

MB61

Mikael Falconnet
Bureau 405
mikael.falconnet@gencpde.cnrs.fr

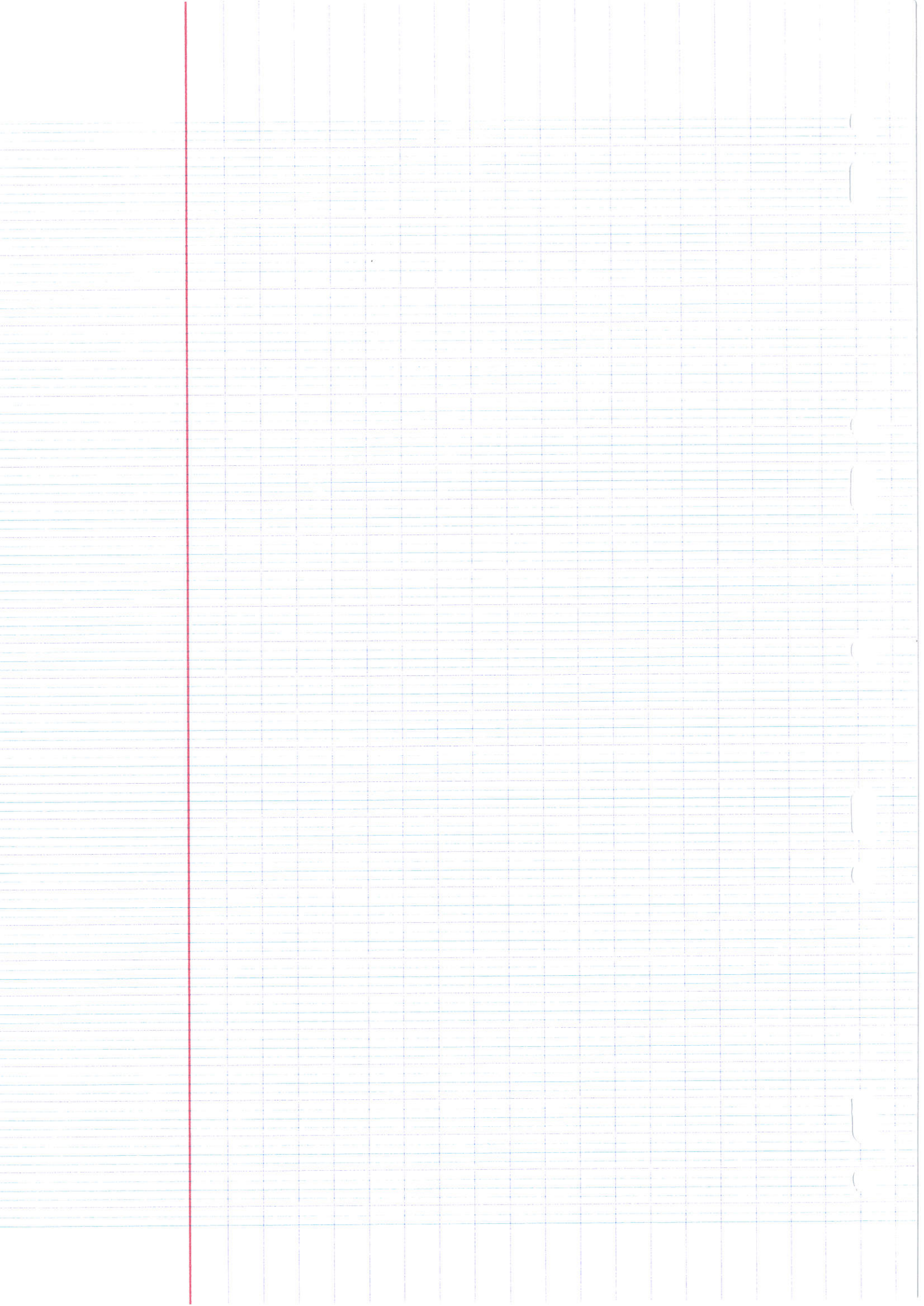
Evaluation

- . Partiel
- . CC₁ (27/02) + CC₂ (26/03 ou 02/04)
- . QCM à chaque TD

Contenu du cours

1. Rappels de probabilité
2. Modelisation
3. Chaînes de Markov
4. A voir

27/02. Tout jusqu'à la fin du chap 2



I Probabilités sur un ensemble fini ou dénombrable.

ach d'élève $\downarrow$ infini mais dénombrable

1.1 Univers, Evénement, Mesure de probabilité

On convient de représenter une expérience aléatoire $\mathcal{E}$, c'est-à-dire dont l'issue est soumise au hasard, par l'ensemble Ω de tous les résultats possibles à cette expérience.

Cet ensemble est appelé univers, espace des possibles ou encore espace fondamental.

Exemples d'univers

- L'univers pour le jeu de pile ou face est $\Omega = \{P, F\}$
- L'univers pour le jeu de pile ou face répété 2 fois.
 $\Omega = \{(P, P); (P, F); (F, P); (F, F)\}$

Un événement aléatoire A lié à une expérience aléatoire $\mathcal{E}$ est représenté par l'ensemble également noté A , des issues pour lesquelles l'événement A est réalisé. C'est un sous-ensemble de Ω .

Exemples d'événements

- "Obtenir pile" dans un jeu de pile ou face
- "Obtenir au moins un pile" lors de plusieurs lancers de pièce

Définition: Soit $\mathcal{E}$ une expérience aléatoire d'univers Ω . On appelle mesure de probabilité (ou plus simplement probabilité) une application P qui associe à tout événement A un nombre réel $P(A)$ telle que (i) $0 \leq P(A) \leq 1$
(ii) $P(\emptyset) = 0$ et $P(\Omega) = 1$

(iii) $A \cap B = \emptyset$ implique que $P(A \cup B) = P(A) + P(B)$

Si Ω est fini ou dénombrable, toute probabilité sur Ω est parfaitement déterminée par les données d'un ensemble de nombres

$$\{P(\omega), \omega \in \Omega\}$$

verifiant

$$- \forall \omega \in \Omega, P(\omega) \geq 0$$

$$- \sum_{\omega \in \Omega} P(\omega) = 1$$

On a alors $P(\{\omega\}) = P(\omega)$ et pour tout $A \subset \Omega$
^{^événement de}

$$P(A) = \sum_{\omega \in A} P(\omega)$$

On note $|E|$ le cardinal d'un ensemble. Si Ω est fini, la probabilité définie sur Ω par

$$P(\{\omega\}) = \frac{1}{|\Omega|}$$

est appelée probabilité uniforme. C'est la probabilité qui rend les issues équiprobables pour tout événement A

$$P(A) = \frac{|A|}{|\Omega|}$$

formule que l'on paraphrase souvent ainsi:

$$P(A) = \frac{\text{nbre de cas probables}}{\text{nbre de cas possibles}}$$

Exemple: Lancers de deux pièces

L'univers Ω associé au lancer de deux pièces est

$$\Omega = \{(P,P), (P,F), (F,P), (F,F)\}$$

$$|\Omega| = 4$$

Si l'on met la proba uniforme un Ω , alors la proba de l'événement A "obtenir au moins un pile" est donnée par

$$P(A) = \frac{|\{(P,P), (P,F), (F,P)\}|}{|\Omega|} = \frac{3}{4}$$

1.2 Variables aléatoires expériences, variances

Une variable aléatoire est une fonction dont la valeur dépend de l'issue d'une expérience aléatoire & d'univers Ω .

On dit qu'une variable aléatoire X est discrete si elle prend un nombre de valeurs fini ou dénombrable.

L'ensemble des issues ω sur lesquelles X prend une valeur fixée x forme un événement

$$[X = x]$$

La probabilité de l'événement $X=x$ est notée $P(X=x)$.

La fonction $p_x: x \mapsto P(X=x)$ est appelée la loi de la variable aléatoire X .

Si $\{x_1, x_2, \dots\}$ est l'ensemble des valeurs possibles pour X , on a

$$p_x(x_i) \geq 0 \quad \text{et} \quad \sum_i p_x(x_i) = 1$$

Exemple

Soit S_2 le nbre de pile obtenu lors du lancé de 2 pièces.

L'ensemble des valeurs possibles pour S_2 est $\{0, 1, 2\}$

Si l'on munit Ω associé à cette expérience aléatoire de la probabilité uniforme P , il vient

$$P(S_2=0) = P(\{F, F\}) = \frac{1}{4}$$

$$P(S_2=1) = P(\{(P, F), (F, P)\}) = \frac{1}{2} \quad \text{et enfin:}$$

$$P(S_2=2) = \frac{1}{4}$$

On remarque qu'il est possible de réécrire l'événement A : "Obtenir au moins un pile" comme $\{S_2 \geq 1\}$

On retrouve alors

$$P(A) = P(S_2 \geq 1) = P(S_2=1) + P(S_2=2) = \frac{3}{4}$$

Lorsqu'elle existe, on appelle espérance ou moyenne d'une variable aléatoire X discrète la quantité notée $E(X)$ définie

$$\begin{aligned} E(X) &= \sum_i x_i P(X=x_i) \\ &= \sum_i x_i P_X(x_i) \end{aligned}$$

Exemple: Si définie comme précédemment

$$E(S_2) = 0 \times P(S_2=0) + 1 \times P(S_2=1) + 2 \times P(S_2=2) = 1$$

Lorsqu'elle existe on appelle variance de X la quantité $\text{Var}(X)$ définie par $\text{Var}(X) = E[(X - E(X))^2]$

Proposition:

$$\text{Var}(X) = E(X)^2 - [E(X)]^2$$

Preuve: Exercice

1.3 Probabilité conditionnelles, indépendance

Soient P une mesure de probabilité sur Ω et B un événement tel que $P(B) > 0$

La probabilité conditionnelle de A sachant B est le réel

$P(A|B)$ défini par $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Remarque: L'application $A \mapsto P(A|B)$ est une mesure de probabilité.

Les événements A et B sont indépendants si:
 $P(A \cap B) = P(A) \times P(B)$.

On remarque que si $P(A) > 0$, $P(B) > 0$, on a les équivalences A et B indépendants $\Leftrightarrow P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B)$

$A_1, \dots, A_n$ sont indépendants (dans leur ensemble) si pour tout $k \in \{1, \dots, n\}$ et pour tout ensemble d'entiers distincts $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$
 $P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \times \dots \times P(A_{i_k})$

Des variables aléatoires $X_1, \dots, X_n$ sont dites indépendantes si les événements $[X_1 = \omega_1], \dots, [X_n = \omega_n]$ sont indépendants pour tout n -uplet "admissible" $(\omega_1, \dots, \omega_n)$.

Proposition 1.11

Soient X et Y deux variables aléatoires discrètes définies sur le même univers Ω .

On a toujours lorsque $E(X)$ et $E(Y)$ existent

$$E(aX + bY) = aE(X) + bE(Y) \text{ pour tous réels } a \text{ et } b$$

Si de plus, X et Y sont indépendantes $\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y)$

1.4 Loi jointes, marginales

Noos avons vu qu'une variable aléatoire X est caractérisée par sa loi p_X . En particulier, si deux variables aléatoires X_1 et X_2 ont même loi, on ne peut pas les distinguer quand on les

considère séparément.

De plus, mis à part le cas où X_1 et X_2 sont indépendantes, connaître les lois de X_1 et X_2 ne suffit pas pour connaître la loi de la variable aléatoire $Y = (X_1, X_2)$.

