Literatur

http://pisa.ipn.uni-kiel.de/Zusfsg_PISA2006_national.pdf

Profil – Das Magazin für Gymnasium und Gesellschaft, 12/2008, S. 6 ff.

Henze, N. ([6]2006): *Stochastik für Einsteiger. Wiesbaden: Vieweg.*

Meyer, J. (1995): *Einfache Paradoxien der beschreibenden Statistik. In: Stochastik in der Schule 15(2), 27–50.*

Getrost, G. (1994); Stein, G: *Fehler und Fallen der Stochastik im Stochastikunterricht. In: Prax. Math 36(6), 249–253.*

Tan, A. (1986): *A Geometric interpretation of Simpsons Paradoxon. In: Call. Math. Jour. 17 (4),* S. 340 f.

Anschrift des Verfassers

Renate Motzer
Didaktik der Mathematik
Universität Augsburg
Universitätsstr. 10
86135 Augsburg
Renate.Motzer@math.uni-augsburg.de

Polyalphabetische Verschlüsselung und Stochastik

Thorsten Mehlich, Bochum

***Zusammenfassung**: Am Beispiel der Vigenère-Verschlüsselung sollen die Themen relative Häufigkeiten und Laplace-Verteilungen verdeutlicht werden. Dazu wird als erstes die Vigenère-Verschlüsselung vorgestellt. Anschließend wird der Kasiski-Test beschrieben, um ein Gefühl zu vermitteln, wie relative Häufigkeiten hier eine Rolle spielen. Danach folgt eine genaue stochastische Betrachtung der Verschlüsselung und die Herleitung des Friedman-Koinzidenzindexes.*

1 Die Vigenère-Verschlüsselung

Definition 1.1 *Eine unverschlüsselte Nachricht nennt man* **Klartext**.

Definition 1.2 *Verschlüsselt man eine Nachricht, so nennt man das Ergebnis* **Kryptogramm** *oder* **Kryptotext**.

Definition 1.3 *Bei einem* **Substitutionsalphabet** *wird jeder Klartextbuchstabe durch genau einen Buchstaben des Alphabets ersetzt.*

Definition 1.4 *Im Fall einer* **polyalphabetischen Verschlüsselung** *kann ein Klartextbuchstabe durch mehr als nur einen Buchstaben ersetzt werden. Wie er substituiert wird, hängt von der Verschlüsselungsmethode ab.*

Bei der Vigenère-Verschlüsselung handelt es sich um eine polyalphabetische Verschlüsselungsmethode. Hierbei benutzt man für verschiedene Buchstaben des Klartextes unterschiedliche Substitutionsalphabete. Ein gutes Konzept hierfür machte bereits Vigenère im 16. Jahrhundert. Er benutzte dazu die nach ihm benannte Vigenère-Tabelle (Tab. 1).

Zur Verschlüsselung benötigt man ein Schlüsselwort, das periodisch wiederholt wird. Sei das Schlüsselwort z. B. „BACH“ und die Nachricht „GEHENACHWESTEN“.

Schlüssel:	B	A	C	H	B	A	C	H
Klartext:	G	E	H	E	N	A	C	H
Kryptogramm:	H	E	J	L	O	A	E	O
Schlüssel:	B	A	C	H	B	A		
Klartext:	W	E	S	T	E	N		
Kryptogramm:	X	E	U	A	F	N		

Jeden einzelnen Buchstaben des Klartextes sucht man in der obersten Zeile (siehe Tab. 1). Dann geht man in der entsprechenden Spalte so weit nach unten, bis man die Zeile mit dem Buchstaben (linker Rand der Tabelle) des Schlüssels findet. Beim Schnittpunkt der Spalte und der Zeile steht der Buchstabe für das Kryptogramm. In dem Beispiel beginnt man mit dem *G* in der obersten Zeile und geht dann zwei Zeilen nach unten zum B-Alphabet. Beim Schnitt der Senkrechten und der Waagerechten findet man das *H* für die verschlüsselte Nachricht. Um eine Nachricht zu dechiffrieren, geht man genau umgekehrt vor.

