Homework 2: Cryptanalysis

This homework is due **Friday, October 13** at **6 p.m.** and counts for 6% of your course grade (4% if you are a graduate student taking CMSC 33250). You will have a budget of four extensions (24-hour periods) over the course of the quarter that you can use to turn assignments in late without penalty and with no questions asked. You cannot consume partial days. Once your extensions are used up, further extensions will only be granted in extraordinary circumstances.

We encourage you to discuss the problems in general terms with other students in the class. However, the answers you turn in must be your own original work, and you are bound by the University's policy on Academic Honesty and Plagiarism. Also, please document any material discussions you had with others about this assignment (e.g., "Note: I discussed this exercise with Jane Smith").

Solutions should be submitted electronically via chisubmit in plain text format using the template found at `hw2/hw2.txt` in the upstream repository.

Solve both of the following problems. You may want to write some short programs to help in a language of your choice; feel free to submit them along with your answers. They may help us understand your answers, but the programs will not be graded.

1. Here is some ciphertext that was produced with a Vigenère cipher:

```
UEMLWSTJNDDBMFLUSLUEPLVQADRQKGIZBSSLRQDQMDMRGBRVFWPINHGBEZIJLGNLVCJLRG
ZRDNHWRFPSKIXUAERLDVGARGYDPYWQOZYIEKNFEHWZGESCHRRJNZHBMGAFOXRYGUBBOWVH
UGWJJEAVNBWOCMAAULUSGGPCHVXUSGCRHVPGZSPKIVFSZQHRPWBTWGZRJFONLIFSARSKMP
ZNFAUIYSGSZWSISEWKXWNKCSYWWVFVBBRVZSGWKQWRUHFEWCFMPVWVHNLNQKQJVVRBPLEY
AGMZDXNAAHAJVVLLOQWLRFGWYDXVGAOJGRBFESLXHVSGWKQQBVRFJFVLHGCCUECZLWJWIE
KRQPVXUWQWOFMCDVBAVSSENHDHQNLVQOFSZHHHAUWPARBYHEAVRZAFXEAPOHHRTAASAUMA
YZCZHVAUEMLWSTJNDDBMFZRORLPLTNGAGSAENHDHQNLVQWOXUWBFUDRQUBALXXRJFQEHRP
WCFWFXVURQNBTGGTFWSLVUNZCRVVLUAODVRVRGEJRRVNFKXRQUBALXXNLVCJDPUSERJHWF
SFGQPTGABBOPEXAAUOXGUSYUKUMGZZGDDVQLBPNHEXAADNDGGAPSXBEAQNRRHVFSEMEWMF
LUSKUIGAPOHOCCGFGEEPRLBPNHEXKHQDDWLKGSIEYGAGWOLRSWNGEEPRLBRKVSOQNBUNRB
OADNDGGAPOHPINFFQNBTGGYCCBVRDNHAGXRUUBKOSTQUOOUEVKRRWQYZTRFKIPRYNZEVWH
WFHDHIYWPHNRRVUSFKQXVWETKXRQSGWKQANKVBRRPIWQWJDGNKRWJWLRMAWPHHFLNHAVAU
APVMXIFLVCJHHJZRHDHVEWDIEUMAYFIOSIPLRRYUMZAAOHVXBHECRLHRLUSEUHRUEMLWMB
FXSUVXBDNKAQJBJPSIHRGAFIJFSAKGWPXXVGAOHWLRWSTWUKHWQHDDXGZVGEVEIABZWWMB
FBTPKIEATVPRJAGGPALRTXBFYHHGGVBYUMZAAOPHSAWFSHIEFYVJAQMALUSBLJGZNAAQHZ
```

Assume that encrypting with the key letter `A` results in no change, `B` results in an increment by one place in the alphabet, `C` results in an increment by two places, etc.

What is the key? (Please show your work.)

2. Here is a table of the relative frequency of letters in English text generally:

A: 0.08167	B: 0.01492	C: 0.02782	D: 0.04253	E: 0.12702	F: 0.02228	G: 0.02015
H: 0.06094	I: 0.06996	J: 0.00153	K: 0.00772	L: 0.04025	M: 0.02406	N: 0.06749
O: 0.07507	P: 0.01929	Q: 0.00095	R: 0.05987	S: 0.06327	T: 0.09056	U: 0.02758
V: 0.00978	W: 0.02360	X: 0.00150	Y: 0.01974	Z: 0.00074		

Here is a specific plaintext:

```
thiscourseintroducestheprinciplesandpracticeofcomputersecurityitaimsto
teachyouhowtomodelthreatstocomputersystemsandhowtothinklikeanattackera
ndadefenderitpresentsstandardcryptographicfunctionsandprotocolsandgive
sanoverviewofthreatsanddefensesforsoftwarehostsystemsnetworksandtheweb
italsotouchesonsomeofthelegalpolicyandethicalissuessurroundingcomputer
securityinareassuchasprivacysurveillanceandthedisclosureofsecurityvuln
erabilitiesthegoalofthiscourseistoprovideafoundationforfurtherstudyinc
omputersecurityandtohelpyoubetterunderstandhowtodesignbuildandusecompu
tersystemsmoresecurelyseetheschedulefordetailsthecourseworkconsistsoff
ivehomeworksfiveprojectsandafinalexaminadditionstudentsenrolledincmscm
ustsubmitaweeklypaperresponsebasedonthereadingswhichareoptionalforunde
rgraduatesallassignmentsmustbedoneindividuallywiththeexceptionofprojec
tsandwhichwillbedoneingroupsyourcoursegradewillbebasedonthefollowingco
mponents
```

The *population variance* of a finite population X of size N and mean μ is given by

$$\mathrm{Var(X)} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{x_i} - \mu)^2.$$

(a) What is the population variance of the relative letter frequencies in English text generally?

(b) What is the population variance of the relative letter frequencies in the given plaintext? (Hint: Make sure to normalize the relative letter frequencies before computing the population variance — i.e., make sure the frequencies sum to 1.)

(c) For each of the following keys — `yz`, `xyz`, `wxyz`, `vwxyz`, `uvwxyz` — encrypt the plaintext with a Vigenère cipher and the given key, then calculate and report the population variance of the relative letter frequencies in the resulting ciphertext. Describe and briefly explain the trend in this sequence of variances.

(d) Viewing a Vigenère key of length k as a collection of k independent Caesar ciphers, calculate the mean of the frequency variances of the ciphertext for each one. (E.g., for key yz, calculate the frequency variance of the even numbered ciphertext characters and the frequency variance of the odd numbered ciphertext characters. Then take their mean.) Report the result for each key in part (c). Is the mean variance like those observed in part (b)? Part (c)? Briefly explain.

(e) Consider the ciphertext that was produced with key uvwxyz. In part (d), you calculated the mean of six variances for this key. Revisit that ciphertext, and calculate the mean of the frequency variances that arise if you had assumed that the key had length 2, 3, 4, and 5. Does this suggest a variant to the Kasiski attack? (Don't say no!) Briefly explain. □