Homework 2: Cryptanalysis

This homework is due **Friday, September 27, 2024** at **11:59 p.m.** and counts for 5% of your course grade. Late submissions will be penalized by 10% plus an additional 10% every 5 hours until received. Late work will not be accepted after 24 hours past the deadline. If you have a conflict due to travel, interviews, etc., please plan accordingly and turn in your homework early.

We encourage you to discuss the problems and your general approach with other students in the class. However, the answers you turn in must be your own original work, and you are bound by the Honor Code. Submit your solution on GradeScope following the instructions at the end of this document.

To solve these problems, you will probably want to write some short programs; submit them with your answers. We recommend Python, but you may use any common language or numerical package.

1. Here is some ciphertext that was produced by a simple substitution cipher that we learned about in class:

   ```
   OBRGXIMYAZZAWCATBNMUYYHAZNVGFCXPVVSIJSVLKIFAVGBIECAZSBWGRGRQWUCHMMOCYE
   FLGQQNKFSHQMGYALNKCIJQVEKVWXNFOYFYQBESGOYTXMAYTXSISNBPMSGOJBKFWRUTTMLS
   BNQMLLRGFNZUAWHZLBRVZGHUVZMCKJEHSLSWGXCNZYEXRIMLPXRIXNUXRNSNRPHFDHBMAY
   WKHTKNGNUXRNJUVGMYNYEYNLYYGPGYFSBNQQWUCHMMSLRWTFDYQRNOJUEWNLVUZIDHWXLH
   TLKNEXMALBRQGUMMGXCUFXLHTLLLRTROJYFIDHLIGADLUBVXENSCALVCDFFIQCFAHISILU
   XXZXNUAMZAWISRNOJYKMQYECGRSBWHAHLUFBBPDPWLJBRYOCYEAYSVYXSISPRKSNZYPHMM
   WKHXMWWMGAZNEOFMDHKORMGOKNUHTAZQRAZPWBRTQXGZFMTJAXUTRNWCAPZLUFRODLFYFL
   GUKHRODLTYRGRYWHNLRIUCNMDXOCGAKIFAQXKUQMVGZFDBVLSIJSGADLWCFGNCFMGTMWWI
   STBIMHGKXBSPVGFVWHRYHNWXSKNGHLBENHYYQPZLXUEXNHDSBGDQZIXGNQKNUXCCKUFMQI
   MMRYEYUNFHEUDIAZVUJWNGQYSFVSDNZYFNOLWGRBLJGLGTMWWISKZJAXVMXCFVEBMAAHTB
   SNGUPENMWCGBRIFFLHMYOBBBRNZIEHTAZFLTBKMUVGSYVQVMGNZYROHFKISPZLOBBVZHLB
   BKNOYBYRTHVYELSUFXGADJJISBSUTFRPZSGZPTQLQCAZHNGHGADMCCYEEODARGDLSFQHDM
   FIGKZCKYNLDWGHQEDPQHRBSBWLNKDBAMFNOJDSJTFIFMYHZXWXZHQYLBNGSQAWRHMWWQNK
   HMVYPEZLWXUXVCDFAHSQSMGXOLWWVHTMLCZXHHOUVMHHYZBKQYAHSHQWWGRGSMFIEPHFDB
   RMTLFBVLZLESOTBEXIEYQYKBFNOJDCRLAOLWEHRMWMGADYFYZRRZJIAMHYJQVMGIMNQXKU
   QNUXUUDORHENAGRMGULCFUDCFANEHNLFRTGYSXBYXIMLBIOIFYAMGUKWBNMNWXSHQGGLRM
   GUFYVMGYJHHFDLAWNEROHYEBNLANLHQNZYABBYKNPTKWMFNMHIFMJBSBJYTTQXLIPHLGAM
   FTQCSN
   ```

   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, and so on.

   What is the key? (Show your work.)

2. Briefly, what is "snakeoil cryptography"? How does it relate to the exercise above?

3. Here is a Python dictionary of the relative frequency of letters in English text:

```
{ "A": .08167, "B": .01492, "C": .02782, "D": .04253, "E": .12702, "F": .02228,
  "G": .02015, "H": .06094, "I": .06966, "J": .00153, "K": .00772, "L": .04025,
  "M": .02406, "N": .06749, "O": .07507, "P": .01929, "Q": .00095, "R": .05987,
  "S": .06327, "T": .09056, "U": .02758, "V": .00978, "W": .02360, "X": .00150,
  "Y": .01974, "Z": .00074 }
```

Here is some plaintext:

```
ethicslawanduniversitypolicieswarningtodefendasystemyouneedtobeabletot
hinklikeanattackerandthatincludesunderstandingtechniquesthatcanbeusedt
ocompromisesecurityhoweverusingthosetechniquesintherealworldmayviolate
thelawortheuniversitysrulesanditmaybeunethicalundersomecircumstancesev
enprobingforweaknessesmayresultinseverepenaltiesuptoandincludingexpuls
ioncivilfinesandjailtimeourpolicyineecsisthatyoumustrespecttheprivacya
ndpropertyrightsofothersatalltimesorelseyouwillfailthecourseactinglawf
ullyandethicallyisyourresponsibilitycarefullyreadthecomputerfraudandab
useactcfaaafederalstatutethatbroadlycriminalizescomputerintrusionthisi
soneofseverallawsthatgovernhackingunderstandwhatthelawprohibitsifindou
btwecanreferyoutoanattorneypleasereviewitsspoliciesonresponsibleuseoft
echnologyresourcesandcaenspolicydocumentsforguidelinesconcerningproper
```

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?

(b) What is the population variance of the relative letter frequencies in the given plaintext?

(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.

Submission Template

Please submit one pdf file to GradeScope, paginated per the specifications below: **HW2.pdf**. Failure to follow the provided formatting guidelines will result in an automatic **5%** deduction from your score. **Page 1** contains the answer and textual explanation of the process you used to derive your answer for Problem 1. **Page 2** contains similar information, but to Problem 2. **Page 3** contains similar information, but to Problem 3. Please use **pages 4 and beyond** to provide code and other required support details for your solutions to Problems 1 and 3. If you used the same code for both problems, please say so. If we can't read your file, we can't grade it. Make sure to comment your code with clear explanations and be sure to *cite any references used*.

```
Filename: HW2.pdf
# Page 1: Problem 1
key=XXXXXXXXXX

brief description of your approach ...

<page break>

# Page 2: Problem 2
briefly_answer_the_questions ...

<page break>

# Page 3: Problem 3
part_a_var_english=0.0000000
part_b_var_plaintext=0.0000000
part_c_var_ciphertexts=[0.0000000, 0.0000000, 0.0000000, 0.0000000, 0.0000000]
part_c_explain="briefly_describe_and_explain_trend ..."
part_d_means=[0.0000000, 0.0000000, 0.0000000, 0.0000000, 0.0000000]
part_d_explain="briefly_compare_and_explain_results ..."
part_e_means=[0.0000000, 0.0000000, 0.0000000, 0.0000000]
part_e_explain="briefly_explain_attack_variant ..."

<page break>

Provide your code and detailed work here for Problems 1 and 3 ...
```