The Unicity Point of a Cryptosystem

Recall that the key equivocation $H(K|C^n)$ measures the remaining uncertainty in the key after the attacker has obtained an $n$-gram of ciphertext.

The *unicity point* of a cryptosystem is the smallest value of $n$ such that $H(K \mid C^n) \approx 0$. 
Recall key equivocation:

\[ H(K|C^n) = H(K) + H(M^n) - H(C^n) \]

Thus the unicity point is the least \( n \) such that,
\[ H(K) = H(C^n) - H(M^n). \]

Can we estimate this for a natural language?
$\mathcal{K} =$ key space

$\mathcal{M} =$ plaintext alphabet

$\mathcal{C} =$ ciphertext alphabet

For any set $\mathcal{X}$, $|\mathcal{X}|$ denotes the number of elements in $\mathcal{X}$. 
Assumptions:

- $H(K) = \log_2 |K|$
- $H(M^n) \approx n \cdot H_L$, at least for large $n$.

Always have: $H(C^n) \leq \log_2 |C^n| = n \cdot \log_2 |C|$

Thus

$$H(K) = H(C^n) - H(M^n) \iff n \geq \frac{\log_2 |K|}{\log_2 |C| - H_L}.$$
If we go further and assume

\[ H(C^n) \approx n \log_2 |C| \]

then we get the following estimate for the unicity point:

\[ U \approx \frac{\log_2 |K|}{\log_2 |C| - H_L}. \]

This is equivalent to assuming individual ciphertext characters are independent
Simple substitution cipher on English:

\[
\log_2 |\mathcal{K}| = \log_2 (26!) \approx 88.4 \\
\log_2 |\mathcal{C}| = \log_2 (26) \approx 4.7 \\
H_E \approx 1.5
\]

Thus \( U \geq \frac{88.4}{(4.7 - 1.5)} = 27.6 \).

For a polyalphabetic cipher of blocksize \( d \), \(|\mathcal{K}| = (26!)^d\), so unicity is approx. \( 27.6d \).
Unicity of a Linear Cipher

\[ H(M^n) = 1.5n \]

\[ H(C^n) = \log_2(26^n) = n \cdot \log_2(26) \approx 4.7n \]

\[ H(K) = \log_2(|K|) \]

Assume we are using a blocksize of \( d \). How many invertible \( d \times d \) matrices mod 26?

\[ |K| = \prod_{j=0}^{d-1} (2^d - 2^j)(13^d - 13^j) \]

© 2011 Clifford Bergman
We get the following estimates of the unicity point for a linear cipher: \( U_d \approx \log_2(|\mathcal{K}|)/3.2 \)

\[
\begin{array}{|c|c|}
\hline
d & U_d \\
\hline
5 & 34 \\
10 & 138 \\
15 & 310 \\
20 & 552 \\
25 & 863 \\
\hline
\end{array}
\]
The Redundancy of a Language

Assume a language $L$ has entropy $H_L$. A random language on the same alphabet, $\mathcal{M}$, would have entropy $\log_2 |\mathcal{M}|$. 
\[
\frac{H_L}{\log_2 |\mathcal{M}|}
\]
measures the efficiency of \( L \).

Define the \textit{redundancy} of \( L \) to be

\[
R_L = 1 - \frac{H_L}{\log_2 |\mathcal{M}|}.
\]

For our model of English, \( R_E \approx 1 - 1.5 / \log_2(26) \approx .68 \).
If we assume that $\mathcal{M} = \mathcal{C}$ and that the distribution of $\mathcal{C}$ is uniform, then unicity and redundancy are related by the formula

$$U = \frac{\log_2 |\mathcal{K}|}{R \log_2 |\mathcal{M}|}.$$
Perfect Secrecy

A cryptosystem is said to have **perfect secrecy** if for all \( n > 0 \),
\[
H(M^n \mid C^n) = H(M^n).
\]
Equivalently, the variables \( M^n \) and \( C^n \) are independent.

**Theorem:** If a cryptosystem is perfect, then there are at least as many keys as messages.

Example: the one-time pad