

## UE PHYS341 “Statistiques et modélisation”

## Projets de suivi du groupe C1

**Sujet 1. Simulation de l'évolution temporelle de l'épidémie d'une grippe**

Un modèle simple de propagation de maladie infectieuse dans une population de taille finie est donné par :

$$\begin{aligned}\frac{dS}{dt} &= -\alpha IS \\ \frac{dI}{dt} &= \alpha IS - \beta I \\ \frac{dR}{dt} &= \beta I\end{aligned}$$

où  $S, I$  et  $R$  désignent respectivement les proportions d'individus “susceptibles”, “infectieux” ou “guéris” (ces derniers ne sont plus contaminants). Le théorème de Kermack-McKendrick établit qu'une épidémie ne peut se développer que si le nombre initial d'individus susceptibles est supérieur à une valeur seuil  $\beta/\alpha$ . Il est demandé dans ce projet de mettre en oeuvre ce modèle (on justifiera notamment la méthode de résolution et la valeur du pas d'intégration choisies); de vérifier “expérimentalement” le théorème de Kermack-McKendrick; d'ajuster les paramètres du modèle aux données épidémiologiques accessibles dans `~bottin/pub/phys341/influenza.data`.

**Sujet 2. Simulation d'une infection**

Ce sujet reprend l'évolution de l'épidémie de grippe sur un espace de dimension finie. On prendra les trois types d'individus “sains” et “malades” et “guéris” sur un réseau à deux dimensions discret. En laissant évoluer au hasard leur position et en choisissant des règles de contamination par contact avec les plus proches voisins du même type que dans le sujet 1, on étudiera l'influence de la taille du système, ainsi que le seuil de probabilité de contamination sur le développement de l'infection.

**Sujet 3. Attracteur de Lorentz**

Pour étudier la convection, le mathématicien et météorologue Edward Lorentz (1963) propose le modèle suivant :

$$\begin{aligned}\frac{dX}{dt} &= \sigma(Y - X) \\ \frac{dY}{dt} &= -XZ + rX - Y \\ \frac{dZ}{dt} &= +XY - bZ\end{aligned}$$

Il montre que l'on peut obtenir un comportement chaotique avec seulement 3 variables (système non-linéaire à trois degrés de liberté). Le projet consiste en la simulation de ce modèle (RK4), et en la représentation de  $(X, Y, Z)$ , et de  $Z(t)$ . On étudiera ensuite l'attracteur obtenu pour les valeurs initialement utilisées par Lorentz ( $\sigma = 10$ ,  $b = 8/3$  et  $r = 28$ ), et étudiera l'évolution temporelle de deux conditions initiales proches par exemple en notant les passages d'une “aile” à l'autre du papillon. Enfin, on fera varier  $r$  entre 145 et 170 et on décrira les différents régimes observés.

#### Sujet 4. Emergence d'un comportement collectif

Comment expliquer la cohérence des comportements observés dans un vol d'oiseaux ou dans un banc de poisson ? Craig Reynolds a proposé un modèle qui explique l'apparition de tels phénomènes à l'échelle du groupe à partir de règles comportementales individuelles élémentaires (chaque individu réagit à tout moment en fonction de la position et de la vitesse de ses plus proches voisins). Dans ce projet, un programme est fourni (`~bottin/pub/phys341/boids-1.sci`) qui simule le déplacement aléatoire d'individus évoluant indépendamment les uns des autres dans un espace à deux dimensions. Il est demandé de rechercher des règles de déplacements individuel qui permettent de générer un déplacement cohérent du nuage d'individus.

#### Sujet 5. Estimation d'un taux de croissance bactérienne par analyse d'images

La croissance d'une population d'*escherichia coli* a été suivie par vidéomicroscopie. L'objectif du projet est de mettre en oeuvre une méthode simple d'analyse d'images (fichiers accessibles dans `~bottin/pub/phys341/images.tar`) permettant d'estimer le taux de croissance de cette population. Pour cela, on devra tout d'abord estimer sur chaque image la proportion de surface occupée par les bactéries. On tracera l'évolution de cette proportion en fonction du temps. On estimera alors par la méthode des moindres carrés les paramètres d'un modèle simple d'évolution de l'effectif de la population.