Exemple :

Soient deux variables aléatoires de Bernoulli de paramètre $\frac{1}{2}$
 X_1 et X_1'

$$\begin{aligned} P(X_1=1) &= P(X_1'=1) = \frac{1}{2} \\ P(X_1=0) &= P(X_1'=0) = \frac{1}{2} \end{aligned}$$

On définit X_2 comme une copie de X_1 c-à-d $X_2 = X_1$

On définit X_2' comme $X_2' = 1 - X_1$

On voit que si $X_1=1$, alors $X_2'=0$, et si $X_1=0$ alors $X_2'=1$.

On remarque que X_2 et X_2' sont aussi des v.a. de Bernoulli de paramètre $\frac{1}{2}$

On définit $Y = (X_1, X_2)$
 $Y' = (X_1', X_2')$

La loi de Y est donnée par $P(Y=(0,0)) = \frac{1}{2}$, $P(Y=(1,1)) = \frac{1}{2}$ tandis que la loi de Y' est donnée par $P(Y'=(1,0)) = \frac{1}{2}$, $P(Y'=(0,1)) = \frac{1}{2}$

Avec cet exemple, nous introduisons le concept de loi jointe

Soient $X_1, X_2, \dots, X_n$ des variables aléatoires définies sur le même univers Ω et respectivement à valeurs dans des ensembles finis ou dénombrables.

$A_1, \dots, A_n$. La loi jointe de $X_{1:n} = (X_1, \dots, X_n)$ est la fonction

$P_{X_{1:n}} : A_1 \times \dots \times A_n \rightarrow [0,1]$ définie par

$$P_{X_{1:n}}(x_1, \dots, x_n) = P(X_1=x_1, \dots, X_n=x_n)$$

↑
"fois"

Dans le cas particulier où $X_1, \dots, X_n$ sont indépendantes dans leur ensemble, on a $P_{X_1, n}(x_1, \dots, x_n) = P(X_1=x_1) \times \dots \times P(X_n=x_n)$

$$= P_{X_1}(x_1) \times \dots \times P_{X_n}(x_n)$$

$$= \prod_{i=1}^n P_{X_i}(x_i)$$

Les lois marginales sont les lois de chacune des variables aléatoires X_i prises séparément.

On peut voir grâce à l'exemple ci-dessus que la loi jointe apporte plus d'informations que les lois marginales.

Loi de Bernoulli $E(X) = p$
 $V(X) = p(1-p)$

Loi géométrique

"loi du premier succès rencontré qd on répète une expérience de Bernoulli"

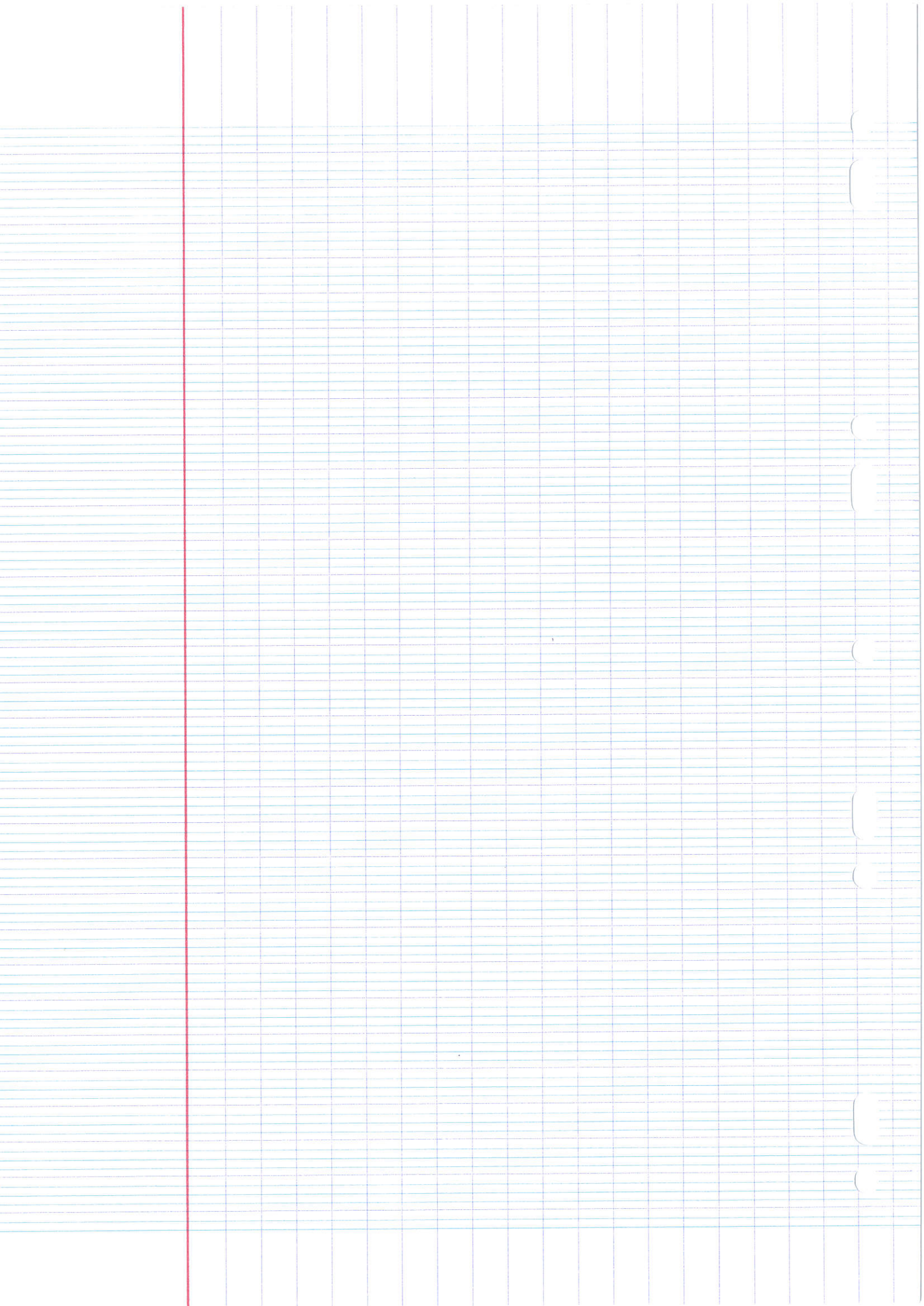
$$E(G) = \frac{1}{p} \quad V(G) = \frac{1}{p^2}$$

$A \perp B$

sachant que

A, B

et, n



II Un peu de modélisation

Lancer de pièces de monnaie:

* Incapacité de prédire le résultat du lancé
→ Justifie l'utilisation d'un modèle probabiliste.

* Résultats possibles de l'expérience?
→ 2 valeurs possibles: pile ou face

* Loi de probabilité:

Un paramètre $p \in [0, 1]$

$$IP(\text{pile}) = p$$

$$IP(\text{face}) = 1 - p$$

On peut se demander si $p = \frac{1}{2}$

La réponse pourra être expliquée grâce à la loi des grands nombres

2.1 Loi des grands nombres pour pile ou face

Dans le cadre d'un grand nombre de lances de la même pièce de monnaie, on s'interroge sur la proportion de pile que on devrait obtenir.

Intuition $\approx \frac{1}{2}$

"Quand le nbre de lances augmente, la proportion converge vers $\frac{1}{2}$ "
Tout ceci dépend d'une expérience aléatoire et il convient de choisir un modèle, puis donner un sens à cette convergence

Pour notre modélisation, on se donne une suite de v.a. $(X_n)_{n \geq 1}$

indépendantes et équiréparties.

$$P(X_n=1)=p \quad \forall n \geq 1$$

$$P(X_n=0)=1-p \quad \forall n \geq 1$$

$$p \in]0,1[$$

$$S_n = X_1 + X_2 + \dots + X_n$$

représente le nombre de piles obtenues au cours des n premiers lancers.

$$\bar{X}_n = \frac{S_n}{n} \quad \text{représente la proportion de piles obtenus au cours des } n \text{ premiers lancers.}$$

Proposition:

La var S_n suit une loi binomiale de paramètres (n, p) . S_n est à valeurs dans $\{0, \dots, n\}$. $P(S_n=k) = \binom{n}{k} p^k (1-p)^{n-k}$ où $\binom{n}{k}$ donne le nombre de sous-ensembles distincts à k éléments que l'on peut former à partir d'un ensemble contenant n éléments.

$$E(S_n) = np \quad \text{Var}(S_n) = np(1-p)$$

Preuve

$[S_n=k]$ est égal à l'événement "obtenir k piles lors de n lancers"

$$\underbrace{(0, 1, 0, 1, 0, 0, 1)}_{n \text{ lancers}} \quad \text{avec } k \text{ fois } 1$$

L'événement $[S_n=k]$ est la réunion sur tous les k -uplets $(i_1, \dots, i_k)$ de $\{1, \dots, n\}$ des événements

$$A_{i_1, \dots, i_k} = \left(\bigcap_{j=1}^k [X_{i_j}=1] \right) \cap \left(\bigcap_{i \notin \{i_1, \dots, i_k\}} [X_i=0] \right)$$

Les événements $A_{i_1, \dots, i_k}$ sont disjoints deux à deux.

$$P(A_{i_1, \dots, i_k}) = p^k (1-p)^{n-k}$$

grâce à l'indépendance des X_i .

Le nombre de k -uplets différents que on peut former est donné par $\binom{n}{k}$

$$\text{Ainsi } P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\begin{aligned} E(S_n) &= E(X_1 + \dots + X_n) \\ &= E(X_1) + \dots + E(X_n) \\ &= \underbrace{p + \dots + p}_{n \text{ fois}} \\ &= np \end{aligned}$$

$$\begin{aligned} \text{Var}(X_1 + \dots + X_n) &= \text{Var}(X_1) + \dots + \text{Var}(X_n) \\ X_1, \dots, X_n \text{ ind} &\rightarrow \underbrace{p(1-p) + \dots + p(1-p)}_{n \text{ fois}} \\ &= np(1-p) \end{aligned}$$

Loi Binomiale

Esperance

$$E = np$$

Variance

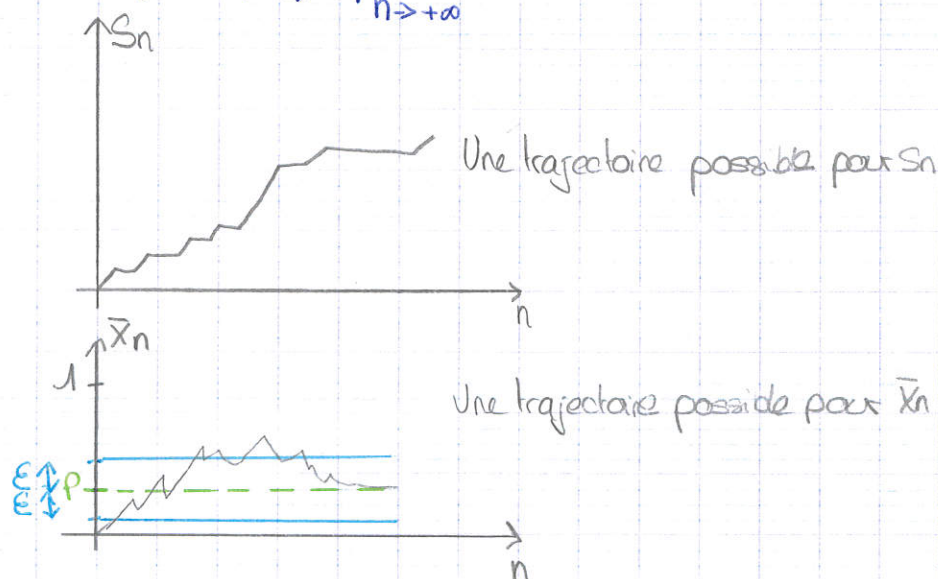
$$\text{Var} = np(1-p)$$

On remarque que $E(\bar{X}_n) = p$

En fait, on a plus fort: lorsque n tend vers l'infini, la proportion de pile $\bar{X}_n$ converge vers p au sens suivant.

