

Atelier scientifique

Lycée Raymond Naves

Sébastien Déjean

Université Paul Sabatier, Toulouse 3
www.lsp.ups-tlse.fr/Dejean

5 janvier 2006



1 Formation

- Curriculum Vitae
- Expérience professionnelle

2 Métier

- Dans les papiers
- Dans les faits

3 Activités

- Analyse de données transcriptomiques
- Analyse textuelle
- Simulation de processus (multi-)fractionnaire

- **1993** - Baccalauréat série C, mathématiques et sciences physiques
- **1994** - Diplôme d'Études Universitaires Généralisées (1ère année) Mathématiques et Informatique
- **1994-1997** - Institut Universitaire Professionnalisé, *Mathématiques Industrielles, Calcul Scientifique et Statistique* (DEUG, licence, maîtrise)
- **1998** - Diplôme d'Études Approfondies en mathématiques appliquées
- **2002** - Doctorat en mathématiques appliquées

Stages "ingénieur"

- Caisse Régionale d'Assurance Maladie (4 mois) - Analyse du calcul du taux de cotisation des accidents du travail
- Siemens Automotive (10 mois) - Modélisation du comportement d'un moteur à essence en vue de réduire la durée d'essais sur banc-moteur

Recherche

Institut National de la Recherche Agronomique - Comparaison de procédures d'estimation dans le cadre des modèles non linéaires à paramètres aléatoires : application à la modélisation de l'évolution temporelle de l'indice de surface foliaire de cultures observées par télédétection spatiale.

Enseignement

Statistique, informatique : IUT, école d'ingénieurs, formation professionnelle.

Ingénieur de recherche en calcul scientifique

Définition de l'emploi type d'après le Référentiel des emplois-types de la recherche et de l'enseignement supérieur (REFERENS) :

L'ingénieur de recherche en calcul scientifique analyse, dans le cadre de projets de recherche, un problème théorique ou une situation d'expérience et d'observation. Il recherche les méthodes d'analyse, conçoit et optimise les outils permettant le traitement du problème. Il s'assure de la pertinence des résultats obtenus.

Statisticien

*Laboratoire de Statistique et Probabilités
Unité Mixte de Recherche 5583, CNRS-UPS-INSA*



Activités au sein du LSP

- Recherche appliquée en statistique et probabilités
- "Relationnel" scientifique et industriel
- Webmaster : www.lsp.ups-tlse.fr

Statisticien

Laboratoire de Statistique et Probabilités
Unité Mixte de Recherche 5583, CNRS-UPS-INSA



Activités au sein du LSP

- Recherche appliquée en statistique et probabilités
- "Relationnel" scientifique et industriel
- Webmaster : www.lsp.ups-tlse.fr

Recherche appliquée en statistique et probabilités



Trois exemples :

- Analyse de données transcriptomiques
- Analyse textuelle
- Simulation de processus (multi-)fractionnaire

Analyse de données transcriptomiques

En collaboration avec des biologistes de l'Institut National de la Recherche Agronomique (INRA) et de l'Institut National de la Santé Et de la Recherche Médicale (INSERM).

