Math 400 Fall 2017
Cryptology HW 2 Solutions
Due Thursday, September 14

1. Decrypt the following message, which was encrypted with a monoalphabetic substitution cipher:

   KZRNK GJKIP ZBOOB XLCRG BXFAU GJBNG RIXRU XAFGJ BXRME MNKNG BURIX KJRXR SBUER
   ISATB UIBNN RTBUM NBIGK EBIGR OCUBR GLUBN JBGRL SJGLN GJBOR ISLRS BAFFO AZBUN
   RFAUS AGGBI NGLXM IAZRX RMNVL GEANG CJRUE KISRM BOOAZ GLOKW FAUKI NGRIC BEBRI
   NJAWB OBNNO ATBZJ KOBRC JKIRR NGBUE BRINK XKBAF QBROA LNMRG MALUF BBG

Letter	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
Frequency	16	32	5	-	7	8	22	-	16	11	13	10	8	20	12	1	1	28	7	3	14	1	2	10	-	6

Letter	B	R	G	N	A	I	U	K	O	J	L	X	M	F	S	E	Z	C	T	W	P	V	Q
Frequency	32	28	22	20	16	16	14	13	12	11	10	10	8	8	7	7	6	5	3	2	1	1	1

Bigram	NG	RI	BU	BR
Frequency	7	7	6	5

   Solution: "'I was, I think, well educated for the standard of my day. My sister and I had a German governess - a Fraulein. A very sentimental creature. She taught us the language of flowers - a forgotten study nowadays, but most charming. A yellow tulip, for instance, means 'Hopeless Love,' while a China aster means 'I Die of Jealousy at Your Feet."

   From "The Tuesday Night Club" by Agatha Christie

2. Compute **by hand** the indices of coincidence of the following strings. Then compute their mutual indices of coincidence. Show me your work.

   (a) `It is a truth universally acknowledged`

   (b) `that a single man in possession of a good`

   (c) `fortune must be in want of a wife`

   Solution: The frequency charts for each string are

(a)

Letter	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
Frequency	3	0	1	2	3	0	1	1	3	0	1	3	0	2	1	0	0	2	2	3	2	1	1	0	1	0

length 33

(b)

(c)

Letter	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
Frequency	4	0	0	1	2	1	2	1	3	0	0	1	1	4	5	1	0	0	5	2	0	0	0	0	0	0

length 33

(d)

(e)

Letter	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
Frequency	2	1	0	0	3	3	0	0	2	0	0	0	1	3	2	0	0	1	1	3	2	0	2	0	0	0

length 26

So the indices of coincidence are

(a)

$$\frac{1}{33 \cdot 32}(3 \cdot 2 + 2 \cdot 1 + 3 \cdot 2 + 3 \cdot 2 + 3 \cdot 2 + 2 \cdot 1 + 2 \cdot 1 + 2 \cdot 1 + 3 \cdot 2 + 2 \cdot 1) = \frac{40}{1056} = .03\overline{78}.$$

(b)

$$\frac{1}{33 \cdot 32}(4 \cdot 3 + 2 \cdot 1 + 2 \cdot 1 + 3 \cdot 2 + 4 \cdot 3 + 5 \cdot 4 + 5 \cdot 4 + 2 \cdot 1) = \frac{76}{1056} = .071\overline{96}.$$

(c)

$$\frac{1}{26 \cdot 25}(2 \cdot 1 + 3 \cdot 2 + 3 \cdot 2 + 2 \cdot 1 + 3 \cdot 2 + 2 \cdot 1 + 3 \cdot 2 + 2 \cdot 1 + 2 \cdot 1) = \frac{34}{650} \approx .05231.$$

To calculate the mutual indices of coincidence, we get

(a) and (b):

$$\frac{1}{33 \cdot 33}(3 \cdot 4 + 2 \cdot 1 + 3 \cdot 2 + 1 \cdot 2 + 1 \cdot 1 + 3 \cdot 3 + 3 \cdot 1 + 2 \cdot 4 + 1 \cdot 5 + 2 \cdot 5 + 3 \cdot 2) = \frac{64}{1089} \approx .05877.$$

(a) and (c):

$$\frac{1}{33 \cdot 25}(3 \cdot 2 + 3 \cdot 3 + 3 \cdot 2 + 2 \cdot 3 + 1 \cdot 2 + 2 \cdot 1 + 2 \cdot 1 + 3 \cdot 3 + 2 \cdot 2 + 1 \cdot 2) = \frac{48}{858} \approx .055944.$$

(b) and (c):

$$\frac{1}{33 \cdot 25}(4 \cdot 2 + 2 \cdot 3 + 1 \cdot 3 + 3 \cdot 2 + 1 \cdot 1 + 4 \cdot 3 + 5 \cdot 2 + 5 \cdot 1 + 2 \cdot 3) = \frac{57}{858} \approx .06653.$$

3. Compute **by hand** the index of coincidence of the following string. What is unusual about this string?

   `the quick brown fox jumps over the lazy dog`

   Solution: This text has 35 letters, with four "o"s, 3 "e"s, 2 "h", "r", "t", "u", and 1 of each other letter. Thus the index of coincidence is

$$\text{IndCo} = \frac{1}{35 \cdot 34}(4 \cdot 3 + 3 \cdot 2 + 2 \cdot 1 + 2 \cdot 1 + 2 \cdot 1 + 2 \cdot 1) = \frac{26}{35 \cdot 34} = \frac{13}{35 \cdot 17} \approx .022.$$

   This is really unusually low—much lower than you'd expect even for completely random text. In a very long text you can't go much below .036, but this text is so short that we can have a very low index of coincidence.