1.1 Angriff auf eine Vigenère-Verschlüsselung

Wenn ein Angreifer eine auf diese Art und Weise verschlüsselte Nachricht dechiffrieren möchte, dann

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Tab. 1: Vignère-Tabelle

wäre es günstig, wenn er zumindest die Schlüsselwortlänge bestimmen könnte. Der deutsche Major F. W. Kasiski veröffentlichte 1863 hierzu eine Methode. In jeder Sprache findet man bestimmte Bigramme häufiger als andere. In der deutschen Sprache besitzt das Bigramm „EN“ die größte relative Häufigkeit. Je öfter ein Bigramm im Klartext vorkommt, desto größer ist die Chance für das Auftreten des folgenden Falles.

Schlüssel:	D	A	C	H	D	A	C	H
Klartext:	G	R	E	N	Z	E	N	S
Kryptogramm:	J	R	G	U	C	E	P	Z
Schlüssel:	D	A	C	H	D	A	C	H
Klartext:	I	N	D	O	F	F	E	N
Kryptogramm:	L	N	F	V	I	F	G	U

Das Bigramm „EN“ steht hier zweimal an der gleichen Stelle des Schlüsselwortes. Dadurch entsteht im Kryptogramm an den jeweiligen Stellen ebenfalls das gleiche Bigramm. Mit welcher Wahrscheinlichkeit tritt dieser Fall ein? Es sei m die Länge des Schlüssselwortes. Ein Bigramm, wie z. B. im obigen Fall „EN“ komme im Klartext x-mal vor. Für ein Bigramm existieren nur m unterschiedliche Positionen im Schlüsselwort. Falls $x > m$ ist, so tritt der obige Fall mindestens einmal auf. Wenn $x < m$ ist, dann tritt der Fall mit der Wahrscheinlichkeit

$$1 \cdot \frac{m-1}{m}\frac{m-2}{m} \cdot \ldots \cdot \frac{m-x+1}{m} \qquad (1)$$

nicht ein. Damit ist die Wahrscheinlichkeit für das Eintreten dieses Falles

$$1 - 1 \cdot \frac{m-1}{m}\frac{m-2}{m} \cdot \ldots \cdot \frac{m-x+1}{m} \qquad (2)$$

Da es in jeder Sprache Bigramme gibt, die häufig auftauchen, wird dieses Ereignis sicherer und öfter auftreten, je länger die Nachricht ist.

In dem Klartext befindet sich die Buchstabenfolge „EN“ mehrmals. Falls zufälligerweise das Schlüsselwort in den jeweiligen Fällen auch an derselben Stelle wiederholt wird (das Schlüsselwort wird bei der Verschlüsselung periodisch verwendet), dann findet man auch in dem verschlüsselten Text eine sich wiederholende Buchstabenfolge. In diesem Beispiel findet man zweimal „GU“ im Kryptogramm. Falls diese Wiederholung durch identische Bigramme im Klartext an den gleichen Stellen des Schlüsselwortes entstanden ist, dann beträgt der Abstand der Bigramme im Kryptogramm ein ganzzahliges Vielfaches der Schlüsselwortlänge. Ein Angreifer kann also im verschlüsselten Text nach sich wiederholenden Bigrammen suchen. Möglicherweise stammen sie von identischen Bigrammen im Klartext, die an gleichen Schlüsselwortpositionen stehen. In diesem Fall kann der Abstand dieser identischen Teile nur ein Vielfaches der Schlüsselwortlänge sein. Beim obigen Beispiel ergibt sich als Abstand für „GU“ $12 = 3 \cdot 4$, also dreimal die Schlüsselwortlänge. Findet man mehrere identische Kryptogrammteile, so bestimmt man für alle deren Abstand und zerlegt den Abstand in seine Primfaktoren. Jedoch resultiert nicht jede Bigrammwiederholung im Kryptogramm aus einer Bigrammwiederholung im Klartext an der gleichen Schlüsselwortstelle. Derartige Wiederholungen tauchen auch zufällig auf. Deshalb sind die am häufigsten vorkommenden Faktoren nur gute Kandidaten für die Schlüsselwortlänge. Auf diese Weise schränkt man die Menge aller möglichen Schlüsselwortlängen

