Informationsgehalt von Zufallssignalen

Diskrete gedächtnislose Informationsquelle

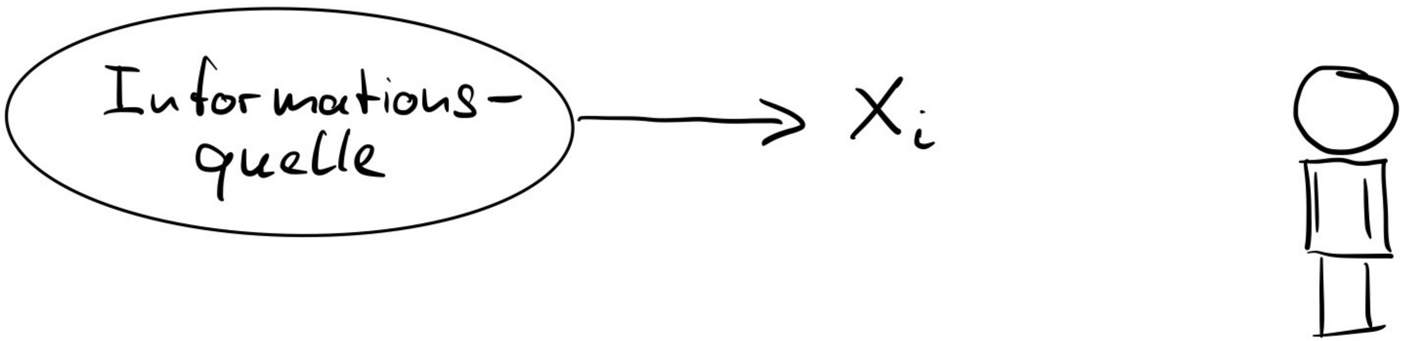
Gegeben sei eine Informationsquelle, die zu diskreten Zeitpunkten $i \in \mathbb{Z}$ Symbole X_i abgibt.

Jedes Symbol stammt aus Symbolvorrat mit M_x unterschiedlichen Symbolen

$$X_i \in \Xi = \{x_1,\, x_2,\, \dots,\, x_{M_x}\}$$

Die a-priori Wahrscheinlichkeiten der einzelnen Symbole sind bekannt und ändern sich nicht über die Zeit

$$P(X_i = x_k) = p_k \qquad k = 1 \dots M_x$$



Shannon'sche Informationsmaß I

Definition des Informationsgewinns $I_\mathrm{S}(x_k)$ durch Beobachtung eines Symbols x_k

Anforderungen an das Informationsmaß:

1. $I_\mathrm{S}(x_k) \in \mathbb{R}$ und $I_\mathrm{S}(x_k) \geq 0$
2. $I_\mathrm{S}(x_k) = f(P(X_i = x_k))$
3. Für Symbolsequenzen $(x_{k_1}, x_{k_2}, x_{k_3}, \ldots, x_{k_N})$ aus einer gedächtnislosen Quelle muss gelten

$$I_\mathrm{S}(x_{k_1}, x_{k_2}, x_{k_3}, \ldots, x_{k_N}) = \sum_{n=1}^{N} I_\mathrm{S}(x_{k_n})$$

Aus 2. und 3. folgt sofort

$$f\left(\prod_{n=1}^{N} P(X_n = x_{i_n})\right) = \sum_{n=1}^{N} f\left(P(X_n = x_{i_n})\right)$$

Shannon'sche Informationsmaß II

Nur die Logarithmusfunktion erfüllt diese drei Bedingungen

$$I_{\mathrm{S}}(x_k) = -\log_2(P(X_i = x_k)) = -\mathrm{ld}(P(X_i = x_k)) = -\mathrm{ld}(p_k) \qquad [I_{\mathrm{S}}(x_k)] = \mathrm{bit}$$

Basis des Logarithmus kann beliebig gewählt werden

Gebräuchliche Wahl ist die Basis 2, womit sich die Einheit *bit* (binary digit) ergibt

