Eine Unterrichtssequenz mit Materialien und Lösungen zur

Chiffrierung mit Tabellen basierend auf Buchstabenhäufigkeiten

Martin Huber, martin.huber3@unifr.ch

28. Januar 2022

erstellt im Rahmen von Fachdidaktik I des Lehrgangs GymInf
betreut durch Prof. Dr. Juraj Hromkovič und Regula Lacher

Chiffrierung mit Tabellen weiterentwickeln

Aufgabe 1 Der folgende Geheimtext basiert auf der Erzeugung von neuen Symbolen für alle 26 Buchstaben A-Z mit einer rechteckigen Tabelle. Dechiffrieren Sie den folgenden Geheimtext:

```
B2 B3 B1 C5 A6 A1 C6 B8 B4 C5 B9 B0 A0 C9 B8 B7 B8 A6 A0 A3 A9 C7 B0 A4
A9 B4 A1 B3 B1 B9 C2 C3 B7 B0 C7 C0 C9 A7 B2 B3 B1 A1 A6 B2 B2 B4 A0
```

Dabei sind als Hilfe die unten stehenden Informationen bekannt:

- Das Symbol B1 codiert H, B8 codiert G, C7 codiert A, A9 codiert R, C5 codiert W, C6 codiert Z.
- Im Klartext kommen folgende Buchstaben häufiger als einmal vor:

Symbol	E	N	I	S	C	G	H	A	R	T	W
abs. Häufigkeit	8	6	5	4	3	3	3	2	2	2	2

- Im Chiffrentext kommen folgende Symbole häufiger als einmal vor:

Symbol	B2	A0	A1	A6	B0	B1	B3	B4	B8	A9	B7	B9	C5	C7	C9
abs. Häufigkeit	4	3	3	3	3	3	3	3	3	2	2	2	2	2	2

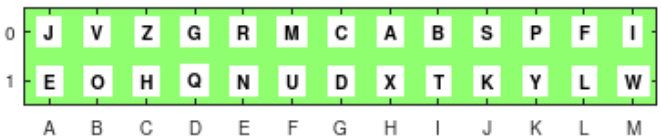
- Der Inhalt des Klartextes erinnert an einen sportlichen Erfolg der Schweiz vom 28.6.2021.

In dieser Unterrichtssequenz lernen Sie einen Ansatz kennen, wie aufgrund von Erkenntnissen über frührere Verschlüsselungsmethoden eine Weiterenwicklung umgesetzt und analysiert wird.

Vorwissen: Monoalphabetische Codierung mit Tabellen & Buchstabenhäufigkeit

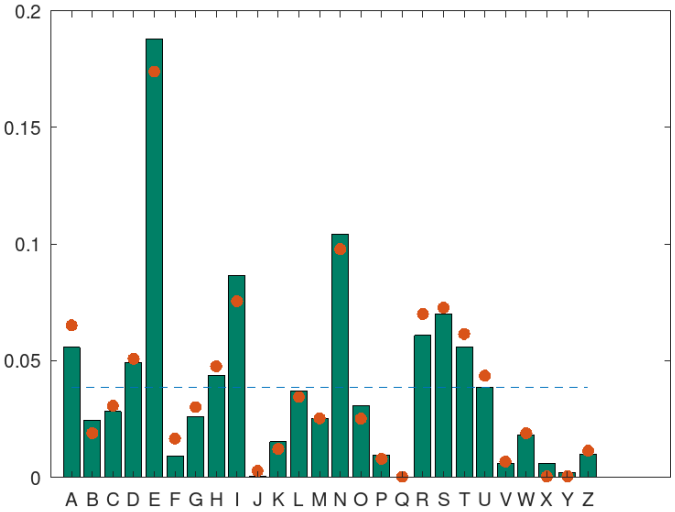
- **Monoalphabetische Verschlüsselung mit Chiffrierungstabellen:** Jeder Buchstabe des Alphabetes wird in eine rechteckige Tabelle eingefügt, welche eine ausreichende Anzahl Zeilen und Spalten hat, damit das ganze Klartextalphabet Platz hat. Die einzelnen Zeilen und Spalten werden mit eigenen Symbolen versehen. Damit kann jeder Buchstabe des Klartextalphabetes durch eine Folge von zwei Symbolen kodiert werden:
  - Erstes Symbol: Symbol der Zeile, in der sich der Klartextbuchstabe befindet.
  - Zweites Symbol: Symbol der Spalte, in der sich der Klartextbuchstabe befindet.

  Beispiel: Eine Chiffrierungstabelle der Grösse 2×13 für Buchstaben A-Z. Der Buchstabe Q wird als 1D codiert.



Da jeder Buchstabe des Klartextalphabets durch genau ein Symbol des Geheimtextalphabets codiert wird, handelt es sich um eine monoalphabetische Verschlüsselung. Die Reihenfolge, wie die Klartextbuchstaben in der Tabelle verteilt werden, ist dabei frei wählbar. Wenn sie alphabetisch eingefügt werden, findet man die grundlegende Idee der Geheimschrift von Polybios wieder.