I	T	E	K	G		Y	X	R	X	E^{10}	I	K	V	H	Y		P	M	Z	H	F^{20}
G	K	M	R	A		P	R	O	G	U^{30}	H	H	S	R	E		A	P	L	W	P^{40}
T	X	P	M	V		L	V	F	I	F^{50}	R	U	J	M	Y		R	M	C	B	J^{60}
T	B	R	Q	J		C	Z	W	W	N^{70}	X	R	N	Y	B		J	C	B	G	U^{80}
G	L	R	W	U		L	V	T	J	M^{90}	N	A	U	Z	G		U	T	P	N	A^{100}
U	A	H	N	X		H	E	F	P	T^{110}	E	A	P	P	I		F	G	U	G	B^{120}
R	A	U	L	G		Y	L	N	Y												

Tab. 2: Kryptogramm

ein. Man bestimmt Testkandidaten, die eher geeignet als andere erscheinen.

Dazu wird ein Beispiel betrachtet. Man hat ein Kryptogramm abgefangen (Tab. 2), das nur eine Buchstabenfolge ohne Verwendung von Leerzeichen ist. Zur besseren Übersicht wird die Nachricht in 5-er-Gruppen geschrieben.

Nun betrachtet man für mehrfach auftauchende Buchstabenfolgen ihren Abstand und davon die Primfaktorzerlegung.

XR 64 = 2^6

JM 37 = 37

CB 19 = 19

GU 50 = $2 \cdot 5^2$; 16 = 2^4; 66 = $2 \cdot 3 \cdot 11$

LV 40 = $2^3 \cdot 5$

Der Faktor, der in drei Fällen auftritt, ist $2^3 = 8$. Dies scheint ein guter Kandidat für die Schlüsselwortlänge zu sein. Das würde acht Substitutionsalphabete ergeben. Für jeweils alle Buchstaben, die einen Abstand von acht Stellen im Kryptogramm haben, erstellt man eine Häufigkeitsanalyse (Tab. 3). Jede Spalte steht für eines der acht Substitutionsalphabete.

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

Tab. 3: Häufigkeitsanalyse der einzelnen Spalten

Man macht den Ansatz, dass in jedem Alphabet der Buchstabe mit der größten Häufigkeit dem *E* entspricht. Dann ergeben sich bei Rückwärtsbetrachtung mit der Vigenère-Tabelle die Schlüsselwortalphabete. Den entsprechenden Buchstaben des Schlüsselwortes erhält man, indem man von dem Substitutionsbuchstaben für das *E* vier Buchstaben zurückzählt, da der Schlüsselwortbuchstabe das verschlüsselte *A* ist. Wird z. B. das *E* durch das *G* substituiert, dann wurde das *C*-Alphabet verwendet. Die nächste Tabelle gibt die häufigsten Buchstaben des *i*-ten Schlüsselwortbuchstabens des vorliegenden Kryptogramms an.

i	1	2	3	4
Schlüsselwortbuchstabe	E, F, I, P, T	L	A, N	W, B

i	5	6	7	8
Schlüsselwortbuchstabe	C	H	C	N

Jetzt kommt der aufwendigste Teil. Für einige Buchstaben des Schlüsselwortes macht man Annahmen. Diese testet man, indem man mit diesen Schlüsselwortteilen die entsprechenden Teile des Kryptogramms entschlüsselt. Ergeben sich sinnvolle Wörter oder Wortteile, so versucht man diese sinnvoll zu ergänzen. Auf diese Weise kann man weitere Schlüsselwortbuchstaben bestimmen. Findet man keine Wörter oder Wortteile, dann verwirft man die Annahme. In der Praxis schreibt man dazu das periodische Schlüsselwort, den verschlüsselten Text und den Klartext untereinander.

Im siebten Schlüsselalphabet erscheint das „C“ als Schlüsselwortbuchstabe unpassend, da „CHCN“ keine deutsche Wortendung ist. Ersetzt man es durch ein „E“, dann ergibt sich die relativ häufige deutsche Wortendung „CHEN“. Dies kann als erste Annahme für die Schlüsselwortendung verwendet werden. Der vermutete Klartext wird hier zunächst klein geschrieben (Tab. 4).