Théorème: Qd n devient grand, $\bar{X}_n$ converge en probabilité vers p , c'est-à-dire que pour tout $\varepsilon > 0$

$$P(|\bar{X}_n - p| \geq \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 0$$



Proposition: (Inégalité de Bienaymé-Tchebychev)

Soit X une v.a. telles que $E(Y)$ et $\text{Var}(Y)$ existent. On alors

$$P(|Y - E(Y)| \geq \alpha) \leq \frac{\text{Var}(Y)}{\alpha^2} \quad \text{Pour tout } \alpha > 0.$$

Preuve Soit Z la v.a. définie par $Z = 1_{\{|Y - E(Y)| \geq \alpha\}}$
c'est à dire valent

$$1 \text{ si } |Y - E(Y)| \geq \alpha$$

$$0 \text{ si } |Y - E(Y)| < \alpha$$

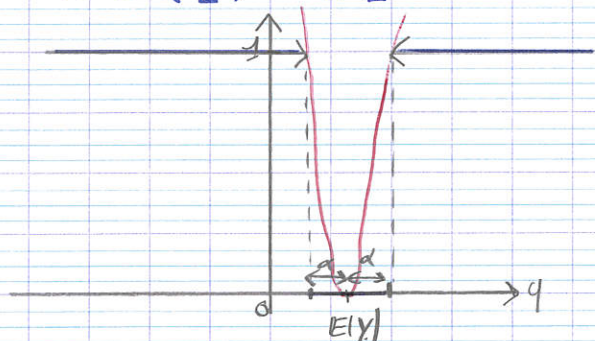
C'est une v.a. de Bernoulli de paramètre $p = P(|Y - E(Y)| \geq \alpha)$.

On a alors

$$E(Z) = p = P(|Y - E(Y)| \geq \alpha)$$

On remarque que

$$\alpha^2 Z \leq [Y - E(Y)]^2$$



$$\begin{aligned} \text{--- } y &\mapsto 1_{\{|y - E(Y)| \geq \alpha\}} \\ \text{--- } y &\mapsto \frac{[y - E(Y)]^2}{\alpha^2} \end{aligned} \quad \uparrow \text{ } \int \text{supérieur à}$$

En effet, si $|Y - E(Y)| < \alpha$, alors le mb de gauche vaut 0 et le mb de droite est positif donc l'inégalité est respectée.

Si $|Y - E(Y)| \geq \alpha$, le mb de gauche est égal à α^2 , et le mb de droite est plus grand que α^2 car

$$|Y - E(Y)| \geq \alpha \Rightarrow [Y - E(Y)]^2 \geq \alpha^2$$

Là encore, l'inégalité est respectée.

$$\text{On a } \alpha^2 Z \leq \sum [Y - E(Y)]^2$$

En prenant l'espérance de part et d'autre, on obtient

$$\begin{aligned} E(\alpha^2 Z) &\leq E(\sum [Y - E(Y)]^2) \\ &= \alpha^2 E(Z) \quad \text{Var}(Y) \\ &= \alpha^2 P(|Y - E(Y)| \geq \alpha). \end{aligned}$$

Le fait que l'inégalité entre deux variables aléatoires $Y_1 \leq Y_2$ soit conservée par passage à l'espérance vient du fait que

$$Y_3 \geq 0 \Rightarrow E(Y_3) \geq 0$$

On applique sa à $Y_2 - Y_1 \geq 0$ et on obtient $E(Y_2 - Y_1) \geq 0$
Par linéarité

$$E(Y_2 - Y_1) = E(Y_2) - E(Y_1)$$

Ainsi

$$\begin{aligned} E(Y_2) - E(Y_1) &\geq 0 \\ \Leftrightarrow E(Y_2) &\geq E(Y_1) \end{aligned}$$

Preuve du théorème

Soit $\varepsilon > 0$, on applique Bienaymé-Tchebychev à $\bar{X}_n$.

$$E(\bar{X}_n) = p$$

$$\begin{aligned} \text{Var}(\bar{X}_n) &= \text{Var}\left(\frac{S_n}{n}\right) \\ &= \frac{1}{n^2} \text{Var} S_n \end{aligned}$$

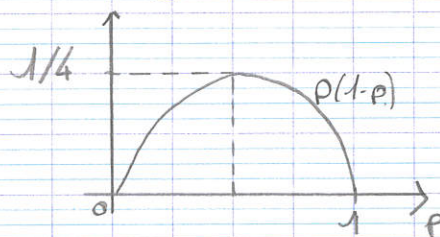
$$= \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) = \frac{1}{n^2} [\text{Var}(X_1) + \dots + \text{Var}(X_n)]$$

$$= \frac{1}{n^2} \underbrace{[p(1-p) + \dots + p(1-p)]}_{n \text{ fois}} = \frac{p(1-p)}{n}$$

$$\frac{1}{n^2} \overbrace{n p(1-p)}^{\rightarrow}$$

$$P(|\bar{X}_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \quad \text{Vient de } P(|Y - E(Y)| \geq \varepsilon) \leq \frac{\text{Var}(Y)}{\varepsilon^2}$$

$$P(|\bar{X}_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0$$



$$p(1-p) = p - p^2 = -p^2 + p + 0$$

On a bien

$$P(|\bar{X}_n - p| \geq \varepsilon) \xrightarrow{n \rightarrow \infty} 0 \quad \text{pour tout } \varepsilon > 0$$

(faible)

2.2. Loi des grands nombres pour un modèle indépendant de génération de séquences d'ADN

Soyons clair, nous ne savons pas comment sont générées les séquences d'ADN.

Tout ce qu'on peut faire c'est imaginer le modèle aléatoire, le tester sur des données et recommencer dès que les résultats mathématiques ne sont pas compatibles avec la réalité biologique.

Nous supposons qu'une séquence de longueur n est modélisée par une suite de variables aléatoires $X_1, \dots, X_n$ à valeurs dans l'alphabet des nucléotides $A = \{A, C, G, T\}$ indépendantes et équidistribuées suivant la loi.

$$P(X_i = x) = p(x), \quad \forall x \in A$$

On a bien sûr

$$\sum_{x \in A} p(x) = 1 \quad \text{et} \quad p(x) > 0, \quad \forall x \in A$$

On note $N_n(x)$ et $R_n(x)$ les v.a. définies par $N_n(x) = \sum_{i=1}^n Y_i(x)$ où $Y_i(x) = \mathbb{1}\{X_i = x\}$ $R_n(x) = \frac{N_n(x)}{n}$

$N_n(x)$ représente le nbre de nucléotides x dans la séquence, $R_n(x)$ représente la proportion de nucléotides x .

Les variables aléatoires $(Y_1(x), \dots, Y_n(x))$ sont indépendantes et équidistribuées suivant une loi de Bernoulli de paramètre $p(x)$.

Proposition

Quand n devient grand $R_n(x)$ converge en probabilité vers $p(x)$, c-à-d pour $\forall \varepsilon > 0$ $P(|R_n(x) - p(x)| \geq \varepsilon) \xrightarrow{n \rightarrow +\infty} 0$

Preuve: Application directe du théorème de convergence pour le pile ou face

On vient de voir que quand la longueur de la séquence est grande, la proportion observée de nucléotides de type x tend vers la proportion théorique $p(x)$ sous le modèle indépendant.

Qu'en est-il de la proportion observée de dinucléotides xy ?

$$N_n(xy) = \sum_{i=1}^{n-1} Y_i(xy) \quad \leftarrow \text{nbre de nucl.} \quad \text{Proportion} \downarrow$$

$$R_n(xy) = \frac{N_n(xy)}{n-1}, \text{ où } Y_i(xy) = \mathbb{1}\{X_i = x, X_{i+1} = y\}$$

Propositi^o: les variables aléatoires $Y_1(xy), \dots, Y_{n-1}(xy)$ sont équidistribuées et suivent une bi de Bernoulli de paramètre $p(x)p(y)$.

En revanche elles ne sont pas indépendantes

$$\begin{matrix} |a|c|g| \\ y_1/y_2 \end{matrix}$$

preuve: les v.a. $Y_i(xy)$ sont bien à valeurs dans $\{0, 1\}$.

$$P(Y_i(xy) = 1) = P(X_i = x, X_{i+1} = y)$$

$$\begin{aligned} \text{indépendance} &: P(X_i = x) P(X_{i+1} = y) \\ &= p(x) p(y) \end{aligned}$$

$$= P(X_i = x) P(X_{i+1} = y)$$

$$= P(x) p(y)$$

Pour voir que les variables aléatoires ne sont pas indépendantes, nous allons montrer

$$P(Y_1 | x, y) = 1, Y_2 | x, y = 1$$

$\neq$

$$P(Y_1 | x, y) = 1 \times P(Y_2 | x, y) = 1$$

$$= [P(x) p(y)]^2$$

par indépendance

$$P(Y_1 | x, y) = 1, Y_2 | x, y = 1$$

$$= P(X_1 = x, X_2 = y, X_2 = x, X_3 = y)$$

Si $x \neq y$, $[X_2 = y]$ est incompatible avec $[X_2 = x]$ donc le probab ci-dessus est nulle,

Si $x = y$, on a $P(Y_1 | x, y) = 1, Y_2 | x, y = 1 = P(X_1 = x, X_2 = x, X_3 = x) = p(x)^3$
 $p(x) \times p(x) \times p(x)$

D'autre part, on a

$$P(Y_1 | x, y) = 1 \times P(Y_2 | x, y) = 1 = p(x)^4$$

$$p(x)^3 = p(x)^4$$

$$\Leftrightarrow p(x)^3 - p(x)^4 = 0$$

$$\Leftrightarrow p(x)^3 [1 - p(x)] = 0$$

$$\Leftrightarrow p(x) = 0 \text{ ou } 1$$

Dans notre cas on n'a jamais

$$p(x)^3 = p(x)^4$$

En conclusion, $Y_1 | x, y$ et $Y_2 | x, y$ ne sont pas indépendantes et ce pour tous $x, y \in \mathcal{X}$.

Theoreme: Quand n devient grand, $R_n | x, y$ converge en probabilité vers $p(x)p(y)$ c'est à dire que $\forall \epsilon > 0$

$$\mathbb{P}(|R_n(y) - p(x)p(y)| \geq \varepsilon) \xrightarrow{n \rightarrow \infty} 0$$

Preuve = Une fois encore, nous allons utiliser l'inégalité de Bienaymé-Tchebychev.