#### Sujet 6. Evolution de Population : le système proie prédateur

Le modèle de Lotka-Volterra décrit les interactions entre deux espèces, une proie (les lapins) et un prédateur (les renards). Soit  $R(t)$  la population de renards à un instant donné et  $L(t)$  celle des lapins. Les équations qui dirigent l'évolution de ces populations au cours du temps sont :

$$\begin{aligned}\frac{dL(t)}{dt} &= +k_L L(t) - AR(t)L(t) \\ \frac{dR(t)}{dt} &= -k_R R(t) + BR(t)L(t)\end{aligned}$$

On simulera l'évolution de ces populations, puis on représentera et discutera cette évolution en variant les paramètres du problème.

#### Sujet 7. Synchronisation d'oscillateurs

Le modèle de Kuramoto qu'il est proposé d'étudier est défini par un ensemble de  $N$  oscillateurs. L'état de chaque oscillateur  $i$  est défini par sa phase  $\phi_i$ . Ces oscillateurs évoluent suivant l'ensemble d'équations suivantes :

$$\frac{d\phi_i}{dt} = \omega_i - \frac{K_0}{N} \sum_{j=1}^N \sin(\phi_i - \phi_j)$$

où les  $\omega_i$  sont distribués aléatoirement. Chaque oscillateur évolue donc avec sa fréquence propre  $\omega_i$  et est couplé aux autres via le terme en sinus. Lorsque le couplage  $K_0$  est faible, les oscillateurs suivent leur fréquence naturelle  $\omega_i$ . Lorsque  $K_0$  devient suffisamment grand, les oscillateurs se mettent en phase. On simulera l'évolution de ces oscillateurs, puis on représentera et discutera cette évolution en fonction des paramètres du problème.

## Evolution des températures mensuelles à Paris

Fichier de données = moyenne mensuelle des températures journalières maximales et minimales relevées à Paris (station météo du Parc, 14ème) : *TempMensMax.txt*, *TempMensMin.txt*

Tracer les moyennes mensuelles des températures journalières maximum et minimum données entre 1873 et 2003 à Paris. Examiner leur distribution. Conclure à un traitement de données approprié. Déterminer si cette température augmente, diminue ou reste plutôt stable.

## Lien entre cycle de progestérone et fécondité

Fichier de données = *PDG.txt*

À partir des données du taux de progestérone au cours du cycle d'ovulation d'une population de femmes, déterminer s'il peut exister un lien entre cette hormone et la fécondité.

## Morphologie de crânes égyptiens

Fichier de données = *CranesEgyptiens.txt*

À partir des mesures de 150 crânes égyptiens appartenant à 5 époques différentes, étudiez l'évolution morphologique de ces crânes dans le temps.

## La marche de l'homme ivre

On étudie le déplacement latéral d'un ivrogne qui marche dans un grand espace découvert. On suppose qu'il est toujours tourné dans la même direction. Pour chaque pas en avant effectué, l'homme dévie d'une petite distance  $d$  constante, vers la gauche avec la probabilité  $p$  (où  $0 < p < 1$ ), vers la droite avec la probabilité  $(1-p)$ . Tracer le parcours de l'homme pour un nombre de pas  $n$  variable et pour différentes valeurs de  $p$ . En générant un grand nombre de parcours, étudier la statistique simple de la position de l'ivrogne sur l'axe latéral. Relier le résultat à une distribution mathématique connue.

## Morphologie du corps humain

Fichier de données = *Body.txt*

À partir des données récoltées lors d'une enquête de grande envergure, étudier les caractéristiques morphologiques du corps humain. On s'intéressera notamment à l'influence du sexe et de l'âge sur ces caractéristiques. Déterminer quelles sont les caractéristiques significativement différentes pour les deux populations hommes et femmes.

## Recrutement militaire aux Etats-Unis en 1970

Fichier de données = *USDraftlottery.txt*

Pour recruter les nouvelles troupes pendant la guerre du Viet-Nam, le Sénat des Etats-Unis avait recours à un système de tirage au sort portant sur la date (jour et mois) de naissance, garantissant en théorie la parfaite égalité des chances pour tous les mobilisables. Mais en 1970, des plaintes ont été déposées, arguant que ceux nés en début d'année avaient été plus recrutés que les autres. À partir des données du tirage au sort, déterminer si ces plaintes sont avérées. On pourra par exemple comparer ces données avec celles obtenues par un grand nombre de tirages au sort numériques. On utilisera également les données du tirage au sort de l'année 1971.

