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 6
03. Mai 2023

Wiederholung

- **Shannonsche Informationsmaße:**

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

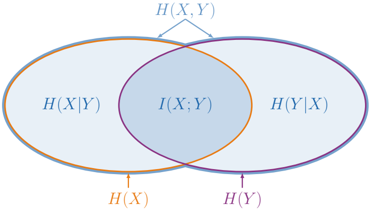
$$H(Y|X) = -\mathbb{E}\Big(\log_2\big(p_{Y|X}(Y|X)\big)\Big)$$

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

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

- **Grundlegende Zusammenhänge, Kettenregeln:**

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



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

Wiederholung

- **Nichtnegativität / Wichtige (Un-)Gleichungen:**

$$\left\{ D(X||Y),\quad H(X),\quad H(X|Y),\quad I(X;Y),\quad I(X;Y|Z) \right\} \geq 0$$

$D(X||Y) = 0 \Leftrightarrow X \sim Y$

$H(X) = 0 \Leftrightarrow X$ deterministisch

$H(X|Y) = 0 \Leftrightarrow X$ Funktion von Y

$I(X;Y) = 0 \Leftrightarrow X, Y$ unabhängig

$I(X;Y|Z) = 0 \Leftrightarrow X, Y$ bed. unabh. geg. Z

$$H(X|Y,Z) \leq H(X|Y) \leq H(X) \leq \log_2 |\mathcal{X}|$$

$H(X|Y,Z) = H(X|Y) \Leftrightarrow X \to Y \to Z$

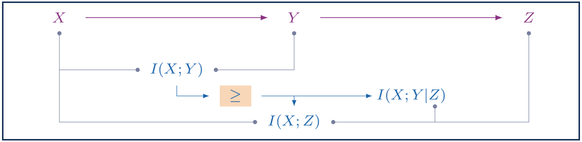
$H(X|Y) = H(X) \Leftrightarrow X, Y$ unabhängig

$H(X) = \log_2 |\mathcal{X}| \Leftrightarrow X$ gleichverteilt

$$H(X,Y) \leq H(X) + H(Y) \qquad H(X,Y|Z) \leq H(X|Z) + H(Y|Z)$$

$H(X,Y) = H(X) + H(Y) \Leftrightarrow X, Y$ unabhängig

$H(X,Y|Z) = H(X|Z) + H(Y|Z) \Leftrightarrow X, Y$ bed. unabh. geg. Z



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
- (2.5) Wichtige (Un-)Gleichungen
- (2.6) Beispiel / Online-Zufallsexperiment

Inhalt Vorlesung 6

- 2. Informationsmaße für diskrete Zufallsgrößen
- (2.7) n-dimensionale Verallgemeinerungen (Folien VL 5)
- (2.8) Asymptotische Größen (Folien VL 5)
- (2.9) Konvexitätseigenschaften (Folien VL 5)

- 3. Eigenschaften von eindeutig decodierbaren und präfixfreien Codes
- (3.1) Kraft-McMillan-Ungleichung
- (3.2) Entropieschranke für eindeutig decodierbare Codes
- (3.3) Eigenschaften von Huffman-Codes

Vorgehensweise / Themenübersicht

(1) Praktische Problemstellung

- systematischer Entwurf effizienter Datenkompression
- systematischer Entwurf effizienter zuverlässiger Datenübertragung

(2) Abstraktion / Mathematische Modellbildung

- stochastisches Datenmodell
  Quelle = Folge von Zufallsgrößen
- deterministisches Kompressionsmodell
  Quellen-Codier-Decodier-Schemata
- operationelle Bewertungskenngrößen
  (minimale) mittlere Codewortlänge / Coderate, Distortionmaß, Rate-Distortion-Funktion

- stochastisches Übertragungsmodell
  Kanal = bedingte Wahrscheinlichkeitsverteilung
- deterministisches Kommunikationsmodell
  Kanal-Codier-Decodier-Schemata
- operationelle Bewertungskenngrößen
  (maximale) Coderate, Decodierfehlerwahrscheinlichkeit, Codierkapazität

