Seminar Kolmogorovkomplexität

Notation und Grundlagen

Nuria Brede

30.10.2009

Notation

- $\mathcal{B}^* = \{0,1\}^*$; $\mathcal{B}^\infty$ Menge der einseitig unendlichen Binärstrings; $\mathcal{N}$ Menge der natürlichen Zahlen; $\mathcal{Z}$ Menge der ganzen Zahlen; $\mathcal{R}$ Menge der reellen Zahlen
- $l(n)$ – Länge der Darstellung einer Zahl n
- $d(A)$ – Kardinalität einer Menge A
- $\log x$ – Logarithmus zur Basis 2; $\ln x$ – Natürlicher Logarithmus
- $\langle \cdot \rangle$ als Standard-Tupelfunktion
- Kennzeichnung d. Stelligkeit einer Funktion: $\phi^{(n)}$ ist n-stellig
- Sei ϕ eine partielle Funktion; $\phi(x) < \infty$: ϕ ist definiert / konvergiert an der Stelle x; $\phi(x) = \infty$: ϕ ist nicht definiert / divergiert an der Stelle x
- 2^A ist die Potenzmenge der Menge A
- Quantoren: $\forall^\infty \equiv$ "für alle bis auf endlich viele"
  $\exists^\infty \equiv$ "es existieren unendlich viele"

Asymptotische Notation

f und g seien Funktionen über den reellen Zahlen.

1. $f(x) = O(g(x))$ falls es Konstanten $c, x_0 > 0$ gibt, so dass $\forall x \geq x_o.\ |f(x)| \leq c|g(x)|$
2. $f(x) = o(g(x))$ falls $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$
3. $f(x) = \Omega(g(x))$ falls $f(x) \neq o(g(x))$
4. $f(x) = \Theta(g(x))$ falls sowohl $f(x) = O(g(x))$ und $f(x) = \Omega(g(x))$

Achtung: Weicht in Punkt (3) und (4) von der in Theorie II benutzten Notation ab!
Vorteil dieser Variante: Ω als Komplement von o

Kombinatorik

- Die Zahl der **Permutationen** n verschiedener Objekte ist $n!$
- Die Zahl der **Variationen** (Reihenfolge wird berücksichtigt) von k aus n Objekten ist n^k
- Die Zahl der Variationen von k **verschiedenen** aus n Objekten ist $(n)_k = \frac{n!}{(n-k)!}$
- Die Zahl der **Kombinationen** (ohne Berücksichtigung der Reihenfolge) ist $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Konkatenation

- $\mathcal{B} = \{0, 1\}$, $\mathcal{B}^* = \{\epsilon, 0, 1, 00, 01, 10, 11, 100, \dots\}$
- **Konkatenation** als binäre Operation auf Elementen von $\mathcal{B}^*$, die einen String xy mit dem geordneten Paar $(x, y) \in \mathcal{B}^* \times \mathcal{B}^*$ assoziiert
- $\mathcal{B}^*$ ist **abgeschlossen unter Konkatenation**: $x, y \in \mathcal{B}^* \Rightarrow xy \in \mathcal{B}^*$
- Konkatenation ist **assoziativ** $(xy)z = x(yz)$ und hat ϵ als **Einselement** $x\epsilon = \epsilon x = x$

Darstellung als Binärstring

- Bijektive Abbildung $\mathcal{B}^* \to \mathcal{N}$ wünschenswert
- Benutzen der **lexikographischen Ordnung** $(\epsilon, 0), (0, 1), (1, 2), (00, 3), (01, 4), (10, 5), (11, 6), \ldots$: Jeder binäre String wird auf seinen **Index** in dieser Ordnung abgebildet
- $x = 2^{n+1} - 1 + \sum_{i=0}^{n} a_i 2^i$ wird durch $a_n \ldots a_1 a_0$ repräsentiert und im Buch als **gleiches Objekt** behandelt
- Die **Länge** $l(x)$ eines Strings ist die **Anzahl seiner Bits**
- Konvention ergibt somit z.B. $l(4) = 2$ wegen $4 = 01$

Standard-Codierung

- Einfache **Präfix-Codes** ergeben sich, wenn man die 0 als **Endmarkierung** benutzt und $x \in \mathcal{N}$ als 1^x0 codiert; indem man jedem Objekt seine **Länge als Präfix** mitgibt und diesen Ansatz **iteriert** erhält man noch **kürzere** Codes:
- $E_i(x) = \begin{cases} 1^x0 \text{ falls } i = 0 \\ E_{i-1}(l(x))x \text{ falls } i > 0 \end{cases}$
- $E_1(x) = 1^{l(x)}0x$ und $l(E_1(x)) = 2l(x) + 1$
- $E_1(x)$ wird im Buch abgekürzt als $\overline{x}$ und als "**selfdelimiting version**" eines binären Strings x bezeichnet
- Manchmal wird das kürzere $E_2(x)$ gebraucht mit
- $E_2(x) = \overline{l(x)}x$ und $l(E_2(x)) = l(x) + 2l(l(x)) + 1$