Calcul de l'espérance $\mathbb{E}(R_n(y))$

$$\begin{aligned} \mathbb{E}(R_n(y)) &= \frac{1}{n-1} \mathbb{E}(N_n(y)) \\ \text{Linéarité de l'espérance} \quad &= \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{E}(Y_i(y)) \quad \left. \begin{array}{l} N_n = \sum_{i=1}^{n-1} Y_i \\ \downarrow \\ p(x)p(y) \end{array} \right\} \\ &= \frac{1}{n-1} \underbrace{[p(x)p(y) + \dots + p(x)p(y)]}_{n-1 \text{ fois}} \\ &= p(x)p(y) = \mathbb{E}(N_n(y)) \end{aligned}$$

Calcul de la variance

$$\text{Var}(R_n(y)) = \frac{1}{(n-1)^2} \text{Var}(N_n(y))$$

On a besoin de calculer $\text{Var}(N_n(y))$

$$\text{Var}(N_n(y)) = \mathbb{E}([N_n(y)]^2) - \mathbb{E}(N_n(y))^2$$

Comme $\mathbb{E}(N_n(y)) = \underbrace{(n-1)}_n \underbrace{p(x)p(y)}_p$, le seul terme qu'on ne connaît pas est $\mathbb{E}([N_n(y)]^2)$.

Pour comprendre l'expression de $[N_n(y)]^2$

$$\begin{aligned} (a_1 + a_2)^2 &= a_1^2 + a_2^2 + 2a_1a_2 \quad (\text{identité remarquable}) \\ (a_1 + a_2 + a_3)^2 &= a_1^2 + a_2^2 + a_3^2 + 2a_1a_2 + 2a_1a_3 + 2a_2a_3 \\ (a_1 + \dots + a_m)^2 &= \sum_{i=1}^m a_i^2 + 2 \sum_{1 \leq i < j \leq m} a_i a_j \end{aligned}$$

$i \backslash j$	1	2	3	...	m
1					
2					
$\vdots$					
m					

Juste rajout
d'esperance

$$[N_n(x_u)]^2 = \sum_{i=1}^{n-1} Y_i^2(x_u) + 2 \sum_{1 \leq i < j \leq n-1} Y_i(x_u) Y_j(x_u)$$

$$\mathbb{E}([N_n(x_u)]^2) = \sum_{i=1}^{n-1} \mathbb{E}(Y_i^2(x_u)) + 2 \sum_{1 \leq i < j \leq n-1} \mathbb{E}(Y_i(x_u) Y_j(x_u))$$

Comme $Y_i^2(x_u) = Y_i(x_u)$, on a $\mathbb{E}(Y_i^2(x_u)) = \mathbb{E}(Y_i(x_u)) = p(x) p(y)$;
 Si U et V st deux v.a. independantes alors $\mathbb{E}(UV) = \mathbb{E}(U) \mathbb{E}(V)$



Si $j-i \geq 2$, alors $Y_i(x_u)$ et $Y_j(x_u)$ sont independantes.

Ainsi si $j-i \geq 2$,

$$\mathbb{E}(Y_i(x_u) Y_j(x_u)) = \underbrace{\mathbb{E}(Y_i(x_u))}_{p(x)p(y)} \underbrace{\mathbb{E}(Y_j(x_u))}_{p(x)p(y)} = p(x)^2 p(y)^2$$

soit $p(x)p(y)p(x)p(y)$

Combien y a-t-il de couples (i,j) tel que

$$* 1 \leq i < j \leq n-1$$

$$* j-i \geq 2$$

$i=1 \rightarrow$ valeurs possibles de j

$$3 \rightarrow n-1$$

$n-2$ termes

$i=2 \rightarrow$ valeurs possibles par j

$$4 \rightarrow n-1$$

$n-3$ termes

$i = n-3 \rightarrow$ valeurs possibles pour j

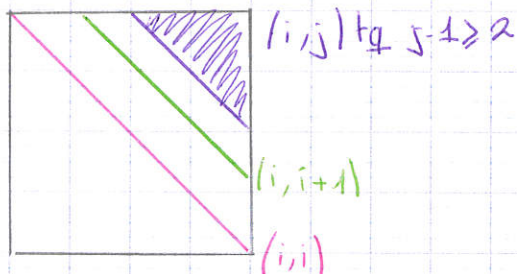
$n-1 \rightarrow n-1$
1 terme

En tout, il y a

$$1+2+\dots+(n-3)+(n-2) \\ = \frac{(n-3)(n-2)}{2} \text{ termes}$$

Ainsi

$$2 \leq \mathbb{E}(Y_i(x_u) Y_j(x_u)) = (n-1)(n-2) p(x)^2 p(y)^2 + \sum_{i=1}^{n-2} \mathbb{E}(Y_i(x_u) Y_{i+1}(x_u)) \\ 1 \leq i \leq j \leq n-1$$



$Y_i(x_u) Y_{i+1}(x_u)$ est ainsi une variable aléatoire de Bernoulli. Son paramètre est donné par $\mathbb{P}(Y_i(x_u) Y_{i+1}(x_u) = 1)$

$$= \mathbb{P}(Y_i(x_u) = 1, Y_{i+1}(x_u) = 1)$$

$$= \begin{cases} 0 & \text{si } x \neq y \\ p(x)^2 & \text{si } x = y \end{cases}$$

Ainsi

$$\sum_{i=1}^{n-2} \mathbb{E}(Y_i(x_u) Y_{i+1}(x_u)) = (n-2) p(x)^2 \delta_x(y)$$

où $\delta_x(y) = \begin{cases} 0 & \text{si } y \neq x \\ 1 & \text{sinon} \end{cases}$

Il vient

$$\mathbb{E}(\|N_n(x_u)\|^2) = (n-1) p(x) p(y) + (n-1)(n-2) p(x)^2 p(y)^2 \\ + (n-2) p(x)^2 \delta_x(y)$$

$$E[N_n(x,y)]^2 = (n-1)^2 p(x)^2 p(y)^2$$

$$\begin{aligned} \text{Var}(N_n(x,y)) &= \overbrace{(n-1)p(x)p(y)}^{E[N_n(x,y)]} + \overbrace{p(x)^2 p(y)^2 (n-1)[(n-2)-(n-1)]}^{=-1} \\ &\quad + (n-2)p(x)^3 p(y) \\ &= (n-1)p(x)p(y) - (n-1)p(x)^3 p(y)^2 + (n-2)p(x)^3 p(y) \end{aligned}$$

$$\begin{aligned} \text{Var}(R_n(x,y)) &= \frac{(n-1)p(x)p(y)[1-p(x)p(y)]}{n-1 \times (n-1)} + \frac{(n-2)p(x)^3 p(y)}{(n-1)^2} \\ &\leq \frac{5}{4(n-1)} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

Ainsi, pour $\varepsilon > 0$

$$P(|R_n(x,y) - p(x)p(y)| \geq \varepsilon) \leq \frac{\text{Var}(R_n(x,y))}{\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0$$

$$R_n(x) \xrightarrow[n \rightarrow +\infty]{\text{Probable}} p(x)$$

$$R_n(y) \xrightarrow[n \rightarrow +\infty]{\text{Probable}} p(y)$$

$$R_n(x,y) \xrightarrow[n \rightarrow +\infty]{P} p(x)p(y)$$

On définit

$$D_n = \frac{\sum_{x,y \in \mathcal{A}} [n N_n(x,y) - N_n(x) N_n(y)]}{n N_n(x) N_n(y)}$$

Proposition: Sous l'hypothèse du modèle indépendant.

D_n converge en loi vers la loi χ^2 à 9 degrés de liberté.

III Chaînes de Markov

Charlotte la marmotte est une grosse dormeuse. Elle cherche un compagnon qui dort beaucoup. Elle rencontre Firmin le lémurien. Elle veut répondre à la question :

"Combien de tps, en moyenne, Firmin dort-il par jour?"

Pour répondre à cette question, on va utiliser un modèle markovien.

La principale caractéristique d'un tel modèle est l'absence de mémoire.

Cela signifie que seul l'état actuel du processus (suite de va) influence son état qui va suivre.

1. Définition et propriété de Markov

Définition : Soient S un ensemble fini ou dénombrable (appelé espace d'états) et $(X_n)_{n \in \mathbb{N}}$ une suite de va définies sur le même univers Ω muni de la mesure de probabilité P et à valeurs dans S .

On dit que X_n est une chaîne de Markov homogène si

* (Propriété de Markov) pour tout entier n et tous $x_0, \dots, x_{n+1} \in S$
$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n)$$

* (Homogénéité) la probabilité de transition $P(X_{n+1} = y | X_n = x)$ ne dépend pas de n , pour tous $x, y \in S$.

On note $p(x, y)$ cette probabilité de transition.

La loi de X_0 est appelée la loi initiale de la chaîne, et on appelle matrice de transition (de dimensions possiblement infinies), la matrice définie par

$$P = (p(x, y))_{x, y \in S^2}$$

La matrice P vérifie

$$p(x, y) \geq 0 \text{ et } \sum_y p(x, y) = 1$$

Une telle matrice est dite stochastique (aléatoire).

Dans notre exemple, un état est une activité pour le lémurien Firmin.

$$S = \{ \text{Dormir, Manger, Se brosser la fourrure} \}$$
$$= \{ D, M, B \}$$

Nous allons supposer que chaque minute, Firmin peut changer d'activité ou la poursuivre, et que cela ne dépend que de l'activité en cours.

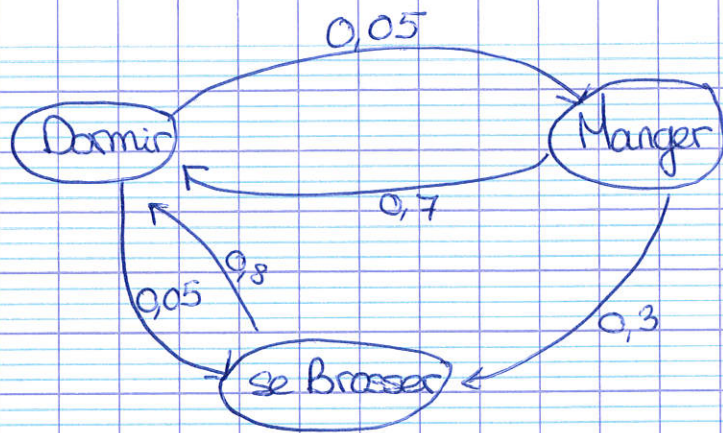
Comportement de Firmin

- * Qd Firmin dort, il ne se réveille pas la minute suivante avec proba 0,9.
- * Si Firmin se réveille, il mange ou se brossse la fourrure avec équiprobabilité.
- * Lorsqu'il mange cela ne dure qu'une minute puis il change d'activité.
- * Après le repas, il se brosse la fourrure avec proba 0,3 ou bien il entame une sieste avec une proba de 0,7.
- * Après 1 min de brossage, Firmin s'écroule avec proba 0,8, ou bien il continue avec proba 0,2.

La matrice P correspondant à la description précédente est

	^{sachant} D	M	B	
→ D	0,9	0,05	0,05	Pour les calculs.
→ M	0,7	0	0,3	
→ B	0,8	0	0,2	

Donc



Pour mieux voir il est possible d'augmenter l'intensité des flèches
 G qui peut confirmer la théorie.

Chaine de Markov, $(X_n)_{n \geq 0}$ essentiellement c'est

- Une loi initiale μ

$$\mu(x) = P(X_0 = x)$$

- une matrice de transition P

$$p(x, y) = P(X_{n+1} = y | X_n = x)$$

Fil conducteur

Firmin; le lemuronien

Dormir, Manger, Se Brosser. voir matrice

"Combient de tps par jour en moyenne dort Firmin?"

