Lehrveranstaltung

Informationstheorie

— Sommersemester 2023 —

Martin Mittelbach (Vorlesung, Tutorium), Anne Wolf (Übung, Tutorium)
{martin.mittelbach, anne.wolf}@tu-dresden.de

Professur für Informationstheorie und maschinelles Lernen, TU Dresden

Vorlesung 5
26. April 2023

Wiederholung

- **Mathematisches Datenmodell:**
  (stationäre) Folge diskreter Zufallsgrößen = (stationäre) Quelle

  $$\boxed{X_1,\ X_2,\ X_3,\ X_4,\ \ldots}$$

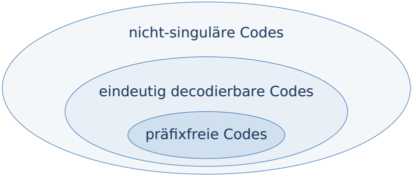
  speziell: (stationäre) gedächtnislose Quelle / (stationäre) Markow-Quelle

- **Codierung** der Werte einzelner Zufallsgrößen X_k mit Alphabet $\mathcal{X} = \{1, 2, \ldots, M\}$ und $\mathbb{W}$-Funktion p mit

  D-wertigem (Quellen-) Code $\mathcal{C}$ variabler Länge.

Quelle (Folge von Zufallsgrößen):	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	...
mögliche Werte der Zufallsgrößen:	1	2	1	1	3	2	1	4	...
	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	...
Codierung:	0	10	0	0	110	10	0	111	...

- **Codeklassen:**



Wiederholung

- **Bewertungskriterium:** Mittlere Codewortlänge

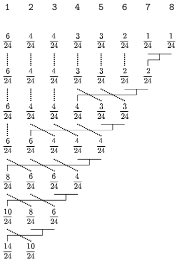
$$\boxed{\bar{\ell} = \bar{\ell}(\mathcal{C}, p) = \sum_{i=1}^{M} p(i)\, \ell(i)}$$

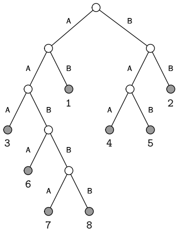
- **Optimale Codes:** Präfixfreie / eindeutig decodierbare Codes mit minimaler mittlerer Codewortlänge

$$\bar{\ell}^*_{\mathrm{ud}} = \bar{\ell}^*_{\mathrm{ud}}(p) = \min\left\{\bar{\ell}(\mathcal{C}, p) : \mathcal{C} \text{ eindeutig decodierbar}\right\}$$

$$\bar{\ell}^*_{\mathrm{pre}} = \bar{\ell}^*_{\mathrm{pre}}(p) = \min\left\{\bar{\ell}(\mathcal{C}, p) : \mathcal{C} \text{ präfixfrei}\right\}$$

- **Huffman Codes:**





Wiederholung

- **Shannonsche Informationsmaße:**

$$H(X) = -\mathbb{E}\Big(\log_2\big(p_X(X)\big)\Big) \qquad H(X,Y) = -\mathbb{E}\Big(\log_2\big(p_{X,Y}(X,Y)\big)\Big)$$

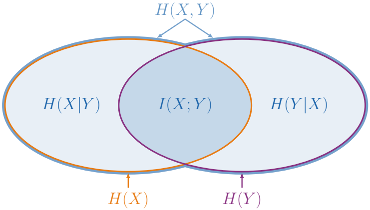
$$H(Y|X) = -\mathbb{E}\Big(\log_2\big(p_{Y|X}(Y|X)\big)\Big)$$

$$D(X||Y) = \mathbb{E}\left(\log_2\left(\frac{p_X(X)}{p_Y(Y)}\right)\right)$$

$$I(X;Y) = \mathbb{E}\left(\log_2\left(\frac{p_{X,Y}(X,Y)}{p_X(X)p_Y(Y)}\right)\right) \qquad I(X;Y|Z) = \mathbb{E}\left(\log_2\left(\frac{p_{X,Y|Z}(X,Y|Z)}{p_{X|Z}(X|Z)p_{Y|Z}(Y|Z)}\right)\right)$$

- **Grundlegende Zusammenhänge, Kettenregeln:**

$$H(X) = H(p_X) = \log_2|\mathcal{X}| - D(p_X||p_U) \qquad I(X;Y) = D(p_{X,Y}||p_X \cdot p_Y)$$



$$I(X,Y;Z) = I(X;Z) + I(Y;Z|X)$$

Inhalt der letzten Vorlesungen

- 1. Verlustlose Datenkompression mit Codes variabler Länge
- (1.1) Einführendes Beispiel, Modellbildung, Problemstellung
- (1.2) Quellen als Datenmodell
- (1.3) Codes variabler Länge
- (1.4) Huffman-Codes

