Gegenüberstellung der metrischen Entropie und shannon Entropie

Bachelorarbeit

vorgelegt am: 04/01/2021

am Fachbereich Mathematik der
Ruhr-Universität-Bochum

Name:	Seil Hong
Matrikelnummer:	108016263063
Studiengang:	Mathematik
Studienjahrgang:	2020
Erstgutachter:	Prof. Dr. Johannes Lederer
Zweitgutachter:	Prof. Dr. Holger Dette

Contents

1 Einleitung

Viele Menschen trinken gerne heißen Kaffee. Man kann jedoch nur kurze Zeit heißen Kaffee genießen. Der Grund liegt daran, dass der Kaffee bei Raumtemperatur schnell abkühlt. Warum kühlt Kaffee so schnell ab? In der Physik antwortet der zweite Hauptsatz der Thermodynamik auf die Frage. In der Physik ist der zweite Hauptsatz der Thermodynamik ein Gesetz, das nicht zu einem Makrozustand mit einer kleineren Entropie übergeht, wenn die Entropie des Makrozustands des Systems zu Zeitpunkt in einem thermisch isolierte System berücksichtigt wird.[1] In anderen Wörtern, Die Energie fließt in Richtung zunehmender Entropie im Laufe der Zeit. Arten von Energie umfassen Wärmeenergie, kinetische Energie, Kernenergie und so weiter. Ein Phänomen, bei dem sich Wärme von einer hohen zu einer niedrigen Temperatur bewegt, und ein anderes Phänomen, bei dem sich Luft vom Hochdruck zum Niederdruck bewegt, sind gute Beispiele für Entropie. Entropie wird jedoch nicht nur in der Physik, sondern auch in der Informatik und in der Statistik verwendet.

Entropie in der Informatik wurde im Jahr 1948 von Claude E. Shannon entwickelt. [7] Entropie aus der Informatik ist inspiriert von folgender Problematik. Man betrachtet einen Kanal, der für den Transport von Daten zuständig ist — beispielsweise eine Telefonleitung — so kann es bei der Übertragung zu Fehlern kommen. Claude E. Shannon bestimmte die Länge der übertragenen Daten durch Entropie, damit der Empfänger die richtigen Daten empfangen kann. Die Anwendungen der Shannon Entropie ist wie folgt: Erstens, die Anzahl der Bits bei der Codierung einer Information wird bestimmt. Zweitens, die Anzahl der Länge einer Information für den Schutz gegen Datenverlust bei der Übertragung der Information über einen Kanal wird entschieden. [8] Shannon Theorem ist deshalb ein wesentliches Element beim Senden und Empfangen von Daten und wird heutzutage in der gesamten Datenkommunikation häufig verwendet. Aus den Gründen wird seine Arbeit sehr häufig zitiert.

In der Statistik wird die Entropie auch verwendet. Die Entropie heißt metrische Entropie. Mit der metrischen Entropie kann nicht nur das Supremum einer Menge, sondern auch das Supremum von empirischen Prozesse. Aber man kann das Theorem nicht direkt zu empirischen Prozesse anwenden. Daher sind einige Vorbereitungen notwendig. In meiner Arbeit werden Bedingungen gezeigt, unter den ein Theorem auf empirische Prozesse angewendet werden kann.

Das Ziel meiner Arbeit ist Shannon Entropie und metrische Entropie und ihre Interpretation und Anwendungen gut zu verstehen, und Zusammenhängen der beiden Entropie herauszufinden. Um das Ziel zu erreichen, stelle ich Shannon Entropie in Kapitel 2 und metrische Entropie in Kapitel 3 vor, und die beiden Definition werden in Kapitel 4 verglichen, um Ähnlichkeiten und Unterschied zwischen der beiden Entropie zu finden. Zusätzlich wird die Gleichheit der beiden Entropie unter bestimmten Bedingungen untersucht.

2 Shannon-Entropie

In diesem Kapitel wird Definition von Shannon Entropie von diskreten und stetigen Zufallsvariablen und ihre Interpretationen und Anwendungen vorgestellt.

2.1 Definition der Shannon Entropie

Entropie von einer Zufallsvariable X ist eine Funktion von der Verteilung der Zufallsvariable X. Durch die Entropie Funktion erhaltet man ein Maß der Zufälligkeit der Verteilung von der Zufallsvariable X.

Definition 2.1. *Die Entropie einer diskreten Zufallsvariable X in Bit ist definiert als*

$$H(X) = -\sum_{x \in X} \Pr(X = x) \log_2 \Pr(X = x).$$

Somit kann die Entropie auch geschrieben werden als

$$H(X) = \mathbb{E}\Big[\log_2 \frac{1}{\Pr(X)}\Big].$$

Definition 2.2. *Die Entropie einer stetigen Zufallsvariable X ist definiert als*

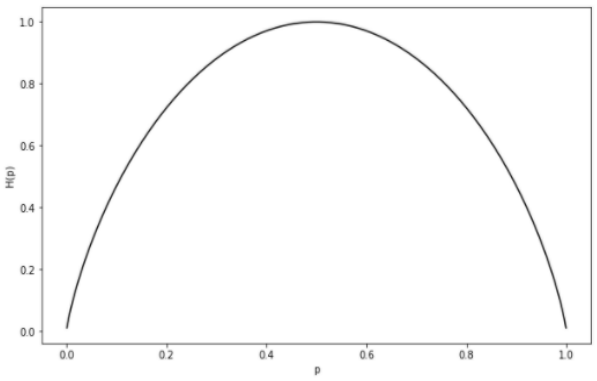
$$H(X) = -\int_{-\infty}^{\infty} \Pr(x) \log_2 \Pr(x) dx$$

Spezialfall. *Die binäre Entropie Funktion $H(p)$ für eine Zufallsvariable X, die nur zwei mögliche Ergebnisse (x_1 und $x_2 \in X$) mit der Wahrscheinlichkeit von x_1 p und von x_2 $1-p$ haben, ist*

$$\mathrm{H}(p) = -p\, log_2 p - (1-p) \log_2 (1-p)$$

Wir definieren, dass $H(0) = H(1) = 0$. Die binäre Entropie Funktion kann als folgender Graph dargestellt werden. Mit dem Graph kann man erkennen, dass die Zufälligkeit maximiert, wenn die Wahrscheinlichkeit 1/2 ist.(Siehe Figur 1.) Daher wird die Wahrscheinlichkeit 1/2 häufig in der Kryptographie verwendet, damit Hacker die encodierte Daten schwer hacken, und Zufälligkeit ist am niedrigsten, wenn Wahrscheinlichkeit 0 oder 1 ist. Die Ergebnisse können nämlich leicht vorhergesagt werden.[3] Ein einfaches Beispiel ist wie folgt.

Beispiel 2.3. *Sei $k \in \{0,1\}^2$ einen Schlüssel für eine verschlüsselte Nachricht $m \in \{0,1\}^b$, wobei a, b natürliche Zahlen sind. Der Schlüssel k ist zufällig generiert für m. Ein Angreifer(Hacker) will den Schlüssel herausfinden, um die Nachricht zu sehen, und er wusste nicht die Länge a der Schlüssel k. Die Wahrscheinlichkeit, dass der Angreifer den Schlüssel herausfinden kann, ist 1/4, wenn die Wahrscheinlichkeit 1/2 ist. Wenn die Wahrscheinlichkeit aber nicht 1/2 ist, wird die Chance höher, dass der Angreifer den Schlüssel erhalten kann. In diesem Beispiel ist die Länge des Schlüssels zu kurz. Die Länge der Schlüssel in der Kryptographie sehr lang. Die Länge der Bits wird mehr als 40 Bits. Daher ist es sehr schwer, Schlüssel herauszufinden.*



Figur 1: Binäre Entropie Funktion

Warum ist dieser Spezialfall wichtig für Shannon Entropie? Dafür gibt es einen folgenden Grund: Das Ziel des Shannon Theorem ist gegen Datenverlust bei der Übertragung über einen Kanal Daten zu schützen. Die Wahrscheinlichkeit eines Bits, das wegen der Fehler geändert wird, ist p oder $1-p$. Bit ist nämlich Binärzahl. D.h. Ein Bit kann nur den Wert '0' oder '1' haben. Daher beruhen Sätze und Theorems in diesem Kapitel auf binäre Entropie Funktion.