- **Relative Buchstabenhäufigkeit für stochastische Kryptoanalyse:** Jede Sprache weist eine charakteristische Häufigkeit der einzelnen Buchstaben auf, sodass in längeren Texten die Buchstabenhäufigkeit nahe an diesen bekannten Werten liegt. Da bei einer monoalphabetischen Verschlüsselung die genau gleichen relativen Häufigkeiten für die Symbole des Geheimtextalphabetes entstehen, kann aufgrund dieser Häufigkeiten systematisch die Entschlüsselung ausprobiert werden (häufigstes Symbol des Geheimtextes durch den häufigsten Buchstaben des Klartextalphabets ersetzen etc.).
  Beispiel: In deutschsprachigen Texten ist der Buchstabe `E` mit ca. 17.4% gefolgt von `N` (ca. 9.8%), `I` (ca. 7.5%) und `S` (ca. 7.3%) der häufigste Buchstabe. Diese charakteristische Häufigkeit ist in untenstehender Grafik mit den roten Punkten bezeichnet. Die gestrichelte Linie markiert die Gleichverteilung.



Für einen deutschsprachigen Beispieltext von rund 25'000 Buchstaben erhält man dann relative Buchstabenhäufigkeiten (grüne Balken in obiger Grafik), welche mit der erwarteten Häufigkeitsverteilung im Grossen und Ganzen übereinstimmt

Um diese Schwäche von monoalphabetischen Verschlüsselungsverfahren zu umgehen, wird folgender Ansatz vorgeschlagen, ausprobiert und analysiert.

Grundidee: Unterschiedliche Codierungen für häufige Buchstaben des Klartextalphabets

Ziel ist es, dass im Geheimtext die Häufigkeit der Symbole möglichst ausgeglichen ist, d.h., möglichst eine Gleichverteilung vorherrscht. Dies wird dadurch erreicht, dass

- ein einzelner Buchstabe, der im Klartext statistisch gesehen häufig vorkommt, mithilfe verschiedener Symbole im Geheimtext codiert wird;
  Beispiel: Der Buchstabe `E` kommt in deutschsprachigen Texten mit 17.4% statistisch gesehen am häufigsten vor. Um im Gegensatz zu einer monoalphabetischen Verschlüsselung zu

vermeiden, dass ein Symbol des Geheimtextalphabets ebenfalls mit 17.4% Häufigkeit vorkommt, codiert man den Buchstaben `E` mithilfe mehrerer Symbole. Wenn man z.B. für `E` fünf unterschiedliche Symbole benutzt und diese etwa gleich häufig zur Codierung nutzt, dann haben diese Symbole alle ca. 17.4% : 5 = ca.3.5% relative Häufigkeit.

- für diese zusätzlichen Symbole im Geheimtextalphabet die Grösse der Chiffrierungstabelle erhöht wird.
  Beispiel: 26 Geheimtextsymbole reichen aus für eine monoalphabetische Chiffrierung der Buchstaben `A-Z`. Sobald man ein grösseres Geheimtextalphabet verwendet, werden die Tabellen auch grösser, z.B. 3×10-Tabelle für 30 Symbole, 5×7-Tabelle für 35 Symbole.

Diese Idee zur Verschleierung der Buchstabenhäufigkeit wurde schon im 17. Jahrhundert entwickelt. Die Verfahren werden *homophone* Verschlüsselungen genannt.

Ausgedrückt mithilfe der Friedman'schen Charakteristik heisst dies:

- Für den Klartext T in deutscher Sprache wird ein typischer Wert ca. $0.0385\ldots$ erwartet, d.h.

$$\mathrm{FC}(T) = \sum_{\Delta \in \text{Klartextalphabet}} \left(h_\Delta(T) - \frac{1}{26} \right)^2 \approx 0.0385\ldots,$$

  wobei $h_\Delta(T)$ die relative Häufigkeit des Symbols Δ aus dem Klartextalphabet mit hier 26 Symbolen bezeichnet.
- Für den Geheimtext G wird ein Wert möglichst nahe bei 0 angestrebt, d.h., man wählt ein Geheimtextalphabet, sodass

$$\mathrm{FC}(G) = \sum_{\Delta \in \text{Geheimtextalphabet}} \left(h_\Delta(G) - \frac{1}{|\text{Geheimtextalphabet}|} \right)^2$$

  möglichst klein ist, wobei |Geheimtextalphabet| die Anzahl Symbole im Geheimtextalphabet bezeichnet.

Lösung zu Aufgabe 1

Als Erstes kann aufgrund der zweistelligen Symbole des Geheimtextes (1. Stelle: `A`, `B`, `C`; 2. Stelle: Ziffern `0` bis `9`) und der Information, dass mit einer Chiffrierungstabelle gearbeitet wurde, geschlossen werden, dass der Verschlüsselung eine 3×10-Tabelle zugrunde liegt, wobei einige Einträge als Starthilfe angegeben wurden. Kombiniert mit der Beobachtung, dass es im Klar- und Geheimtext genau ein Symbol gibt, welches viermal vorkommt, kann man zusätzlich vermuten, dass `B2` den Buchstaben `S` codiert.

	0	1	2	3	4	5	6	7	8	9
A										R
B		H	S						G	
C						W	Z	A		

Somit lautet die bisherige Entschlüsselung

```
S  B3 H  W  A6 A1 Z  G  B4 W  B9 B0 A0 C9 G  B7 G  A6 A0 A3 R  A  B0 A4
R  B4 A1 B3 H  B9 C2 C3 B7 B0 A  C0 C9 A7 S  B3 H  A1 A6 S  S  B4 A0
```

oder nur mit den Klartextsymbolen

```
S.HW..ZG.W....G.G...RA..R...H.....A...S.H..SS..
```

Streichen wir zudem alle benutzten Symbole aus den beiden Häufigkeitstabellen, so erhalten wir für den Klartext noch:

Symbol	E	N	I	C	T
abs. Häufigkeit	8	6	5	3	2

Wir erhalten im Chiffrentext:

Symbol	A0	A1	A6	B0	B3	B4	B7	B9	C9
abs. Häufigkeit	3	3	3	3	3	3	2	2	2

Somit befinden sich in beiden Tabellen noch insgesamt 24 Symbole und damit müssen `E`, `N`, `I`, `C`, `T` durch die neun Symbole aus der Tabelle des Geheimtextalphabetes codiert werden. Demzufolge gibt es mehrere Symbole, welche `E`, `N`, `I` codieren. Dabei gelten folgende Aufteilungen:

- `E`: Die Aufteilung lautet $8 = 3 + 3 + 2$.
- `N`: Mögliche Aufteilungen sind $6 = 2 + 2 + 2 = 3 + 3$. Da aber zwei Symbole mit Häufigkeit 2 schon durch `E` und `T` belegt sind, gibt es nur noch eine mögliche Aufteilung $6 = 3 + 3$.
- `I`: Die Aufteilung lautet $5 = 3 + 2$.

Weitere Informationen können nicht definitiv zugeordnet werden, weshalb jetzt mit Spekulieren und dem Klartextinhalt gearbeitet werden muss. Einige Anhaltspunkte sind:

- `B3` steht zweimal zwischen `S` und `H` sowie einmal direkt vor `H`. Somit ist ein `C` wahrscheinlich.

  ```
  SCHW..ZG.W....G.G...RA..R..CH.....A...SCH..SS..
  ```

- Das Anfangswort scheint `SCHW..Z` zu sein, da dann ein `G` kommt. In den zwei Lücken sind es Symbole mit Häufigkeit 3, weshalb nur noch `E`, `N` und `I` in Frage kommen. Die sinnvolle Variante ist `EI`. Somit ist `A6` ein `E` und `A1` ein `I`.

  ```
  SCHWEIZG.W....G.GE..RA..R.ICH.....A...SCHIESS..
  ```

- `B4` steht zwischen zwei `G` und muss somit ein Vokal sein, d.h., `E` oder `I`, wobei sich Ersteres als passender herausstellt.

  ```
  SCHWEIZGEW....GEGE..RA..R.ICH.....A...SCHIESSE.
  ```

- Das Schlusssymbol `A0` scheint somit ein `N` zu sein:

  ```
  SCHWEIZGEW..N.GEGEN.RA..R.ICH.....A...SCHIESSEN
  ```

- Damit sind die ersten drei Wörter zu erraten: `SCHWEIZ GEWINNT GEGEN`. Somit bedeutet `B9` ein `I`, `B0` ein `N` und `C9` ein `T`.

  ```
  SCHWEIZ GEWINNT GEGEN .RAN.R.ICHI...NA.T.SCHIESSEN
  ```

Es ist an der Zeit, um nochmals Buchhaltung zu machen, um zu sehen, welche Informationen uns noch bekannt sind. Dazu aktualisieren wir wieder die beiden Häufigkeitstabellen.

Klartext:

Symbol	E
abs. Häufigkeit	2

Chiffrentext:

Symbol	B7
abs. Häufigkeit	2

Somit codiert B7 den Buchstaben E, womit wir alle Informationen (bis auf das Datum 28.6.2021) verwendet haben. Dies führt zu:

```
SCHWEIZ GEWINNT GEGEN .RAN.REICHI..ENA.T.SCHIESSEN
```

Mit Probieren oder einer Internetrecherche zum Datum 28.6.2021 erhält man den Klartext:

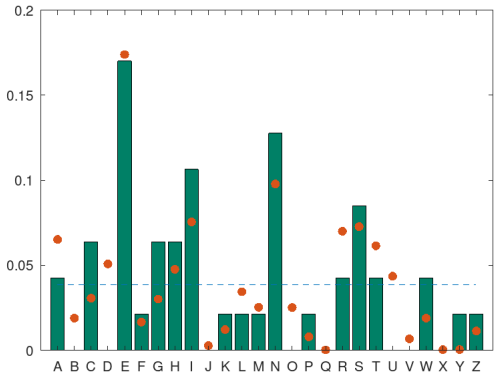
```
SCHWEIZ GEWINNT GEGEN FRANKREICH IM PENALTYSCHIESSEN
```

Fazit: Der Klartext wurde mit der Chiffrierungstabelle

	0	1	2	3	4	5	6	7	8	9
A	N	I	U	F	K	Q	E	Y	D	R
B	N	H	S	C	E	V	X	E	G	I
C	L	J	M	P	O	W	Z	A	B	T

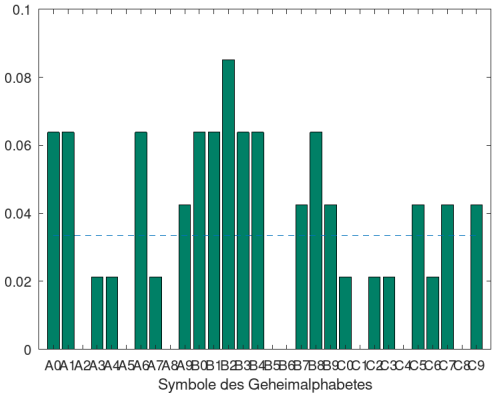
verschlüsselt, weshalb die typisch hohen Buchstabenhäufigkeiten von E, N und I verschleiert wurden. Dies zeigt sich auch in der Übersicht zu den relativen Symbolhäufigkeiten in Klartext und Geheimtext.

Relative Buchstabenhäufigkeit im Klartext T



Die roten Punkte zeigen den statistisch erwarteten Wert für deutschsprachige Texte als Referenz, die gestrichelte Linie die Gleichverteilung. Die Friedman'sche Charakteristik beträgt $\mathrm{FC}(T) = 0.0480\ldots$.