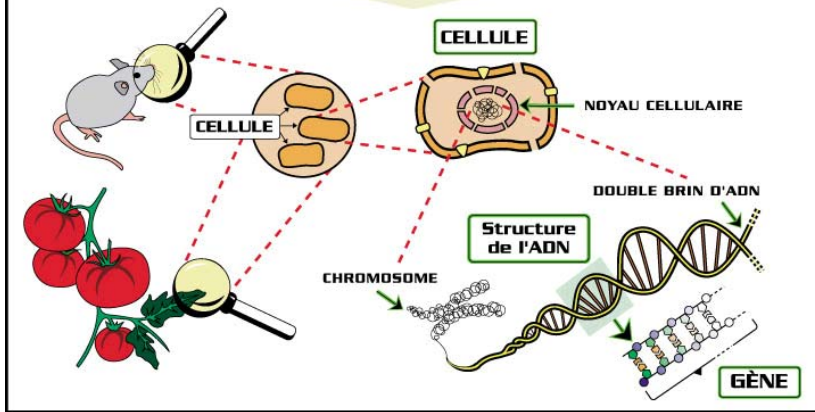
www.lsp.ups-tlse.fr/Biopuces

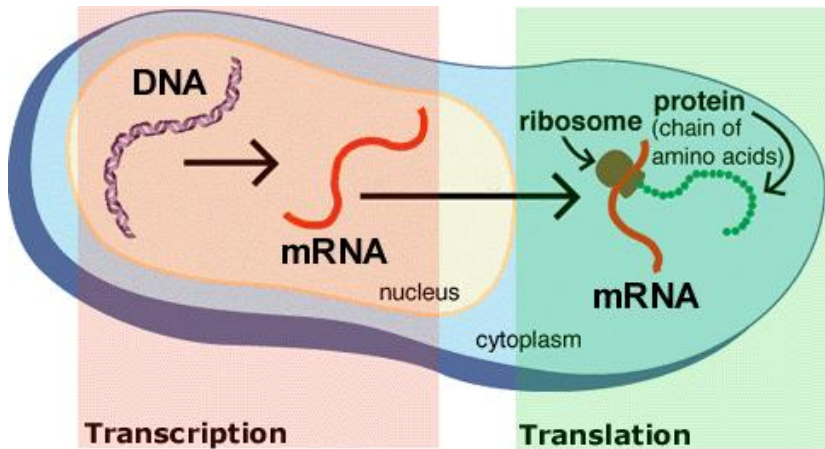
Thématiques biologiques abordées :

- Fermentation de la levure dans différentes conditions de température et au cours du temps
- Étude de nutrition chez la souris
- Analyse transcriptomique du cancer pancréatique humain
- Recherche sur les obésités
- Folliculogénèse chez la truie
- Analyse du transcriptome de *S. meliloti* en culture et en symbiose avec la luzerne
- ...

👉 Point commun : technologie des **biopuces** (*microarrays*), génération d'une quantité importante de données.

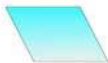
OÙ SE CACHE LE GÈNE ?





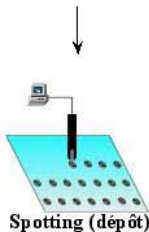
Fabrication des puces à ADN

Lames de verre
recouvertes de polylysine



+

6116 ORFs de levure
amplifiées par PCR



Hybridation

Souche 1



Souche 2



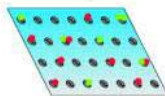
Extraction
des ARN



Cy3

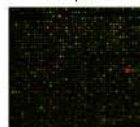
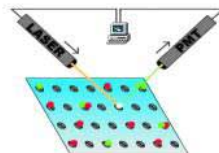
Transcription
des ARNm
en ADNc

Cy5

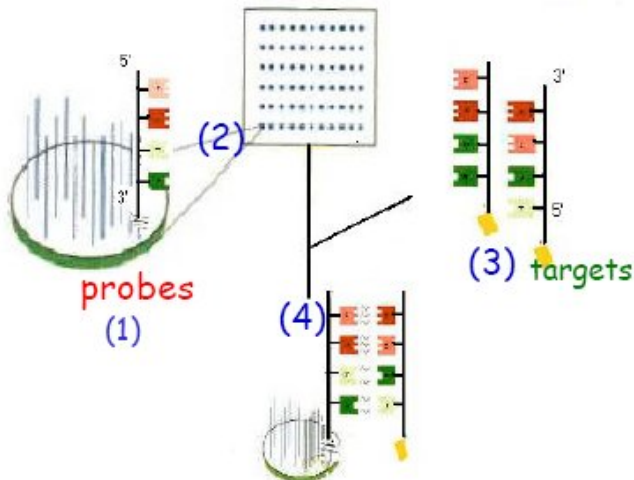


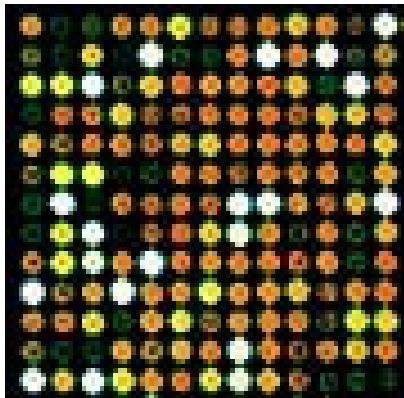
Obtention des résultats

Lecture (scanner)