Beispiel 2.4. *Sei X das Ergebnis eines fairen sechsseitigen Würfel. Somit gilt $X \in \{1, 2, \ldots, 6\}$ und $P(X = i) = 1/6$ für alle $i \in \{1, 2, \ldots, 6\}$. Folglich kann die Entropie von X bestimmt werden als:*

$$\begin{aligned} H(X) &= -\sum_{i=1}^{6} \Pr(X_i = x_i) \log_2 \Pr(X_i = x_i) \\ &= -\frac{1}{6}\log_2\frac{1}{6} - \frac{1}{6}\log_2\frac{1}{6} - \frac{1}{6}\log_2\frac{1}{6} - \frac{1}{6}\log_2\frac{1}{6} - \frac{1}{6}\log_2\frac{1}{6} - \frac{1}{6}\log_2\frac{1}{6} \\ &= -6\Big(\frac{1}{6}\log_2\frac{1}{6}\Big) \\ &= \log_2 6 \approx 2.6 \end{aligned}$$

Sei Y das Ergebnis eines unfairen sechsseitigen Würfel. $Y \in \{1, 2, ..., 6\}$ und $\Pr(Y = i) = \frac{i}{21}$ für alle $i \in \{1, 2, ...6\}$. Folglich kann die Entropie von Y bestimmt werden als:

$$\begin{aligned} \mathrm{H}(Y) &= -\sum_{i=1}^{6} \Pr(Y_i = y_i) \log_2 \Pr(Y_i = y_i) \\ &= -\sum_{i=1}^{6} \frac{i}{21}\log_2\frac{i}{21} \approx 2.4. \end{aligned}$$

Im verglichen zu zwei obigen Beispiel ist die Entropie von X größer als die Entropie von Y. In anderen Wörtern erzeugt die Zufallsvariable X mehr Zufälligkeit als die Zufallsvariable Y. Entropie ist nämlich ein Maß für Zufälligkeit.

2.2 Informationsgehalt

Wenn Sie die Nachrichten im Fernsehen sehen, erhalten Sie eine Wettervorhersage. Viele Leute schauen sich die Wettervorhersage an und überarbeiten ihre Pläne für den Tag. So arbeiten die Leute hier und da, um Informationen zu erhalten. Wie kann man also feststellen, ob die Informationen für sich nützlich sind oder nicht? In den meisten Fällen sind die Informationen wertvoll, wenn ein wirtschaftlicher Nutzen besteht. Der Wert von Informationen in der Informatik ist jedoch etwas anders. In der Informatik bestimmt die Wahrscheinlichkeit des Auftretens der Information den Wert der Information.

Definition 2.5. *Sei x ein Ereignis und die Wahrscheinlichkeit von x $\Pr(X = x) = p$. Dann ist das Informationsgehalt $I(x)$ definiert als:*

$$I(x) = \log \frac{1}{\Pr(X = x)}$$

Beispiel 2.6. *Menschen sehen morgens Wetterdienst im Sommer. Die Wahrscheinlichkeit, morgen sonnig zu sein, $p(sonnig)$ ist 2/3, und die Wahrscheinlichkeit, morgen zu regnen, $p(regnen)$ ist 1/3. Dann ist das Informationsgehalt $I(sonnig)$ ist $\log 3/2$, und die Informationsgehalt $I(regnen)$ ist $\log 3$. Das Informationsgehalt $I(sonnig)$ ist also kleiner als das Informationsgehalt $I(regnen)$. Das heißt, dass $I(regnen)$ wertvollere Information als $I(sonnig)$ in der Informatik ist.*

Mit der Definition von Informationsgehalt können wir die Entropie in anderen Wörtern beschreiben, dass Entropie einer Zufallsvariable X der Erwartungswert von Informationsgehalt der Zufallsvariable X ist.

2.3 Eigenschaften von Shannon Entropie

Lemma 2.7. *Es seien X_1 und X_2 unabhängige Zufallsvariablen, und $Y = (X_1, X_2)$.*
Dann

$$H(Y) = H(X_1) + H(X_2)$$

Beweis. *Die Zufallsvariable Y ist kartesische Produkt von den Zufallsvariablen X_1, X_2 und die Zufallsvariablen X_1, X_2 sind unabhängig. Somit gilt $\Pr(Y = y) = \Pr(X_1 = x_1) \times \Pr(X_2 = x_2)$ für alle $y = (x_1, x_2) \in Y$, $x_1 \in X_1$ und $x_2 \in X_2$. Wir wenden danach folgenden Rechenregel von Logarithmus an.*
(Rechenregel: $\log a \times b = \log a + \log b$)

um dieses Lemma zu beweisen:
$H(Y)$

$$= -\sum_{y \in Y} \Pr(Y = y) \log \Pr(Y = y)$$

$$= -\sum_{x_1 \in X_1} \sum_{x_2 \in X_2} \Pr((X_1, X_2) = (x_1, x_2)) \log \Pr((X_1, X_2) = (x_1, x_2))$$

$$= -\sum_{X_1 \in x_1} \sum_{X_2 \in x_2} \Pr(X_1 = x_1) \Pr(X_2 = x_2) \log \Pr(X_1 = x_1) \Pr(X_2 = x_2)$$

$$= -\sum_{X_1 \in x_1} \sum_{X_2 \in x_2} \Pr(X_1 = x_1) \Pr(X_2 = x_2) \{\log \Pr(X_1 = x_1) + \log \Pr(X_2 = x_2)\}$$

Angewandt: Definition der Shannon Entropie, Unabhängigkeit von X_1 und X_2 und Rechenregel der logarithmischen Funktionen.

$$= -\sum_{X_1 \in x_1} \sum_{X_2 \in x_2} \Pr(X_1 = x_1) \Pr(X_2 = x_2) \log \Pr(X_1 = x_1)$$
$$+ \Pr(X_1 = x_1) \Pr(X_2 = x_2) \log \Pr(X_2 = x_2)$$

Angewandt: distributivgesetz

$$= -\sum_{X_1 \in x_1} \Pr(X_1 = x_1) \log \Pr(X_1 = x_1) \sum_{X_2 \in x_2} \Pr(X_2 = x_2)$$
$$-\sum_{X_2 \in x_2} \Pr(X_2 = x_2) \log \Pr(X_2 = x_2) \sum_{x_1 \in X_1} \Pr(X_1 = x_1)$$

Angewandt: $\sum_{X \in x} \Pr(X = x) = 1$

$$= -\sum_{x_1 \in X_1} \Pr(X_1 = x_1) \log \Pr(X_1 = x_1) - \sum_{x_2 \in X_2} \Pr(X_2 = x_2) \log \Pr(X_2 = x_2)$$
$$= \mathrm{H}(X_1) + \mathrm{H}(X_2)$$

Lemma 2.8.
Die binäre Entropie Funktion $\mathrm{H}(p)$ *hat folgende Eigenschaften:*

$$\max \mathrm{H}(p) = 1, \text{ wenn } p = \frac{1}{2}$$

Beweis.
Binäre Entropie Funktion $H(p)$ *ist* $-plog_2p - (1-p)\log_2(1-p)$ *für* $p \in [0,1]$. *Wir leiten die binäre Entropie Funktion* $H(p)$ *nach* p *ab. Dann bekommen wir folgende Gleichung:*

$\frac{dH(p)}{dp} s$

$$= -log_2p - \frac{p}{p \ln 2} + \log_2(1-p) + \frac{1-p}{(1-p)\ln 2} \mid \textit{Kettenregel}$$

$$= -\log_2 p + \log_2(1-p) + \frac{-p+p^2+p-p^2}{p(1-p)\ln 2} \mid \textit{Die Zaehler } p+p^2+p-p^2 \textit{ ist } 0.$$

$$= -\log_2 p + log_2(1-p)$$

$$= log_2 \frac{1-p}{p} \mid \textit{Rechenregeln von Logarithmus}$$

Die Ableitung $\frac{d\,\mathrm{H}(p)}{dp}$ *ist* 0*, wenn* $p = \frac{1}{2}$*. D.h. Der Punkt* $p = 1/2$ *ist eine Extremstelle von der binären Entropie Funktion* $\mathrm{H}(p)$*. Ich zeige nun, dass die Extremstelle eine Maximalstelle ist. Betrachte Werte von der Ableitung für* $p < \frac{1}{2}$ *und* $p > \frac{1}{2}$*, ob die Werte positiv sind. Die Werte von der Ableitung sind für* $p < \frac{1}{2}$ *positiv und für* $p > \frac{1}{2}$ *negativ. D.h Werte von der binären Entropie Funktion* $\mathrm{H}(p)$ *steigen bis der Punkt* $p = 1/2$ *und sinken nach dem Punkt. Daher ist* $\mathrm{H}(\frac{1}{2})$ *ein Maximum.*

2.4 Interpretationen von Shannon Entropie

Angenommen, eine faire Münze wird geworfen. Dann sind nur zwei mögliche Ergebnisse Kopf oder Zahl. Dies kann so codiert werden, dass Kopf '0' und die Zahl beispielsweise repräsentieren. '010' bedeutet deshalb 'Kopf', 'Zahl' und 'Kopf', wenn wir die faire Münze dreimal werfen. Betrachte nun Entropie des fairen Münzwurfs. Mithilfe der binären Entropie Funktion kann die Entropie leicht berechnet werden, und zwar $\log_2 2 = 1$. Das zeigt auf, dass Entropie durchschnittliche Anzahl der Bits beim Codieren einer Information ist. Das obige Beispiel erfüllt folgende Definition.

