Die Entropie Funktion

Michael Förster

24. Juli 2004

1 Einführung

Die Entropie ist ein Grundkonzept der Informationstheorie. Eine gute Einführung in die Informationstheorie ist in [1] zu finden. Die Informationstheorie behandelt das Konzept der Übertragung von Information von einer Quelle zu einem Empfänger.

Die Entropie mißt die Unbestimmtheit beziehungsweise den Informationsgewinn bei der Codierung von Quellen. Außerdem kann mit ihr die Leistungsfähigkeit von gestörten Kanälen ausgedrückt werden.

Hierbei ist es gleichgültig, ob ein Zufallsexperiment durch die Unsicherheit über den Ausgang **vor** Ausführung oder den Informationsgewinn **nach** Bekanntwerden des Ausgangs beurteilt wird. Beide Größen können durch dieselbe Maßzahl gemessen werden.

Definition 1 *Sei X eine Zufallsvariable mit Werten im Bereich B und sei mit p_b die Wahrscheinlichkeit gemeint, dass X den Wert b annimmt, formal also $P(X = b) = p_b$. Die binäre Entropie von X bezeichnen wir mit $H[X]$ und ist definiert durch:*

$$H[X] := \sum_{b \in B} -p_b \cdot \log_2(p_b)$$

mit der Konvention, dass $0 \cdot \log_2 0 := 0$ gelten soll.

Nützlich für folgende Beweise ist die sogenannte Gibbs Ungleichung.

Lemma 2 *(Gibbs Ungleichung) Seien $p_1, \ldots, p_M$ und $q_1, \ldots, q_M$ Wahrscheinlichkeiten mit $\sum_{i=1}^{M} p_i = \sum_{i=1}^{M} q_i = 1$. Dann*

$$-\sum_{i=1}^{M} p_i \cdot \log_2(p_i) \leq -\sum_{i=1}^{M} p_i \cdot \log_2(q_i)$$

mit Gleichheit genau dann, wenn $p_i = q_i$ für alle $i \in 1, \ldots, M$.

Beweis:
Wir benutzen den natürlichen Logarithmus statt den Logarithmus dualis, weil gilt $\log_2(x) = \frac{ln(x)}{ln(2)}$ und die Behauptung durch eine positive Konstante nicht verändert wird. Weiter gilt $ln(x) \leq x - 1$ und $ln(x) = x - 1 \Leftrightarrow x = 1$

Daraus folgt dann $ln(\frac{q_i}{p_i}) \leq \frac{q_i}{p_i} - 1$ und Gleichheit genau dann, wenn $p_i = q_i$. Nach Voraussetzung gilt $\sum_{i=1}^{M} p_i = 1$, daher können wir die Summe auf beiden Seiten multiplizieren.

$$\sum_{i=1}^{M} p_i \cdot ln(\frac{q_i}{p_i}) \leq \sum_{i=1}^{M} p_i(\frac{q_i}{p_i} - 1) = \underbrace{\sum_{i=1}^{M} q_i}_{=1} - \underbrace{\sum_{i=1}^{M} p_i}_{=1} = 1 - 1 = 0$$

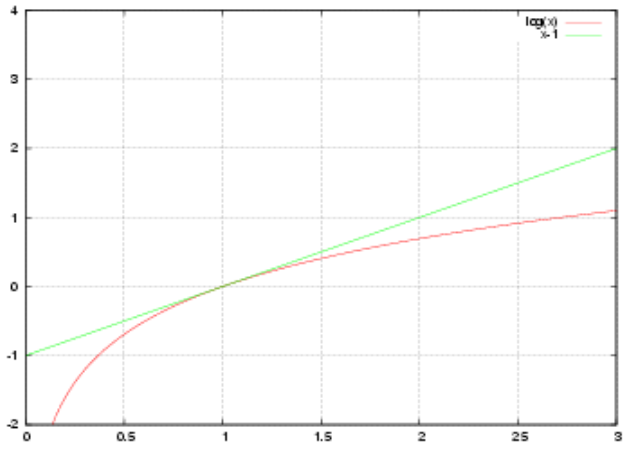

Abbildung 1: $ln(x) \leq x - 1$ *Graphisch dargestellt*

Mit Gleichheit genau dann, wenn $p_i = q_i$ für alle i. Daraus folgt

$$\sum_{i=1}^{M} p_i \cdot ln(\frac{q_i}{p_i}) = \sum_{i=1}^{M} p_i(ln\ q_i - ln\ p_i) = \sum_{i=1}^{M} p_i ln\ q_i - \sum_{i=1}^{M} p_i ln\ p_i \leq 0 \qquad \blacksquare$$

2 Grundlegende Eigenschaften

a.) Wenn $|B| = 2^t$ dann gilt

$$H[X] \leq t \tag{1}$$

Mit Gleichheit genau dann, wenn $P(X = b) = p_b = \frac{1}{2^t}$ für alle $b \in B$
Um dies zu zeigen wenden wir Gibbs Ungleichung an mit $q_b = \frac{1}{2^t}$ für alle $b \in B$. Dazu:

$$\sum_{b \in B} p_b\ \log_2\ p_b \leq -\sum_{b \in B} p_i\ \log_2\ 2^{-t} = (-1) \cdot (-t) = t$$

b.) Entropie hat die **Konzentrations-Eigenschaft**

Wenn $H[X] \leq t$, dann muß es ein b geben so dass gilt:

$$p_b = P(X = b) \geq 2^{-t} \tag{2}$$

Dazu: Annahme für alle $b \in B$ gilt $p_b < 2^{-t}$:
Dann folgt für die Entropie

$$\begin{aligned} H[X] &= \sum_{b \in B} -p_b \cdot \log_2(p_b) > \sum_{b \in B} -p_b \cdot \log_2(2^{-t}) \\ &= (\sum_{b \in B} -p_b) \cdot (-t) = (\underbrace{\sum_{b \in B} p_b}_{=1}) \cdot t = t \end{aligned}$$

Dies ist ein Widerspruch zu (1).

c.) Die Entropie ist **subadditiv**
Wenn $X = (X_1, X_2, \ldots, X_n)$, dann $H[X] \leq \sum_{i=1}^{n} H[X_i]$ (Beweis kommt in Kapitel 3).

