Siegfried Wendt
University of Kaiserslautern
Germany

A More Natural Axiomatic Basis for the Entropy Formula in Information Theorie

In 1948, C.E.Shannon published his paper [1] ”A Mathematical Theory of Communication” which laid the foundation of what later was called *Information Theory*. In this paper, Shannon introduces the formula for the capacity of a discrete channel by definition. He says:

> ”Definition: The capacity C of a discrete channel is given by
>
> $$C = \lim_{T\to\infty} \frac{\log N(T)}{T}$$
>
> where $N(T)$ is the number of allowed signals of duration T. ”

Since then, in all lectures and textbooks on Information Theory, the logarithm in the formula for the capacity of a channel or in the formula for the entropy of a signal generator is introduced by definition, and it is made plausible afterwards that the logarithm was a reasonable choice. Many students feel uncomfortable when they see these formulae for the first time; they have the impression that the logarithm is pulled out like a rabbit from a magician’s hat.

There is no need to introduce these formulae by definition. They can be derived from a much more natural axiomatic basis. This is presented in the following paper.

---

1) C. E. Shannon: A Mathematical Theory of Communication.
Bell System Technical Journal, vol. 27, pp.379–423, 623–656; 1948.

Siegfried Wendt
Universität Kaiserslautern

Quantität der Information

1. Der Ansatz

Es gehört zum Informationsbegriff und seinem Umfeld, daß man die Frage nach der Quantifizierbarkeit der Information untersucht – in Analogie zur Quantifizierbarkeit der Materie in den Maßeinheiten Kilogramm oder Kubikmeter. Information als das Wißbare erlaubt die Anwendung der unscharfen und relativen Quantitätsmaße "viel" und "wenig"; so ist es umgangssprachlich selbstverständlich zu sagen, daß einer viel oder wenig von einer Sache wisse, d.h. daß einer viel oder wenig Information über diese Sache habe. Dies ist Anlaß genug, nach einem scharfen Quantitätsmaß für die Information zu suchen. Dies ist nicht nur eine Frage von rein akademischem Interesse, sondern hat durchaus auch eine praktische Bedeutung für den Bereich der informationstechnischen Systeme; das Fassungsvermögen informationstechnischer Speicher und die Durchsatzgrenzen informationstechnischer Kanäle lassen sich nämlich ohne ein solches Quantitätsmaß nicht sinnvoll erfassen.

Die Frage nach der Quantität der Information bildet den Kern der sogenannten *Informationstheorie*, die 1948 von C.E. Shannon begründet wurde. Insbesondere im Bereich der technischen Nachrichtenübertragungssysteme hat die Informationstheorie zu einem besseren Verständnis der Möglichkeiten und Grenzen der verschiedenen Systemklassen geführt. So kann beispielsweise die qualitative Überlegenheit des UKW–Rundfunks gegenüber dem Mittelwellenrundfunk in der Rundfunktechnik leicht mit Hilfe der Informationstheorie erklärt werden. Im vorliegenden Aufsatz werden allerdings die Anwendungsbereiche der Informationstheorie nicht weiter betrachtet. Es geht hier ausschließlich darum, den Begriff der Informationsquantität zu klären.

Die Vorstellung, daß man Qualität und Quantität trennen kann, wird durch unsere Anschauung der materiellen Welt begründet: Die Definition des Kubikmeters ist unabhängig von der Art der Füllung, sei es Sand, Milch oder Vakuum. Der grundlegende Gedanke, der die Trennung von Qualität und Quantität auch im Falle der Information erlaubt, ist der folgende: Durch die Beantwortung einer *Binärfrage* wird immer das gleiche Quantum an Information übertragen unabhängig davon, wie die Frage lautet und welche der beiden möglichen Antworten gegeben wird. Eine Frage ist genau dann eine Binärfrage, wenn es nur zwei mögliche Antworten gibt, die dem Fragesteller bekannt sind. Der Gefragte kann natürlich auch erwidern, er wisse nicht, welche der beiden möglichen Antworten die richtige sei. Durch diese Erwiderung bleibt die gestellte Binärfrage unbeantwortet und es wird eine andere, nicht explizit gestellte Binärfrage beantwortet: "Können Sie die gestellte Binärfrage beantworten?"

