1.2) Informationsmaße

Die Informationstheorie (Shannon, Wiener ↗ Kybernetik) entstand im 2. Weltkrieg im Zusammenhang mit der Entschlüsselung codierter Nachrichten

Def. Ein Maß μ auf einer σ-Algebra $\mathcal{A}$ ist eine Abb. $\mu: \mathcal{A} \to [0, \infty]$ mit den Eigenschaften

$$\mu(\emptyset) = 0$$

Für disjunkte Ereignisse A_i (d.h. $A_i \cap A_j = \emptyset$)

$$\mu\left(\bigcup_i A_i\right) = \sum_i^\infty \mu(A_i)$$

Im folgenden sei die Ereignisalgebra stets eine σ-Algebra

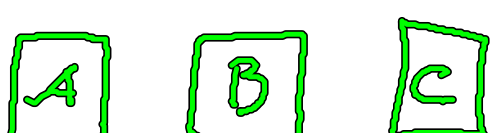
Bsp: Wahrscheinlichkeit P
(speziell $P(A) \leq 1$

Idee der Informationsmaße:

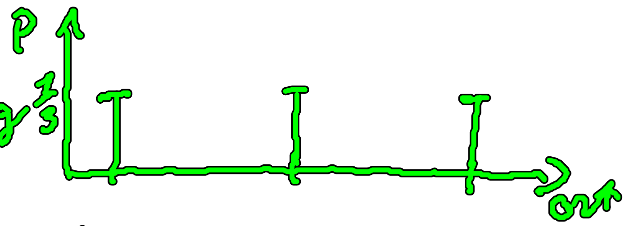
Vergleich verschiedener Wahrscheinlichkeitsverteilungen einer Ereignisalgebra $\mathcal{A}$

Frage: Welche von zwei Verteilungen enthält mehr Information (Kenntnis) darüber, welches Ereignis eintreffen wird?

Beispiel: Hauptgewinn ist hinter einer von 3 Türen versteckt.

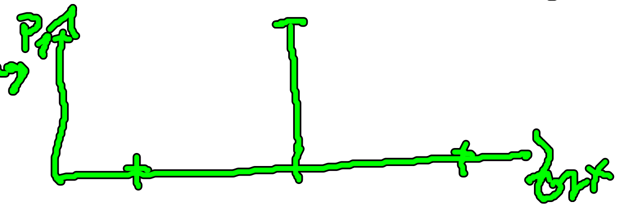


1. Verteilung



Gleichverteilung
(minimale Kenntnis)

2. Verteilung



Scharfe Verteilung
(maximale Kenntnis)
= Sicherheit

Bitzahl:

Ausgangspunkt: diskrete Ereignisalgebra $\mathcal{A} = \{A_i\}$

Frage: Wie lang muß eine Nachricht sein, die einem Beobachter mitteilt, daß *ein* Ereignis eingetreten ist?

Beispiel: Auswahl eines Ereignisses aus $\mathcal{A} = \{A_1, A_2, \dots, A_N\}$, falls der Beobachter keine Vorkenntnis hat.

(i) $\mathcal{A} = \{A_1, A_2\}$: einfache Alternative

= kleinste Informationseinheit

= 1 bit (binary digit) 1942

Nachricht: 0 oder 1

(ii) $\mathcal{A}$ = Menge mit $N = 2^n$ Elementen

n Alternativentscheidungen $\quad N = 2^n \mid \log_2$

z.B. $\underbrace{1100101}_{n \text{ Stellen}}$ $\quad n = \log_2(N)$

Länge der Nachricht: $\boxed{n = \log_2(N)}$ (Bitzahl)

Def: Informationsmaß der Nachricht, falls keine Vorkenntnis vorhanden ist:

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

Verallgemeinerung auf Wahrscheinlichkeitsverteilungen P_i

Falls der Beobachter die P_i kennt, muß ihm nur die fehlende Information mitgeteilt werden $b(P_i)$

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 $\{A_j'\}$ zwei verschiedene, disjunkte sample sets (z.B. 2 Subsys. eines zusammengesetzten Systems)

Für 2 unkorrelierte Subsysteme ist b additiv:

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

wobei nach Def. der Unkorreliertheit (stochastisch unabh.) gilt

$$P''(A_i A_j') = P(A_i)\, P'(A_j') \qquad P_i = P(A_i)$$

(iii) $\underset{\text{sicheres Ereignis}}{P = 1}$ $\quad b(P) = 0$

$$\underset{\text{Maximale Unbestimmtheit}}{b(P) = \log_2 N} \quad \text{für} \quad P = \frac{1}{N} \quad (\text{Gleichverteilung})$$

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

<u>Definiere</u> $b(P) = f(\log P)$ mit noch zu bestimmender Funktion f

Benutze (i) & (ii)

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

$\Rightarrow$ lineare Fkt f

$$\Rightarrow \underline{f(\log P) = k \cdot \log P}$$

aus (iii) folgt für $P = \frac{1}{N}$

$$b(P) = b\left(\frac{1}{N}\right) = k \cdot \log \frac{1}{N} = \underline{-k \cdot \log N \overset{!}{=} \log_2(N)}$$

$\Rightarrow$ $k = -1$

$\log = \log_2$

<u>Konvention</u> Einheit für 1 bit : $\mathrm{lu} = \frac{\ln P}{\log_2 P}$ (bin)

$$\Rightarrow \boxed{b(P_i) = -\mathrm{lu}(P_i)}$$

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

Informationsmaß einer Wahrscheinlichkeitsverteilung $\{P_i\}$

Übermittlung vieler Nachrichten:

A_i tritt mit relativer Häufigkeit P_i auf

mittlere benötigte (da fehlende) Information pro Ereignis:

$$\langle b \rangle = -\sum_i^N P_i \ln P_1$$

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

$$\boxed{I(P) = \sum_i P_i \ln P_i}$$

!

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

I ist Funktional der Verteilung

b ist eine Funktion von P_i

(Ü) Es gilt stets $I(P) \leq 0$

(Ü) $I(P)$ für verschiedene Verteilungen

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

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

Wahrscheinlichkeit für ein Ereignis in der Zelle i:

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

$$I(P) = \sum_i \Delta^d x\, \rho(x^i) \ln\left[\Delta^d x\, \rho(x^i)\right]$$

$$= \sum_i \Delta^d x\, \rho(x^i)\left[\ln \rho(x^i) + \ln \Delta^d x\right]$$

$$= \sum \Delta^d x\, \rho(x^i) \ln \rho(x^i) + \underbrace{\underbrace{\sum \Delta^d x\, \rho(x^i)}_{1} \ln \Delta^d x}_{\text{const}}$$

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

(i) Shannon-Informationsmaß mißt die Kenntnis bezgl. der speziellen Frage: „Welches Ereignis tritt ein?"

Keine Unterscheidung, wie die Verteilung zustande kommt

z.B. bei Gleichverteilung

→ genaue Beobachtung

→ vorurteilsfreie Schätzung

Def: <u>statistische Informationsmaß des Nichtwissens</u>
(fehlende Information)

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

k - geeignete Einheit

→ spätere Interpretation in der Thermodynamik als <u>Entropie</u>