ÉPREUVE COMMUNE DE TIPE 2008 - Partie D

Processus stochastiques

Temps de préparation : ……..…2 h 15 minutes
Temps de présentation devant le jury : …….10 minutes
Entretien avec le jury : ……..10 minutes

GUIDE POUR LE CANDIDAT :

Le dossier ci-joint comporte au total : *12* pages

Document principal : *10 pages*
Documents complémentaires : *2 pages*

Travail **suggéré** au candidat :

Après avoir lu et compris l'ensemble du document le candidat pourra faire un résumé en essayant d'utiliser ses propres connaissances. Il pourra également commenter les simulations présentées en annexe 1. Un glossaire des notions élémentaires de probabilité est en annexe 2.

CONSEILS GENERAUX POUR LA PREPARATION DE L'EPREUVE :

* Lisez le dossier en entier dans un temps raisonnable.
* Réservez du temps pour préparer l'exposé devant le jury.

- Vous pouvez écrire sur le présent dossier, le surligner, le découper … mais tout sera à remettre au jury en fin d'oral.
- En fin de préparation, rassemblez et ordonnez soigneusement TOUS les documents (transparents, *etc.*) dont vous comptez vous servir pendant l'oral, ainsi que le dossier, les transparents et les brouillons utilisés pendant la préparation. En entrant dans la salle d'oral, vous devez être prêts à débuter votre exposé.

A la fin de l'oral, vous devez remettre au jury le présent dossier, les transparents et les brouillons utilisés pour cette partie de l'oral, ainsi que TOUS les transparents et autres documents présentés pendant votre prestation.

Processus Stochastiques

1 Introduction

Sous le nom de processus stochastique on entend un modèle permettant d'étudier un phénomène aléatoire évoluant au cours du temps. Pour le décrire on considère, dans un univers Ω, un ensemble d'états E et une famille (X_t) de variables aléatoires à valeur dans E.

L'ensemble T que décrit l'indice t est dit ensemble des temps.

On se placera ici dans la situation où T est l'ensemble des entiers naturels $\boldsymbol{N}$. On a donc discrétisé le temps.

De même on supposera que l'ensemble E des états des variables aléatoires (X_t) est fini ou dénombrable.

On peut donc interpréter notre modèle comme une suite $X_0, X_1, X_2, \ldots$ de variables aléatoires discrètes.

L'intérêt de ce modèle est de l'appliquer à des situations où les variables aléatoires (X_t) ne sont pas indépendantes.

Exemple 1 : La marche unidimensionnelle symétrique.

Considérons une expérience de pile ou face répétée à chaque incrément de temps de manière indépendante. Associons au jet de la pièce à l'instant j ($j \in \boldsymbol{N}$) la variables aléatoire Y_j définie par

$$Y_j = \begin{cases} 1 \text{ si le j-ème jet donne pile} \\ -1 \text{ si le j-ème jet donne face} \end{cases}$$

Considérons la suite des variables aléatoires à valeur dans $\boldsymbol{Z}$ définies par $X_t = \sum_{j=1}^{t} Y_j$ et $X_0=0$.

Le processus stochastique correspondant à la suite des (X_t) décrit donc l'évolution au cours du temps de l'excès de résultats pile par rapport aux face. Ce processus a aussi une interprétation connue sous le nom de marche unidimensionnelle symétrique. En effet soit une particule se déplaçant sur l'axe des $\boldsymbol{Z}$ par incrément de 1. Imaginons qu'à chaque étape du temps on tire avec équiprobabilité si la particule se déplace vers la droite ou vers la gauche. La particule étant à l'instant initial en 0. La suite des (X_t) représente une marche aléatoire de cette particule suivant ces lois.

On peut facilement obtenir la loi que suit X_t puisque le nombre de déplacement à droite suit une loi binomiale. Mais la connaissance de la loi de chaque X_t n'est pas suffisante pour connaître le processus puisque les (X_t) ne sont pas indépendantes. On aimerait ici pouvoir répondre à des questions sur la dynamique de l'ensemble des (X_t) comme par exemple "Quelle est la probabilité de revenir au point de départ en un temps fini ? " si il y a retour "Quelle est l'espérance du nombre de déplacements nécessaires ? ".

Nous reviendrons plus loin sur ce modèle et son intérêt pour la physique statistique.

