2.8 Prüfzeichencodes

Sei in diesem Abschnitt das Alphabet A immer eine Gruppe mit Operation $*$. Die Menge aller Permutationen von A wird wie üblich mit S_A bezeichnet und ist mit der Verkettung $\circ$ ebenfalls eine Gruppe. Sind nun $\pi_1, \ldots, \pi_n$ aus S_A und $c \in A$, so ist die Menge C aller Wörter $w = w_1 \ldots w_n$ aus A^n, die die *Kontrollgleichung*

$$(K) \qquad \pi_1(w_1) * \pi_2(w_2) * \cdots * \pi_n(w_n) = c$$

erfüllen, ein n-Code, ein sogenannte *Prüfzeichencode* zu der durch die $\pi_1, \ldots, \pi_n$ und c gegebene Kontrollgleichung in der Gruppe A mit $*$. Da man die Kontrollgleichung nach $\pi_i(w_i)$...

$$\begin{aligned} \pi_i(w_i) \quad = \quad & \pi_{i-1}(w_{i-1})^{-1} * \pi_{i-2}(w_{i-2})^{-1} * \cdots * \pi_1(w_1)^{-1} \\ & * c \\ & * \pi_n(w_n)^{-1} * \pi_{n-1}(w_{n-1})^{-1} * \cdots * \pi_{i+1}(w_{i+1})^{-1} \end{aligned}$$

...und folglich nach w_i auflösen kann (beidseitiges Anwenden von π_i^{-1}), ist w_i durch die w_j mit $j \neq i$ bestimmt. Folglich können zwei Wörter aus C sich nicht an genau einer Stelle unterscheiden, also gilt $d(C) \geq 2$.

Die Stärke der Prüfzeichencodes besteht darin, daß man durch geeignete Wahl der π_i (und manchmal auch der Gruppe auf A) speziell strukturierten Fehlern Rechnung tragen kann. Soll ein Code zum Beispiel in der Lage sein, ein Vertauschen der Buchstaben $a \neq b$ an den aufeinanderfolgenden Stellen $i, i+1$ in einem gewissen Wort w zu erkennen, so heißt dies zunächst, daß das Wort

$$w = w_1 \ldots w_{i-1} a b w_{i+2} \ldots w_n$$

im Code enthalten ist, das Wort

$$v = w_1 \ldots w_{i-1} b a w_{i+2} \ldots w_n$$

dagegen nicht (oder umgekehrt). Das ist genau dann der Fall, wenn w die Kontrollgleichung erfüllt und v nicht (oder umgekehrt). Die Kontrollgleichung für w kann man nach $\pi_i(a) * \pi_{i+1}(b)$ auflösen, die für v entsprechend nach $\pi_i(b) * \pi_{i+1}(a)$, und so wird durch die Forderung $\pi_i(a) * \pi_{i+1}(b) \neq \pi_i(b) * \pi_{i+1}(a)$ sichergestellt, daß ein „einfacher Buchstabendreher" erkannt wird.

Hinter den Permutationen π_i steht oftmals schlichte Arithmetik. Betrachten wir zum Beispiel den Körper $A := \mathbb{Z}_{11}$, so ist für jedes $a \in A \setminus \{0\}$ die durch

$$\lambda_a(x) := a \cdot x$$

definierte Abbildung $\lambda_a : A \to A$ eine Bijektion, also aus S_A. Hiermit läßt sich der ISBN-10-Code beschreiben: Die ersten 9 Ziffern $z_1, \ldots, z_9$ werden als

Elemente von $\mathbb{Z}_{11}$ interpretiert; dabei tritt (die Restklasse zu) 10 natürlich nicht auf. Die letzte Stelle z_{10} wird daraus errechnet:

$$z_{10} = 1 \cdot z_1 + 2 \cdot z_2 + \cdots + 9 \cdot z_9,$$

und zwar in $\mathbb{Z}_{11}$, also modulo 11. Hierbei kann die Restklasse zu 10 auftreten, sie wird dann als X (römisch 10) geschrieben. Nun ist $-1 = 10$ in $\mathbb{Z}_{11}$, so daß sich die obige Gleichung umstellen läßt auf

$$1 \cdot z_1 + 2 \cdot z_2 + \cdots + 9 \cdot z_9 + \underbrace{10 \cdot z_{10}}_{=-z_{10}} = 0.$$

Ersetzen wir die Multiplikationen durch die von ihnen bewirkten Permutationen, so erweist sich der ISBN-10-Code als ein Code über der Gruppe $A = \mathbb{Z}_{11}$ mit $+$ und der Kontrollgleichung

$$\lambda_1(z_1) + \lambda_2(z_2) + \cdots + \lambda_{10}(z_1 0) = 0.$$

Grundsätzlich könnte dabei die „Ziffer X" auch an einer der ersten 9 Stellen auftreten, wird aber dort nicht verwendet bzw. vergeben.

