Grundbegriffe der Informationstheorie

Information, Entropie und Redundanz

Aus der täglichen Erfahrung wissen wir: Eine Information (Nachricht) hat einen Neuigkeitswert - sie ist überraschend. Tatsachen, die wir bereits kennen, stellen für uns keine Information dar. Diese Überlegung führt zur Wahrscheinlichkeit, denn Ereignisse mit geringem Informationsgehalt, d.h. Ereignisse die man erwartet, haben eine hohe Wahrscheinlichkeit. Umgekehrt haben unerwartete, überraschende Ereignisse, also Ereignisse mit einer kleinen Wahrscheinlichkeit, einen hohen Informationsgehalt. Informationsgehalt und Wahrscheinlichkeit stehen in gegenläufigem Zusammenhang. Auf diese Erfahrung bauend entwickelte Claude E. Shannon 1948 einen mathematischen Informationsbegriff der folgende typische Fragestellung beantwortet:

> „Eine diskrete (Nachrichten-) *Quelle X* mit dem Zeichenvorrat $X = \{x_1, x_2, \ldots, x_N\}$ sende pro Zeitschritt ein Zeichen. Die Wahrscheinlichkeit (probability) des i-ten Zeichens x_i sei p_i. Welchen Informationsgehalt hat das i-te Zeichen?”

Der Informationsbegriff fußt auf einer axiomatischen Definition

Axiom 1: Der Informationsgehalt I eines Zeichens $x_i \in X$ mit der Wahrscheinlichkeit p_i ist ein nichtnegatives Maß.

$$I(p_i) \geq 0$$

Axiom 2: Die Informationsgehalte unabhängiger Zeichen (x_i, x_j) mit der Verbundwahrscheinlichkeit $p_{i,j} = p_i \cdot p_j$ addieren sich.

$$I(p_{i,j}) = I(p_i) + I(p_j)$$

Axiom 3: Der Informationsgehalt ist eine stetige Funktion der Wahrscheinlichkeiten der Zeichen.

Anmerkung: Axiom 3 bedeutet, dass eine kleine Änderungen der Auftrittswahrscheinlichkeit nur zu einer kleinen Änderung des Informationsgehaltes führen soll.

Im zweiten Axiom wird aus dem Produkt der Wahrscheinlichkeiten die Addition der Informationsgehalte. Dies führt direkt zur Logarithmusfunktion, die die Multiplikation in die Addition abbildet. Man definiert:

Der Informationsgehalt eines Zeichens mit der Wahrscheinlichkeit p ist

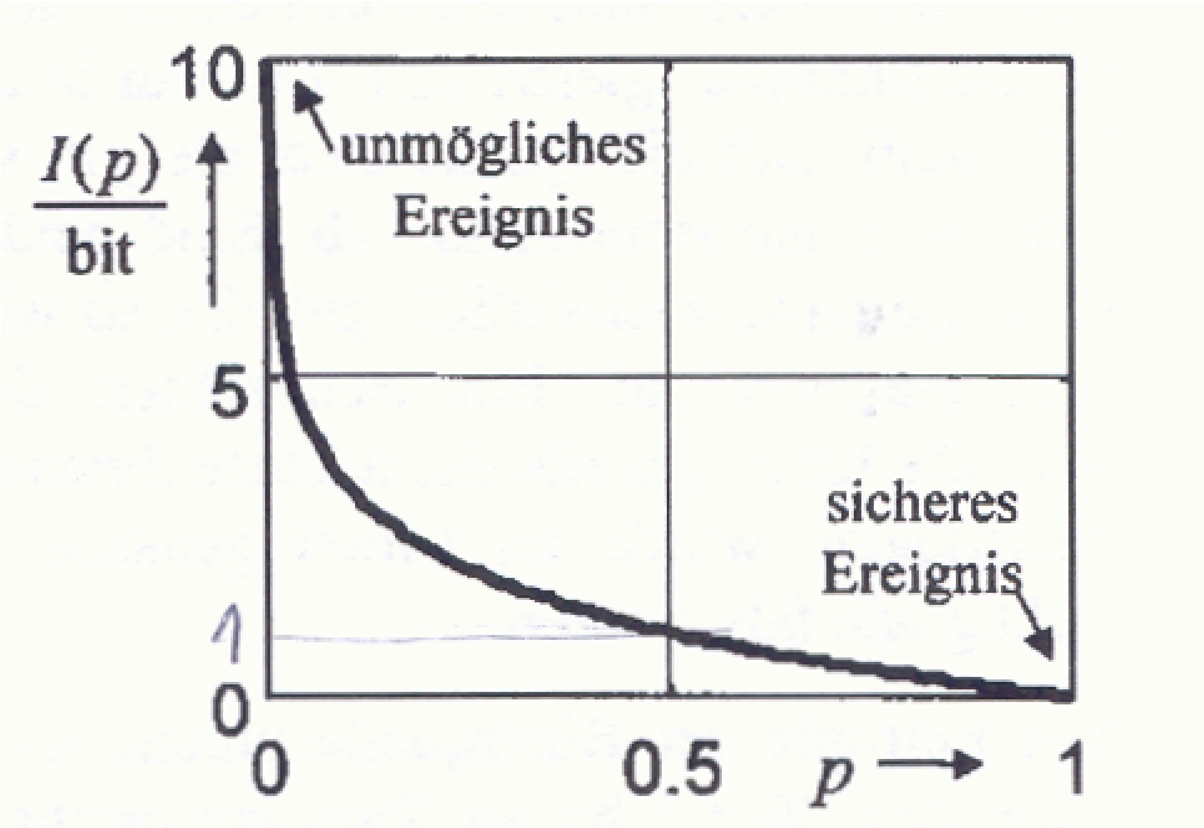
$$I(p) = -ld(p) \text{ mit } [E] = bit.$$