- 2. Informationsmaße für diskrete Zufallsgrößen
- (2.1) Definition der Shannonschen Informationsmaße
- (2.2) Symmetrien
- (2.3) Grundlegende Zusammenhänge, Kettenregeln
- (2.4) Nichtnegativität

Inhalt Vorlesung 5

- 2. Informationsmaße für diskrete Zufallsgrößen
- (2.4) Nichtnegativität (Folien VL 4)

  $\Longrightarrow$ Fortsetzung
- (2.5) Wichtige (Un-)Gleichungen (Folien VL 4)
- (2.6) Beispiel / (Online-Zufallsexperiment)
- (2.7) n-dimensionale Verallgemeinerungen
- (2.8) Asymptotische Größen
- (2.9) Konvexitätseigenschaften

(2.6) Beispiel

Wir werten in diesem Teilabschnitt das dreistufige Online-Zufallsexperiment aus der Vorlesung 3 aus (siehe Folie 21 zu Vorlesung 4).

- **$\mathbb{W}$-theoretisches Modell:**

  Ausgang Zufallsexperiment 1 wird mit Zufallsgröße X_1 mit Alphabet $\mathcal{X}_1$ beschrieben

  Ausgang Zufallsexperiment 2 wird mit Zufallsgröße X_2 mit Alphabet $\mathcal{X}_2$ beschrieben

  Ausgang Zufallsexperiment 3 wird mit Zufallsgröße X_3 mit Alphabet $\mathcal{X}_3$ beschrieben

$$\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}_3 = \{0, 1\}$$

- **Zufallsexperiment 1 (ZE 1):** 2 Münzwürfe

$$p_{X_1}(0) = \mathbb{P}(X_1 = 0) = \mathbb{P}(\text{"zweimal Kopf"}) = 1 - q_1$$
$$p_{X_1}(1) = \mathbb{P}(X_1 = 1) = 1 - p_{X_1}(0) \qquad = q_1$$

  Für den Parameter $q_1 \in [0, 1]$ gilt bei unabhängigen Würfen mit einer fairen Münze

$$q_1 = 3/4.$$

- **Zufallsexperiment 2 (ZE 2):** 1 Münzwurf, Ausgang ZE 2 unabhängig von Ausgang ZE 1

$$p_{X_2}(0) = \mathbb{P}(X_2 = 0) = \mathbb{P}(\text{"einmal Kopf"}) = 1 - q_2$$
$$p_{X_2}(1) = \mathbb{P}(X_2 = 1) = 1 - p_{X_2}(0) \qquad = q_2$$

  Für den Parameter $q_2 \in [0, 1]$ gilt bei einem Wurf mit einer fairen Münze

$$q_2 = 1/2.$$

(2.6) Beispiel

- **Zufallsexperiment 3 (ZE 3):** 2 Münzwürfe, Ausgang ZE 3 abhängig von Ausgang ZE 2 aber nicht von Ausgang ZE 1.

$$
\begin{aligned}
p_{X_3|X_2}(1|0) &= \mathbb{P}(X_3 = 1|X_2 = 0) \\
&= \mathbb{P}(\text{"zweimal Zahl bei ZE 3 unter der Bedingung Kopf bei ZE 2"}) \\
&= \epsilon_1 \\
p_{X_3|X_2}(0|0) &= \mathbb{P}(X_3 = 0|X_2 = 0) \\
&= 1 - p_{X_3|X_2}(1|0) \\
&= 1 - \epsilon_1 \\
p_{X_3|X_2}(0|1) &= \mathbb{P}(X_3 = 0|X_2 = 1) \\
&= \mathbb{P}(\text{"zweimal Kopf bei ZE 3 unter der Bedingung Zahl bei ZE 2"}) \\
&= \epsilon_2 \\
p_{X_3|X_2}(1|1) &= \mathbb{P}(X_3 = 1|X_2 = 1) \\
&= 1 - p_{X_3|X_2}(0|1) \\
&= 1 - \epsilon_2
\end{aligned}
$$