(3) Entwicklung mathematischer Werkzeuge für Modellanalyse

- Informationsmaße
  Entropie (bedingte, relative, differentielle, empirische ∼), Transinformation (bedingte, empirische ∼)
- Eigenschaften, Rechenregeln
  (Symmetrie, Nichtnegativität), Kettenregeln, Konvexität / Konkavität
  Unabhängigkeit / (Gleichverteilung (Normalverteilung)) maximiert Entropie (differentielle ∼)
- Ungleichungen, grundlegende Theoreme
  Kraft-McMillan-Ungleichung, fundamentale Ungleichung, Datenverarbeitungsungleichung, Fano-Ungleichung
  Shannon-McMillan-Theorem, Feinstein-Lemma
- Informationsmaßebasierte Kenngrößen
  Redundanz, Entropierate, Informations-Rate-Distortion-Funktion, Informationskapazität

(4) Herleitung fundamentaler Resultate für abstrahiertes Modell mit math. Werkzeugen

- Quellencodierungstheoreme
  Aussagen über theoretische Grenzen effizienter Datenkompression (bei definierter Güte)
  Sinngemäß: Verlustlose (asymptotisch ∼) Datenkompression ist begrenzt durch Entropie(-rate)
  (Entropie(-rate) $\leq$ min. mittl. Cwl (pro Sym.) / Coderate)
  Sinngemäß: Verlustbehaftete Datenkompression ist begrenzt durch Informations-Rate-Distortion-Funktion
  (Rate-Distortion-Funktion = Informations-∼)

- Kanalcodierungstheoreme
  Aussagen über theoretische Grenzen effizienter Datenübertragung bei definierter Zuverlässigkeit
  Sinngemäß: Maximale Coderate bei zuverlässiger Datenübertragung ist begrenzt durch Informationskapazität
  (Codierkapazität = Informationskapazität)

(5) Anwendung der Resultate auf praktische Problemstellung

Entwurfskriterien für optimale Datenkompression / -übertragung, Systembewertung

3. Eigenschaften von eindeutig decodierbaren und präfixfreien Codes

Vorbemerkungen / Festlegungen zu Abschnitt 3

- Abschnitt 3 ist die Fortsetzung von Abschnitt 1.
- Wir betrachten Modell / Codierung aus Teilabschnitt (1.3), d. h. wir betrachten die symbolweise Codierung der Werte einzelner Zufallsgrößen einer

  stationäre Quelle $X = (X_1, X_2, X_3, X_4, \ldots)$

  mit Alphabet $\mathcal{X} = \{1, 2, \ldots, M\}$

  und der für alle Folgeglieder X_k identischen $\mathbb{W}$-Funktion p

  mittels

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

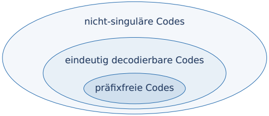
- Bewertung der Effizienz eines D-wertigen Codes $\mathcal{C} = \{c(i), i \in \mathcal{X}\}$ mit Codewortlängen $\ell(i)$, $i \in \mathcal{X}$, für $\mathbb{W}$-Funktion p mit mittlerer Codewortlänge

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

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

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

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



Vorbemerkungen / Festlegungen zu Abschnitt 3

- **Zur Erinnerung** das Einführungsbeispiel aus Abschnitt 1 (Vorlesungen 2/3):
  - Alphabet $\mathcal{X} = \{1, 2, 3, 4\}$ und $\mathbb{W}$-Funktion p

i	1	2	3	4
$p(i)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

  - Codierung mit Code

$$\mathcal{C}_3 = \{0, 10, 110, 111\}.$$

  - Beispiel Codierung Datenstrom:

(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	...
	↓	↓	↓	↓	↓	↓	↓	↓	...
binäre Codierung mit $\mathcal{C}_3$:	0	10	0	0	110	10	0	111	...

  - Der Code $\mathcal{C}_3$ ist:
    - eindeutig decodierbar,
    - präfixfrei,
    - sogar ein Huffman-Code, d. h. ein spezieller präfixfreier Code.

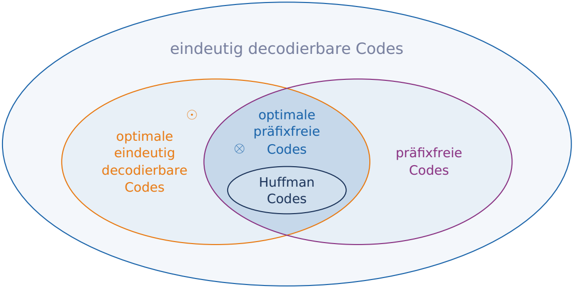
Zusammenfassung wesentlicher Resultate

- **Zusammenfassung** der wesentlichen Ergebnisse aus den folgenden Teilabschnitten (3.1) – (3.3) für D-wertige optimale Codes für festes Alphabet $\mathcal{X}$, feste $\mathbb{W}$-Funktion p und symbolweise Codierung:
  - (Un-) Gleichungen:

$$\boxed{H_D(p) \overset{(3.2.1)}{\leq} \bar{\ell}^*_{\mathrm{ud}}(p) \overset{(3.1.2)}{=} \bar{\ell}^*_{\mathrm{pre}}(p) \overset{(3.3.1)}{=} \bar{\ell}_{\mathrm{Huff}}(p) \overset{(3.3.3)}{<} H_D(p) + 1}$$