Relative Symbolhäufigkeit im Geheimtext G



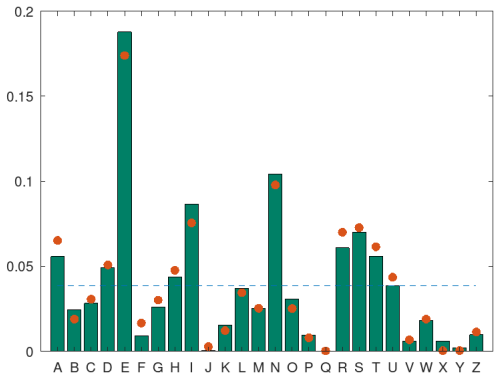
Die Friedman'sche Charakteristik beträgt $\mathrm{FC}(G) = 0.0205\ldots$, welche gegenüber dem Klartext beträchtlich gesunken ist. Die gestrichelte Linie zeigt wiederum die Gleichverteilung.

Dies verhindert den typischen Angriffspunkt bei monoalphabetischen Verschlüsselungen, den Buchstaben E (und dann evtl. auch N und I) relativ einfach erraten zu können. Dies manifestiert sich auch beim Lösen der Aufgabe, wo mithilfe der angegebenen Symbolhäufigkeiten ersichtlich wird, dass über Aufteilungen der Häufigkeiten im Geheimtextalphabet nachgedacht werden muss.

Eine 10×10-Chiffrierungstabelle für deutschsprachige Texte

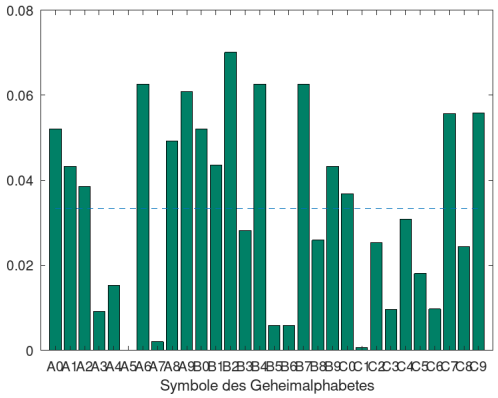
Für die folgenden Betrachtungen nehmen wir einen deutschsprachigen Beispieltext mit rund 25'000 Buchstaben. Als Erstes wird dieser mittels der 3×10-Tabelle aus Aufgabe 1 codiert.

Relative Buchstabenhäufigkeit im Klartext T



Die roten Punkte zeigen den statistisch erwarteten Wert für deutschsprachige Texte als Referenz, die gestrichelte Linie die Gleichverteilung. Die Friedman'sche Charakteristik beträgt $\mathrm{FC}(T) = 0.0417\ldots$.

Geheimtext G **mit** 3×10 **Tabelle**



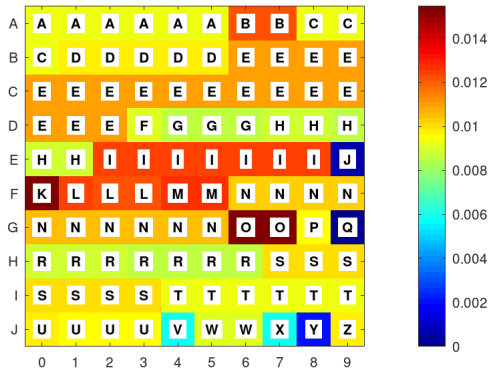
Die Friedman'sche Charakteristik beträgt $\mathrm{FC}(G) = 0.0141\ldots$, welche gegenüber dem Klartext beträchtlich gesunken ist. Die gestrichelte Linie zeigt wiederum die Gleichverteilung.

Um noch näher an eine Gleichverteilung bei den Symbolhäufigkeiten zu gelangen, können die Ungleichheiten in der Häufigkeit der Geheimtextsymbole durch grössere Chiffrierungstabellen noch weiter reduziert werden. Um direkt die Prozentangaben der Buchstabenhäufigkeit zu benutzen, bietet es sich an 10×10-Tabellen betrachtet werden. Die Häufigkeit, wievielmal ein einzelner Buchstabe codiert wird, erhält man dann aus den Prozentzahlen. Dabei muss sichergestellt werden, dass alle Buchstaben mindestens einmal vorkommen. Die restlichen Plätze werden dann anhand der Prozentzahlen verteilt. Dabei sind verschiedene Vorgehensweisen möglich, die Zahlen zu runden etc. Ein Beispiel findet sich in der nächsten Tabelle.

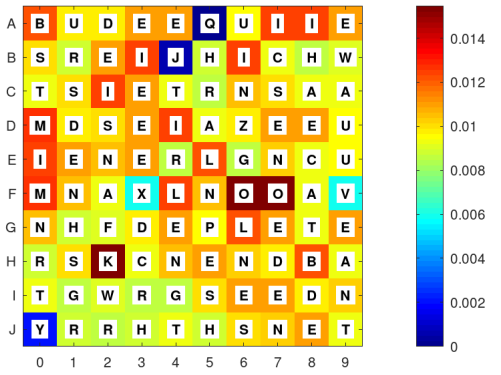
Buchstabe	A	B	C	D	E	F	G	H	I	J	K	L	M
rel. Häufigkeit	6.51%	1.89%	3.06%	5.08%	17.4%	1.66%	3.01%	4.76%	7.55%	0.27%	1.21%	3.44%	2.53%
Anzahl in Tabelle	6	2	3	5	17	1	3	5	7	1	1	3	2
Buchstabe	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
rel. Häufigkeit	9.78%	2.51%	0.79%	0.02%	7%	7.27%	6.15%	4.35%	0.67%	1.89%	0.03%	0.04%	1.13%
Anzahl in Tabelle	10	2	1	1	7	7	6	4	1	2	1	1	1

