

Université Pierre et Marie Curie



UE: Statistique et informatique (LI323)

Projet 2 - Statistique en BioInformatique :

Analyse statistique d'une famille de protéine

Année scolaire : 2013/2014

Professeur chargé de TD/TME :

Martin WEIGT

Etudiants :

Rémi Cadène n°3000693

Joël Fieux-Herrera n°3003174

Sommaire

README	p2
I. Modélisation par PSWM	p3
1. Entropie relative	p3
2. Log-vraisemblance	p4
3. Fraction	p4

Readme

Ce projet Java sous Eclipse contient quatre dossiers :

- bin contient les fichiers compilés,
- src contient les fichiers sources,
- data contient les fichiers annexes donnés en début de projet et générés par notre programme,
- doc contient le sujet et les réponses aux questions.

Afin de compiler et d'exécuter, deux possibilités :

- avec Eclipse :

- ouvrir Eclipse
- sélectionner votre Workspace
- importer (File / Import...)
- sélectionner "General / Existing Projects into Workspace", puis Next
- sélectionner en root directory le dossier contenant notre dossier projet (ex: Téléchargements), puis Finish
- compiler et exécuter (Run)

- avec Terminal :

- Pour compiler :

```
cd <directory>/LI323_P2_CadeneFieux  
javac -cp ./src/BioInfo/*.java -d ./bin/BioInfo/
```

- Pour exécuter :

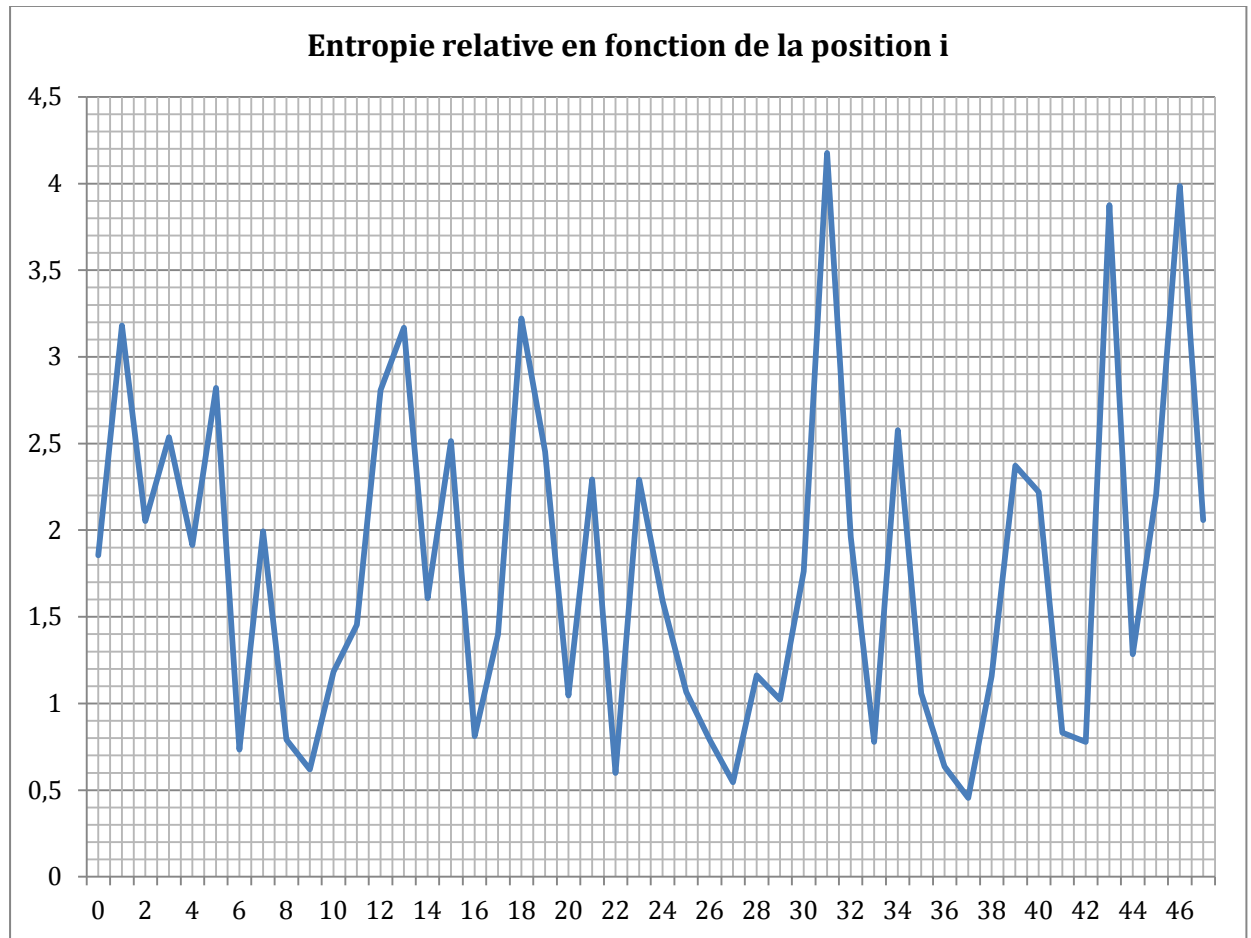
```
java BioInfo
```

I. Modélisation par PSWM

1. Entropie relative

Plus l'entropie relative associée à la colonne est élevée, moins la protéine rencontre de mutation sur celle-ci.