Dans cet exemple les (X_t) ne sont pas indépendantes mais il est clair que X_{t+1} ne dépend que de X_t plus exactement : $P\{X_{t+1} = j\} = ½ P\{X_t = j-1\} + ½ P\{X_t = j+1\}$.

Cette situation très courante correspond à la notion de chaîne de Markov.

2 Chaînes De Markov.

Définition:

Une **chaîne de Markov** est une suite de variables aléatoires X_n $(n \geq 0)$ à valeurs dans E (supposé ensemble d'entiers) telle que la probabilité conditionnelle de X_{n+1} sur les états passés est fonction seulement de X_n ; c'est-à-dire:

$P\{X_{n+1} = j / X_0=i_0, X_1=i_1, \dots X_n=i_n\} = P\{X_{n+1} = j / X_n=i_n\}$.

On peut interpréter cette définition en disant que c'est un processus à temps discret où le futur ne dépend que du présent et pas du passé.

Définition:

Nous dirons qu'une chaîne de Markov est **homogène** si la probabilité $P\{X_{n+1} = j / X_n=i\}$ est en outre indépendante de n.

De très nombreux processus stochastiques vérifient cette propriété et nous nous placerons dans ce contexte de chaînes de Markov homogène.

Il est alors commode d'utiliser la terminologie suivante. On considère qu'on a un système pouvant avoir plusieurs états évoluant dans le temps, l'évènement $\{X_n = i\}$ s'interprétant en disant que le système est dans l'état i à l'instant n. On appelle alors probabilité de transition en une étape de l'état i à l'état j la valeur p_{ij} , indépendante de n, de $P\{X_{n+1} = j / X_n=i\}$.

Dans le cas d'un nombre fini d'états la matrice $P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1q} \\ p_{21} & p_{22} & \cdots & p_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ p_{q1} & p_{q2} & \cdots & p_{qq} \end{pmatrix}$ dont les coefficients sont les probabilités de transition est dite matrice de transition de la chaîne.

Cette matrice est donc une matrice dons les coefficients sont dans *[0,1]* et dont toutes les lignes sont de somme 1 (on parle de matrice stochastique). Inversement à une matrice stochastique et à un état initial on peut associer une chaîne de Markov.

Si au temps n le système est dans l'état i la i-ème ligne de P va représenter la distribution de probabilité entre les états à l'instant $n+1$. Si à l'instant n on a seulement une distribution de probabilité entre les divers états on peut représenter la variable X_n comme un vecteur $X_n=(p1,p2,..)$ avec $\sum_i p_i = 1$. Le produit $X_n.P$ représentera donc la distribution de probabilité à l'instant $n+1$ et donc de façon plus générale à l'instant $n+m$ on obtiendra $X_{n+m}= X_n.P^m$.

En fait la puissance m-nième de la matrice P, la matrice P^m représente les probabilités de transition entre états en m étapes.

Une distribution π est dite **stationnaire** si $\pi P= \pi$. La chaîne (X_n) associée à P et commençant au temps 0 par la distribution π est donc invariante. Nous verrons que sous certaines hypothèses la loi de X_n converge vers une telle distribution stationnaire qui plus est indépendamment de l'état initial.

Exemple 2 : le modèle de diffusion d'Ehrenfest.

Ce modèle a été introduit en 1907 et peut décrire la diffusion d'un gaz placé dans deux récipients à température différente mis en communication.

On considère deux urnes A et B contenant au total a boules. A chaque étape on tire une des boules avec la probabilité $1/a$ et on la change d'urne. On représente l'état du système par X_n le nombre de boules dans l'urne A au temps n.

Ce processus est une chaîne de Markov homogène de matrice de transition de dimension $a+1$.

Par exemple pour $a= 4$ on obtient

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

On peut associer à une chaîne homogène son graphe de transition. Les sommets du graphe sont les états et on dessine une flèche entre les états i et j si $P_{i,j} > 0$ (on parle plus exactement d'arc d'extrémités initiale i et terminale j) .