## Température globale et CO2

Fichier de données = *GMT1.txt, GMT2.txt*

Etudier les données de température globale de la Terre de 1866 à 1996. Compléter ce fichier jusqu'à 2006 en faisant une recherche sur Internet. Comparer la variabilité naturelle présente dans ces données avec le récent réchauffement, et conclure.

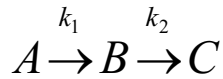
D'autre part, étudier la température globale des 123.000 dernières années *versus* la teneur en CO2 de l'atmosphère. Conclure.

## UE PHYS341 « STATISTIQUES ET MODELISATION »

## Projets de suivi du groupe C1

**1. Simulation d'une réaction en chaîne**

L'objectif de ce projet est d'observer l'évolution des concentrations des réactifs et des produits d'une réaction en chaîne. Nous considérons la chaîne de réaction suivante :



où on fait l'hypothèse de réactions du premier ordre. On envisagera la notion d'étape cinétiquement déterminante.

**2. Statistique d'un cas linéarisable**

On considère l'exemple du lissage des courbes isothermes d'absorption. Dans une solution aqueuse contenant un polluant, on plonge un solide adsorbant (charbon actif sous forme de tissu) qui capture une partie des molécules de la substance polluante. Chaque point d'équilibre est caractérisé par la concentration à l'équilibre et la quantité de polluant adsorbé par unité de masse du charbon actif. Plusieurs modèles sont proposés pour représenter ces isothermes. Ce projet vise à comparer par des critères statistiques les résultats de ces modèles aux données expérimentales.

**3. Modélisation d'une période glaciaire quaternaire**

Le modèle paléoclimatique de Milankovitch, essentiellement basé sur l'origine astronomique de la variabilité de l'insolation solaire de la surface du globe terrestre, décrivant assez fidèlement les changements climatiques passés des dernières périodes glaciaires (cycles interglaciaire glaciaire) du Quaternaire obéit au système d'équations différentielles ordinaires couplées :

$$(1) \quad \dot{x}(t) = -a.x(t), \forall t > 0, x(0) = x_0 > 0, \text{ fini et connu}$$

$$(2) \quad d.\dot{u}(t) = a.x(t) - \beta.u(t) + c.z(t), \forall t > 0, u(0) = 0$$

$$(3) \quad \dot{z}(t) = \beta.u(t) - c.z(t), \forall t > 0, z(0) = 0$$

Le projet vise à obtenir  $u(t)$  afin de le comparer aux mesures effectuées par les différents carottages effectués à Vostok pour la période du Riss/Würm-Würm (env. -135 ma à -20 ma).

#### 4. Comportement d'un oscillateur amorti

On étudiera l'influence de l'amortissement  $h(x)$  au sein de l'équation dynamique d'un oscillateur :

$$m \ddot{x}(t) + h(x) \dot{x}(t) + kx(t) = 0$$

On s'attachera particulièrement aux tracés dans le plan de phase  $(x(t), \frac{dx}{dt})$ .

#### 5. Méthode de Monte-Carlo : écart à l'horizontale

Il est possible de mesurer les propriétés optiques des particules en suspension dans l'air à partir d'un lidar (« light detection and ranging »), qui fonctionne sur le même principe qu'un radar mais pour des longueurs d'onde allant de l'UV à l'IR. Une simple analyse du signal rétrodiffusé lors de tirs horizontaux permet d'accéder au coefficient d'extinction des aérosols, à condition que l'atmosphère soit homogène sur un plan horizontal. Ce projet a pour objectif d'étudier l'influence d'un écart à l'horizontale, suivant une loi gaussienne de probabilité, sur le calcul du coefficient d'extinction.

#### 6. Méthode de Monte Carlo : influence du rapport signal à bruit

On reprend ici le tir lidar horizontal du projet 5 pour mesurer la concentration de particules dans un panache de cheminée. Cette fois, on néglige les écarts d'horizontalité. En revanche, la théorie du signal fait apparaître un bruit de type Poisson sur le nombre de photons rétrodiffusés. Un modèle direct-inverse sera développé pour étudier les erreurs engendrées.