> Man kann sich keine Frage ausdenken, bei deren Beantwortung weniger Information übertragen wird als bei der Beantwortung einer Binärfrage.
> Deshalb wird das Informationsquantum, das bei der Beantwortung einer Binärfrage übertragen wird, als *Maßeinheit der Informationsquantität* festgelegt.

Diese Maßeinheit wird *bit* (basic indissoluble information unit) genannt. Obwohl es keine kleinere Informationsmenge als 1 bit geben kann, gibt es doch sinnvolle Aussagen, worin von nicht ganzzahligen Informationsquantitäten die Rede ist. In diesem Fall kann es sich nur um Aussagen über Mittelwerte handeln. Man betrachte hierzu die folgende Analogie: Obwohl Menschen nur als ganze Einheiten vorkommen, kann man doch sagen: Aus jeder Ehe, die in Deutschland im Jahre 1955 geschlossen wurde, gingen im Mittel 1,9 Kinder hervor.

Selbstverständlich haben wir nicht all unser Wissen in Form von Antworten auf Binärfragen erworben. Um Binärfragen stellen zu können, muß man schon Wissen haben. Um auch dieses Basiswissen quantifizieren zu können, müßte man die Frage beantworten können, welche Alternativen zu diesem Basiswissen es denn gebe. Diese Frage kann sich der Mensch aber nicht beantworten, denn sie lautet ja: ”Was könnte ich wissen anstelle dessen, was ich weiß?” Um diese Frage beantworten zu können, müßte ich etwas wissen können, wovon ich auch noch wüßte, daß ich es nicht weiß. Dies ist offensichtlich zuviel verlangt. Da der Mensch nun also diese Frage nicht beantworten kann, bleibt sein Basiswissen für ihn unquantifizierbar. Man könnte nun natürlich die Frage stellen, ob denn nicht vielleicht das Basiswissen des Menschen für ein Wesen außerhalb der menschlichen Vorstellung quantifizierbar sei, ob also ein solches Wesen die Frage beantworten könne: ”Was könnte der Mensch wissen anstelle dessen, was er weiß?” Diese Frage kann nur von einem Wesen beantwortet werden, welches weiß, wie die Welt sein könnte, wenn sie nicht wäre, wie sie ist. Diese Überlegungen führen in den Bereich der Religion und werden deshalb nicht vertieft.

Es bleibt aber festzuhalten, daß eine Quantifizierung der Information nur möglich ist unter der Voraussetzung eines nicht quantifizierten Basiswissens, welches uns ermöglicht, Binärfragen zu stellen. Unser *Basiswissen* legt einen potentiellen Wissensbereich fest, innerhalb dessen wir unser Wissen durch Binärfragen vermehren können.

Die Informationstheorie baut auf der Erkenntnis auf, daß sich Information immer als Form äußert – als Signalform bzw. als Symbolform – und daß deshalb einer Wissensvielfalt äußerlich eine Formenvielfalt entsprechen muß. *Informieren* bedeutet in diesem Sinne dann nur noch die Identifikation einer bestimmten *Form* aus einem vorab bekannten Repertoire von Formen.

Ein weiteres Merkmal der Informationstheorie besteht darin, daß angenommen wird, daß immer wieder eine neue Auswahl eines Elements aus dem gleichen Repertoire erfolgen muß, so daß man theoretisch eine unendliche Folge von Elementen betrachtet. Ein typisches Beispiel hierfür ist ein Text – beispielsweise der vorliegende Aufsatz –, der eine fast unendliche Folge von Elementen aus dem Repertoire der Schriftzeichen bildet. Eine solche Folge kann eine bestimmte wahrscheinlichkeitstheoretische Charakteristik haben, und diese soll zum Basiswissen gehören.

Durch die Binärfragen soll also das Wissen erfaßt werden, um welche *aktuelle Folge* aus der Menge all derjenigen Folgen es sich handelt, auf welche die *vorgegebene Charakteristik* zutrifft. Da die Folge unendlich ist, braucht man auch unendlich viele Binärfragen, um die Folge zu erfragen. Man interessiert sich jedoch nicht für diese unendlich vielen Binärfragen, sondern nur für das *arithmetische Mittel der Anzahl von Binärfragen*, die pro Element der Folge gebraucht werden. Dieser Mittelwert wird *Entropie* [1)] genannt. Wie die Entropie H aus einer gegebenen Charakteristik berechnet werden kann, soll nun hergeleitet werden.