Es wird meist der Zweier-Logarithmus *ld( )* in Verbindung mit der Pseudoeinheit *”bit”* verwendet. Übliche Schreibweisen sind auch $log_2(x) = 1d(x) = lb(x)$. .

Anmerkung: Die Umrechnung der log-Funktion zu verschiedenen Basen erfolgt mit:

$$log_a(x) = log_b(x)/log_b(a) = log_b(x) \cdot log_a(b).$$

Der Informationsgehalt des sicheren Ereignisses *(p=l)* ist null. Mit wachsender Unsicherheit nimmt der Informationsgehalt stetig zu, bis schließlich im Grenzfall des unmöglichen Ereignisses *(p=0)* der Informationsgehalt gegen unendlich strebt. Der Informationsgehalt spiegelt die eingangs gemachten Überlegungen wieder und erfüllt offensichtlich auch die Axiome 1 und 3



Informationsgehalt $I(p)$ eines Zeichens mit der Auftrittswahrscheinlichkeit p

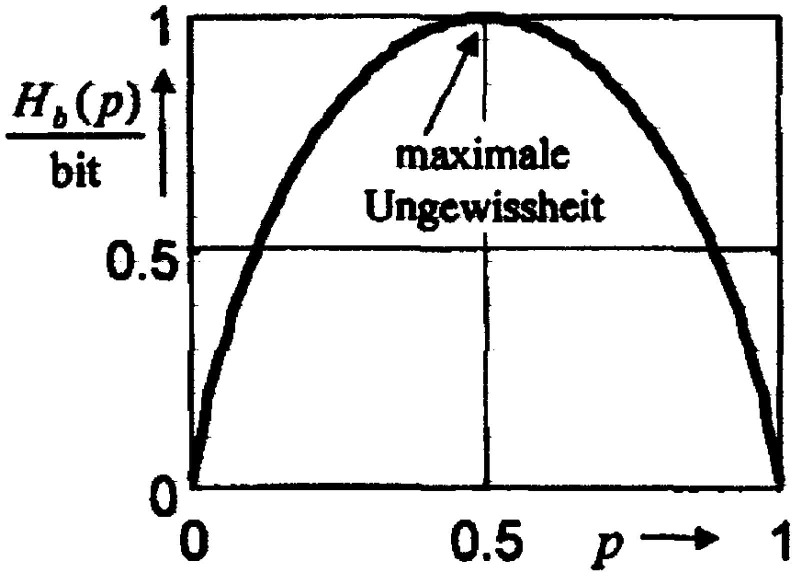
Der Informationsgehalt der Quelle kann mit nun als Erwartungswert der Informationsgehalte aller Zeichen bestimmt werden. Man spricht vom mittleren Informationsgehalt oder in Anlehnung an die Thermodynamik von der Entropie der Quelle. Für den einfachsten Fall einer (endlichen) ***diskreten gedächtnislosen Quelle***, bei der die Wahrscheinlichkeiten der einzelnen Zeichen nicht von den vorhergehenden Zeichen abhängen, definiert man:

> Eine diskrete, gedächtnislose Quelle X mit dem Zeichenvorrat (Alphabet) $X = \{x_1, x_2, \ldots, x_N\}$ und den zugehörigen Wahrscheinlichkeiten $p_1, p_2, \ldots, p_N$ besitzt die ***Entropie***

$$H(X) = -\sum_{i=1}^{N} p_i \cdot ld(p_i)$$

Das einfachste Beispiel einer diskreten gedächtnislosen Quelle ist die ***Binärquelle*** mit dem Zeichenvorrat $X=(0,1)$ und den Wahrscheinlichkeiten $p_0 = p$ und $p_1 = 1-p$, Ihre Entropie $H_b(p)$, auch *Shannonsche Funktion* genannt, ist

$$H_b(p) = -p \cdot ld\left(p\right) - (1-p) \cdot ld(1-p)$$



Entropie der Binärquelle

Der Funktionsverlauf ist im Bild zu sehen. Setzt die Quelle stets das Zeichen ”1” ab (p=0) ist die Entropie gleich null. Da man in diesem Fall weiß, dass die Quelle stetzt die ”1” sendet gibt sie keine Information ab. Sind beide Zeichen gleichwahrscheinlich (p=1/2), so wird die Entropie maximal. Ein Beobachter, der die Aufgabe hätte das nächste Zeichen vorherzusagen würde im mittel geauso häufig richtig wie falsch raten. Die Entropie der Binärquelle nimmt im Maximum den Wert 1 bit an. Dies entspricht genau einer Ja/Nein-Entscheidung (Antwort) um das aktuelle Zeichen zu erfragen (ist das Zeichen ”0”?).
Die Entropie gibt allgemein eine Antwort auf die zwei Fragen:

- Wie viele Ja/Nein-Entscheidungen sind mindestens notwendig, um das aktuelle Zeichen zu erfragen?
- Wie viele Bits benötigt man mindestens, um die Zeichen der Quelle zu codieren?

Um die Bedeutung der Entropie aufzuzeigen, betrachten wir ein Zahlenwertbeispiel. In Tabelle 6-I ist eine diskrete gedächtnislose Quelle mit sechs Zeichen beschrieben. Ihre Entropie beträgt ca. 2,25 bit.

Tabelle 6-1: Diskrete gedächtnislose Quelle mit dem Zeichenvorrat $X = \{a,b,c,d,e,f\} = \{x_1,...,x_6\}$, den Wahrscheinlichkeiten p_i, den Informationsgehalten $I(p_i)$ und der Entropie $H(X)$						
x_i	a	b	c	d	e	f
p_i	0.05	0.15	0.05	0.4	0.2	0.15
$I(p_i)$ / bit	4.32	2.74	4.32	1.32	2.32	2.74
$H(X)$ / bit	≈ 2.25					

Zunächst betrachte man in Tabelle 6-2 die einfache BCD-Codierung anhand des Zeichen-Index. Der BCD-Code ist ein Blockcode mit gleichlangen Codewörtern. Da sechs Zeichen vorliegen, müssen je Codewort 3 Bits verwendet werden.

Tabelle 6-2: BCD-Codierung der Zeichen nach ihren Indizes						
Zeichen	a	b	c	d	e	f
Codewort	001	010	011	100	101	110

Andererseits kann überlegt werden, wie groß die Entropie einer Quelle mit sechs Zeichen maximal sein kann.

Die Entropie einer diskreten gedächtnislosen Quelle mit N Symbolen wird maximal, wenn alle Symbole gleichwahrscheinlich sind (maximale Ungewißheit). Dieses Maximum ist der ***Entscheidungsgehalt*** (des Zeichenvorrats) der Quelle.

$$H_0 = - 1dN \text{ bit}$$

Der Entscheidungsgehalt einer diskreten gedächtnislosen Quelle mit 6 Symbolen ist 2,58 bit. Dem steht im Beispiel die Entropie von 2,25 bit gegenüber.
Die Differenz aus dem Entscheidungsgehalt einer Quel1e und ihrer Entropie wird ***Redundanz*** bzw. *relative* ***Redundanz*** genannt.

$$R = H_0 - H(X) \quad \text{bzw.} \quad r = 1 - \frac{H(X)}{H_0}$$

Die Entropie besagt im Beispiel, dass im Mittel 2,25 Ja/Nein-Entscheidungen notwendig sind und deshalb die Zeichen der Quelle im Mittel mit 2,25 Bits codiert werden können. Ein Verfahren, das einen in diesem Sinne aufwandsgünstigen Code liefert, wird im nächsten Abschnitt vorgestellt.

Quelle mit Gedächtnis

Bei realen Quellen (Sprache, Bild- oder Toninformation) hängen die emittierten Symbole von den Vorausgegangenen ab und sind keine unabhängigen Ereignisse. Ein Modell für solche Prozesse bilden *Markoffketten.* Wir wollen hier die Verhältnisse für den einfachsten Fall, einer Markoffkette 1. Ordnung untersuchen, bei der das emittierte Symbol vom Vorgängersymbol abhängt. Hat eine solche Kette n verschiedene Symbole, können insgesamt n verschiedene Folgepaare auftreten, für die wir wieder die Entropie, in diesem Falle die sogenannte ***Paarentropie H(XY)*** berechnen können. Dabei ist x das vorausgehende und y das nachfolgende Symbol aus den n Symbolen des Qellalphabets. Für die Wahrscheinlichkeit, das ein Paar (x_i, y_j) auftritt, ergibt sich

$$P(x_i y_j) = P(y_j | x_i) \cdot P(x_i)$$

Damit kann jetzt *H(XY)* als Erwartungswert für die Information eines Paares berechnet werden.

$$X(XY) = -\sum_{i=1}^{N}\sum_{j=1}^{N} P(x_i y_j) \cdot ld\left(P(x_i y_j)\right) \qquad \text{(gemessen in bit/Paar)}$$

