Vorlesung 14

Gemeinsame Entropie
und
bedingte Entropie

Wir betrachten ein zufälliges Paar (X, Y)

mit Verteilungsgewichten $\rho(a, b)$

Die Entropie von (X, Y)

(auch *gemeinsame Entropie von* X *und* Y genannt) ist

$$\mathbf{H}[X, Y] := \mathbf{H}[(X, Y)] = -\sum_{a,b} \rho(a, b) \log \rho(a, b).$$

Die gemeinsame Verteilung von X,Y kann man zerlegen:

$$\rho(a, b) = \rho_0(a) P(a, b).$$

Diese Zerlegung überträgt sich auf die Entropie:

$$
\begin{aligned}
&\mathbf{H}[X,Y] \\
&= -\sum_{a,b} \rho(a,b)\left(\log \rho_0(a) + \log P(a,b)\right) \\
&= -\sum_a \rho_0(a) \log \rho_0(a) - \sum_a \rho_0(a) \sum_b P(a,b) \log P(a,b) \\
&= \mathbf{H}[X] + \sum_a \rho_0(a) \mathbf{H}[Y|X=a]
\end{aligned}
$$

mit $\mathbf{H}[Y|X=a] := -\sum_b P(a,b) \log P(a,b)$

(die *Entropie von* Y *gegeben das Ereignis* $\{X=a\}$).

$$\mathbf{H}[X,Y] = \mathbf{H}[X] + \sum_a \rho_0(a)\mathbf{H}[Y|X=a]$$
$$= \mathbf{H}[X] + \mathbf{H}[Y|X]$$

mit $\mathbf{H}[Y|X] := \sum_a \mathbf{P}(X=a)\mathbf{H}[Y|X=a]$

die *bedingte Entropie von* Y*, gegeben* X.

$$\boxed{\mathbf{H}[X,Y] = \mathbf{H}[X] + \mathbf{H}[Y|X]}$$

Sind X und Y unabhängig, dann gilt $\mathbf{H}[Y|X] = \mathbf{H}[Y]$, also

$$\boxed{\mathbf{H}[X,Y] = \mathbf{H}[X] + \mathbf{H}[Y]}$$

Allgemein gilt:

$$
\begin{aligned}
&\mathbf{H}[X] + \mathbf{H}[Y] - \mathbf{H}[X,Y] \\
&\quad = -\sum_{a,b} \mathbf{P}(X=a, Y=b) \log \mathbf{P}(X=a) \\
&\qquad\quad - \sum_{a,b} \mathbf{P}(X=a, Y=b) \log \mathbf{P}(Y=b) \\
&\qquad\quad + \sum_{a,b} \mathbf{P}(X=a, Y=b) \log \mathbf{P}(X=a, Y=b) \\
&\quad = \sum_{a,b} \mathbf{P}(X=a, Y=b) \log \frac{\mathbf{P}(X=a, Y=b)}{\mathbf{P}(X=a)\mathbf{P}(Y=b)} \\
&\quad = D((X,Y)\|(\tilde{X},\tilde{Y}))
\end{aligned}
$$

mit $\tilde{X}$ verteilt wie X, $\tilde{Y}$ verteilt wie Y, $\tilde{X},\tilde{Y}$ unabhängig.

$\mathbf{H}[X, Y]$ ist symmetrisch in (X, Y)

(denn die Verteilungsgewichte von (Y, X) sind

$\bar{\rho}(b, a) := \rho(a, b)$).

Also ist auch

$$\mathbf{H}[Y] \; + \mathbf{H}[X] - \mathbf{H}[X, Y]$$

symmetrisch in X und Y.

Dieses ist gleich

$$\mathbf{H}[Y] - \mathbf{H}[Y \mid X] =: \mathbf{I}[X \| Y],$$

die sogenannte *wechselseitige Information* von X und Y.

Wir hatten $\mathbf{I}[X\|Y]$ als relative Entropie

(von (X,Y) bezüglich seiner “Entkopplung”) erkannt.

Die relative Entropie ist nichtnegativ (Vorlesung 13),

also folgt

$$\mathbf{H}[Y \mid X] \le \mathbf{H}[Y]$$

und analog (vgl. Buch (24.11))

$$\mathbf{H}[Y \mid X, Z] \le \mathbf{H}[Y \mid Z].$$

Merke: “Zusätzliches Bedingen verkleinert die Entropie.”

Beispiel: Stationäre Quellen.

$X_1, X_2, \ldots$ heißt *stationär verteilt*,

wenn für alle $m, n \in \mathbb{N}$ die gemeinsamen Verteilungen

von $X_1, \ldots, X_n$ und von $X_{m+1}, \ldots, X_{m+n}$ übereinstimmen.

