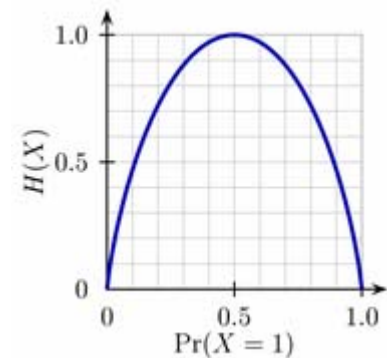


# Information und Kommunikation

Hartmut Klauck  
Universität Frankfurt  
SS 07  
7.5.



# Gemeinsam Typische Sequenzen

- Wir betrachten Zufallsvariablen  $X_1, \dots, X_n, Y_1, \dots, Y_n$
- $X_i, Y_i$  haben für jedes  $i$  die Verteilung  $p(x, y)$
- Die Zufallsvariablen entsprechen Ein- und Ausgaben eines Kanals der  $n$  mal (unabhängig) benutzt wird
- Wir wollen später dekodieren, indem wir zur Ausgabe  $Y_1, \dots, Y_n$  ein  $X_1, \dots, X_n$  finden, so dass beide *gemeinsam typisch* sind

# Gemeinsam Typische Sequenzen

- **Definition 7.1**

- Gegeben sei eine Verteilung  $p(x,y)$  auf  $X \times Y$
- Die Menge  $A_\varepsilon$  der gemeinsam typischen Sequenzen ist die Menge derjenigen strings in  $X^n \times Y^n$  mit:

$$1) \quad \left| -\frac{1}{n} \cdot \log(p_x(x_1, \dots, x_n)) - H(X) \right| \leq \varepsilon$$

$$2) \quad \left| -\frac{1}{n} \log(p_y(y_1, \dots, y_n)) - H(Y) \right| \leq \varepsilon$$

$$3) \quad \left| -\frac{1}{n} \log(p(x_1, \dots, x_n, \dots, y_n)) - H(X, Y) \right| \leq \varepsilon$$

# Gemeinsam Typische Sequenzen

- **Theorem 7.2**

1.  $\text{Prob}(A_\varepsilon) \geq 1 - \varepsilon$  für genügend große  $n$

2.  $|A_\varepsilon| \leq 2^{n(H(X,Y) + \varepsilon)}$

3. Wenn  $U_1, \dots, U_n$  bzw.  $V_1, \dots, V_n$  verteilt sind wie  $p_X$  bzw.  $p_Y$  (Grenzverteilungen von  $p$ ), dann gilt:

$$\text{Prob}(U_1, \dots, U_n, V_1, \dots, V_n \in A_\varepsilon) \leq 2^{-n(I(X:Y) - 3\varepsilon)}$$

$$\text{Prob}(U_1, \dots, U_n, V_1, \dots, V_n \in A_\varepsilon) \geq 2^{-n(I(X:Y) + 3\varepsilon)}$$

(für genügend große  $n$ )

# Beweis

- 1):
  - Wie für typische Sequenzen gilt:  
 $E[-1/n \cdot \log p(X_1, \dots, X_n)] = H(X)$
  - Mit Chebyshev folgt: für alle  $\varepsilon > 0$  gibt es ein  $n_0$ , für alle  $n \geq n_0$ :  
 $\text{Prob}(|-1/n \cdot \log p(X_1, \dots, X_n) - H(X)| \geq \varepsilon) \leq \varepsilon/3$
  - Genauso:  
für jedes  $\varepsilon > 0$  gibt es ein  $n_1$ , für alle  $n \geq n_1$ :  
 $\text{Prob}(|-1/n \cdot \log p(Y_1, \dots, Y_n) - H(Y)| \geq \varepsilon) \leq \varepsilon/3$
  - Und:  
für jedes  $\varepsilon > 0$  gibt es ein  $n_0$ , für alle  $n \geq n_2$ :  
 $\text{Prob}(|-1/n \cdot \log p(X_1, \dots, X_n, Y_1, \dots, Y_n) - H(XY)| \geq \varepsilon) \leq \varepsilon/3$

# Beweis

- Daher sind alle 3 Bedingungen mit Ws.  $1-\varepsilon$  erfüllt, wenn  $n \geq n_0, n_1, n_2$

• 2):

$$I = \sum_{\vec{x}, \vec{y}} P(\vec{x}, \vec{y})$$

$$\geq \sum_{\vec{x}, \vec{y} \in A_\varepsilon} P(\vec{x}, \vec{y})$$

$$\geq \sum_{\vec{x}, \vec{y} \in A_\varepsilon} 2^{-n(H(x, y) + \varepsilon)}$$

$$= 2^{-n(H(x, y) + \varepsilon)} \cdot |A_\varepsilon|$$

# Beweis

• 3)  $U_1, \dots, U_n$  sei mit  $(p_X)^n$  verteilt

-  $V_1, \dots, V_n$ , mit  $(p_Y)^n$

- Es gilt:

$$\text{Prob}(\vec{U}, \vec{V} \in A_\epsilon)$$

$$= \sum_{\vec{x}, \vec{y} \in A_\epsilon} P_X(\vec{x}) \cdot P_Y(\vec{y})$$

$$\leq \sum_{\vec{x}, \vec{y} \in A_\epsilon} 2^{-n(H(X)-\epsilon)} \cdot 2^{-n(H(Y)-\epsilon)}$$