Definition 2.9. *Sei* $|y|$ *die Zahl von Bits in einer Sequenz von Bits* y*. Eine Extraktion Funktion* Ext *nimmt den Wert von einer Zufallsvariable* X *als Eingabe und gibt so eine Sequenz von Bits* y *aus, dass*

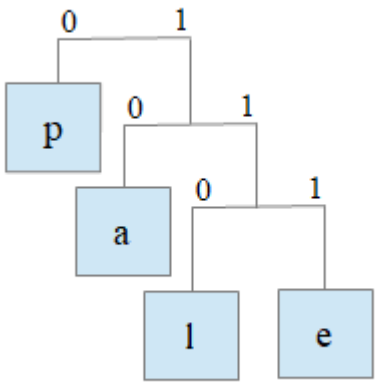
$$\Pr(\mathrm{Ext}(X) = y \mid |y| = k) = \left(\frac{1}{2}\right)^k$$

Beispiel 2.10. *Sei* X *eine Zufallsvariable mit* $\{1,2,3,4,...,9\} \in X$*, und sei* $\Pr(X = k) = 1/9$ *für alle* $k = \{1,2,3,4...,9\}$*. Dann kann die Zufallsvariable* X *so codiert werden: Die Anzahl der Elemente der Zufallsvariable* X *ist großer als* $2^3 = 8$*. Daher sind mindestens 3 Bits für die Codierung der Zufallsvariable* X *erforderlich. Das Elemente '9' kann 2 Bits zuordnet werden. Siehe folgende Tabelle.*

Eingabe	1	2	3	4	5	6	7	8	9
Ausgabe	000	001	010	011	100	101	110	111	00

Tabelle 1: Beispiel von Extraktion

Die obige Tabelle 1 ist perfekt geeignet zu der Zufallsvariable zu codieren.



Figur 2: Darstellung des Huffman Algorithms für das Beispiel 2.11

Es gibt einfachere Methode, die obige Zufallsvariable zu codieren. Betrachte die Anzahl der Elemente der Zufallsvariable. In diesem Fall ist die Anzahl der Elemente der Zufallsvariable 9. 9 liegt zwischen 2^3 und 2^4. Dann kann die Zufallsvariable so codiert werden, wenn die Wahrscheinlichkeit nicht berücksichtigt wird. Wenn die obige Methode verwendet wird, wird die Länge der codierten Daten länger. Man kann es noch kürzer mithilfe 'Kompression'. Es gibt eine häufig verwendete Methode, um Daten zu komprimieren, und zwar Huffman Algorithmus.[2]. Der Schlüssel zu Huffman Algorithmus besteht darin, Ereignissen mit hoher Wahrscheinlichkeit kleine Bits und Ereignissen mit niedriger Wahrscheinlichkeit große Bits zuzuweisen.

Beispiel 2.11. *Eine Nachricht wird beispielsweise 'apple' codiert. In diesem Word ist 'p' am häufigsten mit der Wahrscheinlichkeit* $2/5$ *verwendet. Daher ist p '0' zuzuweisen, und die andere Schriften haben gleiche Wahrscheinlichkeit, und zwar* $1/5$*. Daher wird '10' für 'a', '110' für 'l' und '111' für 'e' zugeordnet 2. Schließlich kann 'apple' 10(a) 0(p) 0(p) 110(l) 111(e) codiert werden. Die durchschnittliche Anzahl der Bits ist 2. Entropie von diesem Beispiel ist* $-3/5\log_2(1/5) - 2/5\log_2(2/5) \approx 1.9$*. Wenn wir den Wort 'apple' wie Beispiel codieren, ist die durchschnittliche Anzahl der Bits ist 3 Bits. Mit dem Huffman Algorithmus kann der Wort 'apple' ein Bit kürzer codiert werden.*

Charakter	a	p	l	e
Wahrscheinlichkeit	1/5	2/5	1/5	1/5

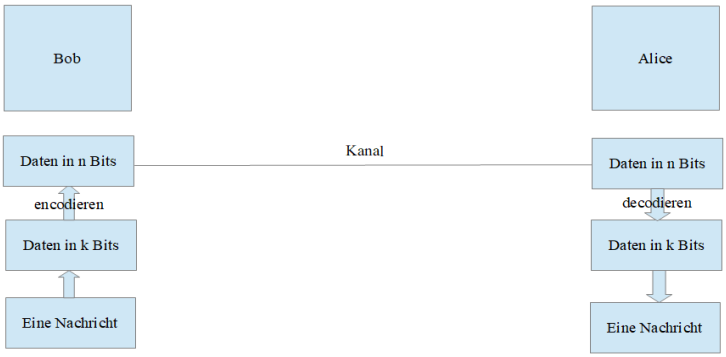
Tabelle 2: Buchstaben 'a','p','l' und 'e' mit Wahrscheinlichkeit

Diese Figur heißt Huffman tree.

2.5 Anwendungen von Shannon Entropie

In der Einleitung haben wir grob zusammengefasst, wie die Shannon Entropie. In diesem Unterkapitel wird die Anwendungen genauer erklärt. Bevor wir die Anwendungen genauer anschauen, müssen wir verstehen, wie die Daten über Kanal überträgt werden. Folgendes Beispiel zeigt gut auf Datensendeschema in Figur 3.

Bob sende eine Nachricht an Alice. Die Nachricht wird in $m \in \{0,1\}^k$ Bits



Figur 3: Datensendeschema

codiert. Wenn wir die Daten durch einen Kanal senden, erhält Empfänger hoch wahrscheinlich nicht richtige Nachricht, weil es einen Fehler bei der Übertragung der Daten vorkommen kann. Um dies zu vermeiden, wird Redundanz der Daten in der Informatik verwendet. Die einfachste Redundanz der Daten ist Wiederholung der Daten. Das ist aber nicht effektiv, weil die Daten zu länger werden kann. So wenden Informatiker folgende Encodierung und Decodierung Funktionen an.

Definition 2.12. *$A(k,n)$ Encodierung Funktion $Enc : \{0,1\}^k \to \{0,1\}^n$ nimmt eine Sequenz von k Bits als Eingabe und gibt eine Sequenz von n Bits. $A(k,n)$ Decodierung Funktion $Dec : \{0,1\}^n \to \{0,1\}^k$ nimmt eine Sequenz von n Bits als Eingabe und gibt eine Sequenz von k Bits.*

Mit der Encodierung Funktion Enc werden die Daten encodiert $Enc(m) \in \{0,1\}^n$ und über Kanal übertragen. Bei der Übertragung der Daten kann ein Fehler mit der Wahrscheinlichkeit p vorkommen, wenn wir annehmen, dass der Kanal binär ist. D.h. Jede Bit kann bei der Übertragung wegen eines Fehlers verändert werden. Beispielsweise wird eine Nachricht '00000' mit der Wahrscheinlichkeit p zu '00001' geändert. Empfänger bekommt wegen des Fehlers veränderte Daten $X \in \{0,1\}^n$ und lässt die Daten decodieren $Dec(X) \in \{0,1\}^k$. Mit der Wahrscheinlichkeit $1-\gamma$, dass γ vorgewählte Konstante ist, erhält der Empfänger richtige Daten. Wenn es keinen Fehler bei der Übertragung

der Daten gibt, kann die originale Nachricht gesendet werden. Ansonsten kann der Sender nur $k = n(1 - H(p)$ Bits Nachricht innerhalb jedes Blockes von n Bits schicken. Diese Länge der Daten ergibt aus Shannon Theorem.