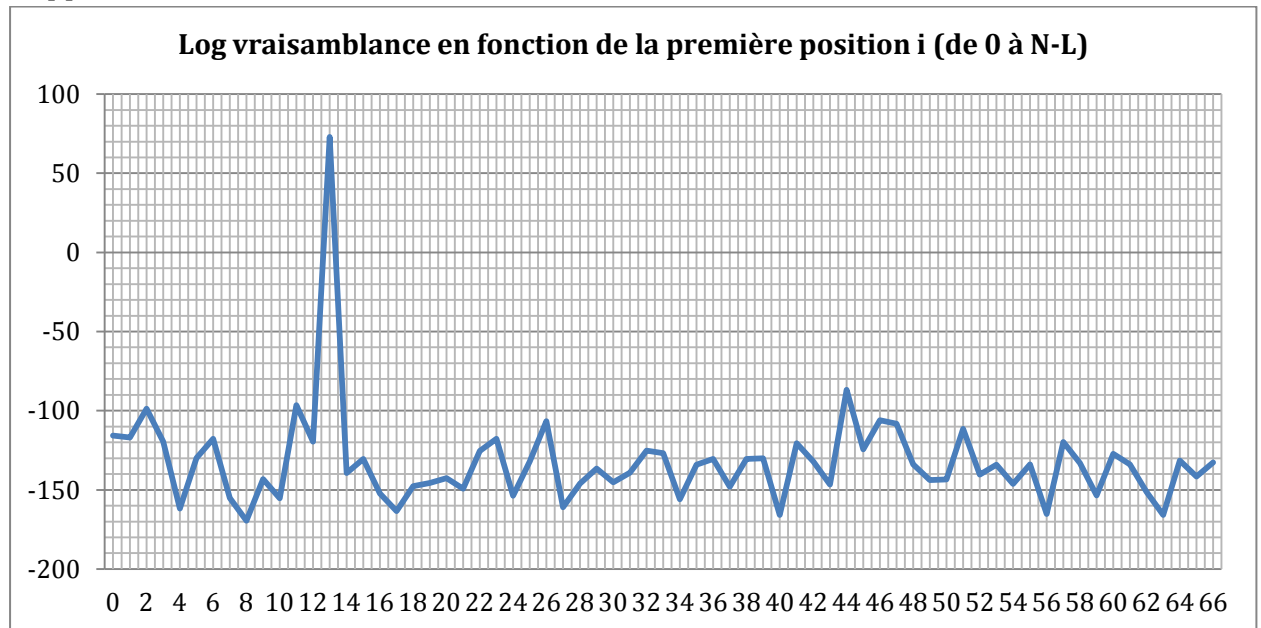
Entropie relative en fonction de la position i



On remarque un maximum en position 31.

2. Log-vraisemblance

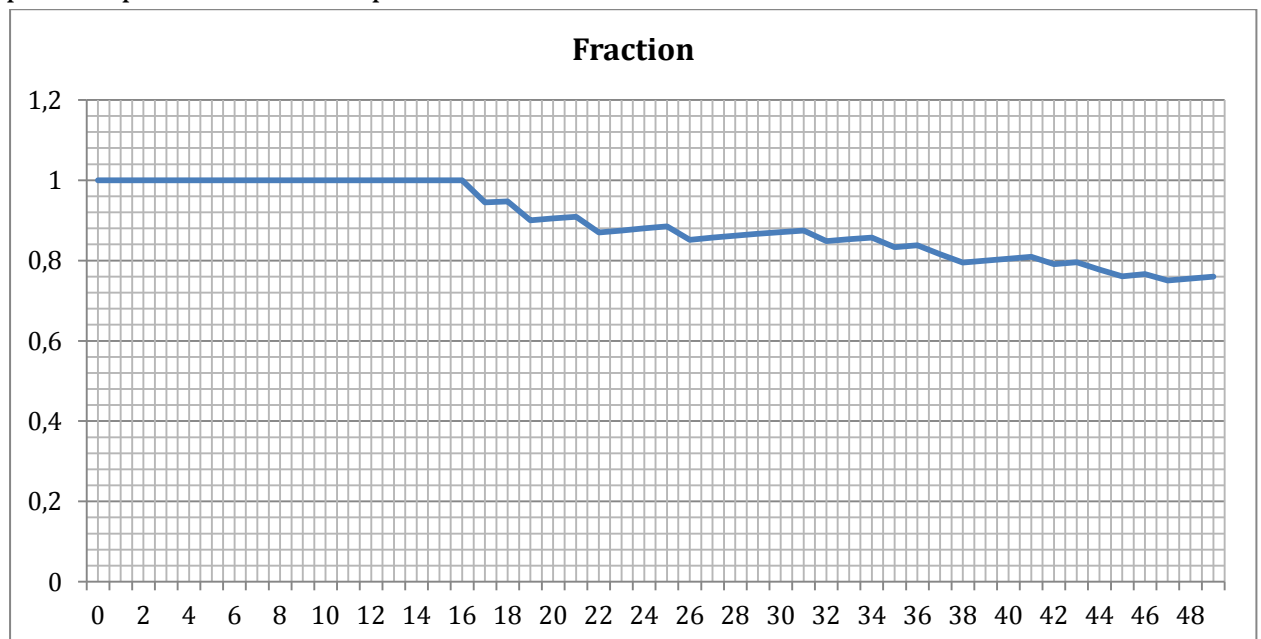
Plus le log vraisemblance de la séquence b est grand, plus celle ci a de chance d'appartenir à la famille.



On remarque donc que la séquence qui commence a partir de la position 13 appartient probablement à la famille.

3. Fraction

Plus la valeur associée au nombre de paires de positions sélectionnées est grande, plus ces paires auront une probabilité élevée d'être en contact.



On remarque que les paires les plus corrélées ont une probabilité élevée d'être en contact.