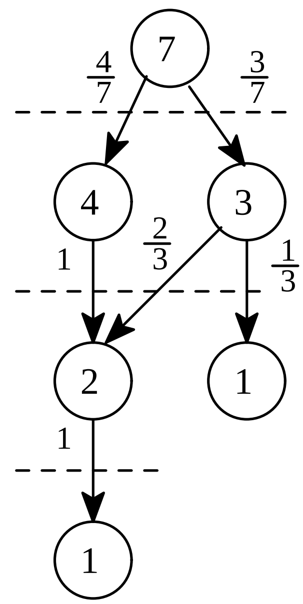
1) Diese Bezeichnung weist auf einen formalen Zusammenhang mit dem Entropiebegriff in der Thermo– dynamik hin, auf den aber hier nicht eingegangen wird.

2. Gleichverteilung und Entkopplung

Zuerst wird ein einfaches, anschauliches Beispiel betrachtet. Man stelle sich zwei Familien vor, die sich seit urdenklichen Zeiten über alle Generationen hinweg jeweils einmal pro Woche zu einem gemeinsamen Essen getroffen haben und die diese Tradition auch in Zukunft unbegrenzt fortsetzen wollen. Die jeweiligen Tage dieser Begegnungen bilden also eine Folge von Elementen aus dem Typenrepertoire der sieben Wochentage. Als Charakteristik dieser Folge soll nun zuerst angenommen werden, daß für alle Wochentage die gleiche Wahrscheinlichkeit gilt und daß die jeweilige Wahl eines Wochentages völlig unabhängig sei vom bisherigen Verlauf der Folge. Es wird also eine Zufallsfolge von 7 gleichwahrscheinlichen und entkoppelten Alternativen angenommen.

Wenn man nun jeweils durch Binärfragen den Wochentag erfragen will, an dem das nächste Essen stattfindet, dann muß man sich für ein Frageschema entscheiden. Bild 1 zeigt ein optimales Frageschema.

<table>
<tr><td></td><td colspan="7">Ist es einer der ersten vier Werktage ?</td><td></td></tr>
<tr><td></td><td colspan="4">ja</td><td colspan="3">nein</td><td>1. Frage</td></tr>
<tr><td></td><td colspan="4">Ist es einer der ersten beiden Werktage ?</td><td colspan="3">Ist es der Sonntag ?</td><td></td></tr>
<tr><td></td><td colspan="2">ja</td><td colspan="2">nein</td><td colspan="2">nein</td><td>ja</td><td>2. Frage</td></tr>
<tr><td></td><td colspan="2">Ist es der Montag?</td><td colspan="2">Ist es der Mittwoch?</td><td colspan="2">Ist es der Freitag ?</td><td rowspan="3">So</td><td></td></tr>
<tr><td></td><td>ja</td><td>nein</td><td>ja</td><td>nein</td><td>ja</td><td>nein</td><td>3. Frage</td></tr>
<tr><td>Erfragter Tag:</td><td>Mo</td><td>Di</td><td>Mi</td><td>Do</td><td>Fr</td><td>Sa</td><td></td></tr>
</table>



$$F = \left(\frac{3}{7}\cdot\frac{1}{3}\right)\cdot 2 + \left(\frac{3}{7}\cdot\frac{2}{3}\cdot 1 + \frac{4}{7}\cdot 1\cdot 1\right)\cdot 3 = \frac{20}{7} = 2{,}857$$