Für die Parameter $\epsilon_1 \in [0, 1]$ und $\epsilon_2 \in [0, 1]$ gilt bei unabhängigen Würfen mit einer fairen Münze

$$\epsilon_1 = 1/4 \quad \text{und} \quad \epsilon_2 = 1/4.$$

(2.6) Beispiel

- Aus den zuvor angegebenen Wahrscheinlichkeiten lassen sich sämtliche gemeinsame / bedingte $\mathbb{W}$-Funktionen berechnen.
- (Gemeinsame) $\mathbb{W}$-Funktion von X_1 und X_2:

$$p_{X_1,X_2}(x_1,x_2) = p_{X_1}(x_1)p_{X_2}(x_2), \qquad x_1,\, x_2 \in \{0,1\}$$

$p_{X_1,X_2}(x_1,x_2)$		x_2: 0	x_2: 1	$p_{X_1}(x_1)$
x_1	0	$(1-q_1)(1-q_2)$	$(1-q_1)q_2$	$1-q_1$
	1	$q_1(1-q_2)$	q_1q_2	q_1
$p_{X_2}(x_2)$		$1-q_2$	q_2	

- (Gemeinsame) $\mathbb{W}$-Funktion von X_2 und X_3:

$$p_{X_2,X_3}(x_2,x_3) = p_{X_2}(x_2)p_{X_3|X_2}(x_3|x_2), \qquad x_2,\, x_3 \in \{0,1\}$$

$p_{X_2,X_3}(x_2,x_3)$		x_3: 0	x_3: 1	$p_{X_2}(x_2)$
x_2	0	$(1-q_2)(1-\epsilon_1)$	$(1-q_2)\epsilon_1$	$1-q_2$
	1	$q_2\epsilon_2$	$q_2(1-\epsilon_2)$	q_2
$p_{X_3}(x_3)$		$(1-q_2)(1-\epsilon_1)+q_2\epsilon_2$	$(1-q_2)\epsilon_1+q_2(1-\epsilon_2)$	

(2.6) Beispiel

- Gemeinsame $\mathbb{W}$-Funktion von X_1, X_2, X_3:

$$\begin{aligned} p_{X_1,X_2,X_3}(x_1,x_2,x_3) &= p_{X_1}(x_1)p_{X_2,X_3}(x_2,x_3) \\ &= p_{X_1}(x_1)p_{X_2}(x_2)p_{X_3|X_2}(x_3|x_2), \qquad x_1,\, x_2,\, x_3 \in \{0,1\} \end{aligned}$$

x_1	x_2	x_3	$p_{X_1,X_2,X_3}(x_1,x_2,x_3)$					
0	0	0	$(1-q_1)$	$\cdot$	$(1-q_2)$	$\cdot$	$(1-\epsilon_1)$	$\frac{3}{32}$
0	0	1	$(1-q_1)$	$\cdot$	$(1-q_2)$	$\cdot$	ϵ_1	$\frac{1}{32}$
0	1	0	$(1-q_1)$	$\cdot$	q_2	$\cdot$	ϵ_2	$\frac{1}{32}$
0	1	1	$(1-q_1)$	$\cdot$	q_2	$\cdot$	$(1-\epsilon_2)$	$\frac{3}{32}$
1	0	0	q_1	$\cdot$	$(1-q_2)$	$\cdot$	$(1-\epsilon_1)$	$\frac{9}{32}$
1	0	1	q_1	$\cdot$	$(1-q_2)$	$\cdot$	ϵ_1	$\frac{3}{32}$
1	1	0	q_1	$\cdot$	q_2	$\cdot$	ϵ_2	$\frac{3}{32}$
1	1	1	q_1	$\cdot$	q_2	$\cdot$	$(1-\epsilon_2)$	$\frac{9}{32}$

Letzte Spalte für $q_1 = \frac{3}{4}$, $q_2 = \frac{1}{2}$, $\epsilon_1 = \epsilon_2 = \frac{1}{4}$.

(2.6) Beispiel

- Das Online-Zufallsexperiment wurde 33 mal durchgeführt.
  Vielen Dank an alle Teilnehmer:innen!
- Gemeinsame $\mathbb{W}$-Funktion p_{X_1,X_2,X_3} für $q_1 = \frac{3}{4}$, $q_2 = \frac{1}{2}$, $\epsilon_1 = \epsilon_2 = \frac{1}{4}$ im Vergleich zu den gemessenen relativen Häufigkeiten des Zufallsexperimentes:

x_1	x_2	x_3	$p_{X_1,X_2,X_3}(x_1,x_2,x_3)$	Relative Häufigkeiten
0	0	0	$\frac{3}{32} \approx 0.09375$	$\frac{4}{33} \approx 0.121212$
0	0	1	$\frac{1}{32} \approx 0.03125$	$\frac{2}{33} \approx 0.060606$
0	1	0	$\frac{1}{32} \approx 0.03125$	$\frac{3}{33} \approx 0.090909$
0	1	1	$\frac{3}{32} \approx 0.09375$	$\frac{3}{33} \approx 0.090909$
1	0	0	$\frac{9}{32} \approx 0.28125$	$\frac{6}{33} \approx 0.181818$
1	0	1	$\frac{3}{32} \approx 0.09375$	$\frac{2}{33} \approx 0.030303$
1	1	0	$\frac{3}{32} \approx 0.09375$	$\frac{2}{33} \approx 0.060606$
1	1	1	$\frac{9}{32} \approx 0.28125$	$\frac{11}{33} \approx 0.333333$

Genauere Übereinstimmung erhält man durch eine größere Stichprobe.

(2.6) Beispiel

- Vergleich "Theorie / Praxis" mit Hilfe der relativen Entropie.
  - Wir fassen die im Zufallsexperiment gemessenen relativen Häufigkeiten als $\mathbb{W}$en auf und nennen die zugehörige $\mathbb{W}$-Funktion $p^{(\mathrm{ZE})}$.
  - Den Unterschied zur theoretischen $\mathbb{W}$-Funktion p_{X_1,X_2,X_3} können wir mit der relativen Entropie quantifizieren.

$$D\left(p^{(\mathrm{ZE})}||p_{X_1,X_2,X_3}\right) = \sum_{x_1,x_2,x_3\in\{0,1\}} p^{(\mathrm{ZE})}(x_1,x_2,x_3)\log_2 \frac{p^{(\mathrm{ZE})}(x_1,x_2,x_3)}{p_{X_1,X_2,X_3}(x_1,x_2,x_3)}$$
$$\approx 0.129848 \text{ (bit)}$$

  - Mit wachsender Stichprobengröße werden die beiden $\mathbb{W}$-Funktionen "immer gleicher" und $D\left(p^{(\mathrm{ZE})}||p_{X_1,X_2,X_3}\right)$ konvergiert gegen 0.

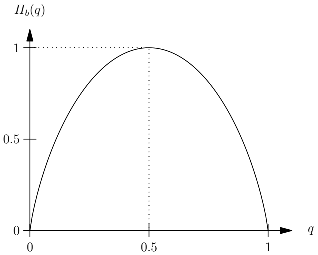
(2.6) Beispiel

- Wir berechnen nun einige Informationsmaße auf Basis der theoretischen $\mathbb{W}$-Funktionen.
- Für die Entropien der Zufallsgrößen X_1, X_2 und X_3 erhalten wir gemäß Definition der Entropie

$$H(X_1) = H_b(q_1), \quad H(X_2) = H_b(q_2), \quad H(X_3) = H_b\Big((1-q_2)\epsilon_1 + q_2(1-\epsilon_2)\Big)$$

wobei H_b die binäre Entropiefunktion darstellt.

$$H_b(q) = -q\log_2 q - (1-q)\log_2(1-q), \qquad q \in [0,1]$$



Diese einfachste Entropiefunktion illustriert bereits die generelle Konkavitätseigenschaft der Entropie.

$\longrightarrow$ allgemein siehe (2.9.2)

- Für die Parameter $q_1 = \frac{3}{4}$, $q_2 = \frac{1}{2}$, $\epsilon_1 = \epsilon_2 = \frac{1}{4}$ erhalten wir

$$H(X_1) \approx 0.811278 \text{ bit}\,, \quad H(X_2) = 1 \text{ bit}\,, \quad H(X_3) = 1 \text{ bit}$$

d. h. die Unbestimmtheit in ZE 1 ist geringer als in ZE 2 und ZE 3.