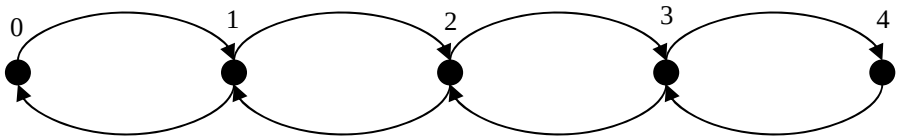


Figure 1: Graphe de transition du modèle d'Ehrenfest à 4 boules

On appellera **chemin** une succession d'arcs tels que l'extrémité terminale d'un arc est l'extrémité initiale du suivant. Beaucoup d'informations peuvent se déduire de l'examen du graphe de transition. On pourra, par exemple, voir si un état x est accessible à partir d'un état y par l'existence d'un chemin entre x et y. On dira que 2 états x et y communiquent si $x=y$ ou si x est accessible à partir de y et y à partir de x. Cette relation est une relation d'équivalence entre états. On considèrera les classes d'équivalence pour cette relation. On pourra réduire le graphe de transition en un graphe sur ces classes.

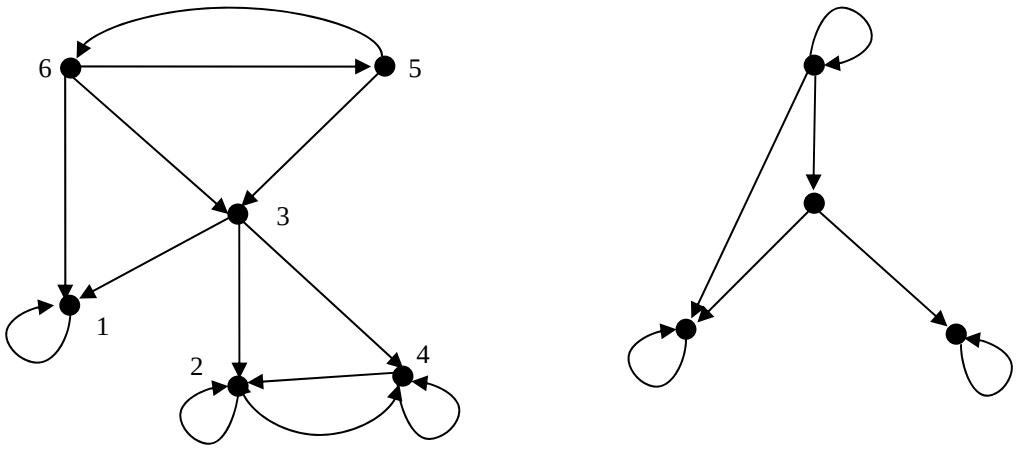


Figure 2: Un graphe de transition et le graphe réduit

Dans cet exemple on voit que si l'état initial était 1 on va rester dans cet état, que si l'état initial était 2 ou 4 on va rester perpétuellement entre les états 2 et 4. On parle de classe d'états finaux. Les états 3,5 et 6 sont d'une autre nature on n'y restera qu'un temps fini avant d'évoluer vers l'une ou l'autre des classes finales et y rester.

Dans le cas d'un nombre fini d'états un théorème d'algèbre linéaire (le théorème de Perron Frobenius) assure qu'une matrice stochastique dont tous les coefficients sont strictement positifs admet 1 comme valeur propre avec un espace propre de dimension 1. Les autres valeurs propres étant de module inférieur à 1. Ceci permet d'assurer que $P^{\infty} = \lim_{m\to\infty} P^m$ existe et est une matrice stochastique de rang 1 donc dont toutes les lignes sont identiques. On en déduit immédiatement que asymptotiquement la variable aléatoire X_n se répartit dans les différents états suivant une distribution limite, distribution stationnaire, indépendante de l'état initial.

Cependant en général la matrice P d'une chaîne de Markov n'est pas strictement positive.

Une chaîne de Markov finie est dite **ergodique** (ou irréductible) si tout état est atteignable en un nombre fini d'étapes à partir de tout autre état. On peut demander une propriété plus forte, le fait qu'il existe un nombre d'étapes qui permet de passer de tout état à tout autre état. Nous dirons donc qu'une chaîne est **régulièr**e si il existe un m tel que tous les coefficients de P^m sont strictement positifs. Ceci se traduit sur le graphe de transition par l'existence, pour tout couple d'états *x, y,* d'un chemin entre *x* et *y* constitué de *m* arcs.