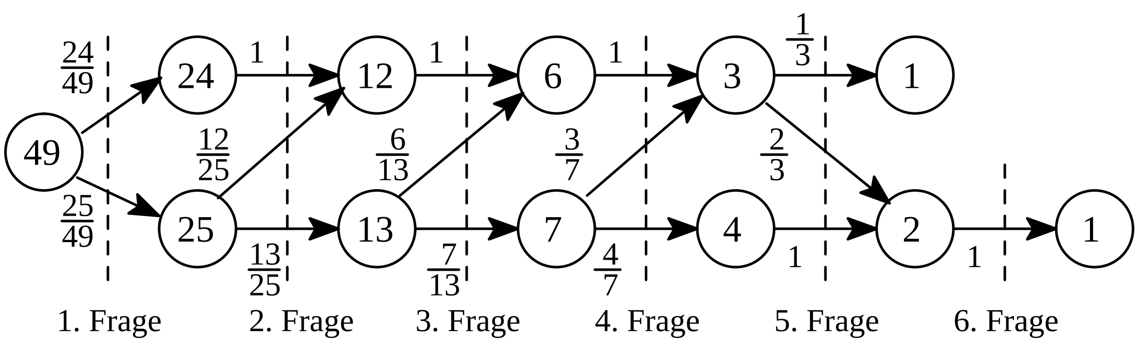
Bild 1 Optimales Frageschema für sieben gleichwahrscheinliche Wochentage

Die Fragen werden hier in Abhängigkeit von den bisherigen Antworten jeweils so gestellt, daß von den verbliebenen noch möglichen Wochentagen möglichst die eine Hälfte auf die Ja–Antwort und die andere Hälfte auf die Nein–Antwort entfallen. Zusätzlich zu der Fragetabelle ist auch noch ein Graph angegeben, der das Schema unabhängig vom Wortlaut der Fragen charakterisiert: In den Knoten ist jeweils die Anzahl der noch verbliebenen Alternativen eingetragen, unter denen man durch die folgenden Fragen noch eine Auswahl treffen muß. An den Kanten ist jeweils die Wahrscheinlichkeit angegeben für diejenige Binärantwort, die in diese Richtung führt.

Wenn man sich nun für ein bestimmtes Frageschema entscheidet und mit diesem alle zukünftigen Elemente der Folge erfragt, dann erhält man eine zu diesem Frageschema gehörige mittlere Anzahl $\overline{F}$ von Binärfragen pro Folgenelement. Diesen Mittelwert $\overline{F}$ kann man leicht aus dem Graphen berechnen: Jeder gerichtete Weg vom Anfangsknoten des Graphen zu einem Endknoten bringt einen Beitrag zu $\overline{F}$, der sich berechnet als Produkt aus den zugehörigen Kantenwahrscheinlichkeiten und der Anzahl der Kanten auf dem Weg. Denn die Anzahl der Kanten auf dem Weg ist gleich der Anzahl der für diesen Weg erforderlichen Fragen, und das Produkt der zugehörigen Kantenwahrscheinlichkeiten ist gleich der Wahrscheinlichkeit, daß die Auswahl längs dieses Weges erfolgt. Der Wert $\overline{F}$ ergibt sich somit als Summe der Beiträge aller unterschiedlichen Wege. Das Frageschema in Bild 1 ist optimal, d.h. es gibt kein Frageschema, das einen kleineren Wert für $\overline{F}$ liefert als dieses optimale Schema.

Der Wert $\overline{F}$ aus dem optimalen Schema in Bild 1 ist aber nicht gleich der Entropie H, d.h. er ist nicht gleich dem gesuchten Mittelwert der Anzahl der pro Folgenelement *mindestens* gebrauchten Binärfragen. Mit dem Schema in Bild 1 wird nämlich jedes Folgenelement einzeln erfragt, und auf diese Weise benötigt man im Mittel pro Folgenelement mehr Binärfragen, als wenn man ein Frageschema anwendet, mit dem man gleich mehrere aufeinanderfolgende Elemente erfragt. Bild 2 zeigt ein optimales Frageschema zur Erfragung von jeweils zwei aufeinanderfolgenden Elementen; hier wird also ein Paar von Essensterminen für zwei aufeinanderfolgende Wochen als eine von 49 gleichwahrscheinlichen Alternativen erfragt. Die hierbei auf einen Essenstermin entfallende mittlere Anzahl von Binärfragen ist kleiner als in Bild 1.