Analyses des résultats





Démarche

- Discussions avec les biologistes : quel est le problème ?
- Étude préliminaire du jeu de données : analyse exploratoire élémentaire
- Discussions avec les biologistes : ce que nous trouvons est-il interprétable ?
- Nouvelles questions plus précises
- Mise en œuvre de techniques statistiques spécifiques (discrimination, étude de données cinétiques...)

Analyse textuelle

Dans le cadre d'un contrat avec le Centre National d'Études Spatiales (CNES) "Méthode et Outils de Data Mining pour les Projets Spatiaux" en collaboration avec des informaticiens de l'Institut de Recherche en Informatique de Toulouse (IRIT) et des linguistes de l'Université de Toulouse 2 Le Mirail.

Démarche

- Données initiales : corpus de documents (quelques dizaines)
- Lemmatisation des documents
- Construire des données numériques (table de contingence)
- Analyse statistique

👉 **Lemmatisation** : opération consistant à regrouper les formes
occurrentes d'un texte ou d'une liste sous des adresses lexicales.

Maître Corbeau, sur un arbre perché,
Tenait en son bec un fromage.
Maître Renard, par l'odeur alléché,
Lui tint peu près ce langage :
"Hé ! bonjour, Monsieur du Corbeau.
Que vous êtes joli ! que vous me semblez beau !
Sans mentir, si votre ramage
Se rapporte votre plumage,
Vous êtes le Phénix des hôtes de ces bois."
A ces mots le Corbeau ne se sent pas de joie ;
Et pour montrer sa belle voix,
Il ouvre un large bec, laisse tomber sa proie.
Le Renard s'en saisit, et dit : " Mon bon Monsieur,
Apprenez que tout flatteur
Vit aux dépens de celui qui l'écoute :
Cette leçon vaut bien un fromage, sans doute. "
Le Corbeau, honteux et confus,

Jura, mais un peu tard, qu'on ne l'y prendrait plus.

maître corbeau arbre percher
tenir bec fromage
maître renard odeur allécher
tenir peu près langage
bonjour monsieur corbeau
être joli sembler beau
sans mentir ramage
rapporter plumage
être phénix hôte bois
mot corbeau sentir pas joie
montrer beau voix
ouvrir large bec laisser tomber proie
renard saisir dire bon monsieur
apprendre flatteur
vivre dépens écouter
leçon valoir bien fromage sans doute
corbeau honteux confus

jurer peu tard prendre plus

👉 Comptage des occurrences

Mots	Occurrences
maître	2
corbeau	4
arbre	1
percher	1
tenir	2
bec	2
fromage	2
renard	2
odeur	1
allécher	1
...	...

👉 Construction d'un tableau de contingence

Mots	Doc-1	Doc-2	...	Doc-n
balise	25	32	...	8
émission	42	21	...	24
transmettre	12	18	...	62
boitier	5	7	...	21
satellite	64	52	...	15
seconde	13	28	...	14
commande	10	21	...	7
transfert	2	41	...	37
amplification	8	7	...	14
batterie	34	12	...	2
...	...	...	...	...

En pratique : quelques dizaines de colonnes et quelques centaines de lignes.