Mit dem Ausdruck für die Paarwahrscheinlichkeit können wir weiter schreiben

$$X(XY) = -\sum_{i=1}^{N}\sum_{j=1}^{N} [P(y_j | x_i) \cdot P(x_i) \cdot ld\left(P(y_j | x_i)\right) + P(y_j | x_i) \cdot P(x_i) \cdot ld\left(P(x_i)\right)]$$

Der erste Teil der Summe liefert einen Ausdruck, den man als die bedingte Entropie ***H(Y|X)*** bezeichnet und der den Erwartungswert für die Information des zweiten Zeichens im Paar angibt, wenn das Zeichen X vorausgegangen ist.

$$H(Y | X) = -\sum_{i=1}^{N}\sum_{j=1}^{N} P(x_i y_j) \cdot P(x_i) \cdot ld\left(P(y_j | x_i)\right)$$

Beachtet man weiter, dass $\sum_{j=1}^{N} P(y_j | x_i) = 1$ gilt, so liefert der zweite Teil der Summe einfach

$$H(X) = -\sum_{j=1}^{N} P(x_i) \cdot ld(P(x_i)) \qquad \text{die Entropie für ein Zeichen allein.}$$

Es gilt also

$$H(XY) = H(X) + H(Y \mid X)$$

wobei H(X) = H und H(Y|X) < H gilt. Gleich H ist die bedingte Entropie nur, wenn die Y unabhängig von den X sind. Davon ausgehend kann man für H(Y|X) auch schreiben

$$H(Y \mid X) = H(Y) - H(X;Y)$$

wobei H(X; Y) die Synentropie, das heißt, der Entropieverlust durch die Abhängigkeit der Zeichen Y von den Zeichen X ist.

$$H(X;Y) = -\sum_{i=1}^{N}\sum_{j=1}^{N} P(x_i y_j) \cdot ld\left(\frac{P(x_i)\cdot P(y_j)}{P(x_i y_j)}\right).$$

Und für H(XY) gilt

$$H(XY) = H(X) + H(Y) - H(X;Y)$$

Hieraus kann man nun auch die mittlere Information angeben, die von einem Zeichen geliefert wird. Wir sprechen bei einer Quelle mit Gedächtnis dabei von der ***Entropierate***
Bei einer Markoff-Quelle 1.Ordnung gilt

$$H = H(XY)/2 \, .$$

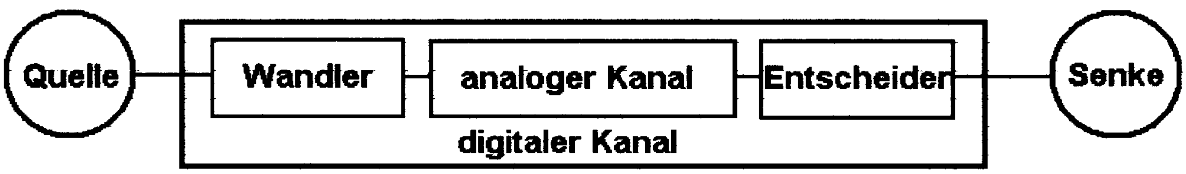
Im allgemeinen Fall von Abhängigkeiten über N Symbole gilt für die Entropierate

$$H = H(X_1 X_2 ... X_N)/N \, .$$

Oder bei unbestimmter Kettenlänge

$$H_\infty = \lim_{N \to \infty} \{H(X_1 X_2 ... X_N)/N\}$$

Nachrichtenkanäle



Ein (diskreter) Nachrichtenkanal besteht aus einem Eingang, an dem alle T Sekunden ein Symbol x_i Î X angelegt wird, und aus einem Ausgang, an dem alle T Sekunden ein Symbol y_i Î Y herausgegeben wird. Man nennt ***X*** das ***Eingangsalphabet*** und ***Y*** das

Ausgangsalphabet - oft sind beide Alphabete identisch. Im allgemeinen sind die statistischen Verknüpfungen zwischen den Ein- und Ausgängen des Kanals invariant gegenüber einer Zeitverschiebung - der Kanal also stationär. Wir setzen dies stets voraus.

In vielen Fällen hängt die Statistik des Ausgangssymbols außer vom momentanen Eingangsymbol auch von der Vergangenheit des Kanals (d.h. von vorangegangenen Ein- und Ausgangswerten am Kanal) ab. Läßt sich die bedingte Wahrscheinlichkeitsmatrix P(Y|X) in Abhängigkeit von k vorangegangenen Ein-/Ausgangswerten angeben, so spricht man von einem (diskreten) Kanal mit einem Gedächtnis k-ter Ordnung. Für die Modellierung eines solchen Kanals sind k Zustandsvariablen und die jeweils zu jedem Zustand gehörigen bedingten Übergangswahrscheinlichkeiten P(Y|X) erforderlich.
Im einfachsten Fall ist der ***Kanal gedächtnislos*** - d.h. er besitzt nur einen Zustand. Bei einem solchen gedächtnislosen Kanal sind die Ausgangswahrscheinlichkeiten durch die bedingte Wahrscheinlichkeitsmatrix P(Y|X) festgelegt. Nimmt man an, dass die Quelle am Eingang eines solchen Kanale stationär ist, so ist, wenn man die Eingangsquelle mit dem Kanal wiederum als eine neue Quelle betrachtet, diese auch stationär.
Beispiel :
Das Eingangsalphabet X und das Ausgangsalphabet Y sind durch

