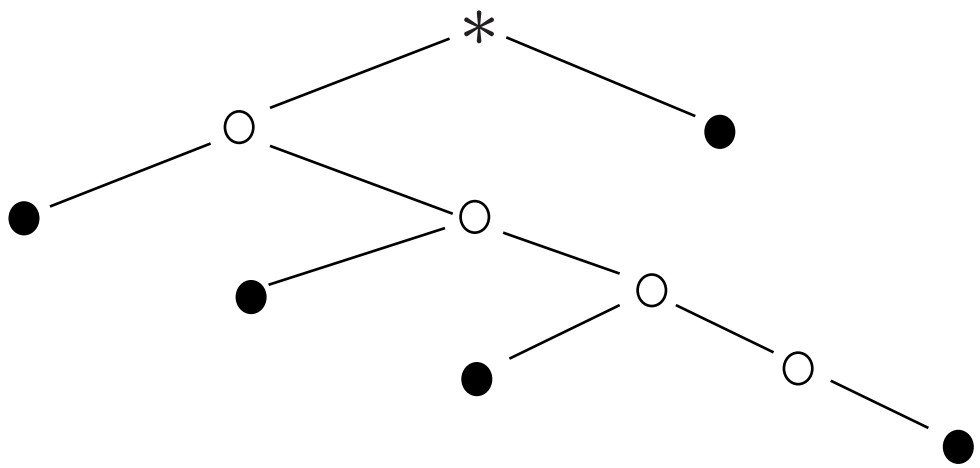
Vorlesung 14a

Quellencodieren und Entropie

1. Binärbäume
als gerichtete Graphen

und

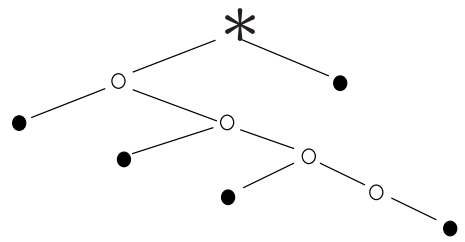
die gewöhnliche Irrfahrt
von der Wurzel zu den Blättern



Wir betrachten planare (d.h. in die Ebene gezeichnete),

verwurzelte Binärbäume

Ihre Merkmale sind:

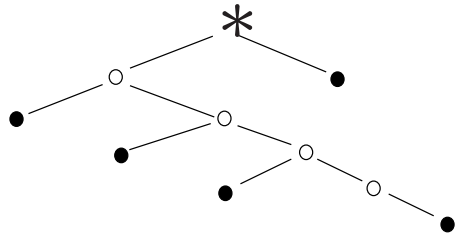


Ein Knoten ist ausgezeichnet als *Wurzel*.

Jeder Knoten (außer der Wurzel) hat genau einen Vorgänger.

Jeder Knoten hat zwei, einen oder keinen Nachfolger.

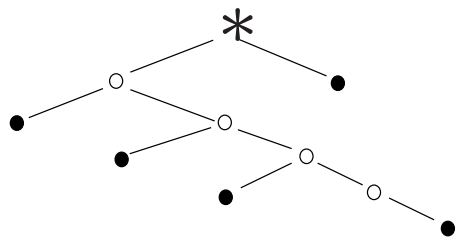
Die Knoten ohne Nachfolger heißen *Blätter*,
die mit Nachfolger heißen *innere Knoten*.



Wir sprechen von einem *vollen Binärbaum,*

wenn jeder innere Knoten zwei Nachfolger hat.

Der obige Binärbaum ist nicht voll.

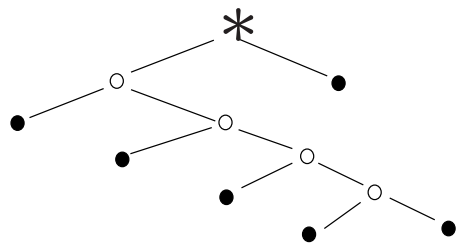


T

Wir sprechen von einem *vollen Binärbaum,*
wenn jeder innere Knoten zwei Nachfolger hat.

Jeder Binärbaum T lässt sich “auf minimale Weise”
zu einem vollen Binärbaum T' ergänzen.

Wir nennen T' die Vervollständigung von T.



T'

Wir sprechen von einem *vollen Binärbaum,*
wenn jeder innere Knoten zwei Nachfolger hat.

