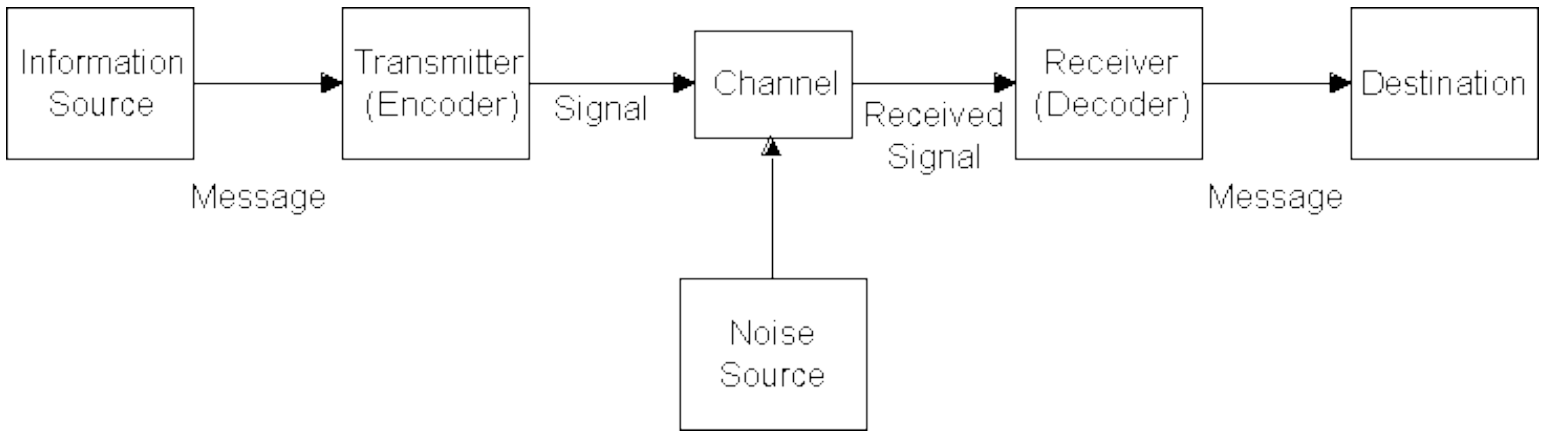
# Begriffe aus der Informatik

## Nachrichten

- Gerhard Goos definiert in *Vorlesungen über Informatik, Band 1, 1995 Springer-Verlag Berlin Heidelberg*:
  - "Die Darstellung einer Mitteilung durch die zeitliche Veränderung einer physikalischen Größe heißt ein *Signal*."
  - "Die dauerhafte Darstellung einer Mitteilung auf einem physikalischen Medium heißt *Inschrift*."
  - "Wenn wir bei der Darstellung und Weitergabe einer Mitteilung vom verwandten Medium und den Einzelheiten der Signale und Signalparameter abstrahieren, heißt die Mitteilung eine *Nachricht*."
  - "Die Kenntnisse, die man benötigt, um einer Nachricht Bedeutung zuzuordnen, nennen wir einen *Kontext* oder ein *Bezugssystem*."
  - "Die zugeordnete Bedeutung heißt eine *Information*."
  - "Man gewinnt sie durch *Interpretation* von Nachrichten auf der Grundlage eines Bezugssystems."
  - Eine *Interpretationsvorschrift*, angewendet auf eine Nachricht, liefert die zugeordnete Information.
  - Das Paar (Nachricht, zugeordnete Information) heißt *Datum*.

# Setting von Shannon-Weaver

- Modell der Kommunikation nach Shannon und Weaver:



# Diskrete, endliche Wahrscheinlichkeitstheorie

- Wir betrachten eine endliche Menge $\Omega$ (die nicht näher spezifiziert ist) und eine Wahrscheinlichkeitsverteilung $p : \Omega \to [0, 1]$, die jedem Element aus $\Omega$ eine Wahrscheinlichkeit zuordnet, derart dass die Summe über alle Wahrscheinlichkeiten 1 ergibt.

# Diskrete, endliche Wahrscheinlichkeitstheorie

- Wir betrachten eine endliche Menge $\Omega$ (die nicht näher spezifiziert ist) und eine Wahrscheinlichkeitsverteilung $p : \Omega \to [0, 1]$, die jedem Element aus $\Omega$ eine Wahrscheinlichkeit zuordnet, derart dass die Summe über alle Wahrscheinlichkeiten 1 ergibt.
- Wir betrachten eine endliche Menge $\Sigma = \{z_1, ..., z_n\}$ (die wir uns als Alphabet bzw. Zeichenvorrat vorstellen) und eine Abbildung (genannt Zufallsvariable) $X : \Omega \to \Sigma$, die ein zufällig ausgewähltes Zeichen modelliert.

# Diskrete, endliche Wahrscheinlichkeitstheorie

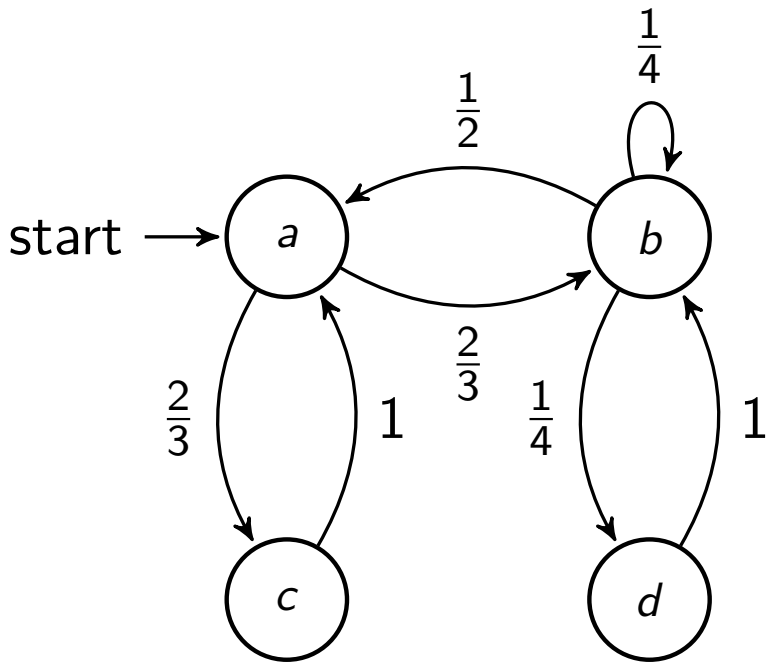
- Wir betrachten eine endliche Menge $\Omega$ (die nicht näher spezifiziert ist) und eine Wahrscheinlichkeitsverteilung $p : \Omega \to [0, 1]$, die jedem Element aus $\Omega$ eine Wahrscheinlichkeit zuordnet, derart dass die Summe über alle Wahrscheinlichkeiten 1 ergibt.
- Wir betrachten eine endliche Menge $\Sigma = \{z_1, ..., z_n\}$ (die wir uns als Alphabet bzw. Zeichenvorrat vorstellen) und eine Abbildung (genannt Zufallsvariable) $X : \Omega \to \Sigma$, die ein zufällig ausgewähltes Zeichen modelliert.
- Der "Zufall" steckt bei diesem Modell in $(\Omega, p)$. Dadurch erhalten wir mit $X$ auch eine Wahrscheinlichkeitsverteilung auf $\Sigma$, nämlich