   (The text is famously a *pangram*: a sentence that contains each letter at least once. It is not a perfect pangram, containing each letter exactly once; such a sentence would have an index of coincidence equal to zero. Perfect pangrams exist, but tend to be strange and extremely awkard in phrasing ("Mr Jock, TV quiz PhD, bags few lynx").)

   For the remaining problems, you may use an index of coincidence calculator like the ones here: `http://cs.colgate.edu/~chris/FSemWeb/tools/coincidence.html`

   You should be able to copy and paste the ciphertexts from the PDF.

4. Which of the following is likely to be a message encrypted with a simple substitution cipher?

   (a) GSZES GNUBE SZGUG SNKGX CSUUE QNZOQ EOVJN VXKNG XGAHS AWSZZ BOVUE SIXCQ NQESX NGEUG AHZQA QHNSP CIPQA OIDLV JXGAK CGJCG SASUB FVQAV CIAWN VWOVP SNSXV JGPCV NODIX GJQAE VOOXC SXXCG OGOVA XGNVU BAVKX QZVQD LVJXQ EXCQO VKCQG AMVAX VWXCG OOBOX VZCSO SPPSN VAXUB DVVAX QJQAJ VSUXC SXXCV OVJCS NSJXV NOJQA MVBSZ VOOSH VSAWX QHGMV GWVSX CSXXC VBSNV ZVNVN SAWQZ ORVXJ CVOQE JCGUW NVA

   (b) DWVQP IIKOP UUYGC ZJDRU ZDSHI CXEAO AKRZC QAMSM DNQLF LUJYI IMJPJ VJZQL GCJSN XTXFL MWOLW IFUQK DBBEY HVMVF ZOJXV FYMJA RDTGT TZQKL YNHPD UPUYU XKNOI DXZXG IHIWK VXZET XFMOS GKIWU EIFDW RLXHH PZXPI VFWKL THEAS IROWC GJAYJ KKODL WXFPI ZVUIK LEXEL IOWVC YMFMR UIZUD CETFE TBPIX SWSPZ MRPKP LIYKL FGSTJ ZTPUH AEBQC QAEPQ GIAKH TDUVM KFGEU MWHAY ZGVSJ LNLJJ MLPEO YEMZU PYMEW XLG

   Solution: We compute the indices of coincidence. The index of coincidence of the first sample is $\approx .063$, and the index of coincidence of the second sample is $\approx .039$. Thus the second looks more like random text, and the first looks like it was encrypted with a simple substitution cipher.

5. Consider the following ciphertext:

   togmg gbymk kcqiv dmlxk kbyif vcuek cuuis vvxqs pwwej koqgg phumt whlsf yovww knhhm rcqfq vvhkw psued ugrsf ctwij khvfa thkef fwptj ggviv cgdra pgwvm osqxg hkdvt whuev kcwyj psgsn gfwsl jsfse ooqhw tofsh aciin gfbif gabgj adwsy topml ecqzw asgvs fwrqs fsfvq rhdrs nmvmk cbhrv kblxk gzi

   (a) Use the Kasiski test (either by shifting the text over one-by-one, or by comparing repeated trigrams) to guess the keyword length. (Hint: the repeated trigrams are LXK at 17 and 232; TWH at 54 and 134; NGF at 149 and 174; and SFS at 156 and 209).

   Solution: We see that the offets that the trigrams have are $232-17 = 215; 124-54 = 80; 174-149 = 35; 209-156 = 53$. Since the first three have 5 as a common factor, we might guess the keyword has length five.

   Alternately, we can offset the text and look for coincidences. With an offset of 2 or 3 there are eight coincidences; with an offset of 4 there are 4; with an offset of 5 there are 16; with an offset of 6 there are 10; with an offset of 7 or 8 there are 9. This gives more weak evidence that the keyword has length 5.

   (b) Use a tool like `http://cs.colgate.edu/~chris/FSemWeb/tools/splitter.html` to split the text up into substrings. Test the index of coincidence of each substring, and use this to determine the key length. Does this match your answer in part (a)?

   Solution: If we use three substrings, the indices are $.044, .041, .044$. If we use four, the indices are $.039, .042, .040, .045$. If we use five, the indices are

.062, .069, .059, .072, .056, which is high enough that this is almost certainly the correct keylength. But for completeness we see that with six substrings we get .037, .045, .046, .044, .031, .035, and with seven we get .046, .029, .032, .041, .036, .032, .046. So at this point we're pretty sure the keyword has length 5.

(c) Find the keyword and decrypt the message. You can use frequency analysis on substrings, or you can use mutual indices of coincidence, but be sure to show intermediate steps so I can see what you did.

Solution: The keyword is `codes` and the message is

> radio, envisioned by its inventor as a great humanitarian contribution, was seized upon by the generals soon after its birth...and impressed as an instrument of war....But radio turned over to the commander a copy of every enemy cryptogram it conveyed....Radio made cryptanalysis an end in itself.

6. Would the method we used on the Vigenère cipher work to decrypt an autokey cipher? Why or why not?

Solution: No, because the keystream does not repeat. As we saw in class this week, the autokey cipher is quite vulnerable, but since the keystream doesn't repeat you can't use substring analysis to find the keyword. However, the keystream comes from *somewhere*, and we can attack that connection.