In der ersten Zeile kann man „auc“ mit einem „h“ ergänzen (auch). Aus dem Klartextbuchstaben „h“ und dem Kryptogrammbuchstaben „M“ folgt der Schlüsselwortbuchstabe „F“. Dann müsste der erste Buchstabe des Schlüsselwortes ein „F“ sein. Als nächstes

				C	H	E	N					C	H	E	N				
I	T	E	K	G	Y	X	R	X	E	I	K	V	H	Y	P	M	Z	H	F
				e	r	t	e					t	a	u	c				
C	H	E	N					C	H	E	N					C	H	E	N
G	K	M	R	A	P	R	O	G	U	H	H	S	R	E	A	P	L	W	P
e	d	i	e					e	n	d	u					n	e	s	c
				C	H	E	N					C	H	E	N				
T	X	P	M	V	L	V	F	I	F	R	U	J	M	Y	R	M	C	B	J
				t	e	r	s					h	f	u	e				
C	H	E	N					C	H	E	N					C	H	E	N
T	B	R	Q	J	C	Z	W	W	N	X	R	N	Y	B	J	C	B	G	U
r	u	n	d					u	g	t	e					a	u	c	h
				C	H	E	N					C	H	E	N				
G	L	R	W	U	L	V	T	J	M	N	A	U	Z	G	U	T	P	N	A
				s	e	r	g					s	s	c	h				
C	H	E	N					C	H	E	N					C	H	E	N
U	A	H	N	X	H	E	F	P	T	E	A	P	P	I	F	G	U	G	B
s	t	d	a					n	m	a	n					e	n	c	o
				C	H	E	N												
R	A	U	L	G	Y	L	N	Y											
				e	r	h	a												

Tab. 4: Erster Test für den Klartext

kann man als vierten Schlüsselwortbuchstaben ein „S“ (das Trigramm „SCH“ findet man oft in der deutschen Sprache) vermuten. Als guten Kandidaten für den zweiten Schlüsselwortbuchstaben hatte man das „L“ bestimmt. Für den dritten Schlüsselwortbuchstaben wurden „A“ und „N“ relativ oft gefunden. Deshalb könnte man jetzt als Annahme für das Schlüsselwort „FLASCHEN“ machen. Testet man dieses Schlüsselwort an dem verschlüsselten Text, so ergibt sich der Klartext: „Dieser Test ist auch ohne die Verwendung eines Computers durchfuehrbar und erzeugt ein brauchbares Ergebnis. Schoen ist das, wenn man keinen Computer hat.“ Diesen Angriff kann man gut in Gaines (1956) noch einmal nachlesen.

1.2 Stochastische Betrachtung

Wie man beim Kasiski-Test gesehen hat, funktioniert der Angriff aufgrund zufälliger Ereignisse. Falls ein Klartext lang genug ist, dann wird z. B. das Bigramm „EN“, das die größte relative Häufigkeit in der deutschen Sprache hat, oft an gleichen Stelle eines Schlüsselwortes wieder auftauchen und dadurch gleiche Verschlüsselungsbigramme erzeugen. Deren Abstand ist dann ein Vielfaches der Schlüsselwortlänge. Ist die Schlüsselwortlänge gefunden, dann kann man durch Überprüfung der relativen Häufigkeiten die einzelnen Schlüsselwortalphabete schnell bestimmen. Dies soll nun stochastich analysiert werden. Die folgende Analyse stammt von einem der bedeutesten Kryptologen William Friedman. Er veröffentlichte 1925 den Kappa-Test oder auch Friedman-Test. Dieser Test dient dazu, die Schlüsselwortlänge bei einer Vigenère-Verschlüsselung zu bestimmen.