Definition 3 *Wenn E ein Ereignis ist, können wir die* ***bedingte Entropie*** *von X unter der Bedingung, dass E eingetreten ist definieren durch*

$$H[X|E] := \sum_{b \in B} -P(X = b|E) \cdot \log_2(P(X = b|E))$$

Auf die gleiche Weise könen wir bedingte Entropie von X mit gegebenen Y definieren, wobei Y eine Zufallsvariable ist, welche Werte aus einem Bereich A annimmt.

$$H[X|Y] := \sum_{a \in A} H[X|Y = a] \cdot P(Y = a)$$

$$= \sum_{a \in A} \sum_{b \in B} -P(X = b|Y = a) \cdot \log_2 \; P(X = b|Y = a) \cdot P(Y = a)$$

$$= \sum_{a \in A} \sum_{b \in B} -P(X = b \cap Y = a) \cdot \log_2 \; P(X = b|Y = a)$$

Wenn bestimmte Werte von Y gegeben sind, stellen wir uns $H[X|Y]$ als Unbestimmtheit von X vor, verteilt über den Wertebereich, den Y annehmen kann.

d.) $H[X|X] = 0$, weil

$$H[X|X] = -\sum_{a \in A} \sum_{a \in A} -P(X = a|X = a) \cdot \underbrace{\log_2 \underbrace{P(X = a|X = a)}_{=1}}_{=0} \cdot P(X = a) = 1$$

e.) $H[X|Y] = 0 \Leftrightarrow X = f(Y)$ für eine Funktion f.
Anders ausgedrückt ist $H[X|Y] = 0$ genau dann, wenn X total abhängig von Y ist.

f.) $H[X|Y] = H[X]$, wenn X und Y unabhängig sind. Dazu:

$$= \sum_{a \in A} \sum_{b \in B} -P(X = b|Y = a) \cdot \log_2 \; P(X = b|Y = a) \cdot P(Y = a)$$

$$= \underbrace{\sum_{a \in A} P(Y = a)}_{=1} \cdot \underbrace{(-\sum_{b \in B} P(X = b) \cdot \log_2 \; P(X = b))}_{=H[X]} = H[X]$$

Definition 4

$$H[X,Y] := -\sum_{a\in A}\sum_{b\in B} P(X=b,Y=a)\cdot \log_2 P(X=b,Y=a)$$

g.) Wenn wir mit mehr als nur einer ZV bedingen also eine Schnittmenge über die Ereignismengen nehmen wird klar, dass gilt:

$$H[X|Y,Z] \leq H[X|Y] \tag{3}$$

Die Haupteigenschaft der bedingten Entropie ist die folgende:

$$H[X,Y] = H[Y] + H[X|Y] \tag{4}$$

Beweis: Der Trick ist hier eine Null-Addition

$$0 = -\log_2(P(Y=a)) + \log_2(P(Y=a))$$

$$H(X,Y) = -\sum_{a\in A, b\in B} P(X=b,Y=a)\cdot \{\underbrace{\log_2 P(X=b,Y=a) - \log_2 P(Y=a)}_{=\log_2 P(X=b|Y=a)} + \log_2 P(Y=a)\}$$

$$= -\sum_{a\in A}\sum_{b\in B} P(X=b,Y=a)\cdot \log_2 P(X=b|Y=a) - \sum_{a\in A}\underbrace{\sum_{b\in B} P(X=b,Y=a)}_{=P(Y=a)}\cdot \log_2 P(Y=a)$$

$$= -\sum_{a\in A}\sum_{b\in B} \underbrace{P(X=b,Y=a)}_{=P(Y=a)\cdot P(X=b|Y=a)} \cdot \log_2 P(X=b|Y=a) - \underbrace{\sum_{a\in A} P(Y=a)\cdot \log_2 P(Y=a)}_{=H[Y]}$$

$$= -\sum_{a\in A} P(Y=a)\cdot \underbrace{\sum_{b\in B} P(X=b|Y=a)\cdot \log_2 P(X=b|Y=a)}_{=H[X|Y=a]\ nach\ Def.\ 1} + H[Y]$$

$$= \underbrace{-\sum_{a\in A} P(Y=a)\cdot H[X|Y=a]}_{=H[X|Y]\ nach\ Def.\ 3} + H[Y]$$

$$= H[X|Y] + H[Y] \qquad \blacksquare$$

3 Subadditivität (Subadditivity)

Satz 5 *Wenn X und Y zwei Zufallsvariablen sind, die nur endlich viele Werte annehmen, dann*

$$H[X,Y] \leq H[X] + H[Y]$$

mit Gleichheit nur wenn X und Y stochastisch unabhängig sind.

