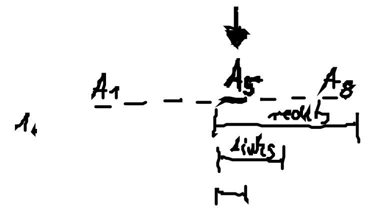
1.2. Informationsmaße

Ziel: Vergleich verschiedener Wahrscheinlichkeitsverteilungen bezüglich ihres Informationsgehaltes.

<u>Frage</u>: Wie lang muss eine Nachricht sein <u>um</u> ein eingetretenes Ergebnis zu übermitteln?

Bsp. ① $\mathcal{A} = \{A_1, A_2\}$ $\to$ ein Bit reicht (rechts / links)
                                                        1   0

② $\mathcal{A} = \{A_1, A_2 \ldots A_N\}$ $\quad N = 2^n$



$\to n$ Alternativentscheidungen

$$\boxed{b(N) = \log_2 N}$$

$\Rightarrow 100$

Nachricht um "A_5" zu übermitteln.

$\triangleright$ Informationsmaß der Nachricht falls keine Vorkenntnis vorhanden ist: $b(N) = \log_2 N$

<u>Verallgemeinerung auf W.-vert. P_i</u>

Falls der Beobachter die P_i kennt $\Rightarrow$ muss <u>nur</u> noch fehlende Information übertragen werden.

Bitzahl $b(P_i)$

<u>Postulate</u> für die Konstruktion von $b(P_i)$

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

(ii) für 2 <u>unkorrelierte</u> sample sets $\{A_i\}$, $\{A_i'\}$ ist die fehlende Info des Gesamtsystems:

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

wobei gilt $p''(A_i A_j') = P(A_i) \cdot P'(A_j')$.

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

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

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

Definiere $b(P) = f(\log P)$ f noch zu bestimmen!

(i) + (ii) $\Rightarrow$ $f(\log P'') = f(\log P + \log P')$

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

$\rightarrow$ linear in $\log P$

$f(\log P) = K \cdot \log P$ K unbekannt.

Aus (iii) folgt: $b(P) = K \cdot \log P = -K \log N \overset{!}{=} \log_2 N$

für $P = \frac{1}{N}$

$\Rightarrow$ $K = -1$
$\log = \log_2$ $\rightarrow$ Konvention für Einheit eines Bits

$\ln 2 = \frac{\ln P}{\log_2 P}$

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

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

Informationsmaß einer W.-vert. $\{P_i\}$: Mittelung über verschiedene Nachrichtenlängen

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

mittlere benötigte Information (oder fehlend) pro Ereignis

Def.: Shannon-Information einer Verteilung $\{P_i\}$

$$I(p) = \sum_{i=1}^{N} p_i \ln p_i \qquad p = (p_1, \dots, p_N)$$

I ist ein Funktional der Vert.f.

b ist Funktion p_i

$I(p) \leq 0$

<u>Maximum</u>: $I(p) = 0$ für $p_i = \delta_{ij}$ scharfe Verteilung

<u>Minimum</u>: Variation der p_i um δp_i unter der Nebenbedingung $\sum_i \delta p_i = 0$

(wegen Normierung $\sum_i p_i = 1$)

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

Addition der NB $\sum_i \delta p_i = 0$ mit Langrange Multiplikator λ

$$\sum_i \underline{(\ln p_i + 1 + \lambda)}\, \delta p_i = 0$$

- unabhängige Variation der δp_i $\longrightarrow$ $\ln p_i = -(1+\lambda) = \text{const.}$ ⊛
- Normierung $\sum_i p_i \overset{⊛}{=} N p_i = 1$ $\quad \boxed{p_i = \frac{1}{N}}$

Gleichverteilung

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

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

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

$$\longrightarrow I(p) = \sum_i \Delta^d x\, \rho(x^i) \ln \rho(x^i) + \underbrace{\sum_i \Delta^d x\, \rho(x^i) \ln \Delta^d x}_{(\underbrace{\sum_i \rho(x^i) \cdot \Delta^d x}_{=1}) \ln \Delta^d x}$$

weglassen da const. für feste Zellengröße

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

Informationsgehalt einer kont. Verteilung

Bemerkung: (i) nur Kenntnis bzgl. der Frage: "Welches Ereignis tritt ein?", keine Unterscheidung woher ich ρ kenne (z.B. geraten oder durch Messung bestätigt)

(ii) Später: Statistisches Informationsmaß des Nichtwissens

k... Boltzmann Konstante

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

Interpretation als Entropie in der Thermodynamik.

(iii) Verallgemeinerte Informationsmaße (Reyi)

$$I_q = -\frac{1}{1-q} \ln\Big(\sum_i p_i^q\Big) \qquad q = 1, 2, \dots$$

$q \to 1$ ergibt Shannon Info

Informationsgewinn:

Maß für die Zusatzinformation einer Wr-vert. $\{p_i\}$ im Vergleich zu einer Referenzverteilung $\{p_i'\}$ über der gleichen 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

Mittelung über p_i:

$$\boxed{K(p, p') = \sum_{i=1}^{N} p_i \ln \frac{p_i}{p_i'}}$$

Informationsgewinn

Kullback-Leibler Information

→ Welche Information geht verloren wenn $\rho'(x)$ verwendet wird um $\rho(x)$ anzunähern: $K(\rho, \rho')$

→ (Wieviel Platz wird verschwendet wenn mit p_i' codiert wird was p_i verteilt ist.)

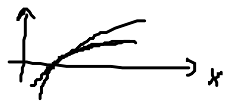
<u>Bemerkungen:</u>

(i) Asymmetrisch bzgl. $p \longleftrightarrow p'$

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

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

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



)

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

(iv) Für $p_i' = \frac{1}{N}$

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

(v) Minimum von K $\overset{?}{\text{-}}$

Variation unter Nebenbedingungen ....

$\rightarrow \boxed{p_i = p_i'}$

$\rightarrow K = 0$