Analyse statistique

- Analyse Factorielle des Correspondances (AFC)
- Classification (*clustering*)

Résultats

- Mise en évidence de groupes de documents homogènes
- Caractérisation de ces groupes en fonction des mots les plus fréquents
- "Apprentissage" du corpus de document sans le lire

Simulation de processus (multi-)fractionnaire

En collaboration avec des enseignants-chercheurs probabilistes du Laboratoire de Statistique et Probabilités.

www.lsp.ups-tlse.fr/FracSim

Démarche

- *Travaux théoriques*
- Compléments pratiques : simulation
- Programmation (langage C, logiciel statistique R)
- Utilisation d'un super-calculateur (programmation parallèle)
- Visualisation (représentation graphique)

Avertissement

L'image qui va suivre est susceptible de choquer les âmes (in)sensibles... aux mathématiques.

$$Y_{h,n}(t) = \sum_{l=1}^n \left(\frac{\cos(2\pi\theta_l) \left[\cos\left(\frac{R_l}{m}t.U_l\right) - 1 \right] + \sin(2\pi\theta_l) \left[\sin\left(\frac{R_l}{m}t.U_l\right) \right]}{R_l^{h(t)+1/2}} \right)$$

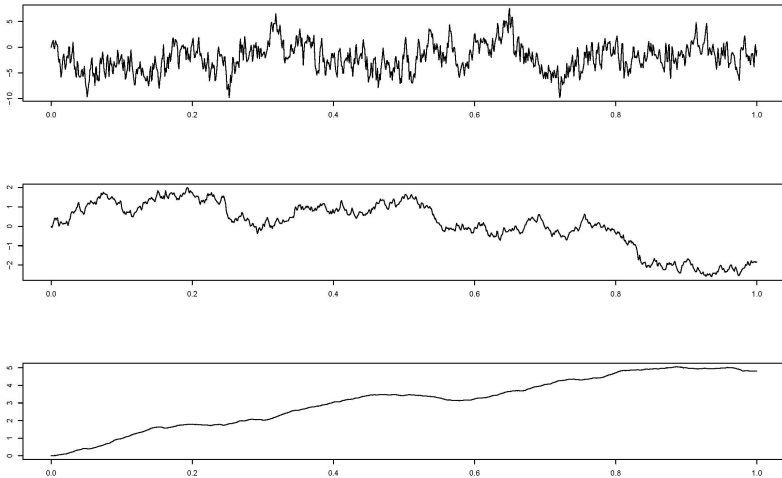
La fonction de régularité $h(t)$ est le paramètre-clé de la simulation.

- $\forall t \in E, 0 < h(t) < 1$
- Si $\forall t \in E, h(t) = h$, alors le processus est **fractionnaire**
- Si $h(t)$ varie quand t varie dans E alors le processus est **multi-fractionnaire**

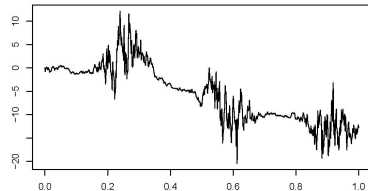
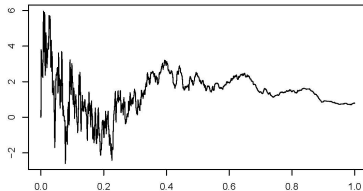
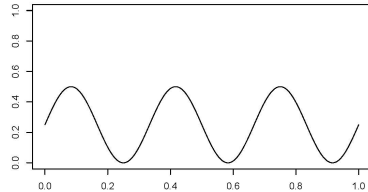
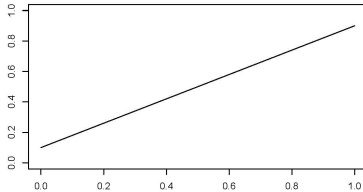
Extrait de programme en langage C

```
...  
for (i=0;i<*k;i++)  
{  
    tmp1 = cos(2*M_PI*theta)*(cos(t[i]*R*U)-1);  
    tmp2 = sin(2*M_PI*theta)*sin(-t[i]*R*U);  
    tmp3 = pow(R,0.5+H[i]);  
    tmp = (tmp1-tmp2)/tmp3;  
    res[i] = res[i] + tmp;  
}  
...
```