Beweis:
Nehmen wir an X und Y nehmen ihre Werte in A bzw. B an. Sei $p_{a,b}$ die Wahrscheinlichkeit $P((X,Y) = (a,b)) = p_{a,b}$ und sei p_a die Wahrscheinlichkeit, dass X den Wert a annimmt also $P(X = a) = p_a$ und $p_b = P(Y = b)$. Da $\sum_{b\in B} p_{a,b} = p_a$ und $\sum_{a\in A} p_{a,b} = p_b$ gilt, folgt

$$\begin{aligned}
H[X] + H[Y] &= \sum_{a\in A} -p_a \log p_a + \sum_{b\in B} -p_b \log p_b \\
&= \sum_{a\in A}\sum_{b\in B} -p_{a,b} \log p_a + \sum_{b\in B}\sum_{a\in A} -p_{a,b} \log p_b \\
&= \sum_{a\in A}\sum_{b\in B} -p_{a,b} \log p_a + \sum_{b\in B}\sum_{a\in A} -p_{a,b} \log p_b \\
&= \sum_{(a,b)\in A\otimes B} -p_{a,b}(\log p_a + \log p_b) \\
&= \sum_{(a,b)\in AxB} -p_{a,b} \cdot log(p_a \cdot p_b)
\end{aligned}$$

Weil $\sum_{a,b} p_a \cdot p_b = (\sum_a p_a) \cdot (\sum_b p_b) = 1 \cdot 1 = 1$ und $\sum_{a,b} p_{a,b} \leq 1$ können wir Gibbs Ungleichung anwenden und bekommen

$$\begin{aligned}
H[X] + H[Y] &= \sum_{(a,b)\in AxB} -p_{a,b} log(p_a \cdot p_b) \overset{La(2)}{\geq} \sum_{(a,b)\in AxB} -p_{a,b} log(p_{a,b}) \\
&= H[X,Y]
\end{aligned}$$

äquivalente Definition für stochastische Unabhängigkeit von X,Y ist. ■

Satz 6 *Sei $X = (X_1, \ldots, X_n)$ ein Zufallsvektor, der Werte in der Menge $B = B_1 \otimes B_2 \otimes \ldots \otimes B_n$ annimmt, wobei X_i Werte in B_i annimmmt. Dann*

$$H[X] \leq \sum_{i=1}^{n} H[X_i]$$

mit Gleichheit genau dann wenn $X_1, \ldots, X_n$ paarweise stochastisch unabhängig sind.

Beweis:
$n = 2$: Lemma 2
$n - 1 \to n$:

$$
\begin{aligned}
& \sum_{i=1}^{n} H[X_i] \overset{IV}{\geq} H[X_1, \ldots, X_{n-1}] + H[X_n] \\
= & \sum_{(b_1,\ldots,b_{n-1}) \in B_1,\ldots,B_{n-1}} -p_{b_1,\ldots,b_{n-1}} \cdot log(p_{b_1,\ldots,b_n-1}) + \sum_{b_n \in B_n} -p_{b_n} \cdot log(p_{b_n}) \\
= & \sum_{(b_1,\ldots,b_n) \in B} -p_{b_1,\ldots,b_n} \cdot log(p_{b_1,\ldots,b_n-1}) + \sum_{(b_1,\ldots,b_n) \in B} -p_{b_1,\ldots,b_n} \cdot log(p_{b_n}) \\
= & \sum_{(b_1,\ldots,b_n) \in B} -p_{b_1,\ldots,b_n} \cdot log(p_{b_1,\ldots,b_n-1} \cdot p_{b_n}) \\
\overset{La(2)}{\geq} & \sum_{(b_1,\ldots,b_n) \in B} -p_{b_1,\ldots,b_n} \cdot log(p_{b_1,\ldots,b_n}) \\
= & H[X_1, X_2, \ldots, X_n] \qquad \blacksquare
\end{aligned}
$$

Eine interessante Erweiterung wurde von Chung, Frankl, Graham und Shearer 1986 bewiesen. Wie zuvor sei $X = (X_1, X_2, \ldots, X_n)$ ein Zufallsvektor der Werte aus $B = B_1 \otimes B_2 \otimes \ldots \otimes B_n$ annimmt und jede Zufallsvariable X_i Werte in B_i annimmt. Nehmen wir weiter an, dass $B_i = \{0, 1\}$ für alle $i \in \{1, \ldots, n\}$. Für eine Teilmenge von Koordinaten S bezeichne X_S eine Zufallsvariable $(X_i)_{i \in S}$.

Satz 7 *(Allgemeine Subadditivität) Sei $X = (X_1, X_2, \ldots, X_n)$ und B wie oben und seien $S_1, \ldots, S_m$ mit $m \leq n$ Teilmengen von $[n] = \{1, \ldots, n\}$ so dass jedes $i \in [n]$ mindestens zu $k \geq 1$ Mengen von $S_1, \ldots, S_m$ gehört. Dann*

$$H[X] \leq \frac{1}{k} \cdot \sum_{i=1}^{m} H[X_{S_i}]$$

Was heißt das?
Beispiel (mit n=5 und m=3):
$X = (X_1, \ldots, X_5)$ und $B = B_1 \otimes B_2 \otimes B_3 \otimes B_4 \otimes B_5$ und $B_i = \{0, 1\}$.
S_1, S_2, S_3 sind Teilmengen von [5]={1,2,3,4,5} so dass $i \in [5]$ zu mindestens k Teilmengen von S_1, S_2, S_3 gehört.

