# Zusammenfassung der 9. Vorlesung (17.12.07)

1.4 *Informationsmaße: Shannon und v. Neumann Entropie*
Das Verschränktheitsmaß für reine bipartite Zustände haben wir als die v. Neumann Entropie der partiellen Spur definiert. Die v. Neumann Enropie ist das Quantenanalogon der klassischen Shannon Entropie, die als Informationsmaß von Zeichenreihen große Bedeutung hat. Für die Quanteninformation hat die v. Neumann Entropie und darauf basierende Informationsmaße ebenso große Bedeutung.

1.4.1 *Die Informationsfunktion*:
Man betrachte ein Alphabet $\mathcal{A} = \{a_i\}_{i=0,1,2,\ldots,(M-1)}$ und eine damit bebildete Zeichenreihe $(x_1, x_2, \ldots, x_N) \in \mathcal{A}^N$, die wir auch *Wort* nennen wollen. Ist die Zeichenreihe unbekannt, dann liefert das zufällige Erkennen von Zeichen aus dieser Reihe Information über das Wort, die durch eine zu bestimmende Informationsfunktion $I$ gemessen werden soll.

Um den Prozess des zufälligen Erkennens von Zeichen zu konkretisieren, kann man sich die $M$ Buchstaben des Alphabets auf Spielkarten geruckt vorstellen. Jedes der möglichen $M^N$ Wörter kann dann mit Hilfe von $N$ solcher Karten gebildet und aufgelegt werden. Sind die Karten verdeckt aufgelegt, entspricht dies der Unkenntnis des Wortes. Das zufällige Erkennen eines Zeichens ist dann gegeben, wenn die Karten vor dem verdeckten Auflegen gemischt werden. Das Aufdecken einer willkürlich gewählten Karte führt zum zufälligen Erkennen eines Zeichens. Nachdem diese Karte zurückgelegt und erneut gemischt wurde, kann dieses Spiel zum zufälligen Erkennen eines weiteren Zeichens wiederholt werden. Dies kann schrittweise beliebig oft fortgesetzt werden, wobei die Information über das Wort immer größer wird.

Bei jedem Schritt ist die Wahrscheinlichkeit, das Zeichen $a_i$ zu entdecken, offenbar $p(a_i) = n((a_i)/N$, wenn $n(a_i)$ mal das Zeichen $a_i$ in dem Wort vorkommt. Überdies sind die Ereignisse des Erkennens eines Zeichens unabhängig, so dass die Wahrscheinlichkeit, schrittweise Zeichen $y_k \in \mathcal{A}$, $k = 1.2. \ldots, K$, zu erkennen, durch

$$p(y_1, y_2, \ldots, y_K) = p(y_1)p(y_2)\ldots p(y_K)$$

gegeben ist. Erstaunlicher Weise legen die folgenden drei Axiome die Informationsfunktion eindeutig fest:

**Axiom 1:** I ist nicht negativ und nur von der Wahrscheinlichkeit $p(y_1, y_2, \ldots, y_K)$ abhängig, d.h. $I : [0,1] \cap \mathbf{Q} \to [0, \infty]$.

Das zweite Axiom legt das Anwachsen der Information fest, wenn das schrittweise Erkennen eines oder mehrerer Zeichen in Stufen eingeteilt wird.

**Axiom 2:** I ist additiv, d.h. $I(p \cdot q) = I(p) + I(q)$

Das dritte Axiom legt die Maßeinheit fest, in der die Informationsfunktion die Information angibt. Wir wählen das Bit als Einheit.