$$X = \{x_1, x_2, ..., x_m\}, m \in N$$

und

$$Y = \{y_1, y_2, ..., y_n\}, n \in N$$

angegeben.
$P(y_j|x_i)$ ist die Wahrscheinlichkeit dafür, dass das Symbol ***y_j*** am Kanalausgang empfangen wird, wenn das Symbol ***x_i*** am Kanaleingang gesendet wird. Die Kanalmatrix sieht wie folgt aus:

$$P(Y \mid X) = \begin{bmatrix} P(y_1 \mid x_1) & P(y_2 \mid x_1) & ... & P(y_n \mid x_1) \\ P(y_1 \mid x_2) & P(y_2 \mid x_2) & ... & P(y_n \mid x_2) \\ : & : & ::: & : \\ P(y_1 \mid x_m) & P(y_2 \mid x_m) & & P(y_n \mid x_m) \end{bmatrix}$$

Ein wichtiges Merkmal einer beliebigen Kanalmatrix ist, dass die Zeilensumme gleich eins ist, z.B. für die erste Zeile gilt:

$$\sum_{i=1}^{N} P(y_i \mid x_1) = 1$$

weil *$\{Y|x_i\}$* ein sicheres Ereignis ist, d.h. dass irgendein y_i Î Y sicher empfangen wird, wenn das Symbol x_i gesendet wird.
Gilt bei einem gedächtnislosen Kanal mit jeweils q Ein- und Ausgängen für die bedingte Wahrscheinlichkeitsmatrix P(Y|X), die den Kanal charakterisiert,

$$P(y_j \mid x_i) = \begin{cases} 1-p, i=j \\ \dfrac{p}{q-1}, i \neq j \end{cases}$$

wobei $0 < p < 1$ gilt, so spricht man von einem symmetrischen Kanal mit der Fehlerwahrscheinlichkeit p.

Empfangsstrategien

Oft besteht bei der Datenübertragung die Aufgabe, aus einem empfangenen Symbol darauf zu schließen, welches Symbol gesendet wurde. Will man die Fehlerwahrscheinlichkeit bei der Auswahl minimieren, so sucht man beim Empfang eines Symbols y_j aus allen möglichen Sendesignalen x_i das Signal x^* aus, für welches gilt

$$P(x^* \mid y_j) \geq P(x_i \mid y_j)$$

Wegen P(x | y)·P(y) = P(y | x)·P(x) erhalten wir

$$\frac{P(y_j \mid x^*) \cdot P(x^*)}{P(y_j)} \geq \frac{P(y_j \mid x_i) \cdot P(x_i)}{P(y_j)}$$

oder

$$P(y_j \mid x^*) \cdot P(x^*) \geq P(y_j \mid x_i) \cdot P(x_i)$$

Die Gleichung zeigt, dass die Auswahl abhängig von der a priori Wahrscheinlichkeit $P(x_i)$ ist. Nimmt man an, dass die Eingangssymbole gleichwahrscheinlich sind, so erhält man als Kriterium

$$P(y_j \mid x^*) \geq P(y_j \mid x_i)$$

Dieses Entscheidungsverfahren wird als ***Maximum-Likelihood-Verfahren*** bezeichnet.

Kanalentropien

Die Entropie der Eingangsquelle an einem Kanal ist

$$H(X) = -\sum_{i=1}^{N} P(x_i) \cdot ld\left(P(x_i)\right).$$

Die Eingangsquelle und der Kanal können zusammen wiederum als eine Quelle betrachtet werden, für deren Entropie gilt

$$H(Y) = -\sum_{i=1}^{N} P(y_i) \cdot ld\left(P(y_i)\right).$$

Wie bei den Markoff-Quellen können wir für einen Nachrichtenkanal auch Verbund- und bedingte Entropien definieren.
Die ***Verbundentropie*** des Kanals ist definiert als

$$H(X,Y) = -\sum_{i,j}^{N} P(x_i, y_j) \cdot ld\left(P(x_i, y_j)\right)$$

Sie ist ein Maß für die in einem Ein-/Ausgangssymbo1paar im Mittel enthaltene Information. (*Wir haben ein Komma zwischen die Symbole gesetzt, um zu verdeutlichen, dass die Symbole nicht wie bei Symbolfolgen zeitlich nacheinander auftreten, sondern als Paar am Kanaleingang und Kanalausgang anliegen. P(x, y) ist lediglich die Wahrscheinlichkeit, dass x und y gemeinsam auftreten. Es ist deshalb auch P(x,y)=P(y,x).)*

Die ***Aquivokation*** oder ***Rücksschlußentropie*** ist definiert

$$H(X \mid Y) = -\sum_{i,j}^{N} P(x_i, y_j) \cdot ld\left(P(x_i \mid y_j)\right)$$

Sie ist ein Maß für die im Mittel in einem Eingangssymbol für einen Beobachter, der den Ausgang kennt, enthaltene zusätzliche Information.