Seien die Teilmengen: $S_1 = \{3, 4\}$, $S_2 = \{1, 2, 3\}$, $S_3 = \{2, 3, 5\}$, Dann gilt

$$\left.\begin{array}{c} 1 \in S_2 \\ 2 \in S_2, S_3 \\ 3 \in S_1, S_2, S_3 \\ 4 \in S_1 \\ 5 \in S_3 \end{array}\right\} k = 1$$

Mit Satz 7 gilt dann: $H[X_1, \ldots, X_5] \leq H[X_{S_1}] + H[X_{S_2}] + H[X_{S_3}] = H[X_3, X_4] + H[X_1, X_2, X_3] + H[X_2, X_3, X_5]$.

Seien die Teilmengen: $S_1 = \{1, 2, 3\}$, $S_2 = \{2, 3, 4, 5\}$, $S_3 = \{1, 3, 4, 5\}$, Dann gilt

$$\left.\begin{array}{c} 1 \in S_1, S_3 \\ 2 \in S_1, S_2 \\ 3 \in S_1, S_2, S_3 \\ 4 \in S_2, S_3 \\ 5 \in S_2, S_3 \end{array}\right\} k = 2$$

In der Notation von oben heißt $X_{S_1} = (X_3, X_4)$, wenn für die Teilmenge S_1 gilt: $S_1 = \{3, 4\}$

Beweis:
Für k=1 folgt die Behauptung aus Satz 6. Sei nun $k > 1$ und v bezeichne die minimale Anzahl von S_i's deren Vereinigung [n] ergibt.

Wir werden die Behauptung durch Induktion über k und v zeigen. Wenn $v = 1$ dann existiert ein S_i mit $S_i = [n]$. O.B.d.A. sei $S_1 = [n]$. Da $k > 1$ liegen jetzt alle Elemente aus [n] in mindestens $k - 1$ Mengen von $S_2, \ldots, S_m$. Nach Induktion über k können wir folgern:

$$\begin{aligned}
& (k-1) \cdot H[X_1, \ldots, X_n] && \leq \sum_{i=2}^{m} H[X_{S_i}] \\
\Leftrightarrow \quad & k \cdot H[X_1, \ldots, X_n] && \leq \sum_{i=2}^{m} H[X_{S_i}] + H[X_1, \ldots, X_n] \\
\underset{S_1=[n]}{\Leftrightarrow} \quad & k \cdot H[X_1, \ldots, X_n] && \leq \sum_{i=2}^{m} H[X_{S_i}] + H[X_{S_1}] \\
\Leftrightarrow \quad & k \cdot H[X_1, \ldots, X_n] && \leq \sum_{i=1}^{m} H[X_{S_i}]
\end{aligned}$$

Sei nun $v > 1$, wir benötigen also v Mengen von $S_1, \ldots, S_m$ vereinigt, um [n] zu erhalten. Seien dies O.B.d.A. $S_1, \ldots, S_v$ also $S_1 \cup S_2 \cup \ldots \cup S_v = [n]$. Seien $S_1' := S_1 \cup S_2$ und $S_2' := S_1 \cap S_2$. Jedes Element von [n] ist in mindestens k Mengen von $S_1', S_2', S_3, \ldots, S_m$, weil

$$\begin{aligned}
p \in S_1 \wedge p \in S_2 \quad &\Rightarrow \quad p \in S_1' \wedge p \in S_2' \\
p \in S_1 \wedge p \notin S_2 \quad &\Rightarrow \quad p \in S_1' \wedge p \notin S_2' \\
p \notin S_1 \wedge p \in S_2 \quad &\Rightarrow \quad p \in S_1' \wedge p \notin S_2' \\
p \notin S_1 \wedge p \notin S_2 \quad &\Rightarrow \quad p \notin S_1' \wedge p \notin S_2'
\end{aligned}$$

Außerdem genügen schon $v - 1$ von diesen Mengen, um [n] zu erhalten, weil $[n] = S_1' \cup S_3 \cup \ldots S_v$. Nach Induktion über v gilt:

$$k \cdot H[X_1, X_2, \ldots, X_n] \leq \sum_{i=3}^{m} H[X_{S_i}] + H[X_{S_1'}] + H[X_{S_2'}]$$

Setze $X = X_{S_1 - S_2}$, $Y = X_{S_1 \cap S_2}$, $Z = X_{S_2 - S_1}$ und nach **??** gilt:

$$\begin{aligned} H[X_{S_1 \cup S_2}] &\leq H[X_{S_1}] + H[X_{S_2}] - H[X_{S_1 \cap S_2}] \\ \Leftrightarrow H[X_{S_1 \cup S_2}] + H[X_{S_1 \cap S_2}] &\leq H[X_{S_1}] + H[X_{S_2}] \qquad (5) \\ k \cdot H[X_1, \ldots, X_n] &\leq \sum_{i=3}^{m} H[X_{S_i}] + H[X_{S_1 \cup S_2}] + H[X_{S_1 \cap S_2}] \\ &\leq \sum_{i=1}^{m} H[X_{S_i}] \qquad \blacksquare \qquad (6) \end{aligned}$$

4 Kombinatorische Anwendungen

Mit Satz (6) wurde von Kleitman, Shearer und Sturtevant (1981) benutzt um verschiedene, interessante Anwendungen in der Extremalen-Kombinatorik zu entwickeln. Die Grundidee kann durch folgendes einfaches Korollar von 6 verdeutlicht werden.