Theorem 2.13. *Für einen binären Kanal mit Parameter $p < 1/2$ und für alle Konstante $\gamma, \delta > 0$, wenn n ausreichend groß ist.*

1. *Für alle $k \leq n(1 - H(p) - \delta)$ existiert es (k,n) Encodierung und Decodierung Funktion, sodass die Wahrscheinlichkeit, dass der Empfänger richtige Nachricht nicht bekommt, ist höchstens γ für jede mögliche Nachricht, der Länge k Bits ist.*

2. *Es gibt keine (k, n) Encodierung und Decodierung Funktion mit $k \geq n(1 - H(p) + \delta)$, sodass die Wahrscheinlichkeit, richtig zu decodieren, ist mindestens γ für eine Nachricht, die Länge k Bits ist und gleichmäßig zufällig gewählt wird.*

Beweis. *Zuerst wird Existenz der geeignete (k,n) Encodierung- und Decodierung-Funktionen, falls $k \leq n(1 - H(p) - \delta)$ durch probabilistische Methode. Am Ende wird die Encodierung- und Decodierung-Funktionen maximal die Wahrscheinlichkeit γ für alle mögliche Eingabe zu besitzen erwünscht.[8] Den Beweis von Shannon Theorem finden Sie [8].*

3 metrischer Entropie

Es gibt verschiedene Arten von Entropie: Entropie in der Physik, Entropie in der Informatik und metrische Entropie in der Statistik. Verschiedene Arten von Entropie wurde klassifiziert, aber ihre Wurzeln sind dieselben. Die Entropie in verschiedenen Bereichen ist in einer untrennbaren Beziehung miteinander verbunden. In diesem Kapitel wird die metrische Entropie vorgestellt. Wie bereits erwähnt, wurde das Phänomen der Entropie in der Physik beobachtet und die Entropie später mathematisch definiert. Basierend auf dieser Theorie wurde dann untersucht, wie Daten in der Informatik sicher übertragen werden können. Die mathematische Entropie ist das Bindeglied zwischen Entropie in der Physik und Entropie in der Informatik. Wie ist diese metrische Entropie definiert? Bevor wir uns die Definition der metrischen Entropie ansehen, müssen wir die Überdeckungszahl und die Verpackungszahl kennenlernen.

3.1 Definition der metrischen Entropie

Definition 3.1. *Betrachten einen positiven Wert $r \in (0, \infty)$, zwei Menge $\mathcal{S}$ und $\mathcal{A}$ mit $\mathcal{S} \subset \mathcal{A}$, und eine Funktion $d : \mathcal{A} \times \mathcal{A} \to [0, \infty]$. Wir definieren zwei Zahlen $\mathrm{cov}[r, \mathcal{S}, \mathcal{A}, d] \in \{1, ..., \infty\}$ als folgendes:*

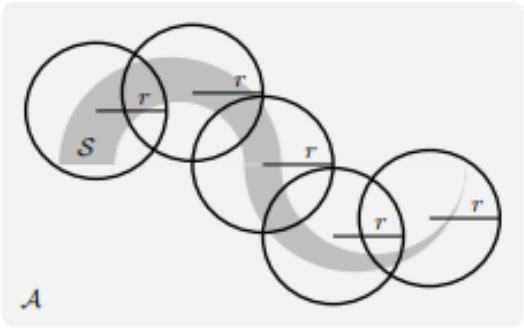
1. *Falls $\mathcal{S} = \emptyset$, setze $\mathrm{cov}[r, \mathcal{S}, \mathcal{A}, d] := pack[r, \mathcal{S}, d] := 1$*

2. *Falls* $\mathcal{S} \neq \emptyset$, *setze* $\mathrm{cov}[r, \mathcal{S}, \mathcal{A}, d] := \min\Big\{m \in \{1,2,...\} : \exists s_1, ..., s_m \in \mathcal{A} :$
$\sup_{s\in\mathcal{S}} \min_{i\in\{1,...,m\}} \max\big\{d[s_i, s], d[s, s_i]\big\} \leq r\Big\}$
Falls das Minimum existiert und ansonsten $\mathrm{cov}[r, \mathcal{S}, \mathcal{A}, d] = \infty$
$pack[r, \mathcal{S}, d] := \min\Big\{m \in \{1,2,...\} :$
$\exists s_1, ..., s_m \in \mathcal{A} : \min_{i,j\in\{1,...,m\}} \max\big\{d[s_i, s_j], d[s_j, s_i]\big\} > r\Big\}$
Falls das Maximum existiert und ansonsten $pack[r, \mathcal{S}, d] = \infty$.

Wir nennen $\mathrm{cov}[r, \mathcal{S}, \mathcal{A}, d]$ *Überdeckungszahl und* $pack[r, \mathcal{S}, d]$ *Verpackungszahl. Wir definieren*

$$ent[r, \mathcal{S}, \mathcal{A}, d] := \log\big[\mathrm{cov}[r, \mathcal{S}, \mathcal{A}, d]\big]$$

mit der Konvention $\log[\infty] := \infty$. *Wir nennen* $ent[r, \mathcal{S}, \mathcal{A}, d]$ *die Entropie Zahl oder die Entropie.*



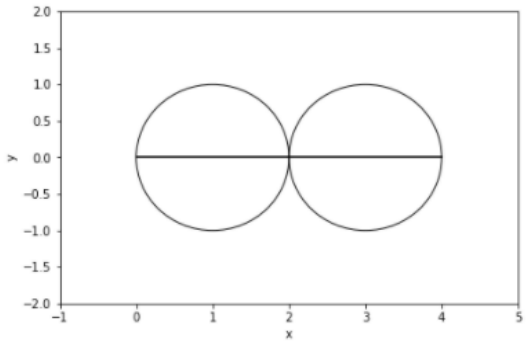
Figur 4: Überdeckungszahl von $\mathcal{S}$ mit Radius r
[5]

Beispiel 3.2. *Sei* $\mathcal{S} = (x, y)$ *für* $x \in [0, 4]$ *und* $y = 0$, *und sei* $\mathcal{A} = \mathbb{R}^2$.
Wir wählen Radius $r = 1$ *und die Distanzfunktion* $d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ *(euklidische Norm) für alle* $(x_1, y_1), (x_2, y_2) \in \mathcal{A}$. *Um die Menge* $\mathcal{S}$ *zu überdecken, sind mindestens zwei Kreise erforderlich. Ein Kreis davon hat den Punkt (1,0) als Kreismittelpunkt mit Radius '1', und der andere Kreis hat den Punkt (3,0) als Kreismittelpunkt mit Radius '1'.*

$$\mathrm{cov}[1, ([0,4], 0), \mathbb{R}^2, d] := \min\Big\{m \in \{1,2,...\} : \exists s_1, ..., s_m \in ([0,4], 0) :$$

$$\sup_{s\in\mathcal{S}} \min_{i\in\{1,...,m\}} \max\big\{d[s_i, s], d[s, s_i]\big\} \leq 1\Big\} \tag{1}$$

Für die Punkte $s_1 = (1, 0)$ *und* $s_2 = (3, 0)$ *gelten* $\sup_{s\in\mathcal{S}} \min_{i\in\{1,2\}} \max\big\{d[s_i, s], d[s, s_i]\big\} \leq$ 1. *Daher ist Überdeckungszahl von* $\mathcal{S}$ *2, und Entropie ist* $\log 2$. *Siehe Figur 5*



Figur 5: Überdeckungszahl von $\mathcal{S}$ in dem Beispiel 3.2

3.2 Eigenschaften der metrischen Entropie

In diesem Kapitel wird die Eigenschaften der Überdeckungszahl und Verpackungszahl und ihre Beziehung vorgestellt.

Lemma 3.3 (Monotonie von Überdeckungszahl und Verpackungszahl)**.** *Für jede Paar von Radius $r_1, r_2 \in (0, \infty)$ mit $r_1 \geq r_2$ und für jede Menge $\mathcal{S}$, Umgebungsmenge $\mathcal{A}$, und Distanzfunktion d, dann gilt :*

$$\text{cov}[r_1, \mathcal{S}, \mathcal{A}, d] \leq \text{cov}[r_2, \mathcal{S}, \mathcal{A}, d] \;\; \textit{und} \;\; pack[r_1, \mathcal{S}, d] \leq pack[r_2, \mathcal{S}, d]$$

Beweis. *Die Überdeckungszahl ist die Anzahl der Kreise, die eine Menge komplett überdecken. Wenn der Radius der Kreise zunimmt, nimmt die Anzahl der Kreise ab, die zum Überdecken einer Menge erforderlich sind. Dies gilt auch für die Verpackungszahl. Sei $\mathcal{S}$ eine Menge auf dem Umgebungsraum $\mathcal{A}$ und sei $d = B(r)$ eine Distanzfunktion. Für Radien $r_1, r_2 \in \mathbb{R}$ mit $r_1 < r_2$ gilt : $|B(r_1)| < |B(r_2)|$. Laut Definition* **??** *gelten $\mathcal{S} \subset \bigcup_{i=1}^{m} B(r_1)$ und $\mathcal{S} \subset \bigcup_{i=1}^{n} B(r_2)$ für minimale Anzahl des Balls $m, n \in \mathbb{N}$.*

Wir nutzen Kontratiktion aus, um folgende Lemmata zu beweisen.