$$\leq 2^{n \cdot (H(X, Y) + \epsilon)} \cdot 2^{-n(H(X)-\epsilon)} \cdot 2^{-n(H(Y)-\epsilon)}$$

$$= 2^{-n(I(X:Y) - 3\epsilon)}$$

# Bemerkungen

- Es gibt also ungefähr  $2^{n H(X)}$  typische  $X$ -Sequenzen,  $2^{n H(Y)}$  typische  $Y$ -Sequenzen, aber nur  $2^{n H(X,Y)}$  gemeinsam typische Sequenzen
- Paare von typischer  $X$ -Sequenz und typischer  $Y$ -Sequenz sind nicht immer gemeinsam typisch
- Die Wahrscheinlichkeit dass ein Paar gemeinsam typisch ist, ist nur  $\approx 2^{-n I(X:Y)}$ , wenn  $X$ -Sequenz und  $Y$ -Sequenz unabhängig gewählt werden

# Shannons Theorem

- **Theorem 7.3**

Gegeben sei ein Kanal mit Kapazität  $C$  durch die Wahrscheinlichkeiten  $p(y|x)$ .

- Alle Raten  $R$  mit  $R < C$  sind erreichbar.
- Alle Raten  $R > C$  sind nicht erreichbar.

- **Anders fomuliert:**

- Für alle  $R < C$  gibt es  $(2^{Rn}, n)$ -Codes, deren maximaler Fehler gegen 0 geht.
- Wenn der Fehler eines  $(2^{Rn}, n)$ -Codes gegen 0 geht, gilt  $R \leq C$ .

# Shannons Theorem

- Wir beginnen mit der ersten Aussage
- Gegeben sei die Matrix der  $p(y|x)$ .  $p(x)$  sei die Verteilung auf dem Eingabealphabet, auf welcher die Kapazität  $C$  des Kanals erreicht wird
- $R < C$
- Wir müssen einen Code konstruieren
- Wir werden den Code einfach probabilistisch wählen
- Dann zeigen wir, dass mit hoher Wahrscheinlichkeit der Fehler gegen 0 geht
- Daher existiert ein guter Code, auch wenn wir ihn nicht explizit konstruieren

# Der Code

- Wir wählen eine  $(2^{Rn}, n)$ -Code so:
  - Wir bestimmen die Menge der Codeworte
  - Dazu füllen wir eine  $2^{Rn} \times n$ -Matrix auf, indem wir jedes Element gemäß  $p(x)$  wählen
  - Die Zeilen der Matrix sind unsere Codeworte
  - Anders gesagt wählen wir  $2^{Rn}$  Codeworte  $x_1, \dots, x_n$ , indem wir  $x_i$  mit Ws.  $p(x_i)$  wählen

# Der Code

- Daher gilt, dass eine bestimmte Codewortmenge mit  $W$ s.

$$P_r(C) = \prod_{w=1}^{2^n \cdot R} \prod_{i=1}^n P(C_i(W))$$

gewählt wird, dabei sei  $C_i(W)$  das  $i$ -te Zeichen des  $W$ -ten Codeworts

- Um einen Code zu erhalten, bilden wir nun einfach die Elemente von  $\{1, \dots, 2^{Rn}\}$  auf die Codeworte ab (Kodierungsfunktion)
- Die Dekodierungsfunktion müssen wir noch beschreiben (Codeworte  $C(W)$  werden wie durch das Inverse der Kodierungsfunktion dekodiert, d.h. zu  $W$ ).
- Mit sehr geringer Wahrscheinlichkeit ist der Code ungültig, das dasselbe Codewort zweimal vorkommt.

# Der Gesamtablauf

- Wir betrachten den Kommunikationsablauf in folgenden Schritten:
  - Ein zufälliger Code wird erzeugt wie beschrieben und Sender und Empfänger erhalten den Code
  - Eine Nachricht  $W$  wird uniform zufällig gewählt, aus  $\{1, \dots, 2^{Rn}\}$  und dem Sender bekannt gemacht
  - Das  $W$ -te Codewort  $C(W)$  wird vom Sender über den Kanal geschickt (den  $n$ -fachen Produktkanal)
  - Der Empfänger erhält  $Y_1, \dots, Y_n$ , gemäß der Verteilung  $\prod p(y_i | C(W))$
  - Der Empfänger dekodiert.

# Dekodierung

- Es gibt eine optimale Methode der Dekodierung: Maximum-Likelihood Decoding
- Das bedeutet, dass der Empfänger, gegeben einen Wert  $y_1, \dots, y_n$  von  $Y_1, \dots, Y_n$  einen wahrscheinlichsten Wert von  $X_1, \dots, X_n$  bestimmt (unter allen Codeworten)
- Dieses Verfahren hat offensichtlich den kleinsten Fehler
- Allerdings ist es schwierig zu analysieren

# Typical-Set-Decoding

- Wir verwenden daher für die Analyse ein anderes Verfahren:
  - Typical-Set-Decoding
- Der Empfänger gibt  $W'$  aus,
  - wenn  $(C(W'); Y_1, \dots, Y_n)$  gemeinsam typisch sind
  - Wenn es kein  $W''$  mit derselben Eigenschaft gibt
- Im letzteren Fall wird ein dummy ausgegeben