Die Einheit des Informationsmaßes ist von der Bezeichnung *Bit* eines *binären Symbols* zu unterscheiden:

- für *Bit* sind nur natürliche Zahlen sinnvoll (z.B. 16 Bit Mikroprozessor)
- Informationsgewinn kann auch eine positive reelle Zahl sein kann (z.B. $1.23\mathrm{bit}$)

Shannon'sche Informationsmaß entspricht logarithmischer Beschreibung von Wahrscheinlichkeiten

Informationsgehalt einzelner Symbole uninteressant

Interessant: Mittlerer Informationsgehalt der Quelle

Mittlerer Informationsgehalt - Entropie

$$H(X) = \mathrm{E}\{I_\mathrm{S}(x_k)\} = -\sum_{k=1}^{M_x} p_k \cdot \mathrm{ld}(p_k) \qquad [H(X)] = \frac{\mathrm{bit}}{\mathrm{Symbol}}$$

Beispiel 1: Binäre Quelle ($M_x = 2$) mit $P(X = 0) = P(X = 1) = 1/2$

$$H(X) = -\frac{1}{2}\mathrm{ld}\left(\frac{1}{2}\right) - \frac{1}{2}\mathrm{ld}\left(\frac{1}{2}\right) = \frac{1}{2} + \frac{1}{2} = 1 \left[\frac{\mathrm{bit}}{\mathrm{Symbol}}\right]$$

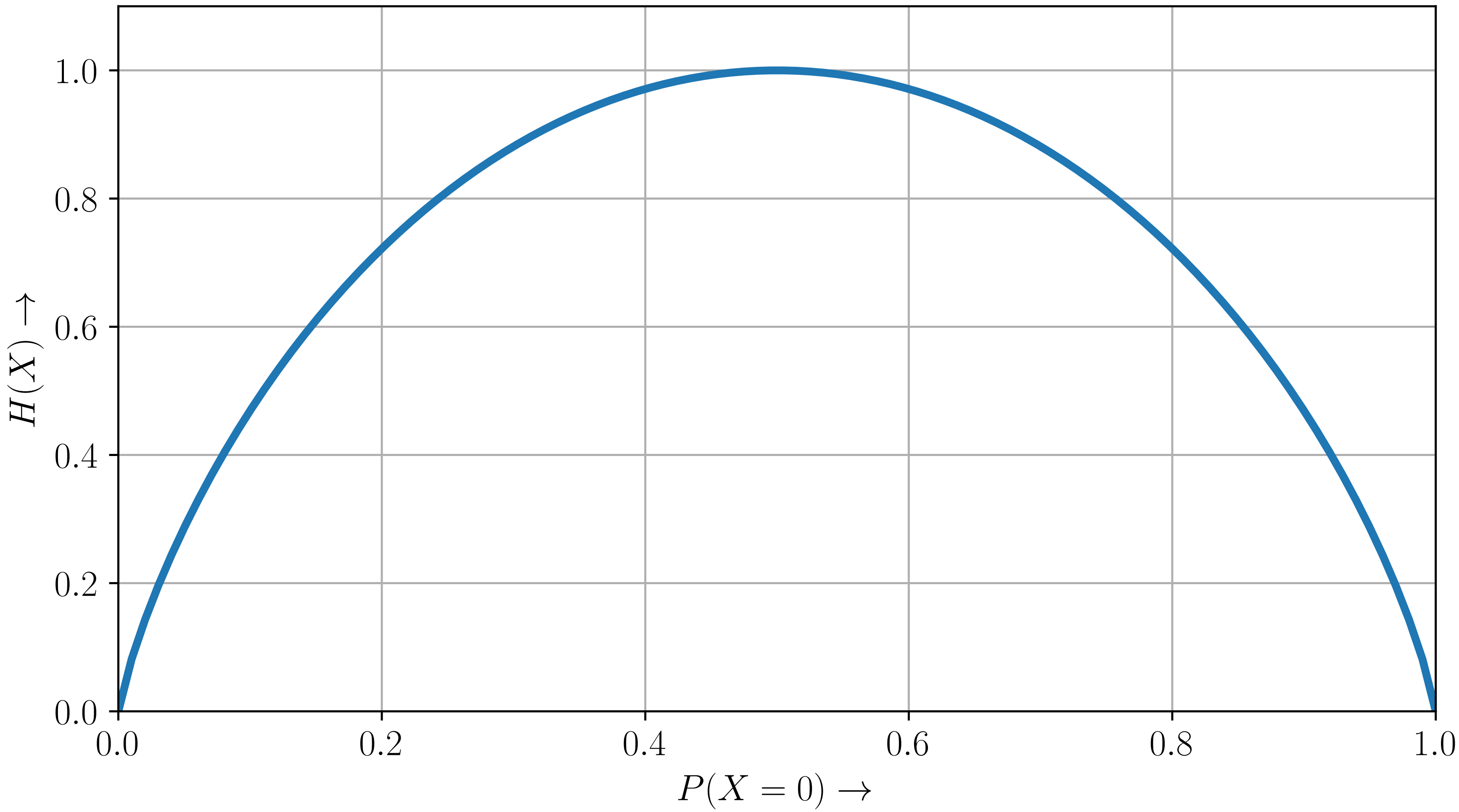
Beispiel 2: Binäre Quelle ($M_x = 2$) mit $P(X = 0)1/4$ und $P(X = 1) = 1 - P(X = 0) = 3/4$