Die ***Streuentropie*** oder ***Irrelevanz*** ist definiert als

$$H(Y \mid X) = -\sum_{i,j}^{N} P(x_i, y_j) \cdot ld\left(P(y_j \mid x_i)\right)$$

Sie ist ein Maß für die im Mittel in einem Ausgangssymbol für einen Beobachter, der den Eingang kennt, enthaltene zusätzliche Information.

Die ***Transinformation*** ist entsprechend der *Synentropie* definiert als

$$H(X;Y) = -\sum_{i,j}^{N} P(x_i, y_j) \cdot \left(ld\left(P(x_i) \cdot P(y_j)\right) - ld\left(P(x_i, y_j)\right)\right)$$

Aus der Definition sieht man, dass die Transinformation in X und Y symmetrisch ist, und man erhält

$$H(X;Y) = H(Y) - H(Y \mid X)$$

und

$$H(X;Y) = H(X) - H(X \mid Y)$$

Die Transinformation ist ein Maß für die im Mittel in einem Ausgangsymbol enthaltene Information verringert um die Streuentropie. Wie bei den Markoff-Quellen gelten auch hier die Beziehungen

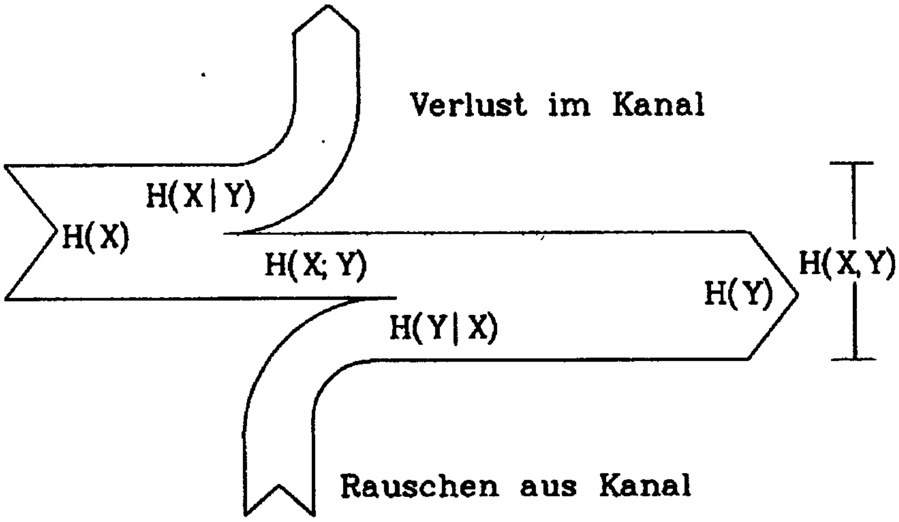
$$\begin{aligned} H(X,Y) &= H(Y \mid X) + H(X) \\ &= H(X \mid Y) + H(Y) \end{aligned}$$

und

$$H(X;Y) = H(Y) - H(Y \mid X)\,.$$

Diese Gleichung lässt sich wie folgt interpretieren. Die Verbundentropie der Ein- und Ausgangssymbole eines Kanals ist die Summe der Einzelentropien verringert um die Transinformation.

Der gesamte Sachverhalt ist im Bild 4.1 dargestellt. *H(X)* ist die Eingangsentropie am Kanal. Sie besteht aus der Rückschlußentropie *H(X | Y),* die im Kanal verloren geht, und der Transinformation *H(X;Y),* die zum Kanalausgang gelangt. Der Kanal fügt die Irrelevanz *H(Y | X)* dem Ausgang zu, so dass am Kanalausgang die Ausgangsentropie als die Summe der Transinformation und der Irrelevanz vorliegt. Die Verbundentropie des Kanals besteht aus der Rückschlußentropie, der Transinformation und der Irrelevanz.