Jeder Binärbaum T lässt sich “auf minimale Weise”
zu einem vollen Binärbaum T' ergänzen.

Wir nennen T' die Vervollständigung von T.

Sei T ein Binärbaum.

Die gewöhnliche Irrfahrt auf T

- macht von jedem inneren Knoten einen Schritt zu einem zufällig gewählten seiner Nachfolger (davon gibt's 1 oder 2)

- und endet in einem der Blätter.

Wenn nicht anderes vermerkt, nehmen wir immer an, dass die Irrfahrt in der Wurzel startet.

Sei T ein voller Binärbaum auf T

und Y die gewöhnliche Irrfahrt auf T

Dann gilt für jedes Blatt b von T mit Tiefe $\ell(b)$:

$$\mathbf{P}(Y \text{ endet in } b) = 2^{-\ell(b)}$$

(denn auf dem Weg von der Wurzel zu b muss $\ell(b)$-mal

das “richtige” Münzwurfergebnis kommen)

Folgerung:

Für jeden vollen Binärbaum T mit Blattmenge B gilt:

$$\sum_{b \in B} 2^{-\ell(b)} = 1.$$

Ist T nicht voll, dann gilt die entsprechende Gleichheit

für die Vervollständigung T' von T.

T' hat dann mehr Blätter als T,

also ist dann

$$\sum_{b \in B} 2^{-\ell(b)} < 1.$$

Dies fassen wir zusammen in der

Bemerkung:

Für jeden Binärbaum T

mit Blattmenge B und Blatttiefen $\ell(b)$ gilt:

$$\sum_{b \in B} 2^{-\ell(b)} \leq 1,$$

mit Gleichheit genau dann wenn T voll ist.

Wir fragen:

Welche Bedingung an natürliche Zahlen $t_1, \ldots, t_m$

ist notwendig und hinreichend dafür,

dass diese Zahlen

als Blatttiefen eines binären Baumes auftreten?

Die Antwort gibt die

Fano-Kraft Ungleichung für binäre Bäume:

$t_1, \ldots, t_m$ seien natürliche Zahlen.

Genau dann gibt es einen binären Baum T,

für den die $t_1, \ldots, t_m$ die Tiefen seiner Blätter sind,

wenn gilt:

$$(*) \quad \sum_{i=1}^{m} 2^{-t_i} \leq 1.$$

Beweis:

" $\Longrightarrow$" siehe die Bemerkung auf Folie 11.

" $\Longleftarrow$": mit Induktion: Für $m = 1$ ist die Behauptung klar.

Schritt vom m zu $m + 1$:

Seien $t_1 \leq t_2 \leq \cdots \leq t_{m+1}$ natürliche Zahlen mit $\sum_{i=1}^{m+1} 2^{-t_i} \leq 1$. Dann gilt

$$\sum_{i=1}^{m} 2^{-t_i} < 1.$$

Nach Induktionsvoraussetzung gibt es

einen Binärbaum T_m mit Blatttiefen $t_1, \ldots, t_m$.

Nach der Bemerkung auf Folie 11 ist T_m nicht voll,

besitzt also einen inneren Knoten (in Tiefe $< t_m$)

mit nur einem Nachfolger.

Zu diesem Knoten fügen wir einen zweiten Nachfolger hinzu

und setzen (falls $t_{m+1} > t_m$) seine "Nachkommenslinie" fort,

bis wir bei der Tiefe t_{m+1} angelangt sind.

Damit ist ein Binärbaum mit den Blatttiefen $t_1, \ldots, t_{m+1}$

konstruiert. $\square$

Die Ungleichung von Fano und Kraft, formuliert für

das bijektive Beschriften der Blattmenge eines Binärbaums:

S sei eine endliche oder abzählbar unendliche Menge

und $t(a),\ a \in S$, seien natürliche Zahlen.

Genau dann gibt es einen Binärbaum, dessen Blätter bijektiv

mit den Elementen $a \in S$ beschriftet sind

und Tiefen $t(a)$ haben,

wenn $\displaystyle\sum_{a \in S} 2^{-t(a)} \leq 1$ gilt.

Die Aussage folgt sofort mit dem vorher gegebenen Induktionsargument,
wenn man die $t(a)$ der Größe nach absteigend ordnet.