**Axiom 3:** $I(1/2) = 1$.

Man folgert unmittelbar, dass diese Axiome von

$$I(p) = -\log_2 p = \log_2 \frac{1}{p}$$

erfüllt werden. Weniger trivial ist es, dass diese Funktion die einzige Funktion ist, die diese Axiom erfüllt: Wegen $0^2 = 0$ und $1^2 = 1$ sind $I(0)$ und $I(1)$ Lösungen der Gleichung $x = 2x$, also sind diese Funktionswerte 0 oder $\infty$. $I(p)$ ist monoton fallend, denn aus $p < q$ folgt $I(p) = I(\frac{p}{q}q) = I(q) + I(\frac{p}{q}) \geq I(q)$. Für $p \in (0,1)$, $\alpha, \beta \in \mathbf{N}$, $(\beta \neq 0)$, $q = \frac{\alpha}{\beta}$, gilt $I(p^q) = qI(p)$, denn $I(p) = I((p^{\frac{1}{\alpha}})^\alpha) = \alpha I(p^{\frac{1}{\alpha}})$, also $I(p^{\frac{1}{\alpha}}) = \frac{1}{\alpha} I(p)$, Betrachte nun eine Folge rationaler Zahlen $0 < p_n \leq 1$ mit $p_n \to 1$. Diese enthält eine Teilfolge $q_n$ mit $q < q_{n+1}$und $q_n \to 1$ und diese wiederum eine solche $r_n$ mit $q_1^{\frac{1}{n}} < r_n$. Nun ist $\frac{1}{n} I(q_1) = I(q_1^{\frac{1}{n}}) \geq I(r_n) \geq 0$ und damit $\lim_{n\to\infty} \frac{1}{n} I(q_1) \geq \lim_{n\to\infty} I(r_n) = \lim_{n\to\infty} I(q_n) = 0$. Sei nun $p \in (0,1)$, rational, und $p_n$ eine Folge rationaler Zahlen mit $p \leq p_n < 1$ und $p_n \to p$, dann gilt $I(p) = I((\frac{p}{p_n})p_n) = I(p_n) + I(\frac{p}{p_n})$. Für $n \to \infty$ gilt $I(p) = \lim_{n\to\infty} I(p_n) + \lim_{n\to\infty} I(\frac{p}{p_n}) = \lim_{n\to\infty} I(p_n) + 0$, weil $\frac{p}{p_n} \to 1$. Analog zeigt man $I(p) = \lim_{n\to\infty} I(p_n)$ für Folgen rationaler Zahlen, die von links gegen $p$ streben. $I(p) = \lim_{n\to\infty} I(p_n)$ gilt deshalb für alle Folgen rationaler Zahlen die gegen $p$ streben. Ist $p$ irrational, dann ist $I(p_n)$ eine Cauchyfolge, und dies dolgt so: $p_n$ ist eine Cauchyfolge, also gibt es für alle $\epsilon > 0$ eine Zahl $N(\epsilon)$ mit $|p_m - p_n| < \epsilon$ für $m, n \geq N(\epsilon)$. Sei o.B.d.A. $p_m > p_n$, dann ist für alle $0 < \tilde{\epsilon} = \epsilon/p_m$ auch $|1 - (p_n/p_n)| < \tilde{\epsilon}$, wenn nur $m, n \geq M(\tilde{\epsilon}) := N(\tilde{\epsilon}/p_m)$ ist. Sei nun $\epsilon_1 > 0$, dann gibt es eine Zahl $N_1(\epsilon_1)$, so dass

$$|I(p_m) - I(p_n)| = |I(\frac{p_n}{p_m})| = |I(\frac{p_n}{p_m}) - I(1)| < \epsilon_1$$

ist, wenn nur $m, n \geq N_1(\epsilon_1)$ ist, denn wir hatten weiter oben gezeigt, dass dass für Folgen rationaler Zahlen, die von links gegen 1 konvergieren, die Folge der Funktionswerte von I gegen $I(1) = 0$ konvergiert. Man kann nun für irrationale $p$ den Wert $I(p)$ als den Grenzwert der Cauchyfolge $I(p_n)$ definieren. Die stetige Fortsetzung von $I$ auf $(0, 1]$ ist damit eindeutig durch die Werte auf den rationalen Zahlen bestimmt.