$$p^X : \Sigma \to [0, 1],\ p^X(z_i) := P(X = z_i) := \sum_{\omega \in X^{-1}(\{z_i\})} p(\omega).$$

# Markov-Ketten

## Graphische Darstellung

- Um eine Markov-Kette anzugeben, gibt man nicht Wahrscheinlichkeiten für die Zeichen an, sondern *Übergangswahrscheinlichkeiten*, für die Wahrscheinlichkeit, dass auf $z_i$ ein $z_j$ folgt. Man kann eine Markov-Kette graphisch darstellen:




- Typische Ausgabe: *acababbdbdbacabd*...

# Normalverteilung

## Definition

- Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ heißt *normalverteilt* ($X \sim \mathcal{N}(\mu, \sigma)$), wenn $P(X \leq k) = \int_0^k \phi(x)dx$, wobei $\phi$ eine Funktion ist, die von den *Parametern* der Normalverteilung abhängt. Diese Parameter sind Erwartungswert $\mu$ und Standardabweichung $\sigma$.

$$\phi(x) = \phi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right).$$

# Normalverteilung

## Definition

- Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ heißt *normalverteilt* ($X \sim \mathcal{N}(\mu, \sigma)$), wenn $P(X \leq k) = \int_0^k \phi(x)dx$, wobei $\phi$ eine Funktion ist, die von den *Parametern* der Normalverteilung abhängt. Diese Parameter sind Erwartungswert $\mu$ und Standardabweichung $\sigma$.

$$\phi(x) = \phi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right).$$

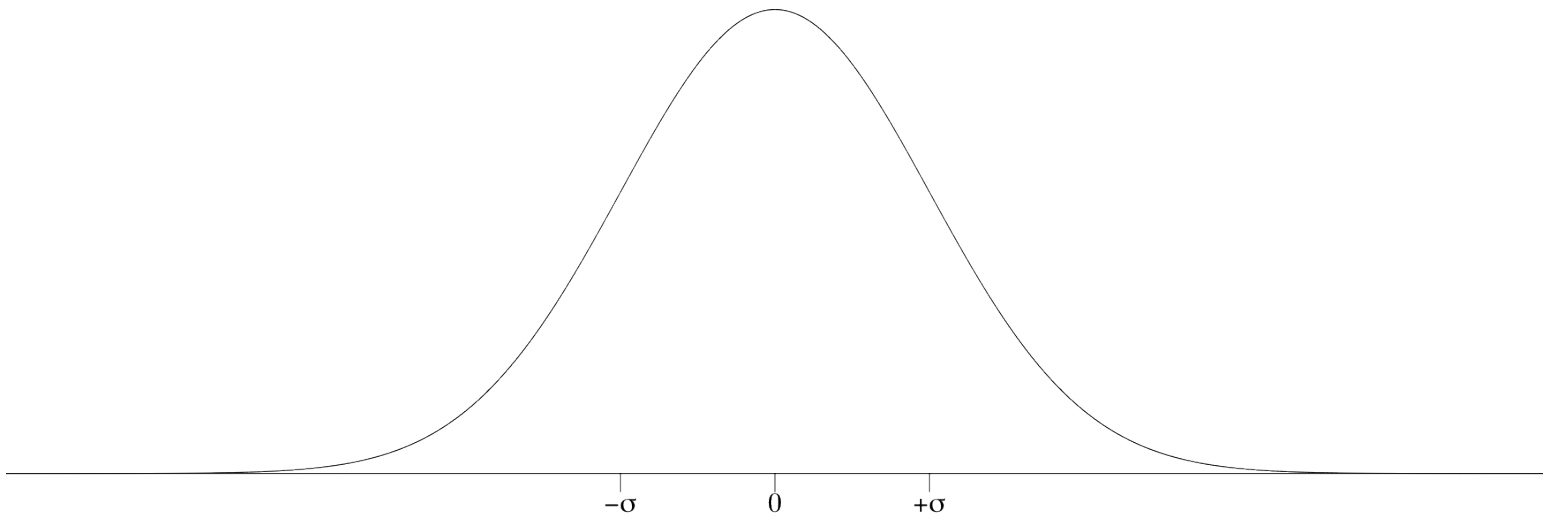
- Im Falle der *Standard-Normalverteilung* $\mathcal{N}(0, 1)$ ist $\mu = 0$ und $\sigma = 1$, dann ist

$$\phi(x) = \phi_{0,1}(x) = \frac{1}{2\pi} e^{-x^2}$$

# Normalverteilung

## Gausssche Glockenkurve

- Der Graph der Dichte der Normalverteilung $\phi$ ist die bekannte *Gausssche Glockenkurve*.




- Man spricht im Kontext der Rauschanalyse nicht von der Normalverteilung sondern von *weissem Gaussschen Rauschen*.

# Wiederholung: Entropie

## Definition

- Entropie ist ein Maß für die *Unsicherheit*, die mit einer Zufallsvariable verbunden wird.