Einseitig unendliche Binärstrings

- $\omega = \omega_1\omega_2\ldots$ und $\omega_{1:n} = \omega_1\ldots\omega_n$
- Zusammenhang zwischen Elementen eines endlichen Alphabets $\mathcal{A} = \{0, 1, \ldots, k-1\}, k \geq 2$ und $\mathcal{R}$: Sei r eine reelle Zahl mit $0 < r < 1$. Dann gibt es eine Folge $\omega_1\omega_2\ldots$ von Elementen ω_n aus $\mathcal{A}$, so dass $r = \sum_n \frac{\omega_n}{k^n}$ und dass die Folge **eindeutig** ist, ausser wenn r die Form q/k^n hat; in diesem Fall gibt es **genau zwei** Folgen, von denen die eine unendlich viele Nullen hat. Umgekehrt gilt, dass falls $\omega_1\omega_2\ldots$ eine unendliche Folge mit $0 \leq \omega_n < k$ ist, die Reihe $\sum_n \frac{\omega_n}{k^n}$ gegen eine reelle Zahl r mit $0 \leq r \leq 1$ konvergiert. Diese Reihe ist die k-**adische Entwicklung** von r; wenn es zwei verschiedene gibt, wird diejenige mit **unendlich vielen Nullen** gewählt

Einseitig unendliche Binärstrings II

- Definiere die Menge $\mathcal{S} \subseteq \mathcal{B}^\infty$ als die Menge der Folgen, die nicht mit unendlich vielen Stellen "$k-1$" (also "1" im binären Fall) "enden". Dann ist S in **1-1 Korrespondenz mit dem Intervall** $[0,1)$
- Sei x ein endlicher String über $\mathcal{B}$. Die Menge aller einseitig unendlichen Folgen, die mit x anfangen, heisst **Zylinder** und wird Γ_x geschrieben: $\Gamma_x = \{x\omega : \omega \in \mathcal{B}^\infty\}$ mit $x \in \mathcal{B}^*$
- $\Gamma_y \subseteq \Gamma_x$ g.d.w. x **Präfix** von y ist. Die Präfix-Relation induziert eine **partielle Ordnung auf den Zylindern** von $\mathcal{B}^\infty$.

Stichprobenraum

- **Stichprobenraum** (Ergebnisraum) S: Die Menge aller möglichen **Ergebnisse eines Zufallsexperiments**
- S **diskret**: endlich oder abzählbar unendlich
- S **stetig**: überabzählbar unendlich
- Ergebnisraum $\neq$ Ereignisraum!

Kolmogorovs Axiome der Wahrscheinlichkeitstheorie

Sei S der Stichprobenraum.

1. Wenn A und B Ereignisse sind, so sind auch ihre **Vereinigung** $A \cup B$, ihr **Schnitt** $A \cap B$ und ihre **Differenz** $A - B$ Ereignisse.
2. S heisst "**sicheres Ereignis**"; $\emptyset$ heisst "**unmögliches Ereignis**".
3. Jedem Ereignis E ist eine nicht-negative reelle Zahl $P(E)$ zugeordnet, die **Wahrscheinlichkeit** dieses Ereignisses.
4. $P(S) = 1$.
5. A und B heissen **unabhängig,** wenn $P(A \cap B) = P(A)P(B)$; dann gilt auch $P(A \cup B) = P(A) + P(B)$.
6. Für eine absteigende Folge $A_1 \supset A_2 \supset \cdots \supset A_n \supset \ldots$ mit $\bigcap_n A_n = \emptyset$ gilt $lim_{n\to\infty} P(A_n) = 0$.