2. Binäre Präfixcodes

als Blattbeschriftungen von Binärbäumen

S sei eine endliche (oder abzählbar unendliche) Menge

(ein “Alphabet”).

Die Elemente von S nennen wir *Buchstaben*.

Wir wollen die Buchstaben $a, a', \ldots$ durch

01-Folgen $k(a), k(a'), \ldots$ codieren.

Dabei soll so etwas wie

$$k(a) = 001,\ k(a') = 00101$$

ausgeschlossen sein.

Definition:

Eine Abbildung

$$k : S \to \bigcup_{t \geq 1} \{0,1\}^t$$

heißt **binärer Präfixcode**,

wenn kein $k(a)$ Anfangsstück irgendeines $k(a')$, $a \neq a'$, ist.

Ist $k(a) = k_1(a) \ldots k_t(a)$, dann nennt man

$$\ell(a) := t$$

die *Länge* (oder auch die Anzahl der Bits)

des *Codeworts* $k(a)$.

Binäre Präfixcodes kann man sich auch als

(planare (d.h. “in die Ebene gezeichnete”))

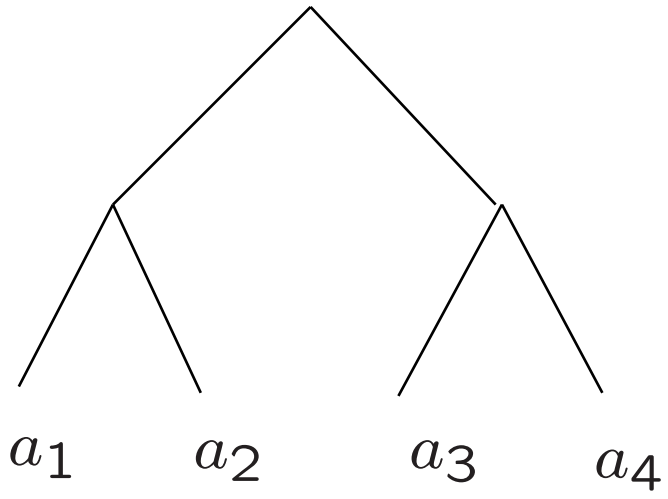
Binärbäume vorstellen,

deren Blätter bijektiv

mit den Buchstaben des Alphabets beschriftet sind.