Die auf einen Essenstermin entfallende mittlere Anzahl F von Binärfragen wird immer kleiner, je mehr Essenstermine für aufeinanderfolgende Wochen man gemeinsam mit einem optimalen Frageschema erfragt. Durch Grenzwertbildung erhält man die Entropie H als minimale mittlere Anzahl von Binärfragen.



$$F = \left[\frac{24}{49}\cdot\frac{1}{3} + \frac{25}{49}\cdot\frac{12}{25}\cdot\frac{1}{3} + \frac{25}{49}\cdot\frac{13}{25}\cdot\frac{6}{13}\cdot\frac{1}{3} + \frac{25}{49}\cdot\frac{13}{25}\cdot\frac{7}{13}\cdot\frac{3}{7}\cdot\frac{1}{3}\right]\cdot 5 \; +$$

$$+\left[\frac{24}{49}\cdot\frac{2}{3} + \frac{25}{49}\cdot\frac{12}{25}\cdot\frac{2}{3} + \frac{25}{49}\cdot\frac{13}{25}\cdot\frac{6}{13}\cdot\frac{2}{3} + \frac{25}{49}\cdot\frac{13}{25}\cdot\frac{7}{13}\cdot\frac{3}{7}\cdot\frac{2}{3} + \frac{25}{49}\cdot\frac{13}{25}\cdot\frac{7}{13}\cdot\frac{4}{7}\right]\cdot 6 = \frac{279}{49} = 2\cdot 2{,}847$$

Bild 2 Optimales Frageschema für 49 gleichwahrscheinliche Alternativen

In der folgenden Betrachtung wird die Symbolisierung $\overline{F}(r, a)$ benutzt für die mittlere Anzahl von Binärfragen pro Folgenelement, die man erhält, wenn das Repertoire die Mächtigkeit r hat und man mit einem optimalen Frageschema a aufeinanderfolgende Folgenelemente gleichzeitig erfragt. Es gilt:

$$\overline{F}(r, a) = \frac{\overline{F}(r^a, 1)}{a}$$

Bei gleichverteilten Folgeelementen gibt es immer ein optimales Frageschema, bei dem nur (k–1) oder k Fragen zum Ziel führen (s. Bild 3). Für ein solches Frageschema gilt also

$$k-1 < F(r^a, 1) \leq k$$

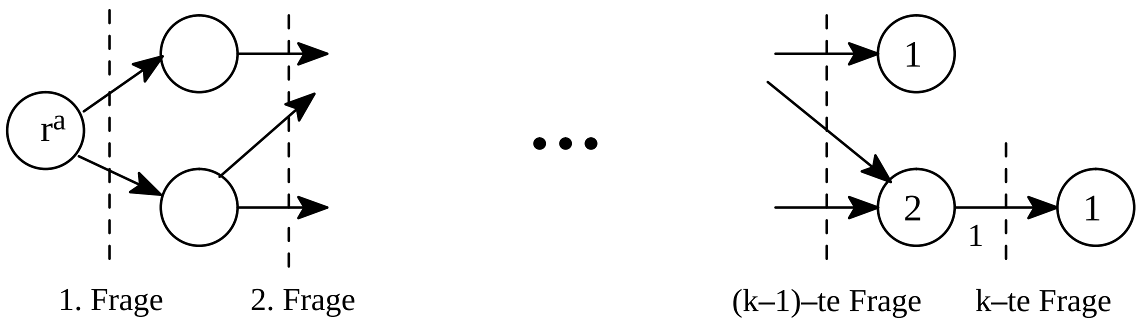


Bild 3 Optimales Frageschema für r^a gleichwahrscheinliche Alternativen

Zwischen r, a und k gilt die Beziehung

$$2^{k-1} < r^a \leq 2^k$$

Dies liegt daran, daß im optimalen Frageschema die Fragen immer so gestellt werden, daß dadurch jeweils die Anzahl der verbleibenden Alternativen möglichst halbiert wird. Daraus ergeben sich die folgenden Konsequenzen:

$$k-1 < a \cdot \mathrm{ld}\, r \leq k \quad ^{1)}$$

$$\frac{k-1}{a} < \overline{F}(r, a) \leq \frac{k}{a} \; ; \qquad \frac{k-1}{a} < \mathrm{ld}\, r \leq \frac{k}{a}$$

$$\boxed{H = \lim_{a \to \infty} \overline{F}(r, a) = \mathrm{ld}\, r}$$

In dem betrachteten Beispiel ist r = 7 und ld 7 = 2,8074. Die Werte aus den Bildern 1 und 2, also $\overline{F}(7, 1)=2{,}857$ und $\overline{F}(7, 2)=2{,}847$ liegen schon recht dicht bei diesem Grenzwert.