#### 7. Approximation lognormale d'une densité de répartition discrète

Les granulomètres sont des instruments permettant de mesurer des concentrations de particules dans l'air. Ils fournissent des distributions en taille, i.e. des nombres de particules (supposées sphériques) par  $\text{cm}^3$  ayant un rayon compris entre  $r$  et  $r+dr$ . On tentera d'approcher au mieux ces densités de répartition (discrètes) par des modèles couramment utilisés afin d'obtenir des fonctions continues sur tout le spectre des rayons de particules. Le modèle le plus connu est la somme de distributions lognormales.

## UE PHYS341 - “Statistiques et modélisation”

**Projets de suivi des groupes 2a et 2b****1. Simulation de l'évolution d'une succession végétale par chaîne de Markov**

Dans les chaînes de Markov, l'évolution du système est discrète, le passage d'un état à un autre se fait selon une probabilité donnée et fixe. La succession d'une série d'états correspond à une trajectoire du système. Soit on passe d'un état à l'autre soit on reste au même état. La probabilité du passage d'un état à un autre est appelée probabilité de transition. L'ensemble des probabilités de transition constitue la matrice de transition. Par calcul vectoriel, on peut déterminer l'état d'une population à un instant  $t$ , à partir de son état initial et de la matrice de transition par l'équation :

$$\mathbf{V}_t = \mathbf{V}_0 \mathbf{P}^t \quad (1.1)$$

$\mathbf{V}_0$  est le vecteur représentant la distribution initiale en probabilité, et  $\mathbf{V}_t$  le vecteur représentant la distribution à l'instant  $t$  en probabilité. Nous étudierons ici avec cette méthode une succession végétale typique du midi de la France.

**2. Modélisation en dynamique des populations d'un système proie-prédateur***2.1. Modèle de Lotka et Volterra*

L'évolution des populations de proies  $x$  et de prédateurs  $y$  selon le modèle Lotka et Volterra est définie par le système d'équations différentielles suivants :

$$\frac{dx}{dt} = rx - gxy \quad (2.1)$$

$$\frac{dy}{dt} = egxy - my \quad (2.2)$$

où  $r$  est le taux d'accroissement de la population de proies et  $g$  le taux d'efficacité de la prédation. Dans ce cas, la population de proies croît de manière exponentielle en l'absence de prédateurs, la présence de prédateur est le seul facteur limitant la croissance des proies, la nourriture des proies est supposée non limitante.  $e$  est un coefficient de conversion spécifiant le nombre de jeunes prédateurs générés par proies capturées.  $m$  est le taux de mortalité des prédateurs, l'idée sous jacente est que sans proies la population de prédateurs décroît. On est dans le cas d'une prédation exclusive, ou le prédateur ne se nourrit que d'une seule proie. Dans ce projet nous allons simuler l'évolution de la population de proies et de prédateur en discrétisant le système.

*2.2. Extension du modèle logistique*

Nous avons traité au TP 1 l'équation logistique pour modéliser la démographie d'une population. Dans ce projet nous étendrons ce modèle pour considérer un système proie-prédateur. Les équations à étudier seront :

$$\frac{dx}{dt} = r_x x(1 - x/K_x) - \mu_x \quad (2.3)$$

$$\frac{dy}{dt} = r_y y(1 - \mu_x y/x) \quad (2.4)$$

où  $r_x$  et  $r_y$  sont, respectivement, les taux de reproduction intrinsèques de chaque espèce et  $\mu_x$  représente le nombre de  $x$  mangés par  $y$  au cours d'un pas de temps.

### 3. Largeur de bande de résonance d'un oscillateur amorti

Un oscillateur amorti forcé est un modèle utilisé dans beaucoup d'applications en physique (en mécanique le système masse-ressort-amortisseur où en électronique un circuit RCL). Nous allons utiliser ce modèle pour prédire la largeur de bande de résonance d'un système d'extraction de l'énergie de la houle par colonne d'eau oscillante.

### 4. Filtrage d'un signal ou d'une mesure expérimentale

Dans plusieurs domaines de la physique, une étape souvent essentielle dans l'observation et analyse des données expérimentales ou des signaux mesurés consiste à s'affranchir de signaux parasites et des valeurs aberrantes. Dans ce projet nous implémenterons différents types de filtres pour les appliquer à des données océanographiques.

