Rapport de questions Labo #1 CRY

Auteur : Harun Ouweis

1. Quel est l’avantage d’utiliser le test du χ2 plutôt que de comparer simplement la lettre la plus fréquente dans le texte chiffré par rapport aux statistiques du langage de base ?

L'avantage d'utiliser le test du χ2 plutôt que de comparer simplement la lettre la plus fréquente dans le texte chiffré aux statistiques du langage de base réside dans la précision et la fiabilité. Le test du χ2 considère la distribution complète des fréquences des lettres, et non pas d'une seule, car en utilisant une seule lettre on pourrait avoir des erreurs car potentiellement elle ne représente pas bien la langue choisie ou que le texte n'a particulièrement pas la même fréquence d'apparition de cette lettre que notre texte de référence. En utilisant χ2, on vérifie alors que l'ensemble du texte soit cohérent.

2. . Pourquoi est-ce que le test du χ2 ne fonctionne-t-il pas directement sur un texte chiffré à l’aide du chiffre de Vigenère ?

Le test du χ2 ne fonctionne pas directement sur un texte chiffré avec le chiffre de Vigenère car ce dernier utilise plusieurs décalages (au lieu d'un seul pour César), rendant la distribution des lettres plus uniforme et diluant l'effet des fréquences naturelles des lettres dans la langue du texte clair. Cela complexifie la distinction des patterns de fréquences qui permettraient de tirer des conclusions sur la clé utilisée.

3. Que mesure l’indice de coïncidence ?

L'indice de coïncidence mesure la probabilité que deux lettres tirées au hasard dans un texte soient identiques. Il reflète la distribution des fréquences des lettres dans un texte, plus l'indice est bas plus le texte est aléatoire. cette indice facilite la cryptanalyse des chiffres comme celui de Vigenère.

4. Pourquoi est-ce que l’indice de coïncidence n’est-il pas modifié lorsque l’on applique le chiffre de César généralisé sur un texte ?

L'indice de coïncidence n'est pas modifié lorsqu'on applique le chiffre de César généralisé sur un texte car ce chiffrement décale simplement toutes les lettres d'une même valeur, préservant les fréquences relatives des lettres. Ainsi, la structure de répétition des lettres reste intacte, tout comme l'indice de coïncidence.

5. Est-il possible de coder un logiciel permettant de décrypter un document chiffré avec le chiffre de Vigenère et une clef ayant la même taille que le texte clair ? Justifiez

Non, ce n'est pas possible, utiliser une clé de Vigenère de la même taille que le texte clair transforme le chiffrement en un système équivalent à un one-time pad. Dans un tel scénario, chaque caractère du texte est chiffré avec une lettre de clé unique et aléatoire, rendant toute forme d'analyse statistique des fréquences de lettres inutiles. Aucun modèle ou répétition exploitable n'existe pour aider à la cryptanalyse, rendant le texte chiffré théoriquement indéchiffrable comme vu en cours sans connaissance préalable de la clé. Cela suppose, bien sûr, que la clé soit parfaitement aléatoire, jamais réutilisée et gardée secrète.

6. Expliquez votre attaque sur la version améliorée du chiffre de Vigenère.

Voici mon processus, expliqué simplement en étapes :

1. **Trouver la bonne combinaison** :
    - J'ai commencé par tester différentes longueurs de clés Vigenère et des décalages de César sur le texte chiffré. J'ai divisé le texte en blocs de la taille de la clé que je testais et ajusté le décalage de César pour chaque nouveau bloc. Cela simulait comment la clé changeait à chaque fois qu'elle était utilisée sur le texte.
2. **Analyser les résultats** :
    - Pour chaque test, j'ai observé comment le texte déchiffré ressemblait à du français grâce au texte de référence en calculant son indice de coïncidence. Je cherchais quelle combinaison me donnait un texte qui, selon cet indice, ressemblait le plus à un vrai texte français.
3. **Retrouver la clé de Vigenère** :
    - Une fois que j'ai trouvé la meilleure combinaison de longueur de clé et de décalage de César, j'ai utilisé ces informations pour retrouver la clé de Vigenère initiale. J'ai inversé le décalage de César sur la clé modifiée pour obtenir sa version originale.
4. **Déchiffrer le message** :
    - Enfin, avec la clé de Vigenère en main et le décalage de César connu, j'ai pu décrypter le message. J'ai appliqué le déchiffrement de Vigenère et ajusté la clé après chaque utilisation complète selon le décalage de César que j'avais déterminé.

Exemple concret : Si je prends un texte chiffré avec la clé "ABCD" et un décalage de César de 1, après chaque utilisation de "ABCD", la clé devient "BCDE". En reconnaissant et en inversant ce processus, j'ai pu retrouver "ABCD" comme la clé initiale et décrypter le texte.

7. Trouvez une méthode statistique (proche de ce qu’on a vu dans ce labo) permettant de distinguer un texte en anglais d’un texte en français. Qu’en pensez-vous ? Testez votre méthode.

J'ai créer une méthode analyze_text_language pour faire cela, le fonctionnement étant d'utiliser la fréquence des lettres et l'IC du texte à analyser que nous comparons aux valeures de références preétablies pour l'anglais et le français, puis la langue est déduite en identifiant à quelle langue les résultats se rapprochent le plus, tant en termes de fréquences qu'en termes d'IC.

Pour tester la méthode il faut modifier le fichier langueTest.txt.

Cela fonctionne pour des textes tests de taille considérables, mais sur des textes d'une taille trop petite il n'est pas possible de détecter la langue assez précisément en utilisant cette méthode car un petit texte augmente la variabilité statistiques, ce serait aussi le cas si les textes de références étaient trop petits aussi.