Propriété de Markov (absence de mémoire)

$$\begin{aligned}
 &P(X_{n+1} = y | X_0 = x_0, \dots, X_n = x_n) \\
 &\quad \text{tout le passé du processus} \\
 &= P(X_{n+1} = y | X_n = x_n)
 \end{aligned}$$

G qui nous intéresse

$$\frac{1}{n} \sum_{k=0}^{n-1} 1_{\{X_k = \text{Dormir}\}}$$

Calcul de la loi d'une trajectoire:

Rappel: Dans le cas indépendant si (X_n) est une suite de v.a. iid de loi μ

$$\begin{aligned} P(X_0 = x_0, \dots, X_n = x_n) &= P(X_0 = x_0) \dots P(X_n = x_n) \\ &= \mu(x_0) \times \dots \times \mu(x_n) \end{aligned}$$

Dans la suite, nous abrégons donc l'expression "une chaîne de Markov de loi initial μ et de bi de probabilité de transition P " par "Markov/ μ, P ".

Théorème: Si un processus discret $(X_n)_{n \geq 0}$ est Markov/ μ, P , alors pour tout entier n et tout état $x_0, x_1, \dots, x_n \in S$

$$P(X_0 = x_0, \dots, X_n = x_n) = \mu(x_0) p(x_0, x_1) p(x_1, x_2) \dots p(x_{n-1}, x_n)$$

Preuve:

Nous allons utiliser une preuve par récurrence.

Soit P_n le prédicat de récurrence suivant:

"Pour tous $x_0, \dots, x_n \in S$

$$\begin{aligned} P(X_0 = x_0, \dots, X_n = x_n) \\ = \mu(x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n) \end{aligned}$$

Montrons que P_0 est vrai:

Pour tout $x_0 \in S$

$$P(X_0 = x_0) = \mu(x_0)$$

* Soit n un entier positif, montrons que $P_n \Rightarrow P_{n+1}$

Supposons que P_n est vraie et montrons que P_{n+1} est vraie.

Soient $x_0, \dots, x_{n+1} \in S$.

$$P(X_0 = x_0, \dots, X_{n+1} = x_{n+1})$$

Propriété de Markov

$$= P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) \times P(X_0 = x_0, \dots, X_n = x_n)$$

$$= P(X_{n+1} = x_{n+1} | X_n = x_n) \times P(X_0 = x_0, \dots, X_n = x_n)$$

Prédicat

$$= P(X_{n+1} = x_{n+1} | X_n = x_n) \times \mu(x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n)$$

Prédicat de Récurrence

transformation en produit

$$= p(x_n, x_{n+1}) \times \mu(x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n) \\ = \mu(x_0) p(x_0, x_1) \dots p(x_n, x_{n+1})$$

2. Loi de X_n

Quelle est la probabilité qu'après n pas, la chaîne de Markov soit dans un état donné?

Dans notre exemple, quelle est la probabilité qu'après n minutes Firmin soit entrain de dormir? Manger? Se brosser la denture?

(D) (M) (S)

Mais allons voir que ce problème se réduit au calcul de la puissance n -ième de P ?

Proposition

Soit (X_n) Markov (μ, P) . Alors

$$P(X_n = x) = [\mu P^n](x), \forall x \in S$$

$$\mu = (\dots)$$

On a $|S|$ coefficients

"Firmin" $\mu = (\underbrace{\mu(D) \quad \mu(M) \quad \mu(B)}_{3 \text{ coefficients}})$

$$P^n = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} \quad |S|$$

$|S|$

colonne de x

$|S|$

$|S|$

$|S|$

$|S|$

$[\mu P^n](x)$

μP^n

Preuve

Par récurrence

Pn: " $P(X_n = x) = [\mu P^n](x)$, pour tout $x \in S$ "

$x P_0$ est vraie

$$P(X_0 = x) = \mu(x) \quad \text{par définition de } \mu \\ \text{car } P = Id \\ \downarrow \\ = [\mu P^0](x)$$

$$x P_n \Rightarrow P_{n+1}$$

Soit $x \in S$,

$$P(X_{n+1}=x) = \sum_{y \in S} P(X_{n+1}=x, X_n=y)$$

$$= \sum_{y \in S} P(X_{n+1}=x | X_n=y) P(X_n=y)$$

$$= \sum_{y \in S} p(y, x) P(X_n=y)$$

Propriété de

recurrence

$$= \sum_{y \in S} p(y, x) [\mu P^n](y)$$

$$= \sum_{y \in S} [\mu P^n](y) \times p(y, x)$$

$$= \int [\mu P^n](y) \times p(y, x) d\mu(y) = \int [\mu P^n] \times p(x) d\mu(y) = \int [\mu P^{n+1}](x) d\mu(y)$$

$$= [\mu P^{n+1}](x)$$

"Firmin"

Appelons $\mu^{(n)}$ la loi de X_n , c'est $\mu^{(n)}(x) = P(X_n=x)$
On vient de voir $\mu^{(n)} = \int \mu P^n$

$$\mu^{(0)} = (1 \ 0 \ 0) \quad \text{loi de } X_0$$

$$\mu^{(1)} = (0,9 \ 0,05 \ 0,05) \quad \text{loi de } X_1$$

$$\mu^{(2)} = (0,885 \ 0,045 \ 0,07) \quad \text{loi de } X_2$$

$$\mu^{(3)} = (0,884 \ 0,045 \ 0,07175) \quad \text{loi de } X_3$$

On peut remarquer que $\mu^{(2)}$ et $\mu^{(3)}$ sont proches

On verra dans la suite que ce n'est pas un hasard.

3 Convergence vers l'équilibre

Beaucoup de propriétés de chaînes de Markov sont liées à la notion de mesure de probabilité invariante.

Définition: On dit qu'une mesure de probabilité λ est invariante par P si $\lambda P = \lambda$

$$(\lambda(1) \dots \lambda(n)) \left(\begin{array}{c} \boxed{1} \\ \vdots \\ \boxed{n} \end{array} \right) \xrightarrow{P(1,1)} \sum_{i=1}^n \lambda(i) p(i,1)$$

$$p(n,n) = (\lambda(1) \dots \lambda(n))$$

$$\lambda P = \lambda \Leftrightarrow \sum_{x \in S} \lambda(x) p(x, y) = \lambda(y), \forall y \in S$$

Dans notre exemple, la mesure définie par $\pi = \left(\frac{160}{181} \quad \frac{8}{181} \quad \frac{13}{181} \right)$ est invariante par P .

$$\frac{160}{181} \approx 88,4\% \quad \frac{8}{181} \approx 4,4\% \quad \frac{13}{181} \approx 7,2\%$$

$$\begin{aligned} \pi^{(1)} &= \pi^{(0)} P = \pi P = \pi \\ \pi^{(n)} &= \pi^{(0)} P^{(n)} = \pi^{(0)} \underbrace{P \cdot P \cdot \dots \cdot P}_{n \text{ fois}} \\ &= \pi \end{aligned}$$

Un calcul de P^3 nous donne

$$P^3 = \begin{pmatrix} 0,884 & 0,04425 & 0,7125 \\ 0,8835 & 0,0435 & 0,073 \\ 0,884 & 0,044 & 0,072 \end{pmatrix}$$

On a l'impression que

$$p^{(n)}(x, y) \xrightarrow{n \rightarrow +\infty} \pi(y)$$

Décomposition en classes

Il est parfois possible de "casser" une chaîne de Markov en composantes plus petites, chacune d'entre elles étant relativement facile à comprendre, et qui toutes ensemble permettent de comprendre le comportement de la chaîne globale.

Ceci est possible grâce aux classes de communication de la chaîne.

Rappelons $p^{(n)}(x, y)$ est défini comme $P(X_n = y \mid X_0 = x)$
 noté comme

$$P_{x,0}(X_n = y)$$

↑ probabilité partant de x (au temps 0) à y à l'instant n

Définition: Soient x et y deux états de S .

On dit que x mène à y , noté $x \rightarrow y$, si $P_{x,0}(X_n = y)$ pour du moins un n est strictement positive.

On dit que x communique avec y , noté $x \leftrightarrow y$, si x mène à y et y mène à x .

Exemple



On voit que

$$\left. \begin{array}{l} * 0 \text{ mène à } 1 \\ * 1 \text{ mène à } 0 \end{array} \right\} 0 \leftrightarrow 1$$

$$\left. \begin{array}{l} * 0 \text{ mène à } 2 \\ * 2 \text{ mène à } 0 \end{array} \right\} 0 \leftrightarrow 2$$

$$\left. \begin{array}{l} * 0 \rightarrow 3 \\ * 3 \rightarrow 0 \end{array} \right\} 0 \leftrightarrow 3$$

$$\left. \begin{array}{l} * 0 \rightarrow 4 \\ 4 \not\rightarrow 0 \end{array} \right\} 0 \not\leftrightarrow 4 \Rightarrow 0 \rightarrow 4$$

Théorème: Pour des états distincts x et y , les propositions suivantes sont équivalentes

- (i) x mène à y
- (ii) $p^{(n)}(x, y) > 0$ pour au moins un n il existe un chemin orienté de longueur n
- (iii) $p(x_0, x_1) p(x_1, x_2) \dots p(x_{n-1}, x_n) > 0$ pour des états $x_0, x_1, \dots, x_n$ avec $x_0 = x$ et $x_n = y$ est un chemin orienté de longueur n

Remarque: (ii) se lit "il existe un chemin ^{orienté} de longueur n reliant x à y " et (iii) " $x_0, x_1, \dots, x_n$ est un chemin orienté reliant x à y "

Preuve

$$p(x, y) \text{ par au moins un } n \geq 0 \\ = P(x) / \bigcup_{n \in \mathbb{N}} [X_n = y]$$

Pour tout n , on a

$$[X_n = y] \in \bigcup_{n \in \mathbb{N}} [X_n = y]$$

Donc

$$P(x, y) \leq P(x) / \bigcup_{n \in \mathbb{N}} [X_n = y] \Rightarrow P(x, y) = p^{(n)}(x, y) \text{ d.l.}$$

Donc

$$p^{(n)}(x, y) \leq P(x) / \bigcup_{n \in \mathbb{N}} [X_n = y]$$

On a donc

$$p^{(n)}(x, y) > 0 \Rightarrow P(x) / \bigcup_{n \in \mathbb{N}} [X_n = y] > 0$$

(ii) $\Rightarrow$ (i) OK! Voir (i) $\Rightarrow$ (ii)

On remarque que

$$P(x) / \bigcup_{n \in \mathbb{N}} [X_n = y] \leq \sum_{n \in \mathbb{N}} P(x) [X_n = y]$$

$$P(A \cup B) \leq P(A) + P(B) \text{ monotonicité}$$

Si $P(x) / \bigcup_{n \in \mathbb{N}} [X_n = y] > 0$, alors $\exists p^{(n)}(x, y) > 0$ oui!

Donc il existe un n tel que
 $p^{(n)}(x, y) > 0$.