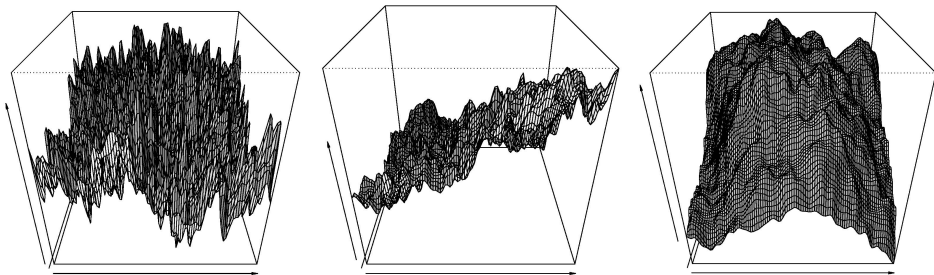
Processus fractionnaire : $h(t) = 0.1 - 0.5 - 0.9$



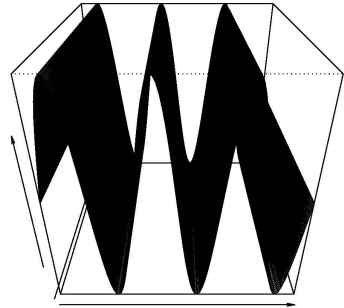
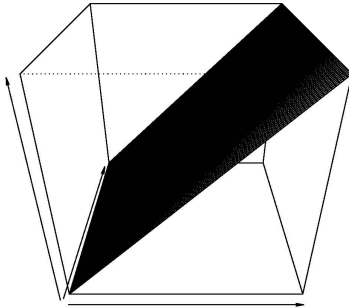
Processus multi-fractionnaire : $h(t)$ croissante - sinusoidale



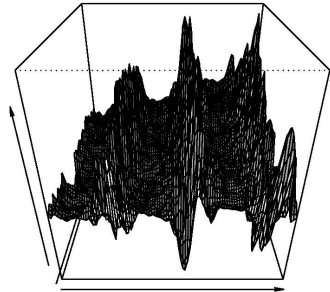
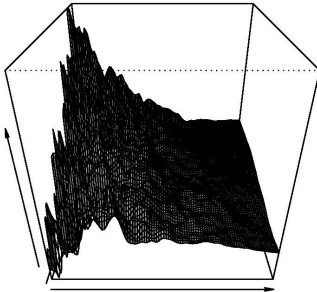
Processus fractionnaire **2D** : $h(t) = 0.1 - 0.5 - 0.9$



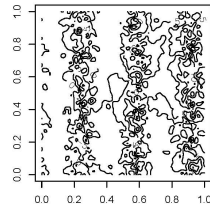
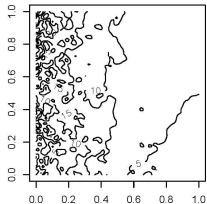
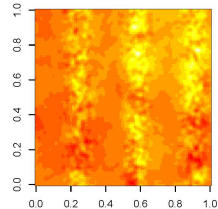
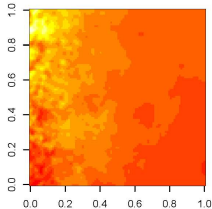
Fonctions de régularité 2D : $h(t)$ croissante - sinusoidale



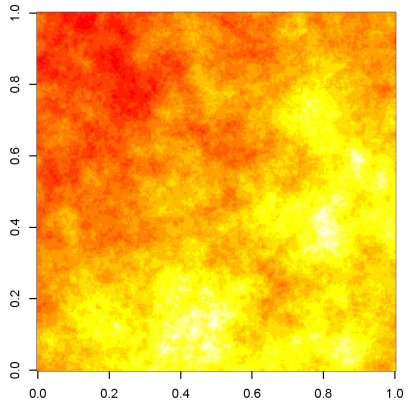
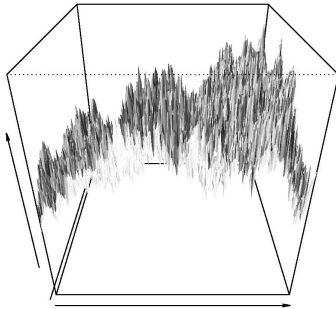
Processus multi-fractionnaire 2D : $h(t)$ croissante - sinusoidale



Autres représentations graphiques



Comparaison d'une surface et d'une image



C'est terminé

Avez-vous des questions ?