Dans le cas d'une chaîne régulière on peut généraliser la théorie précédente et on obtient le résultat:

Théorème

Si P est la matrice de transition d'une chaîne de Markov régulière alors pour toute distribution initiale v on a $\lim_{n\to\infty} vP^n = \pi$, où π est l'unique distribution solution de $\pi P = \pi$.

Ce théorème est puissant puisqu'il permet de prouver l'existence d'une distribution limite et de la trouver sans avoir à calculer la matrice P^m.

2 Chaîne ergodique et retour sur le modèle de diffusion d'Ehrenfest.

Dans l'exemple du modèle de diffusion d'Ehrenfest la chaîne est ergodique mais n'est pas régulière puisque, par exemple pour $a=4$ si on part de l'état où toutes les boules sont en A les états avec un nombre impair de boules dans l'urne A sont injoignables en m étapes si m est pair, et si m est impair ce sont ceux avec un nombre pair de boules que l'on ne peut atteindre. Asymptotiquement la matrice P^m va osciller suivant la parité de m entre

$$P_{2k} = \begin{pmatrix} 1/8 & 0 & 3/4 & 0 & 1/8 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/8 & 0 & 3/4 & 0 & 1/8 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/8 & 0 & 3/4 & 0 & 1/8 \end{pmatrix} \text{ et } P_{2k+1} = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 & 0 \\ 1/8 & 0 & 3/4 & 0 & 1/8 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/8 & 0 & 3/4 & 0 & 1/8 \\ 0 & 1/2 & 0 & 1/2 & 0 \end{pmatrix}.$$

La moyenne des 2 matrices qui est une distribution stationnaire est la matrice

$$\begin{pmatrix} 1/16 & 1/4 & 3/8 & 1/4 & 1/16 \\ 1/16 & 1/4 & 3/8 & 1/4 & 1/16 \\ 1/16 & 1/4 & 3/8 & 1/4 & 1/16 \\ 1/16 & 1/4 & 3/8 & 1/4 & 1/16 \\ 1/16 & 1/4 & 3/8 & 1/4 & 1/16 \end{pmatrix}.$$

Chaque ligne correspond à la distribution binomiale de paramètres a et ½.

On voit sur cet exemple que pour une chaîne ergodique on n'a pas forcément l'existence d'une limite pour P^m. Par contre il existe une distribution π telle que $\pi P = \pi$.

En effet introduisons la matrice $Q = \frac{1}{2}[I + P]]$. Il est immédiat de vérifier que Q est stochastique. Puisque la matrice P est supposée être celle d'une chaîne ergodique soit k le nombre maximal d'étapes nécessaires pour joindre deux états quelconques de la chaîne.

On a $Q^k = \frac{1}{2^k}\left[I + \binom{k}{1}P + \binom{k}{2}P^2 + \ldots + \binom{k}{k-1}P^{k-1} + P^k \right]$.

Pour tout i,j il existe un m ($m \leq k$) tel que $P^m_{ij} > 0$. Comme les autres termes sont positifs ou nuls Q est donc une matrice d'une chaîne régulière. Donc il existe π telle que $\pi Q = \pi$. On en déduit que $\pi P = 2\pi Q - \pi = \pi$ et que de plus π est unique.

Comment interpréter cette distribution stationnaire de P? Une étude approfondie permet de montrer que la proportion de temps passé dans l'état i pendant les n premières étapes admet

pour limite quand n tend vers l'infini la i-ième composante de π. Cette proportion est donc indépendante de l'état de départ.

Définitions

Soit une chaîne ergodique partant de l'état $X_0=i$. On va revenir à cet état. La variable aléatoire $t_i= inf\ \{k>0/\ X_k=i\}$ est dite **temps de premier retour** en i son espérance τ_i est dite **temps de récurrence moyen** en i.

Théorème

Le temps de récurrence moyen en i est l'inverse de la i-ème composante de π: $\tau_i = \dfrac{1}{\pi_i}$.

Ce résultat est conforme à l'intuition puisque quand n tend vers l'infini π_i est la fréquence avec laquelle on est dans l'état i. Donc en n étapes il y a en moyenne $n\pi_i$ retours qui 'doivent' donc se faire en moyenne toutes les $1 / \pi_i$ étapes. Le calcul rigoureux confirme cette interprétation.