Lemma 3.4 (Äquivalenz der Überdeckungszahl und Verpackungszahl)**.** *Für jede Radius $r \in (0, \infty)$, und für jede Menge $\mathcal{S}$, Umgebungsmenge $\mathcal{A}$, und Distanzfunktion d, dann gilt :*

$$\text{cov}[r, \mathcal{S}, \mathcal{A}, d] \leq pack[r, \mathcal{S}, d]$$

Beweis. *Wir möchten zeigen, dass die Überdeckungszahl gleich oder klein als die Verpackungszahl. In dem Fall, dass die Menge $\mathcal{S}$ leer ist, ist die Überdeckungszahl 1 sowohl als auch die Verpackungszahl. Daher gilt trivialerweise die Ungleichheit. Wenn die Menge $\mathcal{S}$ nicht leer ist, können wir annehmen, dass die*

Verpackungszahl $pack[r, \mathcal{S}, \mathcal{A}]$ *eine endliche Zahl* m *ist. Es gibt denn* $s_1, ..., s_m \in \mathcal{S}$*, sodass*

$$\max\left\{d[s_i, s_j], d[s_j, s_i]\right\} > r \text{ für alle } i, j \in \{1, 2, ..., m\}$$

, falls i *und* j *nicht gleich sind. Wir verwenden jetzt Kontradiktion: angenommen, dass die Überdeckungszahl ist größer als die Verpackungszahl ist. Dann es gibt ein* $s \in \mathcal{S}$ *wegen der Definition der Überdeckungszahl, sodass*

$$\min_{i \in \{1,...,m\}} \max\left\{d[s_i, s_j], d[s_j, s_i]\right\} > r$$

Dann hat die Menge $\mathcal{S}$ $m+1$ *Elemente, die obige Ungleichung erfüllt sind. Das zeigt auf, dass die Verpackungszahl* $m+1$ *ist. Das Kontradiktiert, dass die Verpackungszahl* m *ist, die vorher angenommen wurde. Folglich ist die Überdeckungszahl gleich wie oder kleiner als die Verpackungszahl.*

Lemma 3.5. *Für jeden Radius* $r \in (0, \infty)$*, und für jede Menge* $\mathcal{S}$*, Umgebungsmenge* $\mathcal{A}$*, und Distanzfunktion* d*, die Dreiecksungleichung erfüllt. Dann gilt:*

$$pack[r, \mathcal{S}, d] \leq \text{cov}[r/2, \mathcal{S}, \mathcal{A}, d]$$

Beweis. *Zu zeigen, dass die Verpackungszahl* $pack[r, \mathcal{S}, d]$ *ist gleich wie oder kleiner als die Überdeckungszahl* $\text{cov}[r/2, \mathcal{S}, \mathcal{A}, d]$*. Es wird angenommen, dass die Menge* $\mathcal{S}$ *nicht leer ist und die Überdeckungszahl eine endliche Zahl* m *ist. Dann gibt es* $s_1, ..., s_m \in \mathcal{A}$ *wegen der Definition der Überdeckungszahl, sodass*

$$\sup_{s \in \mathcal{S}} \min_{i \in \{1,...,m\}} \max\left\{d[s_i, s], d[s, s_i]\right\} \leq \frac{r}{2}$$

Angenommen, dass $pack[r, \mathcal{S}, d] > \text{cov}[r/2, \mathcal{S}, \mathcal{A}, d]$*. Dann gibt es* $\bar{s_1}, ..., \bar{s_{m+1}} \in \mathcal{S}$*, sodass*

$$\max\left\{d[\bar{s_i}, \bar{s_j}], d[\bar{s_j}, \bar{s_j}]\right\} > r$$

für alle $i, j \in \{1, ..., m+1\}$*, wobei* $i \neq j$*. Wegen der Definition der Entropie und des pigeonhole Prinzips muss ein* $i \in \{1, ..., m\}$ *und ein paar* $j, k \in \{1, ..., m+1\}$*, sodass* $d[\bar{s_j}, s_o] \leq r/2$ *und* $d[s_i, \bar{s_k}] \leq r/2$*. Für die Distanzfunktion* d *gilt die Dreiecksungleichung, sodass*

$$\begin{aligned} d[\bar{s_j}, \bar{s_k}] &\leq d[\bar{s_j}, s_i] + d[s_i, \bar{s_k}] \\ &\leq \frac{r}{2} + \frac{r}{2} \\ &\leq r \end{aligned}$$

Auf diese Weise gilt die Ungleichung $d[\bar{s_k}, \bar{s_j}] \leq r$*. Dies kontradiktiert die Annahme.*

3.3 Interpretation der metrischen Entropie und deren Anwendungen

In der Mathematik wird die metrische Entropie verwendet, obere Schranke für Mengen zu finden. [4] Vor allem in der Statistik findet man das Supremum von empirische Prozesse, metrische Entropie anzuwenden. Das Supremum von empirischen Prozessen, die unendliche oder sogar unzählige Mengen übernehmen, wird kontrolliert. Wir wenden eine Technik 'Chaining' an. Mit der Technik kann das Problem, Maximum über endliche Mengen zu begrenzen, minimieren. [5]. Bevor Wir uns mit diesem Thema befassen, müssen wir Definition von empirischen Prozessen wissen.

Definition 3.6. *Sei* $(\mathcal{A}, \mathfrak{A}, \Pr)$ *ein Wahrscheinlichkeitsraum,* $(\mathcal{B}, \mathfrak{B})$ *ein messbarer Raum, und* $\mathbb{R}$ *mit Borel* σ*-Algebra. Betrachte Zufallsvariablen*

$$X_1, X_2, ... : (\mathcal{A}, \mathfrak{A}, \Pr) \to (\mathcal{B}, \mathfrak{B})$$

eine Menge $\mathcal{F}$ *der messbaren Funktionen* $f : (\mathcal{B}, \mathfrak{B}) \to \mathbb{R}$*, Stichprobegröße* $n \in \{1, 2, 3, ...\}$*, und die Zufallsvariablen*

$$\Pr_n f := \frac{1}{n} \sum_{i=1}^{n} f[X_i] : (\mathcal{A}, \mathfrak{A}, \Pr) \to \mathbb{R}$$

definiert für $f \in \mathcal{F}$*. Wir nennen die Menge von Zufallsvariablen* $\{\Pr_n f : f \in \mathcal{F}\}$ *als empirische Prozesse indiziert bei* $\mathcal{F}$*.*

Wir diskutieren nun Orlicz Normen. Orlicz Normen werden verwendet, maximale Ungleichung zu formulieren.[5]

Definition 3.7 (Young Funktion und Orlicz Normen)**.**
Eine Funktion $f : [0, \infty) \to [0, \infty)$ *ist eine Young Funktion, wenn die Funktion* f *folgende Bedingungen erfüllt.*

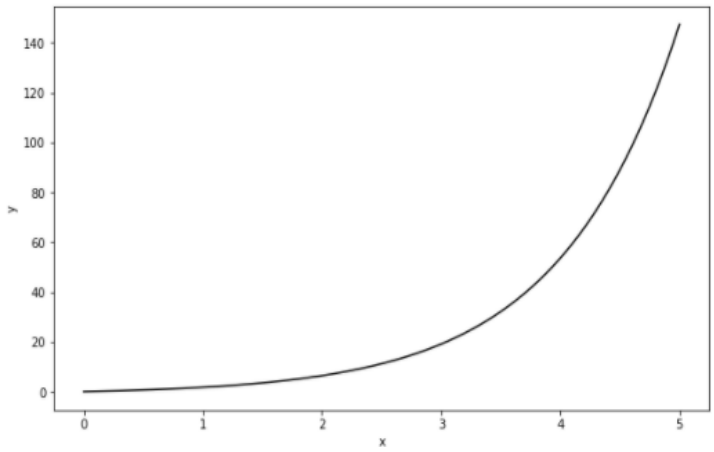
1. f *ist konvex*
2. f *ist nicht fallend.*
3. $f[0] = 0$
4. $\lim_{a \to \infty} f[a] = \infty$

Sei f *eine Young Funktion und* X *eine reellwertige integrierbare Zufallsvariable. Dann,*

$$||X||_f := \inf \Big\{ a \in [0, \infty) : \mathbb{E}\Big[\big[f \frac{|X|}{a} \big] \Big] \leq 1 \Big\}$$

ist Orlicz Norm von X *bezüglich* f*.*

Beispiel 3.8. *In diesem Beispiel wird es gezeigt, dass eine Funktion* $f_q : a \mapsto e^{aq} - 1$ *eine Young Funktion für alle* $q \in [1, \infty)$ *ist.*



Figur 6: exp(x)-1

1. $\exp(a^q) - 1$ *ist konvex. Siehe Figur 6.*
2. *zu zeigen, dass* $\exp(a^q) - 1$ *nicht fallend ist. Betrachte die Ableitung von* $\exp(a^q) - 1$ *nach a. Die Ableitung von* $\exp(a^q) - 1$ *ist* $qa^{q-1}exp(a^q)$*, und die Ableitung von a für* $a \in [0, \infty]$ *ist nonnegativ. Somit ist die Funktion* $\exp(a^q) - 1$ *nicht fallend.*
3. $\exp(0^q) - 1 = 0$.
4. $\lim_{a\to\infty} \exp(a^q) - 1 = \infty$.