Wahrscheinlichkeitsraum

(mit diskretem Stichprobenraum)

- Ein System $\mathcal{F}$ von Mengen $f \subseteq S$, das unter **Vereinigung**, **Schnitt** und **Differenz abgeschlossen** ist, ein **1-Element** (S) und ein **0-Element** $(\emptyset)$ (bezüglich des Schnitts) besitzt, heisst **Ereignisalgebra** oder **Ereignisraum**
- Eine Funktion $P : \mathcal{F} \longrightarrow [0, 1]$ heisst **Wahrscheinlichkeitsmaß** (**Verteilung**) auf $\mathcal{F}$, wenn die Axiome erfüllt sind
- Im Buch wird eine solche Menge von Ereignissen $\mathcal{F}$ zusammen mit dem dem **Wahrscheinlichkeitsmaß** P, auf $\mathcal{F}$, als Tupel $(\mathcal{F}, P)$ **Wahrscheinlichkeitsraum** genannt; üblich ist aber auch die Schreibweise als Tripel $(S, \mathcal{F}, P)$

Wahrscheinlichkeitsraum

(mit stetigem Stichprobenraum)

- Falls der Ereignisraum $\mathcal{F}$ **unendlich** ist und **alle abzählbaren Vereinigungen** $\bigcup A_n$ enthält, wird er **(Borelsche)** σ**-Algebra** genannt
- Eine **Borel Erweiterung** (σ, P^*) ist für **jeden unendlichen Wahrscheinlichkeitsraum** $(\mathcal{F}, P)$ **möglich und eindeutig** durch Erweiterung von $\mathcal{F}$ und P unter abzählbarer Vereinigung, so dass die Axiome 1-6 weiterhin erfüllt werden

Zufallsvariable und Verteilungsfunktion

- **Zufallsvariable**: Messbare Funktion $X : S \to \mathcal{R}$, die den **Ergebnissen** eines Zufallsexperiments **Werte** zuordnet (Bezeichnung durch $X, Y, Z, ...$)
- Zu jeder Zufallsvariable kann man eine **Verteilungsfunktion** $F : \mathcal{R} \to \mathcal{R}$ mit $F(x) = P(X \leq x)$ definieren, die die Wahrscheinlichkeitsverteilung der Zufallsvariable X beschreibt ($P(X \leq x) \hat{=} P(\{\omega : X(\omega) \leq x\})$)
- Ist F eine **Treppenfunktion**, so ist die Zufallsvariable X **diskret**; hat F in allen bis auf abzählbar vielen Punkten eine **stetige Ableitung** f, so wird X **stetig** genannt

Wahrscheinlichkeits- und Wahrscheinlichkeitsdichtefunktion

- X **diskrete Zufallsvariable**: Der Wertebereich von X ist abzählbar $\{x_1, x_2, \dots\}$; die Funktion $P(X = x_i), i = 1, 2, \dots$ heisst **Wahrscheinlichkeitsfunktion**
- X **stetige Zufallsvariable**: Ist die Ableitung f der Verteilungsfunktion F stetig und es gilt $P(\{\omega : a < X(\omega) < b\}) = \int_a^b f(t)dt$, dann heisst f **Wahrscheinlichkeitsdichtefunktion**

Bedingte Wahrscheinlichkeit

- Sind A und B Ereignisse und $P(A) > 0$, so ist die **bedingte Wahrscheinlichkeit** $P(B|A)$ für das Eintreten von B, falls A bereits stattgefunden hat, $P(B|A) = \frac{P(A \cap B)}{P(A)}$
- Somit $P(A \cap B) = P(A)P(B|A)$, woraus sich die **Multiplikations-Regel** ergibt als $P(A \cap B \cap \cdots \cap N) = P(A)P(B|A) \ldots P(N|A \cap B \cap \cdots \cap M)$
- Die **bedingte Wahrscheinlichkeitsverteilung** $P(\cdot|A)$ ist eine Wahrscheinlichkeitsverteilung auf S

Satz von Bayes

- Aus $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$ ergibt sich $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$
- Das **Gesetz der totalen Wahrscheinlichkeit** besagt, dass für disjunkte Ereignisse $A, B, \ldots, N$ mit $A \cup B \cup \cdots \cup N = S$ und ein beliebiges Ereignis X gilt $P(X) = P(A)P(X|A) + P(B)P(X|B) + \cdots + P(N)P(X|N)$
- Daraus folgt für disjunkte Ereignisse $A, B, \ldots, N$, ein beliebiges Ereignis X und $Y \in \{A, B, \ldots, N\}$ der **Satz von Bayes**:

$$P(Y|X) = \frac{P(Y)P(X|Y)}{P(A)P(X|A) + P(B)P(X|B) + \cdots + P(N)P(X|N)}$$

A priori und a posteriori Wahrscheinlichkeit

- Der Satz von Bayes lässt sich als "**Lernen**" auffassen:
- Aus der **a priori Wahrscheinlichkeit** einer Hypothese $P(A)$ wird bei beobachteten Daten B die neue **a posteriori Wahrscheinlichkeit** $P(A|B)$ berechnet