---

1) Mit ”ld” wird der Logarithmus zur Basis 2 bezeichnet : $2^{\mathrm{ld}\, x} = x$ und $\mathrm{ld}(2^y) = y$

3. Ungleichverteilung und Entkopplung

In der bisherigen Betrachtung wurde eine sehr einfache wahrscheinlichkeitstheoretische Charakterstik angenommen, nämlich *Gleichverteilung* und *Entkopplung der Alternativen*. Nun soll die Charakteristik derart modifiziert werden, daß sie einen Schritt näher an den allgemeinsten Fall rückt: Es wird zwar noch vorläufig die Entkopplung der Alternativen beibehalten, aber die Gleichverteilung wird aufgegeben.

Die Herleitung der Formel für die Entropie H für diese Art von wahrscheinlichkeitstheoretischer Charakteristik soll wieder durch die Betrachtung eines Beispiels eingeleitet werden. Dabei wird immer noch angenommen, es gehe um die wöchentliche Auswahl eines Wochentages für ein gemeinsames Abendessen. Es wird nun aber die Wahrscheinlichkeitsverteilung in Bild 4 angenommen, d.h. es wird angenommen, daß die Sonntage mit deutlich größerer Wahrscheinlichkeit vorkommen als die gleichverteilten Werktage zusammen. Für diese Wahrscheinlichkeitsverteilung ist nun nicht mehr das Frageschema in Bild 1, sondern das Frageschema in Bild 4 optimal, wo mit der ersten Frage gleich entschieden wird, ob der zu erfragende Tag ein Sonntag ist. Es ergibt sich nun ein Wert für $\overline{F}$, der kleiner ist als ld 7.

Auch im Fall der Nichtgleichverteilung der Alternativen kann man die mittlere Anzahl der Fragen, die auf ein Element in der Folge entfallen, noch dadurch weiter senken, daß man ein Frageschema anwendet, mit dem gleich mehrere aufeinanderfolgende Elemente erfragt werden. Für den Wert $\overline{F}(r,a)$ gibt es auch hier wieder einen Grenzwert, der aus der vorgegebenen Folgencharakteristik berechnet werden kann. Man findet die Berechnungsvorschrift durch folgende Überlegung: Die Berechnungsvorschrift für den Fall der Gleichverteilung ist bekannt. Also versucht man, den Fall der Ungleichverteilung auf den Fall der Gleichverteilung zurückzuführen. Man macht deshalb aus der Ungleichverteilung künstlich eine Gleichverteilung, indem man künstliche Alternativen einführt, unter denen man eigentlich gar nicht auswählen muß.