Si nous revenons au modèle de diffusion d'Ehrenfest avec a boules au départ toutes en A on voit donc qu'on va tendre vers une distribution limite de moyenne $\dfrac{a}{2}$. Cependant lors des fluctuations de la chaîne on va revenir avec une probabilité égale à 1 dans l'état de départ ! L'espérance du temps de récurrence est de 2^a donc devient considérable dès que a est grand (par exemple pour le nombre d'Avogadro). Ce modèle a permis de clarifier les rapports a priori contradictoires entre les points de vue macroscopique (irréversibilité de l'évolution de l'entropie) et microscopique (à cette échelle les transformations sont réversibles); problématique apparue avec l'interprétation statistique de l'entropie par Boltzmann(1897).

Depuis on a découvert d'autres applications de ce modèle par exemple dans l'étude des marchés en mathématiques financières ou en biologie.

3 Classification des états.

Revenons à l'étude du cas plus général d'une chaîne homogène et intéressons nous à la variable aléatoire N_x qui compte le nombre de passages de la chaîne dans l'état x. Supposons

que la chaîne parte de l'état x et soit q la probabilité de revenir à l'état x en un temps fini. Puisqu'on part de x on a $N_x>0$.

Ce processus stochastique n'ayant pas de mémoire du passé la probabilité de l'évènement $N_x=2$ correspond au produit de q (de x je reviens en x) par $1-q$ (de x je ne reviens jamais en x).

De la même manière de façon générale $P(N_x=k)= q^{k-1}(1-q)$.

Nous dirons qu'un état x est **récurrent** si la probabilité de revenir en x en un temps fini est égale à 1. On voit par le raisonnement précédent que dans le cas d'un état récurrent on va revenir **presque sûrement** une infinité de fois dans l'état initial.

Il est facile de vérifier que le fait d'être récurrent est une propriété compatible avec la notion de classe d'équivalence vue plus haut: Si x et y sont dans la même classe et si x est récurent alors y l'est aussi. On peut montrer une propriété plus forte: si x est récurrent et si y est accessible à partir de x alors y est récurrent et est dans la même classe. Donc quand on est dans une classe récurrente on ne peut plus en sortir; c'est la même notion que celle de classe finale.

Les autres états sont dits transitoires. C'est donc également une propriété commune aux états d'une même classe.

4 Retour sur les marches aléatoires.

Dans le cas d'une marche unidimensionnelle le graphe de transition est infini**.**

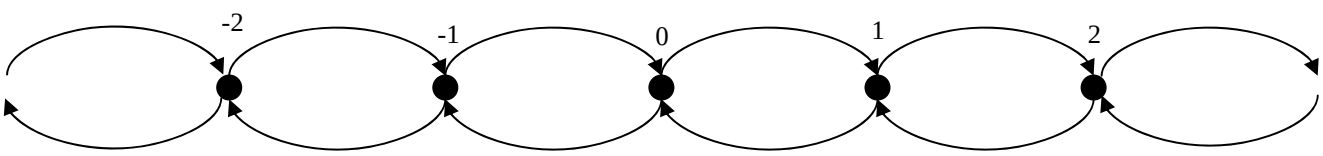


Figure 3: Le graphe de la marche unidimensionnelle

On peut montrer que la probabilité de revenir en 0 en un temps fini est égale à 1. En fait tous les états sont récurrents.

Polya a étudié la généralisation de ce modèle à plusieurs dimensions. Considérons par exemple la grille 2–dimensionnelle $\boldsymbol{Z}^2$ correspondant aux points du plan à coordonnées entières. A chaque étape on évolue avec équiprobabilité d'un point à un de ses 4 voisins

immédiats. Là aussi tous les états sont récurrents. Avec la probabilité 1 on reviendra donc toujours une infinité de fois à l'origine.

Il a montré (1921) qu'a partir de la dimension 3 ce n'est plus vrai. La probabilité de revenir à l'origine pour la marche 3-dimensionnelle symétrique est d'environ 0,34. En quelque sorte il est vrai que sur la terre un chemin au hasard conduit à Rome mais c'est faux dans l'espace!

Ce modèle a été utilisé dans des domaines très variés comme la modélisation du mouvement brownien, de la diffusion des parfums ou les mathématiques financières.