H(X)	Entropie am Kanaleingang
H(Y)	Entropie am Kanalausgang
H(X| Y)	Rückschlußentropie oder Äquivokation
H(X;Y)	Transinformation
H(Y| X)	Streuentropie oder Irrelevanz
H(X,Y)	Verbundentropie

Bild 4.1: Die Entropien am Kanal

Transinformation und Kanalkapazität

Die Transinformation ist im allgemeinen eine Funktion sowohl der bedingten Wahrscheinlichkeiten $P(y_j|x_i)$ die den Kanal charakterisieren, als auch der Wahrscheinlichkeitsverteilung der Kanaleingangssymbole $P(x_i)$. Bildet man das Maximum der Transinformation über alle möglichen (zulässigen) Eingangswahr- scheinlichkeitsverteilungen, so erhält man eine von der Quelle am Eingang unabhängige Größe. Sie ist ein Maß für den Informationsgehalt, den ein Kanal maximal übertragen kann und wird deshalb die Kapazität des Kanals genannt. Die Kanalkapazität ist definiert als

$$C = \max_{P(X)} \{H(Y;X)\}$$

wobei das Maximum über alle zulässigen Eingangswahrscheinlichkeitsverteilungen zu bilden ist. Ein solches Maximum existiert stets, denn die Transinformation ist eine stetige Funktion der n-Variablen $P(x_i)$, und ihr Definitionsbereich ist beschränkt und abgeschlossen (wegen $P(x_i) \geq 0$ und $\sum_i P(x_i) = 1$).

Die Kanalkapazität sagt aus, wie viel bit an Information bei einem gegebenen Kanal maximal mit einem Zeichen übertragen werden können. Besitzt ein solcher Kanal einen Übertragungsbereich von $0 \leq f \leq f_g$ (Tiefpasskanal mit der Bandbreite $B = f_g$) kann maximal mit einer Zeichenrate $R_s = 2 \cdot f_g$ übertragen werden. Für diesen Kanal gilt dann

$$C'/bps = 2 \cdot f_g \cdot \max|_{P(X)} \{H(X;Y)\}$$

Beispiel :

Ein symmetrischer Kanal mit jeweils q- Ein- und Ausgangssymbolen wird beschrieben durch

$$P(Y \mid X) = \begin{bmatrix} 1-p & \frac{p}{q-1} & .... & \frac{p}{q-1} \\ \frac{p}{q-1} & 1-p & .... & \frac{p}{q-1} \\ \vdots & \vdots & :::: & \vdots \\ \frac{p}{q-1} & \frac{p}{q-1} & .... & 1-p \end{bmatrix}$$

Wir betrachten zunächst die Streuentropie $H(Y \mid X)$.

$$H(Y \mid X) = -\sum_{i=1}^{q} P(x_i) \cdot \left[\sum_{j=1}^{q} P(y_j \mid x_i) \cdot ld(P(y_j \mid x_i) \right]$$

Die Summe der eckigen Klammern ist für jedes i gleich groß, weil die Elemente in jeder Zeile bis auf eine Permutation gleich sind. Somit ist

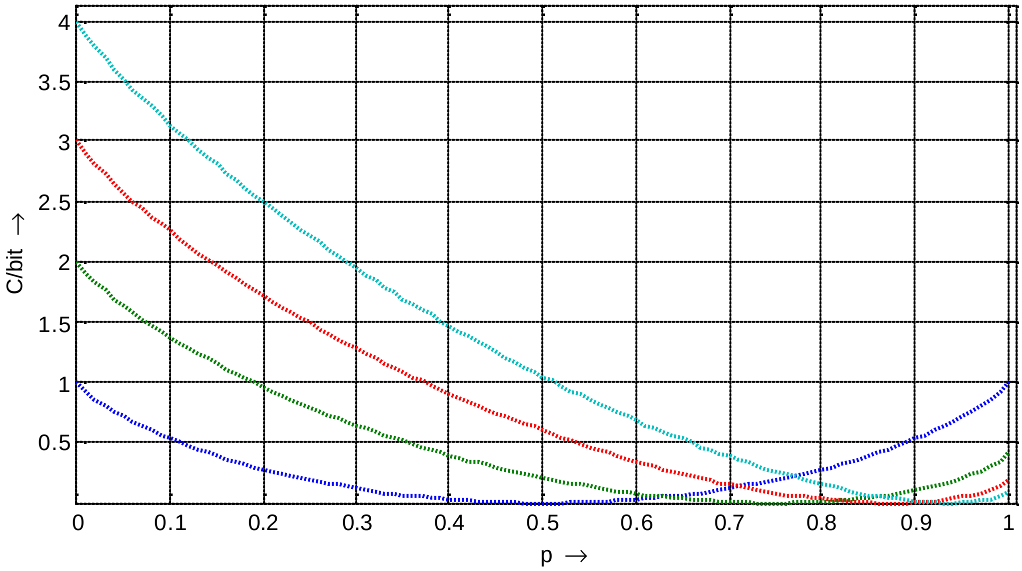
$$H(Y \mid X) = -1 \cdot \left[(1-p) \cdot ld(1-p) + (q-1) \cdot \frac{p}{q-1} \cdot ld(\frac{p}{q-1}) \right]$$

von $P(x_i)$ unabhängig. Der Ausdruck H(Y) – H(Y|X) wird maximal, wenn H(Y) maximal wird. Das ist genau dann der Fall, wenn alle Symbole des Alphabets Y gleichwahrscheinlich sind.

