Fortsetzung Informationsmaße:

Beispiel: $\{A_1, A_2, A_3, A_4 \dots A_8\}$

x

0 x

Nachricht: 1 0 1 $\Leftrightarrow$ A_3 ist eingetreten

Länge der Nachricht $\quad n = \log_2 N \quad$ (Bitzahl)

Def: Informationsmaß der Nachricht $b(N) = \log_2 N$, falls keine Kenntnis vorhanden ist.

Verallgemeinerung auf Wahrscheinlichkeitsv. P_i

Falls der Beobachter die P_i kennt, muss ihm nur die fehlende Information mitgeteilt werden: $\quad$ Bitzahl $b(P_i)$

Postulate für die Konstruktion von $b(P_i)$

(i) $b(P)$ ist universelle Fkt; d.h. hängt von A nur über $P(A)$ ab.

(ii) Seien $\{A_i\}$ und $\{A_j'\}$ 2 verschiedene (disjunkte) Sample Sets

z.B. 2 Subsysteme eines zusammengesetzten Systems,

Für 2 unkorrelierte Subsysteme ist b additiv;

$$b(P'') = b(P) + b(P')$$

wobei für unkorrelierte Ereignisse gilt $P''(A_i A_j') = P(A_i) P'(A_j')$

(iii) $b(P) = 0 \quad$ für $P = 1 \quad$ d.h. das sichere Ereignis

$b(P) = \log_2 N$ für $P = 1/N$ d.h. bei Gleichverteilung

(iv) $b(P)$ ist stetig und wohldefiniert für $0 \le P \le 1$

Definiere $b(P) = f(\log P)$ mit noch zu bestimmender Funktion f

aus (i) und (ii) folgt:

$$f(\log P'') = f(\log P + \log P') \overset{!}{=} f(\log P) + f(\log P')$$

$$\Rightarrow f(\log P) = k \log P \quad \text{(linear in } \log P\text{)}$$

aus (iii) folgt: $b(P) = k \log P = -k \log N \overset{!}{=} \log_2 N$ für $P = \frac{1}{N}$

$\Rightarrow k = -1$

$\log = \log_2$

Konvention: Einheit für 1 bit: $\ln 2 = \frac{\ln P}{\log_2 P}$ ("bin")

$$\boxed{b(P_i) = -\ln P_i}$$

Informationsmaß für die Nachricht, dass A_i eingetreten ist, falls $P_i \equiv P(A_i)$ bekannt ist.

Informationsmaß eines Wahrscheinlichkeitsv. $\{P_i\}$

Übermittlung vieler Nachrichten:

A_i tritt mit relativer Häufigkeit P_i auf mittlere benötigte (d.h. fehlende) Information pro Ereignis

$$\langle b \rangle = -\sum_{i=1}^{N} P_i \ln P_i$$

<u>Def.</u>: <u>Shannon-Information</u> einer Verteilung $\{P_i\}$

$$\boxed{I(P) = \sum_{i=1}^{N} P_i \ln P_i}$$

$P = (P_1 \dots P_N)$

I ist Funktional der Verteilung

b ist Funktion von P_i

Es gilt stets $I(P) \leq 0$

Maximum: $I(P) = 0$ für $P_i = \delta_{ij}$ (scharfe Verteilung mit sicherem Ereignis A_j)

Minimum: Variation der P_i um δP_i unter der Nebenbedingung

$$\sum_i \delta P_i = 0$$

(wegen Normierung $\sum_{i=1}^{N} P_i = 1$)

$$\delta I = \sum_i (\ln P_i + 1)\, \delta P_i = 0$$

Addition der Nebenbedingung $\sum_i \delta P_i = 0$ mit Lagrange-Multiplikator λ.

$$\sum_i (\ln P_i + 1 + \lambda)\, \delta P_i = 0$$

unabhängige Variation δP_i $\Rightarrow \ln P_i = -(1+\lambda) = \text{const.}$

Normierung $\sum_{i=1}^{N} P_i = N P_i \stackrel{!}{=} 1 \quad \Rightarrow \boxed{P_i = \frac{1}{N}}$ Gleichverteilung

<u>Kontinuierliche Ereignismenge ($x \in \mathbb{R}^d$; $\rho(x)$)</u>

Zelleneinteilung des $\mathbb{R}^d$ in Zellen i mit $\Delta^d x$

Wahrscheinlichkeit für Ereignis in Zelle i : $P_i = \rho(x^i)\Delta^d x$

$$I(P) = \sum_i \Delta^d x\, \rho(x^i) \ln \rho(x^i) + \underbrace{\underbrace{\sum_i \Delta^d x\, \rho(x^i)}_{1} \ln \Delta^d x}_{\text{weglassen, da const. für feste Zellengröße}}$$

$\Delta^d x \to 0$ $\boxed{I(\rho) = \int d^d x\, \rho \ln \rho}$

<u>Bemerkungen</u>: (i) Shannon – Informationsmaß mißt die Kenntnis bzgl. der speziellen Frage: "Welches Ereignis tritt ein?"