# Wiederholung: Entropie

## Definition

- Entropie ist ein Maß für die *Unsicherheit*, die mit einer Zufallsvariable verbunden wird.
- Ihre Maßeinheit in der Shannonschen Theorie ist das *bit*, abgekürzt für *binary digit*.

# Wiederholung: Entropie

## Definition

- Entropie ist ein Maß für die *Unsicherheit*, die mit einer Zufallsvariable verbunden wird.
- Ihre Maßeinheit in der Shannonschen Theorie ist das *bit*, abgekürzt für *binary digit*.
- Sei $X$ eine diskrete Zufallsvariable mit Werten in einem endlichen Zeichenvorrat $\Sigma = \{z_1, .., z_n\}$ und Verteilungsfunktion $p : \Sigma \to [0, 1]$. Dann ist die *Entropie* von $X$ definiert als

$$H(X) := \mathbb{E}(I_X) = -\sum_{i=1}^{n} p(z_i) \log_2 p(z_i).$$

# Tukey vs. Shannon

## Definition

- Shannon erwähnt als erster schriftlich den Begriff “bit”, gibt aber an, dass er von Tukey stammt.

# Tukey vs. Shannon

## Definition

- Shannon erwähnt als erster schriftlich den Begriff “bit”, gibt aber an, dass er von Tukey stammt.
- Tukey nutze aber den Begriff “Bit” ausschließlich im Sinne des Speicherns von Daten. Die Zeichenkette 000000 bestand für Tukey aus 6 Bits, selbst wenn von vornherein klar wäre, dass sie nur aus Nullen bestünde.

# Tukey vs. Shannon

## Definition

- Shannon erwähnt als erster schriftlich den Begriff “bit”, gibt aber an, dass er von Tukey stammt.
- Tukey nutze aber den Begriff “Bit” ausschließlich im Sinne des Speicherns von Daten. Die Zeichenkette 000000 bestand für Tukey aus 6 Bits, selbst wenn von vornherein klar wäre, dass sie nur aus Nullen bestünde.
- In Shannons Sinne kommt es auf die Verteilung an. Sind die Zeichenketten, die von einer Quelle stammen etwa so verteilt, dass nur entweder Nullen oder ausschliesslich Einsen erwartet werden können, so enthält die Zeichenkette 000000 genau ein bit an Information (denn sie teilt uns mit, dass eben nicht 111111 gesendet wurde).

# Verbundwahrscheinlichkeiten

## Definition

- Gegeben zwei diskrete Zufallsvariable $X$ und $Y$, bezeichnet man die Wahrscheinlichkeit $P(X = x \text{ und } Y = y)$ als *Verbundwahrscheinlichkeit* und notiert auch $P(X = x, Y = y)$. Mengentheoretisch entspricht dem Wort "und" der Schnitt der Ereignisse $\{X = x\} \cap \{Y = y\}$ innerhalb einer Menge $\Omega$ auf der $X$ und $Y$ definiert sind.

# Verbundentropie

## Entropie eines zusammengesetzten Systems

- Gegeben zwei diskrete Zufallsvariable $X$ und $Y$ mit Werten in einem endlichen Zeichenvorrat $\{z_1, ..., z_n\}$, bezeichnet man die Größe $H(X, Y) := \mathbb{E}(I_{X,Y})$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{n} P(X = z_i, Y = z_j) \log_2 (P(X = z_i, Y = z_j))$$

  als *Verbundentropie* von $X$ und $Y$.
- Verbundentropie im Verhältnis zur Entropie der Teilsysteme:

$$H(X) + H(Y) \geq H(X, Y) \geq \max(H(X), H(Y))$$

# Bedingte Wahrscheinlichkeiten

## Definition

- Mathematisch definiert man für zwei Zufallsvariable $X$ und $Y$ die bedingte Wahrscheinlichkeit

$$P(X = x|Y = y) := \frac{P(X = x, Y = y)}{P(Y = y)}.$$

- Es folgt direkt aus der Definition:

$$P(X = x, Y = y) = P(X = x|Y = y)P(Y = y).$$

# Unabhängige Ereignisse

## Definition

- $\{X = x\}$ und $\{Y = y\}$ heißen *unabhängig*, wenn $P(X = x, Y = y) = P(X = x)P(Y = y)$.
- In diesem Fall ist

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

$$= \frac{P(X = x)P(Y = y)}{P(Y = y)} = P(X = x),$$

also hängt die Wahrscheinlichkeit für $\{X = x\}$ nicht davon ab, ob das Ereignis $\{Y = y\}$ eintritt.

# Gemeinsame Information

- Die Größe

$$I(X;Y) := \sum_{i=1}^{n}\sum_{j=1}^{n} P(X = z_i, Y = z_j)\log_2\left(\frac{P(X = z_i, Y = z_j)}{P(X = z_i)P(Y = z_j)}\right)$$

heißt *gemeinsame Information* von $X$ und $Y$, sie misst die Menge an Information, die $X$ und $Y$ gemeinsam haben.

# Gemeinsame Information

- Die Größe

$$I(X;Y) := \sum_{i=1}^{n}\sum_{j=1}^{n} P(X = z_i, Y = z_j) \log_2 \left( \frac{P(X = z_i, Y = z_j)}{P(X = z_i)P(Y = z_j)} \right)$$

  heißt *gemeinsame Information* von $X$ und $Y$, sie misst die Menge an Information, die $X$ und $Y$ gemeinsam haben.

- Entsprechend lässt sich leicht beweisen:

$$I(X;Y) = H(X) + H(Y) - H(X,Y).$$

# Gemeinsame Information

- Die Größe

$$I(X;Y) := \sum_{i=1}^{n}\sum_{j=1}^{n} P(X = z_i, Y = z_j) \log_2 \left( \frac{P(X = z_i, Y = z_j)}{P(X = z_i)P(Y = z_j)} \right)$$

  heißt *gemeinsame Information* von $X$ und $Y$, sie misst die Menge an Information, die $X$ und $Y$ gemeinsam haben.