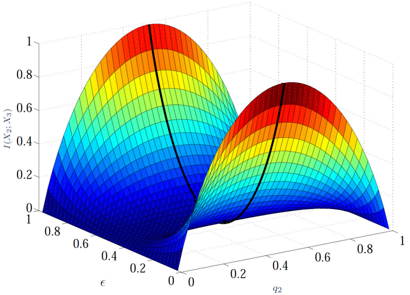
(2.6) Beispiel

- Wegen der Unabhängigkeit von X_1 und X_2 gilt nach (2.5.2)
$$H(X_2|X_1) = H(X_2)$$
und mit der Definition der bedingten Entropie erhalten wir
$$\begin{aligned} H(X_3|X_2) &= p_{X_2}(0)H\Big(p_{X_3|X_2}(\cdot|0)\Big) + p_{X_2}(1)H\Big(p_{X_3|X_2}(\cdot|1)\Big) \\ &= (1-q_2)H_b(\epsilon_1) + q_2 H_b(\epsilon_2). \end{aligned}$$
- Wegen der Unabhängigkeit von X_1 und X_2 gilt nach (2.5.4) und (2.4.4)
$$\begin{aligned} H(X_1, X_2) &= H(X_1) + H(X_2) \qquad \text{und} \qquad I(X_1; X_2) = 0. \\ &= H_b(q_1) + H_b(q_2). \end{aligned}$$
- Mit (2.3.1) erhalten wir
$$\begin{aligned} H(X_2, X_3) &= H(X_2) + H(X_3|X_2) \\ &= H_b(q_2) + (1-q_2)H_b(\epsilon_1) + q_2 H_b(\epsilon_2). \end{aligned}$$
- Mit (2.3.3) folgt
$$\begin{aligned} I(X_2; X_3) &= H(X_3) - H(X_3|X_2). \\ &= H_b\Big((1-q_2)\epsilon_1 + q_2(1-\epsilon_2)\Big) - \Big[(1-q_2)H_b(\epsilon_1) + q_2 H_b(\epsilon_2)\Big]. \end{aligned}$$
- Selbststudium: Warum gilt
$$I(X_1; X_3|X_2) = 0?$$

(2.6) Beispiel

- Für $\epsilon_1 = \epsilon_2 = \epsilon$ erhalten wir die Transinformation

$$I(X_2; X_3) = H_b\Big((1 - q_2)\epsilon + q_2(1 - \epsilon)\Big) - H_b(\epsilon).$$



Diese einfache Transinformationsfunktion illustriert bereits die generellen Konkavitäts-/Konvexitätseigenschaften der Transinformation:
Bzgl. eines Arguments konvex, bzgl. des anderen Arguments konkav.

$\longrightarrow$ allgemein siehe (2.9.3/4)

- Für die Parameter $q_1 = \frac{3}{4}$, $q_2 = \frac{1}{2}$, $\epsilon_1 = \epsilon_2 = \frac{1}{4}$ erhalten wir

$$H(X_3|X_2) \approx 0.811278 \text{ bit} \leq H(X_3) = 1 \text{ bit}$$

$$H(X_1, X_2) \approx 1.811278 \text{ bit}, \quad H(X_2, X_3) \approx 1.811278 \text{ bit}, \quad I(X_2; X_3) \approx 0.188722 \text{ bit}$$

- Vergleich mit Werten der Informationsmaße für "gemessene" $\mathbb{W}$-Funktion $p^{(\mathrm{ZE})}$
$\Longrightarrow$ Selbststudium

(2.7) n-dimensionale Verallgemeinerungen

$X_1, X_2, \ldots, X_n$ seien n diskrete Zufallsgrößen und $X = (X_1, X_2, \ldots, X_n)$ der entsprechende n-dimensionale Zufallsvektor.

- **(2.7.1) Kettenregel Entropie für n Zufallsgrößen:**

$$H(X_1, X_2, \ldots, X_n) = \sum_{k=1}^{n} H(X_k | X_{k-1}, X_{k-2}, \ldots, X_1)$$

- **(2.7.2) Kettenregel Transinformation für n Zufallsgrößen:**

$$I(X_1, X_2, \ldots, X_n; Y) = \sum_{k=1}^{n} I(X_k; Y | X_{k-1}, X_{k-2}, \ldots, X_1)$$

- **(2.7.3) Unabhängigkeit maximiert Entropie für n Zufallsgrößen:**

$$H(X_1, X_2, \ldots, X_n) \leq \sum_{k=1}^{n} H(X_k)$$

$$H(X_1, X_2, \ldots, X_n) = \sum_{k=1}^{n} H(X_k) \quad \Longleftrightarrow \quad X_1, X_2, \ldots, X_n \text{ sind unabhängig}$$

(2.7) n-dimensionale Verallgemeinerungen

- **Herleitung zu (2.7.1):**

(2.7) n-dimensionale Verallgemeinerungen

- **Herleitung zu (2.7.2):**

(2.7) n-dimensionale Verallgemeinerungen

- **Herleitung zu (2.7.3):**

(2.7) n-dimensionale Verallgemeinerungen

- **(2.7.4) Kettenregel bedingte Entropie für n Zufallsgrößen:**

$$H(X_1, X_2, \ldots, X_n|Y) = \sum_{k=1}^{n} H(X_k|X_{k-1}, X_{k-2}, \ldots, X_1, Y)$$