$$H(X) = -\frac{1}{4}\mathrm{ld}\left(\frac{1}{4}\right) - \frac{3}{4}\mathrm{ld}\left(\frac{3}{4}\right) \approx 0.5 + 0.31 = 0.81 \left[\frac{\mathrm{bit}}{\mathrm{Symbol}}\right]$$

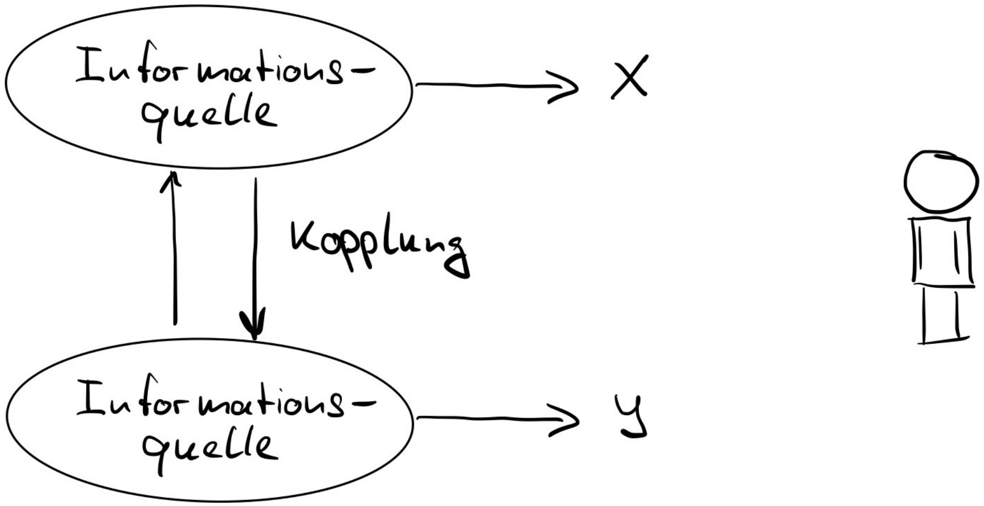
Für diskrete und gedächtnislose Quellen gilt

$$0 \leq H(X) \leq \mathrm{ld}(M_x)$$

Entropie binärer Quellen



Entropie gekoppelter Quellen I



Verbundentropie: Mittlere Informationsgewinn bei der Beobachtung beider Quellen

$$H(X, Y) = -\sum_{k=1}^{M_x} \sum_{l=1}^{M_y} P(X = x_k, Y = y_l)\mathrm{ld}(P(X = x_k, Y_= y_l)) \qquad [H(X, Y)] = \frac{\text{bit}}{\text{Symbolpaar}}$$