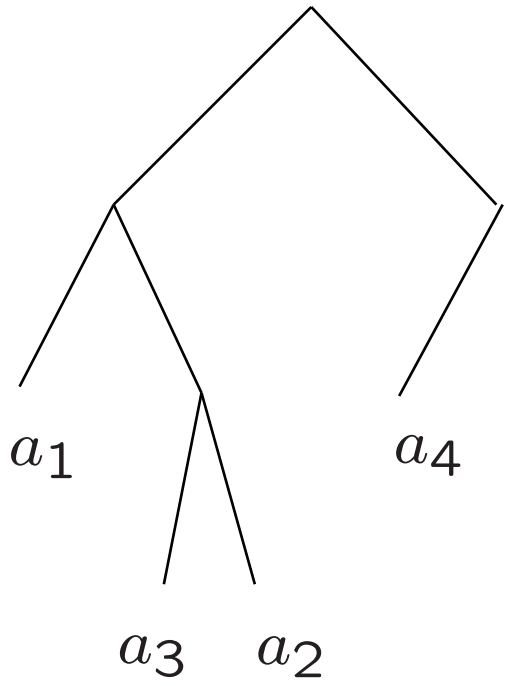
Beispiel: $S = \{a_1, a_2, a_3, a_4\}$

$k(a_1) := 00,\ k(a_2) := 01,\ k(a_3) := 10,\ k(a_4) := 11.$



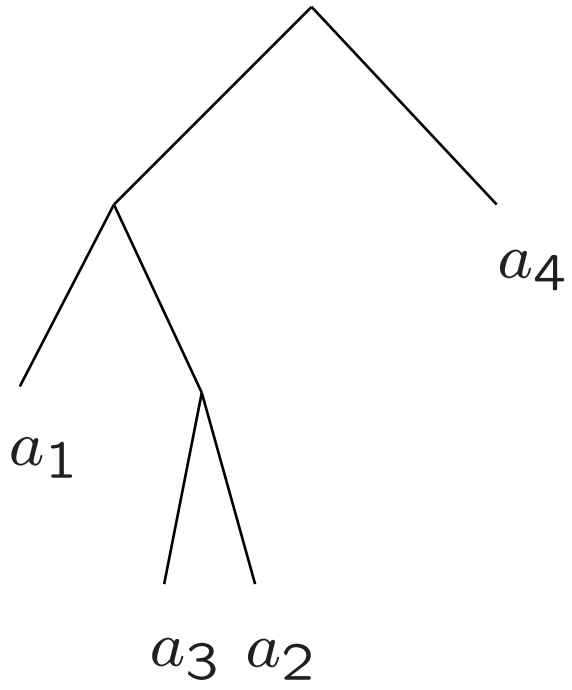
Beispiel: $S = \{a_1, a_2, a_3, a_4\}$

$k(a_1) := 00,\ k(a_2) := 011,\ k(a_3) := 010,\ k(a_4) := 10.$



Beispiel: $S = \{a_1, a_2, a_3, a_4\}$

$$k(a_1) := 00,\ k(a_2) := 011,\ k(a_3) := 010,\ k(a_4) := 1.$$



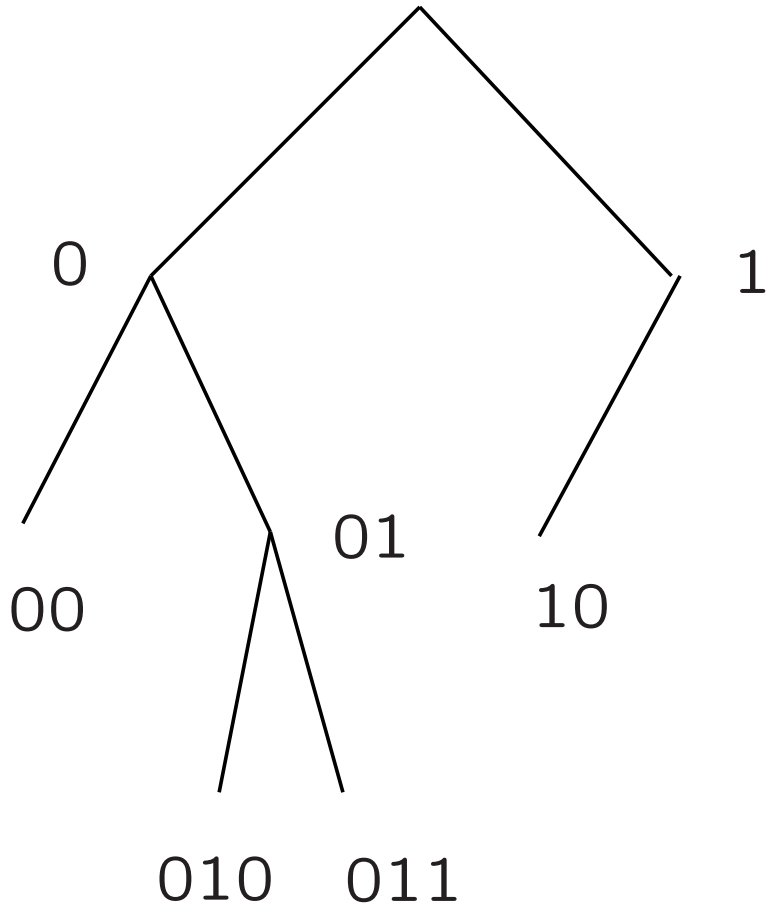
Regeln für die Beschriftung der Knoten mit 01-Wörtern:

Die Wurzel trägt das Wort der Länge 0 (das “leere Wort”).

Das Wort an einem Knoten (außer der Wurzel)
setzt das Wort am Vorgängerknoten um ein Bit fort
(die Wörter in der Tiefe t haben somit die Länge t).

Verschiedene Knoten tragen verschiedene Wörter.

Beispiel:



Wir formulieren die (bereits oben bewiesenen)
zwei Aussagen um die Fano-Kraft-Ungleichung
jetzt noch einmal für binäre Präfixcodes:

Die Fano-Kraft-Ungleichung für binäre Präfixcodes:

(Teil 1) Für jeden binären Präfixcode k gilt:

$$\sum_{a \in S} 2^{-\ell(a)} \leq 1 \ .$$

"Je mehr Buchstaben man codieren will,
um so länger müssen die Codewörter sein."