- **(2.7.5) Kettenregel bedingte Transinformation für n Zufallsgrößen:**

$$I(X_1, X_2, \ldots, X_n; Y|Z) = \sum_{k=1}^{n} I(X_k; Y|X_{k-1}, X_{k-2}, \ldots, X_1, Z)$$

- **(2.7.6) Bedingte Unabhängigkeit maximiert bedingte Entropie für n Zufallsgrößen:**

$$H(X_1, X_2, \ldots, X_n|Y) \leq \sum_{k=1}^{n} H(X_k|Y)$$

$$H(X_1, X_2, \ldots, X_n|Y) = \sum_{k=1}^{n} H(X_k|Y) \iff X_1, X_2, \ldots, X_n \text{ sind bedingt unabhängig gegeben } Y$$

(2.8) Asymptotische Größen

- **(2.8.1) Entropierate:** Die Größe

$$\overline{H}(X) = \lim_{n\to\infty} \frac{H(X_1, X_2, \ldots, X_n)}{n}$$

nennen wir Entropierate der Folge $X = (X_k)_{k\in\mathbb{N}}$ diskreter Zufallsgrößen, sofern der Grenzwert existiert.

$$X = \Bigg(\underbrace{\underbrace{\underbrace{\underbrace{\underbrace{X_1,}\ X_2,}\ X_3,\ \ldots\ ,}\ X_n,}\ X_{n+1},}_{\cdots} X_{n+2}, \ldots \Bigg)$$

Bemerkung:

Die Entropierate ist u. a. eine untere Schranke für die (asymptotisch) verlustlose, "blockweise" Datenkompression, speziell bei gedächtnisbehafteten Modellen.

(2.8) Asymptotische Größen

- **(2.8.2) Transinformationsrate:** Die Größe

$$\overline{I}(X;Y) = \lim_{n\to\infty} \frac{I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n)}{n}$$

nennen wir Transinformationsrate der Folgen $X = (X_k)_{k\in\mathbb{N}}$ und $Y = (Y_k)_{k\in\mathbb{N}}$ diskreter Zufallsgrößen, sofern der Grenzwert existiert.

$$X = \left(X_1,\ X_2,\ X_3,\ \ldots\ ,\ X_n,\ X_{n+1}, X_{n+2}, \ldots\right)$$

$$Y = \left(Y_1,\ Y_2,\ Y_3,\ \ldots\ ,\ Y_n,\ Y_{n+1}\ , Y_{n+2}, \ldots\right)$$

Bemerkung: Mit der Transinformationsrate kann man u. a. eine obere Schranke für die Datenrate für eine zuverlässige Datenübertragung erhalten, speziell bei gedächtnisbehafteten Modellen.

(2.8) Asymptotische Größen

- **(2.8.3) Beispiel: Entropierate einer stationären gedächtnislosen Quelle**
  Die Entropierate einer stationären gedächtnislosen Quelle (siehe (1.2) in VL 2) $X = (X_k)_{k\in\mathbb{N}}$ is gegeben durch

$$\overline{H}(X) = \lim_{n\to\infty} \frac{H(X_1, X_2, \ldots, X_n)}{n} = H(X_1).$$

  $\Longrightarrow$ Konkretes Beispiel siehe 2. Übung, Aufgabe 10 g)

- **Herleitung zu (2.8.3):**

(2.8) Asymptotische Größen

- **(2.8.4) Beispiel: Entropierate einer stationären Markow-Quelle**

  Die Entropierate einer stationären Markow-Quelle (siehe (1.2) in VL 2) $X = (X_k)_{k \in \mathbb{N}}$ is gegeben durch

  $$\overline{H}(X) = \lim_{n \to \infty} \frac{H(X_1, X_2, \ldots, X_n)}{n} = H(X_2|X_1).$$

  $\Longrightarrow$ Konkrete Beispiele siehe 2. Übung, Aufgabe 11 c), sowie 2. Hausaufgabe, Aufgabe 15 d)