Korollar 8 *Sei $\mathcal{F}$ eine Familie von Teilmengen von $\{1, 2, \ldots, n\}$ und sei mit p_i die Menge der Bruchteile von Mengen in $\mathcal{F}$ bezeichnet, die i enthalten. Dann*

$$|\mathcal{F}| \leq 2^{\sum_{i=1}^{n} H(p_i)}$$

wobei $H(y) := -y \cdot \log_2(y) - (1-y) \cdot \log_2(1-y)$

Beweis:
Bringen wir jede Menge $F \in \mathcal{F}$ in Zusammenhang mit ihrem Vorkommen-Vektor v_F, der ein Vektor der Länge n ist.

Beispiel: $\mathcal{F} = \{S_1, S_2, S_3\}$ Familie von Teilmengen von $\{1, \ldots, 5\}$ $S_1 = \{1\} \Rightarrow v_{S_1} = (1,0,0,0,0)$ $S_2 = \{4,5\} \Rightarrow v_{S_2} = (0,0,0,1,1)$ $S_3 = \{2\} \Rightarrow v_{S_3} = (0,1,0,0,0)$

Sei $X = (X_1, \ldots, X_n)$ eine Zufallsvariable, die Werte in $\{0,1\}^n$ annimmt, wobei $P((X_1, X_2, \ldots, X_n) = v_F) = \frac{1}{|\mathcal{F}|}$ für alle $F \in \mathcal{F}$. Klar ist:

$$\begin{aligned} H[X] &= -\sum_{F \in \mathcal{F}} P(X = v_F) \cdot \log_2 P(X = v_F) \\ &= |\mathcal{F}| \cdot (-\frac{1}{|\mathcal{F}|} \cdot \log_2 \frac{1}{|\mathcal{F}|}) \\ &= |\mathcal{F}| \cdot (\frac{1}{|\mathcal{F}|} \cdot \log_2 |\mathcal{F}|) \\ &= \log_2 |\mathcal{F}| \end{aligned}$$

Es gilt hier $p_i = P(X_i = 1)$ und $1-p_i = P(X_i = 0)$ und weil hier $H[X_i] = H(p_i)$ für alle $1 \leq i \leq n$ gilt:

$$\begin{aligned} H[X] &= H[X_1, \ldots, X_n] = \log_2 |\mathcal{F}| \\ &\overset{Satz6}{\leq} \sum_{i=1}^{n} X_i = \sum_{i}^{n} H(p_i) \\ \Rightarrow |\mathcal{F}| &= 2^{\log_2 |\mathcal{F}|} \leq 2^{\sum_{i=1}^{n} H(p_i)} \end{aligned}$$ ■

Dieses Korollar liefert einen schnellen Beweis für die wohlbekannte Abschätzung:

Korollar 9 *Für jede Ganzzahl n und für jede reelle Zahl* $0 < p \leq \frac{1}{2}$ *gilt:*

$$\sum_{i \leq np} \binom{n}{i} \leq 2^{n \cdot H(p)}$$

Beweis: (nach P. Frankl)
Sei $\mathcal{F}$ die Familie aller Teilmengen von $\{1, 2, \ldots, n\}$ mit der Ordnung höchstens $p \cdot n$ also $|\mathcal{F}| = \sum_{i \leq np} \binom{n}{i}$. Sei weiter p_i der Bruchteil der Teilmengen von $\mathcal{F}$, die i enthalten dann gilt $p_1 = p_2 = \cdots = p_n$.
Beispiel mit n=5 und p=$\frac{1}{2}$:
$n \cdot p = \frac{5}{2} \Rightarrow$ Teilmengen haben höchstens die Gröβe 2.

$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{2,3\}, \{2,4\}, \{2,5\}, \{3,4\}, \{3,5\}, \{4,5\}\}$$

$$|\mathcal{F}| = \binom{5}{0} + \binom{5}{1} + \binom{5}{2} = 1 + 5 + 10 = 16$$

n		1	2	3	4	5	
1	$\emptyset$	0	0	0	0	0	$\leq np$
2	$\{1\}$	1	0	0	0	0	$\leq np$
3	$\{2\}$	0	1	0	0	0	$\leq np$
4	$\{3\}$	0	0	1	0	0	$\leq np$
5	$\{4\}$	0	0	0	1	0	$\leq np$
6	$\{5\}$	0	0	0	0	1	$\leq np$
7	$\{1,2\}$	1	1	0	0	0	$\leq np$
8	$\{1,3\}$	1	0	1	0	0	$\leq np$
9	$\{1,4\}$	1	0	0	1	0	$\leq np$
10	$\{1,5\}$	1	0	0	0	1	$\leq np$
11	$\{2,3\}$	0	1	1	0	0	$\leq np$
12	$\{2,4\}$	0	1	0	1	0	$\leq np$
13	$\{2,5\}$	0	1	0	0	1	$\leq np$
14	$\{3,4\}$	0	0	1	1	0	$\leq np$
15	$\{3,5\}$	0	0	1	0	1	$\leq np$
16	$\{4,5\}$	0	0	0	1	1	$\leq np$
$\sum$		5	5	5	5	5	
		$\lvert\mathcal{F}\rvert \cdot p_1$	$\lvert\mathcal{F}\rvert \cdot p_2$	$\lvert\mathcal{F}\rvert \cdot p_3$	$\lvert\mathcal{F}\rvert \cdot p_4$	$\lvert\mathcal{F}\rvert \cdot p_5$	$\sum_{i=1}^{n} \lvert\mathcal{F}\rvert \cdot p_i \leq \lvert\mathcal{F}\rvert \cdot np$

