

Vorlesung 14a

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:

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

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

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

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

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

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

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

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

die sogenannte *wechselseitige Information* von X und Y .

Wir hatten $I[X||Y]$ als relative Entropie
(von (X, Y) bezüglich seiner “Entkopplung”) erkannt.
Die relative Entropie ist nichtnegativ (Vorlesung 13b),
also folgt

$$H[Y | X] \leq H[Y]$$

und analog (vgl. Buch (24.11))

$$H[Y | X, Z] \leq H[Y | Z].$$

Merke: “Zusätzliches Bedingen verkleinert die Entropie.”

Beispiel: Stationäre Quellen.

X_1, X_2, \dots heißt *stationär verteilt*,

wenn für alle $m, n \in \mathbb{N}$ die gemeinsamen Verteilungen von X_1, \dots, X_n und von X_{m+1}, \dots, X_{m+n} übereinstimmen.

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

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

Extremfälle:

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

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

(ii) $X_1 = X_2 = \dots$

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

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

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

Beweis:

$$\begin{aligned} \mathbf{H}[X_1, \dots, X_n] &= \mathbf{H}[X_2, \dots, X_n] + \mathbf{H}[X_1 | X_2, \dots, X_n] = \dots \\ &= \mathbf{H}[X_n] + \mathbf{H}[X_{n-1} | X_n] + \dots + \mathbf{H}[X_1 | X_2, \dots, X_n] . \end{aligned}$$

Mittels Stationarität folgt

$$\mathbf{H}[X_1, \dots, X_n] = \mathbf{H}[X_1] + \mathbf{H}[X_1 | X_2] + \dots + \mathbf{H}[X_1 | X_2, \dots, X_n].$$

“Zusätzliches Bedingen verkleinert die Entropie”,
also fallen die Summanden monoton.

Aus der Konvergenz der Summanden folgt
die Konvergenz des arithmetischen Mittels gegen denselben Grenzwert. \square

Entropiereduktion unter Abbildungen

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

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

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

Weiter ist

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

Simulation diskreter Verteilungen per Münzwurf

Wir betrachten einen vollen Binärbaum.

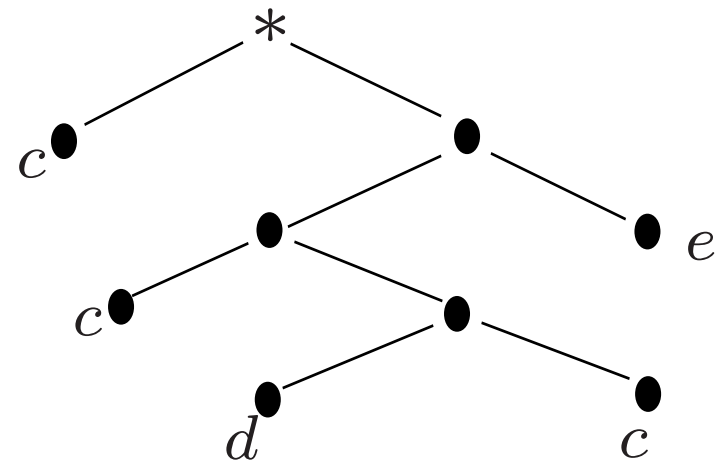
Dieser beschreibt einen fairen Münzwurf, der zufällig abbricht:

Für jedes Blatt b der Tiefe t

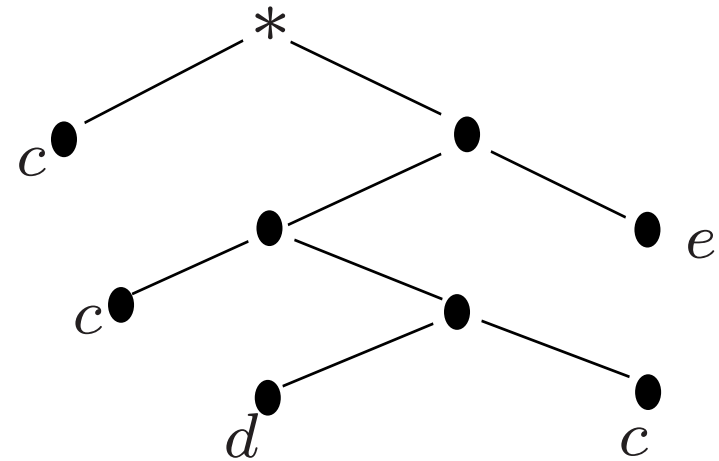
endet das Experiment mit Wahrscheinlichkeit 2^{-t} in b .

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 $1 \leq \ell(b, 1) < \ell(b, 2) < \dots$.

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 der Ungleichung 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

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

Wir werden zeigen:

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

und bekommen so die Abschätzung:

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

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

Beweis: Wir hatten die Dualbruchdarstellung

$$\pi(b) = \sum_j 2^{-\ell(b,j)}, \text{ mit nat\u00fcrlichen Zahlen } 1 \leq \ell(b, 1) < \ell(b, 2) < \dots.$$

Die Bl\u00e4tter a_1, a_2, \dots , die mit b beschriftet sind, lassen sich also so anordnen, dass $\ell(a_1) < \ell(a_2) < \dots$ gilt.

F\u00fcr $p_i := \mathbf{P}(X = a_i | Y = b) = 2^{-\ell(a_i)} / \pi(b)$ folgt $p_{i+1} \leq p_i / 2$.

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

$p_i > 2^{-i}$ f\u00fcr $i \leq k$ und $p_i \leq 2^{-i}$ f\u00fcr $i > k$. Es folgt

$$\sum_{i \geq 1} i(p_i - 2^{-i}) = \sum_{i \geq 1} (i - k)(p_i - 2^{-i}) \leq 0,$$
$$\sum_{i \geq 1} ip_i \leq \sum_{i \geq 1} i2^{-i} = 2.$$

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

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