von Mises Ansatz

- Anhänger der **Frequenztheorie**: Definition von **Wahrscheinlichkeit** benötigt eine Definition von **Zufälligkeit** der Ergebnisse sich wiederholender Zufallsexperimente
- Ansatz: Die Ergebnisse eines Experiments sind nur dann zufällig, wenn sich das Verhältnis der Anzahl erfolgreicher Ausgänge zur Anzahl der Experimente bei *ausreichender* Wiederholung einer *festen Grenze* annähert: Der **Wahrscheinlichkeit** eines erfolgreichen Ausgangs
- **Problem**: *Wie oft* ist "*ausreichend*"?
- Weitere Bedingung für die Zufälligkeit einer Folge ist **Unabhängigkeit**: Der Ausgang des $n - ten$ Experiments darf **nicht aus den bisher beobachteten** $n - 1$ Ergebnissen ableitbar sein

Mises-Wald-Church-Zufälligkeit

- **von Mises**: Eine unendliche Folge $a_1, a_2, \ldots$ aus Nullen und Einsen ist zufällig im Sinne eines **Kollektivs**, wenn Sie folgende Bedingungen erfüllt:
  - Sei f_n die Anzahl der Einsen unter den ersten n Termen der Folge. Dann muss gelten $\lim_{n\to\infty} \frac{f_n}{n} = p$ für ein p mit $0 < p < 1$
  - Jede durch eine *zulässige* **Auswahlregel** ($\hat{=}$ Vorhersagestrategie) gewählte unendliche Teilfolge muss die erste Bedingung ebenfalls erfüllen
- Wenn als "zulässige" Auswahlregel eine **beliebige partielle Funktionen** gewählt werden darf, gibt es **keine** in diesem Sinne zufälligen Folgen
- **Wald**: Wenn man nur **abzählbar viele Auswahlfunktionen** zulässt, gibt es Folgen, die nach obiger Definition zufällig sind
- **Church**: Wähle die **berechenbaren Funktionen** als "zulässige" Funktionen

Martin-Löf-Zufälligkeit

- **Ville**: Es gibt Folgen, die **Mises-Wald-Church-zufällig** sind, aber **nicht** alle Gesetze der Zufälligkeit erfüllen
- **Martin-Löf**: Definiere diejenigen Strings als zufällig, die alle **effektiven Zufälligkeitstests** bestehen; es ist effektiv testbar, ob eine unendliche Folge Gesetze der Zufälligkeit verletzt indem **schrittweise längere Anfangssegmente** der Folge untersucht werden
- Naheliegend: **Effektive** Testbarkeit wird mit **berechenbaren Funktionen** assoziiert
- **Problem** der Definition auf Basis relativer Häufigkeiten für **unendliche** Folgen: In der **Praxis** hat man es mit **endlichen** Folgen zu tun

Zufälligkeit endlicher Strings

Vorschlag Kolmogorovs zur **Zufälligkeit endlicher Folgen**:

- **Verallgmeinerung der Auswahlregel** insofern, dass die Auswahl eines a_i auch von a_j mit $j > i$ abhängen kann
- Sei Φ eine Menge solcher verallgemeinerter Auswahlregeln; eine beliebige Folge a der Länge $n > m$ heisst (m, ϵ)-**zufällig in Bezug auf** Φ, wenn es ein p gibt, so dass die relative Häufigkeit der Einsen in jeder durch ein $\phi \in \Phi$ gewählten Teilfolge $a_{i_1} \dots a_{i_r}$ mit $r \geq m$ sich nicht mehr als ϵ von p unterscheidet
- Falls gilt $d(\Phi) \leq \frac{1}{2} e^{m\epsilon^2}(1 - \epsilon)$, dann existiert für jedes beliebige p und $n \geq m$ eine Folge a der Länge n, die (m, ϵ)-zufällig ist in Bezug auf Φ

Übergang zur Beschreibungskomplexität

- Frühere Versuche konzentrieren sich auf die **Unvorhersagbarkeit der Fortsetzung** einer Folge
- Solche Vorhersagen würden durch **Regelmäßigkeiten** im bereits **beobachteten Anfangssegment** möglich werden
- Veränderung des **Ausgangspunkts**: Benutzen **(nicht) vorhandener Regelmäßigkeiten** in endlichen Anfangssegmenten einer Folge als **Zufälligkeitskriterium**

Quellen


Ming Li and Paul Vitanyi.
An Introduction to Kolmogorov Complexity and its Applications.
3.Auflage, Springer-Verlag, 2008.


W. Feller.
An Introduction to Probability Theory and its Applications, Vol.1.
3.Auflage, Wiley, 1968.


C.P. Schnorr.
Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie.
Vol. 218 of *Lecture Notes in Mathematics*, Springer-Verlag , 1971.