Durch doppeltes Abzählen erhalten wir $\sum_{i=1}^{n} p_i \leq p_n$ und daraus $p_i \leq p$ für alle i. Weil die Funktion H(p) streng monoton steigend ist für $0 < p \leq \frac{1}{2}$ gilt:

$$H(p_i) \leq H(p) \text{ für alle i} \in \{1, \ldots, n\}$$

Mit Korollar 8 erhalten wir:

$$\sum_{i \leq np} \binom{n}{i} = |\mathcal{F}| \leq 2^{\sum_{i=1}^{n} H(p_i)} \leq 2^{n \cdot H(p)} \qquad \blacksquare$$

Die folgende einfache Konsequenz aus Satz 7 sagt uns, dass eine Familie nicht viele Mitglieder haben kann, wenn ihre „Projektionen“ klein sind.

Projektion: $\mathcal{F}_i : \mathcal{F} \to S_i$ mit $\mathcal{F}_i := \{E \cap S_i | E \in \mathcal{F}\}$

Für 1 gilt dann: $\mathcal{F}_1 : \mathcal{F} \to S_1$

$\mathcal{F}_1 := \{E \cap S_1 | E \in \mathcal{F}\}$

Schnitt von Teilmengen von {1,2,3} mit $S_1 = \{1, 2\}$:

$$\left.\begin{array}{c} \emptyset \cap S_1 = \emptyset \\ \{1\} \cap S_1 = \{1\} \\ \{2\} \cap S_1 = \{2\} \\ \{1,2\} \cap S_1 = S_1 \\ \{1,3\} \cap S_1 = \{1\} \\ \{1,2,3\} \cap S_1 = S_1 \end{array}\right\} \mathcal{F}_1 = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}$$

Satz 10 *Sei Ω eine endliche Menge und sei $S = \{S_1, \ldots, S_m\}$ Teilmengen von Ω, so dass jedes Element von Ω in mindestens k>0 Mitgliedern von S enthalten ist. Sei $\mathcal{F}$ eine Familie von Teilmengen von Ω. Für jedes $1 \leq i \leq m$ definieren wir die Projektion von $\mathcal{F}$ auf die Menge S_i durch $\mathcal{F}_i := \{E \cap S_i | E \in \mathcal{F}\}$. Dann*

$$|\mathcal{F}| \leq (\prod_{i=1}^{m} |\mathcal{F}_i|)^{\frac{1}{k}} \tag{7}$$

Beweis:
Nehmen wir an, dass $\Omega = \{1, \ldots, n\}$ und definieren $B_i = \{0, 1\}$ für $1 \leq i \leq n$. Sei $X = (X_1, \ldots, X_n)$ die Zufallsvariable, die Werte in $B = B_1 \otimes B_2 \otimes \cdots \otimes B_n$ annimmt. Für jedes $E \in \mathcal{F}$ sei die Zufallsvariable X mit Wahrscheinlichkeit $\frac{1}{|\mathcal{F}|}$ gleich dem Vorkommen-Vektor von E.

Nach Satz 7 gilt $k \cdot H[X] \leq \sum_{i=1}^{m} H[X_{S_i}]$. Im Beweis zu Korollar 8 haben wir gesehen, dass $H[X] = \log_2 |\mathcal{F}|$. wohingegen $H[X_{S_i}] \leq \log_2 |\mathcal{F}_i|$, wodurch das gewünschte Ergebnis impliziert.

$$\begin{aligned} H[X] &= \log_2 |\mathcal{F}| \\ &\leq \frac{1}{k} \sum_{i=1}^{m} H[X_{S_i}] \\ &\leq \frac{1}{k} \sum_{i=1}^{m} \log_2 |\mathcal{F}_i| \\ &= \frac{1}{k} \log_2 (\prod_{i=1}^{m} |\mathcal{F}_i|) \\ &= \log_2 (\prod_{i=1}^{m} |\mathcal{F}_i|)^{\frac{1}{k}} \end{aligned}$$

Da der Logarithmus streng monoton steigt folgt, dass

$$|\mathcal{F}| \leq \left(\prod_{i=1}^{m} |\mathcal{F}_i| \right)^{\frac{1}{k}} \qquad \blacksquare$$

Die Allgemeine Subadditivität der Entropie Funktion aus Satz 7 kann dazu benutzt werden, um einige nicht trivialen „Schnittmengen-Sätze “ zu beweisen. Die folgenden 3 Ergebnisse wurden von Chung, Frankl, Graham und Shearer 1986 [3]) gezeigt.

Erinnern wir uns, dass eine Familie $\mathcal{F}$ von Teilmengen von irgendeiner Menge Ω intersecting heißt, wenn folgendes gilt: $F \cap F' \neq \emptyset$ für alle $F, F' \in \mathcal{F}$.

Wenn $\mathcal{F}$ intersecting ist, dann $|\mathcal{F}| \leq 2^{|\Omega|-1}$, weil $\mathcal{F}$ nicht eine Menge enthalten kann und gleichzeitig ihr Komplement. Darüber hinaus ist diese Abschätzung optimal. (Man nehme die Familie aller Teilmengen von Ω, die einen Fixpunkt haben.)

Um die Frage etwas interessanter zu machen, können wir fordern, dass die Mitglieder von $\mathcal{F}$ nicht nur intersect sind, sondern dass diese Schnittmengen mindestens eine der gegebenen Konfigurationen enthält.