   Daher ist die Funktion $\exp(a^q) - 1$ *eine Young Funktion.*

Letzte Teil dafür, obere Schranke(Supremum) einer Menge zu finden, ist maximale Ungleichung. Das wichtigste Ziel von Theorie der empirischen Prozesse ist maximale Ungleichung zu Ungleichung für das Supremum über empirische Prozesse zu verallgemeinern.[5]

Lemma 3.9 (maximale Ungleichung)**.**
Sei n eine reelle Zahl, und seien Zufallsvariablen $X_1, ..., x_n$ *auf demselben Wahrscheinlichkeitsraum integrierbar. Für* $\mathcal{L}_q Norm$ *gilt:*

$$||\max X_i||_q \leq n^{1/q} \max_{i\in\{1,...,n\}} ||X_i||_q$$

Beweis. *Der Beweis von diesem Lemma 3.9 finden Sie [5].*

Theorem 3.10. *Sei f eine Young Funktion, Für die Funktion gilt :*

$$\lim \sup_{a,b\to\infty} \frac{f[a][b]}{f[cab]} < \infty$$

für eine Konstante $c \in (0, \infty)$. Dann,

$$||\max_{i\in\{1,\dots,n\}} X_i||_f \leq h_f \cdot f^{-1}[n] \max_{i\in\{1,\dots,n\}} ||X_i||_f$$

für eine Konstante $h_f \in (0, \infty)$, die abhängig nur von f.

Wir haben bereits überprüft, dass metrische Entropie empirische Prozesse und eine Young Funktion ist. Um obiges Theorem für metrische Entropie anzuwenden, eine Technik ist noch erforderlich, das Supremum einer Menge mit metrischer Entropie zu finden. Die Technik ist wie bereits kurz erwähnt 'Chaining'.

Definition 3.11 (Begrenzt und total begrenzt)**.** *Sei zwei Mengen $\mathcal{S}$ und $\mathcal{A}$ mit $\mathcal{S} \subset \mathcal{A}$ und eine Funktion $d : \mathcal{A} \times \mathcal{A} \to [0, \infty]$. Die Paare $(\mathcal{S}, d)$ heißt begrenzt, falls der Durchmesser von $\mathcal{S}$ bezüglich d ist endlich :*

$$\operatorname{diam}[\mathcal{S}, d] := \sup_{s,s'\in\mathcal{S}} d[s, s'] < \infty$$

Das 3-Tupel $(\mathcal{S}, \mathcal{A}, d)$ ist total begrenzt, falls die Überdeckungszahl endlich für alle Radien $r > 0$:

$$\operatorname{cov}[r, \mathcal{S}, \mathcal{A}, d] < \infty \ \text{ für alle } \ r \in (0, \infty)$$

Definition von totaler Begrenztheit ist für Überdeckungszahl verwendet. Für Verpackungszahl wird die Definition auch angewandt, weil Überdeckungszahl und Verpackungszahl äquivalent sind.(Begründung : Lemma 3.3)[5] Für die total Begrenztheit spielt der Umgebungsraum eine wichtige Rolle[5]. Wenn Größe des Umgebungsraums ähnlich wie Größe der Menge auf dem Umgebungsraum, existiert kein Supremum für alle Radien $r \in (0, \infty)$. D.h. total Begrenztheit ist unmöglich für solchen Umgebungsraum. Schließlich können wir das Supremum einer Menge mit folgendem Theorem herausfinden:

Theorem 3.12 (Supremum von empirischen Prozessen)**.**
Sei ein total begrenztes 3 Tupel $(\mathcal{S}, \mathcal{A}, d)$, eine separable Mengen von Zufallsvariablen $\{X_s : s \in \mathcal{S},$ und ein Orlicz Norm $|| \cdot ||_f$ bezüglich einer Young Funktion f. Angenommen, dass für eine Konstante $c_1 \in (0, \infty)$

$$\limsup_{a,b\to\infty} \frac{f[a]f[b]}{f[c_1 ab]} < \infty$$

und, dass für eine Konstante $c_2 \in (0, \infty)$

$$||X_s - X_t||_f \leq c_2 d[s, t] \ \text{ für alle } \ s, t \in \mathcal{S}$$

Dann gibt es eine Konstante $c_3 \in (0, \infty)$, die abhängig nur von f und c_2, sodass

$$\Big|\Big| \sup_{s,t\in\mathcal{S}} |X_s - X_t| \Big|\Big|_f \leq c_3 \int_0^{\operatorname{diam}[\mathcal{S},d]} f^{-1}\big[\operatorname{pack}[r, \mathcal{S}, d]\big] dr$$

generalisiertes Inverses wird so verwendet :

$$f^{-1}[b] := \infty\{a \in [0, \infty) : f[a] > b\}$$

Die total Begrenztheit gibt uns Informationen, dass das Integral in dem Theorem wohldefiniert ist und der Durchmesser endlich ist. Die untere Schranke für die Integration kann mit einer kleinen Konstante ersetzbar sein.[5]

Außerdem können wir durch Entropie(genauer mit Überdeckungszahl und Verpackungszahl) Volumen der Mengen abschätzen. Betrachte eine Menge $\mathcal{S}$ in dem Umgebungsraum $\mathcal{A}$ mit Radius r und Distanzfunktion d. Dann kann die Überdeckungszahl $\text{cov}[r, \mathcal{S}, \mathcal{A}, d]$ bestimmt werden. Wie gesagt, ist die minimale Anzahl der Kreise $B(r)$ (oder Kugel), die Menge zu überdecken, die Überdeckungszahl. Jede Kreise $B(r)$ hat Volumen $Vol(B(r))$. Dann ist das Volumen von $\mathcal{S}$ $Vol(\mathcal{S})$ kleiner als Multiplikation von Überdeckungszahl und Volumen eines Kreises, also $Vol(B(r)) \times \text{cov}[r, \mathcal{S}, \mathcal{A}, d]$, wenn das Maximum existiert. Je kleiner der Radius wird, desto dichter wird die Menge $\mathcal{S}$ mit $B(r)$ gefüllt. Dann kann $Vol(B(r))$ gleich wie $\lim_{r \to 0} Vol(B(r)) \cdot \text{cov}[r, \mathcal{S}, \mathcal{A}, d]$. Eine Art der Entropie (Volumen Entropie) basiert auf die Idee, Volumen mit Überdeckungszahl abzuschätzen.

$$Vol(\mathcal{S}) \leq Vol(B(r)) \cdot \text{cov}[r, \mathcal{S}, \mathcal{A}, d]$$

Die Ungleichung gilt wegen der Definition der Überdeckungszahl. Wenn das 3-Tupel $(\mathcal{S}, \mathcal{A}, d)$ total begrenzt ist, kann die Multiplikation $Vol(B(r)) \cdot \text{cov}[r, \mathcal{S}, \mathcal{A}, d]$ mit klein genug Raidus r das Volumen $Vol(\mathcal{S})$ approximiert werden. Diese Idee kann zu Volumen Entropie entwickelt werden.

4 Gegenüberstellungen über metrische Entorpie und Shannon Entropie

In diesem Kapitel werden die Shannon Entropie und metrische Entropie verglichen, und Beziehungen zwischen metrischer Entropie und Shannon Entropie untersucht.

4.1 Gleichheit zwischen metrischen Entropie und Shannon Entropie

Unter der bestimmten Bedingungen kann metrische Entropie gleich wie Shannon Entropie sein. In diesem Kapitel ist Distanzfunktion d euklidische Norm.

Um das Maximum von Shannon Entropie zu bestimmen, wird das wichtiger Satz 'Lagrange Multiplikator' gebraucht. [2]

Satz 4.1 (Lagragne Multiplikator)**.** *Seien $f : \mathbb{R}^D \to \mathbb{R}$ und $g : \mathbb{R}^D \to \mathbb{R}^C$ stetige und differenzierbare Funktionen. Betrachte das Optimierungsproblem unter der*

Bedingung $g(x) = 0$. Eine Funktion $F : \mathbb{R}^{D+C} \to \mathbb{R}$ wird definiert:

$$F(x, y) = f(x) + y \cdot g(x).$$

Sei $(a, b) \in \mathbb{R}^{D+C}$ ein stationärer Punkt von F. Dann sind folgende Aussagen wahr.

1. *Falls $(a, b) \in \mathbb{R}^{D+C}$ ein stationärer Punkt von F ist, dann ist a ein stationärer Punkt von f unter $g(a) = 0$.*
2. *Falls $a \in \mathbb{R}^D$ ein stationärer Punkt von f unter $g(a) = 0$, Dann existiert ein stationärer Punkt $b \in \mathbb{R}^C$ von F.*

Die Konstante b heißt Lagrange Multiplikator.