4 Conclusion.

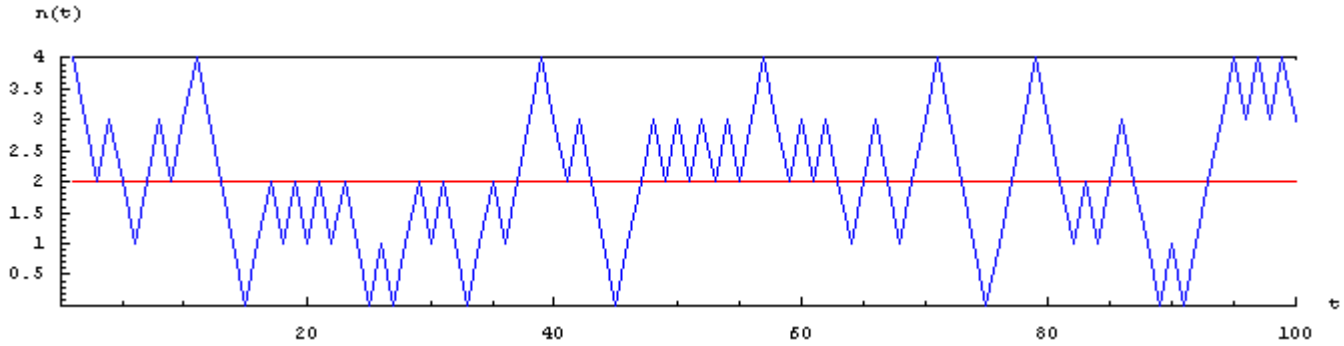
Nous n'avons fait qu'esquisser l'intérêt des chaînes de Markov. En effet nous n'avons pas abordé des développements très riches comme les problèmes d'optimisation où on associe un coût aux différentes transitions possibles et où par des décisions on peut modifier les matrices des coûts et de transition. Les chaînes de Markov apparaissent aussi dans la gestion des files d'attente. Précisons une dernière application dans la théorie de l'information. En effet on peut considérer un texte dans une langue donnée comme une chaine de Markov où la probabilité d'occurrence d'une lettre dépend du caractère précédent, chaîne dite du premier ordre, ou des 2 ou 3 derniers, chaîne d'ordre 2 ou 3. Par une étude statistique on obtient les probabilités de transition. On simule ainsi les différentes langues et Shannon (1951) a utilisé ce modèle pour définir la quantité d'information qu'elles véhiculent.

Voici par exemple le résultat d'une simulation pour la langue française:

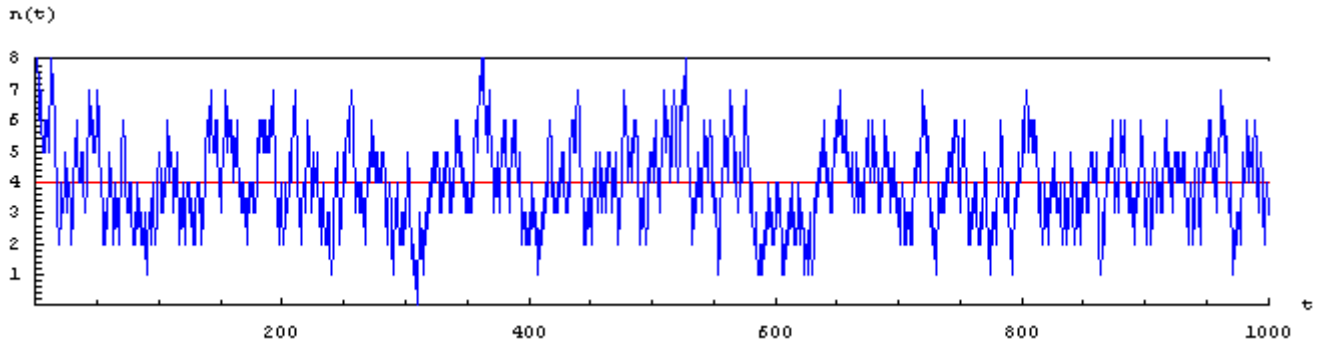
Au 2ème ordre "MAITAIS DU VEILLECALCAMAIT DE LEU DIT"

Au 3ème ordre " DU PARUT SE NE VIENNER PERDENT LA TET".