Wir werden zeigen: $h_Q := \lim_{n\to\infty} \frac{\mathbf{H}_2[X_1, \ldots, X_n]}{n}$ existiert.

$$h_Q := \lim_{n\to\infty} \frac{\mathbf{H}_2[X_1, \ldots, X_n]}{n} \in [0, \mathbf{H}_2(X_1)]$$

Extremfälle:

(i) $X_1, X_2, \ldots$ unabhängig und identisch verteilt:

$$\mathbf{H}_2[X_1, \ldots, X_n] = n\mathbf{H}_2[X_1], \quad h_Q = \mathbf{H}_2[X_1]$$

(ii) $X_1 = X_2 = \ldots$

$$\mathbf{H}_2[X_1, \ldots, X_n] = \mathbf{H}_2[X_1], \quad h_Q = 0$$

$r_Q := 1 - \dfrac{h_Q}{\mathbf{H}_2[X_1]}$ heißt *relative Redundanz* der Quelle.

Behauptung: $h_Q := \lim_{n\to\infty} \frac{\mathbf{H}_2[X_1,\ldots,X_n]}{n}$ existiert.

Beweis:

$$\mathbf{H}[X_1,\ldots,X_n] = \mathbf{H}[X_2,\ldots,X_n] + \mathbf{H}[X_1 \mid X_2,\ldots,X_n] = \cdots$$

$$= \mathbf{H}[X_n] + \mathbf{H}[X_{n-1} \mid X_n] + \cdots + \mathbf{H}[X_1 \mid X_2,\ldots,X_n] \,.$$

Mittels Stationarität folgt

$$\mathbf{H}[X_1,\ldots,X_n] = \mathbf{H}[X_1] + \mathbf{H}[X_1 \mid X_2] + \cdots + \mathbf{H}[X_1 \mid X_2,\ldots,X_n].$$

“Zusätzliches Bedingen verkleinert die Entropie”,
also fallen die Summanden monoton.
Aus der Konvergenz der Summanden folgt
die Konvergenz des arithmetischen Mittels gegen denselben Grenzwert.□

Entropiereduktion unter Abbildungen

$$\mathbf{H}[h(X)] \le \mathbf{H}[X]$$

denn wegen $\mathbf{P}\big(X = a, h(X) = h(a)\big) = \mathbf{P}(X = a)$ ist

$$\mathbf{H}[X] = \mathbf{H}[X, h(X)].$$

Weiter ist

$$\mathbf{H}[X, h(X)] = \mathbf{H}[h(X)] + \mathbf{H}[X \,|\, h(X)] \ge \mathbf{H}[h(X)].$$

Simulation diskreter Verteilungen per Münzwurf

Wir betrachten einen vollen Binärbaum.

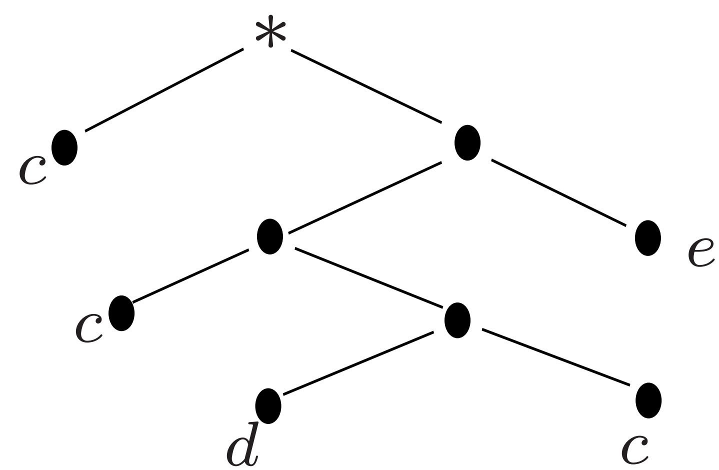
Dieser beschreibt einen fairen Münzwurf, der zufällig abbricht:

Für jedes Blatt b der Tiefe t

endet das Experiment mit Wahrscheinlichkeit 2^{-t} in b.

Die Blätter seien mit Elementen einer abzählbaren Menge S beschriftet.