In der Optimierungstheorie finden wir das lokale Extremum. das Extremum ist eine Teilmenge der stationären Punkte, alle stationäre Punkte müssen gefunden werden und untersucht werden, ob es sich um einem Extremum handelt. Der Lagrange-Multiplikator ist eine Möglichkeit, eingeschränkte Optimierungsprobleme zu lösen. Wir verwandeln ein eingeschränktes Problem in ein nicht eingeschränktes Problem, indem wir dem Wert, den wir optimieren möchten, einen formalen Lagrange-Multiplikator-Term hinzufügen.[2]

4.1.1 das Maximum der Shannon Entropie

Lemma 4.2. *Diskrete Zufallsvariablen maximiert, wenn die Zufallsvariablen gleich verteilt ist.*

Beweis. *Bei dem Fall binäre Entropie Funktion kann man leicht es erkennen, dass obige Aussage wahr ist. Siehe Beispiel 1. Berechnen nun das Maximum der Entropie Funktion einer Zufallsvariable X, die n-Elemente besitzt: Sei X die Zufallsvariable mit $x_i \in X$ für $i \in \{1, ..., n\}$ und sei $\Pr(X = x_i) = 1/n$. Dann ist die Entropie Funktion $H(X) = -\sum_{i=1}^{n} \Pr(X = x_i) \log_2 \Pr(X = x_i)$. Das Maximum der Entropie Funktion kann mithilfe der Lagrange- Multiplikator herausfinden[6]. Siehe folgende Gleichung:*

$$\hat{H}(X) = -\sum_{i=1}^{n} \Pr(X = x_i) \log_2 \Pr(X = x_i) - \lambda\Big(\sum_{i=1}^{n} p(X = x_i) - 1\Big)$$

Die Gleichung wird nach $\Pr(X = x_i)$ abgeleitet, um die Extremstelle zu finden.

$$\frac{d\hat{H}}{d \Pr(X = x_i)} = -\log_2 \Pr(X = x_i) - 1 - \lambda$$

Die obige abgeleitete Funktion wird 0, wenn $\Pr(X = x_i) = \frac{1}{n}$. In dem Punkt $\Pr(X = x_i) = \frac{1}{n}$ maximiert der Wert der Shannon Entropie.

Lemma 4.3. *Shannon Entropie einer stetigen Zufallsvariable maximiert, wenn die Zufallsvariable normal verteilt ist.*

Beweis. *Das Maximum der steigen Zufallsvariable kann auch durch Lagrange Multiplikator berechnet werden. Es gibt folgende Beschränkungen, um das Maximum zu finden.*
Sei X eine stetige Zufallsvariable und μ der Erwartungswert von X und σ^2 die Varianz von X.

1. $\int_{-\infty}^{\infty} \Pr(x)dx = 1$
2. $\int_{-\infty}^{\infty} x \Pr(x)dx = \mu$
3. $\int_{-\infty}^{\infty} (x-\mu)^2 \Pr(x)dx = \sigma^2$

Wende nun die Lagrange Multiplikator an. Dann erhält man:

$$\frac{dH(X)}{dx} = -\int_{-\infty}^{\infty} \Pr(x)\log_2 \Pr(x) + \lambda_1\Big(\int_{-\infty}^{\infty} \Pr(x)dx - 1\Big) + \lambda_2\Big(\int_{-\infty}^{\infty} x\Pr(x)dx - \mu\Big) + \lambda_3\Big(\int_{-\infty}^{\infty} (x-\mu)^2\Pr(x)dx - \sigma^2\Big)$$

Setze die Ableitung gleich wie 0, um die Extremstelle zu finden. Lege Pr *auf der linken Seite und die andere Faktoren auf der rechten Seite. Dann bekommt man*

$$\Pr(x) = \exp\Big\{-1 + \lambda_1 + \lambda_2 x + \lambda_3(x-\mu^2)\Big\}. \tag{2}$$

Die Gleichung (2) *in die obige drei Beschränkungen wird eingesetzt, und bekommt man das folgende Ergebnis*

$$\Pr(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\Big\{-\frac{(x-\mu)^2}{2\sigma^2}\Big\}$$

Die $\Pr(x)$ *ist nicht anderes als die Normalverteilung von X. Die Wahrscheinlichkeit wird in die Gleichung eingesetzt. Daraus ergibt sich dann $H(X) = \frac{1}{2}\Big\{1 + ln(2\pi\sigma^2)\Big\}$. Die Entropie von X ist abhängig nur von der Varianz σ^2 ab. Je größer die Varianz wird, umso größer wird Entropie.*

4.1.2 Gleichheit zwischen Maximum der Shannon Entropie und metrische Entropie

Diskrete Zufallsvariable muss für Gleichheit zwischen metrischer Entropie und Shannon Entropie gleich verteilt sein. Für gleich verteilte diskrete Zufallsvariable X ist Shannon Entropie $\log_2 n$ mit Wahrscheinlichkeit $\Pr(X = x_i) = 1/n$ für alle $i \in 1, ..., n$, und die metrische Entropie von X ist $log_2 \operatorname{cov}[r, \mathcal{S}, \mathcal{A}, d]$, wobei $\operatorname{cov}[r, \mathcal{S}, \mathcal{A}, d] \in \{1, ..., m\}$.

$$H(X) = \log_2 n = log_2 \operatorname{cov}[r, \mathcal{S}, \mathcal{A}, d] = \log_2 m = \text{ metrische Entropie von } X$$

Die Menge $\mathcal{S}$ hat Tupel $(x \in X, \Pr(X = x))$ als Element und $\mathcal{S} \subset \mathcal{A}$. Metrische Entropie einer diskreten Zufallsvariable hängt schwach von der Wahrscheinlichkeit ab. Denn maximale Überdeckungszahl einer diskreten Zufallsvariable

wird durch Elemente der diskreten Zufallsvariable für beliebig r bestimmt. Betrachte einen Graph einer diskreten Zufallsvariable. Der Graph besteht aus zwei Achsen: x-Achse zeigt Ereignisse x_i eienr Zufallsvariable und y-achse zeigt Wahrscheinlichkeiten der Ereignisse $\Pr(X = x_i)$. Wegen der Definition von Überdeckungszahl ist die maximale Anzahl von Bällen, die Punkte $(x_i, \Pr(X = x_i)$ zu überdecken, gleich wie die Anzahl der Ereignisse der Zufallsvariable. In diesem Fall muss die Ereignisse der Zufallsvariable endlich sein. Diese Idee kann man sich intuitiv vorstellen. Aber es ist schwer das Supremum einer Zufallsvariable zu erkennen, wenn die Menge nicht endlich ist.

Es gibt mathematische Methode, das Supremum einer unendlicher Menge zu finden. Die Methode ist nämlich das Theorem 3.12. Diskrete Zufallsvariablen sind abzählbar, weil eine bijektive Abbildung zwischen den Ereignissen einer diskreten Zufallsvariable und den natürlchen Zahlen existiert. Daher kann es höchstens abzählbare Teilmengen, die dicht auf der Zufallsvaraible liegen. Die abzählbare Teilmengen sind nämlich die Überdeckungsbällen der metrischen Entropie.

Stetige Zufallsvariablen sind überabzählbar, weil Definitionsbereich der Zufallsvariablen gleich wie Reelle Zahlen ist. Daher existieren abzählbare Überdeckungen nicht für alle stetige Zufallsvariablen. Wenn die abzählbare Überdeckungen existieren, kann das Theorem 3.12 angewendet werden, um das Supremum einer Menge zu finden.

Im Allgemeinen ist das Supremum einer Menge nicht gleich wie das Maximum der Menge. Deswegen muss abzählbare abgeschlossene und beschränkte (kompakte) Teilmengen, eine Menge zu überdecken, gewählt werden, damit die Vereinigung der Teilmengen abgeschlossen und beschränkt(kompakt) ist. Die Überdeckungsbälle in der metrischen Entropie sind kompakt, weil jeder Überdeckungsball $\max\left\{d(s, s_i), d(s_i, s)\right\} \leq r$ in der Definition 3.1 abgeschlossen und beschränkt ist. Daher kann das Supremum einer Menge, das mit dem Theorem 3.12 gesucht ist, gleich wie das Maximum der Menge sein.