**Moyenne glissante.** Un filtre par moyenne glissante substitue à une valeur la moyenne des valeurs aux alentours. Plus l'intervalle est grand, plus le filtrage est efficace.

**Médiane.** Un filtre par moyenne médiane substitue à une valeur la médiane des valeurs aux alentours. Ce filtrage est efficace pour gommer les valeurs aberrantes.

**Filtrage en fréquence.** L'analyse par Transformé de Fourier montre que toute série de données, dépendant de quelque paramètre que ce soit (temps, variable d'espace, autre variable), est décrite par ses composantes fréquentielles.

### 5. Algorithmes de Vélométrie par Images de Particules (PIV)

Le principe général de cette méthode consiste à prendre deux photos des particules qui tracent l'évolution d'un écoulement et, connaissant l'écart temporel qui les sépare ( $\Delta t$ ), obtenir un champ de vecteurs vitesse en recherchant les pics de la fonction de corrélation. Dans ce projet nous utiliserons la fonction `fft()` de Scilab pour calculer la fonction de corrélation croisée.

### 6. L'attracteur de Lorenz

Edward N. Lorenz., météorologue américain, découvre en 1963 que l'on peut obtenir un comportement chaotique avec seulement trois variables, soit un système non linéaire à trois degrés de liberté. Il montre donc qu'une dynamique très complexe peut apparaître dans un système formellement très simple. Nous allons à présent étudier le système différentiel suivant, dit de Lorenz :

$$\frac{dx}{dt} = \sigma(y - x) \quad (6.1)$$

$$\frac{dy}{dt} = \rho x - y - xz \quad (6.2)$$



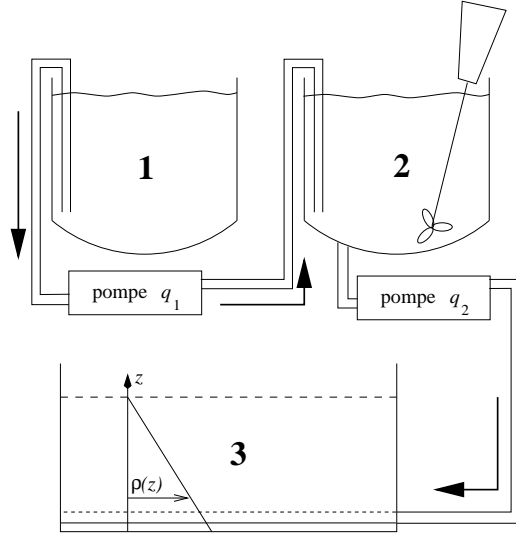


FIG. 1. Schéma d'un système utilisé pour établir un gradient de densité. (1) Réservoir d'eau salée, (2) réservoir où on mélange progressivement l'eau initialement claire avec l'eau salée et (3) cuve expérimentale stratifiée dont le profil de densité est  $\bar{\rho}(z)$ .

$$\frac{dz}{dt} = -\beta z + xy \quad (6.3)$$

## 7. Simulation d'une réaction en chaîne

L'objectif de ce projet est d'observer l'évolution des concentrations des réactifs et des produits d'une réaction en chaîne. Nous considérons donc les réactions décrites par les équations de cinétique chimique suivantes :

$$\left[\frac{dA}{dt}\right] = -k_1[A] \quad (7.1)$$

$$\left[\frac{dB}{dt}\right] = k_1[A] - k_2[B] \quad (7.2)$$

$$\left[\frac{dC}{dt}\right] = k_2[B] \quad (7.3)$$

## 8. Remplissage d'un bassin d'eau salée stratifiée

Dans l'étude en laboratoire des problèmes de dynamique des fluides géophysiques il est souvent désirable de remplir un bassin avec un fluide stratifié, c'est à dire un fluide dont la densité varie avec la hauteur (comme c'est le cas des océans et de l'atmosphère). La figure 1 montre un schéma d'un système à deux réservoirs pour établir un gradient de densité dans une cuve expérimentale remplie d'eau salée. Dans ce projet nous étudierons les équations qui modélisent le système de reservoirs et proposerons un algorithme qui pourra être utilisé par un expérimentateur désirant savoir combien de sel et quels débits sont nécessaires pour produire un profil de densité quelconque.