- Entsprechend lässt sich leicht beweisen:

$$I(X;Y) = H(X) + H(Y) - H(X,Y).$$

- Unabhängige Zufallsvariablen haben daher keine gemeinsame Information, $I(X;Y) = 0$.

# Gemeinsame Information

- Die Größe

$$I(X;Y) := \sum_{i=1}^{n}\sum_{j=1}^{n} P(X = z_i, Y = z_j) \log_2 \left( \frac{P(X = z_i, Y = z_j)}{P(X = z_i)P(Y = z_j)} \right)$$

  heißt *gemeinsame Information* von $X$ und $Y$, sie misst die Menge an Information, die $X$ und $Y$ gemeinsam haben.
- Entsprechend lässt sich leicht beweisen:

$$I(X;Y) = H(X) + H(Y) - H(X,Y).$$

- Unabhängige Zufallsvariablen haben daher keine gemeinsame Information, $I(X;Y) = 0$.
- $I(X;X)$ ist die *Selbstinformation* von $X$. Es ist

$$I(X;X) = H(X).$$

# Bedingte Entropie

## Auch genannt: Equivokation

- Gegeben zwei diskrete Zufallsvariable $X$ und $Y$ mit Werten in einem endlichen Zeichenvorrat $\{z_1, ..., z_n\}$, bezeichnet man die Größe $H(X|Y) := H(X, Y) - H(Y)$ als *bedingte Entropie* von $X$ unter der Bekanntheit von $Y$.

# Bedingte Entropie

## Auch genannt: Equivokation

- Gegeben zwei diskrete Zufallsvariable $X$ und $Y$ mit Werten in einem endlichen Zeichenvorrat $\{z_1, ..., z_n\}$, bezeichnet man die Größe $H(X|Y) := H(X, Y) - H(Y)$ als *bedingte Entropie* von $X$ unter der Bekanntheit von $Y$.
- Man kann auch direkter definieren (und die vorherige Definition daraus herleiten):

$$H(X|Y) := \sum_{i=1}^{n} \sum_{j=1}^{n} P(X = z_i, Y = z_j) \log_2 \left( \frac{P(Y = z_j)}{P(X = z_i, Y = z_j)} \right)$$

# Bedingte Entropie

## Auch genannt: Equivokation

- Gegeben zwei diskrete Zufallsvariable $X$ und $Y$ mit Werten in einem endlichen Zeichenvorrat $\{z_1, ..., z_n\}$, bezeichnet man die Größe $H(X|Y) := H(X, Y) - H(Y)$ als *bedingte Entropie* von $X$ unter der Bekanntheit von $Y$.
- Man kann auch direkter definieren (und die vorherige Definition daraus herleiten):

$$H(X|Y) := \sum_{i=1}^{n} \sum_{j=1}^{n} P(X = z_i, Y = z_j) \log_2 \left( \frac{P(Y = z_j)}{P(X = z_i, Y = z_j)} \right)$$

- Im Falle unabhängiger Zufallsvariablen gilt

$$H(X|Y) = H(X).$$

# Beispiel

## Bedingte Wahrscheinlichkeiten von Zeichenfolgen

- In erster Näherung wollen wir eine Nachricht in englischer Sprache modellieren durch eine Folge unabhängiger Zufallsvariablen $X_i$ mit Werten im Alphabet $\Sigma := \{A, B, C, ..., Y, Z, \_\}$ und Gleichverteilung, also $P(X_i = z) = \frac{1}{27}$ für alle $z \in \Sigma$. Dabei ist _ das Leerzeichen.

# Beispiel

## Bedingte Wahrscheinlichkeiten von Zeichenfolgen

- In erster Näherung wollen wir eine Nachricht in englischer Sprache modellieren durch eine Folge unabhängiger Zufallsvariablen $X_i$ mit Werten im Alphabet $\Sigma := \{A, B, C, ..., Y, Z, \_\}$ und Gleichverteilung, also $P(X_i = z) = \frac{1}{27}$ für alle $z \in \Sigma$. Dabei ist _ das Leerzeichen.
- Die Selbstinformation eines Zeichens $z \in \Sigma$ ist somit $I_i(z) = -\log_2 P(X_i = z) = -\log_2 \frac{1}{27} = \log_2 27 \approx 4,75\text{bit}$.

# Beispiel

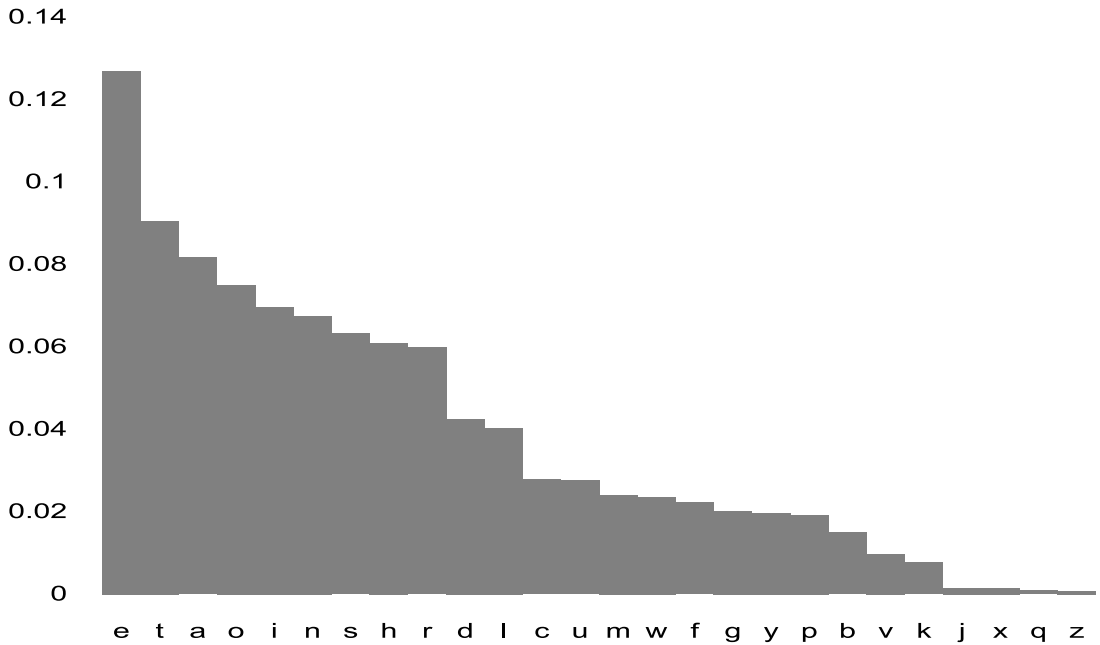
## Bedingte Wahrscheinlichkeiten von Zeichenfolgen