Zusammenfassend kann man sagen, als Erstes sucht man im Kryptogramm nach sich wiederholenden Bigrammen. Vermutlich entstanden einige dieser Bigramme durch identische Bigramme im Klartext, die an der gleichen Stelle des Schlüsselwortes stehen. In diesem Fall ist der Abstand der Bigramme im Kryptogramm ein ganzzahliges Vielfaches der Schlüsselwortlänge. Für die vermutete Schlüsselwortlänge teilt man den Text in ebenso viele Spalten ein und macht für jede Spalte eine Häufigkeitsanalyse. Der Buchstabe mit der größten Häufigkeit in einer Spalte ist vermutlich die Verschlüsselung des E‘s. Falls dies zutrifft, dann hat man den korrekten Substitutionsbuchstaben gefunden. Dies überprüft man, indem man mit dieser Annahme versucht Teile des verschlüsselten Textes zu entschlüsseln.

Als erstes benötigt man ein Lemma.

Lemma 1.1 Sei n die Buchstabenanzahl eines Textes. n_1 = Anzahl der a‘s, n_2 = Anzahl der b‘s, ..., n_{26} = Anzahl der z‘s. Sei A das Ereignis, dass zwei zufällig aus dem Text gewählte Buchstaben sind gleich sind. Dann ist

$$P(A) = \frac{\sum_{i=1}^{26} n_i(n_i-1)}{n(n-1)}. \quad (3)$$

Beweis: Die Anzahl der Möglichkeiten für zwei zufällig ausgewählte *a*'s ist $\frac{n_1(n_1-1)}{2}$, da es für das erste *a* n_1 und für das zweite *a* $n_1 - 1$ Möglichkeiten gibt. Weil die Reihenfolge nicht von Bedeutung ist, dividiert man den Term durch zwei. Betrachtet man nicht nur die *a*'s, sondern alle sechsundzwanzig Buchstaben, dann ist die Anzahl der Möglichkeiten

$$\frac{n_1(n_1-1)}{2} + \frac{n_2(n_2-1)}{2} + \ldots + \frac{n_{26}(n_{26}-1)}{2} = \sum_{i=1}^{26} \frac{n_i(n_i-1)}{2}. \quad (4)$$

Zur Wahl des ersten Buchstabens aus dem Text hat man *n* Möglichkeiten und für den zweiten Buchstaben bleiben dann noch $n - 1$ Möglichkeiten. Da auch hier die Reihenfolge nicht von Bedeutung ist, ergibt sich für die Anzahl aller möglichen Fälle $\frac{n(n-1)}{2}$.

P(A) ist die Anzahl der günstigen dividiert durch die Anzahl aller Fälle.

$$P(A) = \frac{\sum_{i=1}^{n} \frac{n_i(n_i-1)}{2}}{\frac{n(n-1)}{2}} = \frac{\sum_{i=1}^{26} n_i(n_i-1)}{n(n-1)}. \quad (5)$$

q.e.d.

Mit diesen Laplace-Verteilungen kommt man zu folgender Definition für den (Friedman-)Koinzidenzindex.

Definition 1.5 *Die Wahrscheinlichkeit P(A) aus Lemma 1.1 nennt man auch den (Friedman-)Koinzidenzindex*

$$\kappa := \frac{\sum_{i=1}^{26} n_i(n_i-1)}{n(n-1)}. \quad (6)$$

Man benötigt noch eine Definition für das zufällige Auftreten eines Buchstaben an einer Textstelle.

Definition 1.5 *Ein Text bestehe aus n Buchstaben und die Buchstaben A bis Z haben die Häufigkeiten n_1 bis n_{26}. Dann sind die Wahrscheinlichkeiten, dass die Buchstaben A bis Z an einer zufälligen Stelle im Text auftauchen, $p_1 = \frac{n_1}{n}, \ldots, p_{26} = \frac{n_{26}}{n}$.*

Für die einzelnen Sprachen kann man nun den Koinzidenzindex berechnen. Dazu bestimmt man aus einer großen Stichprobe für jede Sprache die relativen Häufigkeiten der einzelnen Buchstaben und berechnet κ. Falls alle Buchstaben in einem Text mit gleicher Wahrscheinlichkeit vorkommen, dann ist $p_i = \frac{1}{26}$ für alle *i*. So ergibt sich

Sprache	Index κ
Englisch	0,0655
Deutsch	0,0762
gleiche Buchstabenw.-keit	0,0385

Für einen Vigenère-verschlüsselten Text kann man κ bestimmen. Jedoch hat man nicht nur ein Alphabet, sondern jeder Schlüsselwortbuchstabe steht für ein Alphabet. Diese Anzahl möchte man bestimmen. Zu diesem Zweck hat Friedman den folgenden Satz bewiesen.