**LP341 Statistiques, informatique et modélisation**

## Projets de suivi du groupe 2a

**Sujet 1. Description statistique d'une dynamique chaotique**

On s'intéresse aux régimes dits *chaotiques* observés pour certaines valeurs du paramètre de contrôle du système dynamique à temps discret défini par l'itération

$$x_{t+1} = f_r(x_t)$$

de l'application logistique  $f_r$  :

$$f_r : [0, 1] \rightarrow [0, 1], \\ x \mapsto r x (1 - x).$$

On s'attachera à caractériser quantitativement ces régimes de chaos temporel par les outils probabilistes adaptés.

**Sujet 2. Modélisation de l'activité électrique d'une fibre musculaire**

Dans certaines conditions expérimentales, on observe que le potentiel électrique de certaines fibres musculaires peut *osciller* en fonction du temps. On se propose d'étudier numériquement la dynamique d'un modèle permettant de reproduire et comprendre l'origine de ces oscillations (Morris et Lecar, 1981). Les variables dynamiques considérées sont le potentiel  $V(t)$  et la proportion de canaux transmembranaires ouverts  $w(t)$ . Elles obéissent aux équations suivantes :

$$\begin{cases} C \frac{dV}{dt} = -g_{Ca} m_\infty (V(t) - V_{Ca}) - g_K w(t) (V(t) - V_K) - g_f (V(t) - V_f) + I_{App}, \\ \frac{dw}{dt} = -\frac{\phi}{\tau} (w(t) - w_\infty); \end{cases}$$

où l'on suppose que les fractions de canaux ouverts  $m_\infty$ ,  $w_\infty$ , ainsi que le temps caractéristique  $\tau$  dépendent du potentiel selon :

$$\begin{aligned} m_\infty &= 0.5 \{1 + \tanh((V - v_1)/v_2)\}, \\ w_\infty &= 0.5 \{1 + \tanh((V - v_3)/v_4)\}, \\ \tau &= 1 / \cosh((V - v_3)/(2v_4)). \end{aligned}$$

**Sujet 3. Analyse statistique des inégalités de revenus dans le monde**

Le coefficient de Gini est un paramètre de dispersion qui quantifie les inégalités de revenus. On étudiera par exemple la variabilité et l'évolution de cet indicateur en s'appuyant sur les données disponibles sur le site :

<http://www.wider.unu.edu/wiid/wiid.htm>

**Sujet 4. Oscillateur de van der Pol**

On caractérisera les oscillations auto-entretenues de l'oscillateur de van der Pol :

$$\frac{d^2x}{dt^2} + \mu (x^2 - 1) \frac{dx}{dt} + \omega_0^2 x = 0,$$

modèle de circuit électronique comportant un composant actif.

**Sujet 5. Caractérisation de nombres aléatoires générés numériquement**

On discutera des «bonnes» propriétés statistiques attendues d'un générateur de nombres aléatoires. Est-il par exemple possible de quantifier le caractère aléatoire des nombres produits par le générateur du logiciel Scilab ?

**Sujet 6. Modélisation d'un signal nerveux**

Un modèle simplifié décrivant l'électrodynamique de la membrane de cellules nerveuses, dit de FitzHugh-Nagumo (1951), est le suivant :

$$\begin{cases} \frac{dv}{dt} = v(v - \beta)(\delta - v) - Cw + I_{\text{App}}, \\ \frac{dw}{dt} = \epsilon(v - \gamma w). \end{cases} \quad (1)$$

On reproduira numériquement l'apparition de trains d'onde dans ce système, en s'appuyant sur une analyse des équations de FitzHugh-Nagumo dans l'espace des phases.

**Sujet 7. Statistiques de vitesse d'un écoulement turbulent**

On analysera à l'aide d'outils statistiques les propriétés d'un signal de vitesse mesuré au sein d'un écoulement hydrodynamique turbulent.

**LP341 Statistiques, informatique et modélisation**

## Projets de suivi du groupe 2b

**Sujet 1. Dynamique d'un oscillateur nonlinéaire forcé**

Un exemple de système mécanique simple susceptible d'une dynamique complexe est l'oscillateur de Duffing forcé :

$$m \frac{d^2x}{dt^2} = -\lambda \frac{dx}{dt} - \frac{\partial \mathcal{V}}{\partial x} + f \sin(\omega t)$$

où  $\lambda$  est une constante d'amortissement positive et où le potentiel  $\mathcal{V}(x)$  est donné par :

$$\mathcal{V}(x) = \frac{1}{2}ax^2 + \frac{1}{4}x^4.$$

Que pouvez vous dire du comportement dynamique de cet oscillateur en fonction des valeurs adoptées par les paramètres de contrôle ?