Betrachten wir zuerst ein einfaches Beispiel. Sei, wie zuvor $[n] = \{1, \ldots, n\}$

Satz 11 *Nehmen wir an, dass $\mathcal{F}$ eine Familie von Teilmengen von $[n]$ ist, so dass die Schnittmenge von irgendwelchen 2 Mitgliedern dieser Familie ein Paar von aufeinanderfolgenden Zahlen enthält, d.h. für alle $F, F' \in \mathcal{F}$ existieren einige $1 \leq i < n$ so dass $F \cap F' \supseteq \{i, i+1\}$. Dann*

$$|\mathcal{F}| \leq 2^{n-2}$$

Bemerkung:
Diese obere Schranke ist optimal: Sei $\mathcal{F}$ Familie aller Teilmengen, die die Menge {1,2} enthalten.

Beweis:
Sei S_0 und S_1 die Menge aller geraden bzw. ungeraden Zahlen in $[n]$.

Betrachte die Projektionen $\mathcal{F}_\epsilon := \{F \cap S_\epsilon | F \in \mathcal{F}\}$ von unserer Familie $\mathcal{F}$ auf diese zwei Mengen mit $\epsilon \in \{0, 1\}$.

Seien $G, G' \in \mathcal{F}_\epsilon$, dann hat ihre Schnittmenge mit dem Distributivgesetz die Form $G \cap G' = (F \cap S_\epsilon) \cap (F' \cap S_\epsilon) = (F \cap F') \cap S_\epsilon$ für irgendwelche $F, F' \in \mathcal{F}$. Nach Voraussetzung enthält $(F \cap F')$ ein Paar aufeinanderfolgender Zahlen und ist daher nicht leer. Daher ist jede der Mengen $\mathcal{F}_\epsilon$ intersecting und so gilt:

$$|\mathcal{F}_\epsilon| \leq 2^{|S_\epsilon|-1} \text{ für beide } \epsilon \in \{0, 1\}$$

Die Voraussetzungen von Satz 10 sind erfüllt und so erhalten wir mit k=1

$$|\mathcal{F}| \leq |\mathcal{F}_0| \cdot |\mathcal{F}_1| \leq 2^{|S_0|-1} \cdot 2^{|S_1|-1} = 2^{|S_0|+|S_1|-2} = 2^{n-2} \qquad \blacksquare$$

Satz 10 hat aber noch weitere Konsequenzen. Zum Beispiel nehme man Ω als Menge von allen $\binom{n}{2}$ Kanten eines vollständigen Graphen K_n. Wir betrachten die Teilmengen F von Ω und sehen diese als markierte Teilgraphen $G = ([n], E)$ von K_n an. Markiert heißt hier, dass wir keine isomorphen Graphen identifizieren.

Satz 12 *Nehmen wir an, dass $\mathcal{F}$ eine Familie von markierten Teilgraphen von K_n ist, so dass für alle $F, F' \in \mathcal{F}$ der Graph $F \cap F'$ keine isolierten Knoten besitzt. Dann*

$$|\mathcal{F}| \leq 2^{\binom{n}{2} - \frac{n}{2}}$$

Beweis:
Wähle S_i als einen Sterngraph am i-ten Knoten, d.h. S_i besteht aus allen n-1 Kanten $\{i, j\}, j \neq i$.

Klar ist, dass jede Kante in genau 2 Mengen von $S_1, \ldots, S_n$. Das ist die Kante, die die 2 gewählten Sterne aus $S_1, \ldots, S_n$ verbindet.

Betrachten wir die Projektion $\mathcal{F}_i := \{F \cap S_i | F \in \mathcal{F}\}$ von $\mathcal{F}$ auf diese Sterngraphen $i = 1, \ldots, n$.

Wir bemerken, dass wenn $G, G' \in \mathcal{F}_i$, dann hat ihre Schnittmenge die Form

$$\begin{aligned} G \cap G' &= (F \cap S_i) \cap (F' \cap S_i) \\ &\underset{Ass}{=} F \cap (S_i \cap F') \cap S_i \\ &\underset{Komm}{=} F \cap (F' \cap S_i) \cap S_i \\ &\underset{Ass}{=} (F \cap F') \cap (S_i \cap S_i) \\ &\underset{Idempotenz}{=} (F \cap F') \cap S_i \end{aligned}$$

für irgendwelche $F, F' \in \mathcal{F}$ und daher ist diese nicht leer, weil der Knoten i, nach Voraussetzung nicht isoliert sein kann im Teilgraph $F \cap F'$.

Da jedes $\mathcal{F}_i$ eine Familie von Teilmengen von S_i ist und die paarweisen Schnittmengen nicht leer sind, sind diese $\mathcal{F}_i$ intersecting und daher

$$|\mathcal{F}_i| \leq 2^{|S_i|-1} = 2^{n-1-1} = 2^{n-2} \text{ für alle } i = 1, \ldots, n \tag{8}$$

Wenden wir Satz 10 mit k=2 (Jede Kante ist in genau 2 Sterngraphen) an und beachten folgende Umrechnung,

$$\binom{n}{2} - \frac{n}{2} = \frac{n!}{2 \cdot (n-2)!} - \frac{n}{2} = \frac{n \cdot (n-1)}{2} - \frac{n}{2} = \frac{n \cdot (n-1-1)}{2} = \frac{n \cdot (n-2)}{2}$$

erhalten wir:

$$|\mathcal{F}| \underset{(7)}{\leq} (\prod_{i=1}^{n} |\mathcal{F}_i|)^{\frac{1}{2}} \underset{(8)}{\leq} (\prod_{i=1}^{n} 2^{n-2})^{\frac{1}{2}} = 2^{n \cdot (n-2)/2} = 2^{\binom{n}{2} - \frac{n}{2}} \qquad \blacksquare$$

Sagen wir, dass eine Familie $\mathcal{F}$ von Teilgraphen von K_n triangle-intersecting heißt, wenn $F \cap F'$ ein Dreieck enthält, für alle $F, F' \in \mathcal{F}$.

