Information und Entropie

- Wie misst man Information?
- Wie viel Information steckt in einer Nachricht?

$\Rightarrow$ Quantifizierung durch Informationstheorie (SHANNON)

Entropie

Die fundamentale Größe ist die Entropie einer ZV, ein Maß für die Unsicherheit oder den Informationsgehalt der ZV.
Bsp.:

$$F \in_R \{0,1\} \textit{ faire Münze}$$

$$W \in_R \{1,2,\ldots,6\} \textit{ Würfel}$$

$$Z \in \{0,a,b,\ldots,h\} \qquad P_Z(0) = \frac{1}{2} \qquad P_Z(a) = P_Z(b) = \ldots = P_Z(h) = \frac{1}{16}$$

$$G \sim geom(p) = geom(\frac{1}{3}) \qquad P[G=n] = (1-p)^{n-1} \cdot p \quad n > 0$$

Versuch für ein Informationsmaß: Supportmenge der ZV X $\{x \in X \mid P_X(x) > 0\}$
$\rightarrow \underbrace{\textit{Kardinalität der Supportmenge}}_{\textit{Logarithmus von}}$

! Alle Logarithmen zur Basis 2

$$K(F) = 1$$

$$K(W) = log(6) \approx 2.58$$

$$K(Z) = log(9) \approx 3.12$$

$$K(G) = \infty$$

Idee: Info.gehalt von X=x sei $\log(\frac{1}{P_X(x)})$

Definition der Entropie
Def.: $H(X) = -\sum P_X(x) \cdot log(P_X(x))$
(Annahme: $\forall x \in X : P_X(x) \geq 0$)
Da der Logarithmus zur Basis 2 ist, so misst die Entropie "Bits"
Alternative Definition: $H(X) = E[-log(P_X(X))]$

Entropie der binären Verteilung

$$\left.\begin{smallmatrix} P[X=0]=p \\ P[X=1]=1-p \end{smallmatrix}\right\} \quad h(p) = . - p \cdot log(p) - (1-p) \cdot log(1-p)$$

$$h(0.5) = 1$$

$$h(0.11) = h(0.89) = 0.5$$

$$h(0) = h(1) = 0$$

Theorem:

$$0 \leq H(X) \leq log \mid X \mid$$

$$mit\ 0 = H(X) \Leftrightarrow \exists x : P_X(x) = 1$$

$$mit\ H(X) = log \mid X \mid \Leftrightarrow P_X(x) = \frac{1}{\mid X \mid} \qquad \forall x$$

<u>Beweis:</u>

$$1.) \qquad 0 \geq H(X)$$

$$Da\ P_X(x) > 0\ f\ddot{u}r\ x \in X:$$

$$P_X(x) \cdot log(P_X(x) = 0 \Leftrightarrow \exists x : P_X(x) = 1$$

$$P_X(x) \cdot log(P_X(x) > 0 \Leftarrow \{^{P_X(x)>0}_{P_X(x)<1}$$

$$2.) \qquad Jensen:\ f(\cdot)\ konkav\ (\cap)$$

$$\Rightarrow E[f(x)] \leq f[E(X)]$$

$$H(X) = E[-log(P_X(X))] = E[log(\frac{1}{P_X(X)})] \leq log(E[\frac{1}{P_X(X)}])$$

$$= log(\sum_{x \in X} P_X(x) \cdot \frac{1}{P_X(x)}) = log \mid X \mid$$

Entropiewerte der Beispiele:

$$H(F) = 1$$

$$H(W) \approx 2.58$$

$$H(Z) = \frac{1}{2} \cdot log(\frac{1}{2}) + 8 \cdot \frac{1}{16} \cdot log(\frac{1}{16}) = 2.5$$

$$H(G) = \frac{h(p)}{p} \approx 2.75$$

<u>**Interpretationen von H(X)**</u>

- Anzahl nötiger bits im Durchschnitt, um einen Wert von X zu codieren
- Anzahl nötiger binärer Fragen (im Durchschnitt), um das Resultat X zu erfahren
- Anzahl der perfekten Zufallsbits, welche aus X extrahiert werden können (im Durchschnitt)

<u>**Bedingte Entropie**</u>

<u>Def.:</u> Bedingte Entropie von X gegeben Ereignis Y=y:

$$H(X \mid Y = y) = -\sum_{x \in X} P_{X|Y=y}(x) \cdot log(P_{X|Y=y}(x)) \qquad \text{"Entropie von } \underline{X|\,Y{=}y}\text{"}$$

<u>Def.:</u> Bedingte Entropie von X gegeben eine ZV Y:

$$H(X \mid Y) = \sum_{y \in Y} P_Y(y) \cdot H(X \mid Y = y)$$

<u>Alternative Definition:</u>

$$H(X \mid Y) = E[-log(P_{X|Y}(X))]$$

$$H(X \mid Y) = -\sum_{x \in X} \sum_{y \in Y} P_{X|Y}(x, y) \cdot log(P_{X|Y=y}(x))$$

	s	w	r
t	0	$\frac{1}{12}$	$\frac{2}{12}$
m	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{12}$
h	$\frac{2}{12}$	0	$\frac{1}{12}$

⇒ Verbleibende Unsicherheit über X, wenn Y bekannt ist.

Bsp. 1: X $\in_R$ {000,001,. . . , 111}

$$X = X_1X_2X_3$$

$$Y = X_1 \oplus X_2 \oplus X_3$$

$$H(X) = 3 \qquad H(Y) = 1$$

$$H(X \mid Y) = 2$$

Bsp. 2: Barometer B∈ {tief, mittel, hoch}

Wetter morgen W∈{sonnig, wechselhaft, regen}

$$H[BW] \approx 1.675$$

$$H[B] = H[\frac{1}{4}, \frac{1}{2}, \frac{1}{4}] = 1.5$$

$$H[W] \approx 1.554$$

$$H[W \mid B = t] = h(\frac{1}{3}) \approx 0.918$$

$$H[W \mid B = m] = log(3) \approx 1.585 \textit{ Konditionierung auf Ereignis kann Unsicherheit erhöhen!!!}$$

$$H[W \mid B = h] = h(\frac{1}{3}) \approx 0.918$$

$$H[W \mid B] = \frac{1}{4} \cdot H[W \mid B = t] + \frac{1}{2} \cdot H[W \mid B = m] + \frac{1}{4} \cdot H[W \mid B = h] \approx 1.252$$

$$H[W] \geq H[W \mid B]$$

Theorem: Für ZV X und Y gilt:

$$H[X \mid Y] \leq H[X] \qquad \textit{mit Gleichheit gdw. X unabhängig von Y}$$

"Zusatzinformation verringert die Unsicherheit"

Theorem: "Kettenregel für Entropie"

$$H[XY] = H[X] + H[Y \mid X]$$

$$H[XY] = H[Y] + H[X \mid Y]$$

Wichtige Spezialfälle:

- X unabhängig von Y ⇔ H[XY] = H[X] + H[Y]
- Y = f(X) ⇔ H[Y|X] = 0 ⇔ H[XY] = H[X]