$$\boldsymbol{H(Y)_{max} = ldq} \; .$$

Die Kanalkapazität errechnet sich nun zu:

$$C = H(Y)_{\max} - H(Y \mid X) = ld(q) + \left[(1-p) \cdot ld(1-p) + pld(\frac{p}{q-1}) \right]$$



Kontinuierliche Kanäle

Bisher haben wir diskrete Kanäle untersucht. Es ist aber auch möglich dieses Konzept auf kontinuierliche Kanäle auszudehnen. Shannon hat 1948 für den sogenannten AWGN (Additive White Gaussian Noise) Kanal gezeigt, das sich dabei die Kanalkapazität zu

$$C'/bps = B \cdot ld\left(1 + \frac{P_S}{P_N}\right)$$

bestimmt. Dieser Zusammenhang ist für die Informationsübertragung von grundlegender Bedeutung, da der AWGN-Kanal ein gutes Modell für die meisten Übertragungskanäle ist und die Shannonsche Kanalkapazität zum einen die physikalischen Grenzen aufzeigt, zum anderen aber auch das technisch Anzustrebende darstellt. Für Kanäle mit großem Signal-Rauschabstand (SNR Signal Noise Ratio / dB) gilt vereinfachend

$$C/bps = \frac{1}{3} B \cdot SNR/dB \qquad \text{mit} \qquad SNR/dB = 10 \cdot \lg\left(\frac{P_S}{P_N}\right)$$

In der weiteren Darstellung wollen wir jetzt auch den Strich am C weglassen und auch

für die mögliche Information pro Zeiteinheit einfach C schreiben.
Die **Shannonsche Gleichung** zeigt:

- Die Kanalkapazität wächst beliebig, wenn der Störabstand beliebig erhöht wird.
- Die Kanalkapazität strebt gegen einen Grenzwert, wenn die Bandbreite beliebig heraufgesetzt wird.

Der letzten Aussage wollen wir uns nun zuwenden. Ursache für dieses Verhalten ist die Abhängigkeit der störenden Rauschleistung von der Bandbreite. Für thermisches Rauschen gilt ganz allgemein

$$P_N = k \cdot T \cdot B = N_0 \cdot B$$

$k = 1{,}38\ 10^{-23}$ J /K Boltzmannkonstante
T = absolute Temperatur in [Kelvin]
B = Bandbreite in [Hz]
$N_0 = k\ T = P_N / B$ spektrale thermische Rauschleistungsdichte

Damit erhalten wir

$$C = B \cdot ld(1 + \frac{P_S}{N_0 \cdot B})$$

Wir wollen jetzt den Grenzwert C_∞ der sich ergibt, wenn wir B nach unendlich streben lassen. Zunächst wandeln wir den $ld(\cdot)$ in den $ln(\cdot)$ und nutzen dann die Näherung $ln(1+x) \approx x$, wenn $x \rightarrow 0$ strebt.

$$C_\infty = \lim_{B \to \infty} B \cdot \ln(1 + \frac{P_S}{N_0 \cdot B}) \cdot \frac{1}{\ln(2)} = \frac{P_S}{N_0 \cdot \ln(2)}$$

Dieser Grenzwert kann nicht überstiegen werden. Bildet man den Quotienten P_S / C_∞, ergibt sich der Wert der Energie für ein bit an Information der für eine fehlerfreie Übertragung unbedingt aufgewendet werden muß. Es sind

$$E_{bit} = k \cdot T \cdot \ln(2) = 2{,}81 \cdot 10^{-21} Ws$$

Die Übertragung von Information erfordert ***Bandbreite***, ***Energie*** und ***Zeit*** (und Geld).
Für den Anwender ist vor allem die Informationsrate ***J*** als Effektivitätsmaß einer Übertragung von Interesse. Sie kann höchstens gleich C werden. Wichtig ist aber auch der Aufwand der pro 1bit getrieben werden muß in Bezug auf:

- Energieaufwand → Energie pro 1bit
- Bandbreiteneffizienz → spektrale Informationsrate [bit s^{-1} /Hz]
- Investitionskosten, Betriebskosten

Den Energieaufwand setzt man häufig ins Verhältnis zur Rauschleistungsdichte

$$\frac{E_{bit}}{N_0} = \frac{P_S}{J} \cdot \frac{1}{N_0} \geq \frac{P_S}{C} \cdot \frac{1}{N_0} = \frac{P_S}{C \cdot k \cdot T}, \left[\frac{Ws}{bit \cdot Ws} = 1 \right]$$

Diese Größe ist dimensionslos und lässt sich in dB ausdrücken. Sie ist ein Maß für den

Störabstand.
Untersuchen wir jetzt die spektrale Effizienz. J/B heißt spektrale Informationsrate und es gilt

$$\frac{J}{B} \leq \frac{C}{B} = ld\left(1 + \frac{C}{B} \cdot \frac{E_{bit}}{N_0}\right), \left[\frac{bit \cdot s^{-1}}{Hz}\right]$$