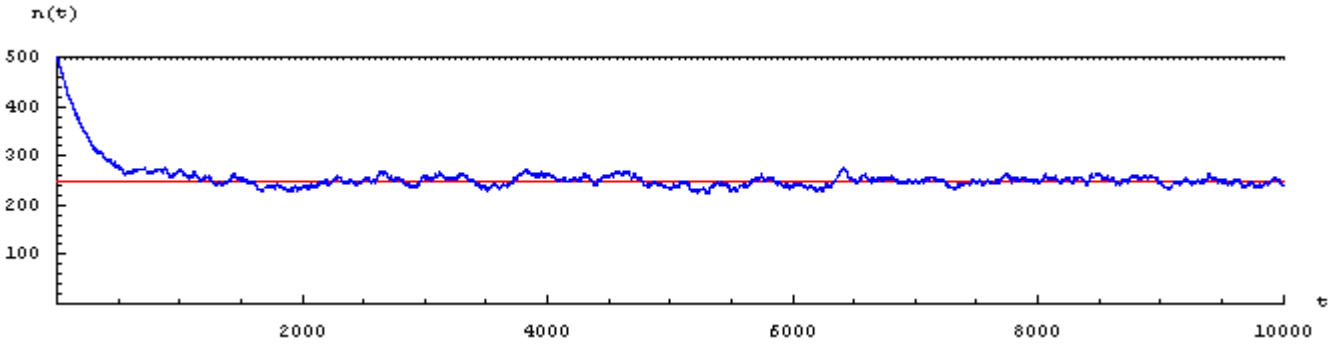
Annexe 1:



Simulation du modèle de diffusion d'Ehrenfest avec 4 boules au départ toutes en A



Simulation du modèle de diffusion d'Ehrenfest avec 8 boules au départ toutes en A



Simulation du modèle de diffusion d'Ehrenfest avec 500 boules au départ toutes en A

Annexe 2: Vocabulaire élémentaire de théorie discrète des probabilités.

On appelle **Univers** Ω un ensemble fini ou dénombrable correspondant à l'ensemble des issues possibles d'une expérience aléatoire.

Un **évènement** A est une partie de Ω.

On associe à chaque élément x de Ω sa **probabilité** $p(x)$; p étant une fonction à valeur dans $[0,1]$ vérifiant $\sum_{x\in\Omega} p(x) = 1$. Pour chaque évènement on définit alors sa probabilité par $p(A) = \sum_{x\in A} p(x)$.

L'évènement impossible est l'ensemble vide et est donc de probabilité nulle. On remarquera que l'inverse n'est pas forcément vrai. De même l'évènement certain Ω est de probabilité 1 et une propriété est dite vraie **presque sûrement** si elle est vérifiée pour un ensemble de probabilité 1. Par exemple si on répète un lancer de pièce jusqu'à obtenir PILE, on s'arrête presque sûrement au bout d'un nombre fini de lancers. Cependant il existe un évènement constitué d'une infinité de lancers FACE.

On notera $P(A/B)$ la probabilité de l'évènement A sachant que B est réalisé. On a donc $P(A/B) = \dfrac{P(A\cap B)}{P(B)}$.

Deux évènements A et B sont dits **indépendants** si la réalisation de l'un n'apporte aucune information sur la réalisation de l'autre autrement dit si $P(A\cap B) = P(A)P(B)$.

Une **variable aléatoire** X est une fonction sur Ω à valeur dans R.

On appelle **espérance** $E(X)$ d'une variable aléatoire X la moyenne de toutes les valeurs possibles de X pondérées par leurs probabilités de réalisation $E(X) = \sum_i k_i\, P(\{X = k_i\})$.

Deux variable aléatoires X et Y sont dites **indépendantes** si pour tout x, y les évènements $\{X=x\}$ et $\{Y=y\}$ sont indépendants.

Considérons la répétition d'une expérience, comme le lancer d'une pièce, qui n'a que deux résultats possibles PILE et FACE de probabilités p et $1-p$. Alors la variable aléatoire qui compte le nombre de PILE sur n expériences vérifie $P(\{X = k\}) = \binom{n}{k} p^k (1-p)^{n-k}$. Cette loi est dite **loi binomiale** de paramètres n et p. Son espérance est np.