Seien P_f und P_g die Wahrscheinlichkeiten, dass der Ball im Sturm des FCB bzw. des GC ist. Dann ist klar, dass erstens die Wahrscheinlichkeiten der Tore $0.2P_g$ und $0.1P_f$ sind, und dass zweitens die Wahrscheinlichkeiten des Ballbesitzes der Verteidiger P_f bzw. P_g sind. Einerseits gelangt nämlich der Ball nur vom Sturm des einen Vereins ins Tor des andern, und andererseits nur vom Tor des FCB oder vom Sturm des GC in die Verteidigung des FCB (und umgekehrt).



Für die Unbekannten P_f und P_g gelten die Gleichungen

$$2.1P_f + 2.2P_g = 1$$

(Summe aller Wahrscheinlichkeiten ist 1) und beispielsweise

$$P_g = 0.8P_f + 0.4P_g$$

(Gleichung des Knotens "Sturm GC"). Die Lösung der Gleichungen ist $P_f = 30/151$ und $P_g = 40/151$, womit alle Wahrscheinlichkeiten der stationären Verteilung bestimmt sind.

b) Der GC gewinnt mit zu erwartendem Torverhältnis $0.2P_g/0.1P_f = 8 : 3$.

c) Die Entropierate der Zustandsfolge ist

$$P_f \cdot h(0.2) + P_f \cdot h(0.1) + P_g \cdot h(0.4) + P_g \cdot h(0.2) \approx 0.685.$$

5.4 One-Time Pad

- $H(C|KM) = 0, H(K|MC) = 0, H(M|KC) = 0$.
- $I(M;K) = 0$, da der Schlüssel K unabhängig von M generiert wird.
- $H(K) = I(K;C|M)$, da $H(K|MC) = 0$ und $I(M;K) = 0$.
- $H(C) \leq H(K) = N$, denn C und K sind Zufallsvariablen, welche die gleiche Anzahl Werte annehmen, und K ist gleichverteilt.

  $\Rightarrow I(M;C) = 0$.