Satz 1.1 *Sei κ_{sp} der Koinzidenzindex der Sprache eines Vigenère-verschlüsselten Klartextes. Weiterhin seien n die Buchstabenanzahl des Textes, κ der Koinzidenzindex des Kryptogramms und s die Anzahl der unterschiedlichen Buchstaben im Schlüsselwort. Dann gilt nährungsweise für große n*

$$s = \frac{n(\kappa_{sp} - 0{,}0385)}{\kappa(n-1) + \kappa_{sp} - 0{,}0385n}. \quad (7)$$

Beweis: Durch ein Schlüsselwort der Länge *s* kann man die Nachricht in *s* Spalten einteilen. Alle Buchstaben in einer Spalte wurden mit dem gleichen Substitutionsalphabet verschlüsselt. Es werden die Anzahl der Buchstabenpaare aus gleichen Spalten gezählt und die Anzahl der Paare aus unterschiedlichen Spalten. Ein Paar kann auftauchen, weil es aus einer Spalte mit dem gleichen Substitutionsalphabet stammt. Der Grund hierfür ist die verwendete Sprache, weswegen der Faktor κ_{sp} ist. Ein Paar kann aber auch aus Spalten unterschiedlicher Substitutionsalphabete stammen. In diesem Fall ergibt sich als Faktor κ der Wert für die gleiche Buchstabenwahrscheinlichkeit ($\kappa = 0{,}0385$).

Die Anzahl der Paare mit gleichen Buchstaben, die vom gleichen Substitutionsalphabet des Schlüsselwortes stammen, sind

$$\frac{n\left(\frac{n}{s} - 1\right)}{2} = \frac{n(n-s)}{2s}. \quad (8)$$

Zur Wahl des ersten Buchstabens gibt es *n* Möglichkeiten. Dieser Buchstabe liegt in einer bestimmen Spalte. Aus dieser Spalte stehen noch $\frac{n}{s} - 1$ weitere Buchstaben zur Verfügung, um mit dem ersten Buchstaben ein Paar zu bilden. Da die Reihenfolge nicht von Bedeutung ist, dividiert man noch durch zwei.

Des Weiteren ist die Anzahl der Paare von Buchstaben aus verschiedenen Substitutionsalphabeten des Schlüsselwortes

$$\frac{n(n - \frac{n}{s})}{2} = \frac{n^2(s-1)}{2s}. \quad (9)$$

Hat man einen der *n* Buchstaben ausgewählt, so bleiben noch $n - \frac{n}{s}$ Buchstaben aus einer anderen Spalte

übrig. Da auch hier die Reihenfolge nicht von Bedeutung ist, dividiert man wieder durch zwei.

Die durchschnittliche Anzahl A an Paaren gleicher Buchstaben ist somit

$$A = \frac{n(n-s)}{2s} \cdot \kappa_{sp} + \frac{n^2(s-1)}{2s} \cdot 0{,}0385. \qquad (10)$$

Dann ist die Wahrscheinlichkeit *P(A)*, ein Paar gleicher Buchstaben zu treffen,

$$\begin{aligned} \kappa &= \frac{A}{\frac{n(n-1)}{2}} = \frac{n-s}{s(n-1)} \cdot \kappa_{sp} + \frac{n(s-1)}{s(n-1)} \cdot 0{,}0385 \\ &= \frac{(\kappa_{sp} - 0{,}0385)n + s(n \cdot 0{,}0385 - \kappa_{sp})}{s(n-1)} \\ &\Leftrightarrow \quad s(\kappa(n-1) - (n \cdot 0{,}0385 - \kappa_{sp})) = n(\kappa_{sp} - 0{,}0385) \\ &\Leftrightarrow \quad s = \frac{n(\kappa_{sp} - 0{,}0385)}{\kappa(n-1) + \kappa_{sp} - n \cdot 0{,}0385}. \end{aligned}$$

q.e.d.