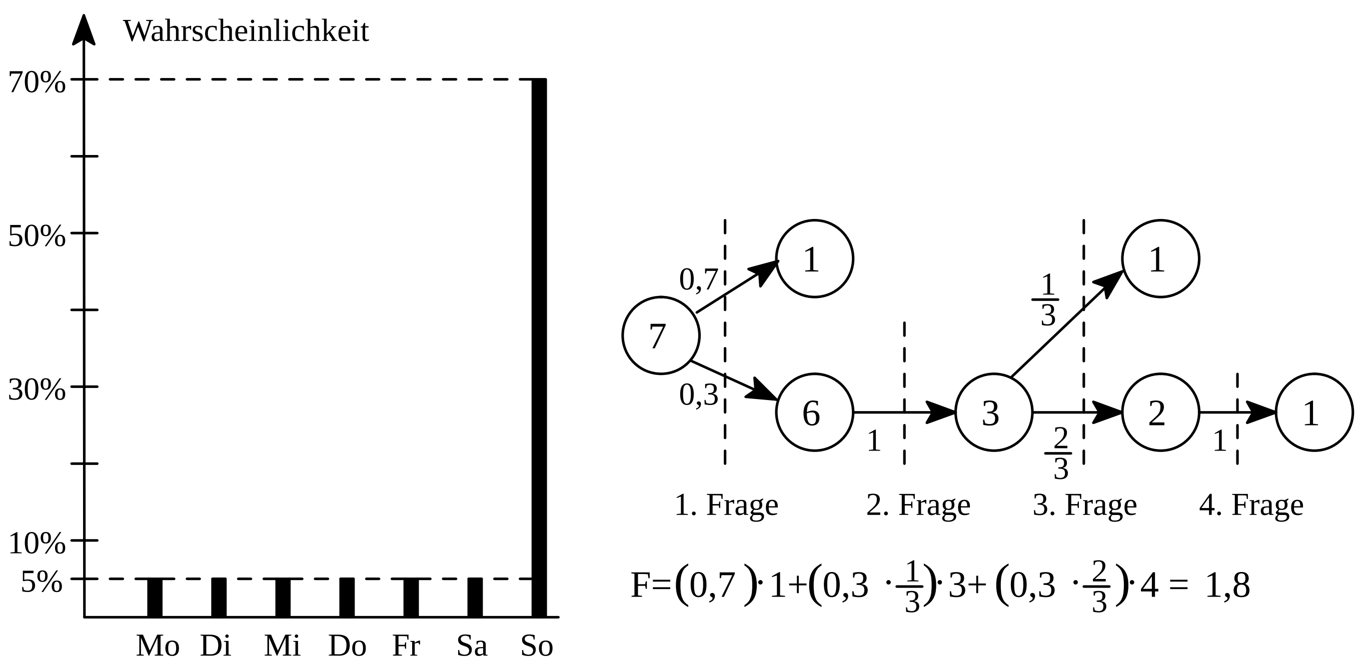


Bild 4 Wahrscheinlichkeitsverteilung der Elemente in der Folge und zugehöriges Frageschema

Im Falle der Verteilung in Bild 4 erhält man eine solche künstliche Gleichverteilung, indem man 14 alternative Sonntage einführt, von denen jeder mit einer Wahrscheinlichkeit von 5% auftritt, also mit der gleichen Wahrscheinlichkeit wie jeder Werktag. Dadurch hat man die Gesamtzahl der Alternativen künstlich von 7 auf 20 erhöht. Der Grenzwert für die mittlere Anzahl von Binärfragen zu diesen 20 gleichwahrscheinlichen Alternativen beträgt ld 20. Im Falle eines Werktags ist die aus den 20 Alternativen ausgewählte tatsächlich auch eine echte und keine künstliche Alternative. In diesen 30% der Fälle hat man also nicht unnötig viel gefragt. In den restlichen 70% der Fälle hätte man aber zuviel gefragt; man hätte nämlich eine Alternative aus 14 künstlichen erfragt. Da die 14 künstlichen Sonntagsalternativen gleichverteilt sind, bräuchte man als Grenzwert eine mittlere Anzahl von ld 14 Binärfragen für die Auswahl. Diese ld 14 Binärfragen hätte man also zuviel gefragt, wenn man mit ld 20 Binärfragen eine Sonntagsalternative erfragt hätte. Also ist man nach (ld 20 – ld 14) Binärfragen im Mittel bereits mit der Fragerei fertig, falls das aktuelle Element ein Sonntag ist, während man im anderen Falle tatsächlich im Mittel ld 20 Binärfragen braucht, bis man weiß, welcher Werktag es ist. Somit ergibt sich als Entropie zu der Verteilung in Bild 4 das gewichtete Mittel aus den Grenzwerten für die Erfragung eines Werktags bzw. eines Sonntags:

$$H = 0{,}3 \cdot (\, \mathrm{ld}\ 20\,) + 0{,}7 \cdot (\, \mathrm{ld}\ 20 - \mathrm{ld}\ 14\,) = 1{,}657$$

Der Wert von $\overline{F}$ in Bild 4 liegt schon recht dicht an diesem Grenzwert.