Die eindeutige stetige Fortsetzung der Informationsfunktion auf $(0, 1)$ ist auch differenzierbar: Für $p \in (0, 1)$ sei $1 > p_n > p_n + 1 > p$ und $p_n \to p$, dann gilt mit $\kappa_n := 1 - \frac{p}{p_n}$

$$\begin{aligned} \frac{T(p_n) - I(p)}{p_n - p} &= \frac{1}{p_n - p} I(\frac{p}{p_n}) = -I((\frac{p}{p_n})^{\frac{1}{p_n - p}}) = -\frac{1}{p_n} I((\frac{p}{p_n})^{\frac{p_n}{p_n - p}}) \\ &= -\frac{1}{p_n} I((1 - \kappa_n)^{\frac{1}{\kappa_n}}) \qquad \longrightarrow \qquad -\frac{1}{p} I(\frac{1}{e}). \end{aligned}$$

Analog folgt dies für die linksseitige Ableitung. Damit ist. Damit ist $I'(p) = -\frac{1}{p} I(\frac{1}{e})$ und, wegen $I(1) = 0$ ist $I(p) = -I(\frac{1}{e}) \ln p$ die die eindeutig bestimmte Lösung der Differentialgleichung. Hierbei kann noch über $I(\frac{1}{e})$ verfügt werden. Aus $e^{\alpha x} = 2^x = y$ folgen $x = \log_2 y$, $\alpha x = \ln y$ und $\alpha = \ln 2$, und damit $log_2 y = \frac{1}{\ln 2} \ln p$. Setzt man also $I(\frac{1}{e}) = \frac{1}{\ln 2}$, dann ist

$$I(p) = -\log_2 p$$

und $I(\frac{1}{2}) = 1$, wie im Axion 3 festgelegt wurde. Damit ist folgender Satz bewiesen :

**Satz:** Die Axiome 1 - 3 legen die Informationsfunktion eindeutig als die Einschränkung von $I(p) = -\log_2 p$ auf rationale Argumente fest.

Die Informationsfunktion hängt nur von den Häufigkeiten ab. mit denen einzelne Zeichen in einer Zeichenreihe vorkommen, nicht aber von deren Anordnung. $I(p) = -\log_2 p$ wird auch für beliebige Verteilungen mit nichtrationalen Wahrscheinlichkeiten $p_i \geq 0$, $\sum_i p_i = 1$, die idealisiert bei unendlichen Zeichenreihen auftreten können, verwendet.

1.4.2 *Die Shannon Entropie*:
Die Shannon Entropie ist der Erwartungswert der Informationsfunktion

$$H = \sum_{i=0}^{M-1} p_i I(p_i) = -\sum_{i=0}^{M-1} p_i \log_2 p_i, \quad \text{wobei} \quad 0 \log_2 0 = 0$$

gerechnet wird. Besteht das Wort aus paarweise gleichen Zeichen, dann ist $H = 0$. Sind alle Zeichen gleichverteilt, dann ist $H = \log_2 M$.

$$-1 \log_2 1 = 0 \quad \leq \quad H \quad \leq \quad -\sum_{i=0}^{M-1} \frac{1}{M} \log_2 \frac{1}{M} = \log_2 M.$$

Letzterer Wert ist stationär, denn mit $0 = d1 = d(\sum_{i=0}^{M-1} p_i) = \sum_{i=0}^{M-1} dp_i$ ist $p_i = (1/M)$ Lösung von $dH = -\sum_{i=0}^{M-1}(\log_2 p_i + \frac{1}{\ln 2})dp_i = 0$. Nun ist $(\partial^2 H/\partial p_i^2) = -\frac{1}{p_i \ln 2} < 0$ und $(\partial^2 H/\partial p_i \partial p_k) = 0$ für $i \neq k$, so dass die quadratische Form $\sum_{i,k=0}^{M-1}(\partial^2 H/\partial p_i \partial p_k)dp_i dp_k$ strikt negativ ist. $H$ ist desahalb strikt konkav und nimmt bei $p_i = (1/M)$ ihren maximalen Wert, $\log_2 M$, an.