Man beachte, dass der Satz 1.1 nur die unterschiedliche Anzahl an Schlüsselwortbuchstaben angibt. Bei Schlüsselwörten wie „BANANA“ liefert der Friedman-Test $s = 3$. Das Schlüsselwort besteht aber aus sieben Buchstaben. Nach einem durchgeführten Friedman-Test, wendet man den Kasiski-Test an und sucht speziell nach Werten in der Nähe von s. Beide Tests zusammen ergeben ein gutes Verfahren zur Bestimmung von guten Kandidaten für die Schlüsselwortlänge.

Zum Abschluss soll noch der Friedman-Test für das vorherige Beispiel durchgeführt werden. Die verwendete Sprache ist Deutsch und somit $\kappa_{sp} = 0{,}0762$. Des Weiteren ist $n = 129$. Für die Buchstabenanzahl der n_i's ergibt sich

n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
7	5	3	0	5	5	9	6	4

n_{10}	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}	n_{16}	n_{17}	n_{18}
5	3	6	6	6	1	9	1	10

n_{19}	n_{20}	n_{21}	n_{22}	n_{23}	n_{24}	n_{25}	n_{26}
1	6	9	4	4	5	6	3

Damit erhält man $\kappa \approx 0{,}041$ und daraus folgt

$$s = \frac{129 \cdot (0{,}0762 - 0{,}0385)}{0{,}041 \cdot (129-1) + 0{,}0762 - 0{,}0385 \cdot 129} \approx 13{,}89. \qquad (11)$$

Dieser Wert ist noch nicht besonders gut, da der gesuchte Wert acht ist. Der Grund hierfür ist der kurze Text. Zum Nachlesen ist Beutelspacher (2007) empfehlenswert. Die stochastischen Grundlagen kann man in Dehling und Haupt (2003) nachschlagen.

4 Schlussbemerkungen

An dem Beispiel der Vigenère-Verschlüsselung wurde das Prinzip der relativen Häufigkeiten und ihre Anwendung erklärt. Zuerst wendet man den Kasiski-Test an, wobei man ein intuitives Verständnis für die relativen Häufigkeiten entwickelt, wenn man Bigramme des Klartextes und deren sich wiederholende Verschlüsselungen im Kryptogramm betrachtet. Dabei erkennt man, dass es zu identischen Verschlüsselungen gleicher Bigramme aus dem Klartext kommen kann, wodurch es möglich ist, auf die Schlüsselwortlänge zu schließen. Mit der Durchführung einer Häufigkeitsanalyse bestimmt man dann Kandidaten für Substitutionsalphabete des Schlüsselwortes. Dieses intuitive Verständnis der Bedeutung der relativen Häufigkeiten wird dann stochastisch genau ausgearbeitet, was zum Friedman-Test führt. Diese ist ein schönes Beispiel aus der Praxis für die Anwendung von relativen Häufigkeiten und Laplace-Verteilungen.

Die stochastische Betrachtung der Vigenère-Verschlüsselung kann in der Schule nach einer Einführung der relativen Häufigkeiten und Laplace-Verteilungen behandelt werden. Es ist eine gute Übung für die Verwendung von relativen Häufigkeiten und Laplace-Verteilungen. Hierbei ist es wichtig, dass die Schüler sich klar machen, was jeweils die absoluten Häufigkeiten und die Gesamtzahl aller Möglichkeiten sind. Für die Einführung in das Thema ist jedoch viel Zeit notwendig.

Literatur

Beutelspacher, A. (2007): Kryptologie. Wiesbaden: Friedr. Vieweg & Sohn Verlag | GWV Fachverlage GmbH.

Gaine, H. (1956): Cryptanalysis.

Schneier, B. (1997): Handbook of Applied Cryptography. CRC Press LCC.

Dehling, H./??? Beate, H. (2003): Einführung in die Wahrscheinlichkeitstheorie und Statistik. Berlin/ Heidelberg 2003: Springer.

http://www.cryptool.de

htpp://www.cryptoportal.org

Anschrift des Verfassers

Thorsten Mehlich
Fakultät für Mathematik
Ruhr-Universität Bochum, NA 3/28
44801 Bochum
thorsten.mehlich@ruhr-uni-bochum.de