Die Art und Weise, wie in diesem Beispiel die Entropie berechnet wurde, beruht auf der Voraussetzung, daß alle Wahrscheinlichkeitswerte p_i der einzelnen Alternativen A_i aus dem gegebenen Repertoire rationale Zahlen sind. Nur dann gibt es einen Wahrscheinlichkeitswert p_0, aus dem sich die Wahrscheinlichkeiten p_i durch Multiplikation mit ganzzahligen Faktoren m_i gewinnen lassen. Die Wahrscheinlichkeit einer echten Alternative A_i ist dann also $p_i = m_i \cdot p_0$.

Es gilt der Zusammenhang

$$\sum_{i=1}^{r} m_i = \frac{1}{p_0}$$

Nun kann man anstelle jeder einzelnen Alternative A_i aus dem gegebenen Repertoire eine individuelle Anzahl m_i fiktiver Alternativen betrachten, für die alle die einheitliche Wahrscheinlichkeit p_0 gilt.

Damit ergibt sich die Entropie des fiktiven Falls zu

$$H_{\mathrm{fiktiv}} = \mathrm{ld}\ \frac{1}{p_0}$$

Um die tatsächliche Entropie zu erhalten, muß man den Wert für die überflüssigen Fragen zur Unterscheidung der fiktiven Alternativen subtrahieren:

$$H = H_{fiktiv} - \sum_{i=1}^{r} p_i \cdot ld\, m_i = ld \frac{1}{p_0} - \sum_{i=1}^{r} p_i \cdot ld \frac{p_i}{p_0}$$

$$= ld \frac{1}{p_0} - \sum_{i=1}^{r} p_i \cdot ld\, p_i - \sum_{i=1}^{r} p_i \cdot ld \frac{1}{p_0}$$

$$= ld \frac{1}{p_0} + \sum_{i=1}^{r} p_i \cdot ld \frac{1}{p_i} - ld \frac{1}{p_0} \sum_{i=1}^{r} p_i$$

$$\boxed{H = \sum_{i=1}^{r} p_i \cdot ld \frac{1}{p_i}}$$

Da in dieser Formel kein p_0 vorkommt, ist die für die Herleitung angenommene Einschränkung, daß alle p_i rationale Zahlen sein sollen, für die Gültigkeit der Formel irrelevant.

4. Ungleichverteilung und Verkopplung

Nun wird in einem letzten Schritt die angenommene Folgencharakteristik noch weiter verallgemeinert, indem Bindungen zwischen den Elementen der Folge eingeführt werden.

Wenn es Bindungen zwischen den Folgenelementen gibt, dann kann man die jeweils aktuell vergangenen Folgenverläufe in Klassen einteilen, wobei zu jeder Klasse k eine andere Wahrscheinlichkeitsverteilung V_k für die Alternativen des nächsten zu erfragenden Folgenelements gehört. Die Wahrscheinlichkeit, daß ein vergangener Folgenverlauf zur Klasse k gehört, ist p_k. Dieser Wert p_k ist somit auch die Wahrscheinlichkeit dafür, daß die Verteilung V_k für die Erfragung des nächsten Folgenelements relevant wird. Die jeweilige Verteilung V_k gibt für jede der r Alternativen im Repertoire eine Wahrscheinlichkeit p_{ki} an. Wenn die Verteilung V_k relevant ist, tritt also die i–te Alternative mit der Wahrscheinlichkeit p_{ki} als nächstes Folgenelement auf.

Zu jeder Verteilung V_k kann man einen Entropiewert H_k berechnen, der gleich der Entropie H der ganzen Folge wäre, wenn es nur eine einzige Klasse für alle möglichen vergangenen Folgenverläufe gäbe, d.h. wenn die Verteilung V_k die einzige wäre, die man betrachten müßte. Wenn aber mehrere Verteilungen V_k betrachtet werden müssen, von denen jede nur mit ihrer eigenen Wahrscheinlichkeit p_k relevant wird, dann trägt diese Verteilung V_k auch nur mit einem entsprechenden Bruchteil von H_k, nämlich mit $p_k \cdot H_k$ zur gesamten Entropie H bei.

$$\boxed{H = \sum_{k=1}^{\text{Anzahl der Verteilungen}} p_k \cdot H_k = \sum_{k=1}^{\text{Anzahl der Verteilungen}} p_k \cdot \left(\sum_{i=1}^{r} p_{ki} \cdot ld \frac{1}{p_{ki}} \right)}$$