    (Fundamentale und Kraft-McMillan-Ungleichung sind die wichtigsten Herleitungswerkzeuge.)

  - Relationen:



Beispiele für ⊙ und ⊗: $\longrightarrow$ Selbststudium

(3.1) Kraft-McMillan-Ungleichung

- **(3.1.1) Kraft-McMillan-Ungleichung:**
  - (i): Jeder eindeutig decodierbare D-wertige Code $\mathcal{C} = \{c(1), c(2), \ldots, c(M)\}$ mit Codwortlängen $\ell(1), \ell(2), \ldots, \ell(M)$ erfüllt die Ungleichung

$$\sum_{i=1}^{M} D^{-\ell(i)} \leq 1.$$

  - (ii): Ist für die Codewortlängen $\ell(1), \ell(2), \ldots, \ell(M)$ die Ungleichung aus (i) erfüllt, so existiert ein D-wertiger präfixfreier Code mit denselben Codewortlängen.
- **Herleitung zu (3.1.1):** Siehe z. B. S. 43 und S. 47 in R. W. Yeung: A First Course in Information Theory, Springer, 2002.

- **(3.1.2) Folgerungen:**
  - Zu jedem eindeutig decodierbaren D-wertigen Code existiert ein D-wertiger präfixfreier Code, mit denselben Codewortlängen.

    Diese Aussage erhält man aus der Kombination von (3.1.1), (i), und (3.1.1), (ii).
  - Für die minimalen mittleren Codewortlängen $\bar{\ell}^*_{\mathrm{ud}}(p)$ und $\bar{\ell}^*_{\mathrm{pre}}(p)$ eindeutig decodierbarer bzw. präfixfreier D-wertiger Codes für die $\mathbb{W}$-Funktion p gilt

$$\boxed{\bar{\ell}^*_{\mathrm{ud}}(p) = \bar{\ell}^*_{\mathrm{pre}}(p).}$$

    Diese Aussage erhält man mit der vorhergehenden Folgerung.

(3.1) Kraft-McMillan-Ungleichung

- **Historisches:**
  - Die Kraft-McMillan-Ungleichung ist nach Leon Gordon Kraft und Brockway McMillan benannt.
  - Kraft veröffentlichte diese 1949 in seiner Masterarbeit (MIT), beschränkte sich jedoch auf den präfixfreien Fall.
  - Die Verallgemeinerung für eindeutig decodierbare Codes wurde 1956 von McMillan vorgenommen.

Kraft, Leon G. (1949): A device for quantizing, grouping, and coding amplitude modulated pulses, Cambridge, MA: MS Thesis, Electrical Engineering Department, Massachusetts Institute of Technology.

McMillan, Brockway (1956): Two inequalities implied by unique decipherability, IEEE Trans. Inf. Theory, 2(4):115-116.

(3.2) Entropieschranke für eindeutig decodierbare Codes

- **Vorbemerkung:** Wir haben die Entropie in (2.1.1) bezüglich des Logarithmus $\log_2$ zur Basis 2 definiert. Für die Analyse von D-wertigen Quellencodes betrachten wir die Entropie bezüglich des Logarithmus $\log_D$ zur Basis D. Als Bezeichung verwenden wir $H_D(\cdot)$ anstatt $H(\cdot)$. Es gilt aufgrund der Eigenschaften des Logarithmus

$$H_D(\cdot) = \frac{H(\cdot)}{\log_2(D)}.$$

- **(3.2.1) Entropieschranke für eindeutig decodierbare Codes:**
  - (i) Die minimale mittlere Codewortlänge $\bar{\ell}^*_{\text{ud}}(p)$ eindeutig decodierbarer D-wertiger Codes für die $\mathbb{W}$-Funktion p erfüllt die Ungleichung

$$\boxed{H_D(p) \leq \bar{\ell}^*_{\text{ud}}(p).}$$

  - (ii) Für die mittlere Codewortlänge $\bar{\ell}(\mathcal{C}, p)$ eines eindeutig decodierbaren D-wertiger Codes $\mathcal{C} = \{c(1), c(2), \dots, c(M)\}$ gilt

$$\bar{\ell}(\mathcal{C}, p) = H_D(p)$$

    genau dann, wenn für alle $i = 1, 2, \dots, M$ die Codewortlängen $\ell(i)$ die Gleichung

$$\ell(i) = -\log_D p(i)$$

    erfüllen.