- In erster Näherung wollen wir eine Nachricht in englischer Sprache modellieren durch eine Folge unabhängiger Zufallsvariablen $X_i$ mit Werten im Alphabet $\Sigma := \{A, B, C, ..., Y, Z, \_\}$ und Gleichverteilung, also $P(X_i = z) = \frac{1}{27}$ für alle $z \in \Sigma$. Dabei ist _ das Leerzeichen.
- Die Selbstinformation eines Zeichens $z \in \Sigma$ ist somit $I_i(z) = -\log_2 P(X_i = z) = -\log_2 \frac{1}{27} = \log_2 27 \approx 4,75\text{bit}$.
- Die Entropie der Zufallsvariablen $X_i$ ist damit auch etwa $4,75\text{bit}$ (wegen Gleichverteilung).

# Beispiel

## Eine realistischere Verteilung der Zeichen

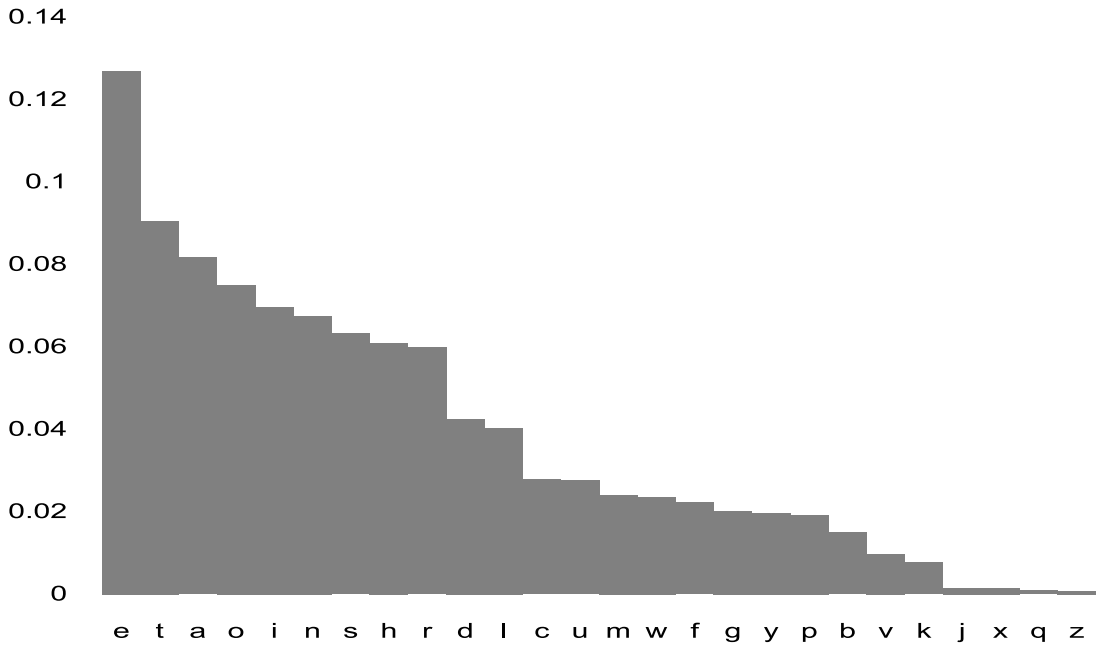
- Die Häufigkeit der Buchstaben in der englischen Sprache ist nicht gleichverteilt:



# Beispiel

## Eine realistischere Verteilung der Zeichen

- Die Häufigkeit der Buchstaben in der englischen Sprache ist nicht gleichverteilt:

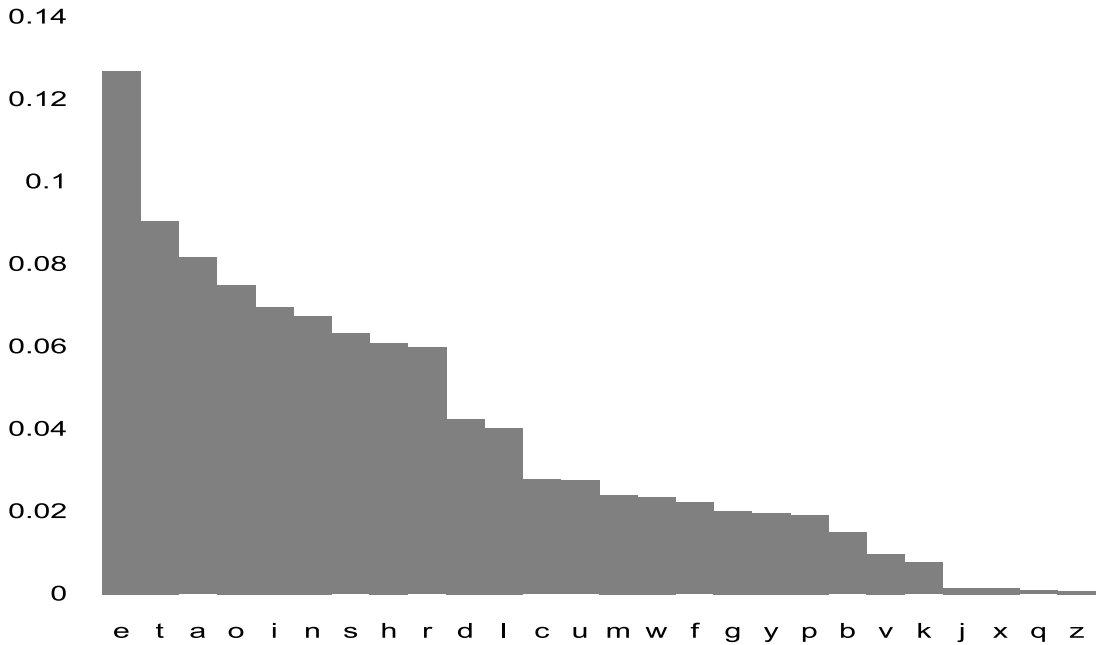



- Die Selbstinformation eines Zeichens $z \in \Sigma$ ist somit z.B. $I_i(e) = -\log_2 P(X_i = e) = 3\text{bit}$ oder aber $I_i(v) = 6,64\text{bit}$.

