1.2. Informationsmaße

Ziel: Vergleich verschiedener Wahrscheinlichkeitsverteilungen bezüglich ihres Informationsgehaltes.

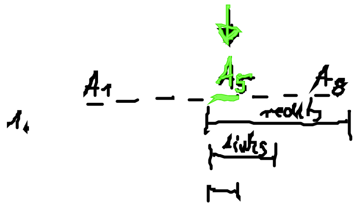
Frage: Wie lang muss eine Nachricht sein um ein eingetretenes Ergebnis zu übermitteln?

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

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

—> n Alternativentscheidungen

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



=> 100

Nachricht um "A_5" zu übermitteln.

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

Verallgemeinerung auf W.-vert. P_i

Falls der Beobachter die P_i kennt => muss nur noch fehlende Information übertragen werden.

Bitzahl $b(P_i)$

Postulate für die Konstruktion von $b(P_i)$

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

(ii) für 2 unkorrelierte 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 $0 \leq p \leq 1$

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 $\{P_i\}$

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

$P = (P_1, \dots, P_N)$

$\underline{I}$ 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 Lagrange Multiplikator λ

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

- unabhängige Variation der δP_i $\longrightarrow$ $\ln P_i = -(1+\lambda) = \text{const.}$
- Normierung $\sum_i P_i \overset{⊛}{=} N P_i = 1$

$$\boxed{P_i = \frac{1}{N}}$$

Gleichverteilung

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

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

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

$$\longrightarrow I(P) \approx \sum_i \Delta^d x\, \varrho(x^i) \ln \varrho(x^i) + \underbrace{\sum_i \Delta^d x\, \varrho(x^i) \ln \Delta^d x}_{(\underbrace{\sum_i \varrho(x^i) \cdot \Delta^d x}_{\overset{\shortparallel}{1}}) \ln \Delta^d x}$$

weglassen da
const. für feste
Zellengröße

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

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)

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

$q \to 1$ 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 mi p_i' codiert wird was p_i verteilt ist.)

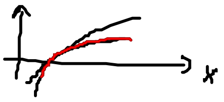
Bemerkungen:

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

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

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

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



$\Big)$

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

(iv) Für $p_i' = \frac{1}{N}$

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

(v) Minimum von K ?

Variation unter Nebenbedingungen ....

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

$\longrightarrow K = 0$