(Teil 2) Ist $\ell : S \to \mathbb{N}$ eine Abbildung mit $\sum_{a \in S} 2^{-\ell(a)} \leq 1$,

dann gibt es einen binären Präfixcode,

für den $\ell(a)$ die Länge von $k(a)$ ist für alle $a \in S$.

"Jede nicht allzu kurzwertige Abbildung ℓ tritt als
Codewortlängenabbildung auf."

3. Sparsames Codieren zufälliger Buchstaben.

Sei X eine S-wertige Zufallsvariable

(ein “zufälliger Buchstabe”)

mit Verteilungsgewichten $\rho(a) = \mathbf{P}(X = a)$.

Gefragt ist nach einem binären Präfixcode für S,

der in dem Sinn möglichst günstig ist,

dass die erwartete Codelänge von X,

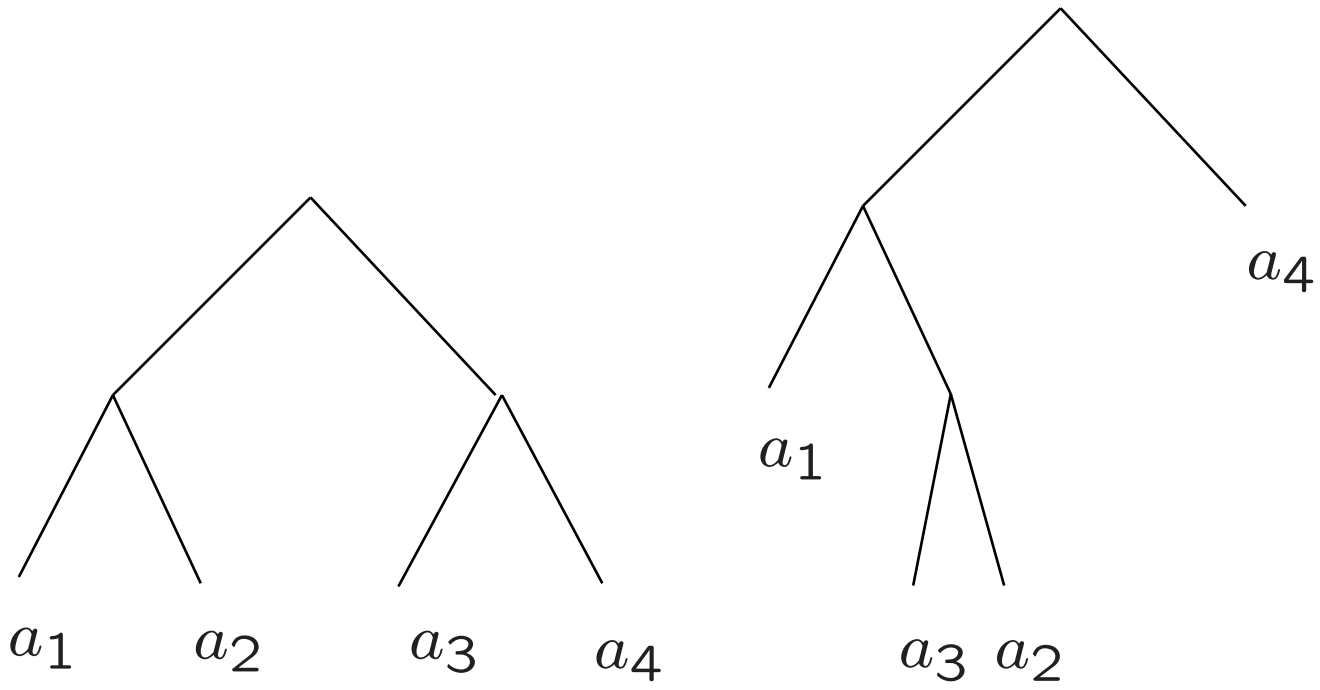
$$\mathbf{E}[\ell(X)] = \sum_{a \in S} \ell(a)\rho(a,)$$

möglichst klein ist.

Beispiel: Beispiel: $S = \{a_1, a_2, a_3, a_4\}$.

Sind die vier Ausgänge gleich wahrscheinlich,

dann ist der Code links günstiger als der rechts.