# Beispiel

## Eine realistischere Verteilung der Zeichen

- Die Häufigkeit der Buchstaben in der englischen Sprache ist nicht gleichverteilt:




- Die Selbstinformation eines Zeichens $z \in \Sigma$ ist somit z.B. $I_i(e) = -\log_2 P(X_i = e) = 3\text{bit}$ oder aber $I_i(v) = 6,64\text{bit}$.
- Die Entropie der Zufallsvariablen $X_i$ und damit $H(X)$ lässt sich anhand dieser Tabelle berechnen: $\approx 4\text{bit}$.

# Besser: Markov-Ketten

# Kullback-Leibler-Divergenz

## Ein geringfügig allgemeinerer Begriff als *bedingte Entropie*

- Für zwei Wahrscheinlichkeitsmaße (z.B. Verteilungen von diskreten Zufallsvariablen oder aber irgendwelche stetigen Maße) $P$ und $Q$ auf einem Raum (genau genommen einer $\sigma$-Algebra) $\Omega$ sodass $Q$ *absolutstetig* bezüglich $P$ ist (im Falle diskreter Verteilungen genügt $P(i) > 0 \implies Q(i) > 0$), definiert man die *Kullback-Leibler-Divergenz* (auch: *Informationsgewinn*) von $P$ bezüglich $Q$ als

$$D_{KL}(P||Q) := -\int_{\Omega} \log \frac{dQ}{dP} dP,$$

wobei $\frac{dQ}{dP}$ die *Radon-Nikodym-Ableitung* von $Q$ nach $P$ ist.

# Satz von Bayes

- Der *Satz von Bayes* besagt, dass für zwei Zufallsvariable $X$ und $Y$ gilt:

$$P(X = x|Y = y) = \frac{P(Y = y|X = x)P(X = x)}{P(Y = y)}$$

# Bayessche Inferenz und Abduktion

# Bayes für bedingte Entropie

## Korollar

- Der Satz von Bayes
  $P(X = x|Y = y) = P(Y = y|X = x)\frac{P(X=x)}{P(Y=y)}$ impliziert

$$H(X|Y) = H(Y|X) + H(X) - H(Y)$$

# Bayes für bedingte Entropie

## Korollar

- Der Satz von Bayes
  $P(X = x|Y = y) = P(Y = y|X = x)\frac{P(X=x)}{P(Y=y)}$ impliziert

$$H(X|Y) = H(Y|X) + H(X) - H(Y)$$

- Es folgt durch weitere Anwendung dieses Satzes auf $-H(Z|X)$ und Addition:

$$H(X|Y) - H(Z|X) = H(Y|X) - H(X|Z) + H(Z) - H(Y),$$

eine Formel, in der $H(X)$ nicht mehr auftaucht!

# Korrelation

## Definition

- Für zwei Zufallsvariable $X$ und $Y$ heißt

$$\mathrm{Cov}(X, Y) := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

die *Kovarianz* von $X$ und $Y$.

# Korrelation

## Definition

- Für zwei Zufallsvariable $X$ und $Y$ heißt

$$\mathrm{Cov}(X, Y) := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

  die *Kovarianz* von $X$ und $Y$.
- Es ist $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

# Korrelation

## Definition

- Für zwei Zufallsvariable $X$ und $Y$ heißt

$$\mathrm{Cov}(X, Y) := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

  die *Kovarianz* von $X$ und $Y$.
- Es ist $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
- $X$ und $Y$ heißen *unkorreliert*, wenn $\mathrm{Cov}(X, Y) = 0$ ist.

# Korrelation

## Definition

- Für zwei Zufallsvariable $X$ und $Y$ heißt

$$\mathrm{Cov}(X, Y) := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

  die *Kovarianz* von $X$ und $Y$.
- Es ist $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
- $X$ und $Y$ heißen *unkorreliert*, wenn $\mathrm{Cov}(X, Y) = 0$ ist.
- Für unabhängige Zufallsvariable ist $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, also sind unabhängige Zufallsvariable stets unkorreliert.