Für unabhängige Quellen (d.h. $P(X, Y) = P(X) \cdot P(Y)$) gilt

$$H(X, Y) = H(X) + H(Y)$$

Entropie gekoppelter Quellen II

Entropie von X falls ein konkretes Symbol y_l beobachtet wurde

$$H(X|y_l) = -\sum_{i=k}^{M_x} P(X = x_k|Y = y_l)\text{ld}(P(X = x_k|Y = y_l))$$

Bedingte Entropie

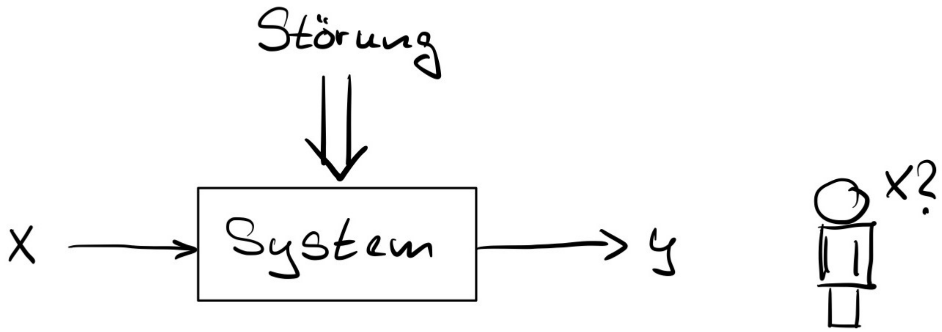
$$H(X|Y) = \sum_{l=1}^{M_y} P(Y = y_l) \cdot H(X|Y = y_l) = -\sum_{k=1}^{M_x} \sum_{l=1}^{M_y} P(X = x_k, Y = y_l)\text{ld}(P(X = x_k|Y_= y_l))$$

Beobachtung eines zusätzlichen Symbols kann Unsicherheit bezügliche eines Zufallsexperimentes nie erhöhen

$$H(X) \geq H(X|Y)$$

Wechselseitige Information

Welchen Informationsgewinn erhält man bezüglich der einen Zufallsvariable durch Beobachtung der anderen?



$$I(X;Y) = H(X) - H(X|Y)$$
$$= H(X) + H(Y) - H(X,Y)$$

Die Wechselseitige Information ist stets positiv: $I(X;Y) \geq 0$

Für statistisch unabhängige Zufallsvariablen X und Y gilt: $I(X;Y) = 0$

Data Processing Theorem: Durch Informationsverarbeitung kann niemals Information gewonnen werden (meist geht dabei Information verloren).

Differentielle Entropie

Bisherige Betrachtung: Diskrete Zufallsvariablen

Definition der differentiellen Entropie für kontinuierliche Zufallsvariablen

$$h(X) = -\int_{-\infty}^{\infty} f_X(x) \cdot \mathrm{ld}(f_X(x))\mathrm{d}x$$

Differentielle Entropie ist *kein* interpretierbares Informationsmaß

Allerdings lässt sich damit wechselseitige Information für kontinuierliche Zufallsvariablen bestimmen

$$I(X;Y) = h(X) - h(X|Y)$$

Folgerungen aus der Analyse von differentieller Entropie und wechselseitiger Information:

– Bei gegebener Varianz ist Gauß-verteilte Zufallsvariable die *zufälligste* aller Zufallsvariablen

– Gaußverteiltes weißes Rauschen ist *Worst-Case*-Modell für eine Störung