Die Buchstaben kommen nun alle in bestimmter Anzahl in der Tabelle vor. Für die Verteilung auf die 10×10-Tabelle ist die Reihenfolge der Buchstaben beliebig wählbar. Wir betrachten im Folgenden zwei Varianten: einmal Zeile für Zeile alphabetisch eingefügt und einmal zufällig eingefügt. Nun wird der Beispieltext mit ca. 25'000 Buchstaben mit beiden Tabellen codiert. Die Farben zeigen die relative Häufigkeit der Symbole für den deutschsprachigen Beispieltext mit den rund 25'000 Buchstaben.

10×10-Chiffrierungstabelle alphabetisch ausgefüllt



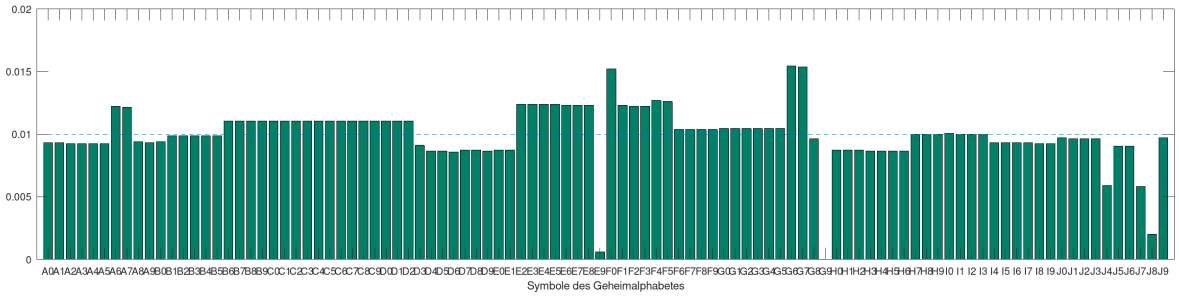
10×10-Chiffrierungstabelle zufällig ausgefüllt



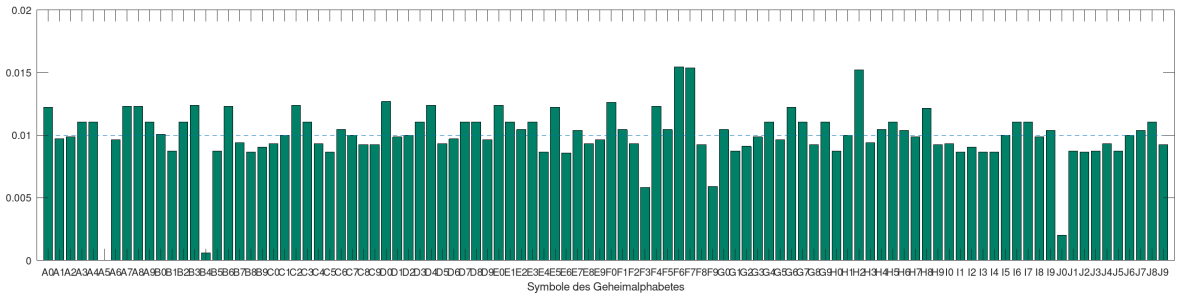
Dabei ergibt sich in beiden Fällen eine Symbolhäufigkeitsverteilung, welche sehr nahe an der Gleichverteilung ist. Die Friedman'sche Charakteristik beträgt nur noch $\mathrm{FC}(G) = 0.0005\ldots$, d.h., sie ist sehr nahe am angestrebten Wert von 0.

Graphisch dargestellt ergeben sich die folgenden Symbolhäufigkeiten für die beiden Tabellen:

10×10-Chiffrierungstabelle alphabetisch ausgefüllt



10×10-Chiffrierungstabelle nach dem Zufallsprinzip ausgefüllt



In der Aufgabe 8 wird der Zusammenhang der beiden untersucht und dann daraus eine mögliche Variante entwickelt, um auch diese nahezu gleichverteilte Chiffrierung zu knacken.

Aufgabe 2 Entschlüsseln Sie den folgenden Text mit der 10×10 Chiffrierungstabelle "zufällig gefüllt" aus dem vorhergehenden Abschnitt.

```
G1 F2 B0 C0 D1 A1 H2 E1 E0 G0 G5 G2 B2 H0 A2 C1 F6 F1 C2 D0 F0 G3 A3 E2 C3 H1 D3 F4
```

Aufgabe 3 Verschlüsseln Sie den folgenden Text

`BAECKEREIEIERZEUGNIS IM INNENHOF`

mit der 3×10-Chiffrierungstabelle aus Aufgabe 1.