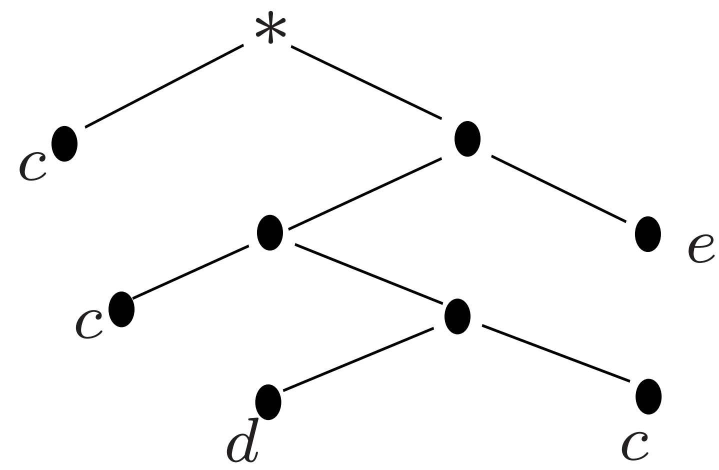
Z. B. für $S = \{c, d, e\}$:



Es sei X das zufällige Blatt,

in dem das Experiment endet,

und $h(X)$ seine Beschriftung.



Dann hat die Verteilung von $Y = h(X)$ die Gewichte

$$\pi(c) = \frac{1}{2} + \frac{1}{8} + \frac{1}{16}, \quad \pi(d) = \frac{1}{16}, \quad \pi(e) = \frac{1}{4}\,.$$

Das Experiment erzeugt per Münzwurf

eine Zufallsvariable Y mit Verteilung π.

Jede Verteilung π auf einer abzählbaren Menge S

lässt sich auf diese Weise simulieren:

Man schreibe die Gewichte von π als Dualbruch,

$$\pi(b) = \sum_j 2^{-\ell(b,j)} ,$$

mit natürlichen Zahlen $1 \leq \ell(b,1) < \ell(b,2) < \cdots$.

Dann gilt

$$\sum_{b,j} 2^{-\ell(b,j)} = \sum_b \pi(b) = 1 .$$

$$\pi(b) = \sum_j 2^{-\ell(b,j)} \ , \quad \sum_{b,j} 2^{-\ell(b,j)} = \sum_b \pi(b) = 1 \ .$$

Nach dem Lemma von Fano-Kraft

gibt es also einen vollen binären Baum,

der für jedes Paar (b, j) ein Blatt der Tiefe $\ell(b, j)$ frei hält.

Versieht man alle diese Blätter mit der Beschriftung b,

dann endet der Münzwurf in einem mit b beschrifteten Blatt

mit der Wahrscheinlichkeit $\sum_j 2^{-\ell(b,j)} = \pi(b)$.

Wir stellen jetzt eine Beziehung her

zwischen der erwarteten Anzahl der

für die Simulation von π benötigten Münzwürfe

und der binären Entropie von π.

Zur Erinnerung:

X ist das zufällige Blatt, in dem das Experiment endet,

und $Y = h(X)$ seine Beschriftung.

Ist $\ell(a)$ die Tiefe des Blattes a, dann gilt:

$$\mathbf{P}(X = a) = 2^{-\ell(a)}$$

Also folgt: $\mathbf{E}[\ell(X)] = \mathbf{H}_2[X]$

Für $Y = h(X)$ gilt

$$\mathbf{H}_2[X] = \mathbf{H}_2[X, Y] = \mathbf{H}_2[Y] + \mathbf{H}_2[X \mid Y]$$

Wir werden zeigen:

$$\mathbf{H}_2[X \mid Y] \leq 2$$

und bekommen so die Abschätzung:

$$\mathbf{H}_2[Y] \leq \mathbf{E}[\ell(X)] \leq \mathbf{H}_2[Y] + 2$$

Behauptung: $\mathbf{H}_2[X \mid Y] \leq 2$

Beweis: Wir hatten die Dualbruchdarstellung

$\pi(b) = \sum_j 2^{-\ell(b,j)}$, mit natürlichen Zahlen $1 \leq \ell(b,1) < \ell(b,2) < \cdots$.

Die Blätter $a_1, a_2, \ldots$, die mit b beschriftet sind, lassen sich also so anordnen, dass $\ell(a_1) < \ell(a_2) < \cdots$ gilt.

Für $p_i := \mathbf{P}(X = a_i \mid Y = b) = 2^{-\ell(a_i)}/\pi(b)$ folgt $p_{i+1} \leq p_i/2$.

Dies impliziert, dass es eine Zahl k gibt, so dass

$p_i > 2^{-i}$ für $i \leq k$ und $p_i \leq 2^{-i}$ für $i > k$. Es folgt

$$\sum_{i \geq 1} i\Big(p_i - 2^{-i}\Big) = \sum_{i \geq 1} (i-k)\Big(p_i - 2^{-i}\Big) \leq 0, \quad \text{und somit}$$

$$\textstyle\sum_{i \geq 1} i p_i \leq \sum_{i \geq 1} i 2^{-i} = 2.$$

Mit Entropieschranke (Vorlesung 13): $\mathbf{H}_2[X \mid Y = b] \leq 2$.

Es folgt $\mathbf{H}_2[X \mid Y] \leq 2 \quad \square$.