- **Herleitung zu (2.8.4):**

(2.8) Asymptotische Größen

- **Herleitung zu (2.8.4):**

(2.9) Konvexitätseigenschaften

- **(2.9.1) Konvexkombination von $\mathbb{W}$-Funktionen**
  - Sei X eine diskrete Zufallsgröße mit dem Alphabet $\mathcal{X}$.
  - Seien $p_X^{(1)}$ und $p_X^{(2)}$ zwei $\mathbb{W}$-Funktionen von X auf dem Alphabet $\mathcal{X}$.
  - Dann definieren wir für $\lambda \in [0,1]$ die $\mathbb{W}$-Funktion

$$p_X = \lambda p_X^{(1)} + (1-\lambda) p_X^{(2)}$$

elementweise, d. h. für alle $x \in \mathcal{X}$

$$p_X(x) = \lambda p_X^{(1)}(x) + (1-\lambda) p_X^{(2)}(x).$$

  - Die $\mathbb{W}$-Funktion p_X heißt Konvexkombination der $\mathbb{W}$-Funktionen $p_X^{(1)}$ und $p_X^{(2)}$.
  - Zahlenbeispiel: Alphabet $\mathcal{X} = \{1,2,3,4\}$ und $\mathbb{W}$-Funktionen

x	1	2	3	4
$p_X^{(1)}(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

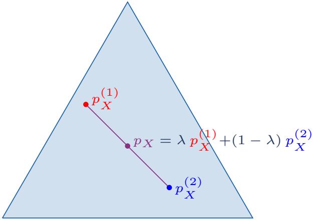
x	1	2	3	4
$p_X^{(2)}(x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

$\Longrightarrow$ Konvexkombination

x	1	2	3	4
$p_X(x)$	$\frac{1}{4}(1+\lambda)$	$\frac{1}{4}$	$\frac{1}{8}(2-\lambda)$	$\frac{1}{8}(2-\lambda)$

(2.9) Konvexitätseigenschaften

- **(2.9.1) Konvexkombination von $\mathbb{W}$-Funktionen** $[\ldots]$
  - Illustration:



Menge der $\mathbb{W}$-Funktionen auf dem Alphabet $\mathcal{X}$

  - Eine Konvexkombination von $\mathbb{W}$-Funktionen entsteht folgendermaßen:

X_1 diskrete Zufallsgröße mit endlichem Alphabet $\mathcal{X}$ und $\mathbb{W}$-Funktion p_{X_1}

X_2 diskrete Zufallsgröße mit endlichem Alphabet $\mathcal{X}$ und $\mathbb{W}$-Funktion p_{X_2}

Z diskrete Zufallsgröße mit binärem Alphabet $\{1,2\}$ und $\mathbb{W}$-Funktion p_Z

$$p_Z(1) = \mathbb{P}(Z=1) = \lambda \quad \text{und} \quad p_Z(2) = \mathbb{P}(Z=2) = (1-\lambda)$$

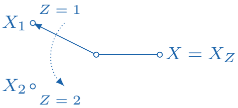
Die Zufallsgröße $X = X_Z$ hat die $\mathbb{W}$-Funktion

$$p_X = \lambda p_{X_1} + (1-\lambda) p_{X_2}.$$

(2.9) Konvexitätseigenschaften

- **(2.9.1) Konvexkombination von $\mathbb{W}$-Funktionen** $[\ldots]$
  - Erläuterungen: Die Zufallsgröße Z hat die Wirkung eines Schalters, der zwischen den Zufallsgrößen X_1 und X_2 mit den $\mathbb{W}$en λ und $1-\lambda$ auswählt.

Falls $Z = 1$, dann gilt $X = X_1$.



Falls $Z = 2$, dann gilt $X = X_2$.

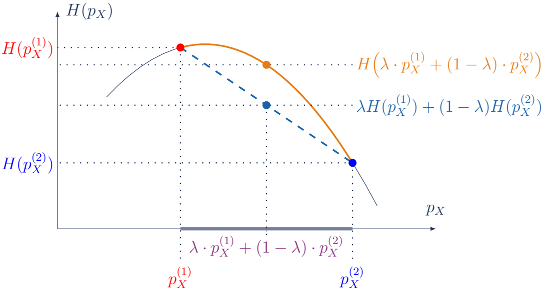
Die $\mathbb{W}$-Funktion p_X von $X = X_Z$ erhält man als Rand-$\mathbb{W}$-Funktion von $p_{X,Z}$.

$$
\begin{aligned}
p_X(x) &= \sum_{z \in \mathcal{Z}} p_{X,Z}(x,z) \\
&= \sum_{z \in \mathcal{Z}} p_Z(z) p_{X|Z}(x|z) \\
&= p_Z(1) p_{X|Z}(x|1) + p_Z(2) p_{X|Z}(x|2) \\
&= \lambda p_{X_1}(x) + (1-\lambda) p_{X_2}(x)
\end{aligned}
$$