**Sujet 2. Statistiques de l'énergie dissipée par un écoulement turbulent**

Un écoulement turbulent est généralement forcé (injection d'énergie) à grande échelle, et dissipe l'énergie à petite échelle par des mécanismes liés au caractère visqueux du fluide. À partir de la connaissance d'une composante du champ de vitesse  $v_x$  – relativement aisément accessible à la mesure – il est possible d'estimer l'énergie dissipée à une échelle  $r$  selon :

$$\epsilon_r(x) = \frac{15\nu}{r} \int_x^{x+r} \left( \frac{dv_x}{dx'} \right)^2 dx',$$

où  $\nu$  est la viscosité cinématique du fluide. À partir de données expérimentales, on caractérisera la statistique de  $\epsilon_r$  en fonction de l'échelle  $r$ .

**Sujet 3. Bifurcations d'un système dynamique déterministe**

On déterminera numériquement le diagramme de bifurcation du système dynamique à temps discret défini par l'itération

$$x_{t+1} = f_r(x_t)$$

de l'application logistique  $f_r$  définie par

$$f_r : \begin{array}{l} [0, 1] \rightarrow [0, 1], \\ x \mapsto r x (1 - x), \end{array}$$

où le paramètre  $r$  est choisi tel que  $0 \leq r \leq 4$ . On pourra s'attacher à caractériser plus précisément certaines des bifurcations, ou encore la dynamique observée dans certains domaines du paramètre de contrôle  $r$ .

### Sujet 4. Dynamique du cycle cellulaire

Un modèle minimal du cycle cellulaire a été proposé en 1991 par Goldbeter. Le schéma biochimique retenu fait intervenir trois protéines, la cycline B, la kinase Cdc2, et une protéase de la cycline, dont les concentrations sont notées  $c$ ,  $m$  et  $x$  respectivement. Le système d'équations différentielles correspondant s'écrit :

$$\begin{cases} \frac{dc}{dt} &= \nu_i - \nu_d x \frac{c}{K_d + c} - k_d c, \\ \frac{dm}{dt} &= V_1 \frac{1-m}{K_1 + 1-m} - V_2 \frac{m}{K_2 + m}, \\ \frac{dx}{dt} &= V_3 \frac{1-x}{K_3 + 1-x} - V_4 \frac{x}{K_4 + x}, \end{cases}$$

avec  $V_1 = V_{M1} \frac{c}{K_c + c}$  et  $V_3 = V_{M3} m$ . On pourra montrer numériquement que les oscillations observées dans ce système correspondent à un cycle limite.

### Sujet 5. Produits de variables aléatoires

On considère le produit  $\Pi$  de  $n$  variables aléatoires  $\pi_1, \dots, \pi_n$  indépendantes et identiquement distribuées (par exemple par une loi uniforme ou Gaussienne). On étudiera les propriétés statistiques de  $\Pi$  (distribution, moments, ...), en particulier pour les grandes valeurs de  $n$ .

### Sujet 6. Modélisation d'un signal nerveux

Le premier modèle ayant décrit la propagation d'un signal nerveux le long d'un axone (en l'occurrence l'axone géant du calmar) est celui de Hodgkin et Huxley (1952). Le potentiel de membrane  $V(t)$  obéit au système d'équations suivant :

$$\begin{cases} C \frac{dV}{dt} &= -g_{Na} m^3 h (V - V_{Na}) - g_K n^4 (V - V_K) - g_{fuite} m^3 (V - V_{fuite}) + I_{App}, \\ \frac{dm}{dt} &= -\frac{1}{\tau_m} (m - m_\infty), \\ \frac{dh}{dt} &= -\frac{1}{\tau_h} (h - h_\infty), \\ \frac{dn}{dt} &= -\frac{1}{\tau_n} (n - n_\infty), \end{cases}$$

où les variables  $m(t)$ ,  $h(t)$  et  $n(t)$  contrôlent l'activité des canaux ioniques transmembranaires impliqués dans ce mécanisme. On pourra s'attacher dans un premier temps à reproduire numériquement la formation du potentiel d'action.