On peut remarquer que:

$$p^{(n)}(x, y) = \sum_{x_1, \dots, x_{n-1}} p(x, x_1) p(x_1, x_2) \dots p(x_{n-1}, y)$$

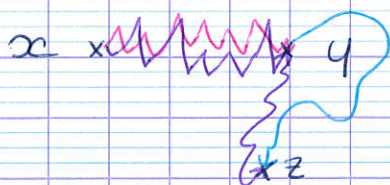
Le théorème dit simplement que si l'on regarde le diagramme de la chaîne de Markov le fait que x mène à y équivaut à la présence d'un chemin débutant en x et finissant en y .

La relation $\leftrightarrow$ est une relation d'équivalence sur S .

En effet,

Transitivité

* Si x communique avec y et si y communique avec z alors x communique avec z .



Reflexif

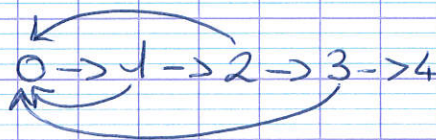
* l'état x communique avec lui-même car
 $P_{xx}(X_0 = x) = 1 > 0$

Symétrie

* Si x communique avec y alors y communique avec x .

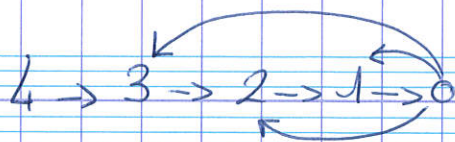
Ainsi, on peut décomposer S en classes de communication.

Exemple

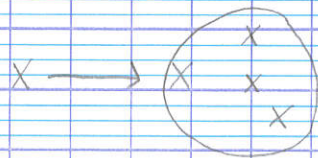


Cette chaîne à 2 classes de communication
 $\{0, 1, 2, 3\}$ et $\{4\}$



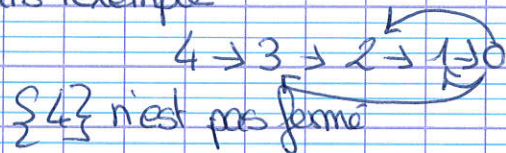


Cette chaîne a les mêmes classes de communication



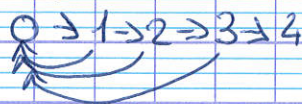
Définition: On dit qu'une classe C est fermée si: $x \in C \Rightarrow y \in C$
 $x \rightarrow y$

Dans l'exemple



Par contre $\{0, 1, 2, 3\}$ est fermé

Dans l'exemple



$\{4\}$ est fermé

$\{0, 1, 2, 3\}$ n'est pas fermé

Définition: un état x est dit absorbant si $\{x\}$ est une classe fermée.

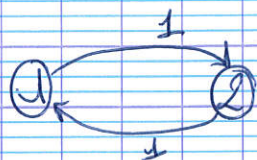
Une chaîne de Markov avec matrice de transition P est dite irréductible si elle possède d'une seule classe de communication.

On pense que $P(X_n = x) \xrightarrow[n \rightarrow \infty]{} \pi(x)$ où π invariant pour P . (Convergence)

Aperiodicité

On cherche le comportement limite des probabilités de transition $p^{(n)}(x, y)$ quand n devient grand.

Regardons un exemple par lequel un tel comportement n'est pas possible.



$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

on a $P^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Donc pour tout $n \geq 0$

$$P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ et } P^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

En particulier pour tout n .

$$P^{(2n)}(1,1) = 1$$

$$P^{(2n+1)}(1,1) = 0$$

Ainsi, la suite $(P^{(n)}(x,x))$ ne peut pas converger *il faut faire attention*

Définition: Un état x est apériodique si $P^{(n)}(x,x) > 0$ pour tout n assez grand *pour tout n*

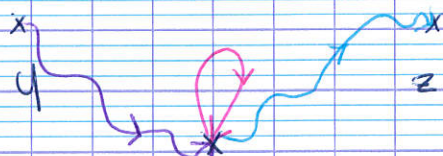
Lemme: Supposons que P est irréductible et possède au moins un état apériodique x . Alors pour tout y et z , on a

$$P^{(n)}(y,z) > 0 \text{ et}$$

pour tout n assez grand.

En particulier, tous les états sont apériodiques

preuve:



Chaine
irréductible

$\Rightarrow$ chemin de longueur n menant de y à $x \Leftrightarrow P^{(n)}(x,y) > 0$

aperiodique de x $\Rightarrow$ chemin de longueur s menant de x à $z \Leftrightarrow p^{(s)}(x, z) > 0$
 $\Rightarrow$ pour tout n assez grand on a une boucle de longueur n partant de $x \Leftrightarrow p^{(n)}(x, x) > 0$

Le chemin $\rightarrow \rightarrow \rightarrow$ mène de y à z ($y \rightarrow z$)

$$\begin{aligned} & p^{(r+n+s)}(y, z) \\ \geq & \underbrace{p^{(r)}(y, x)}_{>0} \underbrace{p^{(n)}(x, x)}_{>0} \underbrace{p^{(s)}(x, z)}_{>0} \\ & > 0 \end{aligned}$$

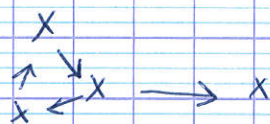
Réurrence et transience

Nous avons vu qu'il était possible de décomposer une chaîne de Markov en classes de communication pour en comprendre le comportement. Nous allons voir que certaines de ces classes sont transitoires tandis que l'on passe infiniment souvent dans d'autres.

Définition Soit (X_n) Markov (μ, P) . On dit qu'un état x est récurrent si

$$P_x(X_n = x \text{ pour une infinité de } n) = 1$$

On dit qu'un état x est transient si $P_x(X_n = x, \text{ pour une infinité de } n) = 0$



transient

récurrent

A hand-drawn diagram of a cell. The cell is roughly rectangular with a wavy outer boundary. Inside, there is a large, dark, oval-shaped nucleus at the top left. To its right is a smaller, bean-shaped mitochondrion with internal folds. Below the nucleus is a Golgi apparatus, depicted as a series of stacked, flattened sacs. To the right of the Golgi is a lysosome, a small, circular organelle. At the bottom right is a large, clear vacuole. Labels with arrows point to each of these structures: Nucleus, Mitochondrion, Golgi apparatus, Lysosome, and Vacuole.





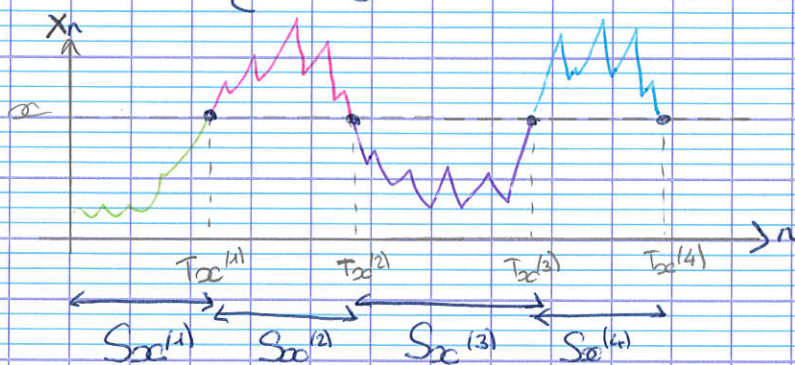
1. $\frac{1}{2}$

1

$$T_x^{(r+1)} = \inf \{ n \geq T_x^{(r)} + 1; X_n = x \}$$

On peut définir la longueur des excursions en dehors de x par

$$S_x^{(r)} = \begin{cases} T_x^{(r)} - T_x^{(r-1)} & \text{si } T_x^{(r-1)} < \infty \\ 0 & \text{sinon} \end{cases}$$



Lemme: Pour tout $r \geq 2$, conditionnellement à $T_x^{(r-1)} < \infty$, $S_x^{(r)}$ est indépendante de

$$\{X_m : m \leq T_x^{(r-1)}\}$$

et

chaîne de Markov.

$$\begin{aligned} \mathbb{P}(S_x^{(r)} = k \mid T_x^{(r-1)} < \infty) \\ = \mathbb{P}_x(T_x = k) \end{aligned}$$

$$[T_x = k] = [X_1 \neq x, X_2 \neq x, \dots, X_{k-1} \neq x, X_k = x]$$

Soit V_x le nbre de visite en x

$$V_x = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = x\}}$$

Avec cette définition, on voit qu'un état x est récurrent si et seulement si

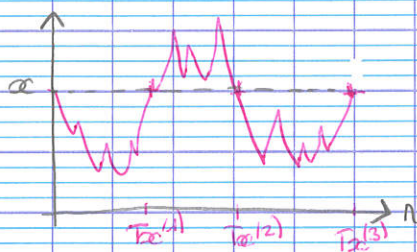
$$\begin{aligned} \mathbb{P}_x(V_x = \infty) &= \lim_{r \rightarrow \infty} \mathbb{P}_x(V_x > r) \\ &= 1 \end{aligned}$$

On peut remarquer que

$$\mathbb{E}_x(V_x)$$

$$\begin{aligned}
&= \mathbb{E} \left(\sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = \infty\}} \right) \\
&= \sum_{n=0}^{\infty} \mathbb{E} \left(\mathbb{1}_{\{X_n = \infty\}} \right) \quad \text{Ici On peut échanger somme et espérance} \\
&= \sum_{n=0}^{\infty} \mathbb{P}(X_n = \infty) \\
&= \sum_{n=0}^{\infty} p^{(n)}(\infty, \infty)
\end{aligned}$$

Lemme: Pour tout entier r , la probabilité $\mathbb{P}_x(V_x > r) = \mathbb{P}_x(T_x < \infty)^r$



Preuve :

Remarquons que si l'on part de x le nombre de visites en x est strictement plus grand que r si et seulement si le temps du r ième passage en x est fini.

Ainsi

$$\mathbb{P}_x(V_x > r) = \mathbb{P}_x(T_x^{(r)} < \infty)$$

On a donc en remarquant que

$$T_x^{(r+1)} = T_x^{(r)} + S_x^{(r+1)}$$

$$\begin{aligned}
\mathbb{P}_x(V_x > r+1) &= \mathbb{P}_x(T_x^{(r+1)} < \infty) \\
&= \mathbb{P}_x(T_x^{(r)} < \infty, S_x^{(r+1)} < \infty)
\end{aligned}$$

$$\begin{aligned}
\text{Lemme} \quad &= \mathbb{P}_x(S_x^{(r+1)} < \infty \mid T_x^{(r)} < \infty) \times \mathbb{P}_x(T_x^{(r)} < \infty) \\
&= \mathbb{P}_x(T_x < \infty) \times \mathbb{P}_x(T_x^{(r)} < \infty)
\end{aligned}$$

Ainsi

$$\begin{aligned}
\mathbb{P}_x(V_x > r+1) &= \mathbb{P}_x(T_x < \infty) \times \mathbb{P}_x(V_x > r) \\
&= \mathbb{P}_x(T_x < \infty)^2 \times \mathbb{P}_x(V_x > r-1)
\end{aligned}$$

$$= \dots$$

$$= P_x(T_x < \infty)^{n+1} \times 1$$

Rappel : $E_x(V_x) = \sum_{n=0}^{\infty} \frac{p^{(n)}(x, \infty)}{P_x(X_n = \infty)}$

On s'intéresse à

$P_x(X_n = \infty)$ pour une infinité de n et on a vu
 $P_x(V_x \geq r) = P_x(T_x < \infty)^r$

On va admettre que si V est une variable aléatoire positive et à valeurs entières.