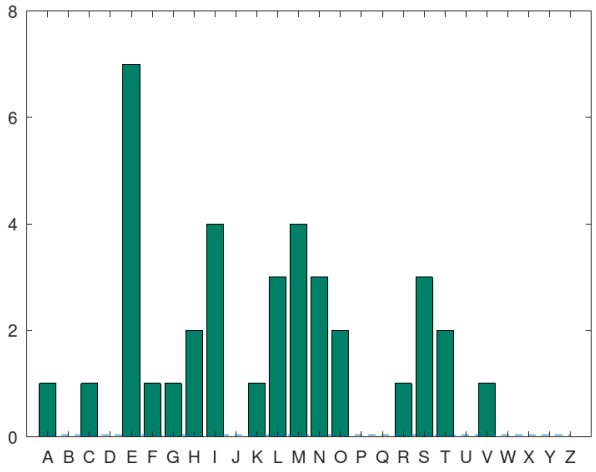
	0	1	2	3	4	5	6	7	8	9
A	N	I	U	F	K	Q	E	Y	D	R
B	N	H	S	C	E	V	X	E	G	I
C	L	J	M	P	O	W	Z	A	B	T

Aufgabe 4 Welche praktischen Probleme haben sich bei der Ver- und Entschlüsselung in den Aufgaben 2 und 3 ergeben? Wie haben Sie diese gelöst?

Aufgabe 5 Gegeben ist der folgende Klartext:

`ES IST NOCH KEIN MEISTER VOM HIMMEL GEFALLEN`

Er weist folgende absoluten Buchstabenhäufigkeiten auf:



Erstellen Sie ein Geheimtextalphabet speziell angepasst an den oben angegebenen Klartext. Kreieren Sie dazu zwei Varianten einer Chiffrierungstabelle:

a) zuerst einmal mit genau 24 Symbolen
b) und einmal mit einer von Ihnen gewählten Anzahl Symbolen

sodass die Friedman'sche Charakteristik $\mathrm{FC}(G)$ des Geheimtextes für die Meldung "Es ist noch kein Meister vom Himmel gefallen" möglichst klein ist. Vergleichen Sie Ihre erzielten Werte innerhalb der Klasse.

Aufgabe 6 Erachten Sie die Idee aus Aufgabe 5 als sinnvoll, dass Geheimtextalphabete auf Klartexte zugeschnitten werden? Welche Vor- und Nachteile finden Sie?

Aufgabe 7 Die DNA kann als Nachricht in einer Sprache mit dem Klartextalphabet $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$ interpretiert werden, wobei die Buchstaben für Adenin, Cytosin, Guanin und Thymin stehen. So gesehen ist eine Sequenz eines DNA-Stranges wie zum Beispiel

```
...ATG CGC AAT GCG ATA TAC...
```

als Nachricht in diesem speziellen Alphabet mit vier Symbolen zu verstehen. Die relative Häufigkeit der vier Buchstaben in der DNA kennt folgende Gesetzmässigkeiten:

- `A` und `T` sind gleich häufig.
- `G` und `C` sind gleich häufig.
- Die konkreten Prozentsätze unterscheiden sich von Spezies zu Spezies: Die Häufigkeit von `A` beim Menschen beträgt 30%, bei der Grünalge 20% und beim Weizen 27%.

Erstellen Sie für die Spezies Mensch eine Chiffrierungstabelle (Grösse frei wählbar), sodass die Symbole im Geheimtext möglichst gleichverteilt sind.

Aufgabe 8 In dieser Aufgabe vergleichen wir die Resultate der beiden 10×10-Chiffrierungstabellen (einmal alphabetisch gefüllt und einmal dem Zufallsprinzip folgend ausgefüllt) und versuchen ein Vorgehen zu entwickeln, wie auch diese Verschlüsselung geknackt werden könnte.

a) Welche Vor- und Nachteile sehen Sie, wenn die Tabelle alphabetisch ausgefüllt wird?
b) Wie hängen die graphischen Darstellungen der Symbolhäufigkeiten bezüglich der beiden unterschiedlichen Tabellen zusammen?
c) Mithilfe der 10×10-Tabellen wurde erreicht, dass die Symbolhäufigkeit sehr nahe an einer Gleichverteilung ist. Somit kann man nicht mehr erraten, dass das häufigste Symbol den Buchstaben `E` codieren "sollte". Wie könnte man trotzdem herausfinden, welche Symbole zum Buchstaben `E` gehören?
d) Welcher der beiden Plots hilft Ihnen dabei besser? Wie könnte man den anderen ebenfalls dafür nutzen?
e) Skizzieren Sie ein Vorgehen, wie Sie einen Geheimtext chiffriert mit der in diesem Kapitel untersuchten Methode "knacken" könnten.

Kommentar für Lehrpersonen: Einbettung und Ziele

Die Unterrichtssequenz leistet einen Beitrag zum fundamentalen Themenkomplex der Informatik "Informations- und Datendarstellung", zu welchem die Datensicherheit und somit Verschlüsselungsmethoden gehören. Die Voraussetzungen für die Unterrichtssequenz sind, dass die Lernenden. . .

- monoalphabetische Verschlüsselungsmethoden mit Chiffrierungstabellen kennen;
- monoalphabetisch verschlüsselte Geheimtexte mithilfe der Analyse von Buchstabenhäufigkeiten geknackt haben.

Die angestrebten Ziele der Unterrichtssequenz sind:

- Eine Methode zur Verschleierung der charakteristischen relativen Symbolhäufigkeit kennenlernen.
- Die Qualität einer Verschlüssung mittels Buchstabenhäufigkeiten und der Friedman'schen Charakteristik bewerten.
- Die Schwächen von bekannten Verschlüsselungsmethoden gründlicher verstehen, indem man versucht, diese zu beseitigen.
- Die Lernenden erfinderisch aktiv werden zu lassen, indem sie selbst neue Chiffrierungstabellen für Verschlüsselungen entwickeln.
- Eigene Alphabete (ein Kerngebiet der Informatik) mit gewünschten Eigenschaften selbst entwickeln.
- Den Entwicklungsprozess von neuen Verfahren in den Grundzügen miterleben, indem sie mittels Vorwissen und Kreativität neue Ansätze vorschlagen, testen, analysieren und wiederum weiterentwickeln.

