CRYPTOGRAPHY HANDOUT 03

SUBSTITUTION CIPHER

Substitution Example

The following example is from Douglas Stinson's Cryptography: Theory and Practice.

Given the ciphertext (encrypted with a substitution cipher):

```
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR
```

The following steps will walk through how to do the cryptanalysis.

1. Do a frequency count for the text.

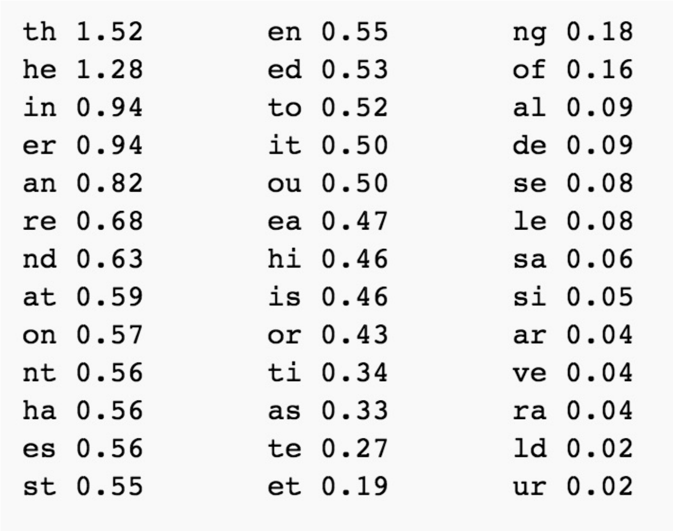
Letter	Count	Letter	Count
A		N	
B		O	
C		P	
D		Q	
E		R	
F		S	
G		T	
H		U	
I		V	
J		W	
K		X	
L		Y	
M		Z	

2. Which letter occurs most? This letter likely corresponds with e, the most frequency-occurring English letter.

3. The next set of most-frequent letters aren't as easy to match up. Let's look at the **bigrams** or digrams instead (pairs of letters). Count the following bigrams:

Bigram	Count	Bigram	Count
DZ		ZW	
NZ		ZU	

4. You should find that `DZ` and `ZW` occur the most often, so what are some guesses (based on the English language) that the letters corresponding to `D` and `W` are what? Use the bigram frequency table for reference:



th	1.52	en	0.55	ng	0.18
he	1.28	ed	0.53	of	0.16
in	0.94	to	0.52	al	0.09
er	0.94	it	0.50	de	0.09
an	0.82	ou	0.50	se	0.08
re	0.68	ea	0.47	le	0.08
nd	0.63	hi	0.46	sa	0.06
at	0.59	is	0.46	si	0.05
on	0.57	or	0.43	ar	0.04
nt	0.56	ti	0.34	ve	0.04
ha	0.56	as	0.33	ra	0.04
es	0.56	te	0.27	ld	0.02
st	0.55	et	0.19	ur	0.02

(From `https://en.wikipedia.org/wiki/Bigram#Bigram_frequency_in_the_English_language`.)

5. We have a choice here. Suppose `W` corresponds to the plaintext letter of `d`. Since `ZRW` and `RZW` both occur at the beginning, and since `RW` occurs again later on and *nd* is a common digram, let's try saying that `R` corresponds to `n`. At this point, we have 3 letters deciphered. What does your text look like so far?

YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR

6. We can keep looking at bigrams and frequently-occurring letters to slowly fill in the rest of the letters until we get to the following message. Can you fill in the final letters?

```
o-r-riend-ro--arise-a-inedhise--t---ass-it

hs-r-riseasi-e-a-orationhadta-en--ace-hi-e

he-asnt-oo-in-i-o-redso-e-ore-ineandhesett

-ed-ac-inhischair-aceti-ted--to-ardsthes-n
```