Für jedes $t \in \mathbb{N}$ ist glaubhaft

(und in der Tat auch eine Folgerung des in Abschnitt 4. formulierten und bewiesenen Quellencodierungssatzes):

Gilt $\#S = 2^t$ mit $\rho(a) = 2^{-t},\ a \in S$,

dann ist es am günstigsten, alle 2^t Buchstaben

mit den 01-Folgen der Länge t zu codieren.

In diesem Fall gilt

$\ell(a) = t = -\log_2 \rho(a)$ für alle $a \in S$.

4. Shannon-Codes und der Quellencodierungssatz

Gegeben seien

eine (endliche oder abzählbar unendliche) Menge S

und Verteilungsgewichte $\rho(a)$ auf S.

Wir denken uns

eine S-wertige Zufallsvariablle X mit Verteilung ρ.

Ein *Shannon-Code* für (S, ρ) ist ein Präfixcode, bei dem jeder Buchstabe $a \in S$ mit einer 01-Folge codiert wird, deren Länge $\ell(a)$ durch Aufrunden von $-\log_2 \rho(a)$ auf die nächste ganze Zahl entsteht, also

$$-\log_2 \rho(a) \leq \ell(a) < -\log_2 \rho(a) + 1 \ .$$

Solche Codes gibt es immer,

denn für die so festgelegten $\ell(a)$ folgt

$$\sum_{a \in S} 2^{-\ell(a)} \leq \sum_{a \in S} \rho(a) = 1 \ ,$$

die Fano-Kraft Bedingung ist also erfüllt.

Wir werden zeigen:

Shannon-Codes

verfehlen die minimal mögliche erwartete Codelänge

um höchstens ein Bit.

Sei dazu

$$\mathbf{H}_2[X] := -\sum_{a \in S} \rho(a) \log_2 \rho(a),$$

mit $0 \log 0 := 0$.

Satz

(Quellencodierungssatz von Shannon)

a) Für jeden binären Präfixcode gilt

$$\mathbf{E}[\ell(X)] \geq \mathbf{H}_2[X].$$

b) Für binäre Shannon-Codes gilt außerdem

$$\mathbf{E}[\ell(X)] < \mathbf{H}_2[X] + 1.$$

Beweis:

Für Shannon-Codes gilt

$$\ell(a) < -\log_2 \rho(a) + 1$$

und folglich

$$\mathbf{E}[\ell(X)] < \sum_a \rho(a)(-\log_2 \rho(a) + 1) = \mathbf{H}_2[X] + 1.$$

Dies beweist erst einmal Teil b) des Satzes.

Jetzt zu Teil a):

Für beliebige Codes gilt

$$\mathbf{H}_2[X] - \mathbf{E}[\ell(X)] = \sum_{a:\rho(a)>0} \rho(a) \log_2 \frac{2^{-\ell(a)}}{\rho(a)}$$

Nun ist die Logarithmusfunktion konkav

und liegt unterhalb ihrer Tangente im Punkt 1.

Folglich gilt $\log_2 x \le c \cdot (x-1)$ mit geeignetem $c > 0$, und

$$\begin{aligned}\mathbf{H}_2[X] - \mathbf{E}[\ell(X)] &\le c \sum_{a:\rho(a)>0} \rho(a)\Big(\frac{2^{-\ell(a)}}{\rho(a)} - 1\Big) \\ &\le c \cdot \Big(\sum_a 2^{-\ell(a)} - 1\Big).\end{aligned}$$

Nach Fano-Kraft ist die rechte Seite ≤ 0, also

$$\mathbf{H}_2[X] - \mathbf{E}[\ell(X)] \le 0. \quad \square$$

5. Die Entropie

Die im Quellencodierungssatz auftretende Größe

$$\mathbf{H}_2[X] := -\sum_{a \in S} \rho(a) \log_2 \rho(a)$$

heißt die **Entropie** von X (zur Basis 2).

Die Entropie ist *der* fundamentale Begriff
der Informationstheorie.

Nach dem Quellenkodierungssatz gibt die Entropie
fast genau die mittlere Anzahl von Ja-Nein Fragen an,
die notwendig und
- bei guter Wahl des Codes - auch hinreichend ist,
um den unbekannten Wert von X von jemandem zu erfragen,
der X beobachten kann.

Dies ist gemeint, wenn man die Entropie beschreibt als den
Grad von Unbestimmtheit oder Ungewissheit
über den Wert, den X annimmt.