Lösungen zu den Aufgaben

Lösung zur Aufgabe 2

Der Klartext lautet: “Hast du kein Pferd, so nimm den Esel.”

Lösung zur Aufgabe 3

Dazu sind mehrere Lösungen möglich, je nachdem wie man damit umgeht, die Symbole für die mehrfach vorkommenden Buchstaben `E`, `N` und `I` auszuwählen. Denn wenn der Buchstabe `E` zu verschlüsseln ist, so stehen drei verschiedene Symbole zur Auswahl. Da das Ziel ist, die Buchstabenhäufigkeit zu verschleiern, sollten möglichst alle zur Verfügung stehenden Symbole gleich häufig benutzt werden. Eine Beispiellösung wäre:

```
C8 C7 B4 B3 A4 A6 A9 B7 A1 B4 B9 A6 A9 C6 B7
A2 B8 A0 A1 B2 B9 C2 A1 B0 A0 B4 B0 B1 C4 A3
```

Die obige Lösung geht eine nach der anderen Variante durch, z.B. für das `E` gibt es repetitiv die Abfolge `B4` → `A6` → `B7` → `B4` . . . .

Lösung zur Aufgabe 4

Die Entschlüsselung stellt kein Problem dar, die Verschlüsselung grundsätzlich auch nicht. Jedoch ist bei der Verschlüsselung zentral, die unterschiedlichen Symbole für den gleichen Buchstaben so zu wählen, dass sie annäherend gleichverteilt sind. Folgende Herangehensweisen sind denkbar:

- Die Liste der möglichen Symbole für einen bestimmten Buchstaben wird repetitiv sequentiell durchgegangen. Jedoch werden wir sehen, dass eine absolute Gleichverteilung der Symbole für einen einzigen Buchstaben auch Nachteile hat (s. Aufgabe 8).
- Mittels einer gleichverteilten Zufallszahl wird jeweils eine Position der Liste ausgewählt.

Lösung zur Aufgabe 5

Zu Teil a: Zuerst kann die Grösse der Tabelle gewählt werden. Dazu sind verschiedene Varianten möglich: 1×24, 2×12, 3×8, 4×6. Jedoch braucht man bei Dimensionen grösser als 10 auch entsprechende “neue” Symbole, weswegen hier ein Beispiel für eine 4×6-Matrix gegeben wird. Insgesamt kommen im Text 16 verschiedene Buchstaben vor. Somit können 8 Plätze auf die häufig vorkommenden Buchstaben verteilt werden. Dazu gibt es verschiedene denkbare Varianten (vgl. das Problem ist vergleichbar mit dem Verteilen von Sitzen bei Proporzwahlen, wo es ebenfalls verschiedene Vorgehen gibt, um Restmandate zu verteilen).

Buchstaben	Häufigkeit im Text	Minimum	verbleibend	Prozentsatz verbleibend	beansprucht	vergeben	**Total Plätze**
A	1	1	0				**1**
C	1	1	0				**1**
E	7	1	6	0.29	2.29	2	**3**
F	1	1	0				**1**
G	1	1	0				**1**
H	2	1	1	0.05	0.38	1	**2**
I	4	1	3	0.14	1.14	1	**2**
K	1	1	0				**1**
L	3	1	2				**1**
M	4	1	3	0.14	1.14	1	**2**
N	3	1	2	0.10	0.76	1	**2**
O	2	1	1	0.05	0.38	1	**2**
R	1	1	0				**1**
S	3	1	2	0.10	0.76	1	**2**
T	2	1	1	0.05	0.38		**1**
V	1	1	0				**1**
Total	37		21				**24**

Plätze verfügbar	Davon Gebraucht	Verbleibend			Davon Gebraucht	
24	16	8			8	

Hier wurde der Ansatz verfolgt, dass man berechnet, wie gross das prozentuale Anrecht an den 8 verbleibenden Plätzen für jeden Buchstaben noch ist und dies in eine Platzanzahl umrechnet. Diese Werte werden dann gerundet. Die gelb markierten Zellen hätten gemäss Rundung keinen weiteren Platz mehr, jedoch würden dann zwei Plätze leer bleiben. Diese werden dann je einem dieser Buchstaben mit den gelb markierten Zeilen zugeteilt. Für den Buchstaben `E` lautet die Rechnung wie folgt: Absolute Häufigkeit im Text: 7; ein Platz grundsätzlich zugesprochen; sechs Vorkommen sind damit noch nicht abgedeckt. Diese machen 29% der restlichen 8 freien Plätze aus, das wären 2.29 Plätze. Dies wird abgerundet und `E` bekommt damit noch zwei zusätzliche Plätze, weshalb der Buchstabe `E` schliesslich dreimal in der Tabelle vorkommt. Damit erreicht man eine Friedman'sche Charakteristik von $\mathrm{FC}(G) = 0.0072\ldots$ im Gegensatz zum Klartext mit $\mathrm{FC}(T) = 0.0543\ldots$.

Zu Teil b: Da der Text aus insgesamt 37 Zeichen besteht, kann man mit jeder Chiffrierungstabelle mit mehr als 37 Feldern erreichen, dass jedes Symbol des Geheimtextalphabetes maximal einmal vorkommt.

Lösung zur Aufgabe 6

Folgender einziger Punkt kann als Vorteil angegeben werden:

- Die Verschlüsselung ist genau auf den Klartext zugeschnitten, was eine hohe Übertragungssicherheit garantieren kann.