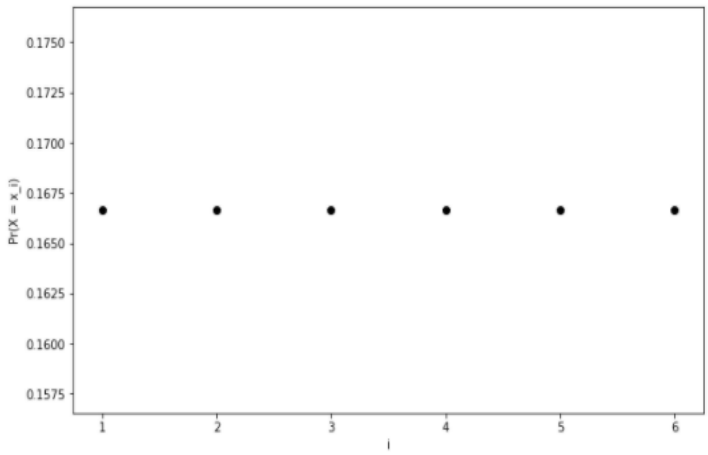
Beispiel 4.4. *Sei X eine Zufallsvariable mit $\Pr(X = x_i) = 1/6$ für $i \in \{1, 2, ..., 6\}$. Diese Zufallsvariable kann als folgender Graph dargestellt werden : Die Menge $\mathcal{S} = X$ und der Umgebungsraum $\mathcal{A} = \mathbb{R}^2$. Dann ist die Menge $\mathcal{S}$ begrenzt und sogar total begrenzt :*
sei $a, b \in (x \in X, \Pr(X = x))$ mit $a = (x_i, \Pr(X = x_i))$ und $b = (x_j, \Pr(X = x_j)$ für $i, j \in \{1, ..., 6\}$

$$\operatorname{diam}[X, d] := \sup_{a,b \in X} d[a, b] < 6 < \infty$$

$(X, \mathbb{R}^2, d)$ ist total begrenzt, weil die maximale Überdeckungszahl

$$\max \operatorname{cov}[r, X, \mathbb{R}^2, d] = 6$$

für beliebige r. D.h. $\operatorname{cov}[r, X, \mathbb{R}^2, d] < \infty$ für alle Radien r. Daher ist das Maximum der metrischen Entropie von X $\log_2 6$. Berechne nun die Shannon



Figur 7: Zufallsvariable X mit $\Pr(X = x_i)$

Entropie. Shannon Entropie von X ist laut der Definition 2.1 wie folgt :

$$\begin{aligned} H(X) &= -\sum_{i=1}^{6} \log_2 \Pr(X = x_i)^{\Pr(X=x_i)} \\ &= -6\frac{1}{6}\log_2 \frac{1}{6} \\ &= \log_2 6 \end{aligned}$$

Daher ist die metrische Entropie von X gleich wie die Shannon Entropie von X für beliebiegen Radien r und euklidische Norm d. Das Maxmium der metrischen Entropie von X ist $log_2 6$.

4.2 Anwendung der Shannon Entropie und der metrischen Entropie in der Codierungstheorie und ein Beispiel

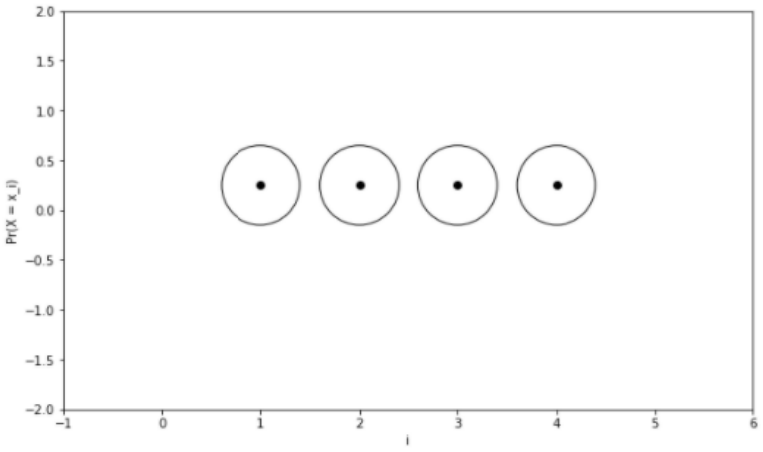
Wir wissen, dass die metrische Entropie ist das Maximum der Shannon Entropie bezüglich einer Zufallsvariable. Das Wissen wird angewendet, Komprimierungsrate in der Informatik zu berechnen.

Definition 4.5. *Seien Daten ein String. Kompermierunsrate der Daten ist definiert als*

$$Komprimierungsrate = \frac{Shannon\ Entropie}{Metrische\ Entropie}.$$

Shannon Entropie zeigt die durchschnittliche Anzahl der Bits des Strings beim Codieren, und metrische Entropie zeigt die maximale Shannon Entropie des Strings.

Mit Huffman Algorithmus im Kapitel 2.4 kann ein String(oder Daten) komprimiert werden, und durchschnittliche Anzahl der Bits des Strings, die es zu codieren erforderlich ist, kann durch die Anwendung der Shannon Entropie bestimmt werden. In dem Beispiel 2.11 war die Shannon Entropie des Strings 'apple' ungefähr 1.9. Berechne nun maximale Shannon Entropie von X. Der String 'apple' besteht aus 4 Buchstaben ,und zwar 'a','p','l' und 'e'. Bezeichne die vier Buchstaben als Elemente einer Zufallsvariable X. Die Shannon Entropie maximiert, wenn die vier Buchstaben gleiche Wahrscheinlichkeit $1/4$ besitzt**??**. Das Maximum der Shannon Entropie von X ist also $\log_2 4 = 2$ für 4 Elemente von X. Bestimme nun die metrische Entropie von X. Die Abstandfunktion d wird euklidische Norm gewählt. Jede Ereignisse der Zufallsvariable wird zu einer natürlichen Zahl zuordnet, die auf x -achse in Graph liegt, wenn wir die Zufallsvariable als einen Graphen Figur 8 dargestellt haben. Dann ist der Abstand zwischen der Ereignisse mehr als 1 für eine Zufallsvariable X. Daher wird Radius r kleiner als 0.5 gesetzt. Dann ist das Maximum der Überdeckungszahl von X gleich wie die Anzahl der Elemente der diskreten Zufallsvariable, und zwar $\log_2 4$. Daraus können wir schließen, dass die maximale Shannon Entropie ist gleich wie die metrische Entropie für die maximale Überdeckungszahl von X. Nun können wir die Komprimierungsrate bestimmen. Die Shannon Entropie wird durch metrische Entropie geteilt. Für das Beispiel 2.11 ist die Komprimierungsrate $\approx 1.9/2.0 = 0.95$. Das bedeutet, dass die Daten 5 Prozent komprimiert wird, wenn die Daten durch Huffman Algorithmus komprimiert wird.



Figur 8: Zufallsvariable mit Überdeckungsbällen

5 Fazit

Shannon Entropie ist die durchschnittliche Anzahl von Bits von Daten, die auf verschiedenen Weisen codiert worden sind. Eine Methode, Daten zu codieren, ist Huffman Algorithmus. Shannon Entropie kann der Erwartungswert des Informationsgehalts von Daten interpretiert werden. Shannon Theorem gibt uns Anzahl von Bits von Block der codierten Daten, um die Daten durch einen Kanal sicher zu übertragen. Shannon Theorem wird daher heutzutage in der Kommunikation sehr häufig verwendet. Metrische Entropie ist definiert als die Logarithmus der Überdeckungszahl einer Menge. Überdeckungszahl wird so bestimmt, wie viele Bälle mindestens erforderlich sind, eine Menge zu überdecken. Dabei spielt der Umgebungsraum der Menge eine große Rolle. Mit der metrischen Entropie kann Supremum einer Menge und sogar Supremum von empirischen Prozessen mit dem Technik 'chaning' bestimmt werden. Unter bestimmten Bedingungen kann die Shannon Entropie und metrische Entropie gleich sein. Für die Gleichheit müssen diskrete Zufallsvariablen gleich verteilt und stetige Zufallsvariablen normal verteilt sein. Dann maximiert Shannon Entropie von den Zufallsvarialben. Die maximale Shannon Entropie kann gleich wie metrische Entropie sein. Anwendung der Zusammenhang zwischen Shannon Entropie und metrischen Entropie ist Komprimierungsrate. Komprimierungsrate ist definiert als Teilung von Shannon Entropie durch metrische Entropie. Wenn ein String mit einem Algorithmus komprimiert wird, kann man sehen, wie stark der String mit dem Algorithmus komprimiert ist.

References

[1] R. Clausius. The mechanical theory of heat – with its applications to the steam engine and to physical properties of bodies. 1867.

[2] D.A. Huffman. Proceedings of the i.r.e. 1952.

[3] Yehuda Lindell Jonathan Katz. Introduction to modern cryptography. 2008.

[4] Johannes Lederer. Bounds for rademacher processes via chaining. 2010.

[5] Johannes Lederer. Modern mathematical statistics. 2020.

[6] Phil Lucht. Method of lagragne multipliers. 2016.

[7] Claude E Shannon. The bell system technical journal. 1948.

[8] Michael Mizenmacher. Eli Upfal. Probability and computing, randomized algorithms and probabilistic analysis. 2005.