# Zusammenfassung der wichtigsten Größen

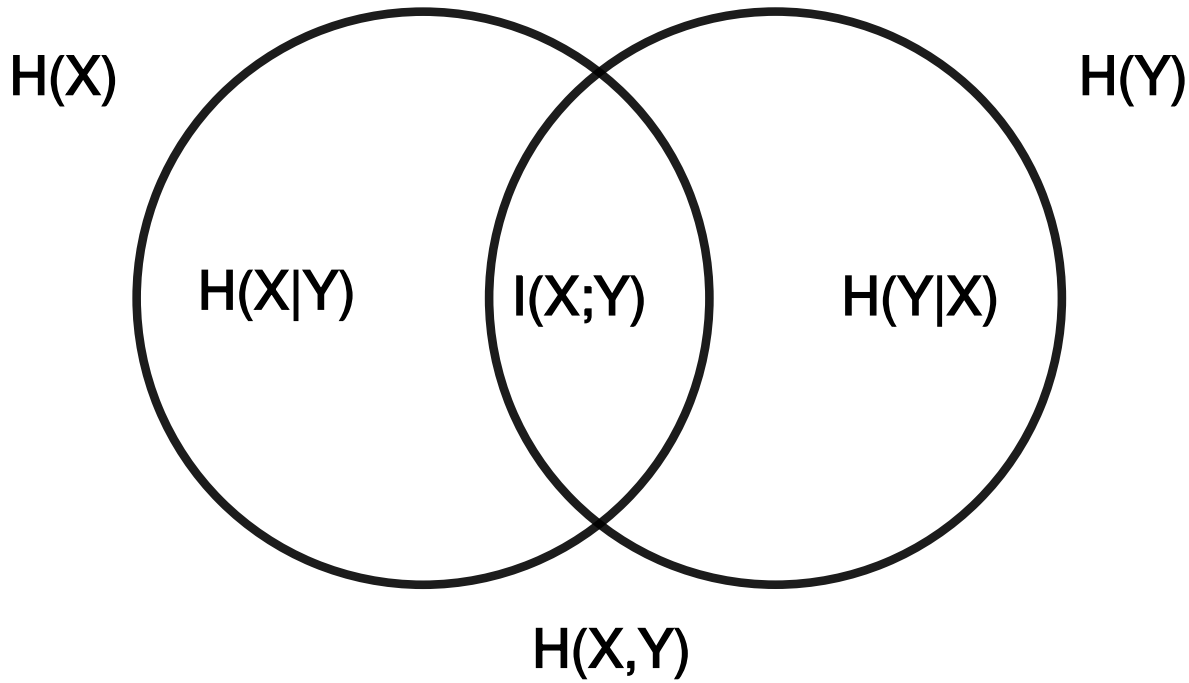
## Mathematisch

- Eine Zufallsvariable $X : \Omega \to \Sigma = \{z_1, ..., z_n\}$ ist bestimmt durch ihre *Verteilungsfunktion* $z \mapsto P(X = z)$.
- Die *Selbstinformation* von $X$ ist die Funktion $I_X : z \mapsto -\log_2 P(X = z)$.
- Die *Entropie* von $X$ ist der Erwartungswert der Selbstinformation: $H(X) := \mathbb{E}(I_X)$.
- Die *Verbundentropie* von $X$ und $Y$ ist die Entropie der komponierten Zufallsvariablen: $H(X, Y) := H((X, Y))$.
- Die *gemeinsame Information* von $X$ und $Y$ ist $I(X; Y) = H(X) + H(Y) - H(X, Y)$.
- Die *bedingte Entropie* von $X$ bezüglich $Y$ ist $H(X|Y) = H(X, Y) - H(Y)$.
- Alle Definitionen lassen sich auf den Begriff der bedingten Entropie zurückführen (und dies ist ein Spezialfall der Kullback-Leibler-Divergenz, und lässt sich somit auf stetige Zufallsvariablen übertragen).

# Zusammenfassung der wichtigsten Größen

## Diagrammatisch, allgemeiner Fall

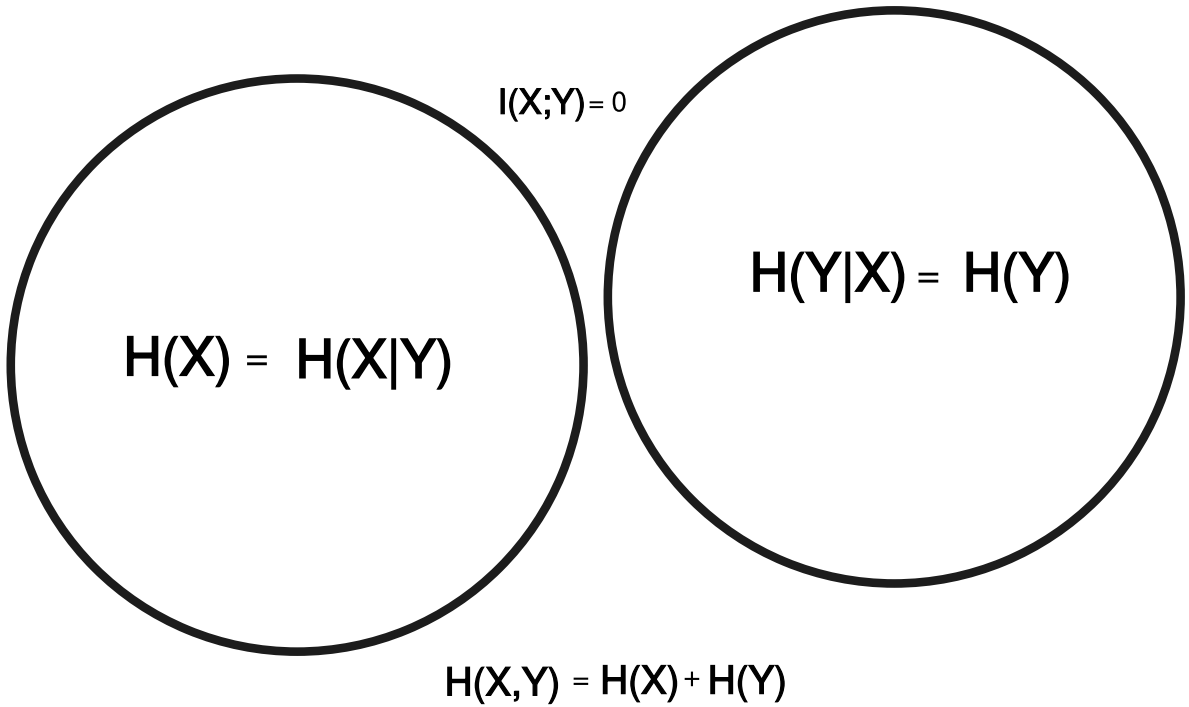
- Für zwei Zufallsvariablen $X$ und $Y$ lassen sich die grundlegenden informationstheoretischen Größen in einem Diagramm darstellen:



# Zusammenfassung der wichtigsten Größen

## Diagrammatisch, Grenzfall: Unabhängigkeit

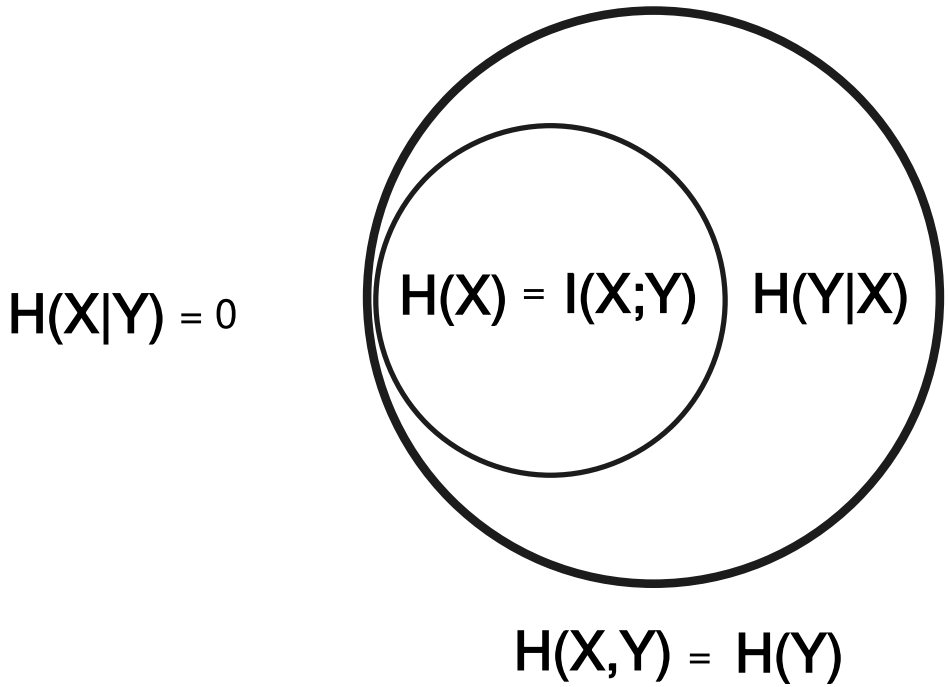
- Für zwei **unabhängige** Zufallsvariablen $X$ und $Y$ sieht das Diagramm so aus:



# Zusammenfassung der wichtigsten Größen

## Diagrammatisch, Grenzfall: Unabhängigkeit

- Für zwei **vollständig abhängige** Zufallsvariablen $X$ und $Y$ sieht das Diagramm (bis auf evtl. Umbenennung) so aus:



# Redundanz und Kodierung

- Die optimale Kodierung im rauschfreien Kanal enthält keine Redundanz (minimale Kodierungslänge).
- Bei Rauschen ist Redundanz unbedingt notwendig, um fehlerfrei eine Nachricht zu rekonstruieren (zu dekodieren).

# Redundanz in der natürlichen Sprache

# Beispiele für Codes

## Morse-Code, ISBN, MP3

# Die Bedeutung des Satzes von Bayes

## In der Kodierungstheorie

- Gegeben (unbekannte) gesendete Daten $x$, modelliert durch eine Zufallsvariable $X$ mit bekannter Verteilung (gegeben durch Bekanntheit des verwendeten Codes) und gestörte (bekannte) Daten $y$, modelliert durch eine Zufallsvariable $Y$ mit bekannter Verteilung (im Zweifelsfall nimmt man hier $Y = X + \mathcal{N}$ an, ein weisses Rauschen), so ist das gesuchte, aber unbekannte $x$ dasjenige, welches die Größe

$$P(X = x|Y = y) = P(Y = y|X = x)\frac{P(X = x)}{P(Y = y)}$$

maximiert (denn sonst wäre der Code ungeeignet gewählt für das gegebene Rauschniveau).

# Kanalkapazität

- Die *Kanalkapazität* ist definiert als die stärkste obere Schranke für die Menge an Information, die über einen Kommunikationskanal übertragen werden kann.
- Formal definiert man

$$C := \sup_X(I(X;Y)),$$

$X$ modelliert die gesendete Nachricht, $Y$ die empfangene (verrauschte).

# Shannons Quellenkodierungstheorem

- Ein Code, der im Mittel mehr bits pro Zeichen verwendet als die Entropie der Zeichenverteilung, ist redundant.
- Das Quellenkodierungstheorem sagt nun, dass man bei Abwesenheit von Rauschen nicht auf Redundanz verzichten kann, diese jedoch beliebig klein halten kann.

# Shannons Theorem für verrauschte Kanäle

- Im Gegensatz zum unverrauschten Kanal lässt sich bei Anwesenheit von Rauschen keine Kodierung finden, die beliebig nah an das Limit der Entropie eines Zeichens herankommt.
- Es ist notwendig, zur Fehlerkorrektur einen Code mit längerer Kodierungslänge zu wählen. Man fügt also Redundanz hinzu.
- In der gesprochenen Sprache als Kanal für geschriebene Sätze wird die Redundanz der natürlichen Sprache zur Fehlerkorrektur eingesetzt.

# Satz von Wolfowitz

## Klippen-Effekt

- Der Satz von Wolfowitz (1957) besagt, dass es eine endliche positive Konstante $A$ gibt, sodass

$$P_{error} \geq 1 - \frac{4A}{n(R-C)^2} e^{-n(R-C)}.$$

- Daraus folgt, dass für Kommunikation bei Raten oberhalb der Kanalkapazität die Fehlerrate exponentiell gegen 1 geht.
- Darum brechen Handy-Gespräche bei schlechtem Empfang oft auch abrupt vollständig ab, anstatt zunächst mit schlechterer Übertragungsqualität weiter zu arbeiten.

# Weisses Rauschen als Modell

- Weisses Rauschen ist nicht nur dann ein gutes Modell, wenn die Kommunikation nur durch Thermalstrahlung gestört wird.

# Weisses Rauschen als Modell

- Weisses Rauschen ist nicht nur dann ein gutes Modell, wenn die Kommunikation nur durch Thermalstrahlung gestört wird.
- Die Summe (und damit das arithmetische Mittel) unabhängiger, normalverteilter Zufallsvariablen ist wieder Normalverteilt.

# Weisses Rauschen als Modell

- Weisses Rauschen ist nicht nur dann ein gutes Modell, wenn die Kommunikation nur durch Thermalstrahlung gestört wird.
- Die Summe (und damit das arithmetische Mittel) unabhängiger, normalverteilter Zufallsvariablen ist wieder Normalverteilt.
- Nach dem zentralen Grenzwertsatz konvergiert die Verteilung von $n$ unabhängigen identisch (nicht notwendig normal) verteilten Zufallsvariablen für großes $n$ gegen eine Normalverteilung.