Jedoch wird diese Übertragungssicherheit durch fast sinnlosen Aufwand erhalten:

- Die Chiffrierungstabelle kann erst generiert werden, wenn der Klartext bekannt ist. So wird dann eine Tabelle erstellt und diese muss dann auf sicherem Kanal ausgetauscht werden. Bei kurzen Klartexten ist der Aufwand zum Schlüsselaustausch fast so gross wie für die Nachricht selbst.
- Die Chiffrierungstabelle ist wohl für einen einmaligen Einsatz gedacht (vor allem wenn auch noch gewisse Buchstaben fehlen). Dazu wäre das Verfahren One-Time-Pad besser geeignet, da dort immerhin der Schlüssel verteilt werden kann, bevor der Klartext bekannt ist.

Lösung zur Aufgabe 7

Gemäss den Zusammenhängen zwischen den verschiedenen Nukleotiden erhält man die Häufigkeiten:

Nukleotid	A	C	G	T
rel. Häufigkeit	30%	20%	20%	30%

Somit würde eine 2×5-Tabelle mit je drei Plätzen für A und T sowie je zwei Plätzen für C und G zu einer Verteilung nahe der Gleichverteilung führen.

Lösung zur Aufgabe 8

a) Das dem Alphabet A–Z folgende Ausfüllen der Chiffrierungstabelle führt nur zu einer einzigen Geheimschrift, da die Grösse der Tabelle aus den verwendeten Geheimtextsymbolen ersichtlich ist und die Anzahl Geheimtextsymbole pro Klartextbuchstaben grösstenteils eindeutig aus der Sprache des Klartextes rekonstruiert werden kann. Somit würde der Schlüsselraum aus einem (oder nur sehr wenigen) Schlüsseln bestehen, wodurch die Verschlüsselung nicht sicher wäre. Hingegen erreicht man durch ein beliebiges Befüllen der Chiffrierungstabelle ein Kryptosystem mit einem grossen Schlüsselraum, welches einzig schon durch die Geheimhaltung des Schlüssels sicher ist. Das Kryptosystem kann demzufolge (bis auf den Schlüssel) offen kommuniziert werden.
   Weiter hat das Ausfüllen nach dem Zufallsprinzip den Effekt, dass nicht gerade alle Symbole für z.B. E im Geheimtextalphabet benachbart sind. Somit sind Doppelbuchstaben wie z.T. EE nicht gerade von der Form C1 C2. Somit kann aus einer ähnlichen "Ordnung" des Symbols im Geheimtextalphabet nicht drauf geschlossen werden, dass sie den gleichen Klartextbuchstaben codieren.

b) Die Befunde aus der vorigen Teilaufgabe sind jedoch bei längeren Texten eher von kosmetischer Natur, da eine Häufigkeitsanalyse zu den gleichen Resultaten führt: Die Länge der Balken in den Diagrammen ist genau die gleiche, nur die Positionen sind verändert. Das heisst, dass eine Sortierung der Balken in beiden Diagrammen aufgrund ihrer Länge zum genau gleichen Bild führen würde.

c) Je nach Text (Länge, statistische Buchstabenhäufigkeit etc.) kann es geschehen, dass es ein Cluster von 17 Balken der ca. gleich grossen Länge gibt, wie man sie im Diagramm der Häufigkeiten für die alphabetisch befüllte Tabelle findet. So könnte man ausprobieren, ob dies die Symbole für E sein könnten. Ähnlich könnten dann mit einem Cluster von ca. 10 ähnlich grossen Balken die Symbole für N erraten und ausprobiert werden. Wie schon angemerkt, hängt dieser Ansatz stark vom verschlüsselten Text ab. Denn es könnten gut auch noch weitere Geheimtextsymbole genau die gleich hohen (oder sehr ähnlichen) Häufigkeiten haben. Wenn es z.B. 15 gleich grosse Balken hat, können höchstens ca. 10 davon zu N gehören und die restlichen Symbole müssten dann (mittels Ausprobieren...) anderen Buchstaben zugewiesen werden.

d) Grundsätzlich ist der Plot mit der alphabetisch befüllten Tabelle für ein visuelles Erraten geeigneter. Vor allem weiss man dort, dass die Cluster aufgrund der Ordnung sowieso schon nebeneinander liegen. Dies erleichtert das Auffinden der Cluster erheblich. Vor allem können so auch zwei Cluster mit ähnlicher Länge gefunden werden, solange dazwischen ein Buchstabe mit einer unterschiedlichen Häufigkeit vorhanden ist. Beim anderen Plot der zufällig befüllten Tabelle müssten die Balken zuerst aufgrund ihrer Länge sortiert werden.

e) Folgende Ideen sind möglich:
- Zum einen können Cluster von gleich langen Balken identifiziert werden und gemäss ihrer Grösse dann den häufigsten Buchstaben zugeordnet werden (s. vorige Teilaufgaben).
- Ein weiterer Angriffpunkt sind die speziell seltenen Buchstaben wie zum Beispiel `Q`, welches auch bei seltenen Texten immer noch nicht ein Geheimtextsymbol darstellt, welches gleich häufig wie die anderen vorkommt. Da in deutschsprachigen Texten das `Q` stets als Bigramm `QU` vorkommt, können so auch die Symbole für `U` bestimmt werden.
- Für längere Texte gibt es demnach Angriffsvarianten mittels seltenen Buchstaben oder dem Versuch, die Cluster für die häufigsten Buchstaben zu bilden. Kürzere Texte haben hingegen kaum Angriffspunkte.