- Keine Unterscheidung, wie die Verteilung zustande kommt, z.B. Gleichverteilung: genaue Beobachtung <u>oder</u> vorurteilsfreie Schätzung bei gänzlich fehlender Kenntnis

(ii) Def.: Statistisches Informationsmaß des <u>Nichtwissens</u> (fehlende Info)

$$S(\rho) = -k \int d^d x\, \rho \ln \rho$$ (k geeignete Einheit)

(Interpretation in der Thermodynamik als Entropie – s. später)

(iii) verallgemeinertes Informationsmaß (Renyi):

$$I_q = -\frac{1}{1-q} \ln\left(\sum_i p_i^q\right) \quad q = 1, 2, \dots \quad (q \to 1 \Rightarrow \text{Shannon}).$$

<u>Informationsgewinn</u>

Maß für die Zusatzinformation einer W.verteilung $\{p_i\}$ im Vergleich zu einer Referenzverteilung $\{p_i'\}$ über derselben Ereignismenge

$$b(p_i') - b(p_i) = \ln \frac{p_i}{p_i'}$$ notwendige Bitzahl um p_i' in p_i zu verwandeln durch eine Nachricht

mittlere Bitzahl (mit der korrigierten W.vert. gewichtet)

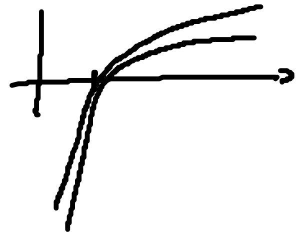
$$K(p, p') = \sum_{i=1}^{N} p_i \ln \frac{p_i}{p_i'}$$ <u>Informationsgewinn</u> (Kullback-Information)

<u>Bem.</u>: (i) Asymmetrisch bzgl. $p \longleftrightarrow p'$!

(ii) Es gilt $K(p, p') \geq 0$, da

$$\sum_i p_i \ln \frac{p_i}{p_i'} \geq \sum_i p_i \left(1 - \frac{p_i'}{p_i}\right) = \underbrace{\sum_i p_i}_{1} - \underbrace{\sum_i p_i'}_{1} = 0$$

$$\ln x \geq 1 - \frac{1}{x} \quad \text{für } x > 0$$



(iii) $p_i' = 0$ ist auszuschließen damit $K < \infty$

(iv) Für $p_i' = \frac{1}{N}$ (Gleichverteilung):

$$K(p, p') = \underbrace{\sum_{i=1}^{N} p_i \ln p_i}_{I(p)} + \underbrace{\sum_{i=1}^{N} p_i}_{1} \ln N = I(p) - I(\tfrac{1}{N})$$

(v) <u>Minimum</u> von K: Variation der p_i um δp_i unter Nebenbedingung $\sum_i \delta p_i = 0$

$$\delta K = \sum_i \left(\ln \frac{p_i}{p_i'} + 1 \right) \delta p_i$$

$$\sum_i \left(\ln \frac{p_i}{p_i'} + 1 + \lambda \right) \delta p_i = 0 \quad \Rightarrow \ln \frac{p_i}{p_i'} = -(1+\lambda) = \text{const.}$$

$$\Rightarrow p_i \sim p_i'$$

Normierung $\Rightarrow p_i = p_i' \Rightarrow K = 0$

(vi) $K(p, p')$ ist konvexe Fkt. von p, da

$$\frac{\partial^2 K}{\partial p_i \partial p_j} = \frac{\partial}{\partial p_j}\left(\ln \frac{p_i}{p_i'} + 1\right) = \frac{1}{p_i}\delta_{ij} \geq 0$$

somit ist auch $I(p) = K(p, \frac{1}{N}) - \ln N$ konvex!

Kontinuierliche Ereignismenge $(x \in \mathbb{R}^d, \rho(x))$

$$p_i = \rho(x^i)\, \Delta^d x$$

$$K(p, p') = \sum_i \Delta^d x\, \rho(x^i) \ln \frac{\rho(x_i)}{\rho'(x_i)}$$

invariant gegen Trafo $x \to \tilde{x}$

$\rho(x) = \tilde{\rho}(x)\, \mathrm{Det}\left(\frac{\partial x}{\partial \tilde{x}}\right)$

$\Delta^d x \to \Delta^d \tilde{x} \cdot \left[\mathrm{Det}\left(\frac{\partial x}{\partial \tilde{x}}\right)\right]^{-1}$

$$\Delta^d x \to 0 : \quad \boxed{K(\rho, \rho') = \int d^d x\, \rho \ln \frac{\rho}{\rho'}}$$

$I(p)$ ist nicht invariant

<u>Bem.</u>: Interpretation von $-k\dot{K}(\rho, \rho')$ in der Thermodynamik als Entropieproduktion, von $kT\, K(\rho, \rho')$ als Exergie (availability).