Im allgemeinen ist die Abbildung $x \mapsto a \cdot x$ vom Restklassenring $\mathbb{Z}_n$ in sich selbst genau dann eine Bijektion, wenn a ein multiplikativ Inverses besitzt (a heißt dann *Einheit*). Zum Beispiel sind die Einheiten von $\mathbb{Z}_{10}$ die Elemente $1, 3, 7, 9$, und so kann man den ISBN-13-Code ebenfalls als Prüfzeichencode beschreiben: Die 13 Ziffern eines Codeworts werden alle als Elemente von $\mathbb{Z}_{10}$ interpretiert, und die Prüfziffer z_{13} an letzter Stelle ergibt sich aus den vorangegangenen zwölf Ziffern $z_1, \ldots, z_{12}$ durch

$$z_{13} = 10 - (z_1 + z_3 + z_5 + z_7 + z_9 + z_{11}) - 3 \cdot (z_2 + z_4 + z_6 + z_8 + z_{10} + z_{12}),$$

wobei in $\mathbb{Z}_{10}$ gerechnet wird. Zur Übung möge man daraus die Kontrollgleichung (über der Gruppe $\mathbb{Z}_{10}$ mit $+$) herleiten.

2.9 Überdeckungen

Die Herstellung von n-Codes mit möglichst großer Informationsrate bei vorgegebenem Minimalabstand ist ein klassisches Packungsproblem: Man packe möglichst viele disjunkte Kugeln vom Radius d in den n-dimensionalen Hamming-Würfel A^n. Stattdessen kann man von der anderen Seite herangehen und fragen, mit wie wenigen Kugeln vom Radius d der Raum überdeckt werden kann. Während die Hamming-Schranke eine obere Schranke für das Packungsproblem ist, ist sie eine untere Schranke für das Überdeckungsproblem: Soll nämlich A^n durch Kugeln $B_d(x)$, $x \in C$, mit $C \subseteq A^n$ überdeckt werden, so folgt

$$|A^n| = |\bigcup_{x \in C} B_d(x)| \leq \sum_{x \in C} |B_d(x)| = |C| \cdot B_d(0),$$

also $|C| \geq |A|^n/|B_d(0)|$ (rechts steht die Hamming-Schranke). Die Gleichheit ist genau dann gegeben, wenn die Kugeln paarweise disjunkt sind, also C perfekt d-fehlerkorrigierend ist. Das Minimum

$$K_A(n,d) := \min\{|C| : C \subseteq A^n, \bigcup_{x \in C} B_d(x) = A^n\}$$

heißt *Überdeckungszahl* von A^n für Kugeln vom Radius d; es gilt $K_A(n,d) \geq |A|^n/|B_d(0)|$ mit Gleichheit genau dann, wenn es einen perfekt d-fehlerkorrigierenden n-Code über A gibt. Bezeichnet man mit

$$P_A(n,d) := \max\{|C| : C \subseteq A^n, B_d(x) \cap B_d(y) = \emptyset \text{ für alle } x \neq y \text{ aus } C\}$$

die *Packungszahl* von A^n für Kugeln vom Radius d, so gilt ganz entsprechend $P_A(n,d) \leq |A|^n/|B_d(0)|$ mit Gleichheit genau dann, wenn es einen perfekt d-fehlerkorrigierenden n-Code über A gibt. Diejenigen d-fehlerkorrigierenden n-Codes C über A mit $|C| = P_A(n,d)$ heißen gelegentlich *optimal*.

Bei vielen Sportwetten geht es darum, für eine bestimmte Zahl n von Spielen eines gewissen Zeitraums in der Zukunft auf jeweils ein Ereignis $x \in A$ zu setzen; vielfach ist $|A| = 3$ (Sieg von Heim- bzw. Gastmannschaft, unentschieden) und $n = 11$ oder 13 (Elfer- bzw. Dreizehner-Wette). Die zulässigen Tips sind dann Wörter aus A^n. Man gewinnt einen k-ten Preis, wenn der eigene Tip w an genau $k-1$ Stellen vom tatsächlichen Spielergebnis s abweicht, also $d(w,s) = k-1$ gilt. Spielt man alle Tips einer Überdeckung C von A^n mit Kugeln vom Radius d, kann man folglich sicher sein, daß einer der Tips einen wenigstens $(d+1)$-ten Preis[1] gewinnt. $K_A(n,d)$ ist folglich die Anzahl Tips, die man wenigstens abgeben muß, um einen wenigstens $(d+1)$-ten Preis zu gewinnen. Packungs- und Überdeckungszahlen, die von der Hamming-Schranke abweichen, sind nur für wenige Parametersätze bekannt.

[1] Gemeint ist natürlich ein $(d+1)$-ter oder *besserer* Preis, also ein k-ter Preis für ein $k \leq d+1$.