$$E(V) = \sum_{v \in V} v \cdot P(V=v)$$

$$= \sum_{v=0}^{\infty} P(V \geq v)$$

$$E(V) = \sum_{v=0}^{\infty} v \cdot P(V=v)$$

$$= \underbrace{P(V=1)} + \underbrace{P(V=2)} + \underbrace{P(V=2)} + \underbrace{P(V=3)} + \underbrace{P(V=3)} + \underbrace{P(V=3)} \dots$$

$$= \underbrace{P(V=1)} + \underbrace{P(V=1) + P(V=2)} + \dots$$

$$+ \underbrace{P(V=2) + P(V=3)} \dots$$

$$+ \underbrace{P(V=3)} + \dots$$

$$= \underbrace{P(V \geq 1)}$$

$$+ \underbrace{P(V \geq 2)}$$

$$+ \underbrace{P(V \geq 3)} \dots$$

$$= \underbrace{P(V > 0)}$$

$$+ \underbrace{P(V > 1)}$$

$$+ \underbrace{P(V > 2)}$$

Théorème : On a la dichotomie suivante :

(i) Si $P_x(T_x < \infty) = 1$, alors l'état x est récurrent et

$$E_x(V_x) = \sum_{n=0}^{\infty} p^{(n)}(x, \infty) = \infty$$

(ii) Si $P_{xx}(T_{xx} < \infty) < 1$, alors l'état x est transient et

$$E_{xx}(V_{xx}) = \sum_{n=0}^{\infty} p^{(n)}(x, x) < \infty$$

Preuve

$$\begin{aligned} (i) P_{xx}(V_{xx} > r) &= \frac{P_{xx}(T_{xx} < \infty)^r}{=} \text{I moins (i)} \\ &= 1 \end{aligned}$$

Ainsi $P_{xx}(V_{xx} = \infty) = 1$

$$\text{car } P_{xx}(V_{xx} = \infty) = \lim_{r \rightarrow \infty} P_{xx}(V_{xx} > r)$$

(ii) Comme

$$E_{xx}(V_{xx}) = \sum_{r=0}^{\infty} P_{xx}(V_{xx} > r)$$

$$= \sum_{r=0}^{\infty} P_{xx}(T_{xx} < \infty)^r$$

$$= 1 + P_{xx}(T_{xx} < \infty) + P_{xx}(T_{xx} < \infty)^2 + \dots$$

$$= \frac{1}{1 - P_{xx}(T_{xx} < \infty)} < \infty$$

Si $0 < a < 1$

$$\sum_{r=0}^{\infty} ar = \frac{1}{1-a}$$

$$\sum_{r=0}^n ar = \frac{1-a^{n+1}}{1-a}$$

Ainsi

$$E_{xx}(V_{xx}) < \infty$$

D'où

$$P_{xx}(V_{xx} = \infty) = 0$$

Theoreme : Soit C une classe de communication.

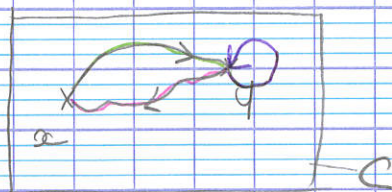
Alors, soit tous les états de C sont transients, soit ils sont tous récurrents.

Preuve : Soit x un état transient de C et y un autre état de C

Montrons que y soit transient

Pour cela regardons

$$\sum_{n=0}^{\infty} p^{(n)}(y, y)$$



$$\rightarrow p^{(s)}(x, y) > 0$$

$$\rightarrow p^{(n)}(y, x) > 0$$

On a

$$p^{(s+n+t)}(y, y) \geq p^{(s)}(x, y) p^{(n)}(y, x) p^{(t)}(y, x)$$

ce qui se réécrit

$$p^{(n)}(y, y) \leq \frac{1}{p^{(s)}(x, y) p^{(t)}(y, x)} \times p^{(s+n+t)}(x, x)$$

Ainsi

$$\sum_{n=0}^{\infty} p^{(n)}(y, y) \leq \frac{1}{p^{(s)}(x, y) p^{(t)}(y, x)} \sum_{n=0}^{\infty} p^{(s+n+t)}(x, x)$$

$$\text{Or } \sum_{n=0}^{\infty} p^{(s+n+t)}(x, x) = \sum_{k=s+t}^{\infty} p^{(k)}(x, x)$$

et

$$\sum_{k=s+t}^{\infty} p^{(k)}(x, x) \leq \sum_{k=0}^{\infty} p^{(k)}(x, x) < \infty$$

car x est transient

$$\text{Donc } \sum_{n=0}^{\infty} p^{(n)}(y, y) < \infty$$

Donc y est transient.

Theoreme : Tant une classe de communication C finie et fermée est récurrente.

preuve : Supposons que la chaîne de Markov commence en C . Comme C est fermée, la chaîne reste à l'intérieur de C au cours du temps.

Comme C est finie, il existe un x tel que

$$P(X_n = x \text{ pour une infinité de } n) > 0$$

En effet dans le cas contraire, on aurait pour chaque $x \in C$ un entier (aléatoire) n_x tel que

$$P(X_n \neq x, \forall n \geq n_x) = 1$$

$N = \max(n_x)$ est défini
 $x \in C$

car on a un nombre fini d'états.

D'où

$$P(X_n \in C, \forall n \geq N) = 1$$

ce qui contredit le fait que la classe est fermée.

On a

$$P(X_n = x \text{ pour une infinité de } n)$$

$$= P(T_x < \infty) P_x(X_n = x \text{ pour une infinité de } n)$$

Mais on vient de voir que

$$P(X_n = x \text{ pour une infinité de } n) > 0$$

donc

$$P_x(X_n = x \text{ pour une infinité de } n) > 0$$

Mais on sait que

$$P_x(X_n = x \text{ pour une infinité de } n) \text{ vaut } 0 \text{ ou } 1.$$

On en conclut que x_0 est récurrent.

Lemme. Supposons que P est irréductible et récurrente (automatique si S est fini).
Alors pour tout $y \in S$ $P(T_y < \infty) = 1$

Théorème. Soit $(X_n)_{n \geq 0}$ une chaîne de Markov de mesure initiale μ et de matrice de transition P avec P irréductible, apériodique et possédant une mesure invariante π .

Alors la $P(X_n = \infty) \xrightarrow{n \rightarrow +\infty} \pi(\infty)$, $\forall x \in S$
et

$$p^{(n)}(x, y) \xrightarrow{n \rightarrow +\infty} \pi(y), \forall x, y \in S$$

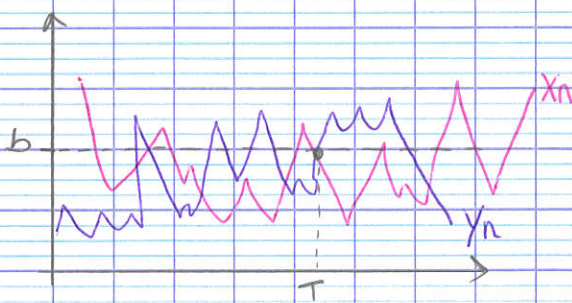
Preuve: uniquement quand S est fini.

Soit (X_n) Markov (μ, P)

Soit (Y_n) Markov (π, P) indépendante de X_n

On fixe un état $b \in S$ et on définit

$$T = \inf \{ n \geq 1, X_n = Y_n = b \}$$



Etape 1. Montrons que la
 $P(T < \infty) = 1$

On définit une nouvelle chaîne de Markov $(W_n)_{n \geq 0}$ sur $S \times S$

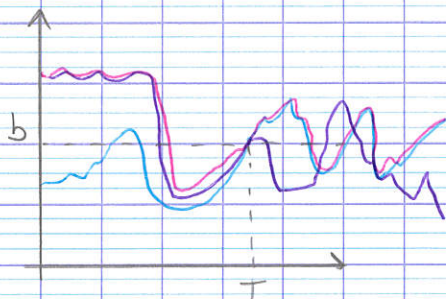
$$W_n = (X_n, Y_n)$$

La matrice de transition $\tilde{P}$ est donnée par
 $\tilde{P}((x,u),(y,v)) = p(x,y)p(u,v)$

La loi initiale λ est donnée par
 $\lambda(x,u) = \mu(x)\pi(u)$

Le temps T peut être vu comme le tps de passage en (b,b) de $(X_n)_{n \geq 0}$
 Comme $S \times S$ est fini, il suffit que $\tilde{P}$ soit irréductible pour que
 $P(T < \infty) = 1$

$\tilde{P}^{(n)}((x,u),(y,v))$
 $= p^{(n)}(x,y)p^{(n)}(u,v) > 0$
 si n assez grand car P est aperiodique
 Donc $\tilde{P}$ est irréductible.
 Donc $P(T < \infty) = 1$



Etape 2 : On définit

$$Z_n = \begin{cases} X_n & \text{si } n < T \\ Y_n & \text{si } n \geq T \end{cases}$$

Alors Z_n est Markov (μ, P)

$$P(Z_n = \infty) \\
= P(X_n = \infty, n < T) + P(Y_n = \infty, n \geq T)$$

Maintenant

$$\begin{aligned} & |P(X_n = \infty) - \pi(\infty)| \\ &= |P(X_n = \infty) - P(Y_n = \infty)| \quad \left. \begin{array}{l} \downarrow Y_n \sim \pi \\ \downarrow X_n \sim Z_n \end{array} \right\} \\ &= |P(Z_n = \infty) - P(Y_n = \infty)| \\ &= |P(X_n = \infty, n < T) + P(Y_n = \infty, n \geq T) - P(Y_n = \infty)| \quad \text{par (*)} \end{aligned}$$

Mais

$$P(Y_n = \infty) = P(Y_n = \infty, n < T) + P(Y_n = \infty, n \geq T)$$

Ainsi

$$\begin{aligned} & P(Y_n = \infty, n \geq T) - P(Y_n = \infty) \\ &= -P(Y_n = \infty, n < T) \end{aligned}$$

Donc

$$\begin{aligned} & |P(X_n = \infty) - \pi(\infty)| \\ &= |P(X_n = \infty, n < T) - P(Y_n = \infty, n < T)| \\ &\leq P(X_n = \infty, n < T) + P(Y_n = \infty, n < T) \\ &\leq P(n < T) + P(n < T) \\ &\leq 2P(T > n) \\ &\quad \xrightarrow[n \rightarrow \infty]{\rightarrow 0} \\ &\text{car } P(T < \infty) = 1 \end{aligned}$$

Remarque : La vitesse de convergence vers la loi stationnaire est de l'ordre $|S|^n$ où S est la valeur propre de P différente de 1 et de plus grand module (qui est strictement plus petit que 1).

D'après le théorème précédent et de la modélisation que nous avons adopter pour Firmin / son comportement, nous pouvons déduire à ce stade de réflexion :

" Rapidement (qq minutes), la probabilité que Firmin soit entrain de dormir chaque min est de l'ordre de 88,40 % "

Ceci ne signifie pas que Firmin darme en moyenne 88,4% du tps.

Pour une telle affirmation, il est nécessaire de disposer d'une loi des grands nbres comme dans un modèle de pile ou face.

4) Théorème ergodique.

Le théorème affirmant que

$$R_n(x) \xrightarrow[n \rightarrow \infty]{\text{Proba}} p, \quad x \in \{P, F\}$$

la proposition affirmant que

$$R_n(x) \xrightarrow[n \rightarrow \infty]{\text{Proba}} p(x), \quad x \in \{a, t, c, g\} \text{ sont des}$$

application de la loi des grands nombres pour des variables aléatoires indépendantes.