(3.2) Entropieschranke für eindeutig decodierbare Codes

- **(3.2.2) Folgerung:**
  - Gilt für die mittlere Codewortlänge $\bar{\ell}(\mathcal{C}, p)$ eines eindeutig decodierbaren D-wertiger Codes $\mathcal{C} = \{c(1), c(2), \ldots, c(M)\}$ die Gleichung

    $$\bar{\ell}(\mathcal{C}, p) = H_D(p),$$

    dann ist $\mathcal{C}$ ein optimaler eindeutig decodierbarer Code.

    Diese Aussage erhält man direkt mit (i) in (3.2.1).

- **Bemerkungen:**
  - Ob ein für die $\mathbb{W}$-Funktion p optimaler, D-wertiger, eindeutig decodierbarer Code $\mathcal{C}$ (d. h. $\bar{\ell}(\mathcal{C}, p) = \bar{\ell}^*_{\mathrm{ud}}(p)$) die Gleichung $\bar{\ell}(\mathcal{C}, p) = H_D(p)$ erfüllen kann, hängt von der konkreten $\mathbb{W}$-Funktion p ab.
  - Für den optimalen binären Code $\mathcal{C}_3$ ist die Gleichung $\bar{\ell}(\mathcal{C}_3, p) = H(p)$ erfüllt.

(3.2) Entropieschranke für eindeutig decodierbare Codes

- **Herleitung zu (3.2.1):**

(3.3) Eigenschaften von Huffman-Codes

- **(3.3.1) Optimalität von Huffman-Codes:**
  Die mittlere Codewortlänge $\bar{\ell}_{\mathrm{Huff}}(p)$ eines D-wertigen Huffman-Codes für die $\mathbb{W}$-Funktion p erfüllt die Gleichung

$$\bar{\ell}_{\mathrm{Huff}}(p) = \bar{\ell}^*_{\mathrm{pre}}(p),$$

  d. h. Huffman-Codes sind optimale präfixfreie Codes.

- Vorgehensweise zur Herleitung:
  - Wir beschränken uns bei der Herleitung der Optimalität von Huffman-Codes der Übersichtlichkeit halber auf den 2-wertigen (d. h. binären) Fall. Der Nachweis für den D-wertigen Fall ist analog.
  - Wir verwenden für die Herleitung von (3.3.1) folgende einfache Eigenschaften optimaler (binärer) Codes, die bei Huffman-Codes aufgrund der Konstruktion stets erfüllt sind.

- **(3.3.2) Eigenschaften optimaler Codes:**
  - (i) Bei einem optimalen (eindeutig decodierbaren oder präfixfreien) D-wertigen Code werden kürzere Codewörter größeren $\mathbb{W}$en zugeordnet.
  - (ii) Es existiert ein optimaler präfixfreier 2-wertiger Code, bei dem die Codewörter, die den beiden kleinsten $\mathbb{W}$en zugeordnet sind, gleich lang sind und sich nur im letzten Symbol (Bit) unterscheiden.

(3.3) Eigenschaften von Huffman-Codes

- **Herleitung zu (3.3.1) (Optimalität von Huffman-Codes):**
  (binärer Fall, D-wertiger Fall analog)

(3.3) Eigenschaften von Huffman-Codes

- **Herleitungen zu (3.3.2) (Eigenschaften optimaler Codes):**

(3.3) Eigenschaften von Huffman-Codes

- **(3.3.3) Obere Entropieschranke für Huffman-Codes:**

  Die mittlere Codewortlänge $\bar{\ell}_{\mathrm{Huff}}(p)$ eines D-wertigen Huffman-Codes für die $\mathbb{W}$-Funktion p erfüllt die Ungleichung

  $$\bar{\ell}_{\mathrm{Huff}}(p) < H_D(p) + 1.$$

  Diese obere Schranke ist die beste, die nur von der Entropie abhängt.

- **Herleitung zu (3.3.3):**