(2.9) Konvexitätseigenschaften

- **(2.9.2) Konkavität der Entropie:**
  Die Entropie $H(X) = H(p_X)$ einer diskreten Zufallsgröße X ist konkav bezüglich der $\mathbb{W}$-Funktion p_X. D. h. für alle $\mathbb{W}$-Funktionen $p_X^{(1)}$ und $p_X^{(2)}$ von X und alle $\lambda \in [0,1]$ gilt

$$H\left(\lambda p_X^{(1)} + (1-\lambda)p_X^{(2)}\right) \geq \lambda H\left(p_X^{(1)}\right) + (1-\lambda)H\left(p_X^{(2)}\right).$$



Bemerkung: Die Konkavität ist relevant bei der Maximierung der Entropie.

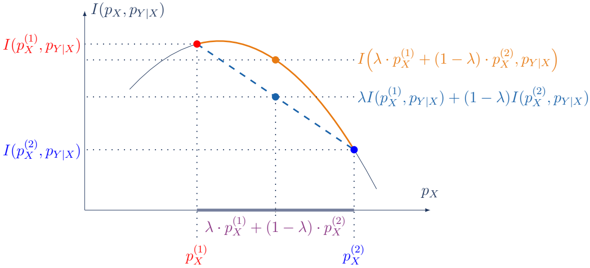
(2.9) Konvexitätseigenschaften

- **Herleitung zu (2.9.2):**

(2.9) Konvexitätseigenschaften

- **(2.9.3) Konkavität der Transinformation $I(p_X, p_{Y|X})$ bezüglich p_X:**
  Die Transinformation $I(X;Y) = I(p_X, p_{Y|X})$ zwischen den diskreten Zufallsgrößen X und Y ist konkav bezüglich der $\mathbb{W}$-Funktion p_X bei fester bedingter $\mathbb{W}$-Funktion $p_{Y|X}$. D. h. für alle $\mathbb{W}$-Funktionen $p_X^{(1)}$, $p_X^{(2)}$ von X und alle $\lambda \in [0,1]$ gilt

$$I\left(\lambda p_X^{(1)} + (1-\lambda) p_X^{(2)}, p_{Y|X}\right) \geq \lambda I\left(p_X^{(1)}, p_{Y|X}\right) + (1-\lambda) I\left(p_X^{(2)}, p_{Y|X}\right).$$



Bemerkung:

Die Konkavität ist relevant bei der Maximierung der Transinformation bzgl. p_X.

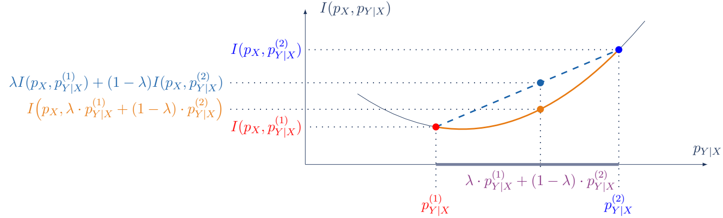
(2.9) Konvexitätseigenschaften

- **Herleitung zu (2.9.3):**

(2.9) Konvexitätseigenschaften

- **(2.9.4) Konvexität der Transinformation $I(p_X, p_{Y|X})$ bezüglich $p_{Y|X}$:**
  Die Transinformation $I(X;Y) = I(p_X, p_{Y|X})$ zwischen den diskreten Zufallsgrößen X und Y ist konvex bezüglich der bedingten $\mathbb{W}$-Funktion $p_{Y|X}$ bei fester $\mathbb{W}$-Funktion p_X. D. h. für alle bedingten $\mathbb{W}$-Funktionen $p^{(1)}_{Y|X}, p^{(2)}_{Y|X}$ von Y unter der Bedingung X und alle $\lambda \in [0,1]$ gilt

$$I\Big(p_X, \lambda p^{(1)}_{Y|X} + (1-\lambda) p^{(2)}_{Y|X}\Big) \leq \lambda I\Big(p_X, p^{(1)}_{Y|X}\Big) + (1-\lambda) I\Big(p_X, p^{(2)}_{Y|X}\Big).$$



Bemerkung:
Die Konvexität ist relevant bei der Minimierung der Transinformation bzgl. $p_{Y|X}$.
Herleitung: Bei Bedarf später.