Le théorème affirmant que

$$R_n(x, y) \xrightarrow[n \rightarrow \infty]{\text{Proba}} p(x) p(y), \quad x, y \in \{a, t, c, g\} \text{ est une loi}$$

des grands nombres pour des variables aléatoires faiblement dépendantes.

Dans le cadre des chaînes de Markov, il existe aussi une loi des grands nbres sous réserve que la chaîne satisfasse de bonnes hypothèses.

Théorème: Supposons que S est fini. Soit P une matrice de transition irréductible et aperiodique. Alors,

* P possède une unique mesure de probabilité invariante λ .

* Pour toute fctⁿ bornée $f: S \rightarrow \mathbb{R}$ pour tout $\varepsilon > 0$,

nous avons:

$$P \left(\left| \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) - \int \right| \geq \varepsilon \right) \xrightarrow[n \rightarrow +\infty]{} 0$$

$$= \frac{1}{n} (f(x_0) + f(x_1) + \dots + f(x_{n-1}))$$

avec $\int = \sum_{x \in S} f(x) \pi(x)$.

Ainsi, en prenant la fonction

$$f(x) = \mathbb{1}_{\{x = 0\}}$$

on obtient, pour tout $\varepsilon > 0$

$$R_n(D) \xrightarrow[n \rightarrow +\infty]{} \pi(D) \approx 88,4\%$$

c'est à dire que la proportion de lps que Firmin passe à dormir converge vers $\pi(D) \approx 88,4\%$.

On pourrait opposer à notre raisonnement le fait que le choix de la matrice de transition P en amont rendant une telle conclusion inévitable. C'est juste mais si l'on fait uniquement l'hypothèse que le comportement de Firmin est Markovien d'ordre 1 et du théorème précédent (P irréductible et aperiodique)

Alors,

* il est possible d'estimer les paramètres de la matrice P car (théorème omis)

$$\underbrace{\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{X_i = x, X_{i+1} = y\}}}_{R_n(x,y)} \xrightarrow[n \rightarrow \infty]{} p(x,y) \pi(y)$$

et

$$\frac{1}{n} \sum_{i=0}^{n-1} R_n(\omega_i) \xrightarrow{n \rightarrow \infty} \pi(\omega)$$

Ainsi,

$$\frac{R_n(\omega, y)}{R_n(\omega)} \xrightarrow[n \rightarrow \infty]{\text{probab}} p(\omega, y)$$

TV

IV Modélisation markovienne de séquences d'ADN

Exemple: Dinucléotides CpG

CpG représente $\ll 5' - CG - 3' \gg$

Dans le génome des mammifères, les cytosines impliquées dans un ^{di}CpG sont fréquemment méthylées, et cette méthylation entraîne une hausse de mutations de CpG vers TpG.

Ces dinucléotides intéressent particulièrement les biologistes, car il existe des portions dans le génome appelées îlots CpG, qui possèdent une concentration en CpG plus élevée que dans le reste de la séquence.

Il s'avère que ces régions sont souvent des zones fonctionnelles, et l'existence de telles zones suggèrent que le phénomène de méthylation est réprimé dans les îlots CpG.

Maintenant que nous savons que les dinucléotides sont importants nous sommes intéressés par les modèles probabilistes de lesquels la probabilité d'apparition d'une lettre dépend de la précédente. C'est ainsi que les modèles Markoviens trouvent leur place en analyse de seq. biologique.

Estimation de l'ordre du modèle.

On dit qu'un processus (X_n) est une chaîne de Markov homogène d'ordre m

(Prop de Markov) pour tout $n \geq m$ et $x_1, \dots, x_{n+1} \in \mathcal{S}$

$$\begin{aligned} P(X_{n+1} = x_{n+1} | X_{1:n} = x_{1:n}) \\ = P(X_{n+1} = x_{n+1} | X_{n-m+1:n} = x_{n-m+1:n}) \end{aligned}$$

- (Homogénéité)...

$$P = (p(x, y), x \in S^m, y \in S)$$

$$p(x, y) \geq 0 \text{ et } \sum_{y \in S} p(x, y) = 1$$

Remarque: Quand $S = \mathcal{A}$ et $\mathcal{A} = \{A, T, C, G\}$

le nbre de paramètres libres pour une chaîne de Markov $(\mu, P, 1)$ est 15, 3 pour la mesure initiale et 4×3 pour la matrice de transition. Pour une chaîne de Markov (μ, P, m) ce nbre devient $(4^m - 1) + 4^m \times 3$.

On a vu que la probabilité d'observer une seq d'ADN $x_1 \dots x_N$ sous le modèle $(\mu, P, 1)$ est

$$P(X_{1:N} = x_{1:N})$$

$$= \mu(x_1) p(x_1, x_2) p(x_2, x_3) \dots p(x_{N-1}, x_N)$$

Théorème: (X_n) est Markov (μ, P, m)

Alors $\forall N \geq m$

$$P(X_{1:N} = x_{1:N})$$

$$= \mu(x_{1:m}) p(x_{1:m}, x_{m+1}) p(x_{2:m+1}, x_{m+2}) \dots p(x_{N-m+1}, x_N)$$

Exemple

seq = aa tat age

Sous Markov $(\mu, P, 1)$

$$P(X_{1:8} = \text{seq}) = \mu(a) p(a, a) p(a, t) p(t, a) p(a, t) p(t, a)$$

$$p(a, g) p(g, e)$$

Sous Markov $(\mu, P, 2)$

$$P(X_{1:8} = \text{seq}) =$$

$$\mu(a,a) p(a,a,t) p(a,t,a) p(t,a,t) p(a,t,a) p(t,a,g) p(a,g,c)$$

Sous Markov $(\mu, P, 3)$

$$P(X_1:s = seq) =$$

$$\mu(a,a,t) p(a,a,t,a) p(a,t,a,t) p(t,a,t,a) p(a,t,a,g) p(t,a,g,c)$$

Sous Markov $(\mu, P, 1)$ cette probabilité peut être réécrite

$$P(X_1:s = seq) =$$

$$\mu(a) p(a,a)^1 p(a,t)^2 p(a,g)^1 p(g,c)^1 p(t,a)^2$$

On peut voir que l'exposant correspondant à $p(a,a)$ est égal à 1 qui est exactement le nombre de fois où le dinucléotide aa apparaît dans la séquence

Idem pour $p(a,t)$

En général, pour Markov $(\mu, P, 1)$ $P(X_1:n=s) = \mu(s/1) \prod_{\substack{(x,u) \in A \times A \\ N_s(x,u) > 0}} p(x,u)^{N_s(x,u)}$

$$\text{où } N_s(x,u) = \sum_{n=1}^{N-1} 1_{\{X_n=x, X_{n+1}=u\}}$$

Étant donné une séquence s , on se demande quel modèle Markovien correspond le mieux à s , c'est à dire lequel maximise

$$P(X_1:N=s)$$

Tout d'abord, on peut fixer m et chercher l'ensemble de paramètres $\theta = (\mu, P)$ tel que

$$\ell_m(\theta | s) = \mu(s/1+m) \prod_{\substack{(x,u) \in A^m \times A \\ N_s(x,u) > 0}} p(x,u)^{N_s(x,u)}$$

est maximum.

chaîne de Markov d'ordre 1. $n=1$

Definition:

La fonction $\theta \mapsto \ell_m(\theta/s)$ est appelée la vraisemblance du modèle d'ordre m .

La théorie du maximum de vraisemblance nous dit que l'ensemble des paramètres

$\hat{\theta}_m$ qui maximise cette fonction ℓ_m pour une seq s donnée

$$\hat{\theta}_m = \underset{\theta}{\text{argument max}} \ell_m(\theta/s)$$

vaut:

$$\hat{\theta}_m = (\hat{\mu}_m, \hat{P}_m) \text{ avec } \text{estimeurs}$$

$$\hat{\mu}_m = \frac{1}{m} \sum_{x=s_i:m} \text{ et } \hat{P}_m(x,y) = \frac{N_s(x,y)}{N_s(x, \cdot)}$$

$$\forall x \in \mathcal{A}^m \text{ et } \forall y \in \mathcal{A}$$

Remarque: $\hat{\theta}_m = (\hat{\mu}_m, \hat{P}_m)$ s'appelle l'estimateur du maximum de vraisemblance du modèle Markovien d'ordre m .

$$M(\mu, P)$$

Remarque: On peut calculer explicitement l'estimateur EMV d'un modèle d'ordre m grâce au nbre d'occurrence de tous les mots d'ordre $m+1$ qui apparaissent dans la séquence s , cad grâce à $N_s(x, y) \forall x \in \mathcal{A}^m, \forall y \in \mathcal{A}$.

Propriété: on a

$$\ell_{m+1}(\hat{\theta}_{m+1}/s) \geq \ell_m(\hat{\theta}_m/s)$$

Or plus la vraisemblance est grande et mieux le modèle s'accorde à la séquence.

Mais le nbre de paramètres croît exponentiellement vite avec m .

on va donc pénaliser la vraisemblance par ce coût!

On définit ainsi un critère

BIC (Bayesian Information Criterion) pour choisir l'ordre du modèle

$$BIC(m) = -\log \ell_m(\hat{\theta}_m | s) + \log N \cdot \dim(M_m)$$

$$= BIC - \log \ell_m(\hat{\theta}_m | s) + 4^m \times 3 \times \log N$$

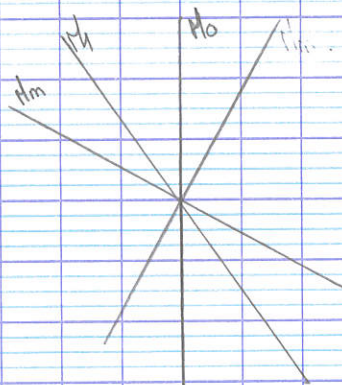
$$\dim(M_m)? P_{\frac{a}{4} \times \frac{b}{4}}^{\frac{a}{4} \times \frac{b}{4}} \rightarrow 4 \times 3$$

$$(m=1) \rightarrow$$

Modèle M_m

$$P_{\frac{a}{4} \times \frac{b}{4}}^{\frac{a}{4} \times \frac{b}{4}} = \begin{pmatrix} a & b & c & d \\ 4^n \times 4 \end{pmatrix}$$

$\rightarrow 4^m \times 3$ paramètres indépendants



Remarque on désigne la log-vraisemblance du modèle

$$\begin{aligned} \ell_m &= \log \ell_m \\ &= \log \left(\prod_{x \in \mathcal{X}} p(x, y) \right)^{N_s(x, y)} \end{aligned}$$

Et pour le BIC, on calcule la log-vraisemblance estimée du modèle

$$\ell_m(\hat{\theta}_m) = \log \left(\prod_{x, y} \hat{p}(x, y) \right)^{N_s(x, y)}$$

$$= \sum_{x,y} N_s(x,y) \log \hat{p}(x,y)$$

$$L_m(\hat{\theta}_m) = \sum_{x,y \in \mathcal{A}^m \times \mathcal{A}} N_s(x,y) \log \frac{N_s(x,y)}{N_s(x)}$$

on voit qu'elle se calcule explicitement avec les nbres d'occurrence des mots de longueur $m+1$ de la séquence s .

on choisit le critère BIC qui est minimum.