### Sujet 7. Analyse statistique de données démographiques

On se propose d'analyser certaines des données disponibles sur le site de l'INSEE, décrivant la situation démographique en 2004 :

[http://www.insee.fr/fr/ppp/ir/accueil.asp?page=SD2004/dd/sd2004\\_pop.htm](http://www.insee.fr/fr/ppp/ir/accueil.asp?page=SD2004/dd/sd2004_pop.htm)

[http://www.insee.fr/fr/ppp/ir/accueil.asp?page=SD2004/dd/SD2004\\_FECONDITE.htm](http://www.insee.fr/fr/ppp/ir/accueil.asp?page=SD2004/dd/SD2004_FECONDITE.htm)

### **Sondages et élections :**

Etudier les résultats d'une élection et les sondages réalisés à son sujet (par exemple le référendum 2005).

Références :

*Théorie des probabilités et scrutins étatiques, Pierre Lemieux, Figaro-économie- 4 sept. 97pXI.*  
<http://www.pierrelemieux.org/artvote.html>

Banque des sondages Ipsos : <http://www.ipsos.fr/CanalIpsos/poll/8067.asp>

Ministère de l'Intérieur : <http://www.interieur.gouv.fr/avotreservice/elections/rf2005/000/000.html>

→ *L'article présente une analyse sur le rôle d'un individu dans une élection. Il y a pas mal de choses de statistiques à décortiquer. Je proposerai d'étendre cette analyse à une étude du référendum et des sondages préliminaires (théorème de Bienaymé-Tchébichev, etc.).*

### **Analyse démographique**

Présenter une étude des caractéristiques de la population française à l'aide des outils statistiques (pyramide des âges, fonction de répartition, etc.). Comparer celle-ci avec la population d'un autre pays de votre choix.

Références :

Institut National d'études démographiques :

<http://www.ined.fr/population-en-chiffres/indexF.html>

Base de données internationale du bureau de recensement américain :

<http://www.census.gov/ipc/www/idbnew.html>



→ *L'idée est de leur faire manipuler les données en vue d'une analyse. Construction d'une pyramide des âges (histogramme), fonction de répartition. Compréhension des différentes caractéristiques de la pyramide (distribution/répartition). Enfin, ce serait bien s'ils pouvaient comparer la population française avec une autre tant qu'à faire différente (à eux de la choisir).*

### **Etude de l'emploi dans les différents pays européens:**

Présenter les caractéristiques de l'emploi dans les différents pays européens. Essayer de trouver si le taux de chômage est corrélé à d'autres variables (ex : proportion des femmes actives, revenu de l'activité agricole, formations tout au long de la vie...) ou des tendances.

Références :

Institut national de la statistique et des études économiques

<http://www.insee.fr>

Office statistique des communautés européennes – Eurostat

<http://europa.eu.int/comm/eurostat>

→ *L'idée est de leur faire chercher/choisir les données eux-même, en utilisant les outils statistiques. Il y a sur les sites fournis plein d'info sur tous les pays européens.*

### **Chocolat et football :**

Dans le cadre d'une opération marketing, une fabrique de chocolat décide de placer des photos des onze principaux joueurs de l'équipe nationale de football dans l'emballage de ces plaques de chocolat. Elle dispose les images aléatoirement dans les plaques. Combien de plaques en moyenne doit acheter un fan pour reconstituer l'équipe au complet ?



Résultat : 33,2 à obtenir par simulation.

→ Il s'agit de leur faire trouver cela par simulation. Ce n'est pas évident à trouver comme cela. Cela souligne l'importance et l'utilité des simulations. Peut-être faut-il les laisser un peu chercher(mais pas trop) comment solutionner cela.

**Les anniversaires :**

On suppose que la probabilité que l'anniversaire d'une personne tombe un mois donné est égale à  $1/12$ . Six amis vont dîner au restaurant. Quelle est la probabilité que les anniversaires tombent tous dans deux mois différents ? Retrouver le résultat par simulation.



Résultat : 4092 cas favorables ;  $p=4092/12^6 = 0.00137$

→ Là, c'est plus facile à trouver. Comme l'autre, ce serait bien qu'ils arrivent à simuler le problème.