Satz 13 *Sei $n \geq 4$ eine gerade Zahl und sei $\mathcal{F}$ eine Familie von (markierten) Teilgraphen von K_n.*

Wenn $\mathcal{F}$ triangle intersecting ist, dann

$$|\mathcal{F}| \leq 2^{\binom{n}{2}-2}$$

Bemerkung:
Es ist nicht bekannt, ob diese Grenze optimal ist.

Beweis:
Wir wählen S_i $1 \leq i \leq m := \frac{1}{2} \cdot \binom{n}{\frac{n}{2}}$ so dass diese aus allen möglichen disjunkten Vereinigungen von 2 kompletten (markierten) Teilgraphen auf $\frac{n}{2}$ Knoten jedes Graphen. Das heißt jedes S_i hat die Form $K_U \cup K_{\bar{U}}$ für irgendeine Teilmenge von Knoten $U \subseteq \{1, \ldots, n\}$ mit $|U| = \frac{n}{2}$. Wir haben $\binom{n}{\frac{n}{2}}$ Möglichkeiten, um aus

n Knoten $\frac{n}{2}$ auszuwählen. Die gewählten Knoten sind in der Menge U und die nicht gewählten in $\bar{U}$.

Die S_i's sind bipartite Graphen. Wir definieren nun wie in den vorherigen Beweisen eine Projektion auf diese bipartiten S_i's: $\mathcal{F}_i := \{F \cap S_i | F \in \mathcal{F}\}$

Jede der Familien $\mathcal{F}_i$ ist intersecting, weil kein Dreieck komplett in einem bipartiten Graph liegen kann, nach dem Satz von König[6]. Deshalb

$$|\mathcal{F}_i| \leq 2^{|S_i|-1}$$

Jeder Graph S_i hat $s := 2 \cdot \binom{n/2}{2}$ Kanten (Disjunkte Vereinigung von 2 vollständigen Teilgrapen mit $\binom{n/2}{2}$).

Jede Kante von K_n ist in $k := \binom{n-2}{n/2}$ S_i's enthalten. Mit Satz 10 folgt

$$|\mathcal{F}| \leq (\prod_{i=1}^{m} 2^{|S_i|-1})^{\frac{1}{k}} = 2^{(s-1)m/k}$$

Substituieren wir die Werte von s, m und k, erhalten wir dass

$$\begin{aligned}
\frac{(s-1)m}{k} &= \frac{(2\binom{n/2}{2}-1)\cdot(\frac{1}{2}\binom{n}{n/2})}{\binom{n-2}{n/2}} \\
&= \frac{\binom{n}{2}! - (\frac{n}{2}-2)!}{(\frac{n}{2}-2)!} \cdot \frac{n!}{2\cdot(\frac{n}{2})!\cdot(\frac{n}{2})!} \cdot \frac{(\frac{n}{2})!\cdot(\frac{n}{2}-2)!}{(n-2)!} \\
&= \binom{n}{2} \cdot \frac{(\frac{n}{2})! - (\frac{n}{2}-2)!}{(\frac{n}{2})!} \\
&= \binom{n}{2} - \frac{\binom{n}{2}\cdot(\frac{n}{2}-2)!}{(\frac{n}{2})!} \\
&= \binom{n}{2} - \frac{\binom{n}{2}}{\frac{n}{2}\cdot(\frac{n}{2}-1)} \\
&= \binom{n}{2} - \frac{\binom{n}{2}}{\frac{n}{2}\cdot(\frac{n}{2}-1)} \\
&= \binom{n}{2} - \frac{n!}{2!\cdot(n-2)!\cdot\frac{n}{2}\cdot(\frac{n}{2}-1)} \\
&= \binom{n}{2} - \frac{n-1}{(\frac{n}{2}-1)} \\
&\leq \binom{n}{2} - \frac{n-1}{(\frac{n}{2}-\frac{1}{2})} \\
&= \binom{n}{2} - 2 \qquad \blacksquare
\end{aligned}$$

Literatur

[1] Mathar,Rudolf - Informationstheorie - Teubner, 1996
- ISBN 3-519-02574-4

[2] Ash,Robert - Information Theory - John Wiley & Sons, 1965

[3] Noga Alon - Probabilistic Methods in Extremal Finite Set Theory
http://www.math.tau.ac.il/∼nogaa/PDFS/set3.pdf

[4] Jaikumar Radhakrishnan - Entropy and Counting, 2001
http://www.tcs.tifr.res.in/∼jaikumar/Papers/EntropyAndCounting.ps

[5] Chung, Frankl, Graham and Shearer,
Some intersection theorems for ordered sets and graphs,
J. Combinatorial Theory, Ser. A 43 (1986), 23-37

[6] Lutz Volkmann - Diskrete Strukturen - Eine Einführung
ISBN 3-86073-647-7